博士論文

Two crystal structures of restriction DNA glycosylase R.PabI in complex with DNA reveal its mechanism of searching for and interrogating the recognition sequence

(制限 DNA グリコシラーゼ R.PabI による認識部位
探索の構造基盤解明)

王　徳龍

博士論文

Two crystal structures of restriction DNA glycosylase R.PabI in complex with DNA reveal its mechanism of searching for and interrogating the recognition sequence

(制限 DNA グリコシラーゼ R.PabI による認識部位探索の構造基盤解明)

東京大学大学院農学生命科学研究科

応用生命化学専攻　食品生物構造学研究室

平成 27 年入学　王　徳龍

指導教員　永田　宏次

Table of Content

Abbreviations

BIS-TRIS: 2-Bis(2-hydroxyethyl)amino-2-(hydroxymethyl)-1,3-propanediol

DNA: Deoxyribonucleic acid

dsDNA: double-stranded DNA

IPTG: Isopropyl-β-D(-)-thiogalactoyranoside

LB: Lysogeny Broth

MES: 2-(*N*-morpholino)ethanesulfonic acid

PCR: Polymerase chain reaction

PDB: Protein Data Bank

PEG: Polyethyleneglycol

r.m.s.d.: Root mean square deviation

SDS: sodium dodecyl sulfate

SDS-PAGE: SDS polyacrylamide gel electrophoresis

ssDNA: Single-stranded DNA

The expressions of amino acid residue and DNA base

Amino acid residues are written as single alphabets; DNA bases are written as three alphabets.

Preface

R.PabI was discovered as a type II restriction enzyme that recognizes the 5'-GTAC-3' sequence and belongs to the HALFPIPE superfamily. Although most restriction enzymes cleave phosphodiester bonds at specific sites by hydrolysis, R.PabI flips the guanine and adenine bases of the recognition sequence out of the DNA helix and hydrolyzes the *N*-glycosidic bond of the flipped adenine in a similar manner to DNA glycosylases. Thus R.PabI has been termed as restriction DNA glycosylase which recognizes specific DNA sequence as a normal restriction enzyme while using glycosylase activity to cleave the sequence. Thereafter, questions are raised about by which mechanism does this novel restriction enzyme search and interrogate its recognition sequence efficiently. In this study, two complex crystal structures have been solved: inactive R.PabI mutant (R32A E63A) in complex with 20bp nonspecific dsDNA and low-active R.PabI mutant (Y68F K154A) in complex with 23bp dsDNA which contains two recognition sequences. The former complex structure reflects a snapshot of R.PabI searching its recognition sequence, the later complex structure shows a binding mode in which R.PabI encounters its recognition sequence and starts to deform it to transform into the cleavage complex.

This thesis is written in two chapters. Chapter 1 mainly discusses 1) the solvation of the structure of the inactive R.PabI mutant (R32A E63A) in complex with 20bp nonspecific dsDNA; 2) the biochemistry experiments that would prove the hypothesis generated by the structure analysis. The complex structure was solved at 1.9 Å, this result, alone with the EMSA and the glycosylase activity assay, indicate that R.PabI binds to nonspecific DNA using a ring-like tetrameric structure, such a structure would facilitate its moving on the DNA. Chapter 2 mainly discusses 1) the solvation of the structure of the low-active R.PabI mutant (K154A Y68F) in complex with 23bp specific dsDNA, 2) the biochemistry experiments that would prove the hypothesis generated by the structure

analysis. The complex structure, which was solved at 2.4 Å, alone with the results of the glycosylase activity assays indicate that R.PabI uses a stepwise mechanism to deform the recognition sequence after encountering it, furthermore, the deformability of the T-A step in the target sequence would facilitate the recognition by R.PabI thus to make the interrogation-recognition process be undertaken more efficiently.

In this research, I have solved two crystal structures of restriction DNA glycosylase R.PabI in complex with DNA, the mechanism by which R.PabI searches and interrogates its recognition sequence have been speculated based on the crystal structures as well as the biochemical experiments related to them. Since no previous research had ever studied the searching, interrogation and cleavage mechanism of the HALFPIPE superfamily of the typeII restriction, this research helps better understanding this previously unstudied superfamily. In addition, since R.PabI recognizes its substrate similar to restriction enzyme while cleaves DNA in a manner closed to glycosylase, elucidation of the recognition mechanism of such an unique enzyme would enlarge our understanding about the protein-DNA interaction.

1. Introduction

1-1 TypeII restriction enzyme and restriction DNA glycosylase R.PabI

A type II restriction enzyme recognizes a specific double-stranded DNA (dsDNA) sequence and cleaves dsDNA at or near the sequence. Because most type II restriction enzymes cleave the phosphodiester bonds of dsDNA by hydrolysis in a sequence-dependent manner, they are also called restriction endonucleases[1]. Restriction enzymes are widely used in the field of biotechnology, such as in gene recombination, to cut dsDNA at specific sites for manipulation or analysis. Based on their structural features, type II restriction enzymes are categorized into several superfamilies: the PD-(D/E)XK superfamily[2–4], the HNH superfamily[5,6], the PLD superfamily[7], the GIY-YIG superfamily[8,9], and the HALFPIPE superfamily[10,11]. Although most restriction enzymes require Mg^{2+} ions for their dsDNA cleavage activities, enzymes within the PLD superfamily and the HALFPIPE superfamily cleave dsDNA without the addition of a divalent cation[7,12]. Restriction enzymes of the PLD superfamily cleave dsDNA using the phospholipase D-like active site, that which not require an Mg^{2+} ion to hydrolyze phosphodiester bonds[7]. On the other hand, restriction enzymes of the HALFPIPE superfamily do not cleave the phosphodiester bonds of dsDNA but cleave the *N*-glycosidic bonds of bases in a manner similar to DNA glycosylases[13].

Since it is already well known that the genes of restriction methylation system always transfer among genomes of different species, therefore the identification of genes of novel restriction methylation system could be performed by comparing the sequence of genes of different related species. If the sequence of one unknown gene is assessed of not being conservative comparing with the other related species, the possibility that this gene is transformed from outside is high, thus this gene might contain restriction methylation system. Utilizing such a mechanism, R.PabI was discovered by comparing

the genome of hyperthermophilic archaea *Pyrococcus abyssi* and its related species *P.horikoshii*[12]. R.PabI was discovered as a type II restriction enzyme and belongs to the HALFPIPE superfamily[10]. R.PabI homologs are only conserved among a hyperthermophilic archaea (*Staphylothermus hellenicus*), a thermophilic bacterium (*Caloramator australicus*), and some mesophilic bacteria such as *Helicobacter* and *Campylobacter* (Fig. 1). R.PabI contains 226 amino acid residues (Molecular weight: 26011), it recognizes the 5'-GTAC-3' sequence and can cleave the *N*-glycosidic bond of the adenine in the recognition sequence. Because the opposing apurinic/apyrimidinic (AP) sites generated by R.PabI are cleaved by β elimination and/or the endogenous AP endonucleases of host cells, R.PabI can cleave dsDNA at a specific site, similarly to restriction endonucleases. The crystal structure of apo R.PabI was solved at 2007[10] and the structure of R.PabI in complex with specific dsDNA was solved at 2014[13]. The structural analysis of R.PabI showed that it forms a dimer and possesses a characteristic highly curved β sheet, called the half-pipe structure, at the protomer-protomer interface[10] (Fig. 2a,b). R.PabI recognizes the 5'-GTAC-3' sequence using the positively charged half-pipe structure. At the recognition sequence, R.PabI bends dsDNA by approximately 90° and flips the adenine and guanine bases out of the DNA helix to recognize the sequence. The *N*-glycosidic bond of the flipped adenine is cleaved by R.PabI by hydrolysis, similarly to DNA glycosylases[13] (Fig. 2c,d). However, since it is difficult to believe that R.PabI would sharply bend every base pair of DNA to search its recognization sequence, other mechanism must be used to search the recognition sequence. In addition, the mechanism by which R.PabI distinguishes its recognition sequence from the other nonspecific sites, as well as R.PabI deforms DNA to transform into the extrusion complex once it encounters its recognition sequence have yet been elucidated.

13

1-2 The mechanism of DNA binding proteins recognizing their target sequence

In prokaryotic cells, restriction enzymes are utilized as protection from imported exogenous DNA. Due to the biological function of restriction enzymes, cells would suffer from lethal problems if restriction enzymes could not find their targets efficiently. In addition to restriction enzymes, other DNA binding proteins must also find their target sites efficiently to maintain cellular function. Therefore, the mechanisms by which DNA binding proteins find their target sites in the "sea" of nonspecific sequences within a short time are important for living cells, and these mechanisms of DNA binding proteins have been analyzed by several approaches including dynamic simulation[14–16], structural analysis[2,4,17] and single-molecule observation[18–20]. The common hypothesis generated by these experiments is that proteins could not feasibly find their target sites directly; instead, they first bind non-specifically and adjust their positions to find their target sites efficiently (facilitated diffusion)[21–24]. Several mechanisms of facilitated diffusion have been proposed: sliding, hopping, and intersegmental transfer. In sliding, proteins bind dsDNA non-specifically and diffuse along it linearly to seek their target sites[25,26]. In hopping, proteins dissociate from dsDNA, then diffuse in the solution and re-associate with nearby dsDNA[27–29]. In intersegmental transfer, proteins bind two dsDNA segments and transfer directly from one site to the other[30,31].

Among restriction enzymes, the facilitated diffusion mechanisms of EcoRV and BamHI, which belong to the PD-(D/E)XK superfamily, have been well studied by several methods[2,4,17,29,32]. The structural analysis of the DNA free states, the specific dsDNA binding states, and the nonspecific dsDNA binding states of these enzymes showed that

these proteins widen their DNA binding clefts when they bind nonspecific dsDNA, and their DNA binding clefts become narrow when they transform their structures into the specific binding states at their recognition sequences (Fig. 3). These structural changes cause differences in the contacts between proteins and DNA; the contact surface areas of the specific dsDNA binding states are larger than the contact surface areas of the nonspecific dsDNA binding states. In addition, the nonspecific dsDNA binding states lack almost all of the base-specific interactions between proteins and DNA compared with the specific dsDNA binding states. The number of hydrogen bonds between proteins and phosphate groups of DNA are also decreased in the nonspecific dsDNA binding states compared with the specific dsDNA binding states. The wider DNA binding clefts, the smaller number of intermolecular hydrogen bonds, and the smaller contact surface area of the nonspecific dsDNA binding states of EcoRV and BamHI allow them to diffuse along dsDNA by lowering the activation energies for breaking and reforming DNA contacts[2,4,17]. Meanwhile, the mechanisms by which the HALFPIPE superfamily restriction enzymes find their recognition sequences have remained unclear because the structure of the nonspecific dsDNA binding state had not been determined.

1-3 The purpose of this research

In this study, the crystal structure of R.PabI in complex with 20bp dsDNA which does not contain recognition sequence has been solved to uncover how R.PabI interacts with nonspecific dsDNA. The structure of the complex showed that R.PabI forms a tetrameric structure to sandwich dsDNA and the tetrameric structure is stabilized by four salt bridges. An electrophoretic mobility shift assay (EMSA) and DNA glycosylase assay of R.PabI mutants showed that the residues that form the salt bridges facilitates the DNA cleavage

activity of R.PabI for dsDNA with an abundance of nonspecific sequences. The formation of the tetrameric structure is predicted to be important for R.PabI to find the recognition sequence efficiently in the sea of nonspecific sequences.
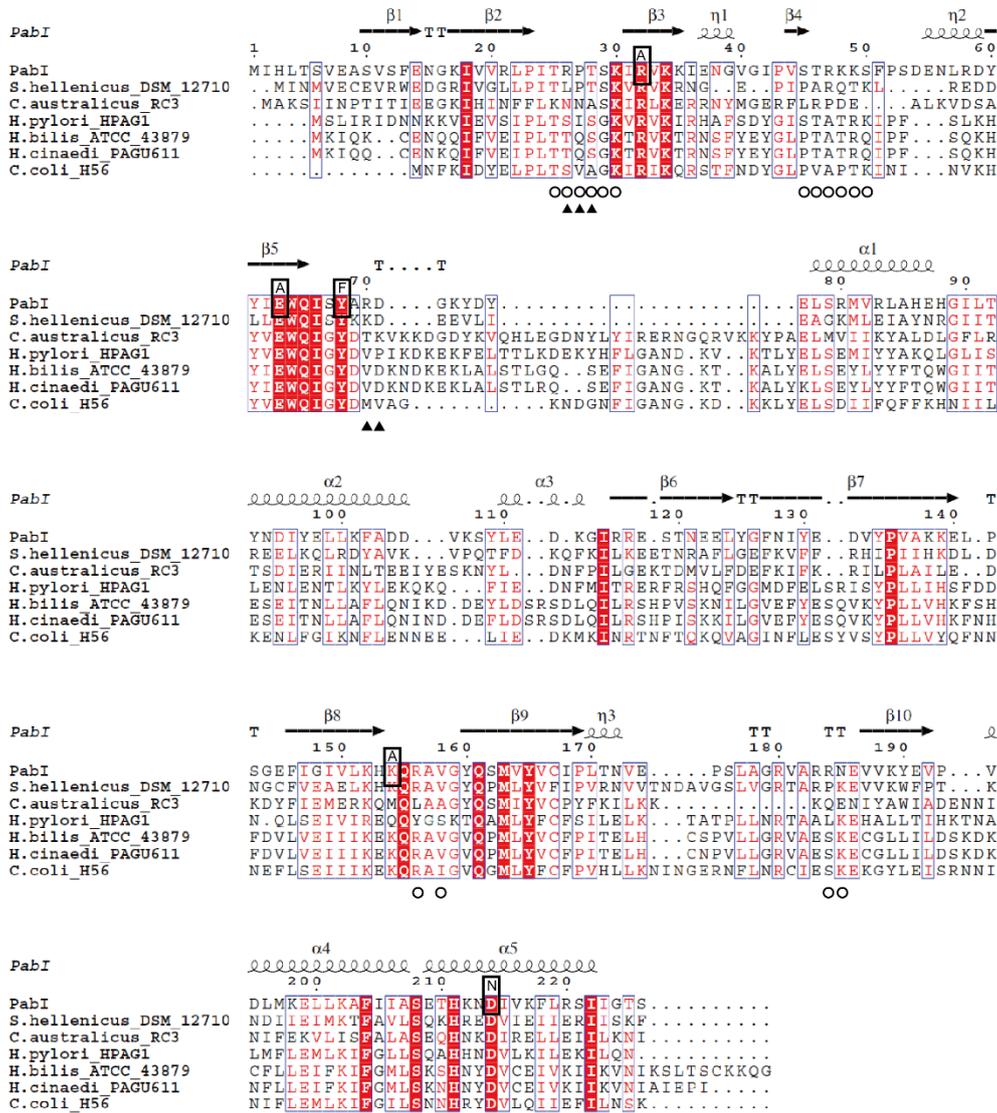
Figure 1. Amino acid sequence alignment of R.PabI and its homologues. Invariant residues are highlighted with red boxes, conserved residues are shown in red text. The secondary structure of R.PabI in the R.PabI-nonspecific dsDNA complex is indicated by helices (α and η (3₁₀)-helices), arrows (β-strand), and TT (β-turn). Residues that were mutated for the structure determination are indicated by black boxes. Residues that interact with the DNA bases and the sugar-phosphate backbones are marked with black and open circles, respectively. Residues analyzed by mutagenesis are marked with black triangles.
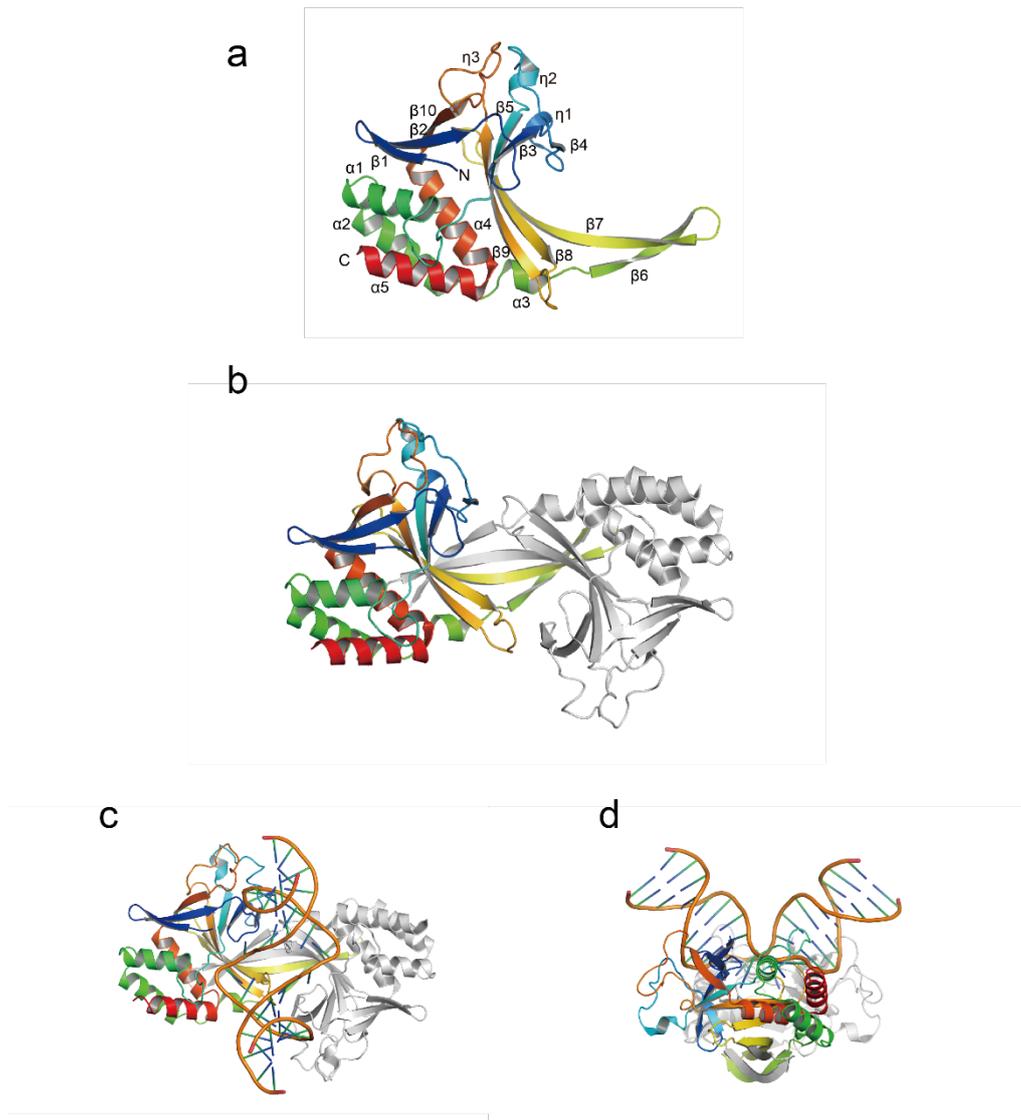
Figure 2. Crystal structures of the DNA-free R.PabI and the R.PabI-specific DNA complex. (**a**) Structure of one protomer of the DNA-free R.PabI. The protomer is colored blue (in the N terminus) to red (in the C-terminus). Secondary structure assignments are labelled on the model. (**b**) Structure of the dimer of the DNA-free R.PabI. One protomer is colored in the same way as (**a**), one protomer is colored grey. (**c**) (**d**) Structure of the R.PabI-specific DNA complex. (**c**) is seen from above of the complex; (**d**) is seen from one side of the complex. The bound dsDNA is colored orange. Each protomer in the complex is colored in the same way as (**b**).
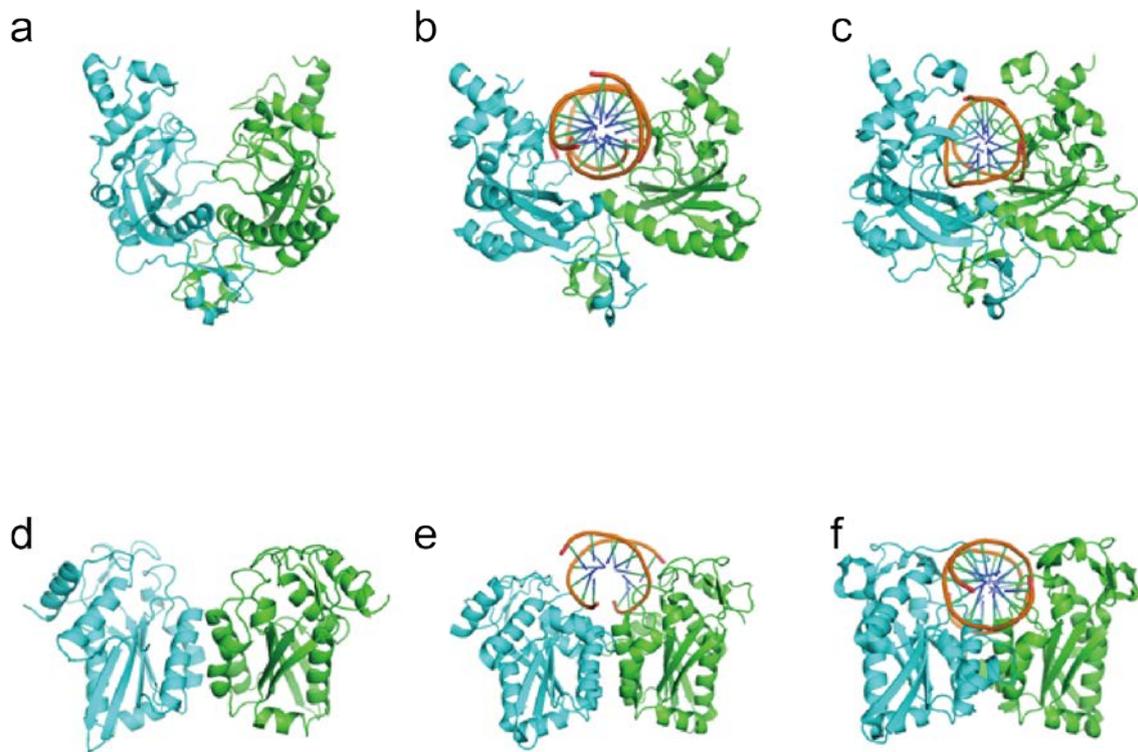
Figure 3. Structural changes of restriction enzymes during the DNA binding. (**a**)~(**c**): The DNA free structure of EcoRV (**a**) (PDB ID: 1RVE); the EcoRV-nonspecific DNA complex structure (**b**) (PDB ID: 2RVE); the EcoRV-specific DNA complex structure (**c**) (PDB ID: 4RVE). (**d**)~(**f**): The DNA free structure of BamHI (**d**) (PDB ID: 1BAM); the BamHI-nonspecific DNA complex structure (**e**) (PDB ID: 1ESG); the BamHI-specific DNA complex structure (**f**) (PDB ID: 1BHM).

2. Materials and Methods

2-1 The expression R.PabI R32A E63A and purification

2-1-1 The design of R.PabI R32A E63A used in the crystallization

The crystal structure of the R.PabI-specific DNA complex showed that R32 and E63, which exist in the half-pipe region, functions by stabilizing the flipped out guanine[13] (Fig. 4a,b, 5), the glycosylase activity assay further indicated that mutating R32 and E63 to alanine would completely deactivate R.PabI (Fig. 4c). In addition, the R32A mutant loses its ability in specifically recognizing and binding its recognition sequence[10] (Fig. 4d). In this research, the double mutant R32A E63A was used to form nonspecific complex with a double stranded DNA which does not contain recognition sequence. Based on the structure of the DNA free R.PabI, the 1-7 residues in the N-terminus could not form stabilized structure, so when the expression vector was constructed, the DNA sequence of the 7 residues were knocked out. The DNA sequence coding 8-226 residues of R.PabI has been inserted into the NdeI-BamHI site of the pET26 vector (Novagen) (Fig. 5a), the vector was proved by the previous researcher

To perform the transformation experiment, the *E.coli* Rosetta(DE3)pLysS (Novagen) was used. Firstly, the Rosetta(DE3)pLysS was taken out from -80°C deep freezer and melted at 4°C, then 90 ng of pET26-R.PabI-ΔN-R32A E63A vector was added into 200 μl of competent cell. After 5 min storing at ice, the mixture was incubated at 42°C for 40 seconds. After storing at ice for another 5 minutes, 600 μl SOC culture medium (2% tryptone, 0.5% yeast extract, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) was added to the mixture, then the mixture was incubated in a shaker for 1 hour at 37°C. After that, the mixture solution was spread on a LB agar plates which contains 2.5% LB medium (Merck), 1.5% agar and antibiotics (20 μg/ml kanamycin and 100 μg/ml chloramphenicol). The plate was incubated at 37°C for overnight, after colony

of the *E.coli* appeared at the plate, it was transferred to 4°C refrigerator for storage.

2-1-2 The over expression of R.PabI-ΔN-R32A E63A

After the transformation, the colony appeared in the plate after incubating for overnight at 37°C. The colony was picked and transformed into the test tubes which contains 10 ml liquid LB culture medium (2.5%) and antibiotics (20 μg/ml kanamycin and 100 μg/ml chloramphenicol). The test tubes were incubated at 37°C for 3 hours. Then, the bacteria solution in the test tubes were transfered into flasks which contained 2L LB culture medium (2.5%) and antibiotics (20 μg/ml kanamycin, 100 μg/ml chloramphenicol). The flasks were incubated at 37°C, after the absorbance of the bacteria solution reached 0.6 at 600 nm, 1 mM IPTG (Wako) was added into the culture medium to induce the expression of the protein. Then, the flasks were incubated at 25°C for overnight. At the next day, the cultured bacteria were collected by centrifugation at 5000 rpm for 15 min. The collected bacteria were stored at -30°C.

2-1-3 The purification of R.PabI-ΔN-R32A E63A

The collected bacteria were resuspended in sonication buffer which contained 25 mM MES (pH 6.0) and 50 mM MgCl$_2$, the sonicator Sonifer 250D (Branson) was used to lysis the resuspended bacteria by 5 min (Power 7). After that, the lysed solution was centrifuged at 40,000 × g for 30 min, the supernatant was collected, 10 μl Cryonase Cold-active Nuclease (TAKARA) was added into the supernatant, the mixture was incubated for 30 min at 4°C to remove contaminant nucleic acids from *E. coli*. The solution was heated at 80°C for 30 min to denature heat-unstable *E. coli* proteins and then centrifuged at 40,000 × g for 30 min to separate the target protein from the denatured proteins. The

supernatant was stirred with 2 ml Toyopearl AF-Heparin-650M (TOSOH) resin for 30 min, then the solution was applied to an Econo-Pac columns (BIO-RAD) to flow through the unbound protein. After the column is washed by 25 ml wash buffer (10 mM MES (pH 6.0), 50 mM $MgCl_2$, 100 mM NaCl), R.PabI protein is eluted by 10 ml Elution buffer (10 mM MES (pH 6.0), 1 M NaCl). The existence of the eluted R.PabI was checked by SDS-PAGE electrophoresis. Then, the eluted R.PabI was dialyzed against 10 mM MES (pH 6.0), 500 mM NaCl for overnight. The dialyzed protein was further purified by a MonoS HR 10/10 (GE Healthcare) column, the column was pre-equilibrated with 10 mM MES (pH 6.0) and the protein was eluted using a linear gradient of 0-1 M NaCl. After the purification, the purity of R.PabI was checked again by SDS-PAGE electrophoresis. After the purity been confirmed, the protein solution was concentrated to ~100 μM (the dimer concentration) and was stored at -80°C until use.

2-2 The double-stranded DNA preparation

The oligonucleotide purification cartridge (OPC)-purified 20bp oligonucleotides (5'-GCACTAGTTCGAACTAGTGC-3', Fig. 6a) were purchased from Eurofins Genomics and were dissolved in the annealing solution containing 2.5 mM MES (pH 6.0), 20 mM NaCl, and 2.5 mM $MgCl_2$ to be 25 μM. The ssDNA samples were annealed by incubating at 368 K and slow cooling to 277 K.

2-3 The crystallization of R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex

Before the protein solution and the DNA solution were mixed, the concentration of R.PabI in the solution was calculated by the absorbance at 280 nm. One R.PabI molecule contains 1 tryptophan (its molar absorption coefficient at 280 nm is 5500 L/(cm·mol)), 14 tyrosine

(1340 L/(cm·mol)) and 7 phenylalanine (190 L/(cm·mol)), based on these numbers, the molar absorption coefficient of R.PabI was calculated to be 25590 L/(cm·mol). The concentration of R.PabI was further calculated by this data.

4 ml of purified R.PabI solution (10 μM as dimer) was mixed with 0.8, 1.2, 1.6 ml of double stranded DNA (25 μM) to make the molar ratio of R.PabI dimer : dsDNA as 2:1, 2:1.5, 1:1. These mixed solutions were concentrated to 156 μM (R.PabI dimer) and ion-exchanged to 10 mM MES (pH 6.0), 100 mM NaCl using Vivaspin 20 (10,000 MWCO) (GE Healthcare).

The mixed solution was used to perform the first screening of crystallization with Crystal Screen HT and Index HT (both are made by Hampton Research), Wizard I&II (Emerald BioStructures), the JCSG+ suite, The JCSG core I Suite, The JCSG Core II suite, The JCSG Core III Suite, The JCSG Core IV Suite (QIAGEN). The Nanolitre Protein Crystallisation Robot mosquito (TTP LABTECH) was used to deliver the reservoir and protein solution to the crystallization plate VCP-1 (Violamo). In each well of the plates, 0.2 μl R.PabI-DNA mixture and 0.2 μl reservoir were mixed, sitting-drop vapour-diffusion method was used to grow the crystal. The plates were incubated at 20°C.

Among the conditions in the first screening, crystal was observed at B3 of the JCSGCore II Suite (0.2 M Calcium acetate, 0.1 M Imidazole (pH 8.0), 10% PEG8000). After that, the concentration of precipitant and pH of buffer were arranged to perform the second screening using the sitting-drop vapour-diffusion method, this time, the Cryschem Plates (Hampton Research) which have 24 wells were used for the second screening. In order to obtain crystal with nice quality, the volume ratio of protein : reservoir was arranged to 1:1, 1.5:1, 1:1.5, 2:1 and 1:2 when performing the second screening.

