Crystal structure of family 4 uracil–DNA glycosylase from *Sulfolobus tokodaii* and a function of tyrosine 170 in DNA binding

Akito Kawai a,⇑, Shigesada Higuchi a, Masaru Tsunoda c, Kazuo T. Nakamura b, Yuriko Yamagata d, Shuichi Miyamoto a

a Faculty of Pharmaceutical Sciences, Sojo University, 4-22-1 Ikeda, Nishi-ku, Kumamoto 860-0082, Japan
b School of Pharmacy, Showa University, 1-5-8 Hatanodai, Shinagawa-ku, Tokyo 142-8555, Japan
c Faculty of Pharmacy, Iwaki Meisei University, 5-5-1 Chuodai-ino, Iwaki 970-8551, Japan
d Graduate School of Pharmaceutical Sciences, Kumamoto University, 5-1 Oe-honmachi, Chu-o-ku, Kumamoto 862-0973, Japan

*Article info*

*Article history:*
Received 20 May 2015
Revised 23 July 2015
Accepted 14 August 2015
Available online 28 August 2015

*Keywords:*
Uracil–DNA glycosylase
Crystal structure
Leucine-intercalation loop
*Sulfolobus tokodaii*

**ABSTRACT**

Uracil–DNA glycosylases (UDGs) excise uracil from DNA by catalyzing the N-glycosidic bond hydrolysis. Here we report the first crystal structures of an archaeal UDG (*sto*UDG). Compared with other UDGs, *sto*UDG has a different structure of the leucine-intercalation loop, which is important for DNA binding. The *sto*UDG–DNA complex model indicated that Leu169, Tyr170, and Asn171 in the loop are involved in DNA intercalation. Mutational analysis showed that Tyr170 is critical for substrate DNA recognition. These results indicate that Tyr170 occupies the intercalation site formed after the structural change of the leucine-intercalation loop required for the catalysis.

@ 2015 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Introduction

Deamination is a common base modification in DNA. When cytosine and adenine are deaminated, they are converted into uracil and hypoxanthine, respectively. Deamination of guanine produces xanthine or oxanine. These deaminated bases are recognized as risk sites for transition mutations in thermophilic archaea. The archaeal family-B DNA polymerase possesses a “read-ahead” scanning mechanism: a running polymerase recognizes the base at position +4 of the DNA template strand and stalls replication if uracil or hypoxanthine is found [1]. Archaeal family-D polymerase are also inhibited by the presence of uracil in the DNA template strand [2]. Therefore, we believe that the study of the repair system in the presence of the deaminated bases is important for further understanding of the DNA replication mechanism in archaea.

Uracil–DNA glycosylase (UDG) is a monofunctional DNA glycosylase that initiates the base excision repair pathway. UDGs are widely identified in archaea, eukaryotes, bacteria, and large DNA viruses and are well-studied examples of the removal of deaminated bases from DNA. UDG-family enzymes are classified into six families on the basis of their substrate specificity, conserved motifs, and structural similarities [3,4]. Archaea commonly carry the genes for UDGs from family 4, 5, and 6 [4–6]. Biochemical studies of these UDGs have demonstrated that family 4 and 5 UDGs possess four conserved cysteine residues required to coordinate the [4Fe–4S] iron–sulfur cluster, and the substrate specificities of family 4, 5 and 6 UDGs are different as follows. Family 4 UDGs remove uracil from both double- and single-stranded DNA [7]. Family 5 UDGs have a broad substrate specificity for uracil, hypoxanthine, and xanthine in double-stranded DNA [8,9]. Family 6 UDGs exhibit a hypoxanthine–DNA glycosylase activity but do not have UDG activity [4]. Moreover, it has been reported that archaeal family 4 UDGs interact with proliferating cell nuclear antigen (PCNA), which is a processivity factor for replicative DNA polymerase [10,11]. These findings suggest a PCNA-mediated...
repair system whereby archaeal family 4 UDG is recruited to PCNA and then removes the uracil base from DNA.

