Published online 12 April 2018

Nucleic Acids Research, 2018, Vol. 46, No. 9 ***4807–4818***
doi: 10.1093/nar/gky261

Crystal structure of the novel lesion-specific endonuclease PfuEndoQ from *Pyrococcus furiosus*

Ken-ichi Miyazono[1], Sonoko Ishino[2], Naruto Makita[2], Tomoko Ito[1], Yoshizumi Ishino[2] and Masaru Tanokura[1,*]

[1]Laboratory of Basic Science on Healthy Longevity, Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan and [2]Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, and Faculty of Agriculture, Kyushu University, Fukuoka 812-8581, Japan

Received February 13, 2018; Revised March 25, 2018; Editorial Decision March 26, 2018; Accepted March 28, 2018

ABSTRACT

Because base deaminations, which are promoted by high temperature, ionizing radiation, aerobic respiration and nitrosative stress, produce mutations during replication, deaminated bases must be repaired quickly to maintain genome integrity. Recently, we identified a novel lesion-specific endonuclease, PfuEndoQ, from *Pyrococcus furiosus*, and PfuEndoQ may be involved in the DNA repair pathway in *Thermococcales* of Archaea. PfuEndoQ recognizes a deaminated base and cleaves the phosphodiester bond 5′ of the lesion site. To elucidate the structural basis of the substrate recognition and DNA cleavage mechanisms of PfuEndoQ, we determined the structure of PfuEndoQ using X-ray crystallography. The PfuEndoQ structure and the accompanying biochemical data suggest that PfuEndoQ recognizes a deaminated base using a highly conserved pocket adjacent to a Zn^{2+}-binding site and hydrolyses a phosphodiester bond using two Zn^{2+} ions. The PfuEndoQ-DNA complex is stabilized by a Zn-binding domain and a C-terminal helical domain, and the complex may recruit downstream proteins in the DNA repair pathway.

INTRODUCTION

The DNA of all organisms is constantly damaged by exogenous and endogenous agents. DNA lesions, such as base or sugar modifications and single- or double-strand breaks, must be repaired quickly because they cause mutations and result in genome instability, cellular senescence or cell death. Organisms have evolved a variety of DNA repair systems (1,2) such as homologous recombination repair (HR) (3), mismatch repair (MMR) (4), base excision repair (BER) (5), nucleotide excision repair (NER) (6) and alternative excision repair (AER) (7) that help maintain cellular functions. The NER and AER pathways are initiated by phosphodiester bond cleavages near lesion sites. While the NER pathway is initiated by dual incisions on both sides of a lesion site, the AER pathway is initiated by a single nick near the lesion site. The AER pathway starts with an endonuclease that recognizes a DNA lesion and that cleaves a phosphodiester bond near the site.

Base deamination is one of the most common types of DNA damage. Deaminations of adenine, guanine and cytosine produce hypoxanthine, xanthine and uracil, respectively. A base deamination occurs spontaneously under physiological conditions and is accelerated by high temperature, ionizing radiation, aerobic respiration and nitrosative stress that is generated by nitrate or nitrite metabolism (8). Because the hydrogen-bonding properties of bases are altered by the amino-keto conversion, base deaminations have strong miscoding properties. For example, a hypoxanthine mispairing with cytosine leads to an A-T to G-C transition mutation during replication (9). Therefore, deaminated sites must be repaired quickly to maintain genomic integrity (Figure 1A).

Because Archaea, the third domain of life, possess DNA repair proteins that are similar to those of eukaryotes, structural and biochemical analyses of archaeal DNA repair proteins have revealed a variety of insights into the eukaryotic DNA repair proteins (10–12). In addition to these proteins with homologous sequences to those of eukaryotes, hyperthermophilic archaea are predicted to possess highly efficient and unique DNA repair systems because they exist in an inhospitable environment; they are always threatened by high temperatures, ionizing radiation and DNA-damaging chemical agents (13,14). Base deaminations are promoted by these environmental stresses, and they need to be quickly repaired by the BER pathway, which is initiated by DNA glycosylases (15), and/or the AER pathway, which is initiated by lesion-specific endonucleases (7). Endonuclease V (EndoV) is a well-characterized lesion-

*To whom correspondence should be addressed. Tel: +81 3 5841 5165; Fax: +81 3 5841 8023; Email: amtanok@mail.ecc.u-tokyo.ac.jp

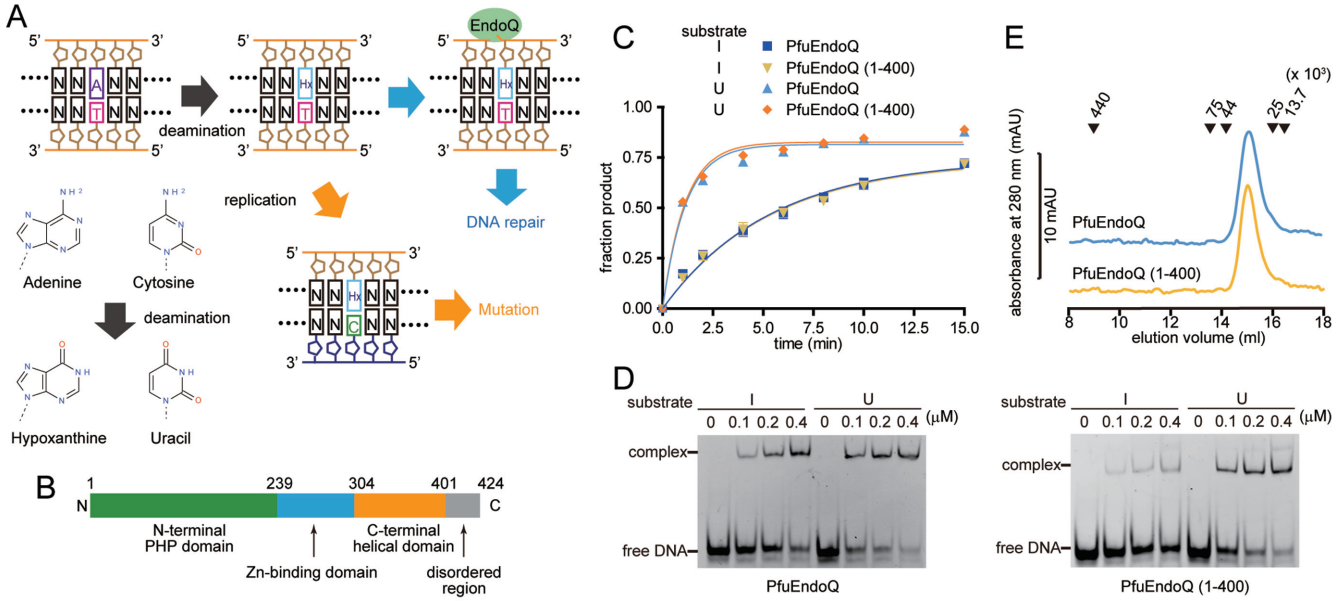

Figure 1. Characterization of PfuEndoQ. (**A**) PfuEndoQ recognizes a deaminated base in DNA and cleaves its 5′ phosphodiester bond. (**B**) Domain structure of PfuEndoQ. PfuEndoQ contains the N-terminal PHP domain, the Zn-binding domain, the C-terminal helical domain and the disordered region. (**C**) DNA cleavage activities of PfuEndoQ and PfuEndoQ (1-400) for dI- or dU-containing substrates (Supplementary Figure S2). 100 nM of the substrate dsDNA and 100 nM of the enzyme were mixed and incubated at 40°C. Plotted values are mean ± SEM ($n = 3$). (**D**) EMSA of PfuEndoQ and PfuEndoQ (1–400) using the dI- and dU-containing probes. 100 nM of the fluorescein-labelled dsDNA and each concentration of PfuEndoQ were mixed and separated using a 10% polyacrylamide gel. (**E**) Gel filtration analysis of PfuEndoQ ($M_r = 47\,652$) and PfuEndoQ (1–400) ($M_r = 44\,916$). The peak positions of the marker proteins are indicated by the black triangles at the top of the chromatogram.

specific endonuclease and is conserved in all three domains of life from bacteria, archaea to human (9,16–20). EndoV from *Pyrococcus furiosus* (PfuEndoV) is probably an endonuclease that is part of the AER pathway, but it is interesting that PfuEndoV specifically recognizes hypoxanthine but not uracil and xanthine and that it cleaves the second phosphodiester bond 3′ of the lesion site using a Mg^{2+} ion (20). In addition to EndoV, we have identified a novel lesion-specific endonuclease EndoQ from *P. furiosus* (PfuEndoQ) that does not share any amino-acid sequence similarity with EndoV (21). The homologues of PfuEndoQ are found only in the thermococcales and some of the methanogens in Archaea and are not present in most members of the Bacteria and Eukarya domains (21). A limited number of bacteria have EndoQ homologs (Supplementary Figure S1), which belong to different phylogenetic tree clusters than the archaeal EndoQ (22). The bacterial EndoQ from *Bacillus pumilus and B. subtilis* were shown to have endonuclease activities similar to PfuEndoQ, and they are predicted to be involved in the AER pathway (22). Although both PfuEndoV and PfuEndoQ are lesion-specific endonucleases, their substrate specificities and cleavage sites are different. PfuEndoQ recognizes uracil, xanthine and an apurinic/apyrimidinic (AP) site, in addition to a hypoxanthine, and cleaves the phosphodiester bond 5′ from the lesion site (21,22). The amino acid sequence of PfuEndoQ suggested that PfuEndoQ possesses an N-terminal polymerase and histidinol phosphatase (PHP) domain, a zinc-binding domain and an uncharacterized C-terminal domain. However, the overall structure of PfuEndoQ and the structural basis for its substrate recognition and DNA cleavage mechanisms remained unclear.

