



OPEN

Crystal structure and DNA cleavage mechanism of the restriction DNA glycosylase R.CcoLI from *Campylobacter coli*

Ken-ichi Miyazono, Delong Wang, Tomoko Ito & Masaru Tanokura

While most restriction enzymes catalyze the hydrolysis of phosphodiester bonds at specific nucleotide sequences in DNA, restriction enzymes of the HALFPIPE superfamily cleave *N*-glycosidic bonds, similar to DNA glycosylases. Apurinic/aprimidinic (AP) sites generated by HALFPIPE superfamily proteins are cleaved by their inherent AP lyase activities, other AP endonuclease activities or heat-promoted β -elimination. Although the HALFPIPE superfamily protein R.PabI, obtained from a hyperthermophilic archaea, *Pyrococcus abyssi*, shows weak AP lyase activity, HALFPIPE superfamily proteins in mesophiles, such as R.CcoLI from *Campylobacter coli* and R. HpyAXII from *Helicobacter pylori*, show significant AP lyase activities. To identify the structural basis for the AP lyase activity of R.CcoLI, we determined the structure of R.CcoLI by X-ray crystallography. The structure of R.CcoLI, obtained at 2.35-Å resolution, shows that a conserved lysine residue (Lys71), which is stabilized by a characteristic β -sheet structure of R.CcoLI, protrudes into the active site. The results of mutational assays indicate that Lys71 is important for the AP lyase activity of R.CcoLI. Our results help to elucidate the mechanism by which HALFPIPE superfamily proteins from mesophiles efficiently introduce double-strand breaks to specific sites on double-stranded DNA.

Restriction enzymes recognize and cleave specific DNA sequences that are not protected by cognate methyltransferases. Among the restriction enzymes, type II restriction enzymes, which recognize specific DNA sequences and cleave at or near sequences, are widely utilized in the field of biotechnology in such applications as genetic engineering and DNA fingerprinting¹. Type II restriction enzymes are classified into five groups depending on their structural motifs²: the PD-(D/E)XK superfamily^{3,4}, the HNH superfamily^{5,6}, the PLD superfamily⁷, the GIY-YIG superfamily⁸, and the HALFPIPE superfamily⁹. R.PabI from *Pyrococcus abyssi* is a useful example of HALFPIPE superfamily restriction enzymes. Although most type II restriction enzymes cleave DNA by their endonuclease activities using Mg²⁺ ions, R.PabI, which recognizes the sequence 5'-GTAC-3', cleaves double-stranded DNA (dsDNA) without the addition of divalent cations at high temperatures¹⁰. In contrast to the general restriction enzymes, R.PabI does not cleave a phosphodiester bond of DNA but cleaves an *N*-glycosidic bond in its recognition sequence, similar to DNA glycosylases¹¹. The apurinic/aprimidinic (AP) site generated by R.PabI is cleaved by its weak AP lyase activities, other AP endonuclease activities or heat-promoted β -elimination^{11,12}. To date, we have determined the structures of R.PabI in the DNA-free state; the sequence-nonspecific DNA-binding state; the intermediate state, which is important for the indirect readout of the recognition sequence; and the product DNA-binding state^{9,11,13,14} (see Supplementary Fig. S1 online). R.PabI forms a homodimer and has a positively charged HALFPIPE structure at the dimer interface⁹. R.PabI recognizes the sequence 5'-GTAC-3' in dsDNA using the HALFPIPE region. The 5'-GTAC-3' site in dsDNA is unwound by the binding of R.PabI dimer, and three residues (Tyr68, His211 and Asp214) of R.PabI catalyze the hydrolysis of the *N*-glycosidic bond between adenine and deoxyribose in the recognition sequence (see Supplementary Fig. S1 online). These results indicate that the HALFPIPE superfamily restriction enzymes are not restriction endonucleases but are restriction DNA glycosylases¹⁵.

DNA glycosylases are enzymes that are utilized in the DNA repair pathways of organisms. In the base excision repair (BER) pathway, a damaged base is recognized by a DNA glycosylase, and an *N*-glycosidic bond of the damaged base is subsequently cleaved. A DNA molecule cleaved by a DNA glycosylase is recognized and repaired by downstream enzymes of the BER pathway^{16,17}. For example, the uracil DNA glycosylase (uracil

Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi Bunkyo-ku, Tokyo 113-8657, Japan. email: amtanok@mail.ecc.u-tokyo.ac.jp

N-glycosylase, UNG) recognizes a uracil in dsDNA, which is generated by cytosine deamination or misincorporation of deoxyuridine monophosphate during replication^{18,19}. The 8-oxoguanine DNA glycosylase (OGG1) recognizes an 8-oxoguanine in dsDNA, which is generated by guanine oxidation^{20,21}. Among DNA glycosylases, some groups of DNA glycosylases, including OGG1, not only cleave *N*-glycosidic bonds but also cleave phosphodiester bonds through their inherent AP lyase activities (bifunctional DNA glycosylases). In contrast, other DNA glycosylases, including UNG, only hydrolyze *N*-glycosidic bonds (monofunctional DNA glycosylases). The AP lyase activities of bifunctional DNA glycosylases require an amine group, including in the N-terminal threonine²² or proline²³ or in the side chain of lysine^{24,25}; after the cleavage of an *N*-glycosidic bond and the generation of a reactive oxocarbenium intermediate, which are promoted by a catalytic acidic residue, an amine group forms an iminium crosslink with C1' of the deoxynucleotide to cleave the DNA backbone through β - and δ -eliminations (see Supplementary Fig. S2 online)^{26,27}.

At medium temperature (40 °C), R.PabI does not cleave the DNA backbone but only hydrolyzes the *N*-glycosidic bond of adenine to produce an AP site¹¹. R.PabI exhibits weak AP lyase activity at high temperatures¹². These observations indicate that R.PabI acts as a monofunctional DNA glycosylase at medium temperature. In contrast, R.PabI homologs from mesophiles, such as R.CcoLI from *Campylobacter coli* and R.HpyAXII from *Helicobacter pylori*, show significant AP lyase activities at 37 °C; R.CcoLI and R.HpyAXII recognize the sequence 5'-GTAC-3' in dsDNA, similar to R.PabI, and cleave the DNA backbone at the adenine site¹⁵. Because AP lyase activities of known DNA glycosylases require an amine group near the active sites, R.CcoLI and R. HpyAXII are predicted to possess amine groups at their active sites that are not conserved in the structure of R.PabI. The amino acid sequence alignment of R.PabI, R.CcoLI and R.HpyAXII shows that R.CcoLI and R. HpyAXII have characteristic insertion residues adjacent to the active site residue (Tyr68 of R.PabI) (Fig. 1a). Although this insertion may be related to the AP lyase activities of R.CcoLI and R. HpyAXII, the structure of the insertion residue has not been determined to date. To demonstrate the structural basis for the AP lyase activity of R.CcoLI, we determined the crystal structure of R.CcoLI at 2.35-Å resolution. The R.CcoLI structure and accompanying biochemical data showed that the characteristic insertion residues of R.CcoLI form an antiparallel β -sheet near the catalytic residues and that Lys71 in the antiparallel β -sheet is utilized for AP lyase activity. The HALFPIPE superfamily restriction enzymes exist both in mesophiles, such as *C. coli* and *H. pylori*, and in hyperthermophiles, such as *P. abyssi*. Comparing the structures and activities of R.CcoLI and R.PabI may help to elucidate the mechanism by which HALFPIPE superfamily restriction enzymes adapt to a wide range of temperature conditions.