2-4 The structure determination of R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex

2-4-1 The X-ray diffraction of R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex

X-ray diffraction data were collected at the NE3A beamline of the Photon Factory (Tsukuba, Japan) under cryogenic conditions (95K). The complex crystal was picked up by Dual-Thickness MicroLoop (MiTeGen). After that, the data was collected at the following condition:

The wavelength of X-ray: 1.000Å

The detector: ADSC Quantum 270

The detector distance: 244 mm

The exposure time: 0.5 s

The oscillation range: 0.5°

The data range: 1~720

2-4-2 The structure determination of R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex

The crystal of the R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex diffracted X-ray to 1.9Å. The X-ray diffraction data were indexed and integrated using the program XDS[33] and scaled using SCALA in the CCP4[34] suite. The crystal of the R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex belongs to the space group $C222_1$ with unit cell parameters of a=72.89, b=261.7, c=65.08 Å. The initial model was determined by the molecular replacement method using the program MOLREP[35] with the coordinates of the crystal structure of the DNA free R.PabI (PDB ID: 2DVY). The initial model was refined and rebuilt using the program Phenix.reine[36] and Coot[37].

2-4-3 The valuation of the complex structure

The interaction between the protein and the DNA was analyzed by PISA[38]. The comparison of different molecules and the display of structures were performed by Pymol (http://www.pymol.org/). The width of the DNA groove was calculated by 3DNA[39]. Molprobity[40] was used to make the Ramachandran plot.

2-5 The design of the R.PabI mutants

2-5-1 The production of the R.PabI-ΔN-Y68F mutant

Based on the structure of the R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex, several important amino acid residues have been identified. To prove the importance of these resides, the mutants of these residues were produced. The pET28a-R.PabI-Y68F was used as the template of these mutants. The Y68 is reported to have important function in the cleavage of the *N*-glycosidic bond, the glycosylase activity of its phenylalanine mutant is decreased so significantly that it could be successfully expressed by the *E.coli*. In addition, its ability in binding to the recognition sequence is nearly the same as the wild type R.PabI[13], which makes it an ideal template to produce mutants which would be used at the DNA binding assays and the glycosylase activity assays. The gene which codes R.PabI-Y68F mutant was cloned into the NdeI-BamHI site of pET28a, this vector was created by the previous researcher. However, in order to produce protein that is similar to that has been used in the crystallization experiment, the 1-7 residues at the N-terminus of R.PabI as well as the His-tag in the vector was knocked out. This pET28a-R.PabI-ΔN-Y68F vector was used as the template to produce all of the other mutants (Fig. 7b).

The deletion mutation experiment was performed according to the protocol of the PrimeSTAR Mutagenesis Basal kit (TAKARA). The primers used to create pET28a-

R.PabI-ΔN-Y68F are written in Table 1. The nucleotides used in this research were all purchased from eurofins. The composition of the PCR reaction solutions and the reaction conditions are written below:

2 x PrimeSTAR Max Premix	25 μl
Primer	10 pmol
Template	200 ng
milliQ	To make the total volume to 50 μl

The reaction condition

95℃	3 min

Repeat the following 3 steps for 30 circles:

98℃	10 s
55℃	15 s
72℃	40 s
72℃	3 min

After the PCR experiment, the agarose gel electrophoresis was performed to confirm that the target sequence is produced. Then, 2 μl DpnI (TAKARA) was added into the PCR product and the mixture was incubated at 37°C for 2 hours to cleave the template. After that, the plasmid was transformed into the *E.coli* XL1-Blue to mass-produce the plasmids. The detailed transformation method was the same as the previous experiments mentioned in 2-1-1. The colony in the plate was transformed into the 10 ml liquid LB culture medium (2.5%) which contains 20 μg/ml kanamycin and was incubated at 37°C for overnight. At

the next day, after the *E.coli* were over reproduced, the Wizard Plus SC Minipreps DNA Purification system (Promega) was used to extracted the vector from the *E.coli*. The sequence of the extracted vector was confirmed by the sequence array service of Fasmac company.

The vector were transformed into *E.coli* Rosetta(DE3)pLysS (Novagan). The transformation was performed in the same way as the previous experiment.

2-5-2 The production of the other mutants

The mutants of the other residues that were considered of having important function were produced by the same method mentioned at 2-5-1. Sequences of the primers used to produce these mutants are written at Table 1. The overexpression as well as purification of these mutants were the same as 2-1.

2-6 The electrophoretic mobility shift assay (EMSA) experiment

Three 5'-fluorescein-labelled 20 bp dsDNAs (5'-GCACTAGTTCGAACTAGTGC-3' (same sequence as the DNA used for co-crystallization, Fig. 6b), 5'-GCATCGATTCGAATCGATGC-3' (nonspecific, Fig. 6c), and 5'-GCATAGCTGTACAGCTATGC-3' (specific, Fig. 6d)) were used as probes. 0.1 μM of each DNA probe and the R.PabI dimer (0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4, and 12.8 μM) were mixed in 10 mM MES (pH 6.0), 300 mM NaCl, and 50 μM single-stranded poly-dT (40 nt) and incubated for 30 min at 4°C. To detect the R.PabI-dsDNA interactions by EMSA, an abundance of ssDNA was added as a competitor. Bound and unbound dsDNA were separated through a 12% polyacrylamide gel in 0.5 × TBE at 4°C, and fluorescence was measured using an LAS4000 Mini system (Fujifilm, Tokyo, Japan).

2-7 The DNA glycosylase activity assay

DNA glycosylase activity assays of the R.PabI mutants were performed using 24 bp, 500 bp, and 3000 bp dsDNAs possessing only one R.PabI recognition sequence at their center regions. The 5'-fluorescein-labelled 24 bp dsDNA (5'-fluorecesin-GGATGCATGAGTACGAGGACCATC-3', Fig. 6e) was purchased from Eurofins. The 500 bp and 3000 bp dsDNA were amplified by PCR using the modified pET26b plasmid, which has only one 5'-GTAC-3' site in the *lacI*-coding region[13], as a template and purified using the QIAquick PCR Purification Kit (QIAGEN) (Sequences of primers used to produce the 500 bp and 3000 bp substrates are written at Table 2). Then, 0.2 μM of the 24 bp dsDNA was mixed with 0.4 μM of the R.PabI dimers in a reaction buffer containing 0.1 M phosphate buffer (pH 6.5), while 5.9 nM of the 500 bp or 3000 bp dsDNA was mixed with 80 nM of the R.PabI dimers in reaction buffer containing 0.1 M phosphate buffer (pH 6.5). The reaction solution was incubated at 45°C for 1, 3, 5, 10, 20, 30, 60, 90, and 120 min (for the 24 bp dsDNA substrate), 5, 10, 20, 30, 45, 60, 90, and 120 min (for the 500 bp DNA substrate), or 5, 10, 20, 30, 45, 60, 90, 120, and 180 min (for the 3000 bp DNA substrate). After the enzymatic reactions, the reaction solutions were supplemented with 0.1 M NaOH to stop the enzymatic reaction. The solutions were then heated at 70°C for 10 min to cleave the products at the 5' and 3' side of the AP sites generated by R.PabI and neutralized by the addition of an equal concentration of HCl. The reaction solutions using 24 bp dsDNA as a substrate were separated on a denaturing 18% polyacrylamide gel in 0.5 × TBE and 7 M urea. The reaction solutions using 500 bp or 3000 bp dsDNA as a substrate were separated on a 1% agarose gel and stained using GelGreen (Biotium). The fluorescence was measured using an LAS4000 Mini system.

The enzymatic rate constant k was obtained from a single-exponential fit to the data from three independent measurements: $f_p = f_p\text{max} \times (1\text{-}e^{-kt})$, in which 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.

Table 1. Primers used in producing the R.PabI mutants

The primers used to produce the R.PabI-ΔN-Y68F	Forward primer: 5'-ATACC<u>ATGGAA</u>GCGAGTGTTTCTTTTGAAAATGG-3'
	Reverse primer: 5'-TCGC<u>TTCCAT</u>GGTATATCTCCTTCTTAAAGTTAAAC-3'

The deletion sequences are underlined

The primers used to produce Y68F R70D	Forward primer : 5'- TTCGCG<u>GAC</u>GATGGAAAATATGATTAT-3'
	Reverse primer : 5'- TCCATC<u>GTC</u>CGCGAAAGATATCTGCCA-3'
The primers used to produce Y68F D71R	Forward primer : 5'- GCGAGG<u>CGC</u>GGAAAATATGATTATGAG-3'
	Reverse primer : 5'- TTTTCC<u>GCG</u>CCTCGCGAAAGATATCTG-3'
The primers used to produce Y68F R70D D71R	Forward primer : 5'- GCGGCC<u>GACCGC</u>AAATATGATTATGAG-3'
	Reverse primer : 5'- TTTTCC<u>GCGGTC</u>CGCGAAAGATATCTG-3'
The primers used to produce Y68F R26A	Forward primer : 5'- ATAACG<u>GCC</u>CCTACATCAAAAATTAGG-3'
	Reverse primer : 5'- TGTAGGG<u>GC</u>CGTTATTGGTAATCTAAC-3'

The mutation sequences are underlined

Table 2. Primers used in producing substrate DNA in the glycosylase assay experiments

Primers used to produce 500bp fragment	Forward primer : 5'-GTCGCGGCGATTAAATC-3'
	Reverse primer : 5'-GCCATCTGATCGTTGGC-3'
Primers used to produce 3000bp fragment	Forward primer: 5'-GAACCATCACCCTAATCA-3'
	Reverse primer : 5'-CTGCCTGTTCATCCGCG-3'
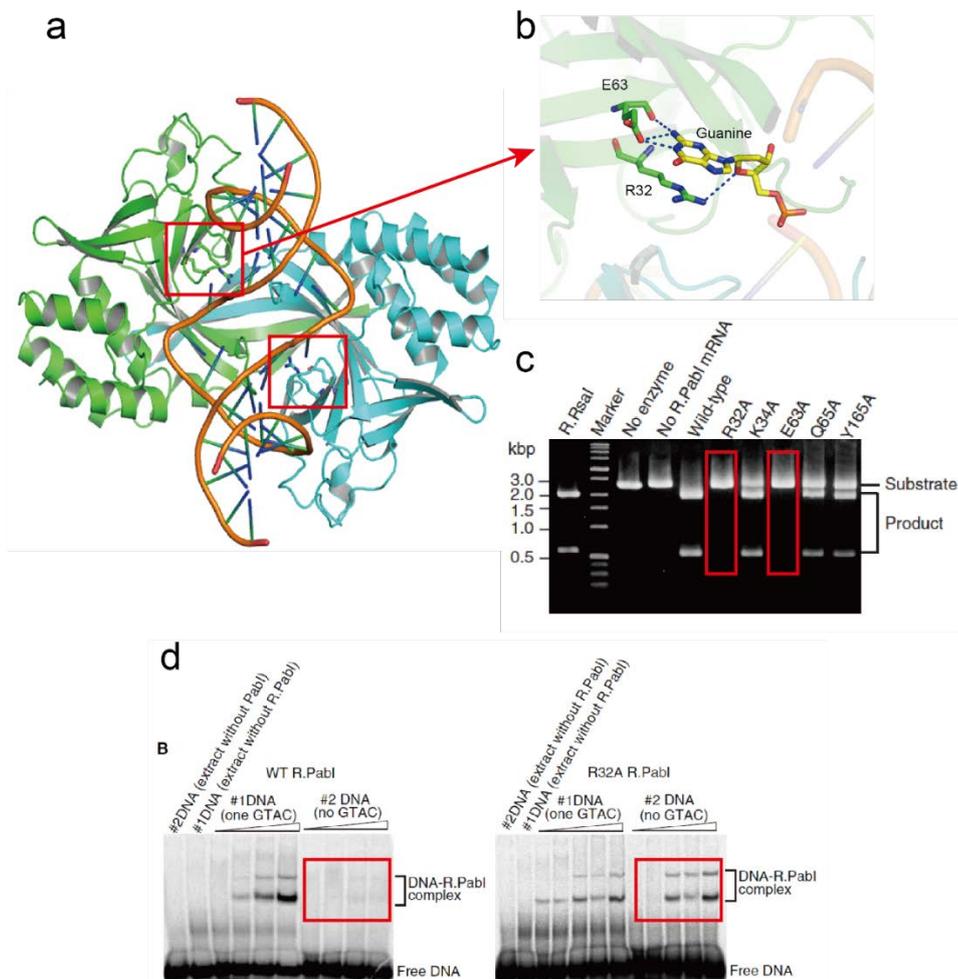
Figure 4. Positions and roles of R32 and E63[10,13]. (**a**) Positions of R32 and E63 in the R.PabI-specific DNA complex. Protomer A and B are colored green and cyan respectively, the dsDNA is colored orange, positions of R32 and E63 in each protomer are shown in red boxes. (**b**) Interactions between DNA base and R32 and E63. R32 and E63 are colored green and the guanine in the DNA is colored yellow. The hydrogen bonds formed between them are shown as green dotted lines. (**c**) Glycosylase activity assay of R.PabI and its mutants performed by the previous researcher. The results of R32A and E63A are shown in red boxes. (**d**) DNA binding assay of wild type R.PabI and the R.PabI R32A mutant. The results of wild type R.PabI and the R32A mutant binding to dsDNA which does not contain recognition sequence are shown in red boxes.

Figure 5. Intermolecular hydrogen bonds between one R.PabI dimer and dsDNA in the previously solved R.PabI-specific dsDNA complex[13]. Hydrogen bonds between protomer A and the dsDNA are shown as red lines. Hydrogen bonds between protomer B and the dsDNA are shown as blue lines. Residues of protomers A and B are colored red and blue, respectively.

a

```
           1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
     5' -  G  C  A  C  T  A  G  T  T  C  G  A  A  C  T  A  G  T  G  C - 3'
     3' -  C  G  T  G  A  T  C  A  A  G  C  T  T  G  A  T  C  A  C  G - 5'
          20'19'18'17'16'15'14'13'12'11'10' 9' 8' 7' 6' 5' 4' 3' 2' 1'
```

b

```
5'- fluorescein -G C A C T A G T T C G A A C T A G T G C - 3'
             3'-C G T G A T C A A G C T T G A T C A C G - fluorescein - 5'
```

c

```
5'- fluorescein -G C A T C G A T T C G A A T C G A T G C - 3'
             3'-C G T A G C T A A G C T T A G C T A C G - fluorescein - 5'
```

d

```
5'- fluorescein -G C A T A G C T [G T A C] A G C T A T G C - 3'
             3'-C G T A T C G A [C A T G] T C G A T A C G - fluorescein - 5'
```

e

```
5'- fluorescein - G G A T G C A T G A [G T A C] G A G G A C C A T C - 3'
               3'- C C T A C G T A C T [C A T G] C T C C T G G T A G - 5'
```

Figure 6. Sequences of dsDNA utilized in the crystallization, EMSA and glycosylase activity assay experiments. (**a**) Sequence of dsDNA used for the crystallization of the R.PabI-nonspecific dsDNA complex. (**b**) Sequence of dsDNA used for the EMSA experiments, the sequence is the same as (**a**). (**c**) Sequence of nonspecific dsDNA used for the EMSA experiments. (**d**) Sequence of specific dsDNA used for the EMSA experiments, the recognition sequence is shown in red box. (**e**) Sequence of specific dsDNA used for the glycosylase activity assays, the recognition sequence is shown in red box.

35
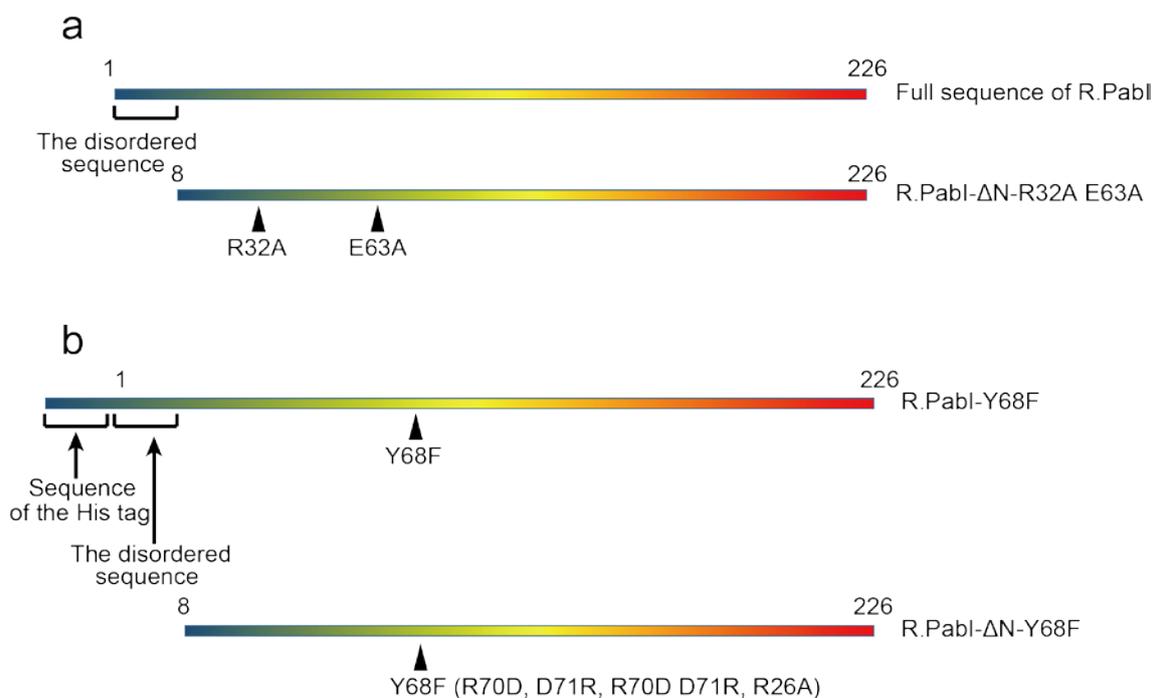
Figure 7. R.PabI construct used in the crystallization and bio-chemistry experiments. (**a**) R.PabI-ΔN-R32A E63A which has been used in the crystallization experiments was produced by deleting the 7 residues in the N-terminus and mutating R32 and E63 to alanine. (**b**) R.PabI-ΔN-Y68F was used as template to create the mutants used in the bio-chemistry experiments. The 7 residues in the N-terminus were deleted and Y68F was mutated to phenylalanine.

3. Results

3-1 The expression, purification and crystallization of the R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex

The over expressed R.PabI-ΔN-R32A E63A was purified by heat treatment, heparin column and cation exchange chromatography, a single peak which contains target protein was eluted during the cation exchange chromatography. SDS-PAGE electrophoresis was also used to check the purity of the eluted fraction, the result showed that a single band appeared at a position below the 30k molecular weight marker, since the molecular weight of R.PabI is about 26 kDa, the eluted protein was confirmed to be R.PabI (Fig. 8,9). About 6 mg of R.PabI dimer could be produced by 2 L of incubated *E.coli*.

In the first screening of the crystallization experiment, crystal appeared at B3 well (0.2 M Calcium acetate, 0.1 M imidazole (pH 8.0), 10% PEG8000) of the crystallization kit JCSG core II suite when R.PabI dimer was mixed with dsDNA in a molar raito of 2:1. Then, each parameter of the reservoir and the volume ratio of protein-DNA complex solution : reservoir were arranged to improve the quality of the crystal, as a result, crystal with the best quality was obtained at the condition of 0.2 M Calcium acetate, 0.1 M MES (pH 5.8), 8% PEG8000 when the complex solution was mixed with reservoir at a volume ratio of 1.5 μl : 1 μl (Fig. 10).

3-2 Structural determination of the R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex

The crystal obtained from 3-1 was used in the X-ray diffraction experiment. The crystal of the R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex diffracted X-ray at 1.9Å (Fig. 11), after the index and integration, the space group to which the crystal belongs was determined to be $C222_1$ with a= 72.89, b=261.7, c=65.08 Å. The detailed information of the data collection is written at Table 3.

38

The molecular replacement was performed using the coordinate of the coordinates of the DNA-free R.PabI structure (PDB ID: 2DVY). After that, the structure of R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex was determined. The asymmetric unit of the crystal contains one R.PabI dimer (protomers A and B), one single-stranded DNA (ssDNA, chain C), and 162 water molecules (Fig. 13a,b, Table 4). The positions of the initiating methionine and 3 residues (G224, T225 and S226) at the C-terminus in protomer A could not be determined because their electron density did not appear in the structure. Similar to this, the positions of the initiating methionine and 3 residues at the C-terminus (G224, T225 and S226) in protomer B also could not be determined. The program Molprobity was used to calculate the validity of the backbone dihedral angles, as a result, 97.2% of residues are included in the favored region, all of the residues are in the allowed region (Fig. 12). The B-factor of each molecule in the asymmetric unit was calculated by the CCP4 suite and the result showed that The B-factor of protomer A is 48.96 Å^2, protomer B is 74.82 Å^2, DNA chain is 57.20 Å^2 and the water molecule is 53.70 Å^2 (Table 4).

Protomer A consists of five α helices, three 3_{10} (η) helices, and ten β strands. Among the β strands, strands β3, β4, β5, β6, β7, β8, and β9 form a half-pipe structure with the corresponding strands of protomer B. The structures of protomers A and B are nearly identical; the root mean square deviation (r.m.s.d.) between the two protomers is 1.7 Å for superposed 210 Cα atoms. The ssDNA is bound to the electropositive half-pipe region. The R.PabI-ssDNA complex observed in the asymmetric unit forms a tetrameric R.PabI structure with a symmetrically related complex generated by the crystallographic two-fold axis (R.PabI protomers A' and B' and ssDNA strand C'). The ssDNA in each asymmetric unit forms a dsDNA between the two R.PabI dimers (Fig. 14a).

In the tetrameric structure, the 20 bp dsDNA is sandwiched by the half-pipe regions of two R.PabI dimers. Although the contact surface area between the two protomers is small (the contact surface areas between the A-A' and B-B' interfaces are 201.1 $Å^2$ and 102.8 $Å^2$, respectively), the two R.PabI dimers are connected by four salt bridges between the protomers (Fig 14b). At the A-A' interface, R70 and D71 of protomer A form two salt bridges with D71' and R70' of protomer A' at distances of 4.06 and 4.06 Å, respectively (the residues and bases of the symmetrically related molecules are indicated by a prime). At the B-B' interface, R70 and D71 of protomer B form two salt bridges with D71' and R70' of protomer B' at distances of 2.92 and 2.92 Å, respectively (Fig 14b,c,d).

3-4 The R.PabI-DNA interactions in the R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex

Similarly to EcoRV and BamHI[2,4,17], the R.PabI dimer interacts with nonspecific dsDNA using small contact surface area and few hydrogen-bonds compared to the specific dsDNA. The interface area between the R.PabI dimer and the dsDNA in the nonspecific complex is 1156 $Å^2$, which is approximately half of the interface area in the specific complex (2166 $Å^2$)[13]. However, because R.PabI binds nonspecific dsDNA as a tetramer (Fig. 14a), the total R.PabI-dsDNA interface area of the nonspecific dsDNA binding state is approximately the same as in the specific dsDNA binding state. In the R.PabI-nonspecific dsDNA complex structure, the dsDNA is recognized by 24 hydrogen bonds between the DNA and each R.PabI dimer (Fig. 15 and Table 5). The number of hydrogen bonds in the nonspecific complex is much lower than in the specific complex; the R.PabI dimer forms 69 hydrogen bonds with the specific dsDNA[13]. In protomer A, T46, R47, K48, K49, and S50 of the β4-η2 loop form eight hydrogen bonds with the phosphate

groups of Ade3' and Cyt4' (Fig. 16a). S29 and K30 of the β2-β3 loop form two hydrogen bonds with the phosphate group of Ade12 (Fig. 16b). In protomer B, T25, S45, R47, R184, and N185 form eight hydrogen bonds with the phosphate groups of Ade16', Gua17', and Thy18' (Fig. 16c). S29 and V158 form hydrogen bonds with the phosphate groups of Thy9 and Thy5', respectively (Fig. 16d,e). Among these hydrogen bonding residues, T46, R47, V158, R184, and N185 also form hydrogen bonds with the phosphate groups of dsDNA in the R.PabI-specific dsDNA complex[13]. In addition to these sequence-nonspecific R.PabI-DNA hydrogen bonds, the side chain of R156 in protomer B is inserted into the major groove of dsDNA and forms two hydrogen bonds with the base groups of Gua7' and Thy8' (Fig. 16e). The structure of dsDNA between the two R.PabI dimers is slightly bent by ~20° at its middle region due to these R.PabI-nonspecific dsDNA interactions (Table 6, Fig. 17a). The minor groove of the dsDNA is expanded from a typical B-form dsDNA in the middle region (~10 Å) due to the dsDNA bending. On the other hand, the major groove of the dsDNA is narrowed from typical B-form dsDNA (~17 Å) in the middle region (Table 6, Fig. 17b).

In addition to these hydrogen bonds, some residues are inserted into the major or minor grooves of dsDNA to recognize its shape or electrostatic potential, although these residues do not form any direct hydrogen bonds with the dsDNA. In the R.PabI-specific dsDNA complex structure, the R.PabI dimer bends dsDNA at the half-pipe region and unwinds the 5'-GTAC-3' sequence for recognition using the β8-β9 loop[13]. In the R.PabI-nonspecific dsDNA complex structure, the β8-β9 loop of the protomer A interacts with the phosphate backbone around the minor groove of Gua7-Cyt10, and the side chains of Q155 and R156, which are part of the sequence-specific interaction in the R.PabI-specific dsDNA complex structure, are inserted into the minor groove (Fig. 16f). Meanwhile, two

arginine residues, R26 of protomer B and R47 of protomer A, are also inserted into the minor and the major groove of dsDNA, respectively (Fig. 16g). Some transcription factors insert arginine residues into minor grooves of dsDNA to recognize the sequence-specific shape and electrostatic potential of dsDNA to increase their sequence specificity[41]. Similarly, λ exonuclease was predicted to slide along dsDNA using an arginine residue inserted into the minor groove of dsDNA as a guide[42]. R26 of protomer B may also play an important role in the R.PabI-dsDNA interaction.

3-5 Structure comparison between binding states

To analyze the structure modification of R.PabI upon the binding of dsDNA, the dimeric structures of R.PabI in the DNA-free state (PDB code: 2DVY)[10], the nonspecific dsDNA binding state, and the specific dsDNA binding state (PDB code: 3WAZ)[13] were superposed. When the structures of R.PabI in the DNA-free state and the nonspecific dsDNA binding state are compared, the structure of each protomer is observed to be slightly twisted in an anticlockwise direction in the nonspecific dsDNA binding state, and the structures of the β2-β3 loop of protomer B, which contains R26 as described above, and of the β4-η2 loop of each protomer are modified to form direct hydrogen bonds with the phosphate groups of the dsDNA (Fig. 18a). On the other hand, the structures of the β8-β9 loops, which play an important role in unwinding dsDNA at the recognition sequence, are not modified by the binding of nonspecific dsDNA. The r.m.s.d. between the R.PabI dimers in the DNA free and the nonspecific dsDNA binding states is 2.2 Å for 424 superposed Cα atoms.

When the structures of R.PabI in the nonspecific dsDNA binding state and the specific dsDNA binding state are compared, the structure of each protomer is further

twisted in an anticlockwise direction in the specific dsDNA binding state, and the structures of β2-β3, β4-η2, and β8-β9 loops of each protomer are modified to form hydrogen bonds with the deformed dsDNA (Fig. 18b). In the specific dsDNA complex, P27, T28, and S29 of the β2-β3 loop protrude into the expanded minor groove of dsDNA to stabilize the unwound structure of dsDNA[13]. Although the position of R26 is slightly modified between the nonspecific dsDNA complex and the specific dsDNA complex, the side chain of R26 is inserted into the minor groove of dsDNA in each dsDNA binding state. The β4-η2 loop is utilized to form direct hydrogen bonds with the phosphate groups of dsDNA in each binding state. Because the dsDNA is bent by approximately 90° at the recognition sequence[13], the positions of the phosphate groups differ between the nonspecific dsDNA complex and the specific dsDNA complex. The structure of β4-η2 loop is modified to optimize the interactions with the phosphate groups. The r.m.s.d. between the R.PabI dimers in the nonspecific dsDNA binding states and the specific dsDNA binding state is 2.4 Å for 427 superposed Cα atoms.

3-6 The result of the electrophoretic mobility shift assay (EMSA)

Although R.PabI binds the specific dsDNA sequence as a dimer[13], the structure of the R.PabI-nonspecific dsDNA complex indicates that R.PabI binds nonspecific dsDNA as a tetramer and the tetrameric structure of R.PabI is stabilized by the four salt bridges. To analyze the DNA binding ability of R.PabI in solution, the EMSA was performed using the R.PabI active site mutant (Y68F), which possesses approximately the same base-specific DNA binding ability as the wild-type R.PabI but has reduced DNA glycosylase activity, as a control[13].

First, the sequence specific and the sequence nonspecific dsDNA binding ability of

the R.PabI Y68F mutant was analyzed (Fig 19a). When the 20 bp dsDNA with the specific sequence was utilized as a probe, two shifted bands were observed: the shifted band observed at the low R.PabI concentration ranges and the super shifted band observed at the high R.PabI concentration ranges. Because R.PabI binds the recognition sequence as a dimer, the shifted band is predicted to present the R.PabI dimer-dsDNA complex. The super shifted band is predicted to present the R.PabI-DNA complex that is formed by sequence nonspecific R.PabI-DNA interactions due to the high concentration of R.PabI. In contrast, when the nonspecific dsDNA was used as a probe, the shifted band was very weak compared to the predominant super shifted band. Although the dsDNA that was utilized for the co-crystallization in this study contains a part of the R.PabI recognition sequences (5'-GT-3', 5'-TA-3' and 5'-AC-3' steps), the DNA binding ability of R.PabI for the dsDNA used for the co-crystallization is approximately the same as that for the dsDNA not containing any 5'-GT-3', 5'-TA-3' and 5'-AC-3' steps (Fig. 19a). This indicates that R.PabI does not recognize 5'-GT-3', 5'-TA-3' and 5'-AC-3' steps in dsDNA except for 5'-GTAC-3' (i.e. the dsDNA used for the co-crystallization is the nonspecific dsDNA for R.PabI).

