HHS Public Access

Author manuscript

DNA Repair (Amst). Author manuscript; available in PMC 2018 September 01.

Published in final edited form as:

DNA Repair (Amst). 2017 September ; 57: 107–115. doi:10.1016/j.dnarep.2017.07.001.

Identification of a prototypical single-stranded uracil DNA glycosylase from *Listeria innocua*

Jing Li[1], **Ye Yang**[1], **Jose Guevara**, **Liangjiang Wang**, and **Weiguo Cao**[*]

Department of Genetics and Biochemistry, Clemson University, Room 060 Life Sciences Facility, 190 Collings Street, Clemson, SC 29634, USA

Abstract

A recent phylogenetic study on UDG superfamily estimated a new clade of family 3 enzymes (SMUG1-like), which shares a lower homology with canonic SMUG1 enzymes. The enzymatic properties of the newly found putative DNA glycosylase are unknown. To test the potential UDG activity and evaluate phylogenetic classification, we isolated one SMUG1-like glycosylase representative from *Listeria innocua* (Lin). A biochemical screening of DNA glycosylase activity *in vitro* indicates that Lin SMUG1-like glycosylase is a single-strand selective uracil DNA glycosylase. The UDG activity on DNA bubble structures provides clue to its physiological significance *in vivo*. Mutagenesis and molecular modeling analyses reveal that Lin SMUG1-like glycosylase has similar functional motifs with SMUG1 enzymes; however, it contains a distinct catalytic doublet S67-S68 in motif 1 that is not found in any families in the UDG superfamily. Experimental investigation shows that the S67M-S68N double mutant is catalytically more active than either S67M or S68N single mutant. Coupled with mutual information analysis, the results indicate a high degree of correlation in the evolution of SMUG1-like enzymes. This study underscores the functional and catalytic diversity in the evolution of enzymes in UDG superfamily.

Keywords

DNA repair; deamination; uracil; single-stranded DNA; mutual information

1. Introduction

As a non-canonical base of DNA, uracil can occur in DNA by misincorporation of dUMP during DNA replication [1, 2] or spontaneous or enzymatic deamination of cytosine [3–5]. Uracil DNA glycosylase (UDG) is the principal enzyme to remove uracil in DNA by hydrolyzing the N-glycosidic bond and initiating base excision repair (BER) pathway [6, 7]. Based on the sequence homology and structural fold, six families were identified. Family 1

[*]Corresponding Author: wgc@clemson.edu; Tel.: (864) 656-4176; Fax: (864) 656-6879.

[1]Contributed equally

Conflict of Interest

The authors declare that there are no conflicts of interest.

UDGs, also known as UNGs, are ubiquitous in bacteria and most eukaryotes. The enzymes exclusively cleave uracil in both double-stranded (ds) and single-stranded (ss) DNA at a very efficient rate [2, 8]. Family 2 UDGs are represented by human thymine DNA glycosylase (hTDG) [9] and *E. coli* mismatch-specific uracil DNA glycosylase (*E. coli* MUG) [10]. They are identified as repair enzymes acting on uracil base in a mispair, xanthine-containing DNA and other modified bases [10–15]. The discovery of human TDG as a 5-formylcytosine (fC) and 5-caboxylcytosine (5-caC) DNA glycosylase places it as an essential enzyme in DNA demethylation [16, 17]. Family 3 enzymes are named as single-strand-selective monofunctional UDG (SMUG1) and are presented in vertebrate, insects and some eubacteria [18, 19]. Family 4 UDGs are a group of prokaryotic thermostable enzymes which excise uracil from DNA strand(s) [20]. Family 5 enzymes are also thermostable and present in archaea and eubacteria, with versatile substrates specificities [21]. Family 6 enzymes (HDG) recognize and excise hypoxanthine in DNA strand(s), exclusively [22].