Results

Determining the structure of R.CcoLI. To determine the crystal structure of R.CcoLI, we overexpressed the D225N mutant of R.CcoLI in *E. coli*. Asp225 of R.CcoLI corresponds to the active site residue of R.PabI (Asp214) (Fig. 1a). Asp214 of R.PabI is employed to stabilize the oxocarbenium ion intermediate, to deprotonate the catalytic water and to bind substrate DNA; the D214N mutation reduces the DNA glycosylase activity and the substrate DNA-binding ability of R.PabI¹¹. Similar to the D214N mutation of R.PabI, the D225N mutation of R.CcoLI is predicted to reduce its DNA glycosylase activity. The D225N mutation was introduced to R.CcoLI to overexpress toxic proteins in *E. coli* cells. After Ni-NTA affinity chromatography, ion-exchange chromatography and gel-filtration chromatography purification, the highly purified R.CcoLI(D225N) mutant was obtained (see Supplementary Fig. S3 online). The gel-filtration analysis showed that R.CcoLI(D225N) forms a homodimer in solution similar to R.PabI (Fig. 1b). The melting temperature of the R.CcoLI(D225N) mutant was 57.5–58.5 °C (Fig. 1c). The R.CcoLI(D225N) mutant retained sequence-specific DNA cleavage activity (Fig. 1d and Supplementary Fig. S4 online). Although we could obtain crystals of the R.CcoLI(D225N) mutant and their X-ray diffraction data, the refinement statistics of the R.CcoLI(D225N) mutant structure were poor. To improve the quality of the crystals, the C189S mutation was introduced to the R.CcoLI(D225N) mutant. Cys189 of R.CcoLI corresponds to Val181 of R.PabI. The side chain of R.PabI Val181 is exposed to a solvent and does not interact with dsDNA^{11,13,14}. The C189S mutation of R.CcoLI was predicted to reduce the heterogeneity of R.CcoLI that was caused by the oxidation of Cys189. The R.CcoLI(C189S-D225N) mutant was overexpressed and purified by a similar method to that employed for the R.CcoLI(C225N) mutant (Fig. 1b). The C189S mutation did not affect the melting temperature or the sequence-specific DNA cleavage activity of R.CcoLI (Fig. 1c,d). These results suggested that the C189S mutation does not affect the function of R.CcoLI.

The R.CcoLI structure was determined at a resolution of 2.35 Å using the crystal of the R.CcoLI(C189S-D225N) mutant. The final model contained one R.CcoLI dimer (chains A and B) in the asymmetric unit (Fig. 2a and Supplementary Fig. S5 online). R.CcoLI consists of five α -helices, fifteen β -strands and one 3_{10} (η)-helix and forms the characteristic HALFPIPE structure using strands β 3- β 2- β 5- β 13- β 12- β 11. Similar to the R.PabI structure, the HALFPIPE region of R.CcoLI has a positively charged surface (Fig. 2b)^{9,11,13,14}. Due to poor electron density, the structure models of the β 2- β 3 and β 3- β 4 loops of chain A, the β 1- β 2 loop, the β 2- β 5 region, the β 6- β 7 loop, the β 11- β 12 loop, the β 12- β 13 loop and the η 1- β 15 region of chain B are not included in the final model. The final model contains only ten water molecules, despite the relatively high-resolution structure (2.35 Å). These properties cause the structure to exhibit a relatively high R_{free} value (Table 1). In the R.CcoLI structure, C189S is exposed to solvent, as predicted (Fig. 2a).

Structure comparison with R.PabI. When the structures of R.CcoLI and R.PabI (the product DNA-binding state, PDB: 3WAZ) were compared, the secondary structures of two regions (sites 1 and 2) were observed to be different between the proteins (Fig. 3a)¹¹. Site 1 exists on helices α 1, α 4 and α 5 of R.CcoLI. In this region of R.CcoLI, three catalytic residues of R.PabI, that is, Tyr68, His211 and Asp214, are conserved as Tyr52, His222 and Asp225, respectively. In the R.CcoLI structure, the insertion residues of R.CcoLI (Fig. 1a) forms an antiparallel β -sheet (β 6- β 8- β 7) adjacent to the active site (Fig. 3b and Supplementary Fig. S5 online). According

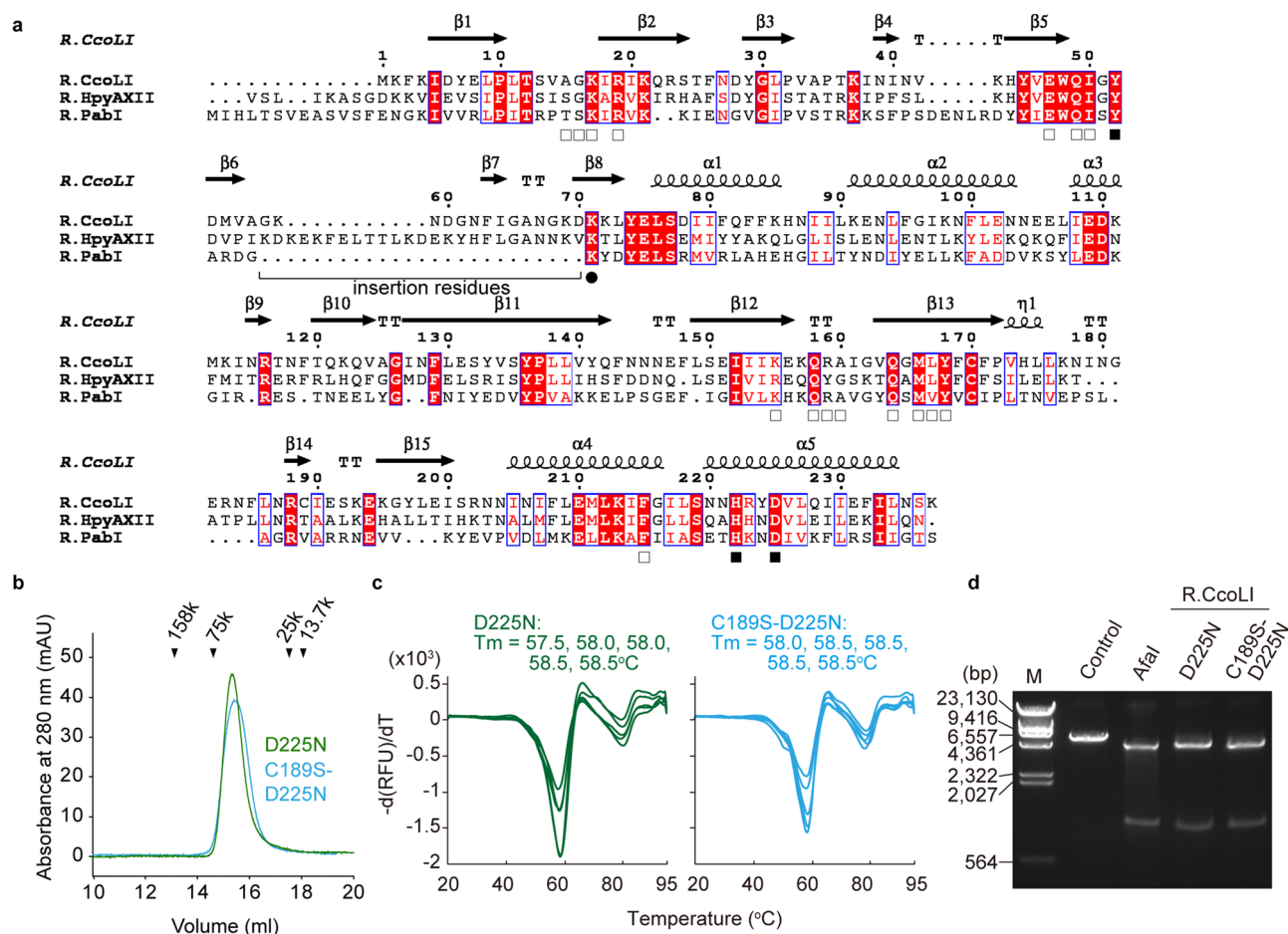


Figure 1. Preparation of R.CcoLI. **(a)** Sequence alignment of R.PabI homologs (R.CcoLI, WP_002830209; R.HpyAXII, ACI43084.1; R.PabI, CAB49082). Invariant residues are highlighted with red boxes, and conserved residues are shown in red text. The secondary structure of R.CcoLI (chain A) is indicated by helices (α and β_{10} (η)-helices), arrows (β -strand), and TT (β -turn). The catalytic triad and base recognition residues of R.PabI are marked with black and open squares, respectively. The residue that is predicted to be involved in the DNA lyase activity of R.CcoLI (Lys71) is marked with a circle. Insertion residues that are not conserved in R.PabI are indicated by a black line. **(b)** Gel filtration of the R.CcoLI(D225N) ($M_r = 28,007$) and R.CcoLI(C189S-D225N) ($M_r = 27,991$) mutants. The peak positions of the marker proteins are indicated by black triangles. Molecular weights of the R.CcoLI(D225N) and R.CcoLI(C189S-D225N) mutants were estimated to be 56 k and 53 k, respectively. Data are representative of three independent experiments. **(c)** Thermostability assay of R.CcoLI mutants. The melting curves of the R.CcoLI(D225N) mutant and the R.CcoLI(C189S-D225N) mutant are shown by green and cyan lines, respectively ($n = 5$). RUF, relative fluorescence unit. **(d)** DNA cleavage activities of the R.CcoLI(D225N) and R.CcoLI(C189S-D225N) mutants. The substrate DNA was treated with R.CcoLI mutants or AfaI, which cleaves the DNA duplex at the same site as R.CcoLI. The image is a cropped gel image. The full image is in Supplementary Fig. S7 online.

to this structure, the side chains of Asp53, Asn67 and Lys71 face the active site of R.CcoLI. Lys71 of R.CcoLI is conserved in R.PabI as Lys73 (Fig. 1a). However, in the R.CcoLI structure, the side-chain orientation of Lys71 is flipped to the active site due to the insertion of the R.CcoLI specific antiparallel β -sheet. In contrast, Asp53 and Asn67 are R.CcoLI-specific residues. Site 2 is the N-terminal region of R.CcoLI. In the structure of R.PabI, the N-terminal region forms a three-stranded antiparallel β -sheet. However, R.CcoLI lacks the first β -strand due to the shortage of the N-terminal residues (Fig. 1a) and forms a two-stranded antiparallel β -sheet (β_1 – β_{15} , Fig. 3c). Because R.CcoLI lacks the first β -strand, the side chains of Phe3, Ile5 and Tyr7 are partially exposed to the solvent. In addition to these two sites, the lengths of two loops (β_{10} – β_{11} and η_1 – β_{14} loops in R.CcoLI) also differ between R.CcoLI and R.PabI. In the R.PabI structure, the lengths of the β_{10} – β_{11} and η_1 – β_{14} loops are shortened by two and five residues, respectively.