Here, we report the structure of PfuEndoQ determined using X-ray crystallography. Based on the structure and on accompanying biochemical data, we predict the mechanism by which PfuEndoQ recognizes a deaminated base in double-stranded DNA (dsDNA) and cleaves the phosphodiester bond 5′ from the lesion site. Our results suggest that PfuEndoQ recognizes a deaminated base using a highly conserved small pocket adjacent to an active site in the N-terminal PHP domain and the PfuEndoQ–dsDNA complex is stabilized using the Zn-binding domain and the C-terminal helical domain. A downstream enzyme involved in the AER pathway might be recruited to the appropriate site by the DNA-bound PfuEndoQ.

MATERIALS AND METHODS

Expression and purification

The gene fragment of PfuEndoQ (Gene ID: 1469427) was amplified by PCR and cloned into the NdeI and NotI sites of the pET21a vector (21). Ile2 of PfuEndoQ, which facilitates degradation in *Escherichia coli*, was mutated to Val2. To obtain a C-terminal deletion construct of PfuEndoQ (PfuEndoQ (1–400)), a stop codon was inserted into the constructed vector using the PrimeSTAR Mutagenesis Basal kit (TAKARA). The *E. coli* Rosetta (DE3) strain was transformed with the vector, and the cells were cultivated in LB medium supplemented with 50 mg/L ampicillin and 25 mg/l chloramphenicol at 37°C until the optical density of the medium at 600 nm reached 0.3. Protein ex-

pression was induced by the addition of 0.1 mM (final concentration) isopropyl β-D-1-thiogalactopyranoside (IPTG), and the medium was further cultivated at 18°C overnight. The harvested cells were resuspended in a buffer containing 50 mM MES (pH 6.0), 50 mM NaCl and 10 mM $MgCl_2$ and were disrupted by sonication. The supernatant after centrifugation at 40 000 × g for 30 min was treated with Cryonase Cold-active Nuclease (TAKARA) at room temperature for 30 min. The treated solution was incubated at 80°C for 30 min to denature the proteins from *E. coli* and was centrifuged at 40 000 × g for 10 min to remove denatured protein aggregates. The supernatant was purified with TOYOPEARL AF-Heparin (TOSOH) resin and a Mono S 10/10 (GE Healthcare) column (for crystallization) or by ammonium sulfate precipitation (80% saturation) and a MonoS 10/10 column (for in vitro assays). The purified protein solutions were dialysed against 10 mM MES (pH 6.0), 100 mM NaCl, 10 mM $MgCl_2$ and 5% glycerol and were stored at −80°C until use.

The PfuEndoQ mutants, except for the Tyr244 and Tyr377 mutants, were prepared by modifying the PfuEndoQ (1–400) expression vector using the PrimeSTAR Mutagenesis Basal kit (TAKARA). The expression and purification of these mutants were performed using the same method used for PfuEndoQ (1–400). The expression and purification of Tyr244 and Tyr377 mutants were performed in basically the same manner as previously described (21).

Endonuclease activity assay

Oligonucleotides were purchased from Eurofins Genomics and dissolved in a solution containing 50 mM Tris–HCl (pH 8.0), 1 mM $MgCl_2$ and 0.01% Tween20 to a final concentration of 100 μM. ssDNA samples were annealed by incubating at 95°C and slow cooling to 4°C. The 5′-fluorescein-labeled 29-bp dsDNAs (5′- fluorescein -GGAATCCTGACGACITGTAGCGAACGATC-3′ and 5′-fluorescein- GGAATCCTGACGACUTGTAGCGAACGATC-3′, Supplementary Figure S2AB) were used as substrates. The dU-containing substrate was treated with uracil–DNA glycosylase (New England Biolabs) to produce the AP site-containing substrate (5′-fluorescein-GGAATCCTGACGAC-AP site-TGTAGCGAACGATC-3′, Supplementary Figure S2C). The AP site-containing substrate was purified using a QIAquick PCR Purification Kit (QIAGEN). The AP site-containing substrate was treated with 0.1 M NaOH, incubated at 70°C for 10 min, and neutralized by the addition of an equal amount of HCl to confirm that the dU site was completely cleaved by the uracil–DNA glycosylase treatment. Then, 100 nM PfuEndoQ and its mutants, except for the Tyr244 and Tyr377 mutants, were mixed with 100 nM substrate dsDNA in a reaction buffer containing 50 mM Tris–HCl (pH 8.0), 1 mM $MgCl_2$ and 0.01% Tween20. For the time course analyses of PfuEndoQ and PfuEndoQ (1–400), the reaction solutions were incubated at 40°C for 1, 2, 4, 6, 8, 10 and 15 min. For the mutation assay of PfuEndoQ (1–400) and its mutants, except for the Tyr244 and Tyr377 mutants, the reaction solutions were incubated at 40°C for 15 min. For the endonuclease activity assay of PfuEndoQ in the presence of ethylenediaminetetraacetic acid (EDTA), 100 nM PfuEndoQ was mixed with 100 nM substrate dsDNA in a reaction buffer containing 50 mM Tris–HCl (pH 8.0), 0.01% Tween20 and 0, 10, 25 or 50 mM EDTA, and the reaction solutions were incubated at 40°C for 15 min. After the reactions, the solutions were supplemented with an equal amount of a solution (10 M urea and 0.2 M EDTA) to stop the reactions. The solutions were separated using a denaturing 18% polyacrylamide gel in 0.5× TBE and 7 M urea. The fluorescence was measured using an LAS4000 Mini system (GE Healthcare). The enzymatic rate constant, k, was obtained from a single-exponential fit to the data from three independent measurements: $f_p = f_p\max \times (1 - e^{-kt})$, in which f_p is the fraction of product, $f_p\max$ is the maximum value of f_p and t is the time of the reaction.

For the endonuclease activity assay of the Tyr244 and Tyr377 mutants, 10 nM 5′-Cy5-labeled 45 bp dsDNA containing dI (5′-dCGAACTGCCTGGAATC CTGACGACITGTAGCGAACGATCACCTCA-3′ and 5′-dTGAGGTGATCGTTCGCTACATGTCGTCAGG ATTCCAGGCAGTTCG-3′) was incubated with 2 nM protein at 75°C for 1.5, 2.5, 5 and 7.5 min. The reactions were terminated with a half volume of stop buffer (98% formamide, 10 mM EDTA, and 0.1% OrangeG). The samples were separated using a denaturing 12% polyacrylamide gel in 1× TBE and 8 M urea. The gel image was visualized by using an image analyser, Typhoon Trio+ (GE Healthcare).

Electrophoretic mobility shift assay

The 5′-fluorescein-labelled 29 bp dsDNAs (5′-fluorescein-GGAATCCTGACGACITGTAGCGAACGATC-3′ and 5′-fluorescein-GGAATCCTGACGACUTGTAGCG AACGATC-3′, Supplementary Figure S2AB) were used as probes. 100 nM of the DNA probe and PfuEndoQ (PfuEndoQ (1–400) and its K320A mutant; 100, 200 and 400 nM) were mixed in 50 mM Tris–HCl (pH 8.0), 1 mM $MgCl_2$, 0.01% Tween20 and 10 μM 29-bp dsDNA (5′-GGAATCCTGACGACATGTAGCGAACGATC-3′). To reduce nonspecific PfuEndoQ–dsDNA aggregates, we added an abundance of the non-labelled dsDNA as a competitor. The mixtures were incubated at 40°C for 30 min. Bound and unbound dsDNA were separated using a 10% polyacrylamide gel in 0.5× TBE at 4°C. The fluorescence was measured using an LAS4000 Mini system (GE Healthcare).

Oligomeric state analysis by gel filtration chromatography

The purified proteins were loaded onto a Superdex 200 HR 10/30 (GE Healthcare) column and eluted with a buffer containing 10 mM Tris–HCl (pH 8.0) and 200 mM NaCl. To estimate the multimerization state of PfuEndoQ, the following standard proteins were used: thyroglobulin (M_r = 669 000), conalbumin (M_r = 75 000), ovalbumin (M_r = 44 000), chymotrypsinogen A (M_r = 25 000) and ribonuclease A (M_r = 13 700).