Firstly isolated from *Xenopus laevis* and human, SMUG1 enzymes show few sequence homologies to other UDG families, but have similar gross structural fold [18]. Later, a genome database searching identified SMUG1 orthologs in a few eubacteria lacking family 1 UNG [19, 23]. In comparison to the robust family 1 UNG, SMUG1 enzymes are less active to uracil-containing DNA, but they are versatile with broader substrate specificity, in which they include uracil, xanthine, 5-formyluracil, 5-hydroxyuracil, 5-hydroxymethyluracil and 3,N^4-ethenocytosine as substrate [19, 24–27]. Depending on reaction conditions, SMUG1 presents different substrate preferences. Notably, in the presence of Mg^{2+} under physiological conditions, SMUG1 switches into double-stranded-selective UDG [28]. Through extensive biochemical and genetic analysis, SMUG1 was identified as the principal 5-formyluracil and 5-hydroxylmethyluracil DNA glycoyslase in mammalian systems [24, 26, 27, 29]. Later on, we reported that bacterial SMUG1, as well as human SMUG1, was not only a UDG, but also a xanthine DNA glycosylase [19]. The structures of the bacterial *Geobacter metallireducens* SMUG1 have been resolved [30].

Through a large-scale phylogenetic analysis of UDG superfamily in more than 1,000 completely sequenced genomes, it is shown that family 3 enzymes can be separated into two clades with the traditional family 3 SMUG1 as one of them (Fig. 1) [31]. UDG enzymes in another clade are present in eubacteria genus including *listeria*, *lactobacillus*, *Streptomyces*, *Amycolatopsis* and *Flavobacteriaceae*. This group of UDGs, which we named as SMUG1-like, is more similar to family 3 SMUG1s and shares a common ancestor with family 1 UNGs (Fig. 1). As identified previously [31], a distinct difference between SMUG1-like and traditional SMUG1 is that the “GMNPGP” in motif 1 of SMUG1 is changed to “GSSPAR” in SMUG1-like enzymes (Fig. 2A). To our knowledge, the biochemical and enzymatic properties of SMUG1-like enzymes are completely unknown. In this study, we characterized a SMUG1-like DNA glycosylase from *Listeria innocua* (Lin). To our surprise, Lin SMUG1-like glycosylase is a single-stranded UDG with little activity on double-stranded uracil-containing DNA. Mutational analysis indicates similar roles of active site residues corresponding to SMUG1. Interestingly, a double substitution of S67-S68 in Lin SMUG1-like DNA glycosylase by M67-N68 in motif 1 to mimic SMUG1 is able to partially rescue the negative effects of the S67M and S68N single mutants, which implies correlation and co-

evolution of the two catalytic residues. The potential physiological role of the SMUG1-like is also discussed.

2. Material and Methods

2.1. Plasmid construction, cloning, expression and purification

The *L. innocua* SMUG1-like gene (GenBank accession number: WP_010991469.1) was amplified by PCR using the forward primer Lin_SMUG1_F (5′-GGG AAT TCC ATA TGG CTA GCA TGA CTG GTG-3′; the NdeI site is underlined) and the reverse primer Lin_SMUG1_R (5′-CCG CTC GAG CCT CTT TAA AGC ACA-3′; the XhoI site is underlined). The PCR reaction mixture (25 μl) consisted of 8 ng of *L. innocua* genomic DNA, 200 nM forward primer Lin_SMUG1_F and reverse primer Lin_SMUG1_R, 1 × Phusion PCR buffer (New England Biolabs), 200 μM each dNTP, and 1 unit of Phusion DNA polymerase. The PCR procedure included a predenaturation step at 98°C for 5 min, 30 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 15 sec, annealing at 50°C for 15 sec and extension at 72°C for 1 min, and a final extension step at 72°C for 10 min. The PCR product was purified by Gel DNA Recovery Kit (Zymo Research). Purified PCR product and plasmid pET21a (+) were digested with NdeI and XhoI, purified by Gel DNA Recovery Kit and ligated according to the manufacturer's instructional manual. The ligation mixture was transformed into *E. coli* strain DH5α competent cells prepared by electroporation. The sequence of the *L. innocua* SMUG1-like gene in the resulting plasmid (pET21a (+)-Lin-SMUG1) was confirmed by DNA sequencing.