The structural comparison between R.CcoLI and R.PabI (the product DNA-binding state) shows that the structures of the dimerization region are also different (Fig. 3a). The structure of the R.PabI dimer is modified depending on the DNA binding state^{9,11,13,14} (see Supplementary Fig. S1 online). When the protomer structure of R.CcoLI (chain A) was compared to those of R.PabI in the DNA-free state (PDB: 2DVY), the sequence-nonspecific DNA-binding state (PDB: 5IFF), the intermediate state (PDB: 6L2O) and the product DNA-binding state (PDB: 3WAZ), the R.PabI structure in the sequence-nonspecific DNA-binding state showed the highest structural

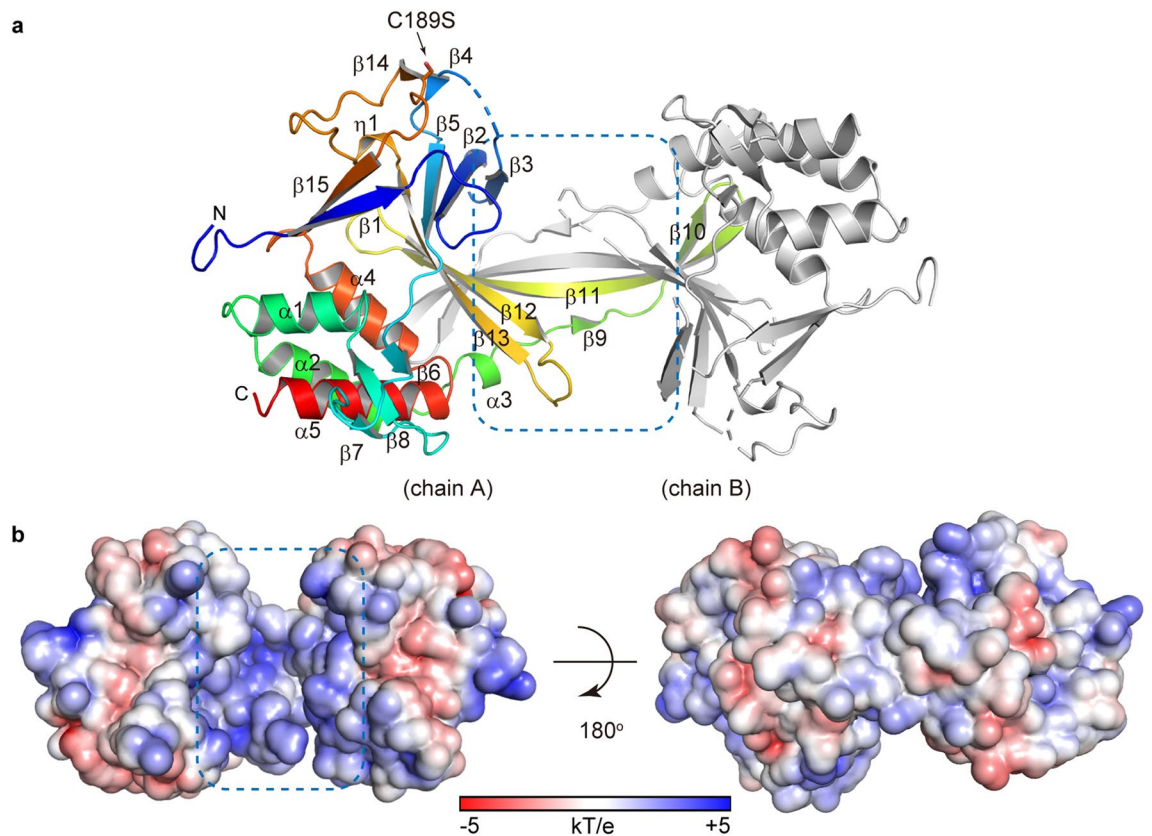


Figure 2. Overall structure of R.CcoLI. **(a)** Ribbon diagram of the R.CcoLI dimer. The ribbon of one R.CcoLI protomer (chain A) is colored blue (in the N terminus) to red (in the C terminus). The other R.CcoLI protomer (chain B) is colored gray. The HALFPIPE region of R.CcoLI dimer is indicated by a blue dotted square. The side chain of C189S is shown by a stick model. **(b)** The ± 5 kT/e electrostatic potential of the R.CcoLI dimer is plotted on the solvent-accessible surface.

similarity (Z-score from the Dali server = 21.2, root-mean-square deviation (RMSD) = 1.9 Å, sequence identity = 26%)²⁸, although the R.CcoLI structure was determined without DNA. The dimeric structure of R.CcoLI was also well superposed with the dimeric structure of R.PabI in the sequence-nonspecific DNA-binding state, and the RMSD was 2.3 Å for 298 superposed Ca atoms (Fig. 3d and Supplementary Fig. S6 online).

Active site structure. The amino acid sequence alignment of R.PabI homologs shows that most of the conserved residues are located near the positively charged HALFPIPE region (Figs. 1a, 2b and 4a). R.CcoLI is predicted to recognize a negatively charged dsDNA in the HALFPIPE region, similar to R.PabI. To elucidate the DNA cleavage mechanism of R.CcoLI, we created the R.CcoLI-product DNA complex model using the coordinates of R.CcoLI and that of R.PabI in the product DNA binding state (PDB: 3WAZ) (Fig. 4a,b)¹¹. The model structure of the R.CcoLI-product DNA complex shows that the catalytic residues of R.CcoLI (Tyr52, His222 and Asp225) clearly exist near the cleaved *N*-glycosidic bond of the adenine in the recognition sequence (Fig. 4b,c). Among the base-recognizing residues of R.PabI, ten residues are conserved in R.CcoLI (Fig. 1a). According to the amino acid sequence similarity, Lys17, Arg19, Glu47, Gln49 and Tyr168 of R.CcoLI are predicted to recognize a guanine base in the recognition sequence; Gln164 and Met166 of R.CcoLI are predicted to recognize a thymine base in the recognition sequence; Ile50 and Phe215 of R.CcoLI are predicted to recognize an adenine base in the recognition sequence, and Gln158 is predicted to recognize a cytosine base in the recognition sequence (Fig. 4d). In addition to these residues, Asp53, Asn67 and Lys71 of R.CcoLI are located near the active site due to the insertion of the characteristic antiparallel β -sheet (Fig. 4b,c). The model structure of the R.CcoLI-product DNA complex shows that the distance between the side-chain amine group of Lys71 and the C1' carbon atom of cleaved deoxyadenosine is approximately 3 Å (Fig. 4b). Because the AP lyase activity of DNA glycosylase is initiated by iminium crosslink formation between C1' and an amine group^{22–27} (see Supplementary Fig. S2 online), Lys71 of R.CcoLI is predicted to be an important residue for the AP lyase activity of R.CcoLI. In contrast, the side chains of Asp53 and Asn67 are located near the O4' oxygen atom of deoxyribose and the phosphate group of the DNA backbone, respectively (Fig. 4b,d). These residues are predicted to be utilized for the stabilization of the R.CcoLI-DNA complex.