Crystallization, data collection and structure determination

The purified PfuEndoQ (1–400) was concentrated to 9.5 mg/mL. Crystallizations of PfuEndoQ (1–400) were performed by the sitting-drop vapour-diffusion method at 4°C.

Each drop was prepared by mixing 2 μl of protein solution and 1 μl of reservoir solution. Crystals of PfuEndoQ (1–400) were obtained with a reservoir solution condition containing 0.1 M Tris–HCl (pH 7.0–7.3), 18–13% PEG3000 and 0.2 M calcium acetate within a few days. The crystals of PfuEndoQ (1–400) were soaked in a reservoir solution supplemented with 2.5 mM $Sm(NO)_3$ for one day to determine the structure by the single-wavelength anomalous dispersion (SAD) method.

X-ray diffraction data sets of the samarium derivative crystal of PfuEndoQ (1–400) was collected in-house using a FR-E SuperBright and an R-AXIS VII (Rigaku). X-ray diffraction data sets of native crystals using X-ray wavelengths below and above the zinc K absorption edge were collected at BL-17A beamline at Photon Factory (Tsukuba, Japan). The crystals were transferred into a reservoir solution supplemented with 30% glycerol before being picked up and flash cooled in a nitrogen stream (93 K (in-house) or 95 K (BL-17A)). The diffraction data sets were indexed, integrated and scaled with XDS (23). The samarium derivative crystal of PfuEndoQ (1–400) belonged to the space group $C2$ with unit cell parameters of $a = 257.35$ Å, $b = 82.25$ Å, $c = 116.58$ Å and $\beta = 109.13°$.

The crystal structure of PfuEndoQ (1–400) was determined by the SAD method using the diffraction data set of the samarium derivative crystal. The samarium substructure and the initial phase were determined with the programs SHELEXC/D (24) and SHARP (25) using the autoSHARP interface (26). The initial phase was improved using the program parrot in the CCP4 suite (27). The initial model of PfuEndoQ (1–400) were built using the program Buccaneer (28). The initial model was refined and rebuilt using the programs phenix.refine (29) and Coot (30). The final model of PfuEndoQ (1–400) was refined to 2.50 Å resolution, with R and R_{free} values of 17.6% and 20.9%, respectively. The geometry of the final model was evaluated with the program MolProbity (31). In the Ramachandran plot, 98.0% of the residues were included in the favoured region, and 99.9% of the residues were included in the allowed region. The data collection, phasing and refinement statistics are summarized in Table 1.

Computational analysis

Structural analysis was carried out using a set of computer programs: DISOPRED2 for the prediction of disordered regions (32), Dali for the search for similar structures from the database (33), Clustal Omega for the amino acid sequence alignment (34), ESpript for the preparation of alignment figures (35), ConSurf for the mapping of the sequence conservation to the protein surface (36), APBS for the calculation of macromolecular electrostatics (37) and Pymol (https://www.pymol.org/) for the depiction of structures.

RESULTS

Structural determination of PfuEndoQ

The full-length PfuEndoQ construct with 424 amino acids did not yield crystals suitable for X-ray diffraction. To determine the PfuEndoQ structure, we prepared a truncated mutant of PfuEndoQ (residues 1–400, Figure 1B), which lacked the C-terminal region from Gly401 to Asn424. This region is predicted to be disordered based on the amino acid sequence (32) and is not conserved among the homologues (Supplementary Figure S1), although this region is utilized for interaction with proliferating cell nuclear antigen (PCNA) (38). In addition, the truncated PfuEndoQ (1–400) showed approximately the same DNA cleavage activity as wild-type PfuEndoQ (Figure 1C and Supplementary Figure S2). The enzymatic rate constants k_{obs} of PfuEndoQ for the dI- and dU-containing substrates were calculated to be 0.19 ± 0.02 min^{-1} and 0.88 ± 0.07 min^{-1}, respectively (mean ± SEM, $n = 3$). The enzymatic rate constants, k_{obs}, of PfuEndoQ (1–400) for the dI- and dU-containing substrates were calculated to be 0.19 ± 0.02 min^{-1} and 0.90 ± 0.06 min^{-1}, respectively (mean ± SEM, $n = 3$). PfuEndoQ cleaves the dU-containing substrate more efficiently than the dI-containing substrate. The electrophoretic mobility shift assay (EMSA) showed that wild-type PfuEndoQ and PfuEndoQ (1–400) bind the dU-containing dsDNA more tightly than the dI-containing dsDNA (Figure 1D). Gel-filtration analysis showed that both the wild-type PfuEndoQ and PfuEndoQ (1–400) form monomers in solution (Figure 1E). These results indicated that truncation of the C-terminal disordered region does not alter the function, conformation and multimerization state of PfuEndoQ.

The PfuEndoQ structure was determined by the SAD method using a samarium derivative crystal. Although we also collected X-ray diffraction data sets of the native crystals, we determined the final structure using the dataset from the samarium derivative crystal because the resulting structure showed the best refinement statistics and the PfuEndoQ structure that was calculated using the dataset of native crystal was partially disordered. Samarium soaking to the crystal might improve the crystal packing of PfuEndoQ. The PfuEndoQ crystal contains five PfuEndoQ monomers in the asymmetric unit (chains A–E). Due to the poor electron density, Ile399 and Ser400 of chain B and Ser400 of chains C and D are not built in the final structure. The monomeric structures of PfuEndoQ are nearly identical. The maximal root-mean-square deviation (r.m.s.d.) between the monomers is 0.854 Å for 398 superposed Cα atoms. We used the chain A structure of PfuEndoQ, which possesses the lowest B-factor, for the images, unless otherwise stated.

Overall structure of PfuEndoQ

The PfuEndoQ structure comprises 16 β strands, 16 α helices and 4 3_{10} (η) helices (Figure 2A and Supplementary Figure S3A). The overall structure of PfuEndoQ is roughly divided into an N-terminal PHP domain (residues 1–238), a Zn-binding domain (239–303) and a C-terminal helical domain (304–400). At the N-terminal PHP domain, 11 β strands (β1–β11) form a barrel-like structure and are surrounded by 10 helices (α1-α8 and η1-η2). The Zn-binding domain consists of 3 β strands (β12-β14) and 3 helices (α9, α10 and η3). A C4-type Zn-binding structure (Cys249, Cys252, Cys268 and Cys271) is formed between the β12 and β13 strands. The β14 strand forms a β sheet with the barrel-like β sheet of the N-terminal PHP domain. The C-terminal helical domain consists of a helix cluster (α11, α12, α13,

Table 1. Data collection, phasing and refinement statistics of PfuEndoQ

Crystal name		Native		Sm derivative
		Crystal 1	Crystal 2	
Data collection				
Beamline		BL-17A		In-house
Space group		*C*2	*C*2	*C*2
Cell dimensions	*a* (Å)	258.87	259.93	257.35
	b (Å)	82.53	82.93	82.25
	c (Å)	116.00	116.34	116.58
	β (°)	109.11	109.21	109.13
Wavelength (Å)		1.28255	1.29083	1.5418
Resolution (Å)		47–2.80 (2.88–2.80)*	47–2.80 (2.88–2.80)	43–2.50 (2.55–2.50)
R_{sym} (%)		6.9 (45.1)	4.2 (29.6)	7.4 (49.4)
$I/\sigma I$		15.2 (2.7)	24.6 (4.2)	32.7 (4.5)
Completeness (%)		98.9 (89.7)	99.0 (89.5)	98.1 (80.6)
Multiplicity		6.4 (5.0)	6.5 (5.1)	13.7 (9.4)
Phasing				
Phasing power				0.744
FOM before/after density modification				0.281/0.733
Refinement				
R/R_{free} (%)				17.6/20.9
No. atoms				
Protein/Zn^{2+} ion/Sm^{3+} ion/water				15806/15/7/415
B-factors (Å^2)				
Protein/Zn^{2+} ion/Sm^{3+} ion/water				44.9/36.2/109.9/39.7
r.m.s deviations				
Bond lengths (Å)				0.002
Bond angles (°)				0.514
Ramachandran plot				
Favoured region (%)				98.0
Allowed region (%)				99.9