The resulting plasmid with wild-type SMUG1-like was used as the template plasmid for all other SMUG1-like mutants. Amplification of mutant DNA and DpnI mediated site-directed mutagenesis procedures were carried out as previously described with modification by using primers carrying the desired mutations [32, 33]. To express the C-terminal His-6-tagged wild-type and mutant Lin SMUG1-like glycosylase gene, the recombinant plasmids were transformed into *E. coli* strain BL21 (DE3 Δ*slyD* Δ*mug* Δ*udg* Δ*nfi* Δ*nth* Δ*ndk*) by electroporation. Protein expression and purification were carried out as previously described [22].

2.2. DNA glycosylase activity assay

The sequence of the oligonucleotides used for DNA glycosylase activity assay is shown in Fig. 2B, and prepared as previously described [19]. DNA glycosylase cleavage assays for Lin SMUG1-like proteins were performed at 42°C for 2 h in a 10 μl reaction mixture containing 10 nM oligonucleotide substrate, 20 μM glycosylase protein unless noted otherwise, 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 5 mM EDTA and 1 mM DTT. The resulting abasic sites were cleaved by incubation at 95°C for 5 min after adding 1 μl of 1 N NaOH. To quantify cleavage products and remaining substrates, the reaction mixtures were analyzed by Applied Biosystems 3130xl sequencer with a fragment analysis module. Cleavage products and remaining substrates were quantified by GeneMapper software.

2.3. Enzyme kinetic analysis

Enzyme kinetic analysis was performed with Lin SMUG1-like glycosylase by incubating different concentrations of respective proteins (1 μM to 100 μM) with 10 nM substrate. Due to the low catalytic efficiency of Lin SMUG1-like, we determined the kinetic parameters under the conditions that enzyme concentration was in excess of substrate concentration as previously described [34]. The linear increase of k_{obs} over enzyme concentrations precluded acquisition of individual k_2 or K_m, but allowed for determination of k_2/K_m (Fig. S1) [34]. The k_2 / K_m values were obtained by ploting k_{obs} against total enzyme concentration by linear regression with equation $k_{obs}=\frac{k_2[E]}{K_m}$.

2.4. Phylogenetic analysis

A total of 28 glycosylase protein sequences were retrieved from Genbank and aligned using the multiple sequence alignment program ClustalX2 [35] and a structure-based alignment program PROMALS3D [36]. Subsequently, the resulting alignment was curated manually to align the catalytic motifs. The phylogenetic tree was generated using the neighbor-joining method within the MEGA v6.0 software package [37] to infer the evolutionary history between different proteins.

2.5. Spontaneous mutation frequency assay

E. coli ung and Lin SMUG1-like glycosylase genes (WT and mutant) were amplified by PCR with added KpnI and HindIII sites. The amplicons and pBluescript-II KS(+) were digested with KpnI and HindIII and ligated. Ligation mixtures were transformed into *E. coli* DH5α to produce constructs pBluescript-II KS (+)-UNG (encoding *E. coli* UNG glycosylase), pBluescript-II KS (+)-WT (encoding Lin SMUG1-like glycosylase) and pBluescript-II KS (+)-S67M (encoding Lin SMUG1-like glycosylase S67M mutant). The sequences of the inserts were confirmed by DNA sequencing. *E. coli* strain MG1655 Δ3 (Δung Δmug Δnth) was transformed with the IPTG inducible constructs pBluescript-II KS (+)-UNG, pBluescript-II KS (+)-WT, pBluescript-II KS (+)-S67M (as control) and pBluescript-II KS (+) (as control), respectively. Single colonies were selected and inoculated into 4 ml liquid LB cultures supplemented with 50 μg/ml ampicillin. After overnight incubation at 30°C, IPTG was added to a final concentration of 0.5 mM. The cultures were incubated at 37 °C for an additional 5 h. For cell counting, the cell cultures were then diluted to 1×10^8 cells per ml and mixed with 3 ml of 0.7% soft agar and then plated on LB plates with both 50 μg/ml ampicillin and 100 μg/ml rifampicin. The plates were incubated at 37°C for 48 h and colony numbers on amp^R and rif^R plates were counted, respectively. The mutation frequency was calculated as the results of rif^R colony number per 10^9 amp^R colony number.