Mutation assay. The model structure of the R.CcoLI-DNA complex suggests that Asp53, Asn67 and Lys71 are important for the catalytic activity of R.CcoLI; in particular, the side-chain amine group of Lys71 is predicted

Data collection	
Space group	$P2_12_1$
Cell dimensions	
a, b, c (Å)	66.49, 89.60, 100.66
Resolution (Å)	47.17–2.35 (2.44–2.35)*
R_{sym} (%)	4.7 (79.2)
Mean ($I/\sigma I$)	23.1 (2.1)
Completeness (%)	99.9 (99.7)
Redundancy	7.2 (6.6)
CC (1/2)	0.999 (0.705)
Refinement	
R/R_{free} (%)	22.9/25.9
No. atoms	
R.CcoLI	3545
Water	10
B -factors (Å ²)	
R.CcoLI	72.10
Water	59.04
R.m.s deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.384
Ramachandran plot	
Favored region (%)	98.0
Allowed region (%)	2.0

Table 1. Summary of data collection and refinement statistics of R.CcoLI. *The numbers in parentheses represent data for the highest-resolution shells.

to be important for the AP lyase activity of R.CcoLI. To analyze the importance of the side-chain atoms of these residues, we prepared the D53A-D225N, N67A-D225N and K71A-D225N mutants and analyzed their DNA glycosylase activities (Fig. 5a–c). In this study, we analyzed the effects of mutations using the D225N mutant as a control. Asp225 of R.CcoLI is predicted to be used to cleave the *N*-glycosidic bond and to generate the oxocarbenium intermediate. The D225N mutation was predicted to reduce the efficiency of this reaction, although the D225N mutant exhibited sequence-specific DNA cleavage activity (Fig. 1d). If the AP lyase activity of R.CcoLI is mediated by the interaction between the side-chain amine group of Lys71 and DNA, the K71A mutant of R.CcoLI acts as a monofunctional DNA glycosylase that hydrolyzes the *N*-glycosidic bond; the oxocarbenium intermediate is not attacked by the amine group of Lys71 but is attacked by water to generate an AP site. The DNA backbone of an AP site is unstable and is easily cleaved by NaOH treatment at the 3' and 5' sides of the AP sites (β - and δ -eliminations, respectively)^{11,15}. The results of the DNA glycosylase assay of the R.CcoLI(D225N) mutant showed that the fractions of cleaved DNA were the same, regardless of whether NaOH was added. This finding indicated that the D225N mutant of R.CcoLI functions as a bifunctional DNA glycosylase. In contrast, the results of the DNA glycosylase assay of the K71A-D225N mutant showed that the fraction of cleaved DNA was highly increased by NaOH treatment. This result indicated that the K71A-D225N mutant of R.CcoLI functions as a monofunctional DNA glycosylase and that the side-chain amine group of Lys71 is important for the DNA cleavage activity of R.CcoLI. The reaction speed of the K71A-D225N mutant was determined to be higher than that of the D225N mutant (Fig. 5c). This finding suggested that the Lys71-dependent β -elimination is the rate-limiting step of the R.CcoLI activity. Notably, the K71A-D225N mutant slightly cleaved the substrate DNA in the absence of NaOH treatment (Fig. 5a–c); the K71A-D225N mutant also functioned as a bifunctional DNA glycosylase.

Although Asp53 does not possess amine groups, the results of the DNA glycosylase assay of the D53A-D225N mutant also showed a cleavage pattern similar to that of the K71A-D225N mutant; the fraction of cleaved DNA was highly increased by NaOH treatment (Fig. 5a–c). Because the side chain of Asp53 is located adjacent to the side chain of Lys71 and the O4' oxygen atom of deoxyribose (Fig. 4b), the D53A mutation might decrease the stability of the side chain of Lys71 and deoxyribose, while the increased flexibility in the D53A-D225N mutant is predicted to reduce the efficiency of iminium crosslink formation between Lys71 and DNA. The results of the DNA glycosylase assay examining the N67A-D225N mutant showed that the cleavage activity was reduced compared to that of the D225N mutant, regardless of whether NaOH was added (Fig. 5a–c). This result suggested that the side chain of Asn67 is not employed for DNA backbone cleavage; also, Asn67 is predicted to be used for DNA stabilization (Fig. 4b,d).

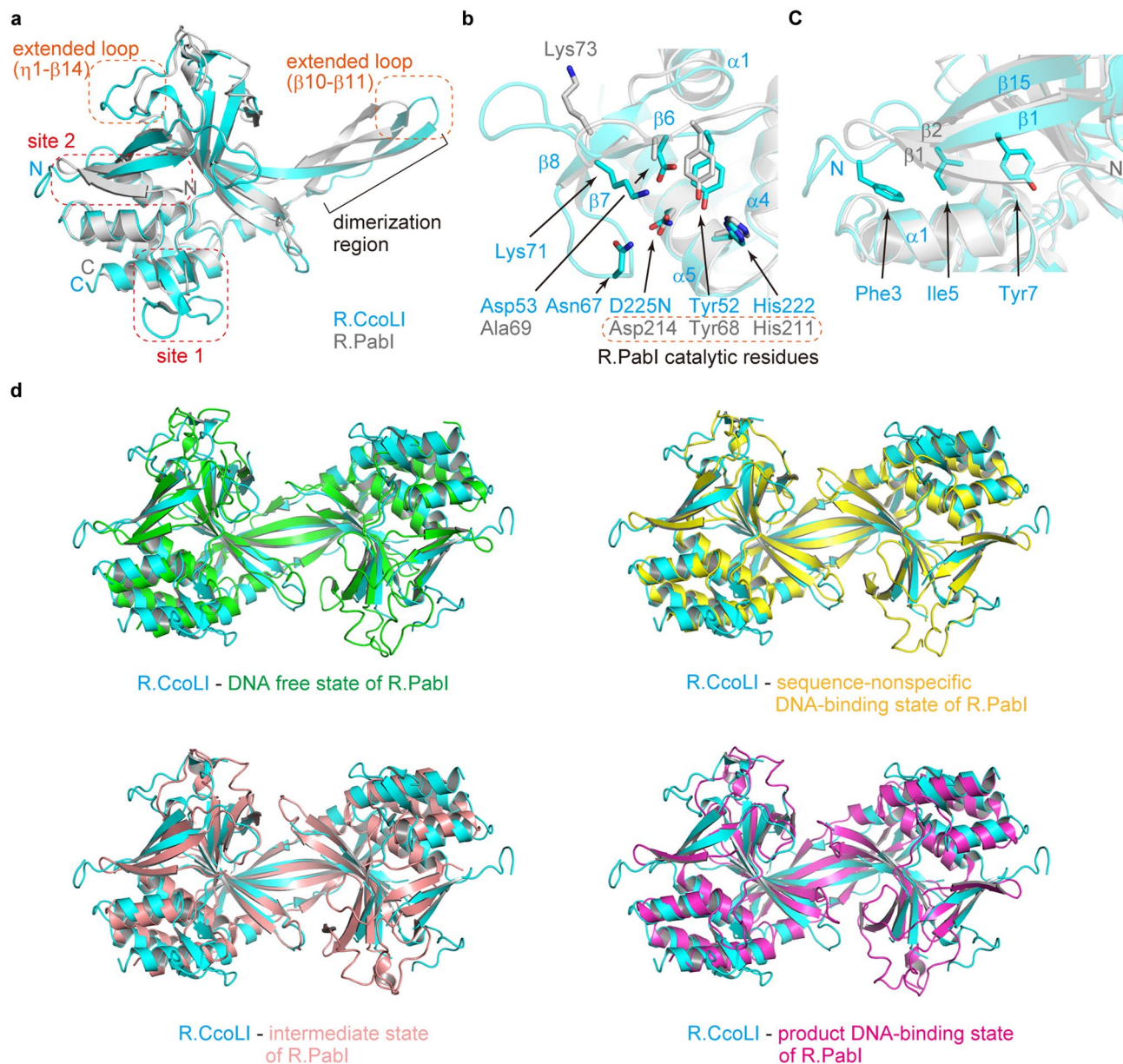


Figure 3. Comparison with R.PabI. **(a)** Characteristic structure of R.CcoLI. The structures of R.CcoLI protomer (chain A) and R.PabI protomer in the product DNA-binding state (PDB ID: 3WAZ, chain A) are superposed and are colored cyan and gray, respectively. Characteristic local structures in R.CcoLI that are not observed in R.PabI are indicated by red dotted squares (sites 1 and 2). The extended loops of R.CcoLI are indicated by orange dotted squares. **(b)** Structure of the insertion residues of R.CcoLI (site 1 in (A)). Residue numbers and secondary structures of R.CcoLI are labeled cyan. Residue numbers of R.PabI are labeled gray. **(c)** Structure of the N-terminal region of R.CcoLI (site 2 in (A)). Secondary structures of R.CcoLI and R.PabI are labeled cyan and gray, respectively. **(d)** Superposition of the dimeric structure of R.CcoLI (cyan) and those of R.PabI in the DNA-free state (PDB ID: 2DVY, green), the sequence-nonspecific DNA-binding state (5IFF, yellow), the intermediate state (PDB ID: 6L2O, salmon pink) and the product DNA-binding state (3WAZ, magenta).