Second, the dsDNA binding ability of the R.PabI mutants were analyzed to evaluate the importance of the salt bridge formations for the tetramerization of R.PabI. To prevent salt bridge formation, the Y68F R70D and the Y68F D71R mutants were produced. These mutations would prevent the tetramerization of R.PabI on nonspecific dsDNA due to the electrostatic repression of their side chains. In addition, the Y68F D71R mutant would also prevent the tetramerization of R.PabI on nonspecific dsDNA due to the steric hindrance of the long side chain of D71R. When the dsDNA binding ability of each mutant was analyzed by EMSA using the 20 bp DNA without the R.PabI recognition

sequence as a probe, the Y68F and the Y68F R70D mutants showed approximately the same band shift pattern; only the super shifted bands were observed at all concentration ranges. In contrast, the Y68F D71R mutant showed a different band shift pattern; the shifted band was observed at the middle R.PabI concentration ranges (0.2 − 1.6 μM of the R.PabI dimer, Fig. 19b) in addition to the super shifted band. Because the structural analysis of R.PabI-nonspecific dsDNA complex showed that R.PabI dimer binds 20 bp dsDNA as a tetramer, the super shifted bands observed in the Y68F and the Y68F R70D mutants were predicted to represent the complex of two R.PabI dimers and a labelled dsDNA, and the shifted bands observed in the result of the Y68F D71R binding assay were predicted to represent the complex of one R.PabI dimer and a labelled dsDNA. These results indicate that the tetramerization of R.PabI is inhibited by the D71R mutation, although the R70D mutation showed no effect on the nonspecific dsDNA binding affinity, and that R.PabI prefers to bind nonspecific 20 bp dsDNA not as a dimer but as a tetramer. Because the side chain length of aspartic acid is shorter than for arginine, the effect of electrostatic repression between R70D and D71 would not be enough to block the binding of the second R.PabI dimer on nonspecific dsDNA. The tetramerization of R.PabI on nonspecific dsDNA is predicted to be mainly accelerated by the deformation of dsDNA structure caused by a binding of one R.PabI dimer and the resulting tetrameric structure of R.PabI is stabilized by the four salt bridges. The deformation of dsDNA structure is an important factor for cooperative bindings of multiple DNA binding proteins[43,44]. The DNA binding ability of the Y68F R70D D71R triple mutant was also analyzed. Because the R70D D71R mutation would form two salt bridges at the interface region, similarly to the wild type enzyme, the inhibition of the tetramerization of R.PabI by the D71R mutation would be recovered by the R70D D71R mutation. The Y68F R70D D71R

mutant showed similar nonspecific dsDNA binding ability to the control Y68F mutant, as expected (Fig. 19b).

3-7 The results of the DNA glycosylase activity assays

To evaluate the importance of the tetramerization of R.PabI on nonspecific dsDNA for its enzymatic activity, the DNA glycosylase activities of the Y68F, Y68F R70D, Y68F D71R, and Y68F R70D D71R mutants were analyzed using dsDNA with different lengths (Fig. 20a, 21a,b 22a,b). The results of the DNA glycosylase assay of the Y68F mutant showed that the activity of the Y68F mutant for the 3000 bp dsDNA was higher than that for the 500 bp dsDNA (Table 7), although the 3000 bp dsDNA contains more nonspecific dsDNA per R.PabI recognition sequence than the 500 bp dsDNA. Similar results were also observed in other restriction enzymes that utilized the facilitated diffusion mechanisms to find their targets; DNA cleavage activities of these enzymes are increased when substrate dsDNA sequences contain more nonspecific sequences[45]. If R.PabI utilizes the tetrameric structure observed in the R.PabI-nonspecific dsDNA complex structure to facilitate diffusion, the DNA glycosylase activities of the mutants that prevent tetramerization on nonspecific dsDNA (the Y68F R70D and Y68F D71R mutants) would be reduced with increasing lengths of nonspecific dsDNA per R.PabI recognition sequence and the reduced activity would be recovered by the R70D D71R mutation. When the 24 bp dsDNA was utilized as a substrate, the Y68F R70D, Y68F D71R, and Y68F R70D D71R mutants showed 127%, 74% and 105% relative activities compared to the control Y68F mutant (Table 7, Fig. 20a, 21b). When the 500 bp and 3000 bp dsDNA were utilized as substrates, the Y68F R70D mutant showed 76% and 19% relative activities compared to the Y68F mutant, respectively. The activity of the Y68F D71R mutant was also decreased

46

to 34% and 19% of the Y68F mutant when using the 500 bp and 3000 bp dsDNA as substrates, respectively. These data showed that the mutations of R70D and D71R result in decreased DNA glycosylase activities for longer dsDNA substrates. Meanwhile, the decreased activities for longer dsDNA substrates were recovered by the double mutation of R70D D71R as expected; the activities of the Y68F R70D D71R mutant for the 500 bp and 3000 bp dsDNA were 71% and 40% of the activity of the Y68F mutant (Table 7, Fig. 20a, 22a,b).

The DNA glycosylase activity of the Y68F R26A mutant was also analyzed for dsDNA with different lengths. In the structure of the nonspecific dsDNA complex, Arg26 is located above the minor groove of dsDNA (Fig.16g) and is predicted to play an important role in the R.PabI-dsDNA interaction. When the 24 bp dsDNA was utilized as a substrate, the Y68F R26A mutant showed 65% relative DNA glycosylase activity compared to the control Y68F mutant (Table 7, Fig 20b, 21b). The Y68F R26A mutant showed 55% and 27% relative activities compared to the Y68F mutant when using the 500 bp and 3000 bp dsDNA as substrates, respectively (Table 7, Fig. 20b, 22a,b). R26 is also predicted to play an important role in cleaving long dsDNA substrates efficiently.

Table 3. Summary of the data collection statistics of the R.PabI-ΔN-R32A E63A-

nonspecific dsDNA complex

Data collection	
X-ray source	Photon Factory AR-NE3A
Detector	ADSC Quantum270
Wavelength (Å)	1.0000
Space group	$C222_1$
Unit cell parameters	
a, b, c (Å)	72.89, 261.7, 65.08
α, β, γ (°)	90, 90, 90
Resolution (Å)	43.6-1.90 (1.94-1.90)*
Wilson B-factor (Å^2)	41.5
Total number of reflections	685056 (27971)
Unique reflections	49388 (3110)
R_{sym} (%)	5.1 (111.1)
Mean $I/\sigma I$	28.5 (2.0)
Completeness (%)	99.6 (99.2)
Multiplicity	13.9 (9.0)

*The numbers in parentheses represent data for the highest-resolution shells.

Table 4. Summary of the refinement statistics of the R.PabI-ΔN-R32A E63A-

nonspecific dsDNA complex

Refinement	
Resolution (Å)	43.6-1.90 (1.94-1.90)
No. of reflections in working/test set	49246 / 2444
R/R_{free} (%)	20.62 / 24.81 (34.07 / 34.00)
No. atoms	
Protein/DNA/water	3475 / 407 / 162
B-factors (Å^2)	
Chain A (protomer A)	48.96
Chain B (protomer B)	74.82
DNA (chain C)	57.20
water	53.70
r.m.s.d.	
Bond angle (°)	1.30
Bond length (Å)	0.013
Rotamer outliers (%)	0.80
C-beta outliers	0
Clash score	3.91

*The numbers in parentheses represent data for the highest-resolution shells.

Table 5. Hydrogen bonds between R.PabI and DNA in the R.PabI-ΔN-R32A E63A-

nonspecific dsDNA complex

Protein				H-bonds	DNA			
Chain ID	Residue	Main/ Side	Atom Name	Distance (Å)	Chain ID	Base	Base/ Backbone	Atom Name
A	Thr46	Side	OG1	2.44	C'	Cyt4'	Backbone	OP2
A	Arg47	Main	N	2.96	C'	Cyt4'	Backbone	OP2
A	Arg47	Main	N	3.17	C'	Cyt4'	Backbone	O5'
A	Lys48	Main	N	3.41	C'	Cyt4'	Backbone	OP2
A	Lys48	Side	NZ	2.87	C'	Ade3'	Backbone	OP1
A	Lys49	Main	N	2.68	C'	Ade3'	Backbone	OP2
A	Ser50	Main	N	3.73	C'	Ade3'	Backbone	OP1
A	Ser50	Main	N	3.38	C'	Ade3'	Backbone	OP2
A	Ser29	Main	N	2.98	C	Ade12	Backbone	OP2
A	Lys30	Main	N	3.83	C	Ade12	Backbone	OP1
B	Thr25	Side	OG1	3.62	C'	Ade16'	Backbone	O3'
B	Thr25	Side	OG1	2.51	C'	Gua17'	Backbone	OP1
B	Ser45	Side	OG	3.70	C'	Ade16'	Backbone	OP1
B	Arg47	Side	NE	2.82	C'	Gua17'	Backbone	OP2
B	Arg156	Side	NH1	3.68	C'	Thy8'	Base	O4
B	Arg156	Side	NH1	2.70	C'	Gua7'	Base	O6
B	Val158	Main	N	3.38	C'	Thy5'	Backbone	OP1
B	Val158	Main	N	3.29	C'	Thy5'	Backbone	OP2
B	Arg184	Side	NE	3.24	C'	Thy18'	Backbone	OP2
B	Asn185	Side	N	2.76	C'	Gua17'	Backbone	OP1
B	Asn185	Side	ND2	3.17	C'	Gua17'	Backbone	O3'
B	Asn185	Side	ND2	3.33	C'	Thy18'	Backbone	OP1
B	Ser29	Main	N	2.76	C	Thy9	Backbone	OP1
B	Ser29	Side	OG	2.76	C	Thy9	Backbone	OP1

Table 6. The width of each groove of the dsDNA in the R.PabI-ΔN-R32A E63A-

nonspecific dsDNA complex

Base pair Step	Width of minor groove (Å)	Width of major groove (Å)
1 GC/GC		
2 CA/TG		
3 AC/GT		
4 CT/AG	11.1	19.1
5 TA/TA	11.7	19.5
6 AG/CT	12.7	19
7 GT/AC	12.9	18.3
8 TT/AA	13.1	18.6
9 TC/GA	13.5	17.8
10 CG/CG	13.5	15.1
11 GA/TC	13.5	17.8
12 AA/TT	13.1	18.6
13 AC/GT	12.9	18.3
14 CT/AG	12.7	19
15 TA/TA	11.7	19.5
16 AG/CT	11.1	19.1
17 GT/AC		
18 TG/CA		
19 GC/GC		

Table 7. DNA glycosylase activities of the mutants

Mutants	Substrate dsDNA	k (min^{-1})*	Relative activity (%)
Y68F (control)	24 bp	0.062 ± 0.003	100
	500 bp	0.038 ± 0.007	100
	3000 bp	0.052 ± 0.006	100
Y68F R70D	24 bp	0.079 ± 0.004	127
	500 bp	0.029 ± 0.008	76
	3000 bp	0.010 ± 0.001	19
Y68F D71R	24 bp	0.046 ± 0.003	74
	500 bp	0.013 ± 0.004	34
	3000 bp	0.010 ± 0.002	19
Y68F R70D D71R	24 bp	0.065 ± 0.005	105
	500 bp	0.027 ± 0.007	71
	3000 bp	0.021 ± 0.003	40
Y68F R26A	24 bp	0.040 ± 0.001	65
	500 bp	0.021 ± 0.006	55
	3000 bp	0.014 ± 0.002	27

*The enzymatic rate constants and their standard errors were obtained from a single-exponential fit to the data from three independent measurements.
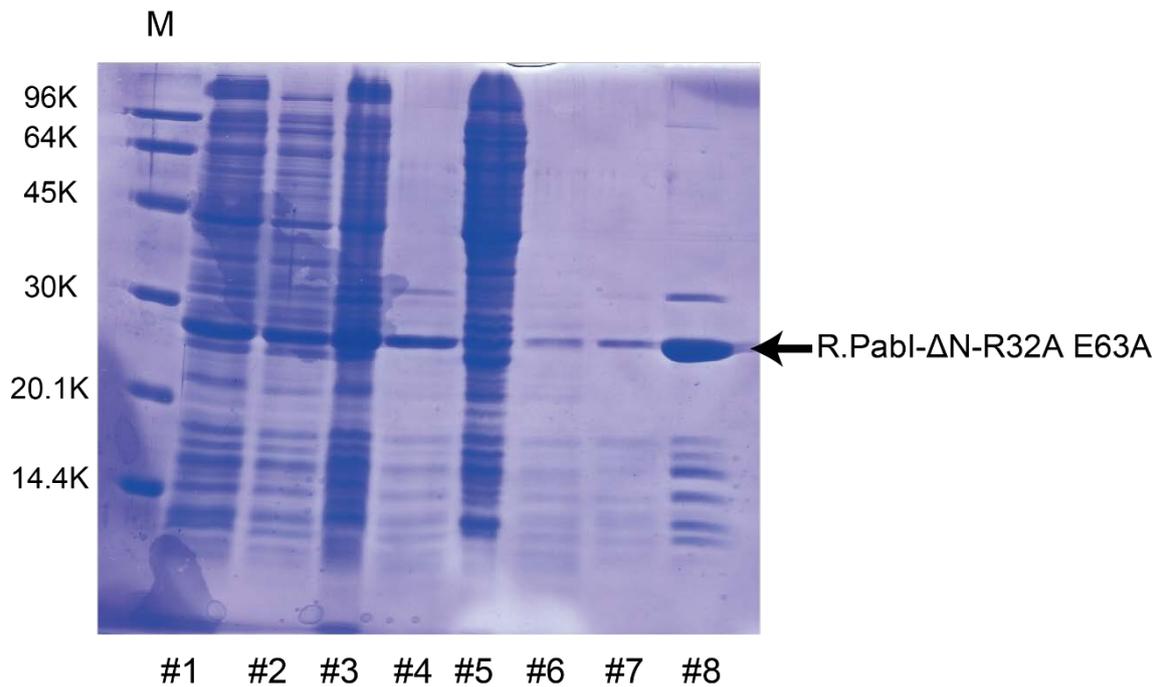
Figure 8. The SDS-PAGE result of the purification of R.PabI-ΔN-R32A E63A.

M: The molecular weight marker

#1 The lysis solution of the *E.coli*   #2 The supernatant after the centrifugation

#3 The precipitate after the centrifugation   #4 The supernatant after the heat treatment

#5 The precipitate after the heat treatment   #6 The unbound fraction of the Heparin purification