Mutation rates to rifampicin resistance (rif^R) were also estimated using fluctuation tests as described [38]. Strains used in spontaneous mutation frequency assay as described above were grown in LB broth supplemented with 50 ug/ml ampicillin, 50 ug/ml kanamycin and 1 mM IPTG. Mutation rates from fluctuation tests were calculated using the Ma-Sandri-Sarkar maximum likelihood method implemented in the FALCOR web tool (www.mitochondria.org/protocols/FALCOR.html) [39, 40].

2.6. Homology modeling

Position-Specific Iterated BLAST (PSI-BLAST) was used to generate the alignment between the amino acid sequence from chain A of the *Xenopus laevis* SMUG1 structure (PDB 1OE5) and SMUG1-like protein sequence from *Listeria innocua* (GenBank accession WP_010991469.1), which showed 19% identity and 31% similarity. The resulting sequence alignment and the *Xenopus laevis* SMUG1 structure were input to the MODELLER 9.15 software package [41]. A homology model was constructed for Lin SMUG1-like protein with uracil base. Using the CHARMM (ver. c36b1) package, the complex was refined by harmonic restraints with a force constant of 10 kcal/mol/Å^2 to fix coordinates of amino acids within 10 Å of the uracil base, following by Newton-Raphson minimization of 200 steps to remove van der Waals clashes around uracil. After removing Harmonic restrains, a generalized Born implicit solvent was carried out, and the complex was refined for a further 1000 steps. The coordinates of uracil base were refined using the Autodock 4.6.2 package.

2.7. Mutual information (MI) analysis

The mutual information (MI) analysis was performed as previously described with modifications [34]. Briefly, to encompass the six known families of the UDG superfamily in the analysis, we compiled sequences of the families 2, 3 5 and 6 and added them to the existing set of families 1 and 4 sequences [34]. Subsequently, a multiple sequence alignment was constructed using MEGA6 [37]. The alignment was used as the input for the mutual information analysis using the MISTIC web server [42]. The reference sequence selected was the SMUG1-like glycosylase from *Listeria innocua* (GenBank accession WP_010991469.1). The parameters were set to those used previously [34].

3. Results and Discussion

3.1. SMUG1-like enzyme as UDG

Listeria innocua genome contains a traditional family 1 UNG (GenBank accession number EHN61868.1) and a SMUG1-like glycosylase and two AP endonucleases (GenBank accession numbers WP_003769869.1 and KJR54533.1), indicating that it possesses a complete base excision repair pathway. The SMUG1-like group as represented by the enzyme in *L. innocua* shows high degree of sequence homology to SMUG1 and possesses similar motifs except for noted difference in motif 1 (Figs. 1–2). To determine its potential DNA repair function, we assayed the DNA glycosylase activity using the fluorescently labeled oligonucleotide substrates (Fig. 2B). The recombinant protein was cloned, expressed in *E. coli* and purified as shown in Fig. 2C. Among the five damaged bases tested, the Lin SMUG1-like enzyme only showed DNA glycosylase activity toward uracil-containing DNA substrate under the conditions that the enzyme was in excess over the substrate (Fig. 2D). No enzymatic activity was detected with double-stranded and single-stranded hypoxanthine-, xanthine-, 5-hydroxymethyluracil (hmU)-, 5-hydroxyuracil (OHU)-, 8-oxoguanine-, 8-oxoadenine-, 5,6-dihydroxyuracil (DHU)-, 5-hydroxycytosine-, thymine glycol-, N6-methyladenine-, O6-methylguanine-, and ethenoadenine-containing DNA (Fig. 2D and data not shown). Lin SMUG1-like enzyme did not show binding affinity to these substrates either as measured by gel mobility shift analysis (data not shown). Gme SMUG1 is still active with uracil-containing substrates with 1 nM enzyme and 10 nM substrate [19]. However, Lin

