Published online 30 December 2014 *Nucleic Acids Research, 2015, Vol. 43, No. 2* ***1081–1089***
doi: 10.1093/nar/gku1332

A structural determinant in the uracil DNA glycosylase superfamily for the removal of uracil from adenine/uracil base pairs

Dong-Hoon Lee[1], Yinling Liu[2], Hyun-Wook Lee[1], Bo Xia[1], Allyn R. Brice[2], Sung-Hyun Park[1], Hunter Balduf[1], Brian N. Dominy[2] and Weiguo Cao[1,*]

[1]Department of Genetics and Biochemistry, South Carolina Experiment Station, Clemson University, 049 Life Sciences Facility, 190 Collings Street, Clemson, SC 29634, USA and [2]367 Hunter Laboratories, Department of Chemistry, Clemson University, Clemson, SC 29634, USA

Received September 16, 2014; Revised December 05, 2014; Accepted December 09, 2014

ABSTRACT

The uracil DNA glycosylase superfamily consists of several distinct families. Family 2 mismatch-specific uracil DNA glycosylase (MUG) from *Escherichia coli* is known to exhibit glycosylase activity on three mismatched base pairs, T/U, G/U and C/U. Family 1 uracil *N*-glycosylase (UNG) from *E. coli* is an extremely efficient enzyme that can remove uracil from any uracil-containing base pairs including the A/U base pair. Here, we report the identification of an important structural determinant that underlies the functional difference between MUG and UNG. Substitution of a Lys residue at position 68 with Asn in MUG not only accelerates the removal of uracil from mismatched base pairs but also enables the enzyme to gain catalytic activity on A/U base pairs. Binding and kinetic analysis demonstrate that the MUG-K68N substitution results in enhanced ground state binding and transition state interactions. Molecular modeling reveals that MUG-K68N, UNG-N123 and family 5 *Thermus thermophiles* UDGb-A111N can form bidentate hydrogen bonds with the N3 and O4 moieties of the uracil base. Genetic analysis indicates the gain of function for A/U base pairs allows the MUG-K68N mutant to remove uracil incorporated into the genome during DNA replication. The implications of this study in the origin of life are discussed.

INTRODUCTION

Enzymes in the uracil DNA glycosylase (UDG) superfamily are well known for their role in the removal of deaminated base damage in DNA repair. So far, six families in the superfamily have been discovered and studied to varied extent (1–7). Family 1 UNG is a highly efficient enzyme that excises uracil from all double-stranded uracil-containing base pairs and single-stranded uracil-containing DNA (2,5). *Escherichia coli* MUG, which belongs to family 2, is named for its UDG activity on mismatched T/U, G/U and C/U base pairs (8–11). In addition to its UDG activity on mismatched base pairs, *E. coli* MUG is also a robust xanthine DNA glycosylase (12). Like family 1 UNG, family 3 SMUG1 can remove uracil from double- and single-stranded DNA albeit it does so much less efficiently (13,14). Family 4 UDGa, found in prokaryotic organisms, is a UDG but the full spectrum of its activity against all deaminated bases is not known (15). Family 5 UDGb, also found in prokaryotic organisms, is not only a UDG but also a hypoxanthine and a xanthine DNA glycosylase (16). However, the UDG activity from family 5 UDGb is limited to double-stranded uracil-containing DNA and the activity on A/U base pairs is lower than that on mismatched base pairs (16). The newly discovered family 6 enzymes are hypoxanthine DNA glycosylases (3).

Within the UDG superfamily, different families have evolved different specificities toward deaminated bases and apply a multitude of catalytic elements to catalyze the breakage of the glycosidic bond associated with pyrimidine and purine deaminated bases (3,12,16). Two important motifs have been identified in the UDG superfamily, in which motif 1 contains residues that form the base recognition pocket and a water activating residue (D64 in *E. coli* UNG and N18 in MUG) and motif 2 includes a catalytically important histidine residue that forms a low barrier hydrogen bond with O2 of uracil (1,6,17). Both family 1 UNG and family 2 MUG contain UDG activity, however one of the fundamental differences between the two families is that family 2 MUG does not demonstrate any enzymatic activity on A/U base pairs. Even though crystal structures of the *E. coli* MUG enzyme are available, to our knowledge there has been no experimental investigation to elucidate the basis for its lack of UDG activity on A/U base pairs.

*To whom correspondence should be addressed. Tel: +1 864 656 4176; Fax: +1 864 656 0393; Email: wgc@clemson.edu

Based on a structural comparison of *E. coli* MUG and UNG enzymes, we identified Lys-68 as a potential structural element located outside of motifs 1 and 2 that can determine the UDG activity on A/U base pairs. Mutational analysis presented here demonstrates that a K68N substitution not only allows *E. coli* MUG to act on A/U base pairs, but also increases its catalytic efficiency on all other double-stranded uracil-containing DNA. Binding analysis shows that the substitution enhances the binding affinity of K68N to all uracil-containing double-stranded DNA. Genetic analysis suggests that the K68N mutant can act as a UDG to remove uracil from A/U base pairs, which are formed by misincorporation of dUMP into genomic DNA. Molecular modeling analysis provides structural information on interactions between the Asn in the K68 position and a uracil base. Interestingly, while the substitution in *E. coli* UNG (N123A) substantially reduces its UDG activity on A/U base pairs and other double-stranded uracil-containing base pairs, the A111N mutation in the same position in family 5 UDGb from *Thermus thermophiles* increases its activity toward double-stranded uracil-containing base pairs with the most notable increase occurring on A/U base pairs. These results underscore the role of this position as an important structural determinant for UDG activity and specificity in multiple families. The structural basis that underlies the gain of enzymatic function and its potential implications for the evolution of a 'leaky MUG' during the origin of life are discussed.

MATERIALS AND METHODS

Plasmid construction, cloning and expression of *E. coli* MUG and *E. coli* UNG

The *E. coli* MUG gene (Uniprot accession number P0A9H1) was amplified by polymerase chain reaction (PCR) using the forward primer Ec.MUG F (5′-TGG GGT ACC CCA TGG GTT GAG GAT ATT TTG GCT CCA GGG- 3′; the NcoI site is underlined) and the reverse primer Ec.MUG R (5′-CCC GGA TCC TTA TCG CCC ACG CAC TAC CAG CGC CTG GTC-3′; the BamHI site is underlined). The PCR reaction mixture (50 μl) consisted of 8 ng of *E. coli* genomic DNA, 200 nM forward primer Ec.MUG F and reverse primer Ec.MUG R, 1 x Taq PCR buffer (New England Biolabs), 200 μM each dNTP and 5 units of Taq DNA polymerase (New England Biolabs). The PCR procedure included a predenaturation step at 94°C for 3 min, 30 cycles of three-step amplification with each cycle consisting of denaturation at 94°C for 40 s, annealing at 60°C for 40 s and extension at 72°C for 1 min, and a final extension step at 72°C for 10 min. The PCR product was purified with DNA gel extraction kit (Zymo Research). Purified PCR product and plasmid pET32a without the thioredoxin tag (pET32a(-)) were digested with NcoI and BamHI, purified with Gene Clean 2 Kit and ligated according to the manufacturer's instructional manual. The ligation mixture was electroporated into *E. coli* strain JM109 competent cells. The sequence of the *E. coli* MUG gene in the resulting plasmid (pET32a(-)-MUG) was confirmed by DNA sequencing.