#7 The wash fraction of the Heparin   #8 The elution fraction of the Heparin purification
purification
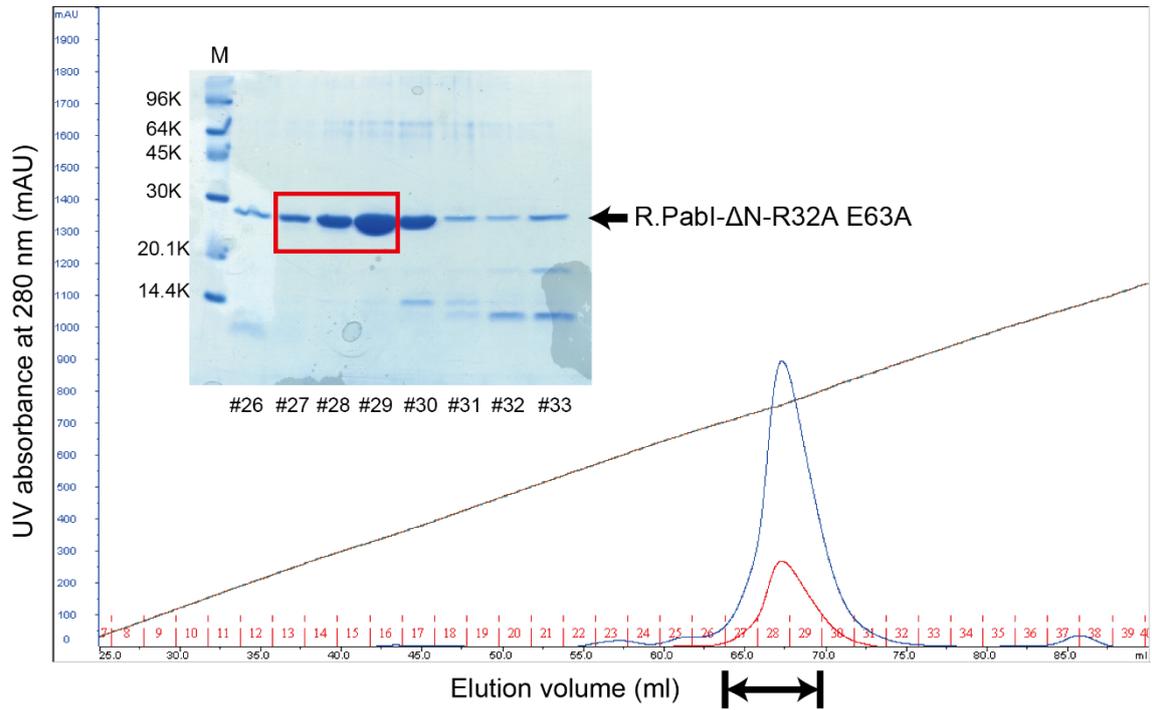
Figure 9. The result of the monoS column purification. The blue line and the red line represent for the absorbance at 280 nm and 260 nm, respectively. The #26~#33 fractions in the chromatogram were checked by the SDS-PAGE, the #27~#29 fractions were collected. R.PabI was eluted at a condition which contains 620 mM NaCl.
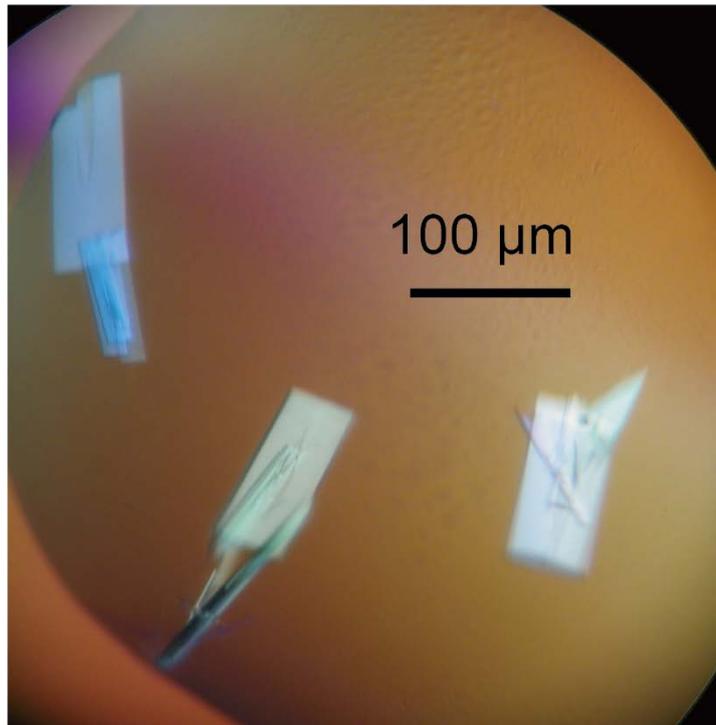
Figure 10. The crystal of the R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex.
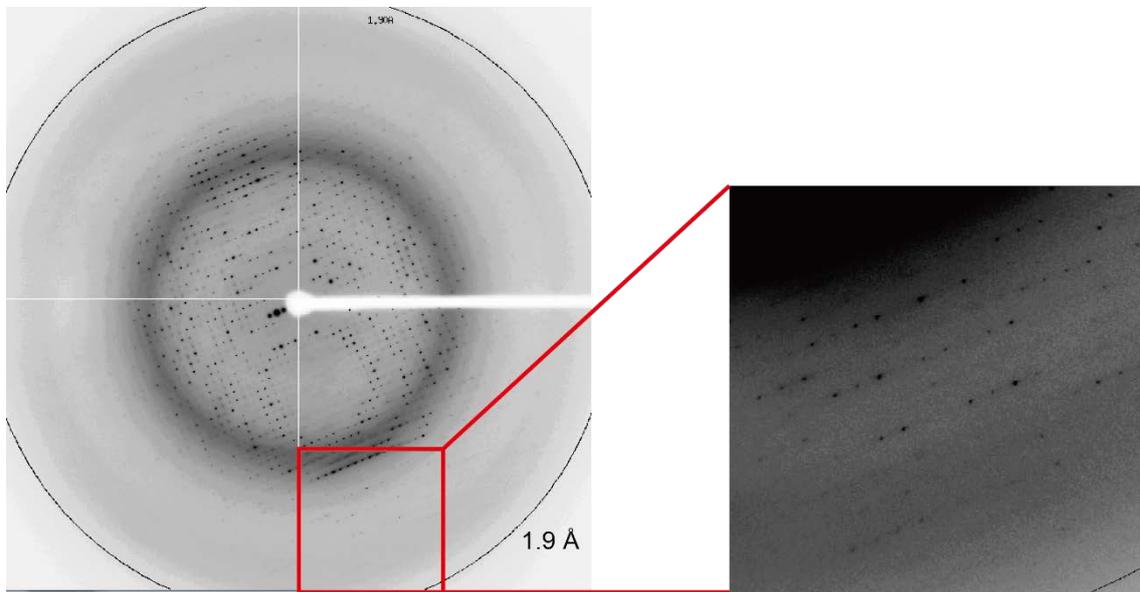
Figure 11. The X-ray diffraction image of the R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex crystal. The position corresponding to 1.9 Å is exhibited as a black circle.
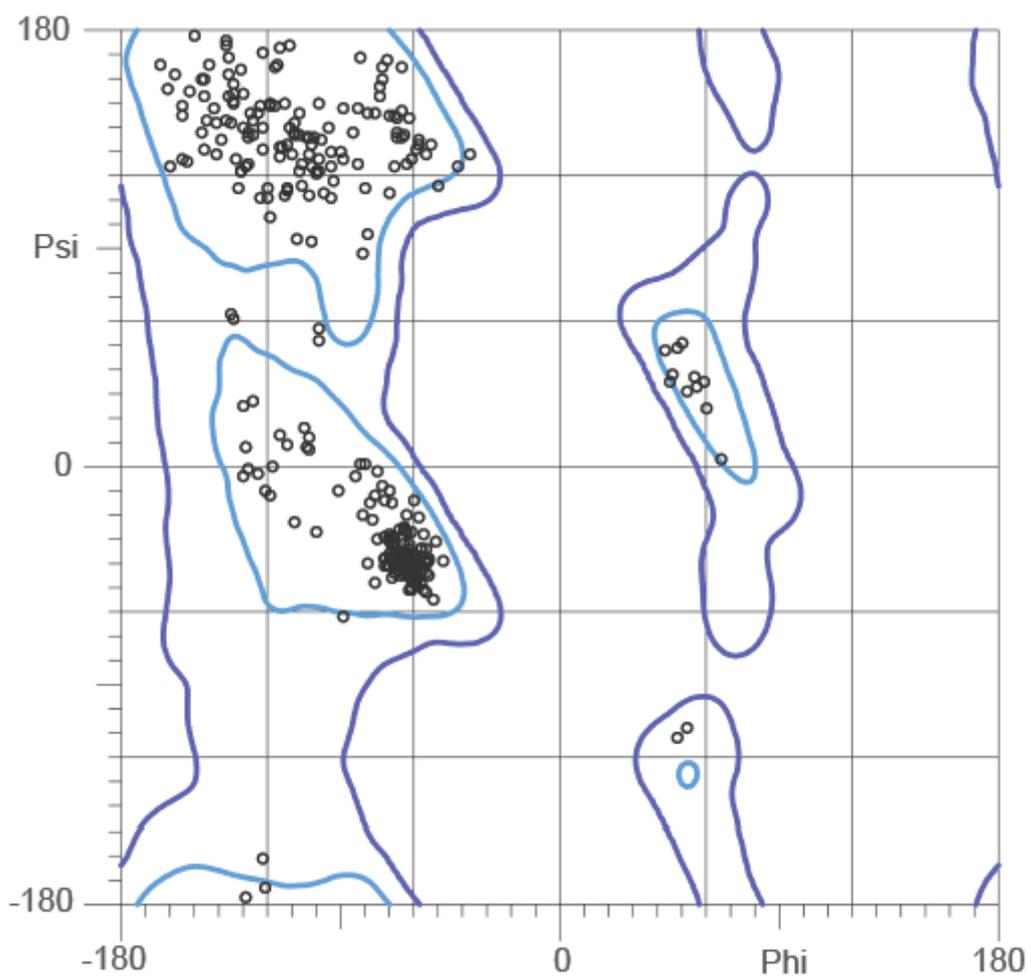
Figure 12. The Ramachandran plot of the R.PabI-ΔN-R32A E63A-nonspecific dsDNA complex structure. 97.2% residues are in favored region; 100% residues are in allowed region.
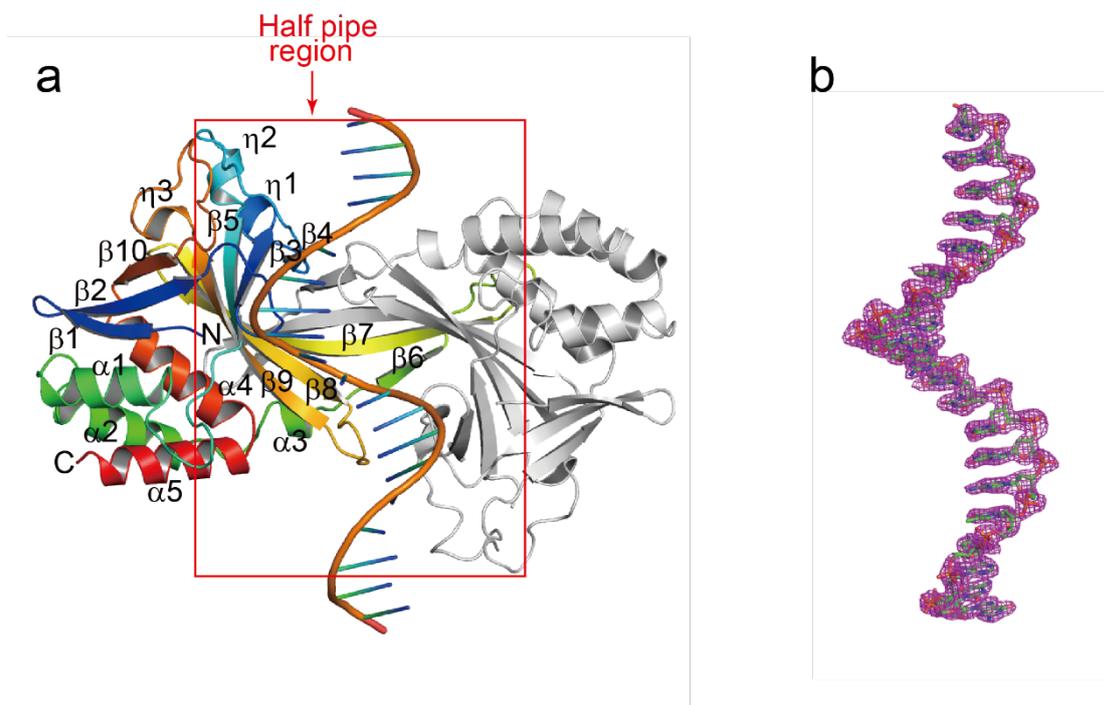
Figure 13. Overall structure of the R.PabI-nonspecific dsDNA complex. (**a**) Complex structure in the asymmetric unit. One R.PabI protomer is colored blue (in the N-terminus) to red (in the C-terminus). The other R.PabI protomer is colored grey. The bound ssDNA is colored orange. Secondary structure assignments of protomer A are labelled on the model. The half-pipe region is indicated by a red box. (**b**) Composite omit map (1 σ, magenta) of the ssDNA in the asymmetric unit. The composite omit map was created by CCP4[34].
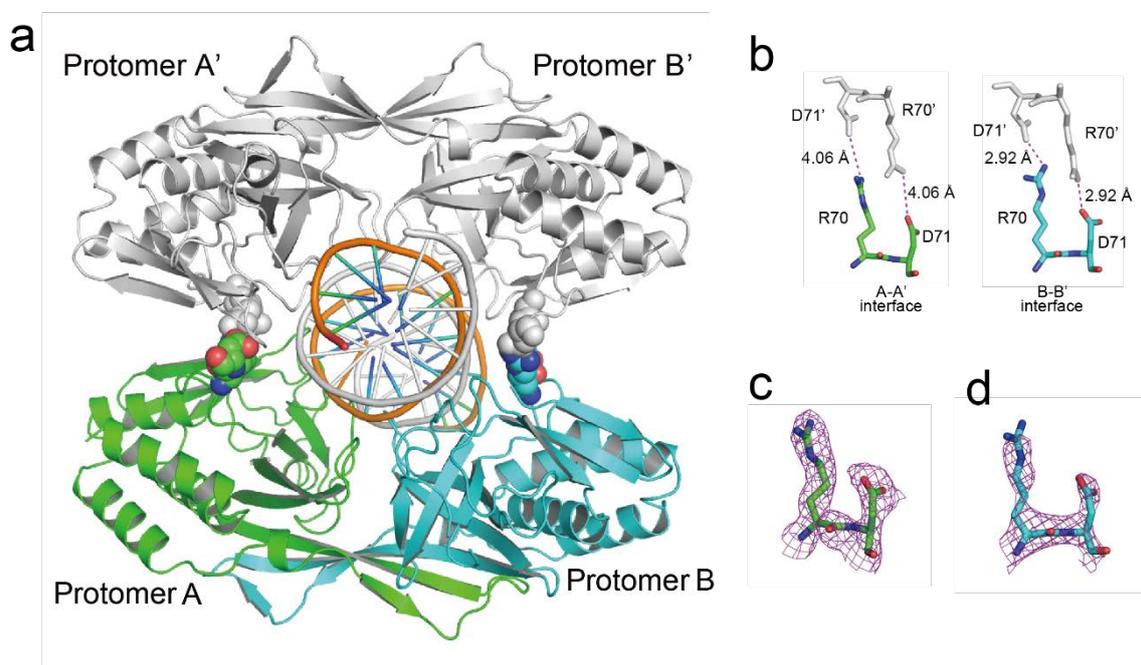
Figure 14. Overall structure of the R.PabI-nonspecific dsDNA complex. (**a**) Tetrameric structure of the R.PabI-nonspecific dsDNA complex. R.PabI protomers A and B in one asymmetric unit are colored green and cyan, respectively. R.PabI protomers generated by a symmetry operation (protomers A' and B') are colored grey. The dsDNA between the two R.PabI dimers is colored orange (chain C) and grey (chain C'). R70 and D71, which form inter-dimer salt bridges, are shown by a sphere model. (**b**) Salt bridges between the two R.PabI dimers. Residues of protomers A and B are shown as green and cyan stick models, respectively. Residues of protomers A' and B' are shown as a grey stick model. Intermolecular salt bridges are shown as magenta dotted lines. (**c**) Composite omit map (1σ, magenta) of Arg70 and Asp71 of protomer A. (**d**) Composite omit map (1 σ, magenta) of Arg70 and Asp71 of protomer B.
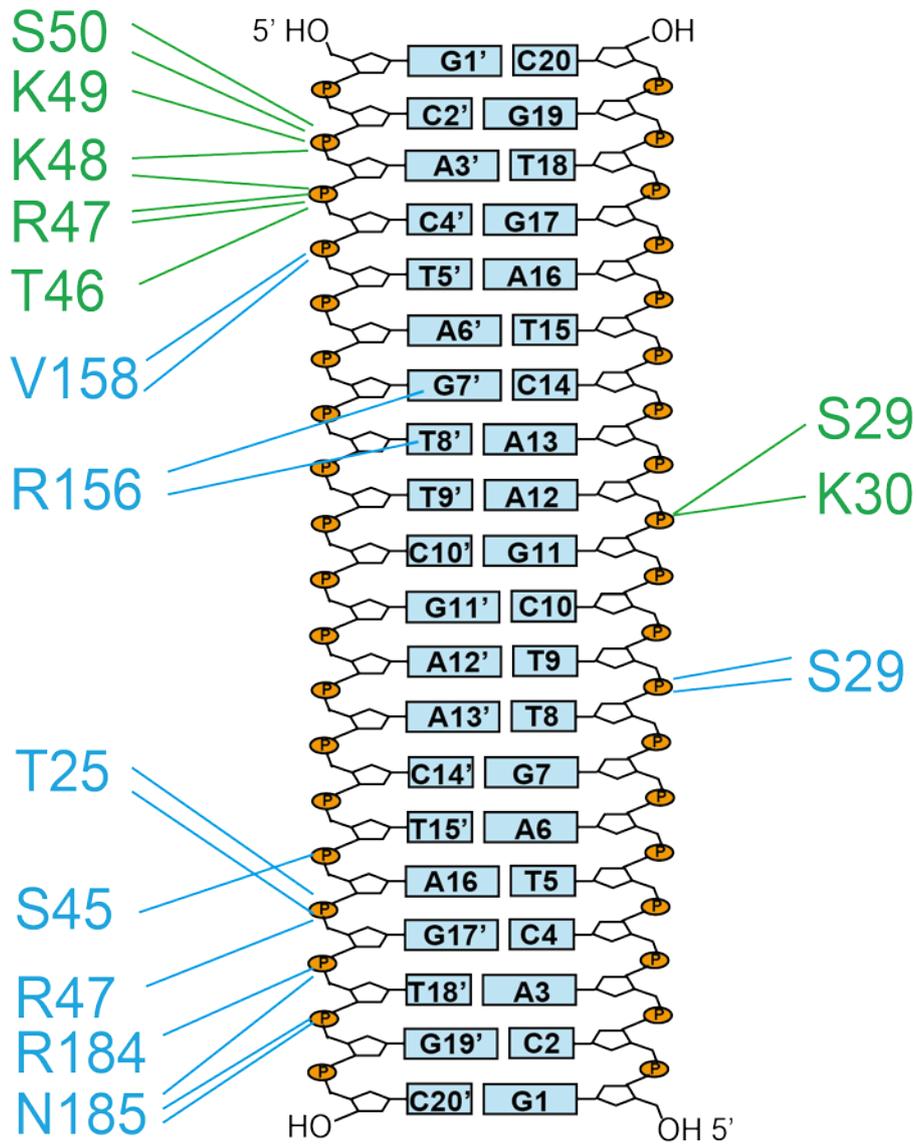
Figure 15. Intermolecular hydrogen bonds between one R.PabI dimer and dsDNA. Hydrogen bonds between protomer A and the dsDNA are shown as green lines. Hydrogen bonds between protomer B and the dsDNA are shown as cyan lines. Residues of protomers A and B are colored green and cyan, respectively.
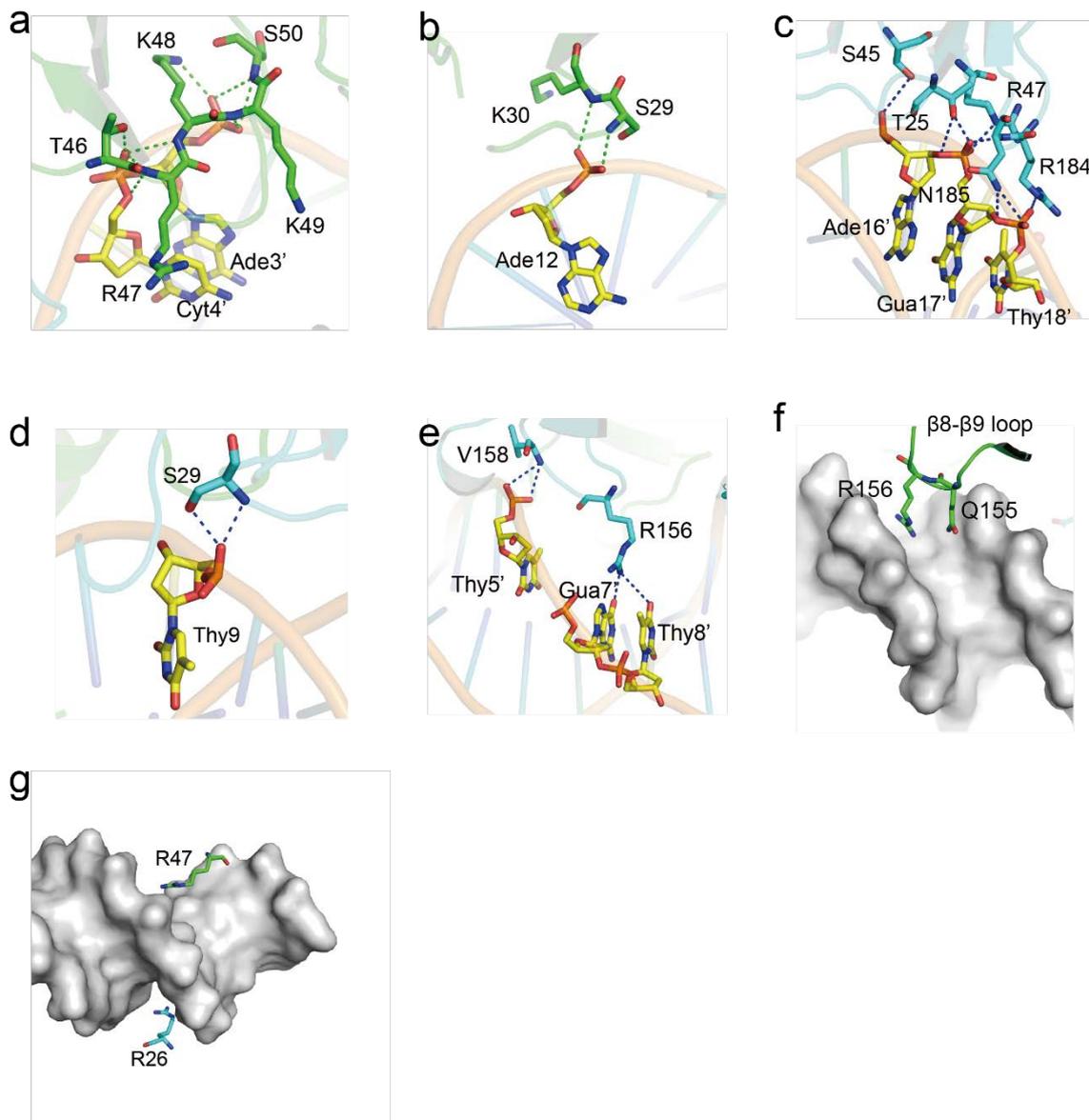
Figure 16. R.PabI-DNA interaction. (**a**, **b**) Intermolecular hydrogen bonds between R.PabI protomer A and dsDNA. Hydrogen bonds are shown as green dotted lines. Residues of protomer A and DNA bases are colored green and yellow, respectively. (**c~e**) Intermolecular hydrogen bonds between R.PabI protomer B and dsDNA. Hydrogen bonds are shown as blue dotted lines. Residues of protomer B are colored cyan. (**f**, **g**) R.PabI residues located above the major or minor groove of the dsDNA. The dsDNA is shown as a grey surface.
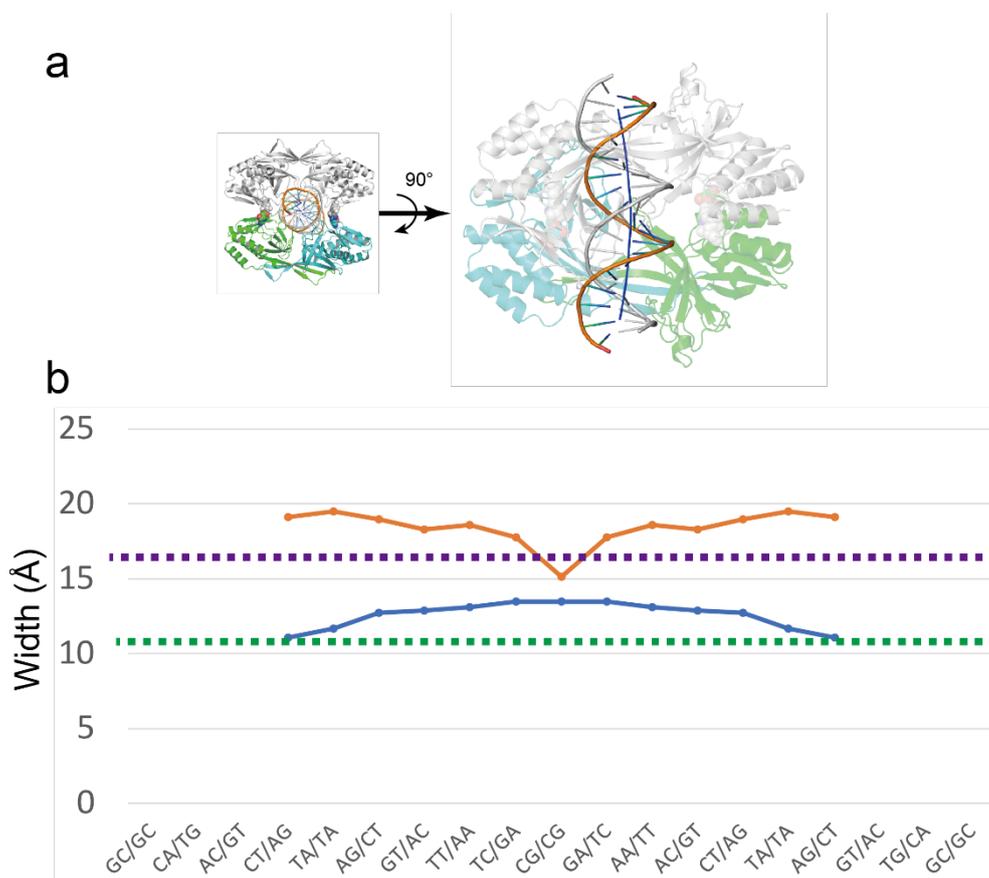
Figure 17. DNA structure in the R.PabI-nonspecific dsDNA complex. (**a**) The structure of dsDNA between the two R.PabI dimers. R.PabI and DNA are colored in the same way as Fig. 14a. DNA axis is shown as a blue line. (**b**) Plots of major (orange line) and minor (blue line) groove width as a function of DNA base step. The major and minor groove widths of ideal B-form dsDNA are shown as purple and green dotted lines, respectively.
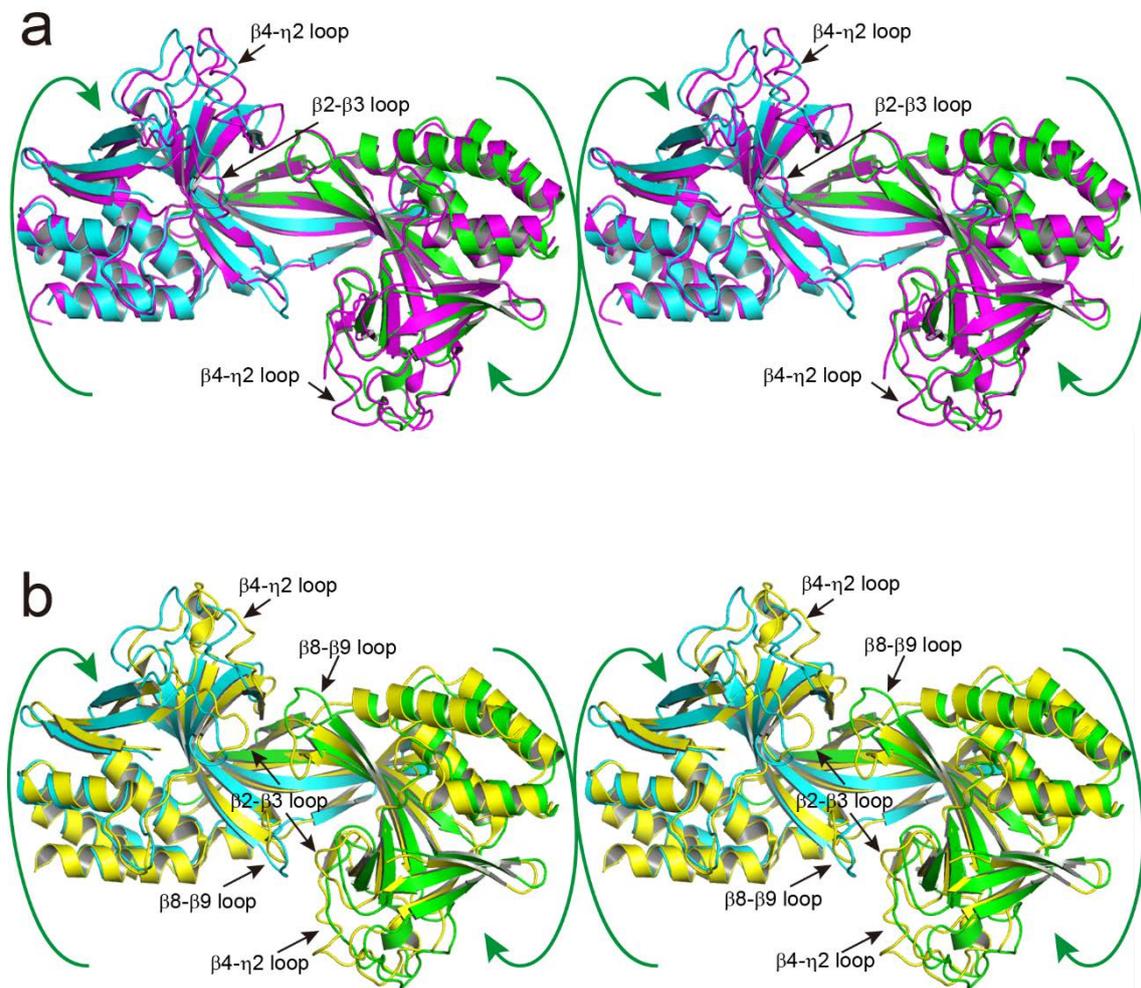
Figure 18. Structure comparison of R.PabI in the DNA-free state, the nonspecific dsDNA binding state, and the specific dsDNA binding state. (**a**) Wall-eyed stereo image of the superposition of the DNA-free state (magenta) and the nonspecific DNA binding state (green and cyan) of the R.PabI dimers. (**b**) Wall-eyed stereo image of the superposition of the nonspecific DNA binding state (green and cyan) and the specific DNA binding state (yellow) of the R.PabI dimers.
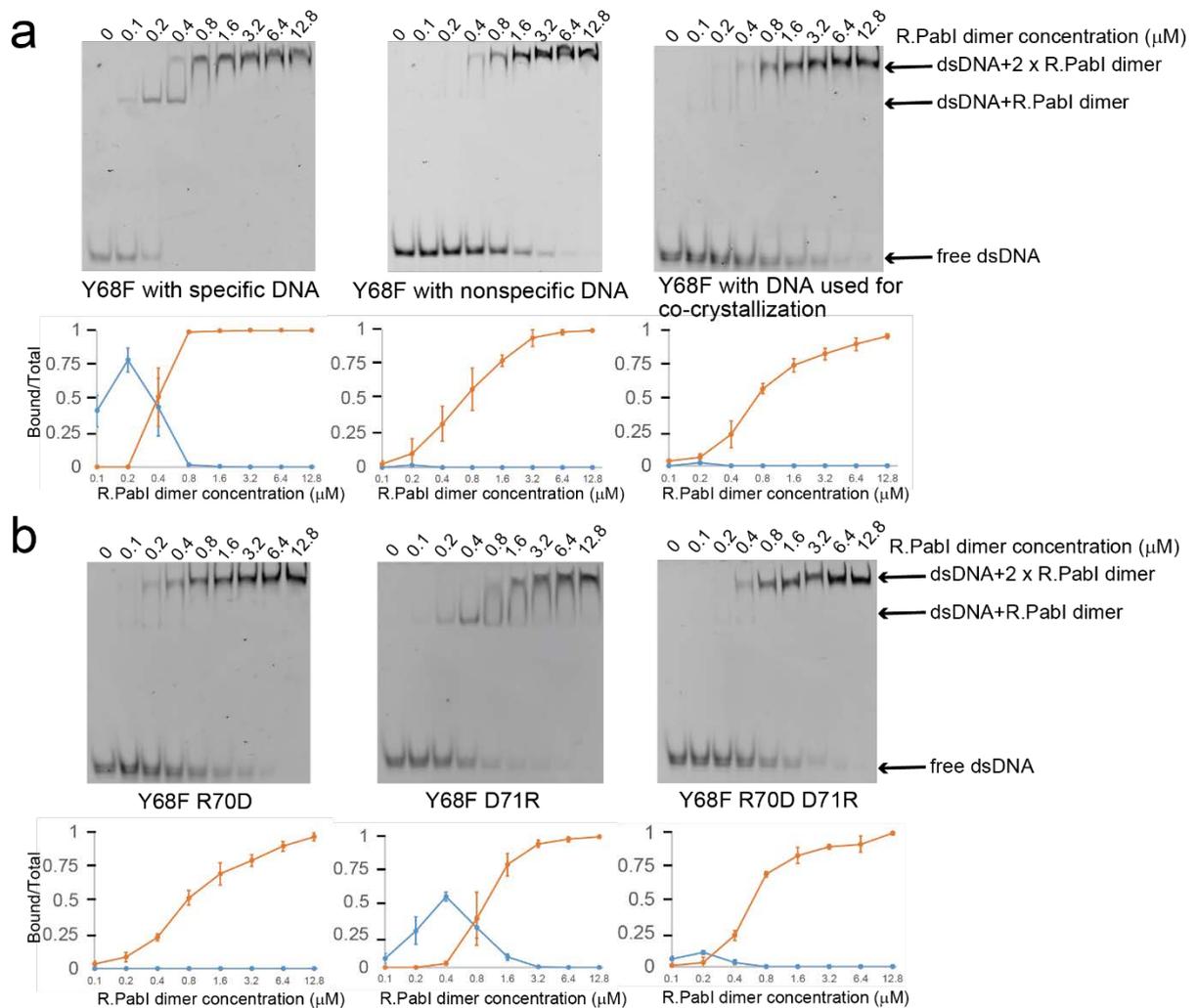
Figure 19. Electrophoretic mobility shift assay. (**a**) 0.1 μM of fluorescein-labelled dsDNA and each concentration of R.PabI were mixed and separated on a 12% polyacrylamide gel. Quantifications show the ratio of DNA bound by the R.PabI dimer (blue) and the R.PabI tetramer (orange). Plotted values are mean ± SD (n=3). (**b**) 0.1 μM of fluorescein-labelled dsDNA (the same sequence as the DNA used for co-crystallization) and each concentration of R.PabI mutants were mixed and separated on a 12% polyacrylamide gel.
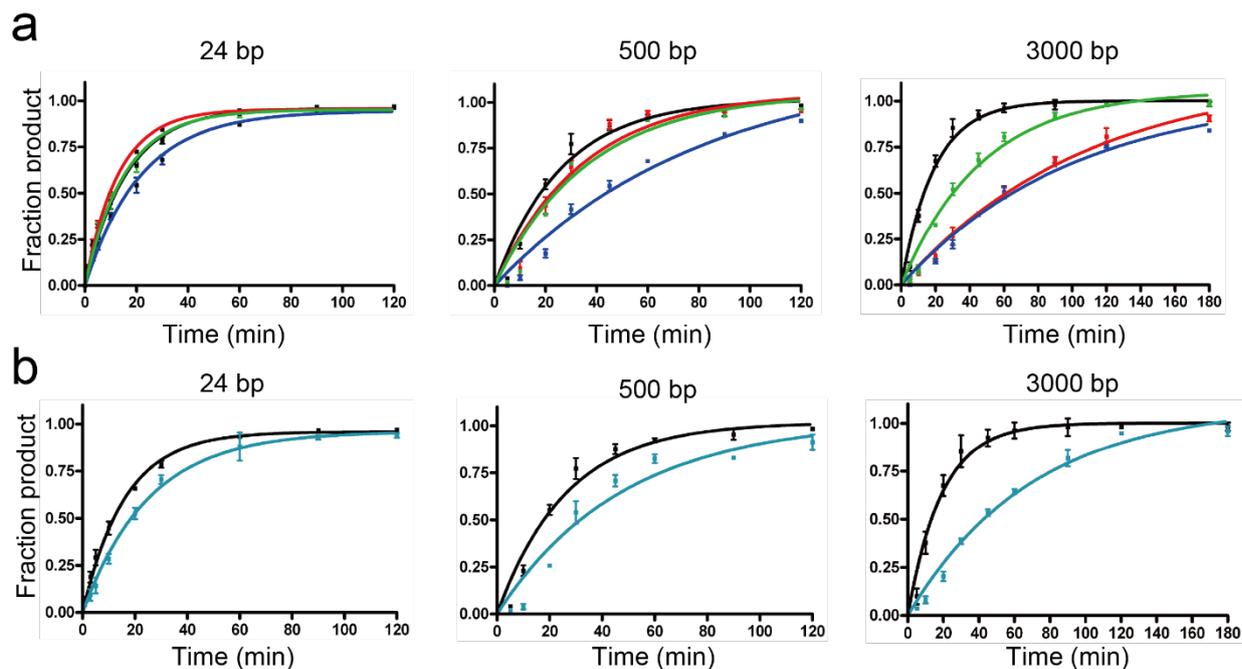
Figure 20. DNA glycosylase assay using various lengths of dsDNA. (**a**) DNA glycosylase assays of the Y68F (black), Y68F R70D (red), Y68F D71R (blue), and Y68F R70D D71R (green) mutants using 24 bp, 500 bp, and 3000 bp dsDNA as substrates. Plotted values are mean ± SD (n=3). (**b**) DNA glycosylase assays of the Y68F (black) and R26A Y68F (cyan) mutants using 24 bp, 500 bp, and 3000 bp dsDNA as substrates. Plotted values are mean ± SD (n=3).
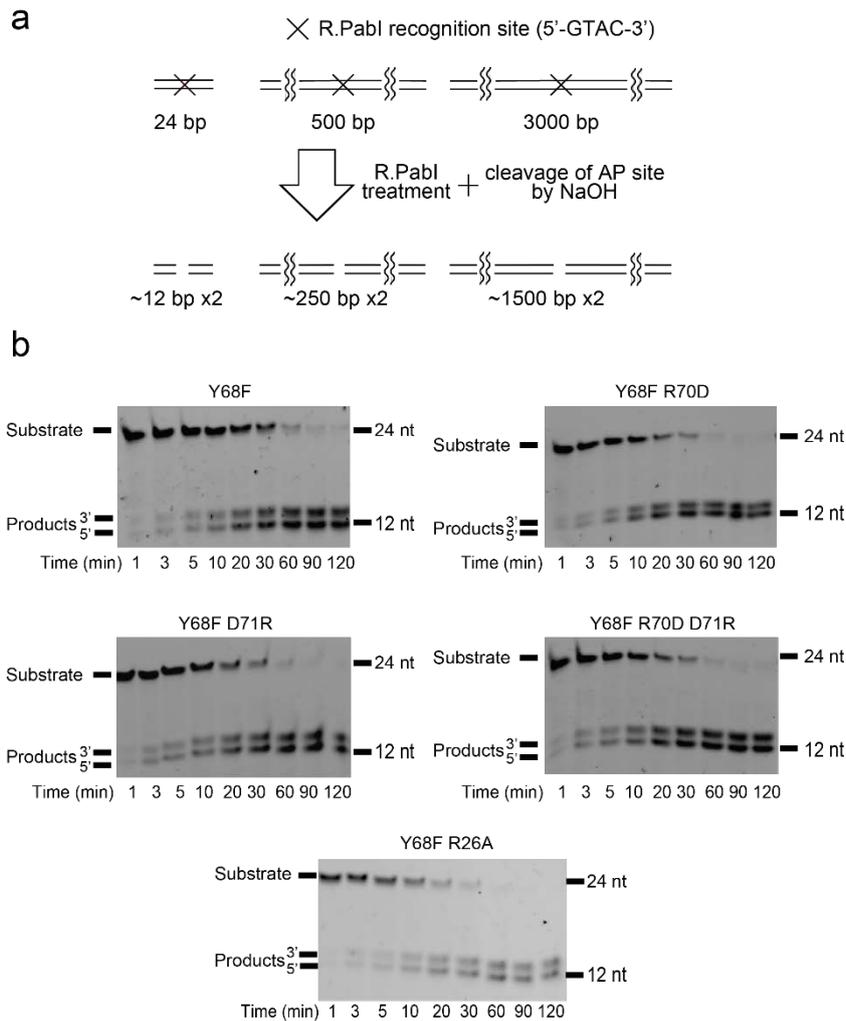
Figure 21. DNA glycosylase activities of the R.PabI mutants. (**a**) Schematic diagram of three R.PabI substrates (24 bp, 500 bp, and 3000 bp) used in the DNA glycosylase assays. The recognition sequence (5'-GTAC-3') is located at the centre of each substrate. After R.PabI treatment and the cleavage of AP sites by NaOH, these substrates are separated into two products with equal lengths. (**b**) DNA glycosylase activity assay using 24 bp dsDNA as a substrate. First, 0.2 μM of the 24 bp substrate dsDNA and 0.4 μM of the R.PabI dimer were mixed and incubated at 45°C. Each experiment was repeated three times.

66
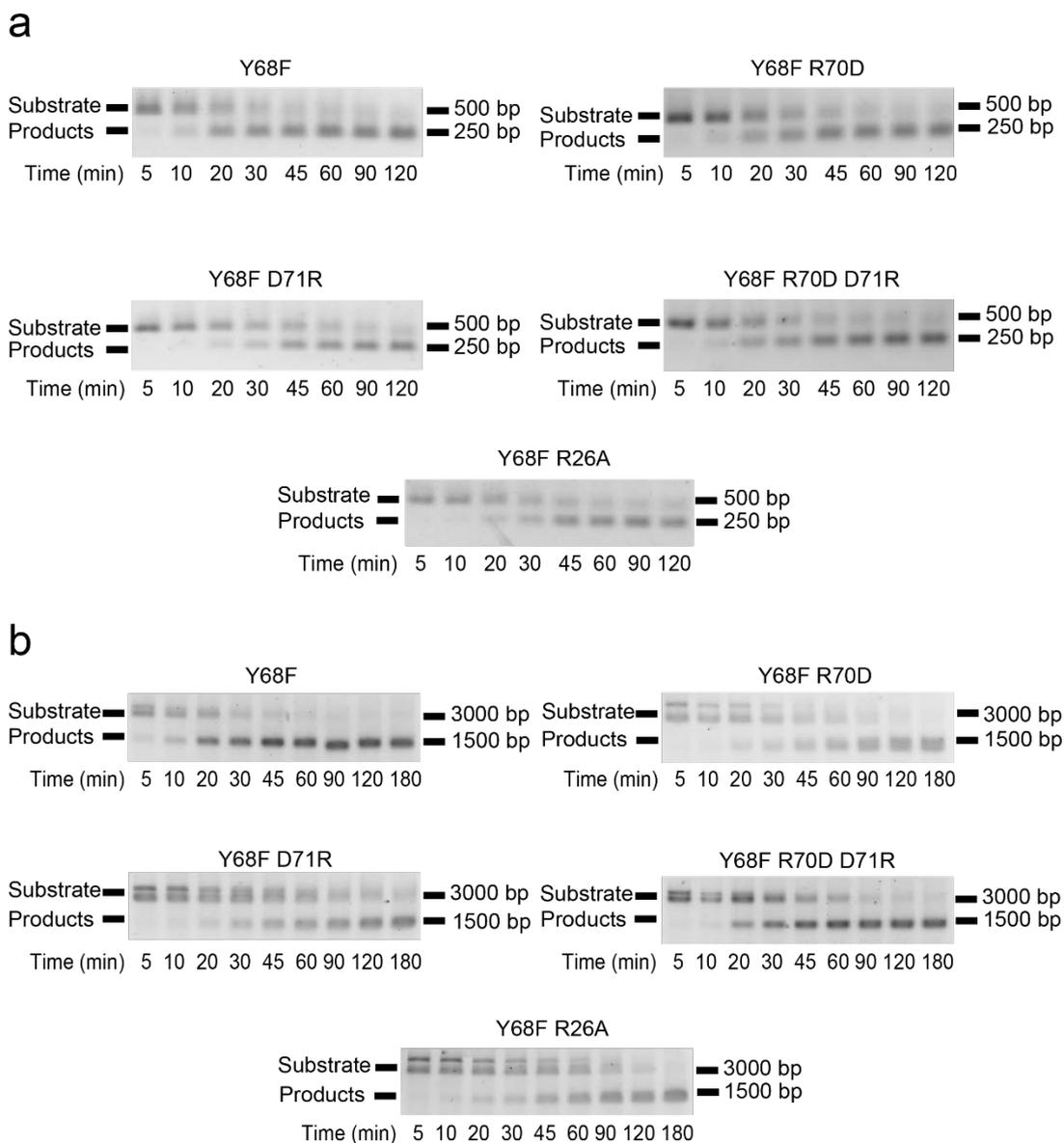
Figure 22. DNA glycosylase activities of the R.PabI mutants. (**a**) DNA glycosylase activity assay using 500 bp dsDNA as a substrate. First, 5.9 nM of the substrate DNA and 80 nM of the R.PabI dimer were mixed and incubated at 45°C. Each experiment was repeated three times. (**b**) DNA glycosylase activity assay using 3000 bp dsDNA as a substrate. First, 5.9 nM of the substrate DNA and 80 nM of the R.PabI dimer were mixed and incubated at 45°C. Each experiment was repeated three times.