SMUG1-like glycosylase required excess enzyme to show detectable activity on uracil-containing substrates, suggesting that the overall enzymatic activity was not as robust as Gme SMUG1 [19]. Unexpectedly, Lin SMUG1-like glycosylase was more active with single-stranded uracil-containing DNA and showed only low-level activity with double-stranded uracil-containing DNA (Fig. 2D).

To examine whether the strong preference for single-stranded U-containing DNA was simply the outcome of assay conditions, we examined UDG activity under different conditions. The enzyme showed optimal ss UDG activity at pH 7.5; however, the ds UDG activity remained very low across all pH values tested (Fig. 3A). Likewise, salt titration with KCl did not change its ss U preference (Fig. 3B). Lin SMUG1-like glycosylase had highest activity at 42°C with ss U but still with minimal activity on ds U at all temperatures examined (Fig. 3C). The glycoslylase activity on ss U was enhanced with Mg^{2+} but suppressed by transition metal ions Co^{2+}, Fe^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+} and Cd^{2+} (Fig. 3D). These results indicate that the strong preference for ss U is an intrinsic property of the enzyme rather than a reflection of assay conditions. We tested the affinity of this glycosylase to single-stranded uracil-containing DNA, but it did not show retarded bank in gel mobility shift analysis (data not shown). Because the overall enzymatic activity of Lin SMUG1-like glycosylase was low, we measured k_2/K_m under the condition that the enzyme concentration was in excess (Fig. S1). This assay method was used previously to study family 4 UDGa [34]. The k_2/K_m value for ss U was 3.5×10^{-6} min^{-1} nM^{-1}, confirming that the overall kinetic efficiency of the enzyme was low (Table 1).

To further characterize the strong preference to ss U, we tested the UDG activity toward different bubble structures. A bubble structure is a double-stranded DNA containing a unpaired segment (Fig. 4). We designed 7-nucleotide, 13-nucleotide and 19-nucleotide bubbles (Fig. 4A–B). A variant of the 19-nt bubble was also included by placing the uracil base 3-nt from the nearest double strand junction (Fig. 4B, 19-nt shift Bubbled DNA). The UDG activity followed the order of ss U > 19-nt bubble > 13-nt bubble > 7-nt bubble > 19-nt shift bubble (Fig. 4C). The k_2/K_m value of the 19-nt bubble is was 2-fold lower than that of the ss U substrate (Table 1). These results revealed that the closer to the double strand junction the uracil is, the lower the UDG activity. Hence, Lin SMUG-like glycosylase is adapted to act on ss U-containing DNA.

3.2. In vivo analysis

To test whether the UDG activity could play role *in vivo*, we initially measured mutation frequencies using the rifampicin-based assay [43, 44]. *E. coli* family 1 UNG is well known for its extremely robust UDG activity. As expected, the presence of *E. coli* UNG reduced the mutation frequency to 32 Rif^R colonies per 10^9 viable cells (Table 2). In the presence of Lin SMUG1-like glycosylase, the mutation frequency reached 464 Rif^R colonies per 10^9 viable cells (Table 2), which was about 15-fold higher than in the presence of *E. coli* UNG but 2-fold lower than its absence (Table 2, compare pBS-Lin-WT with pBS-Lin-S67M and pBS). To more precisely measure the mutation rates per cell division, we performed fluctuation tests as previously described [38]. The presence of *E. coli* UNG reduced the mutation frequency to 24 Rif^R colonies per 10^9 viable cells (Table 2). In the presence of Lin SMUG1-