E. coli UNG gene (Uniprot accession number P12295) was amplified by PCR using the forward primer UDG-NdeI (5′-GGG AAT TC CAT ATG GCT AAC GAA TTA ACC TGG CAT GAC-3′; the NdeI site is underlined) and the reverse primer UDG-HindIII (5′-CCC AAG CTT CTC ACT CTC TGC CGG TAA TAC TGG-3′; the HindIII site is underlined). The PCR reaction mixture (50 μl) consisted of 8 ng of *E. coli* genomic DNA, 200 nM forward primer UDG-NdeI and reverse primer UDG-HindIII, 1 x Taq PCR buffer (New England Biolabs), 200 μM each dNTP and 5 units of Taq DNA polymerase (New England Biolabs). The PCR procedure included a predenaturation step at 94°C for 3 min, 30 cycles of three-step amplification with each cycle consisting of denaturation at 94°C for 40 s, annealing at 60°C for 40 s and extension at 72°C for 1 min, and a final extension step at 72°C for 10 min. The PCR product was purified with DNA gel extraction kit (Zymo Research). Purified PCR product and plasmid pET21a were digested with NdeI and HindIII, purified with Gene Clean 2 Kit and ligated according to the manufacturer's instructional manual. The ligation mixture was electroporated into *E. coli* strain JM109 competent cells. The sequence of the *E. coli* UNG gene in the resulting plasmid (pET21a-UNG) was confirmed by DNA sequencing.

The N-terminal His-6-tagged *E. coli* pET32a(-)-MUG and C-terminal His-6-tagged *E. coli* UNG were electroporated into *E. coli* strain BH214 (*mug*$^-$, *ung*$^-$). An overnight *E. coli* culture was diluted 100-fold into Luria–Bertani (LB) medium supplemented with 100 μg/ml ampicillin. The *E. coli* cells were grown at 37°C while being shaken at 250 rpm until the optical density at 600 nm reached ∼0.6. The cultures were cooled down to 22°C, induced with 1 mM isopropy beta-D-thiogalactopyranoside (IPTG) and grown at 22°C for 20 h. The cells were collected by centrifuging at 4000 rpm at 4°C for 20 min with JLA-8.1000 rotor in Avanti J-26S XPI (Beckman Coulter) and washed once with precooled sonication buffer [20 mM Tris-HCl (pH 7.5), 1 mM ethylenediaminetetraacetic acid (EDTA) (pH 8.0), 0.1 mM dithiothreitol (DTT), 0.15 mM phenylmethylsulfonyl fluoride and 50 mM NaCl].

To purify the *E. coli* MUG, the cell pellet from a 500-ml culture grown to late exponential phase was suspended in 7 ml of sonication buffer and sonicated at output 5 for 3 × 1 min with 1-min rest on ice between intervals using Q125 sonicator (Qsonica). The lysate was clarified by centrifuging at 12000 rpm at 4°C for 20 min. The supernatant was transferred into a fresh tube and loaded into a 1 ml HiTrap chelating column. The column was washed with chelating buffer A [20 mM sodium phosphate (pH 7.4), 500 mM NaCl and 2 mM imidazole]. The bound protein in the column was eluted with a linear gradient of 0–100% chelating buffer B (chelating buffer A and 500 mM imidazole without adjusting pH). Fractions of elution were analyzed by 12% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and those fractions containing MUG protein at around 23.5 kD were pooled. The partially purified MUG protein was loaded onto HiTrap SP column equilibrated with HiTrap SP column buffer A [20 mM HEPES (pH 8.0), 1 mM EDTA and 0.1 mM DTT] and eluted with HiTrap SP column buffer B [20 mM HEPES (pH 8.0), 1 mM EDTA, 0.1 mM DTT and 1 M NaCl]. The MUG protein was considered highly purified as there were no other pro-

tein bands detected by SDS-PAGE analysis (Supplementary Figure S1). The pooled *E. coli* MUG were concentrated, quantified and stored at −20°C. *E. coli* UNG proteins were purified using the same protocol as described above for *E. coli* MUG proteins.

Site-directed mutagenesis

Site-directed mutagenesis of *E. coli* MUG was performed using an overlapping extension PCR procedure (18). The first round of PCR was carried out using pET32a(-)-MUG as template DNA with two pairs of primers, Ec.MUG-F and K68N-R (5′-ACG GTC TAC CAG ATT GGT GAC GCC ACA ACG ATA ATC-3′) and Ec.MUG-R and K68N-F (5′-CGT TGT GGC GTC ACC AAT CTG GTA GAC CGT CCA ACG-3′), respectively. The PCR mixture (50 μl) contained 10 ng of pET32a(-)-MUG as a template, 200 nM each primer, 200 μM each dNTP, 1× *Taq* DNA polymerase buffer and 5 units of *Taq* DNA polymerase. The PCR procedure included a predenaturation step at 95°C for 3 min, 30 cycles of three-step amplification with each cycle consisting of denaturation at 95°C for 50 s, annealing at 65°C for 50 s and extension at 72°C for 1 min, and a final extension step at 72°C for 10 min. The resulting two expected DNA fragments were used for overlapping PCR to introduce the desired mutation. This second run of PCR reaction mixture (100 μl), which contained 1 μl of each of the first run PCR products, 100 μM each dNTP, 1× *Taq* DNA polymerase buffer and 5 units of *Taq* DNA polymerase, was initially carried out with a predenaturation at 95°C for 2 min, five cycles with each cycle of denaturation at 95°C for 30 s and annealing and extension at 60°C for 4 min, and a final extension at 72°C for 5 min. Afterward, 100 nM outside primers (Ec.MUG-F and Ec.MUG-R) were added to the above PCR reaction mixture to continue the overlapping PCR reaction under the same reaction condition with 25 cycles. The PCR product was cloned into pET32a(-) as described above. The recombinant plasmid (pET32a(-)-MUG-K68N) containing the desired mutated gene was confirmed by DNA sequencing and electroporated into *E. coli* strain BH214 (*mug*$^-$, *ung*$^-$).

The *E. coli* UNG-N123A mutant was constructed similarly. The first round of PCR was carried out using pET21a-UNG as the DNA template and two pairs of primers, UDG-NdeI and N123A-R (5′-CGT CAA CAC AGT TGC GAG TAG CAG AAC GCC CTG ACG-3′) and UDG-HindIII and N123A-F (5′-GGC GTT CTG CTA CTC GCA ACT GTG TTG ACG GTA CGC-3′), respectively. The PCR mixture (50 μl) contained 10 ng of pET21a-UNG as a template, 200 nM each primer, 200 μM each dNTP, 1× *Taq* DNA polymerase buffer and 5 units of *Taq* DNA polymerase. The PCR procedure included a predenaturation step at 95°C for 3 min, 30 cycles of three-step amplification with each cycle consisting of denaturation at 95°C for 50 s, annealing at 65°C for 50 s and extension at 72°C for 1 min, and a final extension step at 72°C for 10 min. The resulting two expected DNA fragments were used for overlapping PCR to introduce the desired mutation. This second run of the PCR reaction mixture (100 μl), which contained 1 μl of each of the first run PCR products, 100 μM each dNTP, 1× *Taq* DNA polymerase buffer and 5 units of *Taq* DNA polymerase, was initially carried out with a predenaturation at 95°C for 2 min, five cycles with each cycle of denaturation at 95°C for 30 s and annealing and extension at 60°C for 4 min, and a final extension at 72°C for 5 min. Afterward, 100 nM outside primers (UDG-NdeI and UDG-HindIII) were added to the above PCR reaction mixture to continue the overlapping PCR reaction under the same reaction condition with 25 cycles. The PCR product was cloned into pET21a as described above. The recombinant plasmid (pET21a-UNG-N123A) containing the desired mutated gene was confirmed by DNA sequencing and electroporated into *E. coli* strain BH214 (*mug*$^-$, *ung*$^-$). The *E. coli* MUG and UNG mutant proteins were expressed and purified as described above. The cloning, protein expression and site-directed mutagenesis of *Thermus thermophilus* HB8 (Tth) UDGb gene (Uniprot accession number Q5SJ65) were performed as previously described (16).