4. Discussion

4-1 R.PabI forms tetrameric structure on nonspecific dsDNA

In this study, the crystal structure of the R.PabI-nonspecific dsDNA complex has been determined at 1.9 Å resolution. In the complex structure, R.PabI forms a tetrameric structure on nonspecific dsDNA using four salt bridges, although it binds dsDNA as a dimer at the recognition sequence[13]. The EMSA and DNA glycosylase activity assays using the R.PabI mutants showed that the residues that stabilize the tetrameric structure by the salt bridges are important for finding the recognition sequence in the sea of nonspecific dsDNA sequences. Restriction endonucleases such as EcoRV and BamHI bind nonspecific dsDNA weakly using their widened dsDNA binding cleft to search for their recognition sequences; each restriction endonuclease modifies the structure at their recognition sequences to form base-specific DNA interactions and to hydrolyze phosphodiester bonds at specific sites[2,4,17]. The results indicates that the restriction DNA glycosylase R.PabI also utilizes the nonspecific dsDNA binding state to facilitate diffusion similarly to EcoRV and BamHI, but the tetramerization on nonspecific dsDNA is a unique feature of R.PabI. EcoRV and BamHI, as well as most other restriction enzymes except for NgoMIV, which forms a tetramer at the product binding state[46], bind dsDNA as dimers in both the specific and the nonspecific dsDNA binding states.

4-2 Tetrameric form of R.PabI may facilitate its sliding on DNA

At the nonspecific dsDNA binding state, two R.PabI dimers bind the same region of nonspecific dsDNA to form the tetrameric structure. The tetrameric structure is stabilized by four salt bridges using R70 and D71 of each protomer. The crystal structure of the R.PabI-nonspecific dsDNA complex shows that the structure of dsDNA between the two R.PabI dimers is deformed by the binding of R.PabI, even in the nonspecific dsDNA

binding state (Fig. 17). Although R.PabI may also bind nonspecific dsDNA as a dimer, similarly to EcoRV and BamHI, the deformed structure of nonspecific dsDNA by the binding of one R.PabI dimer would facilitate the binding of the other R.PabI dimer at the same site because the two R.PabI dimers bind the same deformed nonspecific dsDNA site symmetrically, and the resulting tetrameric structure is stabilized by salt bridges (Fig. 14a,b). Actually, the results of EMSA showed that R.PabI binds nonspecific dsDNA as a tetramer, except for the Y68F D71R mutant, which would interfere with the tetramerization of R.PabI by both electrostatic repulsion and steric hindrance (Fig. 19, 20). Sequence-specific DNA binding proteins search for their target sequences using a sliding mechanism that facilitates 1D diffusion along nonspecific dsDNA[25,26] and/or a hopping mechanism that facilitates 3D diffusion[27–29]. Because the tetramerization of R.PabI stabilizes the R.PabI-nonspecific dsDNA complex using the toroidal structure, the tetrameric structure of R.PabI might facilitate not the hopping efficiency but the sliding efficiency for seeking the recognition sequence.

In this study, the results have shown that the salt bridges formed between R70 and D71 are important for the nonspecific dsDNA dependent tetramerization of R.PabI. However, these two residues are not conserved among the R.PabI homologs except the one from hyperthermophilic archaea *Staphylothermus hellenicus* (Fig. 1). Some thermostable proteins form multimers to increase their thermal stabilities[47–49]. Because R.PabI and its homolog from *S. hellenicus* function at high temperature, these proteins may utilize tetrameric structures to increase the stability of the protein-nonspecific dsDNA complex and thus facilitate sliding along dsDNA to search for their recognition sequences efficiently.

Chapter Two

1. Introduction

1-1 The function and categorization of DNA glycosylase

When the nucleobase of DNA undergoes oxidation or alkylation caused by reactive oxygen species (ROS) or chemical oxidants, the bases would be oxidized, alkylated or modified into aberrant bases. These aberrant bases include 8-OHG, 8oxoG, FapyG, mFapyG and so on. Once these modifications are produced, they would block the polymerase entrance or generate mispairing during the replication. For example, thymine glycol generated by the oxidation of thymine blocks the access of the polymerase, the generation of 8oxoG (8-oxo-7,8-dihydroguanine) leads to the G→T transversion mutation because 8oxoG has due ability of pairing with both A and C[50]. Since usually these mutations would cause lethal problems to the existence of living organisms, a system that could recognize the mutation and repair it in time is very indispensable. Living organisms have evolved several methods of repairing these lesions. Among these methods, the base excision repair (BER) pathway plays the most important role by removing the largest amount of DNA damages. The critical first step in BER is carried out by glycosylase. Glycosylase could recognize the lesion in DNA and cleave the *N*-glycosidic bond between the base and the ribose-phosphate backbone to generate apurinic/apyrimidinic (AP) site. The AP site would be further processed by AP endonuclease which could cleave the phosphodiester bond at 5' and 3' of the AP site[50]. The glycosylase could be categorized into several structural superfamilies: EndoV[51], UDG[52], HhH[53], H2TH[54], AAG[55] and ALK[56]. Although these superfamilies differ from each other in their structures, their mechanism of cleaving the *N*-glycosidic bond are similar: After glycosylase encounters the recognition sequence, a loop which links two α helixes would insert into the minor groove of DNA, the double helix is thereafter unwound and aberrant base is flipped out of the DNA backbone. The flipped out base is

then stabilized in a pocket of glycosylase formed by β sheets, the loop which extruded the base stays inside the double helix to stabilize the phosphate backbone of the flipped out base. After that, the bond between the aberrant base and its phosphate-ribose backbone is cleaved. Previous researchers found that nearly all of the glycosylase cleave DNA in such a mechanism except the AlkD family, which does not possess the base binding pocket hence does not form direct contact with the aberrant base, it is assumed of binding to the complementary strand and kinking the DNA in order to extruding the lesion base out of the double helix, so as to let the contact with solution to cleave the *N*-glycosidic bond[50,56,57].

1-2. The mechanism of glycosylase recognizing and deforming the target sequence

The structures of glycosylases forming recognition complexes with lesion DNA show that the DNA undergoes dramatic conformational change due to the interactions with the proteins. Usually, the DNA would be bent by 30° ~ 90°, the double helix is unwound and the aberrant base is flipped out[50]. Since it is hard to imagine that glycosylase would cause such a huge conformational change on every nonspecific site on DNA to search for its recognition site, a highly efficient mechanism by which glycosylase distinguishes its target site from nonspecific sites seems to be necessary. In addition, according to the induced-fit mechanism, after glycosylase encounters its recognition site, both protein and DNA would go through huge conformational deviations to transform into the base extrusion form. Structural biology had been a very useful method to elucidate the recognition mechanism of glycosylase and the conformational deviations that happens after the recognition. Until now, numbers of crystal structures of glycosylase-DNA complex had been solved. MutM, which recognizes and cleaves the oxoG formed by the

oxidation of guanine, is one of the well-studied glycosylases. Complex structures of MutM forming searching complex with nonspecific DNA as well as encounter complex and extrusion complex with lesion DNA have been solved. Based on the searching complex and the encounter complex structure, MutM recognizes its target aberrant base by discerning the difference of the base-stacking properties between the aberrant base and the normal base[58,59]. Once it encounters its recognition site, it bends DNA by around 80° at the lesion site[60], at the same time, a phenylalanine residue is inserted into the DNA groove to buckle the aberrant base in order to facilitate its extrusion, then, a methionine and a histidine insert into the DNA groove to extrude the aberrant base out of the DNA helix and stabilize it in a binding pocket in MutM[61,62]. Meanwhile, the encounter structure of the restriction DNA glycosylase has yet been solved.

1-3 The purpose of this study

The previous experiments have solved the complex structure of R.PabI binding to both specific and nonspecific DNA. These structures have provided a landscape of R.PabI searching and cleaving its recognition sequence. They showed R.PabI claps DNA using a tetrameric form and slides along it[63], they also showed that R.PabI unwinds DNA to flip the adenine out of the double strand to cleave its *N*-glycosidic bond[13]. However, since R.PabI and the DNA must be going through huge conformational deviations when they change from the nonspecific binding form to the specific binding form, the solvation of the two complex structures is not enough to grasp the whole process of the interaction. Thus, the observation of a binding state which reflects R.PabI encountering its recognition sequence and starting to deform it would be important to understand the whole recognition mechanism of R.PabI. Although several crystal structures have shown the

encounter complex of glycosylase binding to the recognition DNA, however, as has been described above, R.PabI is a restriction DNA glycosylase, it recognizes a specific sequence and cleaves the *N*-glycosidic bond of the adenine, this character makes it being dramatically different from the normal glycosylases, therefore, it is difficult to predict its recognition mechanism by referring to the other glycosylases. Thus, the determination of the encounter complex structure of R.PabI would be necessary. In this research, the crystal structure of a low-active R.PabI mutant (K154A Y68F) forming encounter complex with a dsDNA which contains recognition sequence has been solved. The structural analysis indicates that after encountering its recognitions sequence, R.PabI sharply bends the DNA and opens the minor groove, it protrudes the β2-β3 loops into the expanded minor groove to maintain the expansion, the bending of the DNA and the opening of the minor groove would later facilitate the flipping and cleavage of the bases. In addition, the flexibility of the T-A step in the recognition sequence would facilitate the recognition of R.PabI.

2. Materials and Methods

2-1 The expression and purification of R.PabI Y68F K154A

2-1-1 The design of the R.PabI mutant used in the crystallization

According to the previous study, R.PabI K154A, a mutant which has decreased its activity by 40 times comparing with the wild type, is able to cleave the *N*-glycosidic bond in the crystal[13]. In order to obtain the encounter complex of R.PabI, the DNA is required to be uncleaved, therefore the mutant used in this experiment has to possess activity which is much lower than the K154A mutant. In the other researches, in order to study the encounter complex, the loop in glycosylase which stabilizes the extruded base are disordered to make sure that the glycosylase could recognize DNA but loses its ability of cleaving DNA[60]. As to R.PabI, the previous research has shown that Y68, which locates in the loop between the β5 sheet and the α1 helix, helps stabilizing the cleaved adenine. Meanwhile, D214, which exits at the α5 helix, participates in the cleavage of the *N*-glycosidic bond (Fig. 23a). It had been shown that by mutating the Y68 into phenylalanine, the cleavage activity of R.PabI would decrease by 112 times, meanwhile the DNA specific binding ability is still maintained (Fig. 23b,c). As to D214, the activity of its asparagine mutant is also decreased by 75 times, similar to Y68F, its specific DNA binding ability is also as high as the wild type[13] (Fig. 23b,c). Therefore, the K154A Y68F, K154A D214N and K154A Y68F D214N mutants were utilized in the crystallization experiment, it is expected that they might form encounter complex with DNA which contains recognition sequence.

The pET28a-R.PabI-ΔN-K154A was provided by the previous researcher and was used as a template in this experiment. In this construct, the R.PabI K154A sequence was inserted into the NdeI-BamHI site of pET28a (Novagen), according to the previous research, the 1~7 residue of the R.PabI is disordered and could not form specific

secondary structure[10], so these residues, as well as the His-tag at the N-terminal, were deleted from the plasmid. The three mutants (K154A Y68F, K154A D214N and K154A Y68F D214N) were created by point mutating this template (Fig. 24c). The point mutation experiments were performed in the same way as 2-5-1 in chapter 1. The sequence of primers used in the point mutation are written in Table 8.

2-1-2. The overexpression of the R.PabI mutants

The overexpression of the R.PabI mutants were performed in the same way as 2-1-2 in chapter 1.

2-1-3. The purification of R.PabI mutants

The purification procedure of R.PabI mutants were the same as the previous experiments (2-1-3 in chapter 1). The purified R.PabI K154A Y68F, K154A D214N, K154A Y68F D214N were stored at the -80℃ deep freezer until use.

2-2. The crystallization of R.PabI mutants in complex with 20bp dsDNA

In order to creating the intermediate complex in which R.PabI encounters its recognition sequence and starts to bend it to form the extrusion complex, DNA with recognition sequence (5'-GTAC-3') was used in the experiments. Its sequence was: 5'-CGGCATAGCTGTACAGCTATGC-3' (20bp + 2nt sticky end, with the CG step in the 5' being the sticky end and the recognition sequence in the middle). The oligonucleotides were purchased from Eurofins Genomics, the purification method was OPC® cartridge. The purchased oligonucleotides were dissolved in annealing solution containing 2.5 mM MES (pH 6.0), 20 mM NaCl and 2.5 mM $MgCl_2$ to reach a concentration of 50 μM, then

it was heated to 368K and cooled down to 277K to anneal the oligonucleotide to let them form double strand DNA. The final concentration of the double strand DNA was 25 μM.

The purified R.PabI mutants was mixed with DNA solution and was concentrated to 180 μM as dimer. The molar ratio of R.PabI dimer and DNA in the concentrated solution was 1:1. Before the crystallization, ion-exchange was performed to change the buffer of the solution into 10 mM MES pH 6.0, 100 mM NaCl.

The mixed solution was utilized in the crystallization experiment. Firstly, the first screening was performed using the following kits: Crystal Screen HT and Index HT (both are made by Hampton Research), Wizard I&II (Emerald BioStructures), The JCSG+ suite, The JSGCore I Suite, The JCSG Core II suite, The JCSG Core III Suite, The JCSG Core IV Suite (QIAGEN). In the first screening, the crystallization robot Gryphon (Art Robbins Instruments) was used to deliver the protein solution and the reservoir, the Sitting-drop plate used in the crystallization was VCP-1 (Violamo). Vapor-diffusion method was used to perform the crystallization experiment, the plates were kept in 20°C.

Among all of R.PabI mutants-DNA complexes , the crystal of the R.PabI-ΔN-K154A Y68F-DNA complex has been detected at E3 well of the Index kit (0.1 M BIS-TRIS pH 6.5, 45% v/v (+/-)-2-Methyl-2,4-pentanediol and 0.2 M Ammonium acetate). Based on this condition, the second crystallization screening was performed by optimizing the concentration of the precipitant and the pH of the BIS-TRIS buffer. The plates used in the second screening were 24 well sitting-drop Cryschem Plates (Hampton Research), the method was vapor-diffusion, the same as the first screening. The plates were kept in 20°C.

2-3 The structure determination of R.PabI-ΔN-K154A Y68F in complex with 20bp dsDNA

X-ray diffraction data were collected at the BL17A beamline of the Photon Factory (Tsukuba, Japan) under cryogenic conditions (95K). The complex crystal was picked up by Dual-Thickness MicroLoop (MiTeGen). After that, the data were collected at the following condition:

The wavelength of X-ray: 0.98000 Å

The detector: ADSC Quantum 315r

The detector distance: 500 mm

The exposure time: 0.2 s

The oscillation range: 0.1°

The data range: 1~1800

The crystal of the R.PabI-ΔN-K154A Y68F-dsDNA complex diffracted X-ray to 2.5 Å. The X-ray diffraction data were indexed and integrated using the program XDS[33] and scaled using SCALA in the CCP4[34] suite. The crystal of the R.PabI-ΔN-K154A Y68F-dsDNA complex belongs to the space group $C2$ with unit cell parameters of a=143.48, b=82.77, c=140.43. The initial model was determined by the molecular replacement method using the program MOLREP[35] with the coordinates of the R.PabI-nonspecific DNA complex structure (PDB code: 5iff). The initial model was refined and rebuilt using the program Phenix.reine[36] and Coot[37].

2-4 The crystallization of R.PabI-ΔN-K154A Y68F in complex with 23bp dsDNA

The structure of the R.PabI-ΔN-K154A Y68F-20bp dsDNA complex, although been determined at a relatively high resolution (2.5 Å), did not show clear electron density of the DNA nucleotides so that the order of the base pairs could not be elucidated by referring to the form and size of the electron density. In addition, the electron density in the terminal

of the dsDNA was extremely week, hence it is unclear where does the DNA sequence start and where does it end. These problems made it hard to elucidate the sequence of the dsDNA in this complex. To solve these problems, the length of the DNA sequence was extended by 3 base pairs to enable it to stack with the neighbor DNA, in such a situation, it is assumed that the terminal of the dsDNA sequence would be clearly observed. The sequence of the 23bp double stranded DNA was composed of two complementary single strands: 5'-TCAGCAGTACTAAGTACTGCTGA-3' and 5'-TCAGCAGTACTTAGTACTGCTGA-3' (Fig. 24a), the single strands contain two recognition sequences (5'-GTAC-3') which separate from each other by 3 base pairs. The oligo nucleotides were purchased from the Eurofins Genomics and were purified by the oligonucleotide purification cartridge (OPC). The purchased oligonucleotides were dissolved in annealing solution which contains 2.5 mM MES (pH 6.0), 20 mM NaCl and 2.5 mM MgCl$_2$, two single strands were mixed together, then they were heated to 368K and cooled down to 277K in order to anneal the oligonucleotide to let them form double strand DNA. The final concentration of the double strand DNA was 25 μM.

The purified R.PabI-ΔN-K154A Y68F was mixed with DNA solution and was concentrated to 180 μM as dimer. The molar ratio of R.PabI dimer and DNA in the concentrated solution was 1:1. Before the crystallization, ion-exchange was performed to change the buffer of the solution into 10 mM MES pH 6.0, 100 mM NaCl.

To make sure that the molecules in the crystal would pack in the same pattern as the R.PabI-ΔN-K154A Y68F-20bp dsDNA complex, the same crystallization condition (same composition of reservoir, same temperature, same concentration of protein-DNA solution and same volume ratio of protein : reservoir) was used this time. However, the complex crystal has not been obtained. Therefore, streak seeding was performed to obtain

the crystal. Firstly, 24 conditions of crystallization reservoir which differ from each other by either concentration of the precipitant or the pH of the BIS-TRIS buffer were prepared, all of the conditions were deviated from the original condition of the E3 well (0.1 M BIS-TRIS pH 6.5, 45% v/v (+/-)-2-Methyl-2,4-pentanediol and 0.2 M Ammonium acetate) of the index HT kit (Hampton Research). Secondly, the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex solution was mixed with the reservoir by the volume ratio of 1.5 μl : 1.5 μl. Then, an animal hair was used to transfer the seeds of the previously obtained crystal of the R.PabI-ΔN-K154A Y68F-20bp dsDNA complex to the new wells. Then the plate was sealed and incubated at 20℃. two days after the seeding, crystals were observed in the wells.

2-5 The structure determination of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex
2-5-1 The X-ray diffraction of R.PabI-ΔN-K154A Y68F-23bp dsDNA complex
X-ray diffraction data were collected on the BL17A beamline at the Photon Factory (Tsukuba, Japan) under cryogenic conditions (95K). The complex crystal was picked up by Dual-Thickness MicroLoop (MiTeGen). After that, the data was collected at the following condition:

The wavelength of X-ray: 0.98000 Å

The detector: ADSC Quantum 315r

The detector distance: 485 mm

The exposure time: 0.2 s

The oscillation range: 0.1°

The data range: 1~1800

2-5-2 The structure determination of R.PabI-ΔN-K154A Y68F-23bp dsDNA complex

The crystal of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex diffracted X-ray to 2.4 Å. The X-ray diffraction data were indexed and integrated using the program XDS[33] and scaled using SCALA in the CCP4[34] suite. The crystal of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex belongs to the space group $P3_221$ with unit cell parameters of a=83.38, b=83.38, c=140.32. The initial model was determined by the molecular replacement method using the program MOLREP[35] with the coordinates of the previously obtained R.PabI-ΔN-K154A Y68F-20bp dsDNA complex structure. The initial model was refined and rebuilt using the program Phenix.reine[36] and Coot[37].

2-5-3 The valuation of the complex structure

The interaction between the protein and the DNA was analyzed by PISA[38]. The comparison of different molecules, the measurement of the distances between atoms and the display of structures were performed by Pymol (http://www.pymol.org/). The width of the DNA groove and the angle of the stacking of bases was calculated by 3DNA[39]. Molprobity[40] was used to make the Ramachandran plot.

2-6 The design of the R.PabI mutants

Based on the structure of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex, several important amino acid residues have been identified. To prove the importance of these resides, the mutants of these residues were produced. Similar to the chapter 1, the pET28a-R.PabI-Y68F was used as the template of these mutants (Fig. 24d). Sequences of the primers used to produce these mutants are written at Table 8. The overexpression as well as purification of these mutants were performed using the same method as 2-1 of

the chapter 1.

2-7 The glycosylase activity assay

The glycosylase activity assay of R.PabI mutants were performed using the same method

as 2-7 of the chapter 1.

Table 8. Primers used in producing the R.PabI mutants

The primers used to produce K154A Y68F (R.PabI-ΔN-K154A was used as template)	Forward primer：5'- ATATCT<u>TT</u>CGCGAGGGATGGAAAATAT -3' Reverse primer：5'- CCTCGC<u>GAA</u>AGATATCTGCCATTCAAT -3'
The primers used to produce Y68F T28G (R.PabI-ΔN-Y68F was used as template, the same as the two mutants below)	Forward primer：5'- CGTCCT<u>GGA</u>TCAAAAATTAGGGTGAAA -3' Reverse primer：5'- TTTTGA<u>TCC</u>AGGACGCGTTATTGGTAA -3'
The primers used to produce Y68F P27G T28G	Forward primer：5'- CGTGGA<u>GGATCA</u>AAAATTAGGGTGAAA -3' Reverse primer：5'- TTT<u>TGATCC</u>TCCACGCGTTATTGGTAA -3'
The primers used to produce Y68F T28W	Forward primer：5'- CGTCCT<u>TGG</u>TCAAAAATTAGGGTGAAA -3' Reverse primer：5'- TTTTGA<u>CCA</u>AGGACGCGTTATTGGTAA -3'

The mutation sequences are underlined
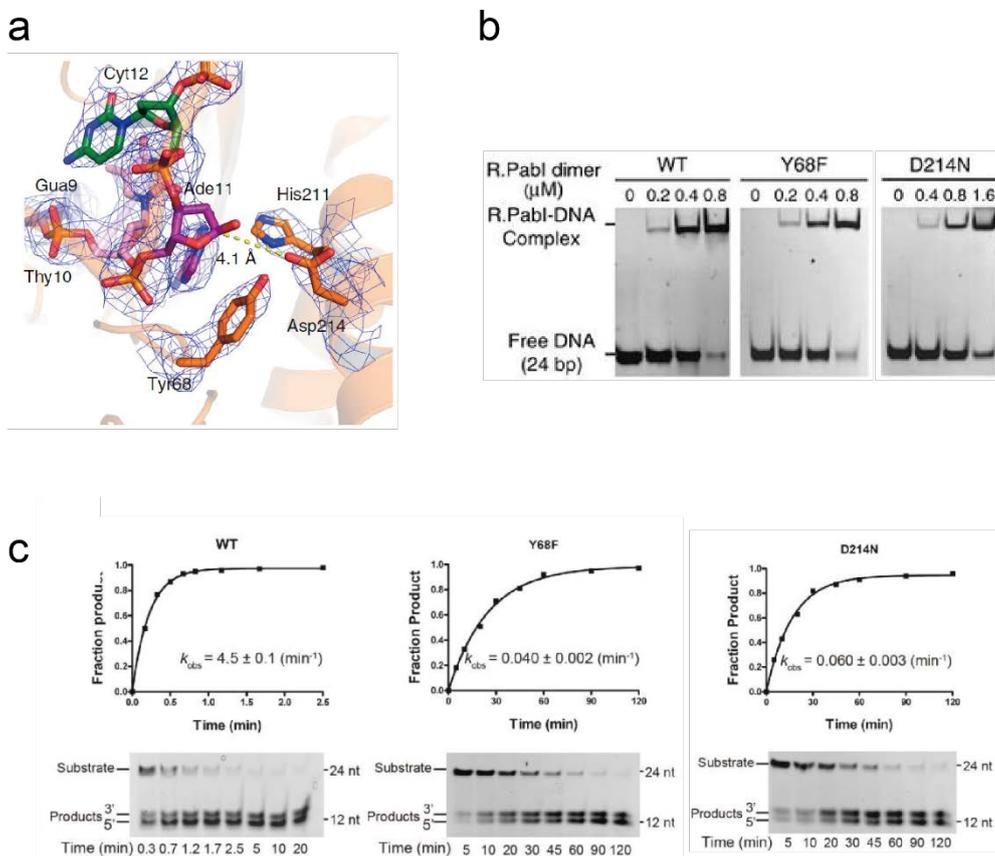
Figure 23. The results from the previous research[13]. (**a**) Positions of Y68 and D214 in the R.PabI-specific dsDNA complex, the *N*-glycosidic bond of Ade11 has already been cleaved. (**b**) DNA binding assay of the R.PabI wild type, Y68F and D21N. (**c**) Glycosylase activity assays of the R.PabI wild type, Y68F and D214N.
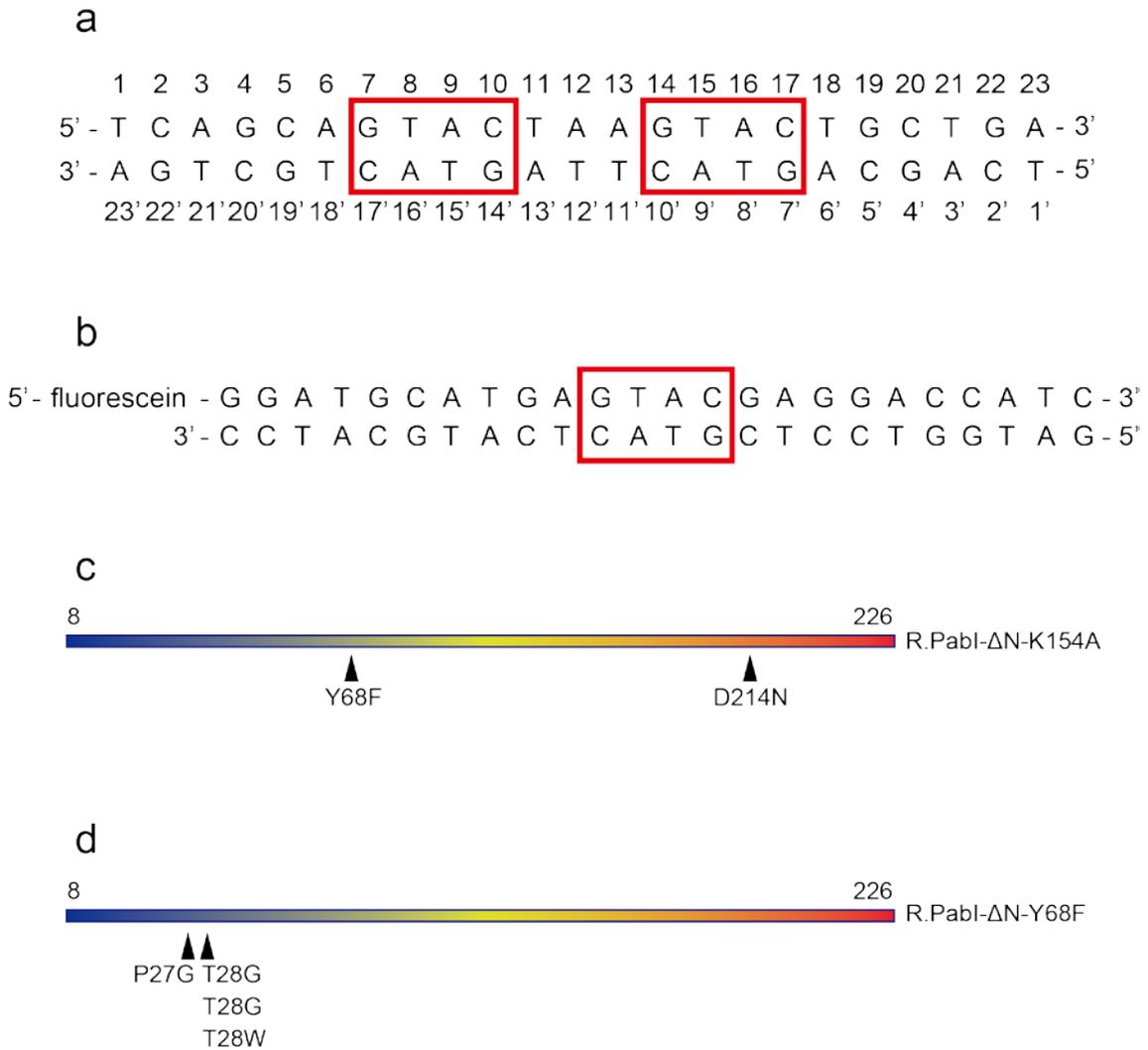
Figure 24. Sequences of dsDNA and constructs of R.PabI. (**a**) Sequence of dsDNA used in the crystallization experiments, the recognition sequences are shown in red boxes (**b**) Sequence of dsDNA used in the glycosylase activity assays. (**c**) Construct of R.PabI mutants used in the crystallization experiments. (**d**) Construct of R.PabI mutants used in the glycosylase activity assays.