Discussion

The HALFPIPE superfamily protein of restriction enzymes was first discovered in the hyperthermophilic archaea *P. abyssi* and was designated R.PabI. Structural studies have identified that HALFPIPE superfamily proteins are not restriction endonucleases, but rather are restriction DNA glycosylases. Because hyperthermophiles, including *P. abyssi*, grow optimally at high temperatures (over 80 °C), proteins from these organisms possess extremely high thermal stability; in fact, R.PabI cleaves dsDNA at temperatures ranging from 60 to 90 °C¹⁰. Mesophiles, such as *Campylobacter* and *Helicobacter*, also have HALFPIPE superfamily proteins. Although R.PabI is still active above 80 °C, R.CcoLI from *C. coli* is denatured at temperatures above approximately 60 °C (Fig. 1c). The

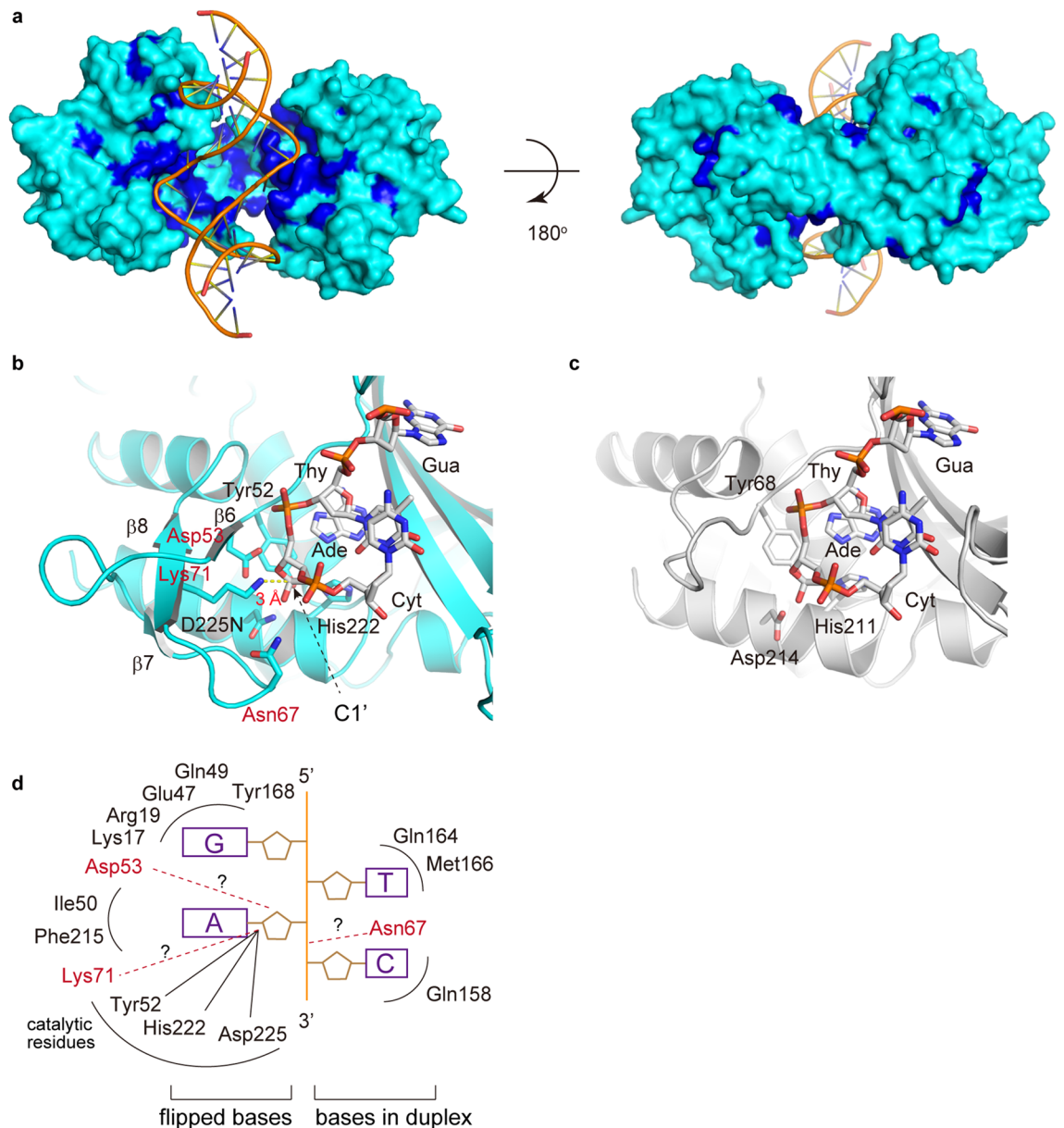


Figure 4. Active site structure. (a) R.CcoLI-DNA binding model. The model was prepared using the coordinates of R.CcoLI and the R.PabI-product DNA complex (PDB ID: 3WAZ). The sequence conservation of Fig. 1A is superposed on the molecular surface of R.CcoLI. The molecular surface of invariant residues is colored blue. The molecular surface of the other residues is colored cyan. The DNA model is shown as an orange cartoon. (b) DNA binding model of R.CcoLI around the active site. R.CcoLI is colored cyan. The DNA model of the sequence 5'-GTAC-3' is shown using a white stick model. The active site residues and residues near the catalytic site are shown by stick models. Residues in the characteristic antiparallel β -sheet region of R.CcoLI are labeled red. (c) Product DNA-binding structure of R.PabI around the active site. R.PabI is colored gray. (d) Plausible DNA recognition mechanism of R.CcoLI. The DNA backbone is shown as an orange line. R.CcoLI residues that are conserved in R.PabI are colored black. R.CcoLI-specific residues are colored red. DNA bases are colored purple.

most striking difference between R.PabI and R.CcoLI is that R.CcoLI cleaves the DNA backbone at medium temperatures, while the DNA backbone cleavage activity of R.PabI is very low at medium temperatures¹¹. Both R.PabI and R.CcoLI are members of the HALFPIPE superfamily of restriction enzymes. However, at medium temperatures, R.PabI acts as a monofunctional DNA glycosylase, while R.CcoLI acts as a bifunctional DNA glycosylase. To demonstrate the structural basis for the functional switching of these proteins, that is, the switching between the monofunctional DNA glycosylase and the bifunctional DNA glycosylase, we determined the crystal structure of the R.CcoLI(C189S-D225N) mutant. The most striking feature of the R.CcoLI structure is that the insertion residues that are not conserved in R.PabI form the characteristic antiparallel β -sheet structure ($\beta 6$, $\beta 7$ and $\beta 8$)