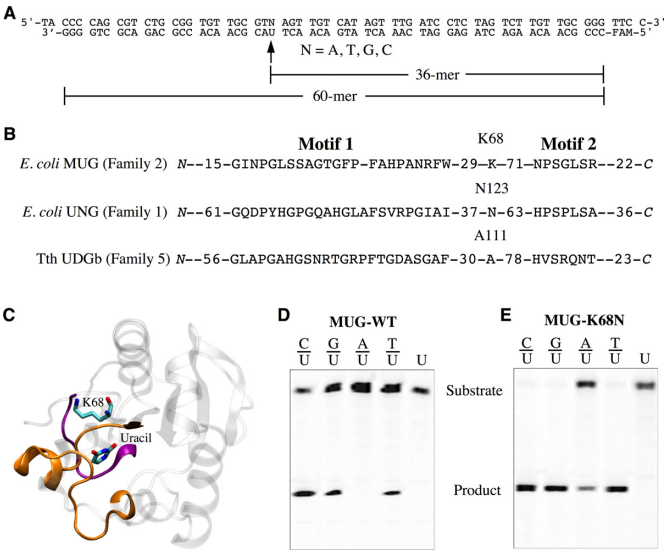

Figure 1. Substrates, sequence alignment, structure and UDG activity. (**A**) Sequences of uracil-containing DNA substrates. (**B**) Sequence alignment of *E. coli* MUG (GenBank accession number P0A9H1.1), *E. coli* UNG (GenBank accession number NP_417075.1) and Tth UDGb (GenBank accession number YP_144415.1). (**C**) Structure of *E. coli* MUG (PDB 1MUG) with uracil. Motifs 1 and 2 are shown in orange and purple, respectively. K68 and uracil are colored by atom type. (**D**) DNA glycosylase activity of MUG-WT on uracil-containing substrates. Cleavage reactions were performed as described in the Materials and Methods section with 100 nM MUG-WT protein and 10 nM substrate. (**E**) DNA glycosylase activity of MUG-K68N on uracil-containing substrates. Cleavage reactions were performed as described in the Materials and Methods section with 100 nM MUG-K68N protein and 10 nM substrate.

DNA glycosylase activity assay

The sequences of the fluorescently labeled oligodeoxynucleotide substrates are shown in Figure 1A and prepared as previously described (12). DNA glycosylase cleavage assays for *E. coli* MUG and UNG were performed at 37°C for 60 min in a 10-μl reaction mixture containing 10 nM DNA substrate, 100 nM DNA glycosylase, 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 5 mM EDTA and 2 mM 2-mercaptoethanol. The glycosylase assay for Tth UDGb was performed as previously described (16). The resulting aba-

sic sites were cleaved by incubation at 95°C for 5 min after adding 0.5 μl of 1 N NaOH. Reactions were quenched by addition of an equal volume of GeneScan stop buffer. Samples (3.5 μl) were loaded onto a 7 M urea-10% denaturing polyacrylamide gel. Electrophoresis was conducted using an Applied Biosystems 3130xl Genetic analyzer. Cleavage products and remaining substrates were quantified using GeneMapper analysis software for Applied Biosystems 3130xl Genetic analyzer.

Gel mobility shift assay

The binding reactions were performed on ice for 10 min in a 10-μl volume containing 50 nM DNA substrate, 20 mM Tris-HCl (pH 7.2), 50 mM NaCl, 5 mM EDTA, 1 mM DTT, 0.1 mg/ml bovine serum albumin, 10% glycerol and the indicated amount of *E. coli* MUG protein. Samples were supplemented with 2 μl of 100% glycerol and electrophoresed at 200 V on a 6% native polyacrylamide gel in 1 x TB buffer (89 mM Tris base and 89 mM boric acid) supplemented with 5 mM EDTA. The bound and free DNA species were analyzed using a Typhoon 9400 Imager (Molecular Dynamics) with the following settings: photomultiplier tube at 600 V, excitation at 495 nm, and emission at 535 nm. To determine the K_D values of the *E. coli* MUG enzymes to G/U base pairs, the mobility shift reactions were performed with 50 nM G/U-containing substrate and 0, 25, 50, 100, 150, 200, 250, 300, 400 or 500 nM of *E. coli* MUG enzyme (MUG-WT or MUG-K68N). To determine the K_D values of the *E. coli* MUG-K68N to the A/U base pair, the mobility shift reactions were performed with 100 nM A/U-containing substrate and 0, 50, 100, 200, 300, 400, 500, 600, 800 or 1000 nM *E. coli* MUG K68N protein. The intensities of the bands were quantified by Image J program. It has been reported that *E. coli* MUG exhibits cooperative binding to DNA (19). Accordingly, Hill plot was used for the K_D calculation using Deltagraph software,

$$[S]_{Free} = [S]_T - \frac{[S]_T \times [E]^h}{K_D^h + [E]^h},$$

where K_D is the equilibrium dissociation constant, $[E]$ is the enzyme concentration, $[S]_T$ is the total substrate concentration, h is the Hill coefficient and $[S]_{Free}$ is the free substrate concentration.

In vivo assay

The plasmids pET32a-MUG and pET32a-MUG K68N were amplified using Ec.MUGPBS-F: 5′-GG CTG CAG GAA TTC CAT GGT TGA GGA TAT TTT GGC TC-3′and Ec.MUGPBS-R: 5′-GAA TTC AAG CTT TTA TCG CCC ACG CAC TAC CAG C-3′ and PCR products were digested with EcoRI and HindIII, and the fragment containing the MUG gene was cloned to pBluescript SK (+) to generate pBS-MUG. The resulting plasmids were confirmed by sequencing and electroporated to *E. coli* BW276 (*Δxth*, *dut-1*, *ung-1*). Tester cultures, inoculated as a single colony, were grown in LB medium (with 100 μg/ml ampicillin and 125 μg/ml thymidine) at 22°C for 48–72 h and diluted 100-fold into prewarmed media at 22°C and inoculated with shaking for 3–4 h until optical density at 600 nm reached 0.6. After adding IPTG to a final concentration of 1 mM, the cultures were incubated at 22°C for additional 4 h. Before plating, 40 μl of 100 mM IPTG was spread on the LB plates containing ampicillin and thymidine. Afterward, 10^6-diluted cells (100 μl) were plated on LB plates containing ampicillin and thymidine. Cell numbers were scored after 24 h of incubation at 42°C or after 72 h of incubation at 22°C. The relative plating efficiencies were taken as the ratios of the cell numbers between 42 and 22°C. Data were calculated based on six independent experiments.