3. Results

3-1 Crystallization and structural determination of the R.PabI-ΔN-K154A Y68F-20bp dsDNA complex

The R.PabI mutants (K154A Y68F, K154A D214N, K154A Y68F D214N) were successfully expressed by *E.coli*. After that, heat treatment, heparin resin purification and the cation exchange chromatography were performed to purify the protein. Finally, the SDS-PAGE was performed to check the purity of the protein. In the SDS-PAGE gel, only one band existed at the position of 26 kDa which is similar to the molecular weight of R.PabI protomer. This result indicated that the R.PabI mutants had been successfully purified. Although the expression amount of each mutant differ from each other, all of them could be produced for more than 5mg by the incubation of 2L *E.coli*.

As has been written in 2-2, all of the three mutants were used to perform the crystallization experiment. The crystal of the R.PabI K154A Y68F-dsDNA complex was observed in the E3 well of the index HT kit. Crystal was not observed in the other conditions, the complex of DNA and other two mutants (K154A D214N and K154A Y68F D214N) also failed to form crystal. To be continued, the condition of E3 (0.1 M BIS-TRIS pH 6.5, 45% v/v (+/-)-2-Methyl-2,4-pentanediol (MPD) and 0.2 M Ammonium acetate) was further optimized in the second screening to obtain bigger crystal with higher quality. As a result, crystals were observed in the second screening plate where the pH of BIS-TRIS ranges from 6.1 to 6.7 and the concentration of MPD ranges from 43% to 48%. After the data collection and the processing, the complex structure was determined at 2.5 Å. In this structure, two R.PabI homodimers bind to double stranded DNA in two separate sites. The dsDNA is sharply bent at the two binding sites, although no base is flipped out of the double helix, the minor groove in the binding sites are largely expanded. The β2-β3 loops of each R.PabI protomer intercalate into the expanded minor groove and T28s

90

which exist in the loops are observed forming hydrogen bonds with DNA bases. Judged by these characters, the complex structure is believed to be in an encounter stage which happens after the searching stage and before the recognition stage. However, problem also exited for this complex: the structure did not show clear electron density of the DNA nucleotides so that the order of the base pairs could not be elucidated by referring to the form and size of the electron density. In addition, the electron density in the terminal of the dsDNA was extremely week, therefore, it is unclear where does the DNA sequence start and where does it end. To improve this problem, crystallization was performed again using dsDNA with different length and sequence.

3-2 Crystallization of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex

The previous reservoir condition was used to crystallize the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex. However, crystal was not obtained. In order to obtain the complex crystal, the streak seeding was performed. 1 day after the seeding, crystals were observed in the 24 well plate where the pH of BIS-TRIS ranges from 5.9 to 6.5 and the concentration of MPD ranges from 45% to 51% (Fig. 25).

3-3 Structural determination of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex

The crystal in the condition which contains 0.2 M Ammonium acetate, 0.1 M BIS-TRIS pH 5.9 and 45% MPD was used in the X-ray diffraction experiment. The crystal of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex diffracted X-ray at 2.4 Å (Fig. 26), after the index and integration, the space group to which the crystal belongs was determined to be $P3_221$ with $a=$ 83.379, $b=$83.379, $c=$140.322 (Table 10).

The molecular replacement was performed using the coordinates of the R.PabI-ΔN-

K154A Y68F-20bp dsDNA complex structure. After that, the structure of R.PabI-ΔN-K154A Y68F-23bp dsDNA complex was determined. The asymmetric unit of the crystal contains electron density corresponding to one R.PabI dimer and one single-stranded DNA. The sequence of DNA was built by the program Coot. The structure was refined to R_{work}=21.86%, R_{free}=26.69%, using the program Phenix.refine. In the complex structure, one homodimer (protomers A and B), one single-stranded DNA (ssDNA, chainC), 9 water molecules and one MPD molecule were observed in one asymmetric unit (Fig. 28, Table 10). The position of the terminal residue S226 in protomer A could not be determined because its electron density did not appear in the structure. On the other hand, the position of the initiating methionine, G16 and the 3 residues in the C-terminal (G224, T225 and S226) in protomer B also could not be determined because their electron density could not be observed. The program Molprobity was used to calculate the validity of the backbone dihedral angles, as a result of the calculation, 95.5% of residues were included in the favored region, 99.3% residues were in the allowed region (Fig. 27). The *B*-factor of each molecule in the asymmetric unit was calculated by the CCP4 suite and the result showed that The *B*-factor of protomer A is 71.89 Å, protomer B is 90.82 Å, DNA chain is 97.55 Å and the water molecule is 60.30 Å (Table 10).

3-4 Overall structure of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex

The structure of each protomer in the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex resembles the structure of protomers in the R.PabI-nonspecific DNA complex (PDB: 2DVY). In addition, the two protomers form homodimer in the same way as the previous structures. The server PISA[38] was used to calculate the contacting area of each molecule in the complex, the result indicated that the interface area between protomer A and DNA

chain is about 670.1 $Å^2$, the area between protomer B and DNA chain is about 611.0 $Å^2$, the area which protomers contact with each other is about 1765.2 $Å^2$.

The R.PabI-ssDNA complex observed in the asymmetric unit forms complex structure with a symmetrically related complex generated by the crystallographic threefold screw axes. The ssDNA in each asymmetric unit forms a dsDNA with two R.PabI homodimers binding to two separated sites. The other ssDNA is named chain C', two protomers in the other homodimer that bind to DNA are named protomers A' and B'. These homodimers bind to dsDNA in two sites. Since potomer A' and B' are generated by crystallographic symmetry operation and are completely identical to protomer A and B, their interactions with dsDNA are also identical to protomer A and B (Fig 29). Therefore, the discussions about the protein-DNA interactions as well as the complex structure comparison would only focus on protomer A, protomer B and the double stranded DNA in the complex.

Similar to the R.PabI-ΔN-K154A Y68F-20bp dsDNA complex structure, the dsDNA in the complex structure is severely bent at the binding sites, in addition, the minor groove at the binding sites are also widened. The terminal of the dsDNA chains (C and C') stack with the DNA chains in the neighbor molecules, this makes it possible to elucidate the DNA sequence since the beginning site and the ending site are clearly grasped (Fig. 30b,c). Furthermore, the electron density of the DNA bases is also unambiguously observed in the complex structure (Fig. 30a). As a result, unlike the previous complex structure, the sequence of the dsDNA in the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex structure has been determined.

3-5 The R.PabI-DNA interactions in the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex

According to the calculation of PISA, the interface area between the homodimer (protomer A and B) and dsDNA is 1281 Å^2, which is slightly larger than the interface area between the homodimer and dsDNA in the R.PabI-nonspecific dsDNA complex (1156 Å^2) but significantly smaller than in the R.PabI-specific dsDNA complex (2166 Å^2). 24 intermolecular hydrogen bonds are formed between the residues of the R.PabI homodimer (protomer A and B) and the nucleotides of the dsDNA (Table 11, Fig. 31). In protomer A, T25, R26, K30 of the β2-β3 loop interact with the ribose phosphate backbones of Ade6, Gua7, Thy8, Cyt20' and Thy21'; T46 and R47 of the β4-η2 loop interact with the phosphate backbones of Cyt20'; R184 and N185 of the η3-β10 loop interacts with the phosphate backbone of Thy21' and Gua22' (Fig. 32a,b). In addition, R156 of the β8-β9 loop and the initiating methionine M7 form hydrogen bonds with the phosphate backbone of Cyt10 and Thy8 respectively (Fig. 32b). On the other hand, in protomer B, T25, R26, K30 of the β2-β3 loop interact with the phosphate backbones of Ade13, Gua14', Thy15' and Ade16'; T46 and R47 of the β4-η2 loop interact with the phosphate backbone of Ade12; R184 of the η3-β10 loop forms hydrogen bond with the backbone of Cyt5' (Fig. 32d,e). In addition, the main chain of R184 in protomer A forms van der waals interactions with the phosphate backbone of Thy21' while R184 in protomer B also forms van der waals interactions with that of Ade13 (Fig.33). It is believed that these interactions formed by the residues of the β4-η2 loops and η3-β10 loops extrude the DNA backbone and force the DNA to bend toward the major groove.

It is noteworthy that T28 of the β2-β3 loop in protomer B forms two hydrogen bonds with Gua 14' and Thy15' (Fig. 32e). Meanwhile, although no hydrogen bonds exit

94

between T28 in protomer A and DNA bases, it is observed to have formed van der waals interaction with Gua7 (Fig. 32c). These are the only interactions being observed between the R.PabI residues and the DNA bases in the complex structure. In the nonspecific complex, T28 does not form any interaction with DNA, nor does the β2-β3 loop intercalate into the minor groove. However, in the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex, the β2-β3 loop of each protomer is positioned whithin the minor groove. In protomer A, the pyrrolidine of P27 is observed of having formed van der waals interaction with the ribose phosphate backbone of Gua19', the main chain of K30 also forms hydrogen bonds with Thy8 (Fig. 32f). In protomer B, van der waals interactions also are formed between the phosphate backbone of Thy11 and the pyrrolidine of P27. Similar to protomer A, hydrogen bond also is formed between the main chain of K30 and the phosphate backbone of Ade16' (Fig. 32g). These results indicate that along with the insertion of the β2-β3 loops into the minor groove, the residues in these loops repulse the DNA backbone outward to maintain the expansion of the minor groove.

3-6 The conformation of the dsDNA in the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex

Because of the interactions with R.PabI, the conformation of dsDNA in the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex is largely deviated from B-form DNA. As has been mentioned in the previous session, the minor groove is largely expanded. 3DNA[39] server was used in calculating the width of the minor groove, as a result, the widest site is expanded to 19.4 Å, while the width in typical B-form DNA is around 10 Å. In comparison to this, the major groove of the dsDNA in the complex does not go through

huge deviation, the width in the widest site is 20.0 Å and the narrowest site is 13.6 Å (the average width in the B-form DNA is around 17 Å) (Fig. 34).

3DNA server was also used to calculate the parameter of position and angle of each base pair. As a result, the T-A step in the recognition sequence (5'-GTAC-3') showed significant difference from the other steps in the Roll degree. Roll represents the bending angle between two base pairs which are next to each other. The Roll degree of the T-A step (42.2°) is significantly larger than the other steps as well as the B-form DNA (0°) (Table 12, Fig. 35a,b). This indicates that the unstacking of the T-A step contributes mostly to the bending of DNA in the complex structure. The base pairs in the recognition sequence have also been significantly untwisted, their average twist degree for one base pair is 21° (Table 12), much lower than the twist degree in a B-form DNA (36°). What is more, the Buckle angle between the T-A pair (24.05°) is also larger than the other base-pairs. The B-factor of each base in the recognition sequence is calculated by the Baverage program in the CCP4 suite and turned out that the B-factor of Ade16' is distinguishably higher than that of Ade9 in the T-A step of the recognition sequence (Fig. 35c). It has already been reported that in the encounter complex of MutM and dsDNA, the would-be-flipped-out base is hugely buckled from its complementary base[60–62]. Together, these results suggest that Ade16' is possibly flipped out earlier than Ade9 when the β8-β9 loops protrude into the minor groove.

3-7 The structure comparison between the R.PabI-nonspecific DNA complex and the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex

As a result of the structure comparison, although the R.PabI dimers in both complexes do not show significant difference (the r.m.s.d. between the R.PabI dimers in the two

complex is 2.8 Å for 399 superposed Cα atoms, Fig. 36a), the conformations of dsDNAs differ from each other hugely. The dsDNA in the nonspecific complex, despite being slightly bent by R.PabI in the middle, remains closed to the B-form DNA. Different from that, the dsDNA in the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex is bent by about 90° at the protein interacting sites (Fig. 36b,c), the minor groove at the protein binding sites are expanded to 19.4 Å (Table 12, Fig. 34), much wider than the minor groove in the R.PabI-nonspecific DNA complex (13.5 Å). Such a difference is assumed to be caused by the sharp bending of the DNA and the protrusion of the β2-β3 loops into the minor groove in the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex. In the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex structure, each residue in the β2-β3 loops moves closer to the minor groove while in the nonspecific complex, they stay outside of the DNA helix (Fig. 37a,b). Besides the positions of the loops, the other noteworthy change within the β2-β3 loops are the directions of the side chains of T28s. Comparing with the nonspecific complex, the direction of their side chains change by about 90° and are deeply inserted into the minor groove to interact with the bases (Fig. 37a,b). This deviation suggests that T28 does not assist the sequence recognition during the searching stage but may have important function after R.PabI encounters its recognition sequence. The role of T28 will be further discussed in the discussion session.

Comparing with the nonspecific complex, the positions of β4-η2 and η3-β10 loops, which take part in repulsing the DNA backbone, move closer to their identical loops in the other protomer (Fig. 36a). Similar to the nonspecific complex, the β8-β9 loops, which extrude the adenine and guanine out of the double helix in the specific complex, are

positioned outside of the helix and do not take part in any interaction with the DNA (Fig. 36a, 37c). This suggests that the β8-β9 loops starts intercalating into the minor groove after the bending of the DNA and the insertion of the β2-β3 loops.

3-7 The structure comparison between the R.PabI-specific DNA complex and the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex

The r.m.s.d. between the R.PabI dimers in the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex and the specific complex is 3.8 Å for 415 superposed Cα atoms. Each R.PabI protomer does not show large conformation deviation from the R.PabI-specific DNA complex (Fig. 38a). The dsDNA in the R.PabI-specific DNA complex also show resemblance to the specific complex: they all show similar bending angle at the recognition sequence (5'-GTAC-3') (Fig. 38c). However, although being seriously bent over, the double helix of dsDNA remains intact, no base is observed being flipped out like the R.PabI-specific DNA complex (Fig. 38b). In the R.PabI-specific DNA complex, the β8-β9 loops unwind the double helix of the dsDNA by deeply protruding into the minor groove and flipping the adenine and guanine outside of the helix [13]. However, the loops in the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex remain outside of the DNA helix and do not form any interactions with the DNA (Fig. 38a, 39c). Together, these results suggest that this binding state happens before the double helix is unwound.

The positions of the β2-β3 loops, which are assumed to have played important role in maintaining the expansion of the minor groove in the R.PabI-specific DNA complex, deviate slightly in the R.PabI-specific DNA complex as they also are deeply inserted into the minor groove and interact with the bases from inside of the minor groove (Fig. 39a,b).

In addition, the directions of the side chains of the residues in the loops are also similar, indicating that the loops do not go through conformational changes when R.PabI transfers to specific binding state with DNA. The β4-η2 and η3-β10 loops which are assumed pushing the DNA backbone inward to bend the DNA in the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex, remain contacting with the DNA backbone in the R.PabI-specific DNA complex[13](Fig. 39a,b), further indicating that they play important role in bending the DNA after R.PabI encounters its recognition sequence.

3-8 The results of the DNA glycosylase activity assays

In the complex structure, T28 of the β2-β3 loop is the only residue that forms intermolecular hydrogen bonds with bases of DNA. In addition, P27, which also exists in β2-β3 loop, is also positioned within the minor groove and forms van der waals interactions with DNA backbone. It is thought that the protrusion of the β2-β3 loops, which cause these residues to form interactions within the groove, stabilize the expanded minor groove so that the β8-β9 loops could insert into the expanded groove to unwind DNA. To evaluate the importance of the β2-β3 loop and T28. the DNA glycosylase activities of the Y68F T28W, Y68F T28G, and Y68F P27G T28G mutants were analyzed using 24 bp dsDNA (Fig. 24b,d). The results show that comparing with the Y68F mutant which was used as control, the activity of the T28G and T28W mutants are reduced to 52% and 37%, respectively (Table 13, Fig. 40, 41). This resut suggests that T28 plays important role in the DNA cleavage. When both P27 and T28 are mutated into glycine in the P27GT28G mutant, the activity is significantly reduced to 14% (Table 13, Fig. 40, 41). It is assumed that since the glycine mutant would eliminate the interactions formed between the side chains of the two residues and the DNA backbone, therefore the β2-β3

loop would lose its ability in maintaining the expansion of the minor groove, as a result, the activity of P27G T28G is severely reduced.

The DNA glycosylase activity of the Y68F K154A mutant was also analyzed using 24 bp dsDNA as substrate. As a result, its activity is reduced to 7% (Table 13, Fig. 40, 41), comparing with the Y68F mutant. This result indicates that although the Y68F K154A possess only little activity, it is still able to recognize and cleave the recognition sequence.

Table 9. Summary of the data collection statistics of the R.PabI-ΔN-K154A Y68F-23bp

dsDNA complex

Data collection	
X-ray source	Photon Factory BL 17A
Detector	ADSC Quantum 315r
Wavelength (Å)	0.9800 Å
Space group	$P3_221$
Unit cell parameters	
a, b, c (Å)	83.38, 83.38, 140.32
α, β, γ (°)	90, 90, 120
Resolution (Å)	46.77-2.40 (2.49-2.40)*
Wilson B-factor (Å^2)	53
Total number of reflections	198369 (14925)
Unique reflections	22650 (2266)
R_{sym} (%)	8.2 (72.2)
Mean $I/\sigma I$	13.1 (1.7)
Completeness (%)	99.7 (97.8)
Multiplicity	8.8 (6.6)

*The numbers in parentheses represent data for the highest-resolution shells.

Table 10. Summary of the refinement of the R.PabI-ΔN-K154A Y68F-23bp dsDNA

complex

Refinement	
Resolution (Å)	46.77-2.40 (2.49-2.40)*
No. of reflections in working/test set	198369 (14925)
R/R_{free} (%)	21.86 / 26.69 (32.31/ 34.86)
No. atoms	
Protein/DNA/water	3535 / 472 / 9
B-factors (Å^2)	
Chain A (protomer A)	71.89
Chain B (protomer B)	90.82
DNA (chain C)	97.55
water	60.30
r.m.s.d.	
Bond angle (°)	1.15
Bond length (Å)	0.006
Rotamer outliers (%)	0.30
C-beta outliers	0
Clash score	14.24

*The numbers in parentheses represent data for the highest-resolution shells.

Table 11. Hydrogen bonds between R.PabI and DNA in the R.PabI-ΔN-K154A Y68F-

23bp dsDNA complex

Protein				H-bonds	DNA			
Chain ID	Residue	Main/ Side	Atom Name	Distance (Å)	Chain ID	Base	Base/ Backbone	Atom Name
A	Met7	Main	N	3.09	C	Thy8	Backbone	OP1
A	Thr25	Side	OG1	2.41	C'	Thy21'	Backbone	OP1
A	Arg26	Side	NH1	3.82	C'	Cyt20'	Backbone	O3'
A	Arg26	Side	NH1	2.87	C'	Thy21'	Backbone	OP1
A	Arg26	Side	NE	2.98	C	Gua7	Backbone	OP2
A	Arg26	Side	NE	3.32	C	Ade6	Backbone	O3'
A	Lys30	Main	N	3.15	C	Thy8	Backbone	OP2
A	Ser45	Side	OG	3.87	C'	Gua19'	Backbone	O5'
A	Thr46	Main	N	3.22	C'	Cyt20'	Backbone	OP1
A	Arg47	Main	N	3.19	C'	Cyt20'	Backbone	OP1
A	Arg184	Side	NH2	3.14	C'	Thy21'	Backbone	OP2
A	Arg156	Side	NE	3.32	C	Cyt10	Backbone	OP1
A	Asn185	Side	ND2	3.54	C'	Gua22'	Backbone	OP2
B	Thr25	Side	OG1	2.56	C	Ade13	Backbone	OP2
B	Arg26	Side	NH2	2.76	C	Ade13	Backbone	OP2
B	Arg26	Side	NH1	3.77	C'	Thy15'	Backbone	OP2
B	Thr28	Side	OG1	3.65	C'	Gua14'	Base	N2
B	Thr28	Side	OG1	2.69	C'	Thy15'	Base	O2
B	Lys30	Main	N	3.69	C'	Ade16'	Backbone	OP1
B	Thr46	Main	N	2.82	C	Ade12	Backbone	OP2
B	Thr46	Side	OG1	3.14	C	Ade12	Backbone	OP2
B	Arg47	Main	N	3.72	C	Ade12	Backbone	OP1
B	Arg47	Main	N	3.76	C	Ade12	Backbone	OP2
B	Arg184	Side	NH2	3.52	C'	Cyt5'	Backbone	OP2

Table 12. The parameters of the dsDNA in the R.PabI-ΔN-K154A Y68F-23bp dsDNA

complex

Base pair Step	Width of minor groove (Å)	Width of major groove (Å)	Roll (°)	Twist (°)
1 TC/GA			5.09	31.12
2 CA/TG			1.20	45.52
3 AG/CT			-3.12	31.24
4 GC/GC	8.8	17.6	-6.34	34.97
5 CA/TG	12.4	20	4.88	40.33
6 AG/CT	17.2	16.2	10.07	25.88
7 GT/AC	19.4	13.6	16.65	18.89
8 TA/TA	19.4	16	42.20	19.48
9 AC/GT	16.8	18.9	7.00	23.44
10 CT/AG	12.2	18.6	-0.09	32.28
11 TA/TA	9.1	16.2	-0.55	43.22
12 TA/TA	9.1	16.2	-0.56	42.23
13 AG/CT	12.2	18.6	-0.09	32.28
14 GT/AC	16.8	18.9	7.00	23.43
15 TA/TA	19.4	16	42.20	19.48
16 AC/GT	19.4	13.6	16.65	18.89
17 CT/AG	17.2	16.2	10.07	25.88
18 TG/CA	12.4	20	4.88	40.32
19 GC/GC	8.8	17.6	-6.35	34.97
20 CT/AG			-3.12	31.24
21 TG/CA			1.20	40.52
22 GA/TC			5.09	31.12

Table 13. DNA glycosylase activities of the mutants

Mutants	k (min^{-1})*	Relative activity (%)
Y68F (control)	0.062 ± 0.003	100
Y68F K154A	0.0046 ± 0.002	7
Y68F T28W	0.023 ± 0.002	37
Y68F T28G	0.032 ± 0.001	52
Y68F P27G T28G	0.0088 ± 0.002	14