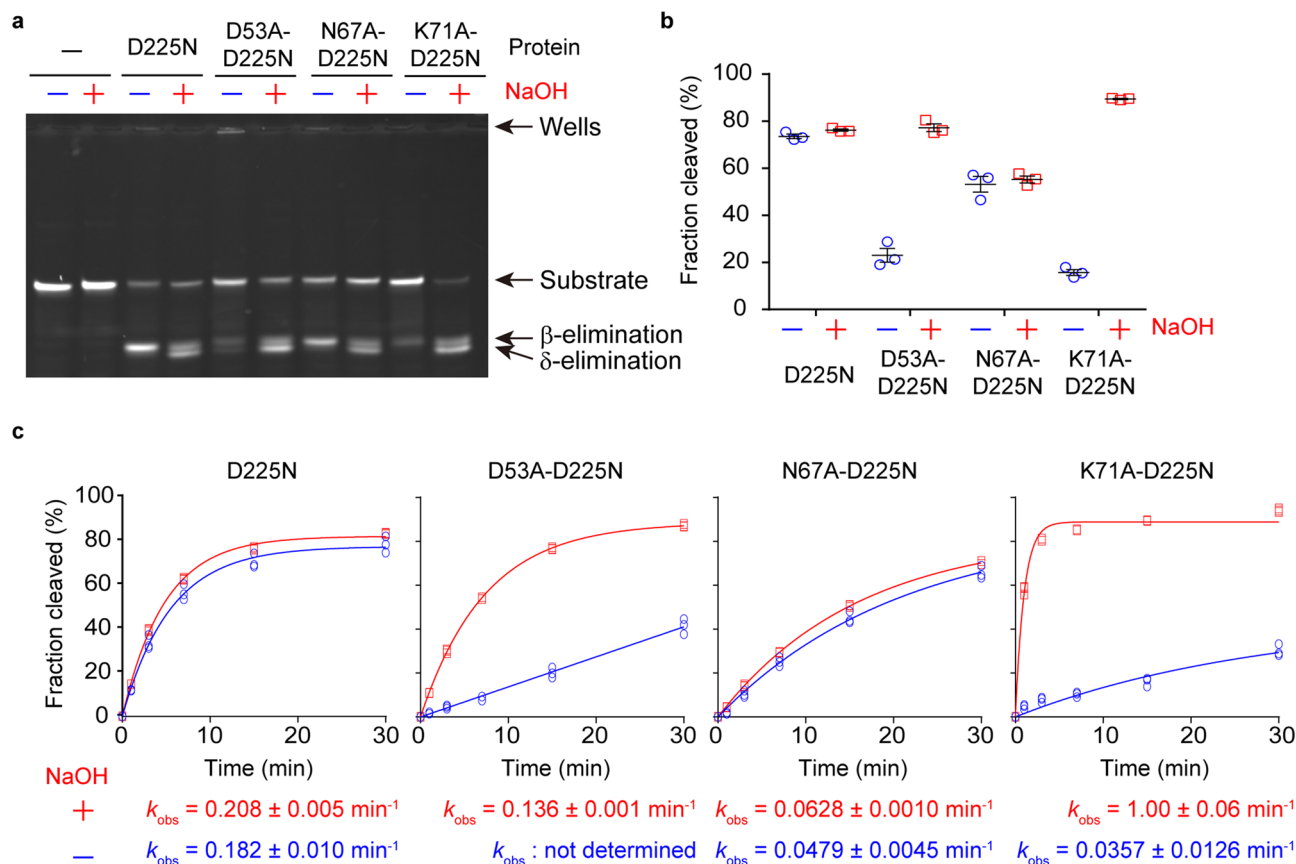


Figure 5. Mutation assay. **(a)** DNA glycosylase activity assay of the R.CcoLI mutants at 37 °C for 15 min. After the reaction, half of the solution was treated with NaOH to generate strand breaks at AP sites (β - and δ -eliminations). Data are representative of three independent experiments. The image is a cropped gel image. The full image is in Supplementary Fig. S7 online. Fraction cleaved = volume of cleaved DNA / (volume of cleaved DNA + volume of uncleaved DNA + volume of DNA stuck in the well) **(b)** Quantification of **a**. The fractions of products are indicated as blue circles and red squares. Data are the means \pm SEM from three independent experiments. **(c)** Kinetics experiments for the DNA glycosylase activities of the R.CcoLI mutants at 37 °C. The enzymatic rate constant and its standard error were obtained from a single-exponential fit to the data from three independent measurements.

(Fig. 3a,b). Due to the formation of the antiparallel β -sheet structure, the side-chain amine group of Lys71 is located near the *N*-glycosidic bond of deoxyadenosine (Fig. 4b). Lys71 of R.CcoLI is conserved in R.PabI as Lys73. The superposition of the structures of R.CcoLI and R.PabI shows that the C α atom of R.CcoLI Lys71 is close to that of R.PabI Lys73 (~ 3 Å). However, the R.CcoLI-specific β -sheet formation causes the inversion of the side chain direction of Lys71 (Fig. 3b). The enzymatic activity assays demonstrated that Lys71 of R.CcoLI is important for AP lyase activity (Fig. 5a–c). These results indicate that the AP lyase activity of R.CcoLI is facilitated by the insertion of the antiparallel β -sheet structure near the active site. The amino acid sequence of the characteristic antiparallel β -sheet structure of R.CcoLI is largely conserved in R.HpyAXII (Fig. 1a), which is the R.PabI homolog from mesophiles that also shows significant AP lyase activity¹⁵. Although the structure of R. HpyAXII has not been determined, this sequence similarity suggests that the corresponding region of R. HpyAXII forms an antiparallel β -sheet structure similar to R.CcoLI. The insertion of the antiparallel β -sheet structure in this region is predicted to be a signature of the HALFPIPE superfamily enzymes with AP lyase activity (bifunctional DNA glycosylase). As mentioned in the results section, the K71A-D225N mutant showed weak DNA cleavage activity. R.CcoLI is also predicted to cleave dsDNA through a Lys71-independent mechanism. However, the Lys71-independent DNA cleavage mechanism of R.CcoLI has not been elucidated to date.

In general, the thermostabilities of proteins have been attributed to several factors: increased numbers of ion-pair networks on protein surfaces, loop shortening and decreased numbers of hydrophobic accessible surface areas^{29–33}. In the R.PabI structure, loop regions that correspond to the β 10– β 11 and η 1– β 14 loops of R.CcoLI are shortened compared to the structure of R.CcoLI (Fig. 3a). These loop shortenings are predicted to be important for the high thermostability of R.PabI. The existence of the additional antiparallel β -sheet structure near the active site is characteristic of R.CcoLI. However, this structure is truncated in R.PabI (Fig. 3a,b). In the R.CcoLI structure, this region shows relatively high temperature factors compared to the protein core region (Fig. 6a,b and Supplementary Fig. S5 online). It is predicted that the truncation of this region is also important for the high thermostability of R.PabI. In the R.CcoLI structure, the hydrophobic residues in the N-terminal region (Phe3, Ile5 and Tyr7) are exposed to the solvent. In contrast, the corresponding region of R.PabI is covered by the additional

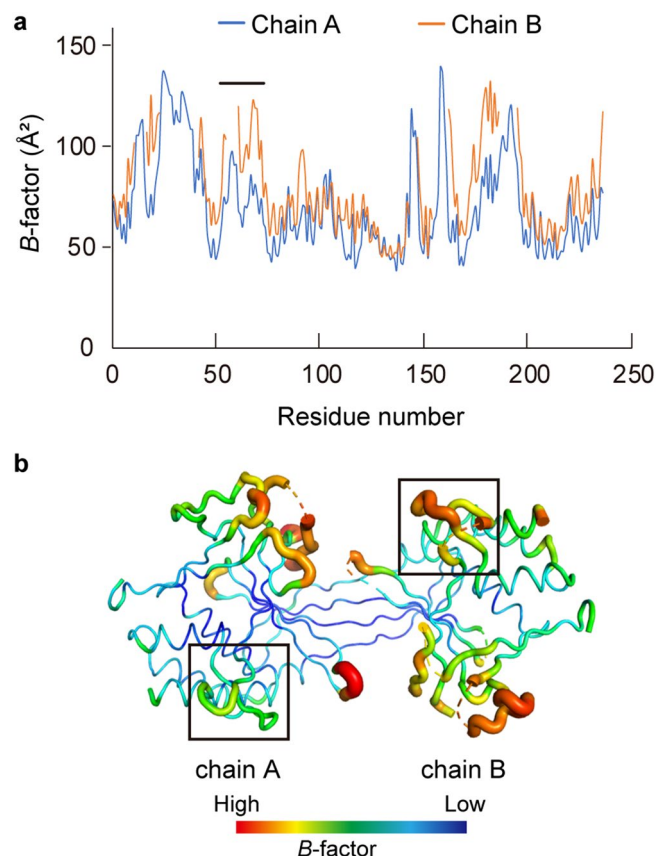


Figure 6. Temperature factor of R.CcoLI. **(a)** Plot of the average B -factor of each residue. The position of the characteristic antiparallel β -sheet region is indicated by a black line. **(b)** The R.CcoLI structure is colored according to the B -factors. The positions of the characteristic antiparallel β -sheet regions are indicated by black boxes.

β -strand (Fig. 3c). Therefore, the hydrophobic surface area of this region is decreased in the R.PabI structure. It is predicted that this difference contributes to the high thermostability of R.PabI. The AP lyase activity of R.CcoLI is mediated by Lys71 in the characteristic antiparallel β -sheet structure. In contrast, R.PabI only shows weak AP lyase activity due to the lack of a lysine amine group near the active site (Fig. 3b). Because *C. coli* is a mesophile, high thermostability is not necessary for R.CcoLI; R.CcoLI can utilize the relatively flexible regions (that is, the antiparallel β -sheet structure containing Lys71) for its catalytic mechanism. Meanwhile, R.PabI from the hyperthermophile *P. abyssi* must possess high thermostability to function at high temperature. Because AP sites are unstable in high-temperature conditions¹¹, AP lyase activity is predicted not to be required for the DNA damaging function of R.PabI. R.PabI might have relinquished its AP lyase activity to obtain high thermostability.