Molecular modeling

The crystal structure of family 2 (*E. coli* MUG), family 1 (*E. coli* UNG) and family 5 (Tth UDGb) UDG was acquired from the RCSB Protein Data Bank (accession codes 1MUG, 2EUG and 2DEM, respectively) and used as a model for subsequent computational analysis. The crystal structure of the human UDG–DNA complex (PDB accession code 1EMH) was used as the DNA model to build the flipped out double-stranded DNA complexed with the protein MUG using the Swiss-Pdb Viewer (SPDBV) program (20). The single amino acid mutants MUG-K68N, UNG-N123A and UDGb-A111N were also made using the mutation tool in the Swiss-Pdb Viewer program and the 'best rotamer' was chosen with the lowest clash score (Supplementary Table S1).

RESULTS

Amino acid substitutions at the K68 position and UDG activity on A/U base pairs

Enzymes in the UDG superfamily contain two motifs that are important for substrate binding and catalysis (Figure 1B). The N123 between the two motifs in *E. coli* UNG interacts with uracil by forming bidentate hydrogen bonds with the N3 and O4 moieties (Figure 1B and Supplementary Figure S2). *E. coli* MUG is a small protein of 168 amino acids with UDG activity that excises uracil from a mismatched base pair. The sequence and structural equivalent position to the N123 in *E. coli* UNG in the MUG enzyme is K68 (Figure 1B). The crystal structure of MUG protein suggests that the bidentate interaction with uracil is lost in MUG due to the occupation of this position by a Lys residue (Figure 1C). To test the role of K68 in determining the catalytic efficiency and substrate specificity, we converted it to Asn in MUG. The MUG-K68N mutant became an enzyme that can remove uracil from an A/U base pair (Figure 1D and E). In addition, the UDG activity on all double-stranded uracil-containing base pairs was enhanced, but no activity on single-stranded uracil-containing DNA was observed under the assay conditions (Figure 1D and E).

Based on this observation, we made 10 additional substitutions at the K68 position (Figure 2). Three hydrophobic amino acid substitutions, K68A, K68M and K68I, reduced the UDG activity, with the most hydrophobic substitution by Ile showing the greatest effect (Figure 2B–D). K68Q, K68D and K68E substitutions all increased the UDG activity on mismatched T/U, G/U and C/U base pairs (Figure 2F–H). It was reported that human UNG-N204D mutation allowed the enzyme to gain cytosine DNA glycosy-

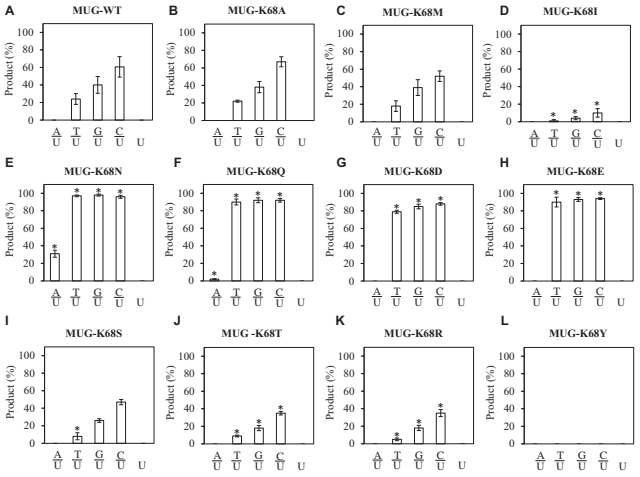

Figure 2. Analysis of DNA glycosylase activity of wt MUG and K68 mutants on uracil-containing substrates. Cleavage reactions were performed as described in the Materials and Methods section with 100 nM *E. coli* MUG protein and 10 nM substrate. Data are the averages of at least three independent experiments. (**A**) MUG-WT. (**B**) MUG-K68A. (**C**) MUG-K68M. (**D**) MUG-K68I. (**E**) MUG-K68N. (**F**) MUG-K68Q. (**G**) MUG-K68D. (**H**) MUG-K68E. (**I**) MUG-K68S. (**J**) MUG-K68T. (**K**) MUG-K68R. (**L**) MUG-K68Y. Statistical analysis was performed using T-test. $*P < 0.05$.

lase (CDG) activity (21). Under the assay conditions, we did not detect CDG activity in *E. coli* MUG-K68D. The K68Q mutant, which bears the closest chemical similarity to K68N, also showed weak activity on the A/U base pair and increased activity toward the mismatched base pairs as compared with the wild-type MUG enzyme (Figure 2F). The K68S, K68T and K68R mutants all reduced the UDG activity, and K68Y abolished the activity to below detection (Figure 2I–L). These results indicate that Asn, within the UDG structural context, is the most optimal choice for the UDG activity.

Binding affinity and catalytic efficiency of K68 mutants to G/U and A/U base pairs

To examine how the K68N substitution may alter the binding affinity toward uracil-containing DNA, we determined the K_D values of the WT and K68N enzymes by gel mobility shift analysis (Figure 3). For the G/U base pair, the K68N substitution reduced the K_D value from 292 to 146 nM, resulting in a 2-fold increase in affinity (Figure 3A and B and Table 1). For the A/U base pair, the wild-type MUG enzyme did not show any noticeable affinity (Figure 3C). However, the K68N substitution increased the binding affinity from not detectable to a K_D value of 566 nM (Figure 3D and Table 1).

To determine the effect of the substitution on catalysis, we measured the single turnover rate constants of the wild-type and the K68N mutants. For the G/U base pair, the K68N substitution improved the k_{st} value from 2.7 per min to 27.7 per min, a 10-fold increase (Table 1). The k_{st} value of the wild-type MUG obtained here for the G/U base pair is similar to that reported in a previous study (11). For the A/U base pair, the K68N mutant exhibited a single turnover rate

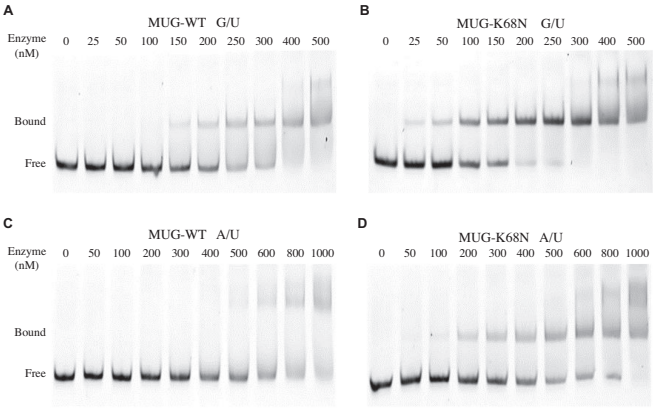

Figure 3. Gel mobility shift analysis of wt MUG and K68 mutants on uracil-containing substrates. Binding reactions were performed as described in the Materials and Methods section with 50 nM G/U base pair and 100 nM A/U base pair substrates and indicated amount of MUG protein. (**A**) MUG-WT with G/U base pair. (**B**) MUG-K68N with G/U base pair. (**C**) MUG-WT with A/U base pair. (**D**) MUG-K68N with A/U base pair. Data are the averages of three independent experiments.