In the present study, we determined the crystal structures of a family 4 UDG isolated from the thermoacidophilic crenarchaeon Sulfolobus tokodaii (stoUDG), in the free form and in the complex form with uracil (stoUDG–uracil complex). To date, approximately eighty crystal structures of UDG-family enzymes isolated from eukaryotes, bacteria, and viruses have been deposited in PDB. These structural studies revealed that the UDG-family enzymes have a common αβα sandwich fold that a four-strands parallel β-sheet is located at the center of the molecule and bordered by the α-helices despite low protein sequence similarities among them [3,12,13]. The crystal structures of family 4 UDGs isolated from the hyperthermophilic bacteria Thermus thermophilus (tthUDG) [14] and Thermotoga maritima (tmaUDG) have been reported and showed that the overall structures and the active site arrangements are similar to those of family 1 UDGs [14]. However, the crystal structure isolated from archaea had never been reported. Our present study provides the first report regarding to the archaeal UDG structures. To characterize the stoUDG structure, we compared it with the crystal structures of the bacterial family 4 UDGs. Our results indicated that the stoUDG structure involved in substrate DNA recognition differs from the bacterial family 4 UDG structures. Thus, to identify the residue important for the substrate DNA recognition, we created model structures of family 4 UDG complexed with DNA (UDG–DNA complex model) and performed mutational analyses of stoUDG.

2. Materials and methods

2.1. Enzyme and DNA preparation

The recombinant wild-type stoUDG for the mutational analyses and the stoUDG mutant truncated the C-terminal region (Tyr195–Lys220) for the crystallizations were overexpressed in Escherichia coli and purified as previously described [15]. The stoUDG mutants for the mutational analyses were constructed by PCR-mediated mutagenesis using KOD-Plus-Ver.2 DNA polymerase (TOYOBO) and DpnI nuclease (NEW ENGLAND BioLabs). Overexpression and purification of stoUDG mutants were performed with the same procedure as used for the test type stoUDG. All of the stoUDG enzymes did not contain the additional amino acids such as affinity-tags in their protein sequences. The oligonucleotide sequences used in the UDG assay and the fluorescence anisotropy-based DNA binding assay are summarized in Supplementary Table S1.

2.2. Crystallization, data collection and structure determination

Crystallizations and data collections for stoUDG crystals in the free form and the stoUDG–uracil complex were performed as pre-

<table>
<thead>
<tr>
<th>Data-collection</th>
<th>Uracil complex form</th>
<th>Free form</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB code</td>
<td>4zby</td>
<td>4zbv</td>
</tr>
<tr>
<td>Source</td>
<td>SPRing-8 BL44XU</td>
<td>MicroMAX 007</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9000</td>
<td>1.5418</td>
</tr>
<tr>
<td>Space group</td>
<td>P2_1 2_1</td>
<td>P2_1 2_1</td>
</tr>
<tr>
<td>Unit-cell</td>
<td>a = 52.60</td>
<td>a = 52.19</td>
</tr>
<tr>
<td>Parameters (Å)</td>
<td>b = 52.22</td>
<td>b = 52.25</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50.0–1.70 (1.76–1.70)</td>
<td>50.0–1.70 (1.76–1.70)</td>
</tr>
<tr>
<td>No. of observed reflections</td>
<td>123,660</td>
<td>168,247</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>23,005 (2,248)</td>
<td>23,097 (2,275)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.4 (5.2)</td>
<td>7.3 (7.1)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (99.6)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>5.6 (39.7)</td>
<td>5.0 (33.0)</td>
</tr>
<tr>
<td>(I/σ(I))</td>
<td>37.6 (4.5)</td>
<td>50.8 (6.4)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>42.7–1.70 (1.78–1.70)</td>
<td>36.9–1.70 (1.78–1.70)</td>
</tr>
<tr>
<td>Reflection used</td>
<td>22,831 (2,807)</td>
<td>23,045 (2,831)</td>
</tr>
<tr>
<td>Rwork (%)</td>
<td>16.5 (20.1)</td>
<td>16.9 (19.1)</td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>20.7 (26.2)</td>
<td>21.6 (24.0)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (99.8)</td>
<td>99.9 (99.4)</td>
</tr>
<tr>
<td>No. of non-hydrogen atoms</td>
<td>1,885</td>
<td>1,818</td>
</tr>
<tr>
<td>Protein</td>
<td>1,657</td>
<td>1,648</td>
</tr>
<tr>
<td>Ligands</td>
<td>28</td>
<td>20</td>
</tr>
<tr>
<td>Water</td>
<td>200</td>
<td>150</td>
</tr>
<tr>
<td>r.m.s deviation from ideality</td>
<td>0.010</td>
<td>0.008</td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>1.195</td>
<td>1.059</td>
</tr>
<tr>
<td>Bond angle (°)</td>
<td>27.5</td>
<td>29.4</td>
</tr>
<tr>
<td>Avg. B-factor</td>
<td>26.3</td>
<td>28.2</td>
</tr>
<tr>
<td>Protein</td>
<td>28.8</td>
<td>40.3</td>
</tr>
<tr>
<td>Ligands</td>
<td>37.2</td>
<td>40.1</td>
</tr>
<tr>
<td>Waters</td>
<td>98.0</td>
<td>98.0</td>
</tr>
<tr>
<td>Allowed region (%)</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Outlier region (%)</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Clashscore</td>
<td>3.5</td>
<td>2.7</td>
</tr>
</tbody>
</table>