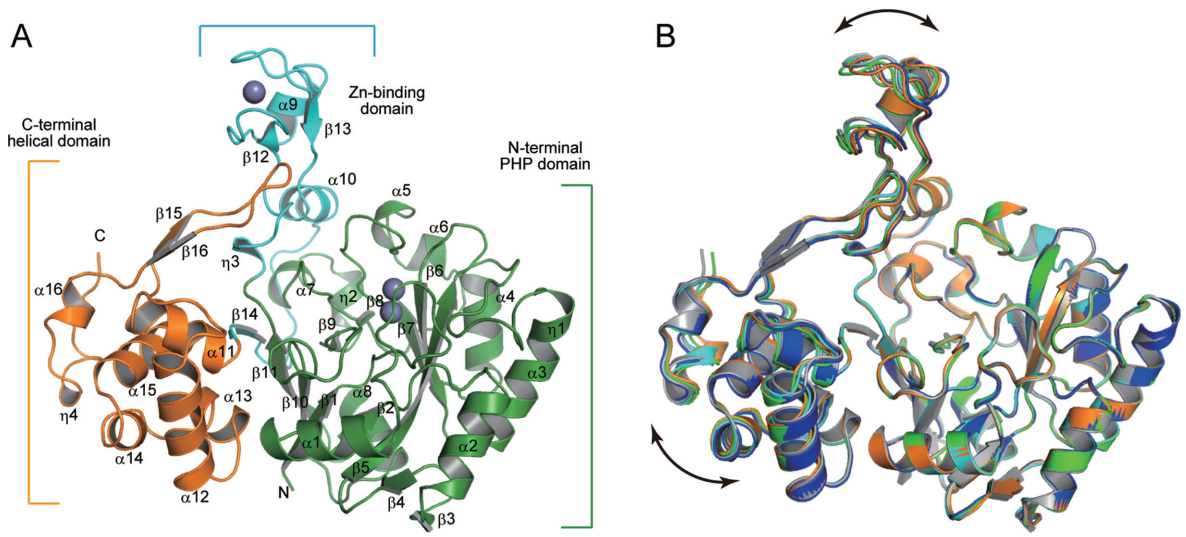

Figure 2. Overall structure of PfuEndoQ. (**A**) Ribbon diagram of PfuEndoQ. Secondary structure assignments are labelled on the ribbon model. The bound Zn^{2+} ions are shown as grey spheres. The N-terminal PHP domain, the Zn-binding domain and the C-terminal helical domain are coloured green, cyan and orange, respectively. (**B**) Superposition of the PfuEndoQ monomers. The five PfuEndoQ structures in the symmetric unit are coloured light green (chain A), blue (chain B), cyan (chain C), orange (chain D) and gray (chain E).

α14, α15, α16 and η4) and a β sheet (β15 and β16) that is sticking out towards the Zn-binding domain. Because the highly conserved β15-β16 loop is sandwiched between the β12 strand and the η3 helix at the stem of the Zn-binding region, this loop is predicted to support the conformational rigidity of the Zn-binding domain. When the PfuEndoQ structures in the asymmetric unit were superposed using their N-terminal PHP domains, the Zn-binding domain and the C-terminal helical domain were relatively flexible to the N-terminal PHP domain (Figure 2B). The relative positions of the three domains might be modified by the binding of a substrate dsDNA.

The electron density map of PfuEndoQ showed that PfuEndoQ binds three metal ions at its surface; two of them

(Zn1 and Zn2) are at the N-terminal PHP domain and one (Zn3) is at the Zn-binding domain. Because PfuEndoQ requires Mg^{2+} for its tight dsDNA binding (21), two metal ions at the N-terminal PHP domain were predicted to be Mg^{2+}. However, the electron density of the metal ions was much higher than that of Mg^{2+}. This indicated that the bound metal ions are not Mg^{2+} but are heavier metal ions. An X-ray fluorescence excitation and wavelength scan suggested that the bound metal ions are Zn^{2+}. The anomalous difference Fourier map using data collected above and below the zinc K absorption edge showed that the three metal ions that bind to PfuEndoQ are Zn^{2+} (Figure 3A), although we did not add Zn^{2+} ions to growth media utilized for the PfuEndoQ overexpression and to buffers utilized for the PfuEndoQ purification. The *B*-factors of Zn^{2+} ions that are calculated assuming that all Zn^{2+}-binding sites are fully occupied (i.e. occupancy of Zn^{2+} ions = 1) are nearly identical with those of Zn^{2+}-coordinating atoms (Supplementary Figure S4A), suggesting that all Zn^{2+}-binding sites are fully occupied. Because PfuEndoQ retained the endonuclease activity in the presence of low concentrations of EDTA (Supplementary Figure S2E), PfuEndoQ was predicted to cleave DNA using tightly bound Zn^{2+} ions, although PfuEndoQ can also utilize Mn^{2+}, Mg^{2+} and Ca^{2+} ions for DNA cleavage (39). At the N-terminal PHP domain, two Zn^{2+} are located at one side of the β-barrel-like structure and are coordinated by His8, His10, Glu76, His84, His139 and Asp193 (Figure 3BC and Supplementary Figure S3B). These metal coordinating residues are completely conserved among the PfuEndoQ homologues (Supplementary Figure S1) (22). This metal binding site is predicted to be a catalytic site of PfuEndoQ. At the Zn-binding domain, four cysteine residues coordinate one Zn^{2+} (Figure 3DE and Supplementary Figure S3C) and form the C4 type Zn-binding structure. Because the cysteine residues in this region are also highly conserved among the PfuEndoQ homologues (Supplementary Figure S1), the structure of the Zn-binding domain that is stabilized by the binding of Zn^{2+} is also predicted to be important for the enzymatic activity of PfuEndoQ.

Structural comparison

To predict the structural basis for DNA cleavage of PfuEndoQ, we compared the PfuEndoQ structure with preexisting protein structures. A database search using the Dali server (33) showed that the N-terminal PHP domain of PfuEndoQ showed low structural similarity to the putative metal-dependent phosphoesterase from *Bifidobacterium adolescentis* (BadPE, Protein Data Bank (PDB) code: 3O0F, *Z*-score = 15.6, r.m.s.d. = 2.9 Å, sequence identity = 20%) that possesses a PHP fold at its active site (Figure 4A and Supplementary Figure S3D) (40). In addition, the PfuEndoQ showed structural similarity to the DNA polymerase III α subunit from *Thermus aquaticus* (PDB code: 2HPI, *Z*-score = 12.3, r.m.s.d. = 4.9 Å, sequence identity = 11%) (41). BadPE has neither DNA polymerase activity nor DNA-proofreading activity (40). However, BadPE hydrolyses the phosphoester bond of *para*-nitrophenyl phosphate. In contrast, the Zn-binding domain and the C-terminal helical domain showed no significant structural similarity to other protein structures. Overall, PfuEndoQ possesses a quite unusual structure to recognize and cleave DNA.

When the structure of PfuEndoQ was compared to that of BadPE, the Zn^{2+} binding site of PfuEndoQ superimposed well with the metal-binding active site of BadPE (Figure 4B and Supplementary Figure S3E). Around the Zn^{2+}-binding site of PfuEndoQ, the Zn^{2+}-coordinating residues of PfuEndoQ, His8, His10, Glu76, His84, His139 and Asp193, superposed well with His17, His19, Glu74, His85, His202 and Asp260 of BadPE, respectively. However, BadPE coordinates iron ions, rather than zinc ions, at the same site. In addition, BadPE binds a zinc ion near the two iron ions using Asp24, His49 and His262 (40). Although His262 of BadPE is conserved in PfuEndoQ as His195, Asp24 and His49 of BadPE are not conserved in PfuEndoQ. Instead, these residues are modified to Ser14 and Arg114, respectively. This structural difference suggests that PfuEndoQ does not bind a third metal ion at the same site. The structure of BadPE in complex with an AMP and a phosphate (40) suggests that the Zn^{2+}-binding site of PfuEndoQ hydrolyses the phosphodiester bond of DNA.

DNA binding model of PfuEndoQ

The structural comparison with BadPE suggests that PfuEndoQ cleaves a phosphodiester bond using its highly conserved Zn^{2+}-binding site in the N-terminal PHP domain (Supplementary Figure S1). In addition to the Zn^{2+}-binding site, the protein surface of the Zn-binding domain and the C-terminal helical domain adjacent to the Zn^{2+}-binding site in the N-terminal PHP domain show relatively high sequence conservation (Figure 5A). These two regions possess positively charged protein surfaces (Figure 5B). In contrast, the backside of PfuEndoQ shows low amino acid sequence conservation and possesses negatively charged protein surfaces. These observations suggest that a deaminated base of dsDNA is recognized near the Zn^{2+}-binding site of the N-terminal PHP domain and the two positively charged protein surfaces of the Zn-binding domain and the C-terminal helical domain interact with the negatively charged DNA backbone (Figure 5C).