like glycosylase, the mutation frequency reached 146 Rif^R colonies per 10^9 viable cells (Table 2), which was about 6-fold higher than in the presence of *E. coli* UNG but about 2-fold lower than its absence (Table 2, compare pBS-Lin-WT with pBS-Lin-S67M and pBS). These results suggest that Lin SMUG1-like glycosylase can reduce mutation *in vivo* although not as efficiently as family 1 *E. coli* UNG. The single-stranded UDG activity may help repair of base damage occurred during DNA replication and transcription, as previously suggested [45].

3.3. Site-directed mutagenesis of Lin SMUG1-like glycosylase in motifs 1, 2 and 3

UDG enzymes have three conserved motifs that are important for their DNA glycosylase activity (Fig. 2A). As shown in the multiple sequence alignment, SMUG1-like enzymes share some sequence conservation in the three motifs, in particular F81 in motif 1, N128A in motif 3 and H210 in motif 2 (Fig. 2A). In addition, as described above, a unique feature of the SMUG1-like enzymes is the substitution of GMNP with GSSP in motif 1. To investigate the role of these residues in Lin SMUG1-like enzyme, we substituted F81, N128 and H210 with alanine, and S67 and S68 with several residues (Fig. 2A). Alanine substitution at F81, N128 and H210A positions rendered the enzyme much less active compared with the WT enzyme (Table S1). Similar effects of these substitutions in SMUG1 enzymes were reported previously [23]. To understand the mutational effects, we modeled the structure of Lin SMUG1-like glycosylase using Gme SMUG1 as a template. Molecular modeling indicates that these three residues interact with uracil in a way similar to what have been observed in family 3 SMUG1 [46, 47]. The highly conserved F81 in motif 1 stacks with uracil to stabilize the flipped out base (Fig. 5A). Likewise, F109 in Gme SMUG1 and F98 in human SMUG1 play a similar role (Fig. 5B) [46, 47]. Situated in the same location as N174 in Gme SMUG1 and N163 in Human SMUG1 [46, 47], N128 in Lin SMUG1-like glycosylase could interact with N3 and O4 of the uracil base (Fig. 5A–B). The first histidine residue in motif 2 is proposed to form a hydrogen bond with O2 of uracil to stabilize the negatively charged uracil leaving group in family 1 UNG, family 4 UDGa and family 5 UDGb [21, 34, 48]. It is likely that H210 plays a similar catalytic role in Lin SMUG1-like glycosylase (Fig. 5A–B).

In Lin SMUG1-like glycosylase, S67 is located in the similar positions as M95 in Gme SMUG1 (Figs. 2A and 6C–D). The mainchain amino group of M95 or S67 interacts with the O2 of uracil (Fig. 5C–D). Substitution of S67 with Ala, Ile and Met all resulted in loss of glycosylase activity (Table S1). Even though the equivalent position in family 2 *E. coli* MUG and family 3 Gme SMUG1 is Ile and Met, respectively, the S67I and S67M mutants were inactive. These results suggest that although Ile and Met residues are compatible in family 2 and family 3 enzymes, they are not in SMUG1-like enzymes.

In Lin SMUG1-like glycosylase, S68 is located in the similar position as N96 in Gme SMUG1 (Figs. 2A and 6C–D). In *E. coli* MUG, the mainchain amino group and the sidechain amide group of N18 located in the equivalent position are proposed to position a water molecule for catalysis [11, 49]. Likewise, N96 in Gme SMUG1 and S68 in Lin SMUG1-like glycosylase could play a similar catalytic role by using the mainchain and sidechaine to position the catalytic water (Fig. 5C–D). Substitution of S67 with Ala resulted in loss of glycosylase activity (Table S1). However, substitution of S68 with Asn, Asp, Glu

and Gln retained UDG activity to various degrees (Table S1). These results may indicate the need of functional sidechain to position the water molecule (Fig. 5C). Overall, Lin SMUG1-like glycosylase adopts a similar catalytic mechanism as SMUG1, in which H210 is responsible for promoting the departure of the uracil leaving group and S68 is needed to position a water molecule to attack the N-glycosidic bond (Fig. 6E).