The structures of restriction endonucleases are frequently modified by the binding of DNA. For example, the structures of such restriction endonucleases as EcoRV and BamHI, which belong to the PD-(D/E)XK superfamily, show that these proteins widen their DNA binding clefts when they weakly bind sequence-nonspecific dsDNA, and the clefts become narrow when they tightly bind their recognition sequences in dsDNA; the weak sequence-nonspecific dsDNA-binding states are utilized for facilitated diffusions on dsDNA^{3,4,34}. Our previous studies also demonstrated that the structure of R.PabI is modified by the binding of dsDNA^{11,13,14}. Although the structure of R.CcoLI was determined in the absence of DNA, the R.CcoLI structure is most similar to the R.PabI structure in the sequence-nonspecific dsDNA-binding state (Fig. 3d and Supplementary Fig. S6 online). This structural similarity might indicate that the R.CcoLI structure is not modified by the binding of sequence-nonspecific dsDNA. However, the precise DNA recognition mechanism of R.CcoLI will be clarified by the determination of the R.CcoLI-dsDNA complex structure. In this study, we analyzed the structure and function of R.CcoLI using the D225N mutant. Structural and functional studies utilizing the wild-type enzyme may demonstrate the function of R.CcoLI more precisely.

Methods

Expression and purification. The gene fragment of R.CcoLI (NCBI Reference Sequence: WP_002830209) was synthesized by GenScript (see Supplementary Table S1 online). Each codon in the synthesized gene was optimized for expression in *E. coli*. The R.CcoLI gene fragment was amplified by PCR using primers in Supplementary Table S2 online (Cloning-F, R), and it was cloned into the SmaI-HindIII site of pET48b (pET48b-

R.CcoLI) to express R.CcoLI with an N-terminal thioredoxin tag. To reduce the cytotoxicity of R.CcoLI, the D225N mutation, which corresponds to the D214N mutation of R.PabI (Fig. 1a), was introduced to pET48b-R.CcoLI plasmid using the PrimeSTAR Mutagenesis Basal Kit (TAKARA) and primers (D225N-F, R) in Supplementary Table S2 online (pET48b-R.CcoLI(D225N)). For R.CcoLI expression, the pET48b-R.CcoLI(D225N) plasmid was transformed into *E. coli* Rosetta (DE3) pLysS (Novagen). The recombinant *E. coli* cells were cultivated at 37 °C in LB medium supplemented with 20 µg/ml kanamycin and 50 µg/ml chloramphenicol until the optical density of the medium at 600 nm reached 0.8. The expression of R.CcoLI(D225N) was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.1 mM. After overnight cultivation at 18 °C, the cells were harvested by centrifugation at 5000 × g for 10 min at 4 °C.

The harvested cells were resuspended in 50 mM Tris HCl pH 8.0, 10 mM imidazole and 1 mM Tris(2-carboxyethyl)phosphine (TCEP) and lysed by sonication. After centrifugation at 40,000 × g for 30 min at 4 °C, the supernatant was mixed with 2 ml of Ni-NTA Superflow resin (QIAGEN). The resin was washed using 20 ml of 50 mM Tris HCl pH 8.0, 20 mM imidazole and 1 mM TCEP, and the bound protein was eluted using 10 ml of 50 mM MES pH 6.0, 200 mM imidazole, 50 mM MgCl₂ and 1 mM TCEP. The eluted protein was treated with HRV3C protease at 4 °C overnight to remove the N-terminal thioredoxin tag. The protein that the expression tag had been removed from was purified using a Mono S HR 10/10 (GE Healthcare) column pre-equilibrated with 10 mM MES pH 6.0, 200 mM NaCl and 1 mM TCEP, and the protein was eluted with a linear gradient of 0.2–1.5 M NaCl. The protein was further purified using a Superdex 200 h 10/30 (GE Healthcare) column pre-equilibrated with 10 mM MES pH 6.0, 200 mM NaCl and 1 mM TCEP. The protein sample was stored at –80 °C until use.

Expression vectors of R.CcoLI(D225N) mutants (the D53A-D225N, N67A-D225N, K71A-D225N and C189S-D225N mutants) were prepared using the PrimeSTAR Mutagenesis Basal Kit (TAKARA) and primers in Supplementary Table S2 online. Each mutant was expressed and purified using the same method as that described above.

Oligomeric state analysis by gel filtration chromatography. The samples (2.3 µM) were loaded onto a Superdex 200 h 10/30 (GE Healthcare) column and eluted with buffer containing 10 mM MES pH 6.0, 200 mM NaCl and 1 mM TCEP. To estimate the oligomeric state of R.CcoLI, the following standard proteins were used: aldolase ($M_r = 158,000$), conalbumin ($M_r = 75,000$), chymotrypsinogen A ($M_r = 25,000$), and ribonuclease A ($M_r = 13,700$).

DNA cleavage assay. A modified pET26b plasmid, possessing only one 5'-GTAC-3' site, was employed as a substrate for R.CcoLI (see Supplementary Fig. S4 online)¹¹. The modified pET26b plasmid was cut with HindIII (TAKARA) to linearize the plasmid. To analyze the DNA cleavage activity of R.CcoLI, 0.2 µg of R.CcoLI mutants and 0.2 µg of the linearized plasmid were mixed in 0.1 M sodium phosphate buffer pH 6.5 and 1 mM TCEP and were incubated at 37 °C for 30 min. The cleaved DNAs were separated by electrophoresis through a 1% agarose gel. The DNAs were visualized with blue-LED light after GelGreen (Biotium) staining. Products by AfaI (TAKARA), which cleaves the sequence 5'-GTAC-3', were also separated as a control.

DNA glycosylase activity assay. DNA glycosylase activity assays of R.CcoLI mutants were performed using 24-bp dsDNA containing one 5'-GTAC-3' sequence (5'-fluorescein-GGATGCATGAGTACGAGGACCATC-3', see Supplementary Fig. S4 online). A total of 0.2 µM of the substrate dsDNA was mixed with 0.8 µM of the R.CcoLI dimer in a reaction buffer (0.1 M sodium phosphate buffer pH 6.5, 1 mM TCEP). The reaction solutions were incubated at 37 °C for 15 min or for 1, 3, 7, 15 and 30 min. After the enzymatic reaction, the solutions were supplemented with 0.1 M NaOH or 0.1 M HCl to terminate the enzymatic reaction. To cleave the 5' and 3' sides of the AP sites generated by R.CcoLI, the solutions supplemented with NaOH were heated at 70 °C for 10 min. The reaction solutions were neutralized by the addition of an equal concentration of HCl or NaOH and separated on a denaturing 18% polyacrylamide gel in 0.5 × TBE and 7 M urea. The fluorescence was measured using an Amersham Imager 680 (GE Healthcare). Data were quantified using the program Amersham Imager 680 Analysis Software (GE Healthcare). The enzymatic rate constant k_{obs} was obtained from a single-exponential fit to the data from three independent measurements: $f_p = f_{p,\text{max}} \times (1 - e^{-k_{\text{obs}}t})$, where f_p is the fraction of product, $f_{p,\text{max}}$ is the maximum value of f_p , and t is the time of the reaction (min).

Denaturation assay. For the denaturation assay of R.CcoLI, proteins (10 µM) and 2.5 × SYPRO Orange (Thermo Fisher Scientific) were mixed in 0.1 M sodium phosphate buffer pH 6.5 and 1 mM TCEP. Denaturation assays were performed using a CFX Connect Real-Time PCR Detection System (Bio-Rad Laboratories). Fluorescence was measured from 20 to 95 °C in 0.5 °C steps (excitation, 450–490 nm; detection, 560–580 nm). Data were analyzed using Bio-Rad CFX Manager 3.0 software.

Crystallization, data collection and structure determination. The purified protein was concentrated to 18 mg/ml using Vivaspin 6 (MWCO 30 k Da, Sartorius) for crystallization. Crystallization experiments of the R.CcoLI(C189S-D225N) mutant were performed using the sitting-drop vapor-diffusion method at 20 °C. The crystals of R.CcoLI(C189S-D225N) were obtained using a reservoir solution of 0.1 M MES pH 6.0 and 8% PEG6000 one day later. X-ray diffraction data of the R.CcoLI(C189S-D225N) crystal were collected on the AR-NW12A beamline at the Photon Factory (Tsukuba, Japan) under cryogenic conditions (95 K). For cryoprotection, the R.CcoLI(C189S-D225N) crystal was soaked in a reservoir solution supplemented with 40% ethylene glycol for several seconds. The R.CcoLI(C189S-D225N) crystal diffracted X-rays to 2.35-Å resolution. The X-ray diffraction data were indexed, integrated and scaled with XDS³⁵. The R.CcoLI(C189S-D225N) crystal belonged to the space group $P2_12_1$ with unit cell parameters of $a = 66.49$ Å, $b = 89.60$ Å and $c = 100.66$ Å. The

crystal contains two R.CcoLI(C189S-D225N) molecules per asymmetric unit according to the Matthews coefficient ($V_M = 2.70 \text{ \AA}^3 \text{ Da}^{-1}$)³⁶.