Because PfuEndoQ cleaves the phosphodiester bond 5′ from the lesion site, PfuEndoQ are predicted to recognize a deaminated base adjacent to its active site. EndoV and DNA glycosylases, which initiate the AER and BER pathways, respectively, flip damaged bases out of the DNA helix, and the flipped bases are recognized by small pockets adjacent to their active sites. Therefore, PfuEndoQ is also predicted to possess a deaminated base recognition pocket adjacent to the Zn^{2+}-binding site of the N-terminal PHP domain. The PfuEndoQ structure shows that there is a small pocket adjacent to the Zn^{2+}-binding site of the N-terminal PHP domain (Figure 5D). This pocket is surrounded by the side chain atoms of His139, Thr142, Glu167, Leu170 and Asp193, and by the main chain atoms of Leu168, Gly169 and Asn191. These residues are highly conserved among the homologues (Supplementary Figure S1). To discriminate a damaged base from other undamaged DNA bases, EndoV utilizes a rigid pocket for the lesion recognition and forms a tight interaction with damaged bases using their

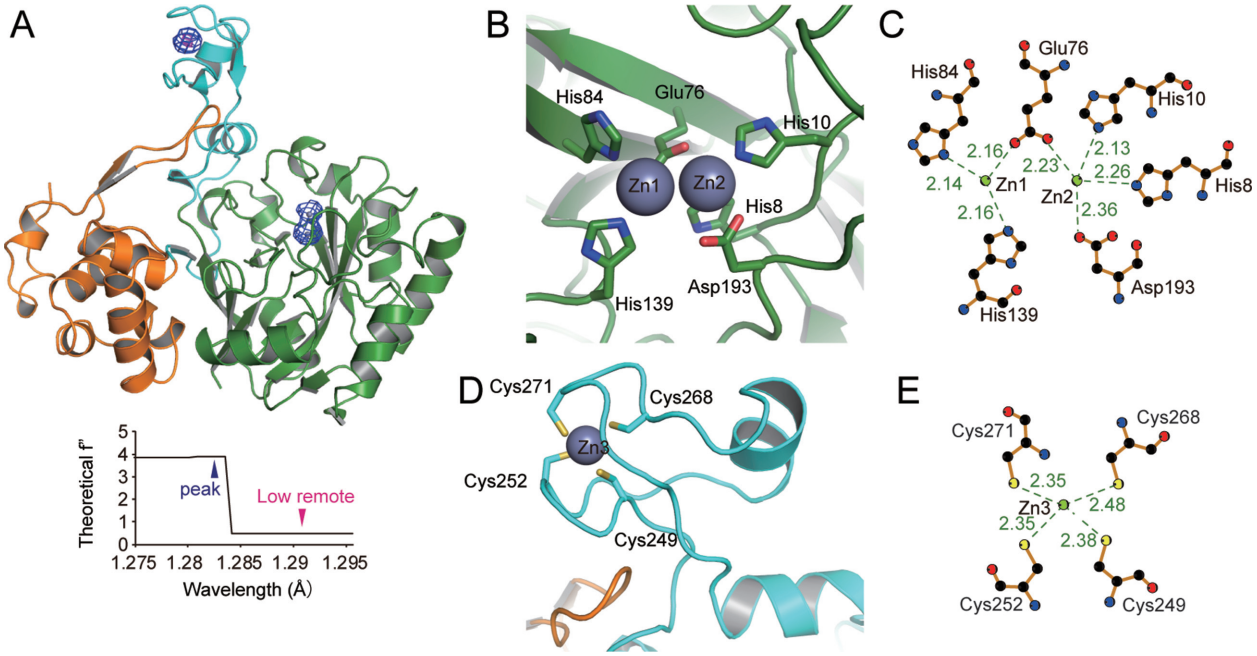

Figure 3. Metal binding sites of PfuEndoQ. (**A**) Anomalous difference Fourier electron density maps of PfuEndoQ collected at X-ray wavelengths above and below the zinc K absorption edge. Blue (peak) and magenta (low remote) meshes show anomalous difference density contoured at 5σ. The theoretical f" for zinc is plotted below the PfuEndoQ structure. (**B**) Zn^{2+}-binding site of the N-terminal PHP domain. The Zn^{2+}-coordinating residues are shown as sticks. The Zn^{2+} ions are shown as gray spheres. (**C**) Schematic diagram of the Zn^{2+}-binding mechanism at the N-terminal PHP domain. (**D**) Zn^{2+}-binding site of the Zn-binding domain. (**E**) Schematic diagram of the Zn^{2+}-binding mechanism at the Zn-binding domain.

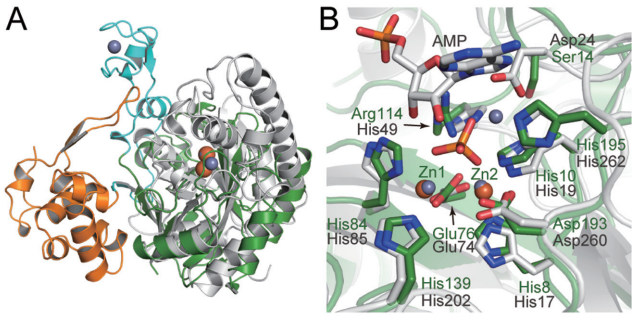

Figure 4. Structure comparison. (**A**) Superposition of the structures of PfuEndoQ and the metal-dependent phosphoesterase from *Bifidobacterium adolescentis* (BadPE). The N-terminal PHP domain, the Zn-binding domain and the C-terminal helical domain of PfuEndoQ are coloured green, cyan and orange, respectively. BadPE is coloured gray. (**B**) Close-up view of the metal binding sites. Residues from PfuEndoQ and BadPE are labeled green and gray, respectively. The AMP and phosphate ion bound to BadPE are shown as sticks.

main chain atoms (18). The PfuEndoQ structure shows that the pocket adjacent to the Zn^{2+}-binding region possesses a lower temperature factor than the other parts of PfuEndoQ, indicating that the pocket is rigid, similar to the hypoxanthine binding pocket of EndoV (Supplementary Figure S4B). PfuEndoQ is predicted to recognize a deaminated base using this pocket.

DNA glycosylases, DNA methylases and EndoV insert amino acid residues into DNA duplexes to flip out target DNA bases and fill the resulting space to stabilize protein-dsDNA complexes (18,42,43). For example, *Thermotoga maritima* EndoV utilizes the highly conserved wedge (the PYIP motif) to divide the DNA duplex adjacent to the lesion site, and Trp80 and Pro82 of the wedge stack with DNA bases before and after the lesion site (18). In the PfuEndoQ structure, the relatively conserved β7-α5 loop locates adjacent to the putative deaminated base-binding pocket. Trp144 of this loop might be inserted into the DNA duplex and stack with DNA bases as a substitute for a deaminated base (Figure 5E).

Mutation analysis

To investigate the importance of the residues at the active site of PfuEndoQ, we created point mutants and analysed their endonuclease activities. The PfuEndoQ structure shows that His8, His10, Glu76, His84, His139 and Asp193 coordinate two Zn^{2+} at the N-terminal PHP domain. The endonuclease activity of PfuEndoQ was decreased by the H8A, H10A, E76Q, H84A, H139A and D193N mutations (Figure 6A). Among the mutants, the E76Q, H84A, H139A and D193N mutants almost abolished their endonuclease activities, although the H8A and H10A mutants retained approximately half of the wild-type activity. These results indicate that the two Zn^{2+} at the N-terminal PHP domain are indispensable for the endonuclease activity of PfuEndoQ. Although PfuEndoQ shows structural similarity to BadPE, the third metal binding residues of BadPE, Asp24, His49 and His262, are not conserved in PfuEndoQ. In addition, PfuEndoQ does not bind any atom at the site that corresponds to the third metal binding site of BadPE. However, PfuEndoQ activity was almost abolished by mutations to Arg114 and His195, which correspond to His49 and His262