Table 1. Binding affinity and rate constants of *E. coli* MUG wild type and K68N on G/U- or A/U-containing substrates[a]

		K_D (nM)	k_{ST} (min^{-1})
WT	G/U	292.3 ± 16.5	2.7 ± 0.2
K68N		146.0 ± 15.3	27.7 ± 1.9
WT	A/U	n.d.[b]	n.d.
K68N		565.6 ± 25.1	1.8 ± 0.06

[a]To determine the single turnover rate of the wild-type MUG and the K68N mutant enzymes with the G/U substrate, the reaction mixtures containing 5000 nM *E. coli* MUG enzyme, 50 nM G/U substrate and the remaining components as described above were incubated at 37°C. Samples were withdrawn at 0 s, 2 s, 5 s, 10 s, 20 s, 30 s, 1 min, 2 min and 5 min. To determine the single turnover rate of the K68N mutant enzyme with the A/U substrate, the reaction mixtures containing 10000 nM *E. coli* MUG-K68N enzyme, 50 nM A/U substrate and the remaining components as described above were incubated at 37°C. Samples were withdrawn at 0 s, 10 s, 20 s, 30 s, 1 min, 2 min, 3 min and 5 min. Based on the K_D values for the G/U or A/U substrates, the enzyme and the substrate reached at least 95% saturation in the single turnover measurements. The rate constants were determined by curve fitting using the integrated first-order rate equation (Deltagraph), $P = P_{max}$ (1-e^{-kt}), where P is the product yield, P_{max} is the maximal yield, t is time and k is rate constant. Data are the averages of three independent experiments.
[b]Not determined.

of 1.8 per min (Table 1). For comparison, the wild-type enzyme did not show any detectable UDG activity on the A/U base pair under the assay conditions.

In vivo repair of A/U base pair by K68N mutant

UDG may remove uracil bases in DNA that are generated from two sources. First, deamination of cytosine in a G/C base pair will convert cytosine to uracil, resulting in a G/U base pair. Second, dUMP may be misincorporated from dUTP in the nucleotide pool into DNA to become an A/U base pair, especially when the dUTP concentration is high. Given that K68N gained enzymatic activity on an A/U base pair, we would like to know whether the activity could remove uracil from misincorporation. Previously,

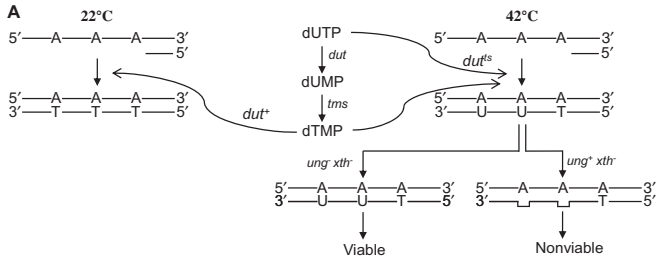

Figure 4. The survival of *E. coli* BW276 containing mug-wt and mug-K68N genes at 42 and 22°C. (**A**) Principle of the *in vivo* assay. *dut*: gene for dUTPase; tms: gene for thymidylate synthase; TS: temperature sensitive. (**B**) The survival of *E. coli* BW276 containing pBS-MUG-WT and pBS-MUG-K68N at 42 and 22°C. Cells with A_{600} value of 1 were diluted 1×10^6 times. Diluted cells (100 μl) were spread on the LB plates containing ampicillin and thymidine and incubated for 24 h at 42°C or 72 h at 22°C. (**C**) Quantitative analysis of survival. The relative plating efficiencies were calculated by the ratio of the cell numbers between 42 and 22°C. The data are the averages of six independent experiments.

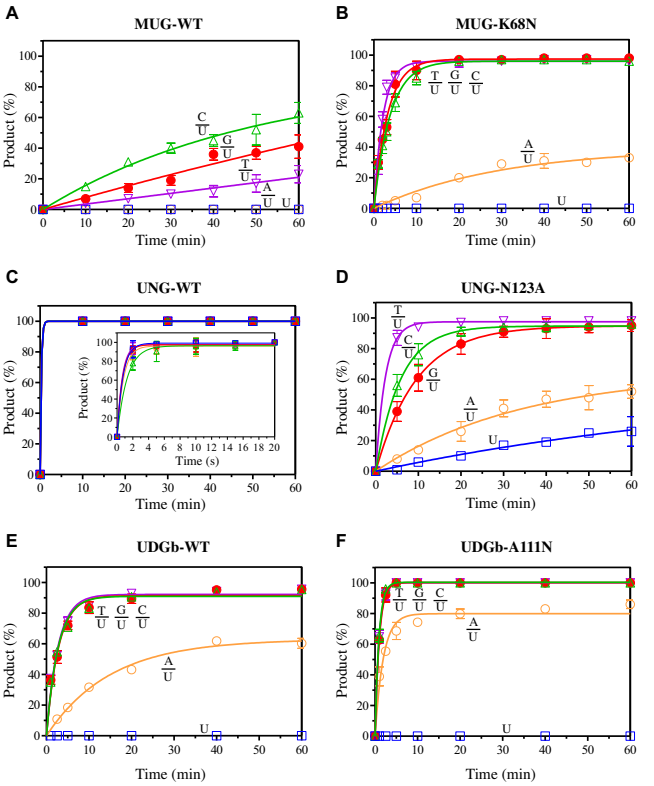

Figure 5. Effect of *E. coli* MUG-K68N, *E. coli* UNG-N123A and Tth UDGb-A111N substitutions on uracil DNA glycosylase activity. Cleavage reactions were performed as described in the Materials and Methods section with 100 nM *E. coli* MUG, *E. coli* UNG or Tth UDGb protein and 10 nM substrate. Time course analysis was conducted from 0 to 60 min. Data are the averages of three independent experiments. (△) C/U; (●) G/U; (○) A/U; (▽) T/U; (□) single-stranded U. (**A**) Time course analysis of uracil DNA glycosylase activity of *E. coli* MUG-WT on uracil-containing substrates. (**B**) Time course analysis of uracil DNA glycosylase activity of *E. coli* MUG-K68N mutant on uracil-containing substrates. (**C**) Time course analysis of uracil DNA glycosylase activity of *E. coli* UNG-WT on uracil-containing substrates. Inset: time course analysis between 0 and 20 s. (**D**) Time course analysis of uracil DNA glycosylase activity of *E. coli* UNG-N123A mutant on uracil-containing substrates. (**E**) Time course analysis of uracil DNA glycosylase activity of Tth UDGb-WT on uracil-containing substrates. (**F**) Time course analysis of uracil DNA glycosylase activity of Tth UDGb-A111N mutant on uracil-containing substrates.

we established an *E. coli*-based genetic system to test the repair *in vivo* (16). The basic premise of the system is that it contains *ung*⁻ *xth*⁻ (*xth* encodes the major AP endonuclease called exonuclease III) *dut*ts. Because the dUTPase gene (*dut*) is temperature sensitive, uracil will be incorporated into DNA at higher temperatures (Figure 4A). DNA polymerases can incorporate dUMP with similar efficiency as dTMP to form A/U base pairs (22). *E. coli* cells can tolerate a limited amount of uracil in its genomic DNA (23). On the other hand, the introduction of a UDG with activity on A/U base pairs would continuously remove uracil bases, resulting in AP sites, which would render the cells unviable (Figure 4A). When the MUG-K68N was introduced into the *E. coli* BW276 cells, its relative plating efficiency (a ratio of cell counts at 42°C versus 22°C) was reduced compared to the MUG-WT (Figure 4B and C). These results indicated that the UDG activity on A/U base pairs obtained through the K68N substitution enabled MUG to remove uracil resulting from polymerase misincorporation during DNA replication.