To test whether S67 and S68 are correlated, we constructed an S67M-S68N double mutation. S67M-S68N mutant turned out to be more active than the S67M or S68N single mutant with single-stranded uracil-containing DNA (Table S1), suggesting that these two positions are correlated. No detectable glycosylase activity was found with double-stranded uracil-containing DNA. To more quantitatively compare the mutational effects of the single mutants, double mutant with the WT Lin SMUG1-like glycosylase, we determined the enzyme kinetics parameters. S68N single mutation reduced the k_2/K_m value by two orders of magnitude (Table 1). Even though S67M-S68N double mutation reduced the k_2/K_m value by 25-fold as compared with the wt enzyme, it increased the k_2/K_m value by 2-fold over the S68N mutation (Table 1). Compared with S67M single mutation, even though S67M did not show any detectable UDG activity under the assay conditions, adding S68N mutation to it apparently reactivated the UDG activity (Table 1). These results indicate that S67M and S68N can work as a correlated pair in Lin SMUG1-like glycosylase, although less efficiently as the natural S67-S68 pair. To assess whether the enhanced glycosylase activity on ss U was due to a relaxation of specificity, we examined the activity using double-stranded and single-stranded hypoxanthine-, xanthine-, 5-hydroxymethyluracil (hmU)-, 5-hydroxyuracil (OHU)-, 8-oxoguanine-, 8-oxoadenine-, 5,6-dihydroxyuracil (DHU)-, 5-hydroxycytosine-, thymine glycol-, N6-methyladenine-, O6-methylguanine-, and ethenoadenine-containing DNA with 20 μM enzyme and 10 nM substrate. No detectable activity was found under the assay conditions, suggesting that the S67M-S68N doublet mutant retained its specificity.

3.4. Mutual information analysis

Previously, we used mutual information-based method to investigate amino acid correlation in family 4 UDGa and family 1 UNG enzymes [34]. Applying similar analysis to the SMUG1-like enzymes using six families in the UDG superfamily, we found that S67 and S68 in Lin SMUG1-like DNA glycosylase were also correlated (Fig. 6). This is consistent with our experimental results showing S67M-S68N is more active as a UDG than either S67M or S68N single mutant (Table 1). The equivalent positions to S67-S68 in family 4 Tth UDGa are the E41-G42 pair. Similar to the results reported here, the E41Q-G42D double mutant is more active than E41Q or G42D single mutant [34]. Taken together, these results underscore the correlated nature of the two neighboring residues in different UDG families. As explained above, these two positions play important roles in protein-DNA interactions and catalysis. Understandably, this amino acid doublet may work in concert to carry out its catalytic function in UDG enzymes.

3.5. Concluding remarks

Even though *E. coli* UNG, as the first discovered DNA glycosylase, is a narrow specificity enzyme acting on uracil-containing DNA [6], now we know that enzymes in UDG superfamily are quite diverse in their specificities. Enzymes in families 2, 3, 4, and 6 can