* Rmerge = 100 × \sum_{hkl} ||\langle F(hkli)\rangle − \langle F(hkli)\rangle||/\sum_{hkl} ||\langle F(hkli)\rangle||, where ⟨F(hkli)⟩ is the mean value of F(hkli).

* Rwork = 100 × \sum_{hkl} ||F_0(hkli) − |F_c(hkli)||/\sum_{hkl} ||F_0(hkli)||, where F_0 and |F_c(hkli)|, the observed and calculated structure factors, respectively.

* Rfree is calculated as for Rwork, but for the test set comprising 5% reflections not used in refinement.
viously described [15]. To examine the anomalous signals of sulfur atoms, data set from a crystal of stoUDG in the free form was collected using a RAXIS IV++ image-plate detector and Cu Kα radiation (wavelength of 1.54 Å) from a MicroMax 007 generator (Rigaku). This data set was indexed, processed and scaled using the XDS package [16]. The crystals belonged to the orthorhombic space group P2₁₂₁₂₁ (Table 1) and contained one stoUDG molecule in the asymmetric unit. The initial stoUDG structure in the free form was solved by the MR method using MOLREP [17] from the CCP4 program suite [18], and manual model rebuilding was performed with COOT [19]. Structure refinement was performed using CNS [20,21] and subsequently using phenix.refine [22] from the PHENIX package [23], including the refinement of atomic displacement parameters using TLS method, with a monomer in the asymmetric unit treated as a single TLS group. Structure validation was performed using MolProbity [24]. Superimposition of UDG structures were performed using SUPERPOSE [25]. Secondary structure predictions were obtained using the Jpred3 server [26]. All molecular graphics were prepared using PyMOL [27].

2.3. UDG assay

Purified enzyme (20 nM) was incubated in a 20 µl of reaction buffer (250 nM uracil-containing DNA, 50 mM Tris–HCl pH 8.0, 5 mM EDTA, 1 mM DTT, and 0.1 mg/ml BSA) at 37 °C for 10 min. The reaction was stopped by the addition of 5 µl of 1 M NaOH. The reaction mixture was heated at 90 °C for 10 min to cleave the abasic sites. An aliquot (25 µl) of formamide was added and the reaction mixture was heated at 95 °C for 10 min. After heating, the reaction mixture was rapidly chilled on ice and then resolved on a denaturing 15% (w/v) polyacrylamide gel containing 8 M urea. The gel was visualized using a Molecular Imager FX-Pro System (Bio-Rad).

2.4. Fluorescence anisotropy-based DNA binding assay

The steady-state fluorescence anisotropy-based DNA binding assay was performed using an F-2500 fluorescence spectrophotometer equipped with polarization filters (Hitachi). Double-stranded DNA labeled with HEX was used to monitor DNA binding. Nine concentrations of wild-type stoUDG (0, 0.12, 0.35, 0.70, 1.17, 2.33, 4.67, 7.01, and 11.68 µM) and stoUDG Y170A mutant (0, 0.13, 0.39, 0.79, 1.31, 2.62, 5.24, 7.86, and 13.09 µM) were prepared by mixing each with 50 nM dSpacer-containing DNA in 400 µl of binding buffers (20 mM Tris–HCl pH 8.0, 100 mM NaCl). The mixtures were incubated in the dark for 1 h prior to the measurement. Data were recorded at an excitation wavelength of 535 nm and emission wavelength of 560 nm. The experiment was carried out three times and these results were then averaged to give the anisotropy value for that particular enzyme concentration. The averaged anisotropy values were plotted against the concentrations of the enzymes and the data was fitted to the following binding equation using gnuplot program (http://www.gnuplot.info/).