*The enzymatic rate constants and their standard errors were obtained from a single-exponential fit to the data from three independent measurements.
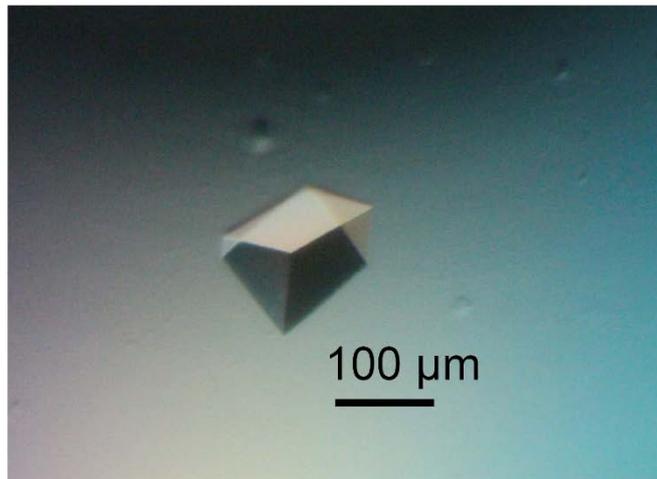
Figure 25. The crystal of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex.
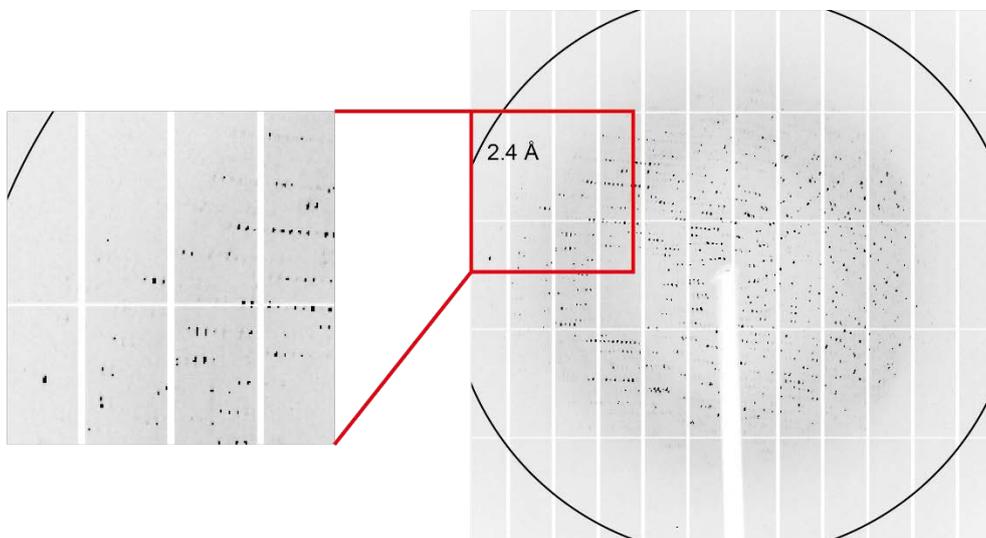
Figure 26. The X-ray diffraction image of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex crystal. The position corresponding to 2.4 Å is exhibited as a black circle.
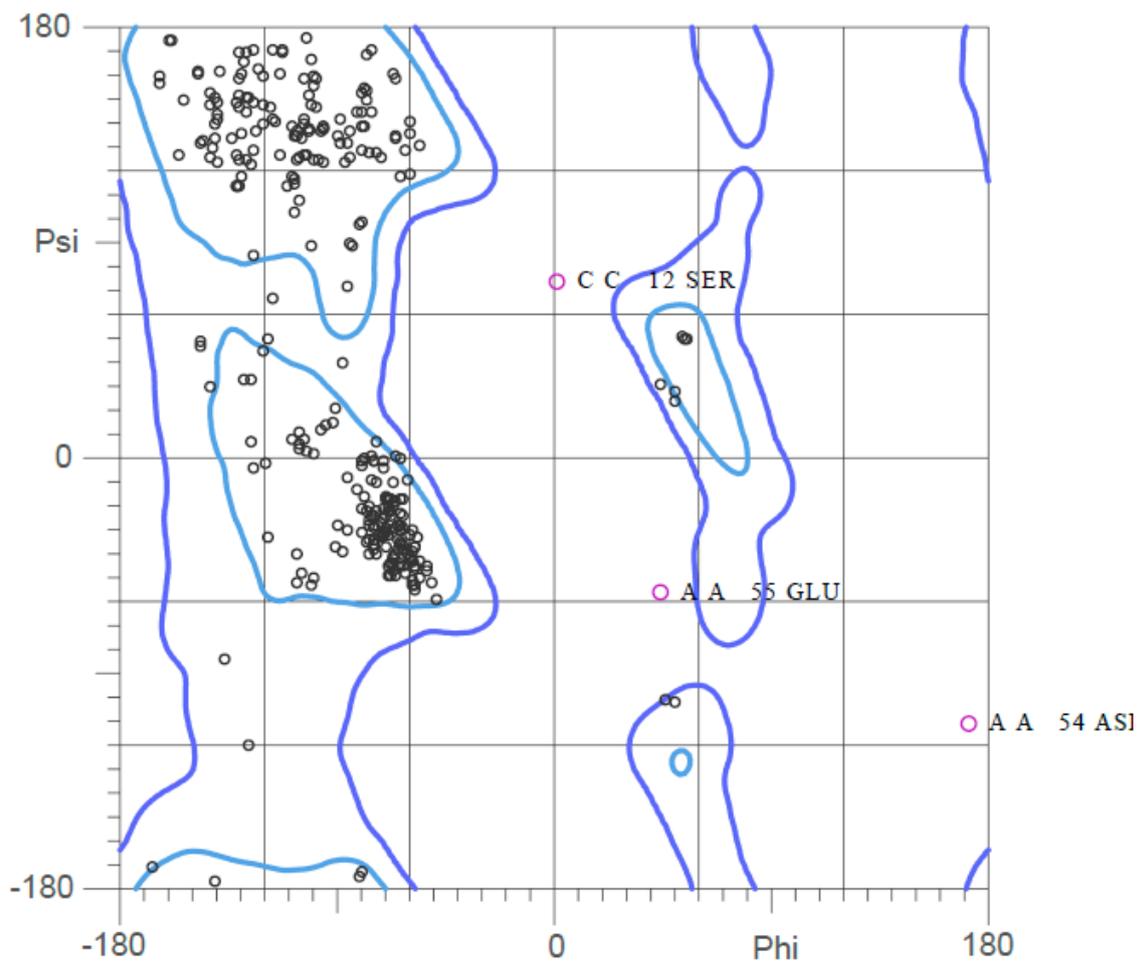
Figure 27. The Ramachandran plot of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex structure. 95.5% residues are in favored region; 99.3% residues are in allowed region.
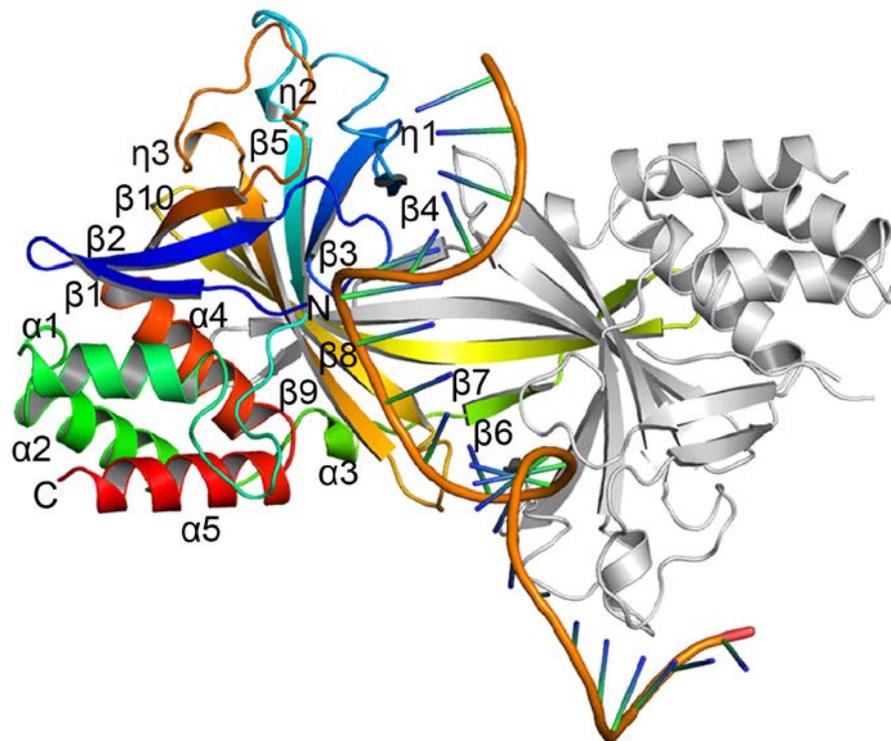
Figure 28. Structure of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex in one asymmetric unit. One R.PabI protomer is colored blue (in the N terminus) to red (in the C-terminus). The other R.PabI protomer is colored grey. The bound ssDNA is colored orange. Secondary structure assignments of protomer A are labelled on the model.
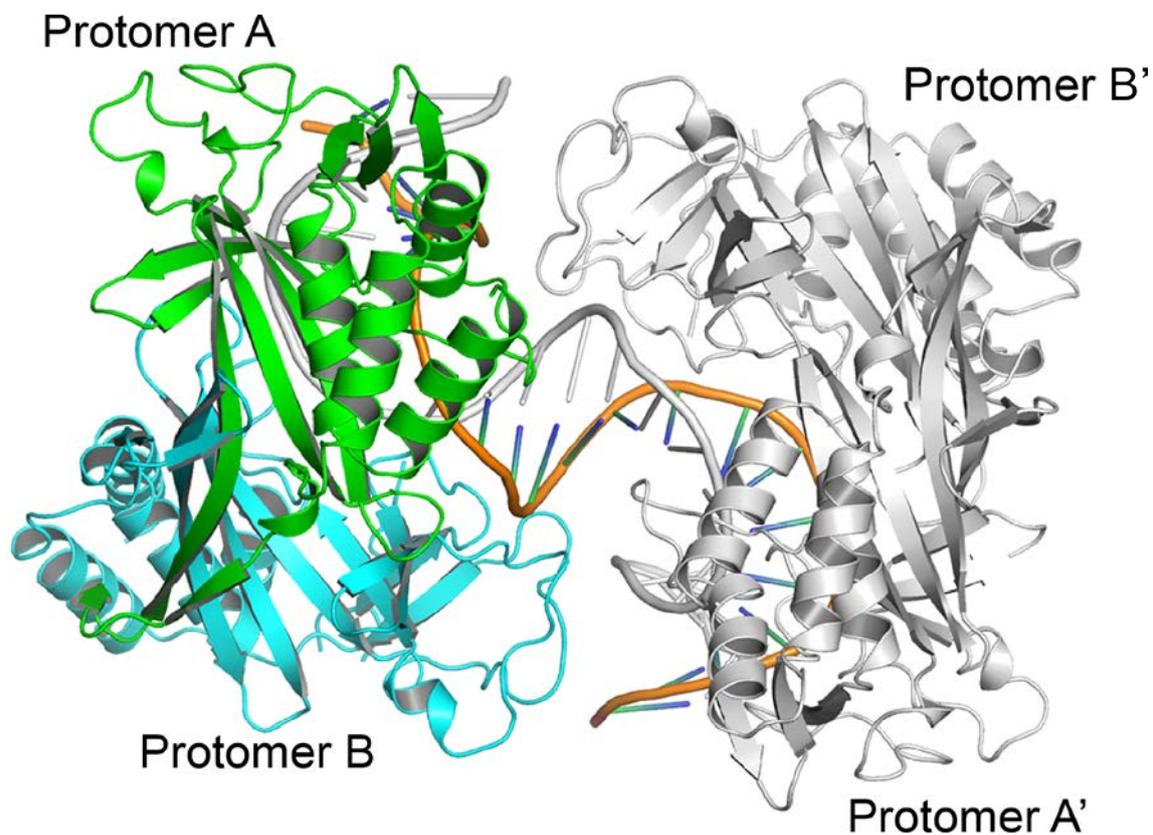
Figure 29. Overall structure of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex. Tetrameric structure of the R.PabI-nonspecific dsDNA complex. R.PabI protomers A and B in one asymmetric unit are colored green and cyan, respectively. R.PabI protomers generated by a symmetry operation (protomers A' and B') are colored grey. The dsDNA between the two R.PabI dimers is colored orange (chain C) and grey (chain C').
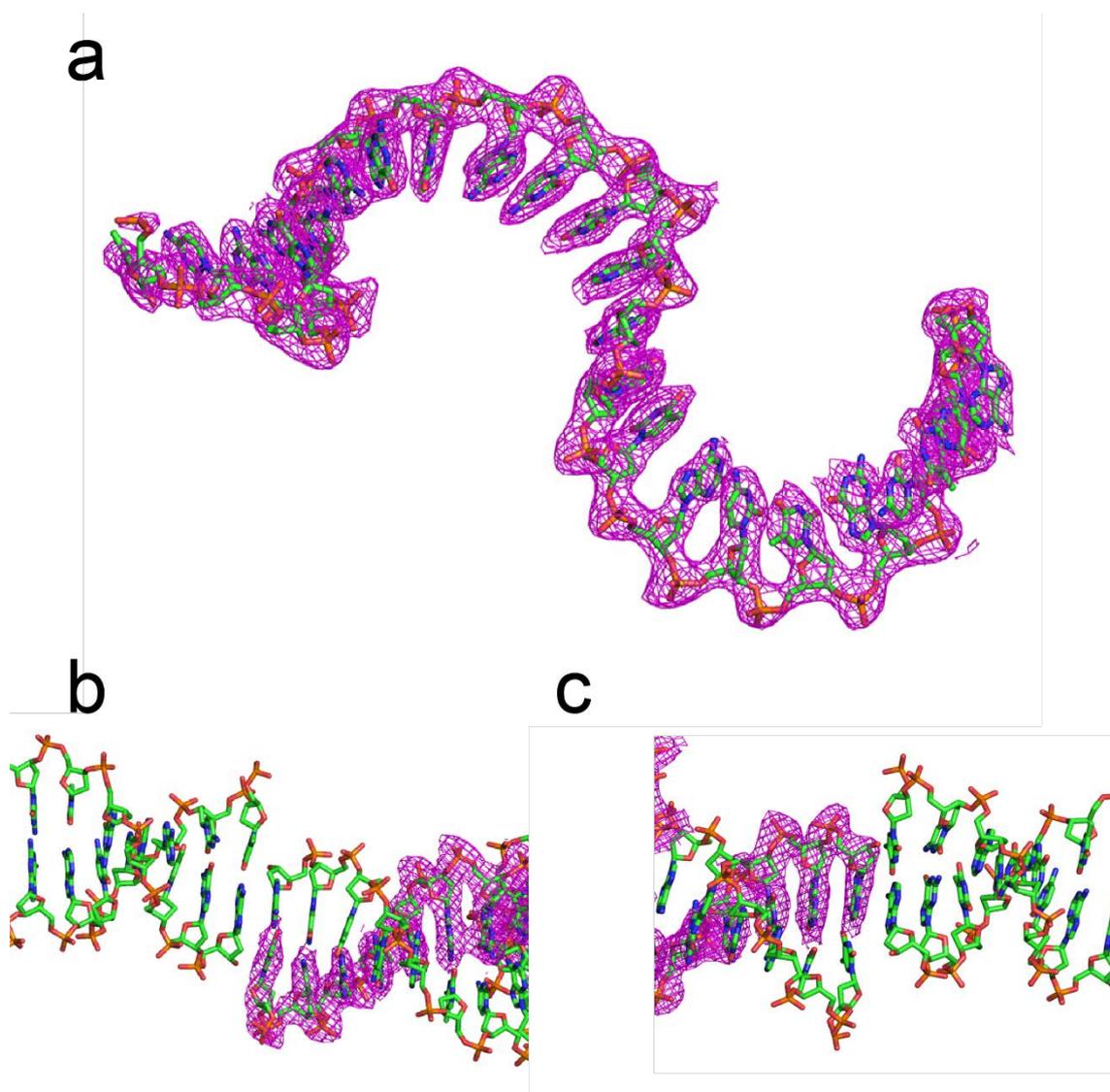
Figure 30. Composite omit map of the ssDNA and its interactions with DNA molecules in the other asymmetric units. (**a**) Composite omit map (1 σ, magenta) of the ssDNA in the asymmetric unit. (**b**, **c**) Interactions between the ssDNA in one asymmetric unit and the DNA molecules in the other asymmetric units.

Figure 31. Intermolecular hydrogen bonds between one R.PabI dimer and dsDNA. Hydrogen bonds between protomer A and the dsDNA are shown as green lines. Hydrogen bonds between protomer B and the dsDNA are shown as cyan lines. Residues of protomers A and B are colored green and cyan, respectively.
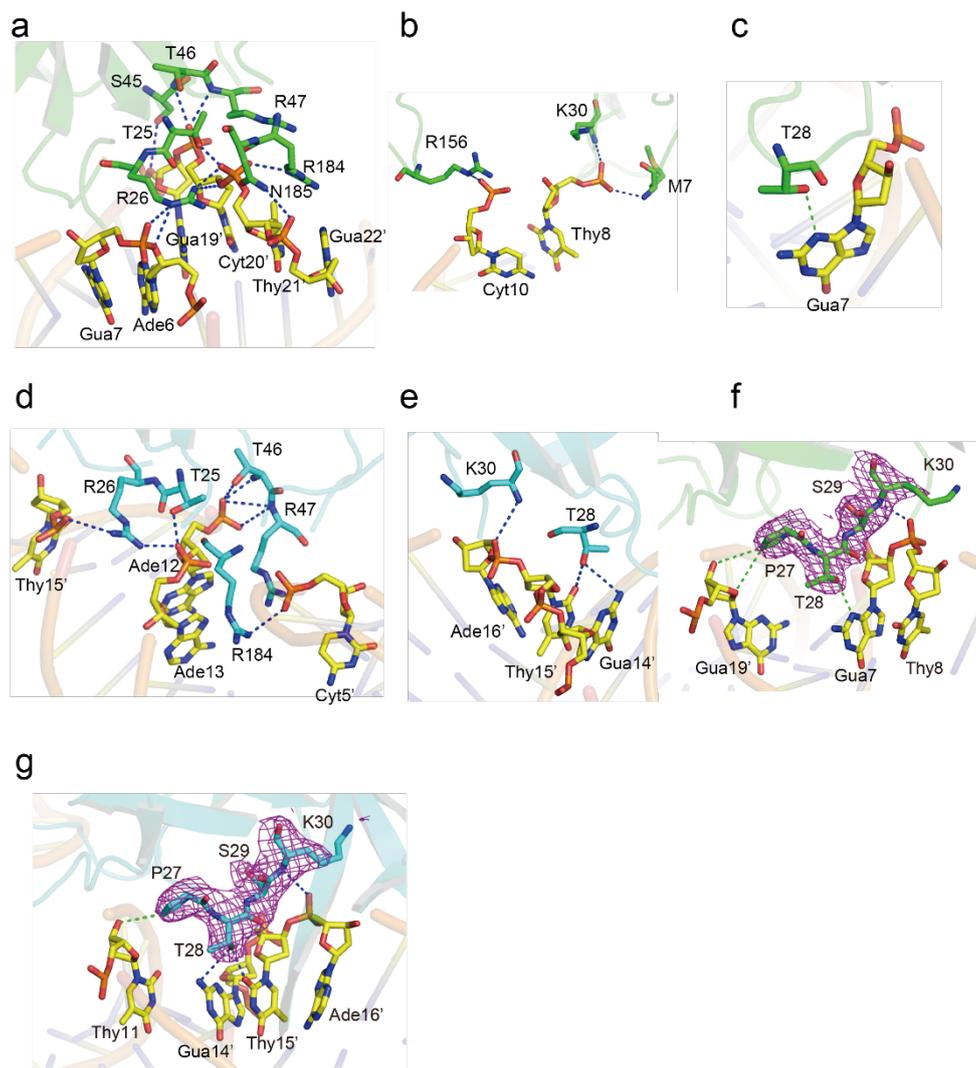
Figure 32. R.PabI-DNA interaction. (**a**, **b**) Intermolecular hydrogen bonds between R.PabI protomer A and dsDNA. Hydrogen bonds are shown as blue dotted lines. Residues of protomer A and DNA bases are colored green and yellow, respectively. (**c**) Van der waals interaction formed between T28 of protomer A and DNA base. Van der waals interaction is shown as green dotted line. (**d, e**) Intermolecular hydrogen bonds between R.PabI protomer B and dsDNA. Hydrogen bonds are shown as blue dotted lines. Residues of protomer B are colored cyan. (**f**) Composite omit map (1 σ, magenta) of the β2-β3 loop

in protomer A. The hydrogen bonds and van der waals interaction formed between the loop and the backbone of DNA are shown as blue and green dotted lines, respectively. (**g**) Composite omit map (1 σ, magenta) of the β2-β3 loop in protomer B. The hydrogen bonds and van der waals interaction formed between the loop and the backbone of DNA are shown as blue and green dotted lines, respectively.
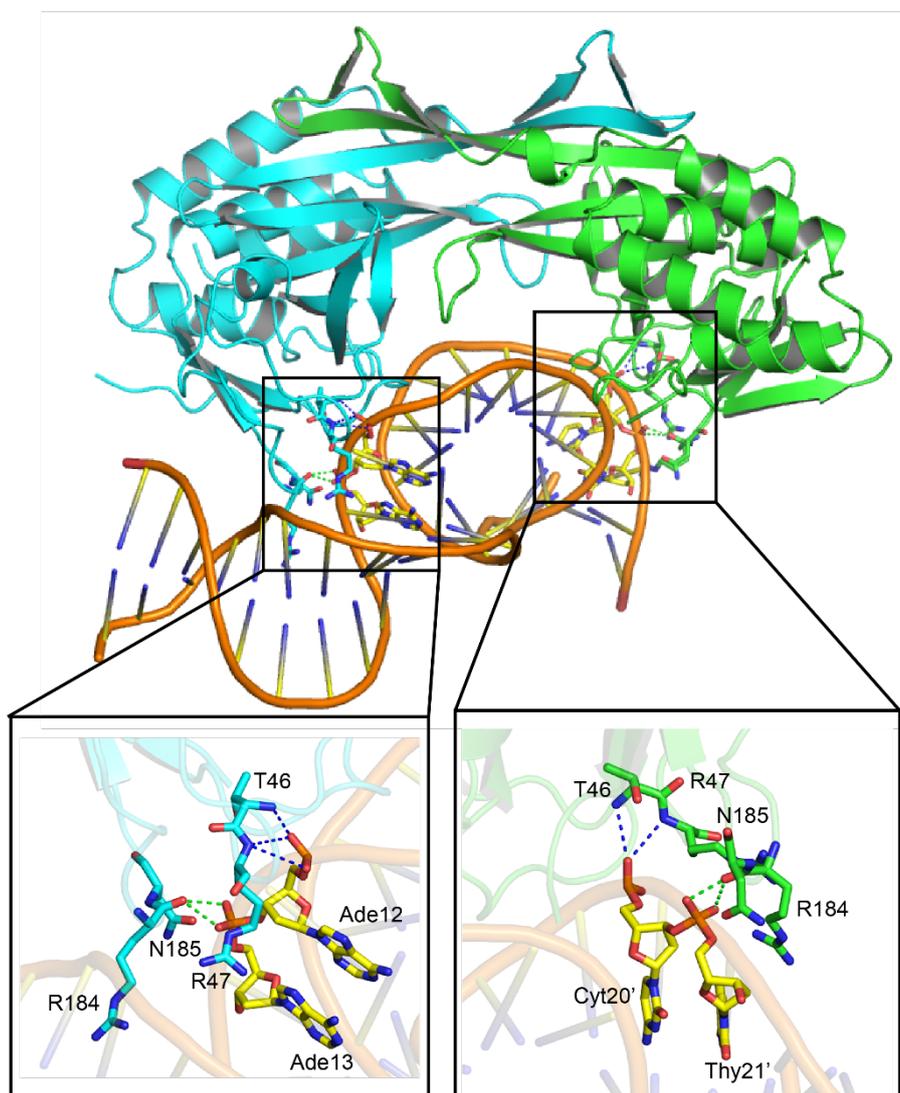
Figure 33. The intermolecular interactions that bend the dsDNA. The residues in β4-η2 loops and η3-β10 loops are shown in sticks. The van der waals interactions are shown in green dotted lines and the hydrogen bonds are shown in blue dotted lines. Protomers A and B and the residues in them are colored green and cyan, respectively. DNA bases are colored yellow, DNA backbone is colored orange.
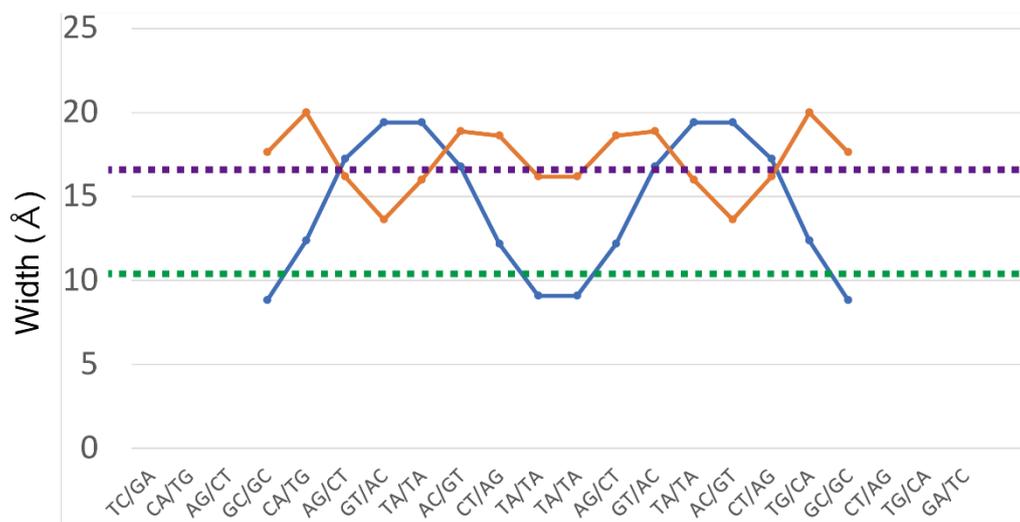
Figure 34. Width of the major groove and the minor groove in the dsDNA. Plots of major (orenge line) and minor (blue line) groove widths are shown as functions of DNA base step. The major and minor groove widths of ideal B-form dsDNA are shown as purple and green dotted lines, respectively.
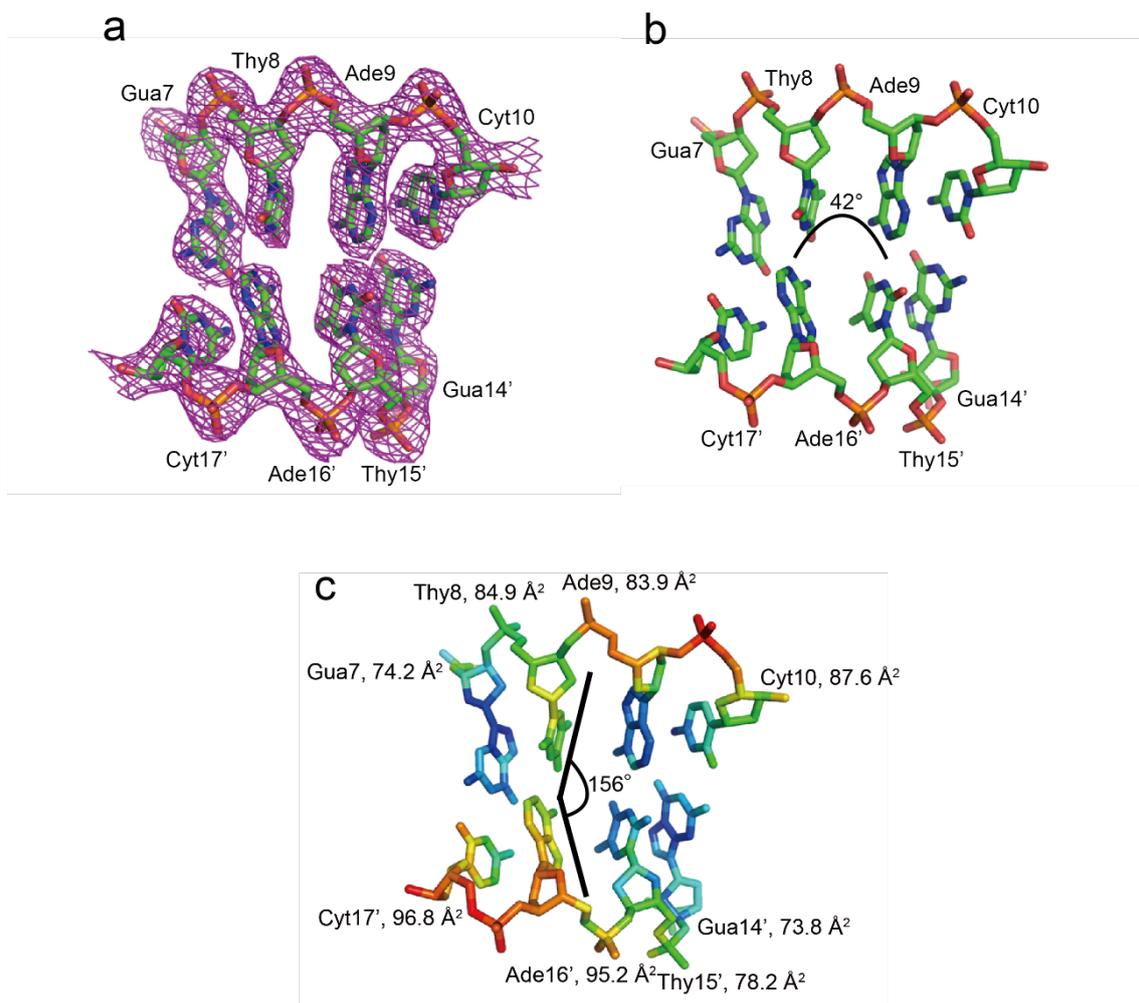
Figure 35. Conformation of the recognition sequence (5'-GTAC-3') in the dsDNA. (**a**) Composite omit map (1 σ, magenta) of the recognition sequence. (**b**) The roll angle between the A-T step is 42°. (**c**) *B*-factor of each base is calculated by CCP4. Bases are colored by blue (low *B*-factor) to red (high *B*-factor). The buckle angle between Thy8 and Ade16' is 24° (the number showed in the figure is its complementary degree).
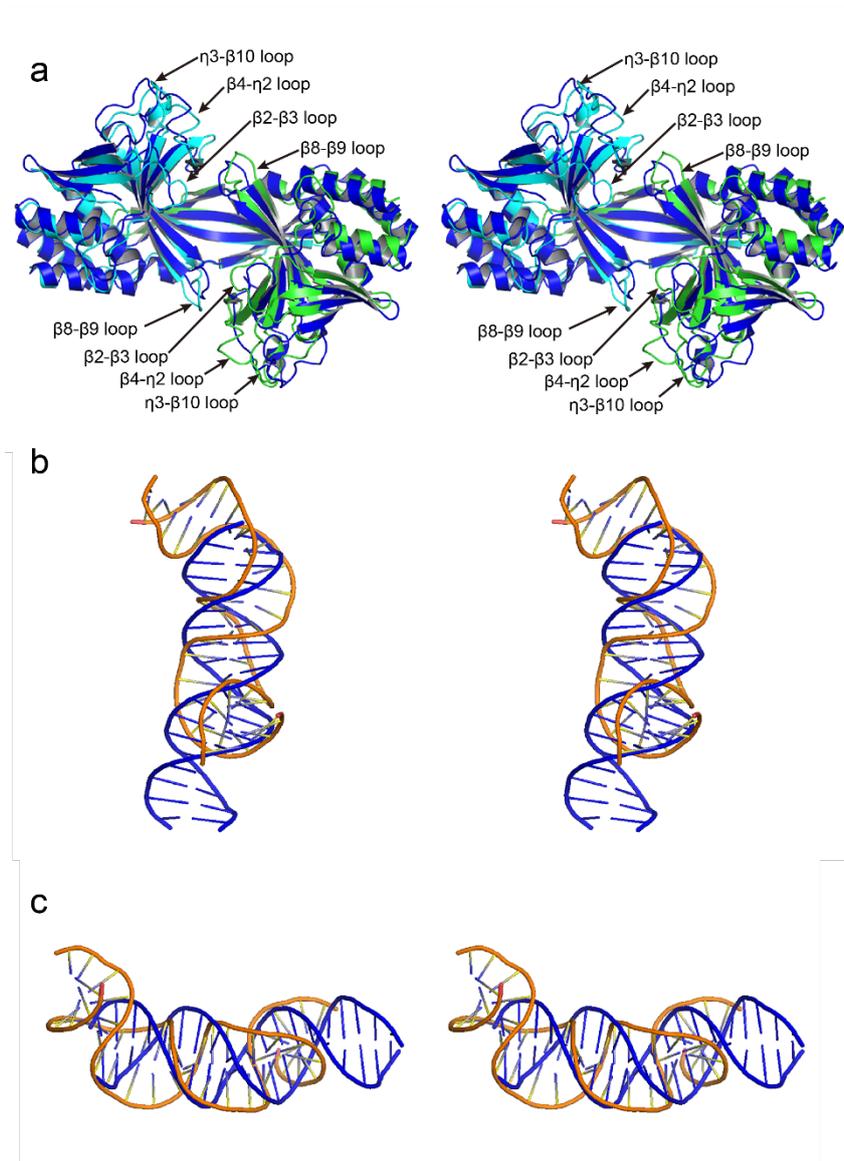
Figure 36. Structure comparison of R.PabI in the nonspecific dsDNA binding state and in the encounter state. (**a**) Wall-eyed stereo image of the superposition of the R.PabI dimers in the nonspecific DNA binding state (blue) and the encounter state (green and cyan). (**b**, **c**) Wall-eyed stereo image of the superposition of the dsDNA in the nonspecific DNA binding state (blue) and the encounter state (orange). (**b**) is seen from above the complex and (**c**) is seen from right side of the complex.
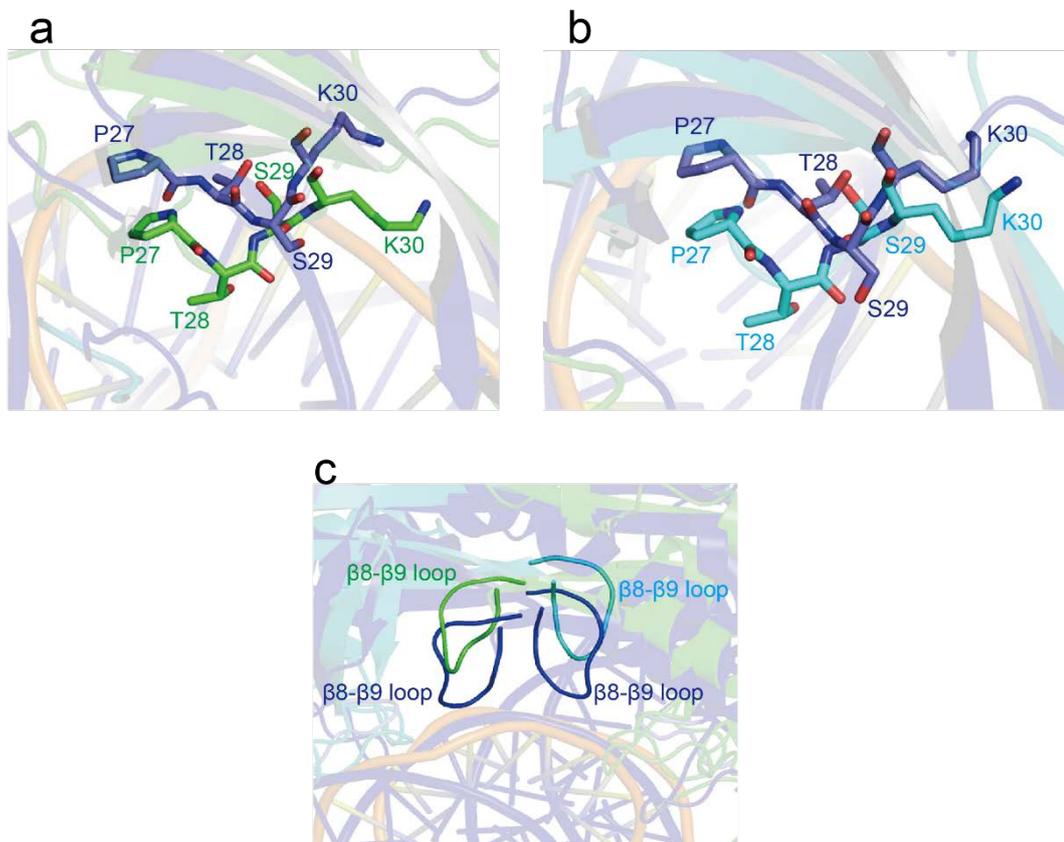
Figure 37. Comparison of the β2-β3 loops and the β8-β9 loops of R.PabI in the nonspecific dsDNA binding state and in the encounter state. Loops and residues in the nonspecific R.PabI-DNA complex are colored blue and those in the encounter complex are colored green and cyan, respectively. (**a**, **b**) Comparison of positions of each residue in the β2-β3 loops in the nonspecific R.PabI-DNA complex and the encounter complex. (**c**) Comparison of positions of β8-β9 loops in the two complexes.
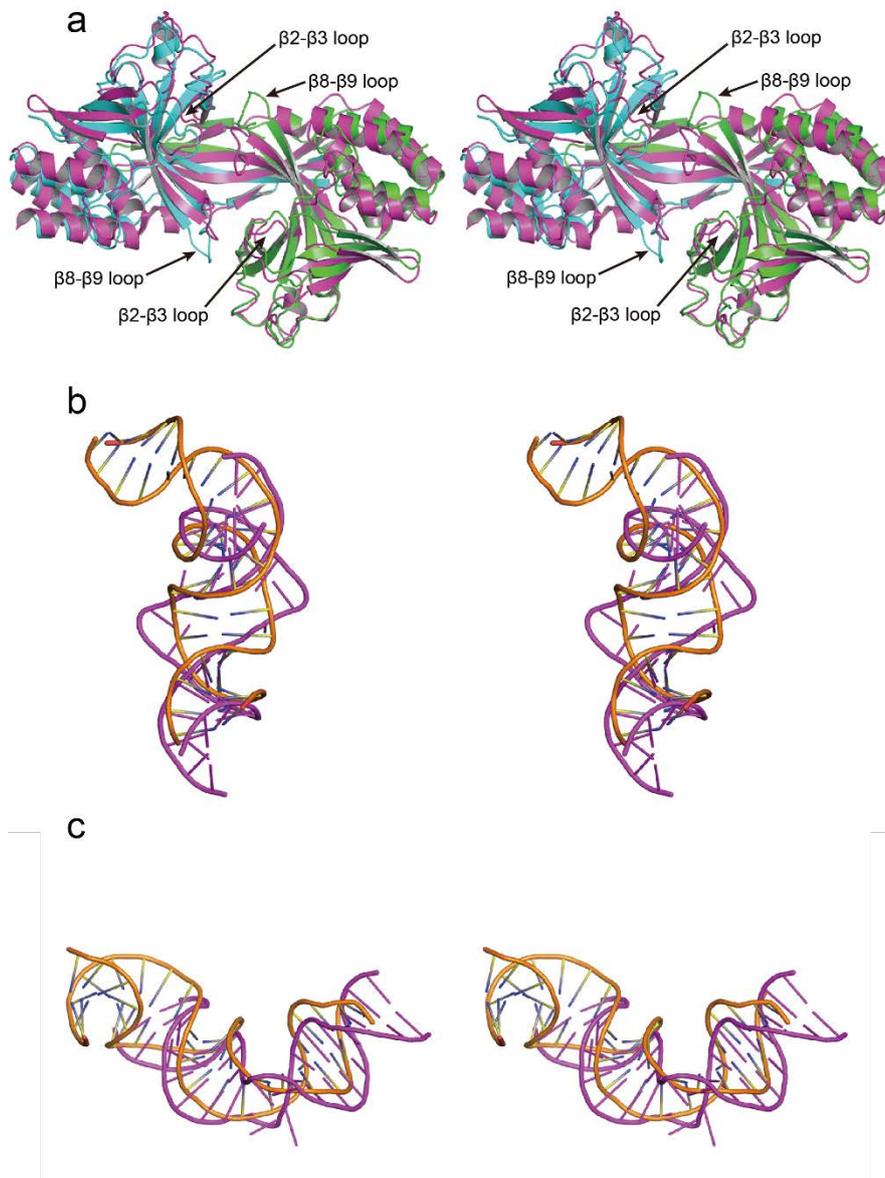
Figure 38. Structure comparison of R.PabI in the encounter state and in the specific dsDNA binding state. (**a**) Wall-eyed stereo image of the superposition of the R.PabI dimers in the encounter state (green and cyan) and the specific DNA binding state (magenta). (**b, c**) Wall-eyed stereo image of the superposition of the dsDNA in the the encounter state (orange) and the specific DNA binding state (magenta). (**b**) is seen from above the complex and (**c**) is seen from right side of the complex.
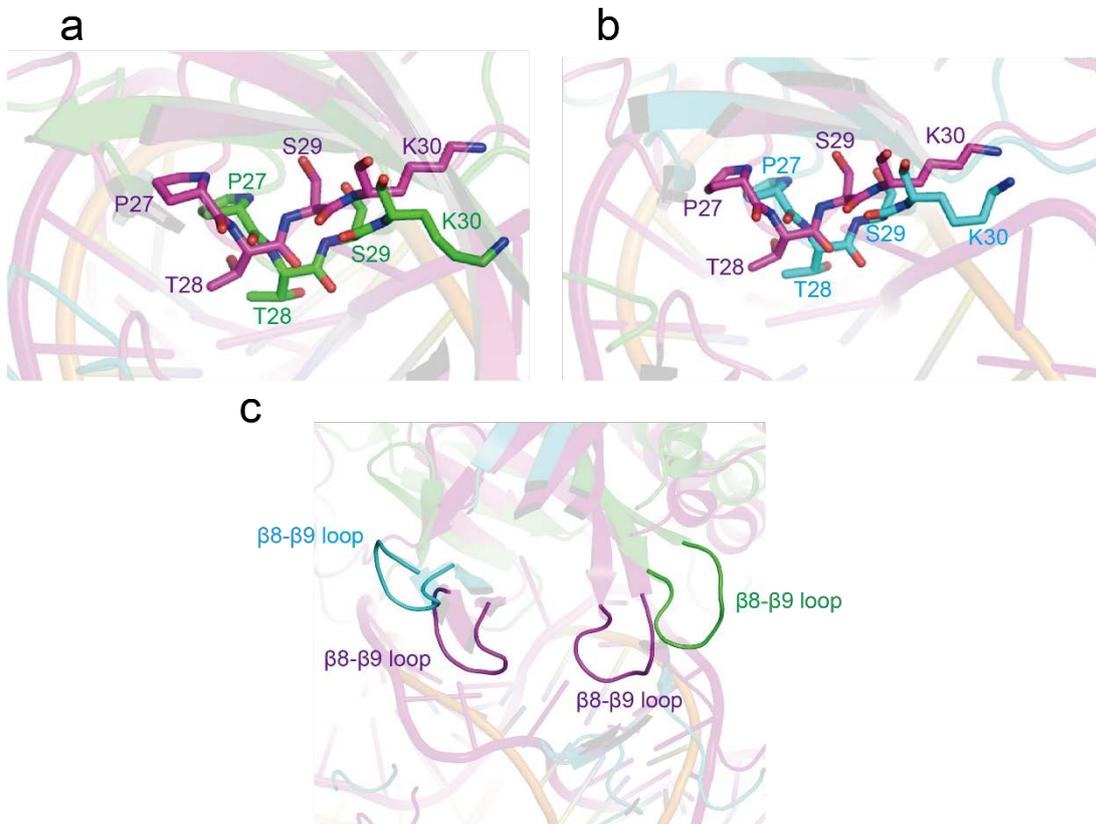
Figure 39. Comparison of β2-β3 loops and β8-β9 loops of R.PabI in the encounter state and in the specific dsDNA binding state. Loops and residues in the encounter complex are coloured green and cyan, respectively, those in the specific R.PabI-DNA complex are coloured magenta. (**a**, **b**) Comparison of positions of each residue of β2-β3 loops in the encounter complex and the specific R.PabI-DNA complex. (**c**) Comparison of positions of β8-β9 loops in the two complexes.
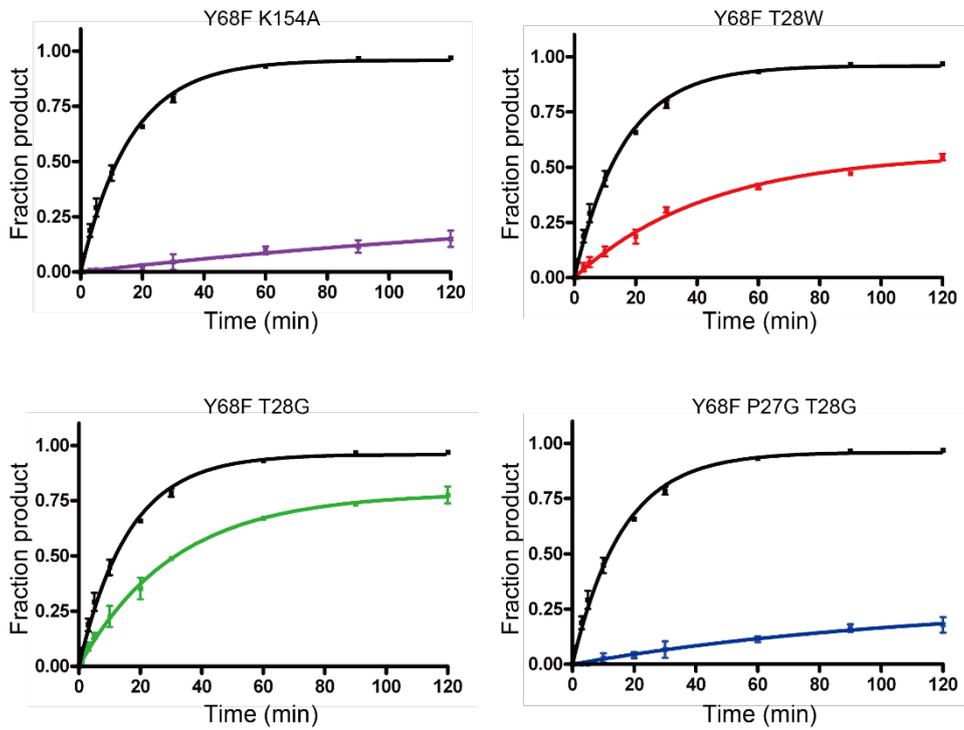
Figure 40. DNA glycosylase assays of the Y68F (black), Y68F K154A (purple), Y68F T28W (red), Y68F T28G (green) and Y68F P27G T28G (blue) mutants using 24 bp dsDNA as substrates. Plotted values are mean ± SD (n=3).
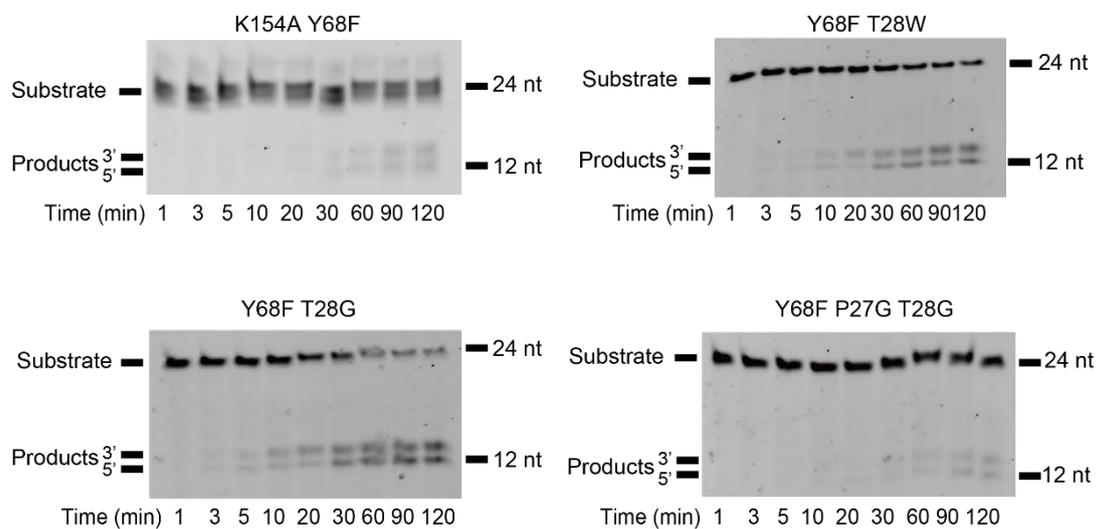
Figure 41. DNA glycosylase activity assay of R.PabI mutants (K154A Y68F, Y68F T28W, Y68F T28G and Y68F P27G T28G) using 24 bp dsDNA as substrate. First, 0.2 μM of the 24 bp substrate dsDNA and 0.4 μM of the R.PabI dimer were mixed and incubated at 45°C. Each experiment was repeated three times.