The initial model of R.CcoLI(C189S-D225N) was determined by the molecular replacement method using the program Phaser³⁷. The ensemble of the R.PabI structures (the DNA-free state (PDB code: 2DVY)⁹, the product DNA-binding state (PDB code: 3WAZ)¹¹ and the sequence-nonspecific DNA-binding state (PDB code: 5IFF)¹³) was used as the search model. The initial model was refined and rebuilt using the programs Phenix.refine³⁸ and Coot³⁹. The final model of R.CcoLI(C189S-D225N) was refined to 2.35 Å resolution with R and R_{free} values of 22.9% and 25.9%, respectively. The geometry of the final model was evaluated using the program MolProbity⁴⁰. In the Ramachandran plot, 98.0% of the residues were in the favored region, and the rest were in the allowed region. The data collection and refinement statistics are summarized in Table 1.

Computational analysis. The protein structure was analyzed using a set of computer programs as follows: PISA⁴¹ for the analysis of the protein interface, surface and assemblies, APBS⁴² for the calculation of macromolecular electrostatics, Dali for the search for similar structures from the database²⁸, Clustal Omega⁴³ for the amino acid sequence alignment, ESprout⁴⁴ for the preparation of alignment figure, and PyMOL (<http://pymol.org>) for the depiction of structures.

Data availability

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession number 7CFA.

Received: 16 July 2020; Accepted: 9 December 2020

Published online: 13 January 2021

References

- Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A., Wilson, G. G. & Murray, N. E. Highlights of the DNA cutters: A short history of the restriction enzymes. *Nucleic Acids Res.* **42**, 3–19 (2014).
- Orlowski, J. & Bujnicki, J. M. Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res.* **36**, 3552–3569 (2008).
- Newman, M., Strzelecka, T., Dörner, L. F., Schildkraut, I. & Aggarwal, A. K. Structure of Bam HI endonuclease bound to DNA: Partial folding and unfolding on DNA binding. *Science* **269**, 656–663 (1995).
- Winkler, F. K. *et al.* The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12**, 1781–1795 (1993).
- Shen, B. W. *et al.* Unusual target site disruption by the rare-cutting HNH restriction endonuclease PacI. *Structure* **18**, 734–743 (2010).
- Saravanan, M., Bujnicki, J. M., Cymerman, I. A., Rao, D. N. & Nagaraja, V. Type II restriction endonuclease R.KpnI is a member of the HNH nuclease superfamily. *Nucleic Acids Res.* **32**, 6129–6135 (2004).
- Grazulis, S. *et al.* Structure of the metal-independent restriction enzyme BfiI reveals fusion of a specific DNA-binding domain with a nonspecific nuclease. *Proc. Natl. Acad. Sci. U S A* **102**, 15797–15802 (2005).
- Sokolowska, M., Czapinska, H. & Bochtler, M. Hpy188I-DNA pre- and post-cleavage complexes—snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res.* **39**, 1554–1564 (2011).
- Miyazono, K. *et al.* Novel protein fold discovered in the PabI family of restriction enzymes. *Nucleic Acids Res.* **35**, 1908–1918 (2007).
- Ishikawa, K. *et al.* Discovery of a novel restriction endonuclease by genome comparison and application of a wheat-germ-based cell-free translation assay: PabI (5'-GTA/C) from the hyperthermophilic archaeon *Pyrococcus abyssi*. *Nucleic Acids Res.* **33**, e112 (2005).
- Miyazono, K. *et al.* A sequence-specific DNA glycosylase mediates restriction-modification in *Pyrococcus abyssi*. *Nat. Commun.* **5**, 3178 (2014).
- Fukuyo, M. *et al.* Restriction-modification system with methyl-inhibited base excision and abasic-site cleavage activities. *Nucleic Acids Res.* **43**, 2841–2852 (2015).
- Wang, D., Miyazono, K. & Tanokura, M. Tetrameric structure of the restriction DNA glycosylase R.PabI in complex with nonspecific double-stranded DNA. *Sci. Rep.* **6**, 35197 (2016).
- Miyazono, K., Wang, D., Ito, T. & Tanokura, M. Distortion of double-stranded DNA structure by the binding of the restriction DNA glycosylase R.PabI. *Nucleic Acids Res.* **48**, 5106–5118 (2020).
- Zhang, Y. *et al.* Restriction glycosylases: Involvement of endonuclease activities in the restriction process. *Nucleic Acids Res.* **45**, 1392–1403 (2017).
- Wallace, S. S. Base excision repair: A critical player in many games. *DNA Repair (Amst)*. **19**, 14–26 (2014).
- Jacobs, A. L. & Schär, P. DNA glycosylases: In DNA repair and beyond. *Chromosoma* **121**, 1–20 (2012).
- Parikh, S. S. *et al.* Uracil-DNA glycosylase-DNA substrate and product structures: Conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl. Acad. Sci. U S A* **97**, 5083–5088 (2000).
- Schormann, N., Ricciardi, R. & Chattopadhyay, D. Uracil-DNA glycosylases—Structural and functional perspectives on an essential family of DNA repair enzymes. *Protein Sci.* **23**, 1667–1685 (2014).
- Bruner, S. D., Norman, D. P. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, 859–866 (2000).
- Banerjee, A., Yang, W., Karplus, M. & Verdine, G. L. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **434**, 612–618 (2005).
- Golan, G. *et al.* Structure of T4 pyrimidine dimer glycosylase in a reduced imine covalent complex with abasic site-containing DNA. *J. Mol. Biol.* **362**, 241–258 (2006).
- Gilboa, R. *et al.* Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *J. Biol. Chem.* **277**, 19811–19816 (2002).
- Fromme, J. C. & Verdine, G. L. Structure of a trapped endonuclease III-DNA covalent intermediate. *EMBO J.* **22**, 3461–3471 (2003).
- Fromme, J. C., Bruner, S. D., Yang, W., Karplus, M. & Verdine, G. L. Product-assisted catalysis in base-excision DNA repair. *Nat. Struct. Biol.* **10**, 204–211 (2003).
- Mullins, E. A., Rodriguez, A. A., Bradley, N. P. & Eichman, B. F. Emerging roles of DNA glycosylases and the base excision repair pathway. *Trends Biochem. Sci.* **44**, 765–781 (2019).
- Drohat, A. C. & Coey, C. T. Role of base excision ‘repair’ enzymes in erasing epigenetic marks from DNA. *Chem. Rev.* **116**, 12711–12729 (2016).

28. Holm, L. & Sander, C. Dali: A network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
29. Karshikoff, A. & Ladenstein, R. Ion pairs and the thermotolerance of proteins from hyperthermophiles: A ‘traffic rule’ for hot roads. *Trends Biochem. Sci.* **26**, 550–556 (2001).
30. Yip, K. S. *et al.* Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamate dehydrogenase family. *Eur. J. Biochem.* **255**, 336–346 (1998).
31. Jaenicke, R. & Böhm, G. The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**, 738–748 (1998).
32. Zhang, X., Meining, W., Fischer, M., Bacher, A. & Ladenstein, R. X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile *Aquifex aeolicus* at 1.6 Å resolution: determinants of thermostability revealed from structural comparisons. *J. Mol. Biol.* **306**, 1099–1114 (2001).
33. Miyazono, K., Sawano, Y. & Tanokura, M. Crystal structure and structural stability of acylphosphatase from hyperthermophilic archaeon *Pyrococcus horikoshii* OT3. *Proteins* **80B**, 439–442 (2004).
34. Viadiu, H. & Aggarwal, A. K. Structure of BamHI bound to nonspecific DNA: A model for DNA sliding. *Mol. Cell* **5**, 889–895 (2000).
35. Kabsch, W. X. D. S. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
36. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).
37. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
38. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
40. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).
41. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
42. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U S A* **98**, 10037–10041 (2001).
43. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
44. Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).

Acknowledgements

The synchrotron radiation experiments were performed at AR-NW12A in the Photon Factory (Tsukuba, Japan) (2016G651). This work was supported by JSPS KAKENHI (Grant Number 26712012).

Author contributions

K.M., D.W. and T.I. conducted the experiments. All the research processes were conducted under supervision of M.T. The draft of manuscript was written by K.M. and revised by M.T.

Competing interests

The authors declare no competing interests.


Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-79537-y>.

Correspondence and requests for materials should be addressed to M.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021