\[
A_0 = A_0 + (A_{\text{obs}} - A_0) \left( \frac{B - (B^2 - 4[DNA][UDG])^{1/2}}{2[DNA]} \right)
\]
with \( B = [\text{DNA}] + [\text{UDG}] + K_d \)

where \( A_0 \) is the observed anisotropy value, \( A_b \) is the anisotropy value of free DNA, \( A_{iso} \) is the anisotropy value of UDG-bound DNA, [DNA] is the total concentration of DNA, [UDG] is the total concentration of the enzyme, \( K_d \) is the dissociation constant of the UDG–DNA complex.

3. Results and discussion

3.1. Overall structure of stoUDG and recognition of uracil

To determine the stoUDG structures, we used a stoUDG mutant truncated at the C-terminal region (Tyr195–Lys220) of stoUDG because the crystallization of wild-type stoUDG was not reproducible [15]. The crystal structures of stoUDG in the free form and of the stoUDG–uracil complex were determined at 1.70–1.90 Å resolution. Data-collection and refinement statistics are summarized in Table 1. Almost all amino acid residues in stoUDG could be assigned to the final models except for the last two amino acid residues (Lys193 and Arg194) of stoUDG in the free form.

Crystal structures of stoUDG in the free form and of the stoUDG–uracil complex show that stoUDG has the \( \alpha/\beta/\alpha \) sandwich fold, consisting of a central six-strands \( \beta \)-sheet bordered by seven \( \alpha \)-helices and four \( 3_1 \)-helices (Fig. 1a). This is a fold common among UDGs. A [4Fe–4S] iron–sulfur cluster, which is one of the characteristic feature of family 4 UDGs, is coordinated to the four conserved cysteine residues (Cys14, Cys17, Cys86 and Cys102) at the N-terminal side of the molecule (Fig. 1a). A difference electron density map derived from the stoUDG–uracil complex diffraction data was calculated with the coordinate of the stoUDG structure in the free form and this initial \( mF_o - DF \) map shows what we interpret as the electron density for a uracil molecule. The uracil-binding pocket of stoUDG is formed by Gly41, Glu42, Ala43, Pro44, Glu48, Phe55, Asn82, and His 164 (Fig. 1b). The specific interactions between stoUDG and uracil are similar in tthUDG and human UNG1 (Figs. 1b and S2) [14,28]. Particularly, family 4 UDGs are active against uracil but inactive against thymine or cytosine. The family 4 UDG–uracil complex structures show that the side chain of the conserved glutamic acid residues (Glu48 of stoUDG and Glu47 of tthUDG) is located near the C5 of uracil. This proximity seems to prevent the binding of C5-substituted pyrimidines such as thymine because of steric hindrance (Fig. 1c) [14]. Moreover, these glutamic acid residues are surrounded by other amino acid residues (Val127, Gly45, Pro44, Pro54 and Val156 of stoUDG); therefore, the rotamer cannot change easily for the escape from the steric hindrance. A similar mechanism for discrimination between uracil and thymine has been observed in human UNG1 (Fig. S2). Substitution of Tyr147 of human UNG1 with alanine, cysteine, or serine results in the additional thymine-DNA glycosylase activity [29]. In contrast to uracil, cytosine has an amino group at C4 position. The crystal structures of the family 1 and family 4 UDG–uracil complexes show that O4 of uracil is recognized as the electron density for a uracil molecule. The uracil-binding pocket of stoUDG is formed by Gly41, Glu42, Ala43, Pro44, Glu48, Phe55, Asn82, and His 164 (Fig. 1b). The specific interactions between stoUDG and uracil are similar in tthUDG and human UNG1 (Figs. 1b and S2) [14,28]. Particularly, family 4 UDGs are active against uracil but inactive against thymine or cytosine. The family 4 UDG–uracil complex structures show that the side chain of the conserved glutamic acid residues (Glu48 of stoUDG and Glu47 of tthUDG) is located near the C5 of uracil. This proximity seems to prevent the binding of C5-substituted pyrimidines such as thymine because of steric hindrance (Fig. 1c) [14]. Moreover, these glutamic acid residues are surrounded by other amino acid residues (Val127, Gly45, Pro44, Pro54 and Val156 of stoUDG); therefore, the rotamer cannot change easily for the escape from the steric hindrance. A similar mechanism for discrimination between uracil and thymine has been observed in human UNG1 (Fig. S2). Substitution of Tyr147 of human UNG1 with alanine, cysteine, or serine results in the additional thymine-DNA glycosylase activity [29]. In contrast to uracil, cytosine has an amino group at C4 position. The crystal structures of the family 1 and family 4 UDG–uracil complexes show that O4 of uracil is recognized through the hydrogen bonds with the main chain nitrogen atom of phenylalanine residue and the side chain Nδ of asparagine residue (Figs. 1b and S2). The uracil recognition by these conserved asparagine residues is important for defining the preference for uracil over cytosine. It has been reported that the substitution of