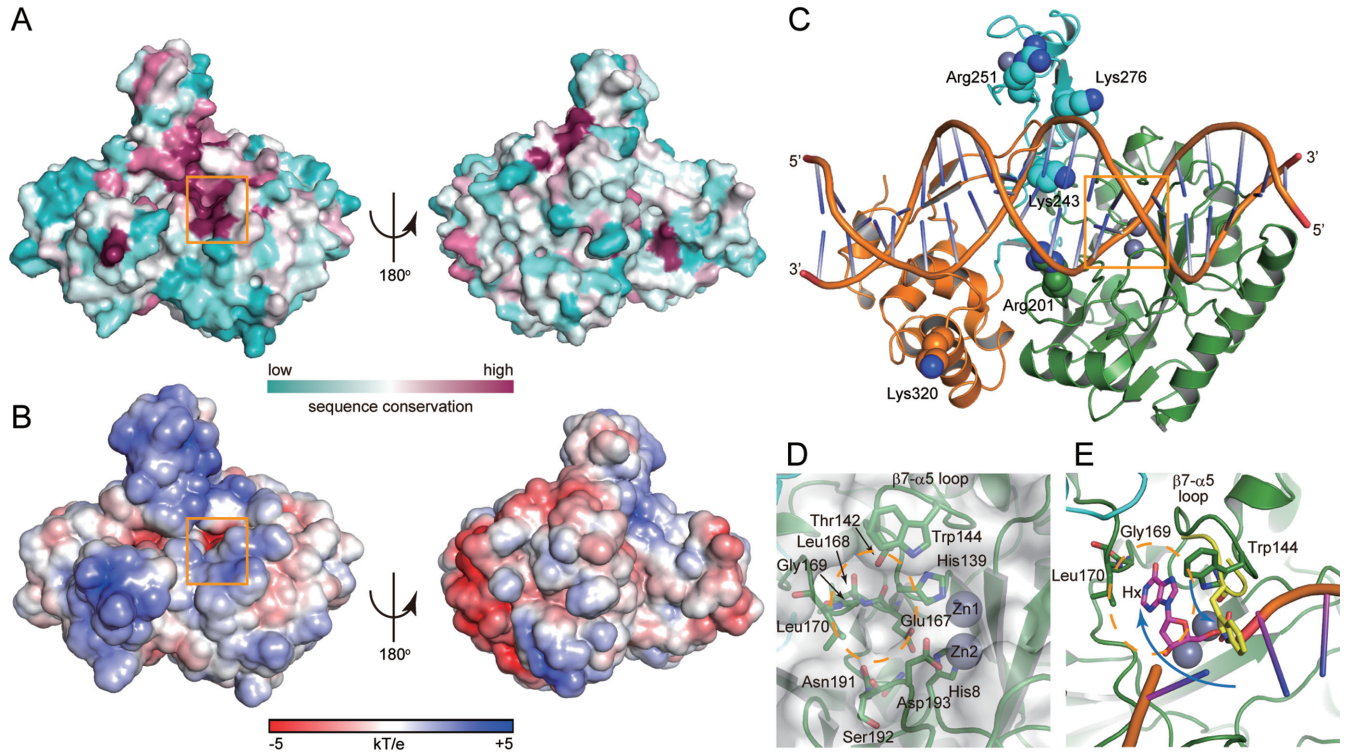

Figure 5. Prediction of the DNA recognition mechanism of PfuEndoQ. (**A**) The sequence conservation of Supplementary Figure S1 is superposed on the molecular surface of PfuEndoQ. The putative active site is indicated by an orange box. (**B**) The ±5 kT/e electrostatic potential of PfuEndoQ is plotted on the solvent-accessible surface. (**C**) PfuEndoQ–dsDNA complex model. The bound dsDNA is shown as an orange cartoon. Positively charged residues analysed by mutation assays are shown by spheres. (**D**) The highly conserved pocket adjacent to the Zn^{2+} binding site of the N-terminal PHP domain. The pocket forming residues are shown as sticks. (**E**) Deaminated base recognition model of PfuEndoQ. Trp144 of the β7-α5 loop (yellow) might be inserted into the dsDNA duplex to flip out a deaminated base (magenta).

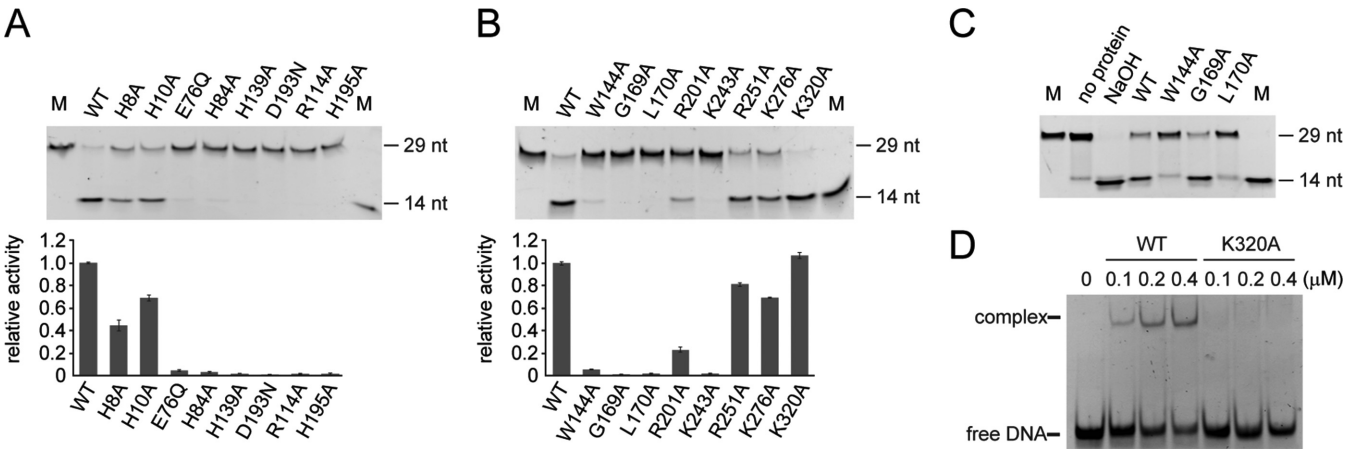

Figure 6. Mutation assay. (**A**) Effect of mutations at the active site for the endonuclease activity. 100 nM of the dI-containing substrate dsDNA and 100 nM of the enzyme (PfuEndoQ (1-400) and its mutants) were mixed and incubated at 40°C for 15 min. Samples were separated through a denaturing 18% polyacrylamide gel. Lane M, ssDNA markers (29 nt and 14 nt). Lower panel shows relative endonuclease activities of the mutants. Plotted values are mean ± SEM ($n = 3$). (**B**) Effect of mutations at the putative deaminated base-binding pocket and the positively charged region for the endonuclease activity. 100 nM of the dI containing substrate dsDNA and 100 nM of the enzyme (PfuEndoQ (1–400) and its mutants) were mixed and incubated at 40°C for 15 min. Samples were separated through a denaturing 18% polyacrylamide gel. Lower panel shows relative endonuclease activities of the mutants. Plotted values are mean ± SEM ($n = 3$). (**C**) Endonuclease activities for the AP site containing substrate. 100 nM of the AP site containing substrate dsDNA and 100 nM of the enzyme (PfuEndoQ (1-400) and its mutants) were mixed and incubated at 40°C for 15 min. Samples were separated through a denaturing 18% polyacrylamide gel. Lane M, ssDNA markers (29 nt and 14 nt). (D) EMSA of PfuEndoQ (1–400) and its K320A mutant using the dI-containing probe. 100 nM of the fluorescein-labelled dsDNA and each concentration of PfuEndoQ were mixed and separated through a 10% polyacrylamide gel.

of BadPE (Figure 6A). This result indicates that Arg114 and His195 are also required for the endonuclease activity of PfuEndoQ, although the functions of these residues remain unclear.

PfuEndoQ possesses a conserved small pocket adjacent to the active site. PfuEndoQ is predicted to recognize a deaminated base using this pocket (Figure 5D and E). The endonuclease activity of PfuEndoQ is highly decreased by mutations to the residues that form the putative deaminated base-binding pocket (G169A and L170A) (Figure 6B). Trp144 is predicted to be inserted into DNA duplex to flip a deaminated base out of the DNA helix. The PfuEndoQ activity was also decreased by the W144A mutation (Figure 6B). The endonuclease activities of the W144A, G169A and L170A mutants were also analysed using the AP site-containing substrate (Supplementary Figure S2C). Because an AP site lacks a base, the mutations that alter the conformation of the deaminated base-binding pocket (G169A and L170A) were predicted not to affect the endonuclease activity for the AP site-containing substrate. The endonuclease activity assay showed that the G169A mutant, which showed highly decreased endonuclease activity for the dI-containing substrate, showed DNA cleavage activity that is broadly comparable to the wild-type activity (Figure 6C). This result suggests that Gly169 is required for the hypoxanthine recognition but is not utilized for AP site recognition. Meanwhile, PfuEndoQ activity was also decreased by the W144A mutation when AP site-containing dsDNA was used as a substrate.

Positively charged residues of the putative dsDNA binding regions were also analysed by mutation assays (Figure 5B). The R201A and K243A mutations highly decreased the endonuclease activity of PfuEndoQ. The R251A and K276A mutations moderately decreased the endonuclease activity of PfuEndoQ. Arg201 at the N-terminal PHP domain and Lys243, Arg251 and Lys276 at the Zn-binding domain are predicted to interact with negatively charged DNA backbone phosphates to stabilize the PfuEndoQ–dsDNA complex. In contrast, mutation of the positively charged residue at the C-terminal helical domain, K320A, did not affect the endonuclease activity of PfuEndoQ. To examine the importance of Lys320 for PfuEndoQ activity in detail, we analysed the DNA binding abilities of the K320A mutants by EMSA. Surprisingly, the DNA binding activity of PfuEndoQ was decreased by the K320A mutation (Figure 6D), although PfuEndoQ activity was not decreased by the K320A mutation. Because PfuEndoQ retains binding to dsDNA after the DNA cleavage reaction, similar to EndoV (18,20,21), PfuEndoQ activity is predicted to be through a single-turnover reaction. The decreased dsDNA-binding ability of the K320A mutant might enable a multi-turnover reaction.