4. Discussion

4-1 The binding state of the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex

In this research, the complex structure of R.PabI-ΔN-K154A Y68F binding with the 23bp

dsDNA which contains two recognition sequences has been determined at 2.4 Å. In the

complex structure, two R.PabI homodimers bind to each recognition sequences separately,

due to these interactions, the DNA is sharply bent by around 90° at each recognition

sequence. This complex structure, especially the conformation of the dsDNA, does not

show considerable resemblance to the previously determined R.PabI-nonspecific DNA

complex and the R.PabI-specific DNA complex. Prior researches have reported that after

the glycosylase MutM finds its recognition complex and before it unwinds the DNA

double helix, it transiently forms an encounter complex in which the DNA double helix

is sharply bent yet all of the bases remain to be intrahelix[58,60,64]. Such a binding state has

also been observed among other glycosylases, such as XPC[65] and AlkD[57]. The

conformation of the DNA in the R.PabI-ΔN-K154A Y68F-23bp dsDNA complex highly

resembles the DNAs in those encounter complexes: it is sharply bent over yet all of its

bases are not flipped out. Therefore, the R.PabI-ΔN-K154A Y68F-dsDNA complex

structure most likely shows a binding state where R.PabI encounters its recognition site

after searching along the DNA and starts to interrogate its sequence in order to transform

into the recognition binding state.

4-2 The function of the β2-β3 loop and T28 in the encounter state

Through the comparison of the encounter complex structure with both the R.PabI-

nonspecific DNA complex and the R.PabI-specific DNA complex, the movement of the

β2-β3 loops have been identified. The location of the β2-β3 loops change from above the

DNA helix in the nonspecific complex, to being deeply inserted into the minor groove in

the encounter complex and the specific complex. What is worthy of note is that in the encounter complex, in which the β8-β9 loops are yet protruded into the minor groove, the β2-β3 loops are already placed within the minor groove and maintain the expansion of the minor groove by pushing the ribose phosphate backbone toward the major groove. This indicates that after R.PabI encounters its recognition sequence, the β2-β3 loops move firstly to protrude into the minor groove, then the β8-β9 loops would insert into the expanded minor groove to unwind the double helix and flip the guanine and adenine. When two residues of the loop which protrude into the minor groove (P27 and T28) were mutated into glycine, the activity of the mutant was decreased by 14%. This result further proved that the β2-β3 loops have important role in cleaving the DNA.

In the encounter complex, T28, which exists in the β2-β3 loops, is the only residue which is observed of having formed hydrogen bonds or van der waals interactions with the guanines in the recognition sequence. The structure comparison of the encounter complex and the R.PabI-nonspecific DNA complex indicated that T28 reorients its side chain direction from above the DNA helix in the nonspecific complex, to being deeply wedged into the minor groove to interact with the guanines in the encounter complex. This result suggests that during sliding on the DNA, T28 is not utilized to interrogate its recognition sequence, it is not until R.PabI stalls at a T-A step and sharply bends DNA does the T28 transfer its side chain direction to interact with the guanine. Thus, it is reasonable to assume that T28 would be utilized to recognize the guanine in the encounter stage. This assumption is further supported by the activity assay which showed that the activity of the glycine and tryptophan mutant of T28 (T28G and T28W) was decreased by 52% and 37% respectively. In addition, previous studies have shown that during the

126

encounter stage, glycosylases usually utilize a wedge residue to insert into the DNA groove and interrogate the recognition site[66–69]. T28 might as well play such a role in the encounter complex of R.PabI. Although it is widely accepted that the final stage of specific recognition of glycosylase is carried out by the extrusion of the base, the recognition at the early encounter stage would help increasing the recognition efficiency. Thus, it is assumed that T28 facilitates the early stage recognition by forming interactions with the guanine in the recognition sequence.

4-3 R.PabI uses sequence dependent flexibility of DNA to detect its recognition site

The encounter complex structure indicates that previous to preceding with the DNA unwinding step observed in the R.PabI-specific DNA complex, the DNA is already sharply bent and considerably untwisted. A prior research had indicated that through the contact with the glycosylase, the DNA is bent over and the helix is untwisted, the untwisting level is directly proportional to the increasing of the DNA bending. At such a condition, even without the protrusion of the loop from the glycosylase, the untwisting of the helix alone could facilitate the flipping of the aberrant base[64]. Based on this theory, the bending and untwisting of the DNA in the encounter complex of R.PabI is assumed to have important role in promoting the flipping out of the base. The DNA base pair calculation has shown that among the recognition sequence, the T-A step has the largest Roll degree, suggesting that the DNA in the encounter complex is mostly bent over at the T-A step. A prior research has calculated the stacking energy of every dinucleotide step in DNA and found out that the stacking interaction between the T-A step is the weakest among all of the dinucleotide steps[70]. This result supports such a hypothesis that the existence of the T-A step in the recognition sequence would largely facilitate the bending

of the DNA thus to make the flipping of the base easier. This hypothesis is similar to studies of the other glycosylases. A glycosylase, such as MutM, recognizes its target DNA site by detecting the difference of flexibility between the damaged base and a normal base. The damaged base would destabilize the stacking interaction with its 5' or 3'-neighboring bases thus reduce the penalty of the flipping of the aberrant base, hence the damaged site would be easier bent over than a normal sequence[60,71–73]. Since R.PabI recognizes a specific dsDNA sequence, thus the flexibility of the T-A step, instead of the aberrant base, might be recognized in such a mechanism.

A prior research has indicated that Rad4, a nucleotide excision repair enzyme, uses a stepwise mechanism to find its recognition site. During searching on DNA, it firstly stalls at sites which are as deformable as its recognition site, then it deforms the DNA site to allow more protein-DNA interactions to be formed in order to distinguish whether it is its recognition site or not[65]. R.PabI may as well be utilizing this process to search for its recognition sequence, because its recognition sequence also contains T-A step which is comparatively easily recognizable due to its flexibility. Such a mechanism would increase the searching efficiency by two or three orders of magnitude, comparing with interrogating each base in the DNA[73].

General Discussion

In this research, I have solved two crystal structures of restriction DNA glycosylase R.PabI in complex with DNA, the mechanism by which R.PabI searches on DNA and interrogates its recognition sequence have been speculated based on the crystal structures as well as the biochemical experiments related to them. In chapter one, the inactive R.PabI mutant binding with nonspecific DNA complex structure has been solved at 1.9 Å. The structure indicates that R.PabI forms a ring-like tetrameric structure to bind to nonspecific DNA and slide on it to increase the searching efficiency for the recognition sequence. In chapter two, the crystal structure of a low-active R.PabI mutant in complex with the specific DNA has been solved at 2.4 Å. The complex crystal structure indicates that after encountering its recognition sequence, R.PabI bends DNA to untwist the recognition sequence and open the minor groove, the β2-β3 loops are protruded into the minor groove to maintain the expansion. The flexible T-A step in the recognition sequence facilitates the bending of the DNA thus makes the flipping of the bases easier.

Together with the previous researches about the DNA-free structure[10] and the R.PabI-specific DNA complex structure[13], the whole procedure by which R.PabI searches and recognizes its target sequence has been unveiled. R.PabI widens its halfpipe region to bind to nonspecific DNA, then it uses the ring-like tetrameric form to facilitate the sliding on DNA to search for its recognition sequence. After it encounters its recognition site or the sites as deformable as the recognition sequence, it bends DNA over and expands the minor groove, the β2-β3 loops are inserted into the opened minor groove to stabilize it. Finally, β8-β9 loops are protruded into the expanded minor groove to flip the guanines and adenines out of the double helix, the *N*-glycosidic bonds of the extruded adenines are excised by hydrolysis.

No research before had ever studied the searching, interrogation and cleavage

mechanism of the HALFPIPE superfamily of typeII restriction, this research, along with its previous researches, have unveiled a complete landscape of R.PabI interacting with DNA. These results would provide an important reference for us to understand the other member of the HALFPIPE superfamily. In addition, since R.PabI recognizes its substrate similar to restriction enzyme while cleaves DNA in a manner closed to glycosylase, elucidation of the recognition mechanism of such an unique enzyme would enlarge our understanding about the protein-DNA interaction.

References.

1.   Pingoud, A., Fuxreiter, M., Pingoud, V. & Wende, W. Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life Sci.* **62,** 685–707 (2005).

2.   Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. & Aggarwal, K. Structure of BamHI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science* **269,** 656–663 (1995).

3.   Kim, Y.C., Grable, J.C., Love, R., Greene, P.J. & Rosenberg, J. Refinement of EcoRI endonuclease crystal structure: a revised protein chain tracing. *Science* **249,** 1307–1309 (1990).

4.   Winkler, F. K. *et al.* The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *The EMBO Journal* **12,** 1781–1795 (1993).

5.   Shen, B. *et al.* Unusual target site disruption by the rare-cutting HNH restriction endonuclease PacI. *Structure* **18,** 734–743 (2010).

6.   Saravanan, M., Bujnicki, J., Cymerman, I., Rao, D. & Nagaraja, V. Type II restriction endonuclease R.KpnI is a member of the HNH nuclease superfamily. *Nucleic Acids Res.* **32,** 6129–6135 (2004).

7.   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).

8.   Ibryashkina, E. *et al.* Type II restriction endonuclease R.Eco29kI is a member of the GIY-YIG nuclease superfamily. *BMC Struct. Biol.* **7,** 48 (2007).

9.   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).

10.   Miyazono, K. *et al.* Novel protein fold discovered in the PabI family of restriction enzymes. *Nucleic Acids Res.* **35,** 1908–1918 (2007).

11.   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).

12.   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).

13.   Miyazono, K. *et al.* A sequence-specific DNA glycosylase mediates restriction-modification in Pyrococcus abyssi. *Nat. Commun.* **5,** 3178 (2014).

14.   Zhou, H.-X. Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 8651–8656 (2011).

15.   Díaz de la Rosa, M., Koslover, E. F., Mulligan, P. J. & Spakowitz, A. J. Dynamic strategies for target-site search by DNA-binding proteins. *Biophys. J.* **98,** 2943–2953 (2010).

16.   Tabaka, M., Burdzy, K. & Hołyst, R. Method for the analysis of contribution of sliding and hopping to a facilitated diffusion of DNA-binding protein: Application to in vivo data. *Phys. Rev. E* **92,** 1–7 (2015).

17.   Viadiu, H. & Aggarwal, A. K. Structure of BamHI bound to nonspecific DNA: A model for DNA sliding. *Mol. Cell* **5,** 889–895 (2000).

18. Gorman, J. & Greene, E. C. Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.* **15,** 768–774 (2008).

19. Blainey, P. C. *et al.* Nonspecifically bound proteins spin while diffusing along DNA. *Nat. Struct. Mol. Biol.* **16,** 1224–1229 (2009).

20. Friedman, L. J., Mumm, J. P. & Gelles, J. RNA polymerase approaches its promoter without long-range sliding along DNA. *Proc. Natl. Acad. Sci.* **110,** 9740–9745 (2013).

21. Redding, S. & Greene, E. How do proteins locate specific targets in DNA? *Chem. Phys. Lett. J.* **570,** 1–11 (2013).

22. Halford, S. E. & Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32,** 3040–3052 (2004).

23. Bauer, M. & Metzler, R. Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys. J.* **102,** 2321–2330 (2012).

24. Halford, S. E. An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.* **37,** 343–348 (2009).

25. Shimamoto, N. One-dimensional diffusion of proteins along DNA. *J. Biol. Chem.* **274,** 15293–15296 (1999).

26. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. & Xie, S. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 5752–5757 (2006).

27. Misteli, T. Protein dynamics: Implications for nuclear architecture and gene expression. *Science* **291,** 843–847 (2001).

28. Gowers, D. M. & Halford, S. E. Protein motion from non-specific to specific

DNA by three-dimensional routes aided by supercoiling. *EMBO J.* **22,** 1410–1418 (2003).

29.   Bonnet, I. *et al.* Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Res.* **36,** 4118–4127 (2008).

30.   Vologodskii, A. & Cozzarelli, N. R. Effect of supercoiling on the juxtaposition and relative orientation of DNA sites. *Biophys. J.* **70,** 2548–2556 (1996).

31.   Schmidt, H. G., Sewitz, S., Andrews, S. S. & Lipkow, K. An integrated model of transcription factor diffusion shows the importance of intersegmental transfer and quaternary protein structure for target site finding. *PLoS One* **9,** e108575 (2014).

32.   Dikic, J. *et al.* The rotation-coupled sliding of EcoRV. *Nucleic Acids Res.* **40,** 4064–4070 (2012).

33.   Kabsch, W. Xds. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 125–132 (2010).

34.   Winn, M. D., Ballard, C. C. & Cowtan, K. D. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 235–242 (2011).

35.   Vagin, A. & Teplyakov, A. MOLREP : an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* **30,** 1022–1025 (1997).

36.   Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 213–221 (2010).

37.   Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60,** 2126–2132 (2004).

38.   Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797 (2007).

39. Lu, X. J. & Olson, W. K. 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31,** 5108–5121 (2003).

40. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 12–21 (2010).

41. Rohs, R. *et al.* The role of DNA shape in protein-DNA recognition. *Nature* **461,** 1248–1253 (2009).

42. Zhang, J., McCabe, K. A. & Bell, C. E. Crystal structures of lambda exonuclease in complex with DNA suggest an electrostatic ratchet mechanism for processivity. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 11872–11877 (2011).

43. Miyazono, K. *et al.* Cooperative DNA-binding and sequence-recognition mechanism of aristaless and clawless. *EMBO J.* **29,** 1–11 (2010).

44. Wong, J. J. W., Lu, J., Edwards, R. A., Frost, L. S. & Glover, J. N. M. Structural basis of cooperative DNA recognition by the plasmid conjugation factor, TraM. *Nucleic Acids Res.* **39,** 6775–6788 (2011).

45. Ehbrecht, H. J., Pingoud, A., Urbanke, C., Maass, G. & Gualerzi, C. Linear diffusion of restriction endonucleases on DNA. *J. Biol. Chem.* **260,** 6160–6166 (1985).

46. Deibert, M., Grazulis, S., Sasnauskas, G., Siksnys, V. & Huber, R. Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.* **7,** 792–799 (2000).

47. Ogasahara, K., Ishida, M. & Yutani, K. Stimulated Interaction between alpha and beta subunits of tryptophan synthase from hyperthermophile enhances its thermal stability. *J. Biol. Chem.* **278,** 8922–8928 (2003).

48.   Thoma, R., Hennig, M., Sterner, R. & Kirschner, K. Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from Thermotoga maritima. *Structure* **8,** 265–276 (2000).

49.   Tanaka, Y. *et al.* How oligomerization contributes to the thermostability of an Archaeon protein. *J. Biol. Chem.* **279,** 32957–32967 (2004).

50.   Brooks, S. C., Adhikary, S., Rubinson, E. H. & Eichman, B. F. Recent advances in the structural mechanisms of DNA glycosylases. *Biochim. Biophys. Acta - Proteins Proteomics* **1834,** 247–271 (2013).

51.   Vassylyev, D. G. *et al.* Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: Structural basis for damaged DNA recognition. *Cell* **83,** 773–782 (1995).

52.   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.* **97,** 5083–5088 (2000).

53.   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).

54.   Fromme, J. C. & Verdine, G. L. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nat. Struct. Biol.* **9,** 544–52 (2002).

55.   Lau, Y., Wyatt, M. D., Glassner, B. J., Samson, L. D. & Ellenberger, T. Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 13573–13578 (2000).

56. Rubinson, E. H., Gowda, S. P., Spratt, T. E., Gold, B. & Eichman, B. F. An unprecedented nucleic acid capture mechanism for excision of DNA damage. *Nature* **468,** 406–411 (2010).

57. Mullins, E. A. *et al.* The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. *Nature* **527,** 254–258 (2015).

58. Banerjee, A., Santos, L.W. & Verdine, G. L. Structure of a DNA Glycosylase Searching for Lesions. *Science* **311,** 1153–1157 (2006).

59. 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).

60. Qi, Y. *et al.* Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature* **462,** 762–766 (2009).

61. Sung, R. J., Zhang, M., Qi, Y. & Verdine, G. L. Sequence-dependent structural variation in DNA undergoing intrahelical inspection by the DNA glycosylase MutM. *J. Biol. Chem.* **287,** 18044–18054 (2012).

62. Sung, R. J., Zhang, M., Qi, Y. & Verdine, G. L. Structural and biochemical analysis of DNA helix invasion by the bacterial 8-oxoguanine DNA glycosylase MutM. *J. Biol. Chem.* **288,** 10012–10023 (2013).

63. 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).

64. La Rosa, G. & Zacharias, M. Global deformation facilitates flipping of damaged 8-oxo-guanine and guanine in DNA. *Nucleic Acids Res.* **44,** 9591–9599 (2016).

65. Velmurugu, Y., Chen, X., Slogoff Sevilla, P., Min, J.-H. & Ansari, A. Twist-

open mechanism of DNA damage recognition by the Rad4/XPC nucleotide excision repair complex. *Proc. Natl. Acad. Sci.* **113,** 2296–2305 (2016).

66. Lee, A. J. & Wallace, S. S. Hide and seek: How do DNA glycosylases locate oxidatively damaged DNA bases amidst a sea of undamaged bases? *Free Radic. Biol. Med.* **107,** 170–178 (2017).

67. Nelson, S. R., Dunn, A. R., Kathe, S. D., Warshaw, D. M. & Wallace, S. S. Two glycosylase families diffusively scan DNA using a wedge residue to probe for and identify oxidatively damaged bases. *Proc. Natl. Acad. Sci.* **111,** 2091–2099 (2014).

68. Kuznetsov, N. A. *et al.* Active destabilization of base pairs by a DNA glycosylase wedge initiates damage recognition. *Nucleic Acids Res.* **43,** 272–281 (2015).

69. Dunn, A. R., Kad, N. M., Nelson, S. R., Warshaw, D. M. & Wallace, S. S. Single Qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA. *Nucleic Acids Res.* **39,** 7487–7498 (2011).

70. Protozanova, E., Yakovchuk, P. & Frank-Kamenetskii, M. D. Stacked-unstacked equilibrium at the nick site of DNA. *Journal of Molecular Biology* **342,** 775–785 (2004).

71. Chiba, J. *et al.* Deformable nature of various damaged DNA duplexes estimated by an electrochemical analysis on electrodes. *Chem. Commun.* **50,** 11126 (2014).

72. Buechner, C. N., Maiti, A., Drohat, A. C. & Tessmer, I. Lesion search and recognition by thymine DNA glycosylase revealed by single molecule imaging. *Nucleic Acids Res.* **43,** 2716–2729 (2015).

73. Yang, W. Poor base stacking at DNA lesions may initiate recognition by many

repair proteins. *DNA Repair (Amst).* **5,** 654–666 (2006).