![Fig. 2. Structure comparison of family 4 UDGs. stoUDG (PDB code 4zby), tthUDGa (PDB code 1ui0) and tmaUDG (PDB code 1vk2) are colored green, yellow and red respectively. To highlight the differences, the regions that are not superimposable are colored.](image)

![Fig. 3. Family 4 UDG–DNA complex models. The family 4 UDG–DNA complex models were created by superimposing the family 4 UDG structure on the crystal structure of tthUDGb–DNA complex (PDB code 2ddg). tthUDGb belongs to family 5 UDG and forms the most similar to family 4 UDGs among the known UDG–DNA complex structures. (a) Superimposition of family 4 UDGs and the tthUDGb–DNA complex. stoUDG (green, PDB code 4zby), tthUDGa (orange, PDB code 1ui0), tmaUDG (pink, PDB code 1vk2), and tthUDGb (light blue, PDB code 2ddg) are shown as ribbon representations. The stoUDG, tthUDGa, and tmaUDG structures are superimposed on the tthUDGb–DNA complex structure with the r.m.s. deviation value of 1.72 Å (176 corresponding Cα atoms), 1.96 Å (168 corresponding Cα atoms), and 1.78 Å (168 corresponding Cα atoms), respectively. (b) Stereo-view of the superimposition structure of stoUDG (PDB code 4zby) and the tthUDGb–DNA complex (PDB code 2ddg). The leucine-intercalation loop of stoUDG is shown as bold-ribbon representation, colored red. The uracil and MES molecules are shown as CPK representations.](image)
Asn204 of human UNG1 with aspartic acid (N204D) reduces the selectivity for uracil, and the UNG1 N204D mutant excises cytosine from DNA [29]. In addition, a MES molecule, which is contained in the crystallization condition of stoUDG, was observed in all the determined stoUDG structures. Further, we confirmed the presence of the MES molecule by examining the anomalous signal from the sulfur atoms (Figs. 1a and S1a). The MES molecule was located at the positively charged DNA binding cleft close to the uracil-binding pocket, and the sulfo group of MES formed hydrogen bonds with the main chain nitrogen atoms of Arg123 and His164 (Fig. S1). These results suggest that the sulfo group of MES mimics the phosphate backbone of the substrate DNA.

The overall r.m.s. deviation value between the Ca positions of stoUDG in the free form and in the stoUDG–uracil complex was 0.29 Å over 192 residues; stoUDG structures in the free form and in the stoUDG–uracil complex did not show any significant conformational changes. These results indicate that a structural change is not required for the recognition of the uracil base by stoUDG. Moreover, the C-terminal Arg194 is located at the opposite side of the uracil-binding pocket (Fig. 1a). We believe that the missing C-terminal amino acid residues for the crystallization do not interfere with DNA binding because they are distant from the DNA binding surface and the stoUDG truncated mutant retained most of the UDG activity compared with wild-type stoUDG [15].