Two tyrosine residues, Tyr244 and Tyr377, are located on the surface of the Zn-binding domain and the C-terminal helical domain, respectively (Supplementary Figure S5A). Tyr residues often function to change the conformation of DNA strands by staking its aromatic ring with the nucleotide bases. Therefore, we made mutants for these two tyrosine residues. The Y244A mutant showed reduced DNA cleavage activity for the hypoxanthine-containing DNA. However, it was very interesting that Y244F, Y377A and Y377F mutants cleaved the same DNA more effectively than the wild-type (Supplementary Figure S5BC). The 244 and 377 positions are probably involved in the cleavage reaction, although their concrete contributions remain to be elucidated.

DISCUSSION

Because hyperthermophilic archaea are always threatened by DNA-damaging agents such as high temperature, ionizing radiation and chemical agents, they are predicted to possess highly sophisticated DNA repair pathways that are not conserved in mesophiles. *Pyrococcus furiosus* is one of the most studied hyperthermophilic archaea and grows optimally at over 100°C (44). *P. furiosus* has extremely efficient DNA repair mechanisms that allow it to survive such an inhospitable environment. For example, chromosome fragmentation caused by ionizing radiation was fully repaired upon incubation at 95°C (45). DNA nucleases are important components of most DNA repair systems. We identified a novel nuclease, PfuExo I from *P. furiosus* (46), and analysed the structure of PhoExoI, the ortholog from *P. horikoshii* (46–48). PfuEndoQ was also discovered from *P. furiosus* (21). Our initial analysis found that EndoQ homologs were found only in the *Thermococcales* and some of the methanogens in Archaea, and were not present in most members of the domains Bacteria and Eukarya. However, our further phylogenetic analyses revealed that some candidates of the EndoQ homologs were found in a small group in Bacteria, and one of the candidates from *Bacillus pumilus* actually showed the EndoQ activity (22). Based on the domain organization, we proposed that EndoQ proteins are classified into three families, in which family 1 consists of archaeal EndoQs, and families 2 and 3 consist of different bacteria, respectively (22). An archaeal EndoQ may be transferred to the bacterial domain and was evolved independently. In this study, we analyzed the crystal structure of PfuEndoQ to understand the structure-function relationships of this endonuclease with unique properties. The structure of PfuEndoQ, as well as PfuExoI as described above, show that they have unique overall structures that are not conserved among other protein structures that recognize their substrate DNA, although their catalytic site is partially similar to pre-existing nuclease folds—PhoExo I and PfuEndoQ possess RNase H-like and PHP-like active site structures, respectively. It could be predicted that during evolution, these nucleases were produced by fusion of a core nuclease domain and other functional domains to recognize and cleave DNA damage which was accelerated by environmental stresses. The crystal structure determined in this study shows that PfuEndoQ fuses the catalytic PHP domain with the Zn-binding domain and the C-terminal helical domain. The cleavage assay using site-specific mutants showed that the electropositive surface of the Zn-binding domain is required for the endonuclease activity of PfuEndoQ. In addition, the EMSA showed that the C-terminal helical domain stabilizes the PfuEndoQ–dsDNA complex. The stable complex with nicked DNA is thought to be important for EndoV to recruit downstream enzymes in its DNA repair pathway and PfuEndQ also remains bound to the nicked DNA product, similar to EndoV (18,20,21). The

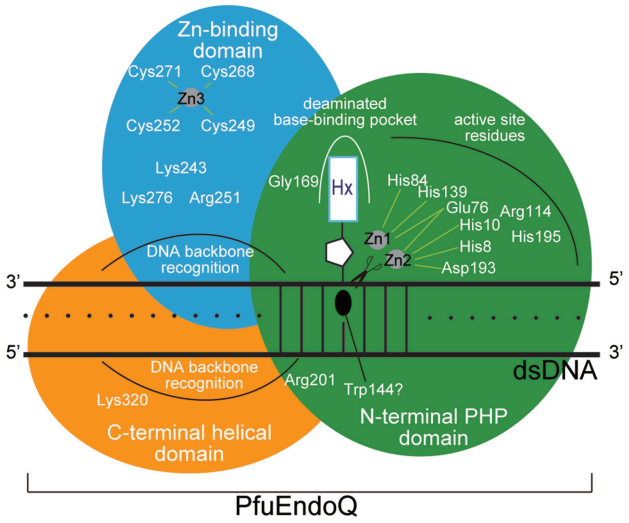

Figure 7. Plausible DNA recognition and cleavage mechanisms of PfuEndoQ. The N-terminal PHP domain, the Zn-binding domain and the C-terminal helical domain are coloured green, cyan and orange, respectively. DNA is shown as a black line.

Zn-binding domain and the C-terminal helical domain are predicted to support a tight PfuEndoQ–dsDNA interaction in order to facilitate its DNA repair pathway, and this tight complex might recruit downstream enzymes of the DNA repair pathway.

Both PfuEndoQ and EndoV recognize deaminated bases in DNA and cleave the phosphodiester bond near the lesion. However, the phosphodiester bonds that are hydrolysed by these endonucleases are different. PfuEndoQ hydrolyses the phosphodiester bond 5′ from the lesion site. In contrast, EndoV hydrolyses the second phosphodiester bond 3′ from the lesion site. This difference can be explained by the relative positions of the nuclease active site and the deaminated base recognition site of these proteins. In the EndoV structure, the active site and the lesion-recognition pocket are separated by ∼8 Å (18). In contrast, the active site of PfuEndoQ is located adjacent to the lesion-recognition pocket, which is surrounded by His139, Thr142, Glu167, Leu168, Gly169, Leu170, Asn191 and Asp193, and the distance between the active site and the lesion-recognition pocket is ∼5 Å. This distance is suitable for hydrolyzing the 5′ phosphate of a deaminated base. The results of the cleavage assay using site-specific mutants support the mechanism for damaged base recognition, in which the deaminated base flips out into the pocket adjacent to the active site. We built the structural model of the PfuEndoQ–dsDNA complex based on amino acid sequence conservation and the electrostatic potential of the protein surface (Figure 7). The validity of this model was supported by the results of mutational analyses using site-specific mutants. However, DNA binding proteins frequently deform the dsDNA structure to form tight protein–DNA interactions (42,49,50). The precise DNA recognition and cleavage mechanisms of PfuEndoQ will be clarified by the determination of the PfuEndoQ–dsDNA complex structure. Structural analyses of the PfuEndoQ-damaged DNA complex will also help us understand how these proteins distinguish the type of base, since EndoV cleaves only hypoxanthine-containing DNA, but EndoQ recognizes uracil, xanthine and AP site, in addition to hypoxanthine.

In this study, we determined the crystal structure of PfuEndoQ, whose C-terminal disordered region was truncated (Figure 1). Although the disordered region-truncated construct (PfuEndoQ (1–400)) showed the approximately the same DNA cleavage activity as the wild-type enzyme, our previous study showed that the QRTLLQYI sequence in the disordered region of PfuEndoQ is the PIP motif that interacts with PCNA. PCNA interacts with EndoQ in its PIP-containing region with a high affinity (apparent K_D = 55 nM) and enhances the EndoQ activity *in vitro* (38). The C-terminal PIP-containing regions of the PCNA binding proteins generally protrude from the core domains, and therefore, the truncated C-terminal region of EndoQ itself does not seem to affect on the binding affinity to the DNA substrate. However, the EndoQ-dsDNA interaction will be stabilized by binding of PCNA. When a PfuEndoQ–dsDNA-PCNA structural model was prepared using the coordinates of the PCNA-DNA complex (PDB ID: 5L7C) (51) and the PCNA-PIP peptide complex (PDB ID: 1ISQ) (52), the C-terminus of the C-terminal helical domain of PfuEndoQ was reasonably located near the PIP–peptide binding region of PCNA (Supplementary Figure S6). The stable EndoQ–dsDNA interaction by the help of PCNA may also enhance recruitment of downstream enzymes of the DNA repair pathway. The precise structural basis for the activation of EndoQ by PCNA could be uncovered by structural determination of the EndoQ–dsDNA–PCNA complex.

DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structure have been deposited in the Protein Data bank under accession number 5ZB8.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The synchrotron-radiation experiments were performed at BL-17A in the Photon Factory (Tsukuba, Japan) (2017G536).

FUNDING

Platform for Drug Discovery, Informatics and Structural Life Science (PDIS) from the Ministry of Education, Culture, Sports, Science and Technology, Japan; JSPS KAKENHI [26242075 to Y.I.]. Funding for open access charge: JSPS KAKENHI [17K19581].
Conflict of interest statement. None declared.

REFERENCES

1. Lindahl,T. and Wood,R.D. (1999) Quality control by DNA repair. *Science*, **286**, 1897–1905.
2. Sancar,A., Lindsey-Boltz,L.A., Unsal-Kaçmaz,K. and Linn,S. (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.*, **73**, 39–85.