E. coli UNG-N123A and Tth UDGb-A111N mutants

It is clear that the K68N substitution greatly increased the UDG activity in *E. coli* MUG (Figure 5A and B). The wild-type family 1 *E. coli* UNG is a highly efficient enzyme. As shown in Figure 5C, the enzyme completed the reaction within seconds. In *E. coli* UNG, the homologous position to MUG-K68 is a highly conserved Asn residue (Figure 1B). To test the effect of losing the bidentate hydrogen bonds in family UNG, we changed UNG-N123 to Ala. The single amino acid substitution caused a dramatic loss of UDG activity. As indicated in Figure 5D, the N123A substitution reduced the activity to such a degree that the reactions with the A/U base pair and the single-stranded uracil substrate could not be completed even with 1-h incubation.

The scenario above represents a case in which the natural amino acid in this position is an Asn. Like *E. coli* MUG, another scenario is that the natural amino acid is not an Asn. The position equivalent to MUG-K68 is an Ala in family 5 Tth UDGb (Figure 1B). The wt UDGb enzyme already contained UDG activity on A/U base pairs (Figure 5E). By substituting A111 with Asn, the UDG activity was increased for C/U, G/U and T/U (Figure 5E and F). The UDG activity on A/U was also higher following the A111N substitution (Figure 5E and F). These results indicate that within the sequence and structural context of family 5 UDGb, the Asn substitution can also enhance the UDG activity especially for A/U base pairs.

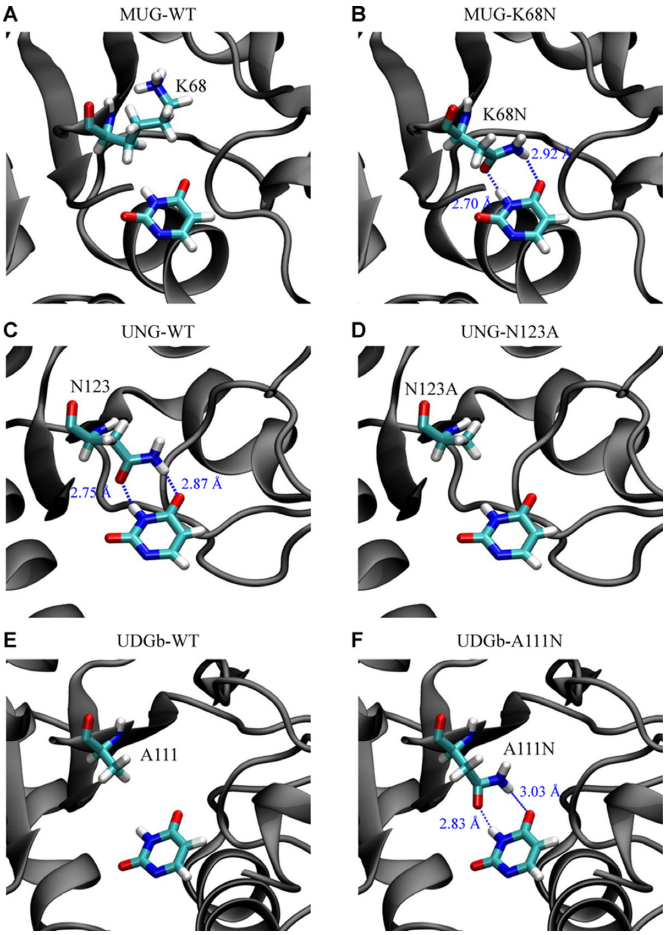

Figure 6. Modeling of interactions between glycosylase and uracil. The hydrogen bonds are shown in blue. (**A**) Interactions between MUG-WT and uracil based on a model built from the solved crystal structure (pdb 1MUG). K68 and uracil are colored by atom type. (**B**) Interactions between MUG-K68N and uracil. MUG-K68N structure was modeled from pdb MUG-K68N. K68N and uracil are colored by atom type. (**C**) Interactions between UNG-WT and uracil based on the solved crystal structure (pdb 2EUG). N123 and uracil are colored by atom type. (**D**) Interactions between UNG-N123A and uracil. UNG-N123A structure was modeled from pdb 2EUG. N123A and uracil are colored by atom type. (**E**) Interactions between Tth UDGb-WT and uracil based on the solved crystal structure (pdb 2DEM). A111 and uracil are colored by atom type. (**F**) Interactions between UDGb-A111N and uracil. UDGb-A111N structure was modeled from pdb 2DEM. A111N and uracil are colored by atom type.

Molecular modeling of MUG-uracil, UNG-uracil and UDGb–uracil interactions

To understand the structural effect of the K68N substitution on MUG, we modeled the Asn substitution to the MUG structure. While the K68 in WT MUG is pointing away from the uracil base, K68N allowed the amide side chain to rotate and form bidentate hydrogen bonds with uracil, as seen in *E. coli* UNG (Figure 6A and B and Supplementary Figure S2A and B). To understand how N123A may alter the interactions with uracil, we modeled UNG-N123A structure. In comparison with UNG-WT, UNG-N123A lost hydrogen bonds with the uracil base (Figure 6C and D and Supplementary Figure S2C and D). Likewise, Tth UDGb-A111N again formed bidentate hydrogen bonds with uracil (Figure 6E and F and Supplementary Figure S2E and F).

DISCUSSION

E. coli MUG differs from *E. coli* UNG in specificity and catalytic efficiency (1,2,11,12,24–26). Structural studies have defined how UNG specifically recognizes a uracil base in the binding pocket (27,28), in which N123 forms bidentate hydrogen bonds with N3 and O4 of uracil. Substitution of N123 with Asp in human and *E. coli* UNG causes a profound loss of uracil excision activity yet one of the human UNG mutants gains the ability to excise cytosine in DNA (21,29), suggesting that N123 in family 1 UNG is an important determinant for the UDG activity. Structural investigations reveal the lack of specific interactions resulting from the lysine (K68) at the equivalent position in *E. coli* MUG (8,9).

This study examined the role of position 68 in determining the UDG specificity and catalytic efficiency in *E. coli* MUG. Conversion of K68 to Asn enables the MUG enzyme to excise uracil from an A/U base pair (Figures 1 and 2). Among the 11 substitutions tested, only K68N and K68Q gained UDG activity on A/U base pairs, indicating the critical role of the bidentate hydrogen bonds in facilitating the excision of uracil from an A/U base pair. Even though K68D and K68E did not show any detectable UDG activity on the A/U base pair, the activity on the mismatched uracil-containing T/U, G/U and C/U base pairs was noticeably enhanced (Figure 2G and H). Likewise, K68N and K68Q not only allowed MUG to gain UDG activity on A/U base pairs, they also enabled MUG to act more efficiently on the three mismatched uracil-containing base pairs (Figure 2E and F). The fact that the WT MUG can only act on mismatched uracil-containing base pairs suggests the reliance of catalysis on spontaneous base flipping by the MUG enzyme. Our previous potentials of mean force analysis has already indicated that the mismatched T/U, G/U and C/U base pairs have a greater tendency to flip out of the DNA helix (12), thus allowing capture by the base recognition pocket in the MUG enzyme. Other studies also underscore the role of thermal stability of base pairs and duplex stability in substrate recognition (30–33). Evidently, the possibility to form bidentate hydrogen bonds offered by the K68N mutant greatly enhances the productive MUG-uracil binding, which leads to more efficient excision of uracil from all the base pairs (Figures 1, 6 and 7).