3.2. Structure comparison of family 4 UDGs

To clarify the structural differences between the family 4 UDGs, the stoUDG structure was compared with that of tthUDGα (PDB code lui0) and tmaUDG (PDB code 1vk2) (Fig. 2). Overall superimpositions of stoUDG on tthUDGα and tmaUDG gave the r.m.s. deviation values of 1.42 Å over 179 corresponding Ca atoms and 1.37 Å over 176 corresponding Ca atoms, respectively. Three regions located at the C-terminus side of the molecules (corresponding residues in stoUDG; region I: Ser131–Ser138, region II: Val149–Glu156 and region III: Leu169–Asn173) did not especially coincide.
with each other, while the structure at the N-terminus side of the molecules and at the central β-sheets resembled one another. UDGs recognize the uracil base in DNA using the extrahelical recognition mechanism whereby the uracil base is expelled from the base stack and then accepted into the active site of the enzyme [30]. In addition, the catalytic mechanisms of family 4 and family 1 UDGs are predicted to be similar [7]. Moreover, in family 1 UDGs, five characteristic motifs involved in the catalytic activity have been identified, and the corresponding amino acid residues in human UNG1 are Asp145–His148 (the water activating loop), His268–Arg276 (the leucine-intercalation loop) [31]. The “region III”, which does not superimpose very well in family 4 UDGs, corresponds to the leucine-intercalation loop. This loop is important for the stabilization of the extrahelical conformation in DNA; the amino acid residue located at the leucine-intercalation loop (Leu272 in human UNG1) penetrates into the flipped-out DNA. It has been reported that, in human UNG1, substitution of Leu272 with alanine reduces enzymatic activity, and alters DNA binding affinity. This effect is probably due to the insusceptibility of the space created by the uracil flipping [31].

3.3. Model structures of the family 4 UDGs–DNA complexes and identification of the intercalating residue of stoUDG

To analyze the effect of the structural differences of the leucine-intercalation loop among family 4 UDGs on DNA binding, we created the UDG–DNA complex models for this family (Fig. 3). The stoUDG–DNA complex model showed that the abasic site of the model is located just above the uracil-binding pocket of stoUDG. The distance C1‘-N1 and the angle C1‘-N1-C4 were 2.49 Å and 120.2°, respectively (Fig. S3a). These uracil and abasic site in the stoUDG–DNA complex model appear to be at the suitable positions because they are placed similar to those in the crystal structure of the human UNG1–DNA complex (Fig. S3b) [31]. Moreover, the phosphate backbone at position +2 in the DNA model coincides with the sufo group of MES, which is considered to be the mimic phosphate backbone of DNA (Fig. S3c). The leucine-intercalation loop in the model is located between the uracil-binding pocket and the phosphate backbone at position +2 (Fig. 3b). Thus, our stoUDG–DNA complex model should be sufficient for the examination of the effect of the structural differences of the leucine-intercalation loop among family 4 UDGs on DNA binding. Fig. 4 shows the close-up view of the region around the leucine-intercalation loop. The crystal structure of the tthUDG–DNA complex showed that the side chain of Gin 194 occupied the space created by the base flipping and revealed the position of the intercalation site (Fig. S4a) [32]. Previous studies on tthUDGs have predicted that the side chain of Arg161 penetrates into the flipped-out DNA and replaces the uracil base [14]. Our tthUDG–DNA complex model supports this prediction and indicates that Arg161 will occupy the intercalation site (Fig. 4a). Moreover, the tmaUDG–DNA complex model suggests the same intercalating mechanism as in tthUDGs. Arg167 in tmaUDG is the identical residue to Arg161 in tthUDG (Fig. 6) and is located at a position suitable for penetration into the flipped-out DNA (Fig. 4b). As shown in Fig. 4c, the stoUDG–DNA complex model could not show the correct position of the intercalating residue since there were no residues, whose side chain can occupy the intercalation site. However, we assumed that the leucine-intercalation loop would undergo a structural change after the recognition of the extrahelical conformation of DNA by

![Fig. 5. Mutational analysis of stoUDG. (a) Comparison of the enzymatic activity of stoUDG and its mutants. (b) Fluorescence anisotropy derived from HEX was used to show the best fit of the data to the binding equation; and the dissociation constant values for wild type stoUDG and Y170A were 1.1 ± 0.2 μM and 12.0 ± 1.4 μM, respectively.](image)