3. Li,X. and Heyer,W.D. (2008) Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res.*, **18**, 99–113.
4. Li,G.M. (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res.*, **18**, 85–98.
5. Zharkov,D.O. (2008) Base excision DNA repair. *Cell. Mol. Life Sci.*, **65**, 1544–1565.
6. Shuck,S.C., Short,E.A. and Turchi,J.J. (2008) Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Res.*, **18**, 64–72.
7. Yasui,A. (2013) Alternative excision repair pathways. *Cold Spring Harb. Perspect. Biol.*, **5**, a012617.
8. Kow,Y.W. (2002) Repair of deaminated bases in DNA. *Free Radic. Biol. Med.*, **33**, 886–893.
9. Schouten,K.A. and Weiss,B. (1999) Endonuclease V protects *Escherichia coli* against specific mutations caused by nitrous acid. *Mutat. Res.*, **435**, 245–254.
10. Kelman,Z. and White,M.F. (2005) Archaeal DNA replication and repair. *Curr. Opin. Microbiol.*, **8**, 669–676.
11. Ishino,Y., Nishino,T. and Morikawa,K. (2006) Mechanisms of maintaining genetic stability by homologous recombination. *Chem. Rev.*, **106**, 324–339.
12. Rouillon,C. and White,M.F. (2011) The evolution and mechanisms of nucleotide excision repair proteins. *Res. Microbiol.*, **162**, 19–26.
13. Watrin,L. and Prieur,D. (1996) UV and ethyl methanesulfonate effects in hyperthermophilic archaea and isolation of auxotrophic mutants of *Pyrococcus* strains. *Curr. Microbiol.*, **33**, 377–382.
14. Jacobs,K.L. and Grogan,D.W. (1998) Spontaneous mutation in a thermoacidophilic archaeon: evaluation of genetic and physiological factors. *Arch. Microbiol.*, **169**, 81–83.
15. Schormann,N., Ricciardi,R. and Chattopadhyay,D. (2014) Uracil–DNA glycosylases—structural and functional perspectives on an essential family of DNA repair enzymes. *Protein Sci.*, **23**, 1667–1685.
16. Aravind,L., Walker,D.R. and Koonin,E. V (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.*, **27**, 1223–1242.
17. Cao,W. (2013) Endonuclease V: an unusual enzyme for repair of DNA deamination. *Cell. Mol. Life Sci.*, **70**, 3145–3156.
18. Dalhus,B., Arvai,A.S., Rosnes,I., Olsen,Ø.E., Backe,P.H., Alseth,I., Gao,H., Cao,W., Tainer,J.A. and Bjørås,M. (2009) Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.*, **16**, 138–143.
19. Zhang,Z., Jia,Q., Zhou,C. and Xie,W. (2015) Crystal structure of *E. coli* endonuclease V, an essential enzyme for deamination repair. *Sci. Rep.*, **5**, 12754.
20. Kiyonari,S., Egashira,Y., Ishino,S. and Ishino,Y. (2014) Biochemical characterization of endonuclease V from the hyperthermophilic archaeon, *Pyrococcus furiosus*. *J. Biochem.*, **155**, 325–333.
21. Shiraishi,M., Ishino,S., Yamagami,T., Egashira,Y., Kiyonari,S. and Ishino,Y. (2015) A novel endonuclease that may be responsible for damaged DNA base repair in *Pyrococcus furiosus*. *Nucleic Acids Res.*, **43**, 2853–2863.
22. Shiraishi,M., Ishino,S., Cann,I. and Ishino,Y. (2017) A functional endonuclease Q exists in the bacterial domain: identification and characterization of endonuclease Q from *Bacillus pumilus*. *Biosci. Biotechnol. Biochem.*, **81**, 931–937.
23. Kabsch,W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
24. Sheldrick,G.M. (2008) A short history of SHELX. *Acta Crystallogr. A Found. Crystallogr.*, **64**, 112–122.
25. Bricogne,G., Vonrhein,C., Flensburg,C., Schiltz,M. and Paciorek,W. (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 2023–2030.
26. Vonrhein,C., Blanc,E., Roversi,P. and Bricogne,G. (2007) Automated structure solution with autoSHARP. In: *Macromolecular Crystallography Protocols*. Humana Press, New Jersey, Vol. **2**, pp. 215–230.
27. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G.W., McCoy,A. *et al.* (2011) Overview of the CCP 4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.
28. Cowtan,K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 1002–1011.
29. 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. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352–367.
30. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
31. Chen,V.B., Arendall,W.B., Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
32. Ward,J.J., Sodhi,J.S., McGuffin,L.J., Buxton,B.F. and Jones,D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.*, **337**, 635–645.
33. Holm,L. and Sander,C. (1995) Dali: a network tool for protein structure comparison. *Trends Biochem Sci.*, **20**, 478–480.
34. Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W., Lopez,R., McWilliam,H., Remmert,M., Söding,J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.
35. Gouet,P., Courcelle,E., Stuart,D.I. and Métoz,F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.
36. Ashkenazy,H., Erez,E., Martz,E., Pupko,T. and Ben-Tal,N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, **38**, W529–W533.
37. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 10037–10041.
38. Shiraishi,M., Ishino,S., Yoshida,K., Yamagami,T., Cann,I. and Ishino,Y. (2016) PCNA is involved in the EndoQ-mediated DNA repair process in *Thermococcales*. *Sci. Rep.*, **6**, 25532.
39. Ishino,S., Makita,N., Shiraishi,M., Yamagami,T. and Ishino,Y. (2015) EndoQ and EndoV work individually for damaged DNA base repair in *Pyrococcus furiosus*. *Biochimie*, **118**, 264–269.
40. Han,G.W., Ko,J., Farr,C.L., Deller,M.C., Xu,Q., Chiu,H.-J., Miller,M.D., Sefcikova,J., Somarowthu,S., Beuning,P.J. *et al.* (2011) Crystal structure of a metal-dependent phosphoesterase (YP_910028.1) from Bifidobacterium adolescentis: Computational prediction and experimental validation of phosphoesterase activity. *Proteins Struct. Funct. Bioinforma.*, **79**, 2146–2160.
41. Bailey,S., Wing,R.A. and Steitz,T.A. (2006) The structure of *T. aquaticus* DNA polymerase III is distinct from eukaryotic replicative DNA polymerases. *Cell*, **126**, 893–904.
42. Bruner,S.D., Norman,D.P. and Verdine,G.L. (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*, **403**, 859–866.
43. Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
44. Fiala,G. and Stetter,K.O. (1986) *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Arch. Microbiol.*, **145**, 56–61.
45. DiRuggiero,J., Santangelo,N., Nackerdien,Z., Ravel,J. and Robb,F.T. (1997) Repair of extensive ionizing-radiation DNA damage at 95 degrees C in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.*, **179**, 4643–4645.
46. Tori,K., Ishino,S., Kiyonari,S., Tahara,S. and Ishino,Y. (2013) A Novel Single-Strand specific 3′-5′ exonuclease found in the hyperthermophilic archaeon, *Pyrococcus furiosus*. *PLoS One*, **8**, e58497.
47. Miyazono,K., Tsutsumi,K., Ishino,Y. and Tanokura,M. (2014) Expression, high-pressure refolding, purification, crystallization and preliminary X-ray analysis of a novel single-strand-specific 3′-5′ exonuclease PhoExo I from *Pyrococcus horikoshii* OT3. *Acta Crystallogr. F Struct. Biol. Commun.*, **70**, 1076–1079.
48. Miyazono,K., Ishino,S., Tsutsumi,K., Ito,T., Ishino,Y. and Tanokura,M. (2015) Structural basis for substrate recognition and

processive cleavage mechanisms of the trimeric exonuclease PhoExo I. *Nucleic Acids Res.*, **43**, 7122–7136.
49. Miyazono,K., Furuta,Y., Watanabe-Matsui,M., Miyakawa,T., Ito,T., Kobayashi,I. and Tanokura,M. (2014) A sequence-specific DNA glycosylase mediates restriction-modification in *Pyrococcus abyssi*. *Nat. Commun.*, **5**, 3178.
50. Mees,A., Klar,T., Gnau,P., Hennecke,U., Eker,A.P., Carell,T. and Essen,L.O. (2004) Crystal structure of a photolyase bound to a CPD-like DNA lesion after in situ repair. *Science*, **306**, 1789–1793.
51. De March,M., Merino,N., Barrera-Vilarmau,S., Crehuet,R., Onesti,S., Blanco,F.J. and De Biasio,A. (2017) Structural basis of human PCNA sliding on DNA. *Nat. Commun.*, **8**, 13935.
52. Matsumiya,S., Ishino,S., Ishino,Y. and Morikawa,K. (2002) Physical interaction between proliferating cell nuclear antigen and replication factor C from *Pyrococcus furiosus*. *Genes Cells*, **7**, 911–922.