Similar to previous studies (29,34,35), we also find N123 in *E. coli* UNG to be profoundly important for UDG activity. Structurally, N123 forms bidentate hydrogen bonds with N3 and O4 of uracil (Figure 6). Previous analysis has emphasized the role N123 plays in bringing an induced fit conformational change to allow UNG to become catalytically competent (29,34–36). Based on the time course analysis presented in Figure 5, we estimated that the UDG activity on mismatched C/U, G/U and T/U base pairs was reduced by two orders of magnitude, and the activity on A/U and single-stranded uracil-containing DNA diminished by three to four orders of magnitude, respectively. These results indicate that the N123A mutation renders the family 1 *E. coli* UNG much like a family 2 MUG with respect to UDG

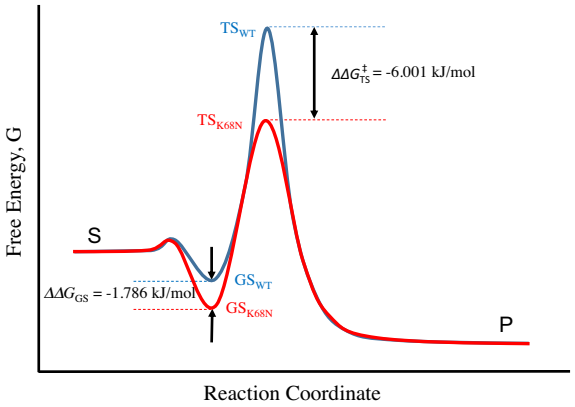

Figure 7. Proposed free energy profile in the course of the uracil DNA glycosylase reaction catalyzed by MUG-WT and MUG-K68N enzymes. GS: ground state. TS: transition state. Blue: MUG-WT. Red: MUG-K68N. $\Delta\Delta G_{GS} = -RT \ln (K_D^{K68N}/K_D^{WT})$. $\Delta\Delta G_{TS} = -RT \ln (k_{st}^{K68N}/k_{st}^{WT})$.

activity. Similar to MUG, the substitution of A111 in family 5 Tth UDGb with Asn increases all UDG activity on double-stranded uracil-containing DNA several-fold with the largest increase seen with the A/U base pair (Figure 5). These results indicate that this position is also an important determinant for the family 5 UDGb. The caveat is that family 5 Tth UDGb possesses UDG activity on A/U base pairs even in the wt enzyme in which position 111 is occupied by an Ala residue. Therefore, this position is a critical but not a sole determinant of the UDG activity on A/U base pairs.

The quantitative analysis of the binding affinity and the effect on catalysis offers insight into how the *E. coli* MUG enzyme becomes a UDG with much enhanced catalytic efficiency on uracil-containing DNA, in particular the A/U base pair. At the ground state, the K68N substitution apparently increased the binding affinities of the mutant toward either G/U and A/U base pairs (Figure 3 and Table 1). For the G/U base pair, the K68N substitution resulted in a 2-fold increase in the ground state binding (Table 1). To determine the effect of K68N substitution on the transition state, we compared the single turnover rate constants between the wild-type enzyme and the mutant. Remarkably, the single amino acid substitution changed the k_{st} by 10-fold for the G/U base pair (Table 1). The transition state catalytic rate analysis, combined with the ground state binding analysis, demonstrates that the K68N substitution greatly enhances the transition state interactions between the enzyme and the uracil, which results in a significant reduction in the $\Delta\Delta G$ at the transition state (Figure 7). Meanwhile, the formation of the bidentate hydrogen bonds facilitates the recognition and retention of the uracil base in the binding pocket to allow MUG to enter a catalytically competent state for glycosidic bond cleavage (Figure 7).

According to the RNA world theory, RNA played a dual role as a storage molecule for genetic material and a catalyst to accelerate biochemical reactions. During the transition from the RNA world to the modern DNA world, uracil was replaced by thymine and ribose was replaced by deoxyribose. The advantage of selecting thymine over uracil in DNA may be 2-fold. First, A/T base pairs confer increased stability to double helical DNA relative to A/U base pairs (37). Second, a prevailing view is that the replacement of uracil with thymine is needed in order to distinguish endogenous uracil in A/U base pairs from uracil generated from cytosine deamination in G/U base pairs (38). This model implies that the prototypic enzyme involved in the repair of cytosine deamination must have possessed activity toward both A/U and G/U base pairs. If endogenous uracil in A/U base pairs is not marked by adding a methyl group to become thymine, it will be subject to constant removal by a UDG, thus, creating a futile cycle (39). It is suggested that the replacement of uracil by thymine is due to the nature of a prototypic UDG (coined as a leaky MUG) that preferentially excised uracil from G/U base pairs yet occasionally removed uracil from A/U base pairs (39). The occasional removal of uracil from A/U base pairs, although tolerated, provides a selection force to replace endogenous uracil with thymine. Data presented here suggest that a leaky MUG can be generated through a small number of mutations starting from a family 1 UNG-like enzyme or a family 2 MUG-like enzyme. Within the sequence and structural context of *E. coli* MUG, a single amino acid substitution at the K68 position can create a leaky MUG that maintains a low-level activity toward endogenous A/U base pairs as demonstrated by the *in vitro* and *in vivo* analysis. Likewise, an ancestral family 1 UNG-like enzyme may lack N123 residue, thereby acting similarly as a leaky MUG-type UDG. Understanding the evolutionary history of UDG superfamily will reveal valuable insight regarding this hypothesis.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We thank members of Cao laboratory for assistance and discussions.

FUNDING

South Carolina Experiment Station [SC-1700274, technical contribution No. 6289]; Department of Defense [W81XWH-10-1-0385]; National Institutes of Health [GM090141]; National Science Foundation Career Award [MCB-0953783 to B.N.D.]. The open access publication charge for this paper has been waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

1. Pearl,L.H. (2000) Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res.*, **460**, 165–181.
2. Parikh,S.S., Putnam,C.D. and Tainer,J.A. (2000) Lessons learned from structural results on uracil-DNA glycosylase. *Mutat. Res.*, **460**, 183–199.
3. Lee,H.W., Dominy,B.N. and Cao,W. (2011) New family of deamination repair enzymes in uracil-DNA glycosylase superfamily. *J. Biol. Chem.*, **286**, 31282–31287.
4. Cortazar,D., Kunz,C., Saito,Y., Steinacher,R. and Schar,P. (2007) The enigmatic thymine DNA glycosylase. *DNA Rep.*, **6**, 489–504.