![Fig. 6. Protein sequence alignments around the leucine-intercalation loop in family 4 UDGs. The protein sequences of family 4 UDGs with confirmed UDG activity were obtained from UniProt database [33]. Their IDs are as follows; Sulfolobus tokodaii, Q56YD0 [15]; Pyrococcus furiosus, Q8U144 [11]; Archaeoglobus fulgidus, Q28007 [34]; Aeropyrum pernix, Q9YF11 [35]; Pyrococcus furiosus, Q8ZSY2 [36]; Thermus thermophilus, Q5SK55 [14]; Thermotoga maritima, Q9WWY1 [37]; and Deinococcus radiodurans, Q9RTK9 [38]. *(Ar)* and *(Ba)* mean archaea and bacteria, respectively. The asterisk symbol indicates the intercalating residue position. The protein sequences labeled with the dagger symbol indicate the predicted secondary structure. The amino acid residues in the bold italic font indicate the regions from the intercalating residue to the next α-helix, and the underlined amino acid residues indicate the helical regions.](image)
stoUDG. Therefore, we focused on three amino acid residues in the loop (Leu169, Tyr170, and Asn171) that are close to the intercalation site and have a branched side chain that can occupy the space created by the uracil flipping (Figs. 4c and S4b). To identify the intercalating residue of stoUDG, we constructed the stoUDG mutants by substituting Leu169, Tyr170, and Asn171 with alanine (L169A, Y170A and N171A, respectively) and measured their enzymatic activities. These results demonstrated that Y170A abolished the enzymatic activity, whereas L169A and N171A retained most of the UDG activity (Fig. 5a). Additionally, the results of the DNA binding assay revealed that Y170A had reduced binding affinity compared with the wild-type stoUDG (Fig. 5b). These results indicate that Tyr170 is the intercalating residue of stoUDG. In family 4 UDGs, the intercalating residue is located at C-terminal side of the short-helix in the leucine-intercalation loop, whereas the intercalating residues of family 1 UNG1 and family 5 thtUDGb are located at the N-terminal side of the short-helix (Fig. S4). These results suggest that the location of the intercalating residue at the short helix is a characteristic feature of family 4 UDGs. The distances of the Cα position in the free form structure between Tyr170 of stoUDG and Arg161 of thtUDGb or Arg167 of tmaUDG are 3.92 Å and 2.98 Å, respectively. Moreover, the leucine-intercalation loop of stoUDG is a part of a rigid helix-loop-helix motif, in which the loop consists of only three amino acid residues (Fig. S4b). However, our results suggest that a structural change in the leucine-intercalation loop of stoUDG will be required. In addition, the tmaUDG structure around the leucine-intercalation loop showed that the last C-terminal α-helix of tmaUDG slanted slightly when comparing to those of stoUDG and thtUDGb (Fig. S5). These results suggest that DNA binding manners of family 4 UDGs involved in the leucine-intercalation loop structure are classified into at least three types: requiring the structural change of the leucine-intercalation loop (stoUDG type), the C-terminal α-helix slanting (tmaUDG type), and no alteration (thtUDG type). The protein sequences around the leucine-intercalation loop of the other family 4 UDGs, which is confirmed the UDG activity, were aligned based on the sequence similarities and secondary structure, and they were classified into the three types (Fig. 6). The protein sequences could be sorted to the three types of DNA binding manner by the loop structure length and existence of the proline residue in the leucine-intercalation loop. This result may suggest that the structural change required for DNA binding, and the proline residue might define the conformation of the leucine intercalation loop and the C-terminal α-helix orientation. The sequence alignments also indicated that the archaeal family 4 UDGs possess a short loop structure and the proline residue in the leucine-intercalation loop. These findings suggest that the archaeal family 4 UDGs may require a structural change of the leucine-intercalation loop when recognizing the substrate DNA.

In conclusion, the study presented here is the first to provide the structural features of an archaeal UDG. The overall structure and the uracil recognition mechanism of stoUDG are similar to those of other UDGs. However, the leucine-intercalation loop structure is characteristic to the stoUDG structure, and Tyr170 located at the leucine-intercalation loop is the intercalating residue of stoUDG. These findings suggest that, in stoUDG, the side chain of Tyr170 occupies the intercalation site created by the uracil flipping after the structural change in the leucine-intercalation loop required for the catalysis.

Acknowledgements

We are grateful to Dr. Masato Yoshimura, Dr. Eiki Yamashita, Professor Atsushi Nakagawa, and the beamline staff for their support at the BL44XU in the SPring-8. Synchrotron experiments were performed with the approval of the Joint Research Committee of the Institute for Protein Research, Osaka University, and the Japan Synchrotron Radiation Research Institute (Proposal No. 2007B6927). This work was partly supported by the National Project on Protein Structural and Functional Analysis by the Ministry of Education, Culture, Sports, Science and Technology of Japan (Protein 3000 project). This work was also partly supported by KUMAYAKU Alumni Research Fund.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2015.08.019.

References