5. Krokan,H.E., Drablos,F. and Slupphaug,G. (2002) Uracil in DNA–occurrence, consequences and repair. *Oncogene*, **21**, 8935–8948.
6. Sartori,A.A., Fitz-Gibbon,S., Yang,H., Miller,J.H. and Jiricny,J. (2002) A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. *EMBO J.*, **21**, 3182–3191.
7. Lucas-Lledo,J.I., Maddamsetti,R. and Lynch,M. (2011) Phylogenomic analysis of the uracil-DNA glycosylase superfamily. *Mol. Biol. Evol.*, **28**, 1307–1317.
8. Barrett,T.E., Savva,R., Panayotou,G., Barlow,T., Brown,T., Jiricny,J. and Pearl,L.H. (1998) Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell*, **92**, 117–129.
9. Barrett,T.E., Scharer,O.D., Savva,R., Brown,T., Jiricny,J., Verdine,G.L. and Pearl,L.H. (1999) Crystal structure of a thwarted mismatch glycosylase DNA repair complex. *EMBO J.*, **18**, 6599–6609.
10. Gallinari,P. and Jiricny,J. (1996) A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature*, **383**, 735–738.
11. O'Neill,R.J., Vorob'eva,O.V., Shahbakhti,H., Zmuda,E., Bhagwat,A.S. and Baldwin,G.S. (2003) Mismatch uracil glycosylase from Escherichia coli: a general mismatch or a specific DNA glycosylase? *J. Biol. Chem.*, **278**, 20526–20532.
12. Lee,H.W., Brice,A.R., Wright,C.B., Dominy,B.N. and Cao,W. (2010) Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase. *J. Biol. Chem.*, **285**, 41483–41490.
13. Haushalter,K.A., Todd Stukenberg,M.W., Kirschner,M.W. and Verdine,G.L. (1999) Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Curr. Biol.*, **9**, 174–185.
14. Mi,R., Dong,L., Kaulgud,T., Hackett,K.W., Dominy,B.N. and Cao,W. (2009) Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG. *J. Mol. Biol.*, **385**, 761–778.
15. Haas,B.J., Sandigursky,M., Tainer,J.A., Franklin,W.A. and Cunningham,R.P. (1999) Purification and characterization of Thermotoga maritima endonuclease IV, a thermostable apurinic/apyrimidinic endonuclease and 3′-repair diesterase. *J. Bacteriol.*, **181**, 2834–2839.
16. Xia,B., Liu,Y., Li,W., Brice,A.R., Dominy,B.N. and Cao,W. (2014) Specificity and catalytic mechanism in family 5 uracil DNA glycosylase. *J. Biol. Chem.*, **289**, 18413–18426.
17. Stivers,J.T. and Jiang,Y.L. (2003) A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.*, **103**, 2729–2759.
18. Ho,S.N., Hunt,H.D., Horton,R.M., Pullen,J.K. and Pease,L.R. (1989) Site-directed mutagenesis by overlap extension using the polymease chain reaction. *Gene*, **77**, 51–59.
19. Grippon,S., Zhao,Q., Robinson,T., Marshall,J.J., O'Neill,R.J., Manning,H., Kennedy,G., Dunsby,C., Neil,M., Halford,S.E. *et al.* (2011) Differential modes of DNA binding by mismatch uracil DNA glycosylase from Escherichia coli: implications for abasic lesion processing and enzyme communication in the base excision repair pathway. *Nucleic Acids Res.*, **39**, 2593–2603.
20. Guex,N. and Peitsch,M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
21. Kavli,B., Slupphaug,G., Mol,C.D., Arvai,A.S., Peterson,S.B., Tainer,J.A. and Krokan,H.E. (1996) Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *EMBO J.*, **15**, 3442–3447.
22. Patel,P.H. and Loeb,L.A. (2000) Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J. Biol. Chem.*, **275**, 40266–40272.
23. Warner,H.R., Duncan,B.K., Garrett,C. and Neuhard,J. (1981) Synthesis and metabolism of uracil-containing deoxyribonucleic acid in Escherichia coli. *J. Bacteriol.*, **145**, 687–695.
24. Hang,B., Downing,G., Guliaev,A.B. and Singer,B. (2002) Novel activity of Escherichia coli mismatch uracil-DNA glycosylase (Mug) excising 8-(hydroxymethyl)-3,N4-ethenocytosine, a potential product resulting from glycidaldehyde reaction. *Biochemistry*, **41**, 2158–2165.
25. Saparbaev,M., Langouet,S., Privezentzev,C.V., Guengerich,F.P., Cai,H., Elder,R.H. and Laval,J. (2002) 1,N(2)-ethenoguanine, a mutagenic DNA adduct, is a primary substrate of Escherichia coli mismatch-specific uracil-DNA glycosylase and human alkylpurine-DNA-N-glycosylase. *J. Biol. Chem.*, **277**, 26987–26993.
26. Saparbaev,M. and Laval,J. (1998) 3,N4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for Escherichia coli double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 8508–8513.
27. Savva,R., McAuley-Hecht,K., Brown,T. and Pearl,L. (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature*, **373**, 487–493.
28. Slupphaug,G., Mol,C.D., Kavli,B., Arvai,A.S., Krokan,H.E. and Tainer,J.A. (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature*, **384**, 87–92.
29. Handa,P., Acharya,N. and Varshney,U. (2002) Effects of mutations at tyrosine 66 and asparagine 123 in the active site pocket of Escherichia coli uracil DNA glycosylase on uracil excision from synthetic DNA oligomers: evidence for the occurrence of long-range interactions between the enzyme and substrate. *Nucleic Acids Res.*, **30**, 3086–3095.
30. Liu,P., Theruvathu,J.A., Darwanto,A., Lao,V.V., Pascal,T., Goddard,W. III and Sowers,L.C. (2008) Mechanisms of base selection by the Escherichia coli mispaired uracil glycosylase. *J. Biol. Chem.*, **283**, 8829–8836.
31. Liu,P., Burdzy,A. and Sowers,L.C. (2002) Substrate recognition by a family of uracil-DNA glycosylases: UNG, MUG, and TDG. *Chem. Res. Toxicol.*, **15**, 1001–1009.
32. Valinluck,V., Liu,P., Burdzy,A., Ryu,J. and Sowers,L.C. (2002) Influence of local duplex stability and N6-methyladenine on uracil recognition by mismatch-specific uracil-DNA glycosylase (Mug). *Chem. Res. Toxicol.*, **15**, 1595–1601.
33. Yang,W. (2006) Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA Rep.*, **5**, 654–666.
34. Jiang,Y.L. and Stivers,J.T. (2002) Mutational analysis of the base-flipping mechanism of uracil DNA glycosylase. *Biochemistry*, **41**, 11236–11247.
35. Kwon,K., Jiang,Y.L. and Stivers,J.T. (2003) Rational engineering of a DNA glycosylase specific for an unnatural cytosine:pyrene base pair. *Chem. Biol.*, **10**, 351–359.
36. Berti,P.J. and McCann,J.A. (2006) Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer. *Chem. Rev.*, **106**, 506–555.
37. Howard,F.B. (2005) The stabilizing contribution of thymine in duplexes of (dA)24 with (dU)24, (dT)24, (dU12-dT12), (dU-dT)12, (dU2-dT2)6, or (dU3-dT3)4: nearest neighbor and next-nearest neighbor effects. *Biopolymers*, **78**, 221–229.
38. Pearl,L.H. and Savva,R. (1996) The problem with pyrimidines. *Nat. Struct. Biol.*, **3**, 485–487.
39. Poole,A., Penny,D. and Sjoberg,B.M. (2001) Confounded cytosine! Tinkering and the evolution of DNA. *Nat. Rev. Mol. Cell Biol.*, **2**, 147–151.