work on purine deamination damage such as xanthine or hypoxanthine DNA glycosylase. As for the UDG activity, family 1 and 4 enzymes can all remove uracil from double- and single-stranded DNA [34]. Some families 2 and 5 enzymes prefer double-stranded DNA [13, 21]. Family 3 SMUG1 enzymes were first identified as a single-stranded selective UDG [18], yet later UDG activity on double-stranded DNA was detected [19, 23, 25, 47, 50]. Unlike family 3 SMUG1 enzymes, the *Listeria* SMUG1-like UDG indeed prefers single-stranded uracil-containing DNA under all assay conditions we tested. The genome of *L. innocua* is 3.09 million bp long with an average G+C content of 37.4% [51]. The weak stacking interactions between A/T pairs can cause higher frequency of opening up DNA to create intermittent single-stranded domains, so-called DNA bubbles and once opened longer lifetime [52, 53]. Spontaneous cytosine deamination occurs in single-stranded DNA is 100-fold greater in comparison to that in the DNA duplex [54, 55]. *Listeria* can also live in high temperature environment [56], at which base deamination is accelerated. The SMUG1-like DNA glycosylase with strong preference for single-stranded DNA may be needed to counter the mutagenic cytosine deamination and maintain the integrity of this low G/C genome. The SMUG1-like enzyme from *Listeria* represents the first example of a single-stranded DNA glycosylase in UDG superfamily.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Min Cao (Clemson University) for the *Listeria innocua* strain and Dr. Richard P. Cunningham (State University of New York, Albany, New York) for *E. coli* strains. We also thank members of Cao laboratory for assistance and discussions.

Funding

This project was supported in part by the National Institutes of Health (GM121997 to W.C.) and the Greenwood Genetic Center (to L.W.).

References

1. Brynolf K, Eliasson R, Reichard P. Formation of Okazaki fragments in polyoma DNA synthesis caused by misincorporation of uracil. Cell. 1978; 13:573–580. [PubMed: 207436]
2. Krokan HE, Drablos F, Slupphaug G. Uracil in DNA–occurrence, consequences and repair. Oncogene. 2002; 21:8935–8948. [PubMed: 12483510]
3. Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993; 362:709–715. [PubMed: 8469282]
4. Neuberger MS, Harris RS, Di Noia J, Petersen-Mahrt SK. Immunity through DNA deamination. Trends Biochem Sci. 2003; 28:305–312. [PubMed: 12826402]
5. Delker RK, Papavasiliou FN. You break it, you fix it: functions for AID downstream of deamination. Nat Immunol. 2013; 14:1112–1114. [PubMed: 24145783]
6. Lindahl T. An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. Proceedings of the National Academy of Sciences of the United States of America. 1974; 71:3649–3653. [PubMed: 4610583]
7. Pearl LH. Structure and function in the uracil-DNA glycosylase superfamily. Mutation research. 2000; 460:165–181. [PubMed: 10946227]

8. Parikh SS, Putnam CD, Tainer JA. Lessons learned from structural results on uracil-DNA glycosylase. Mutat Res. 2000; 460:183–199. [PubMed: 10946228]
9. Gallinari P, Jiricny J. A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. Nature. 1996; 383:735–738. [PubMed: 8878487]
10. Lee HW, Brice AR, Wright CB, Dominy BN, Cao W. Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase. The Journal of biological chemistry. 2010; 285:41483–41490. [PubMed: 20852254]
11. Barrett TE, Savva R, Panayotou G, Barlow T, Brown T, Jiricny J, Pearl LH. Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. Cell. 1998; 92:117–129. [PubMed: 9489705]
12. Barrett TE, Scharer OD, Savva R, Brown T, Jiricny J, Verdine GL, Pearl LH. Crystal structure of a thwarted mismatch glycosylase DNA repair complex. EMBO J. 1999; 18:6599–6609. [PubMed: 10581234]
13. Lee DH, Liu Y, Lee HW, Xia B, Brice AR, Park SH, Balduf H, Dominy BN, Cao W. A structural determinant in the uracil DNA glycosylase superfamily for the removal of uracil from adenine/uracil base pairs. Nucleic acids research. 2015; 43:1081–1089. [PubMed: 25550433]
14. Hang B, Downing G, Guliaev AB, Singer B. Novel activity of Escherichia coli mismatch uracil-DNA glycosylase (Mug) excising 8-(hydroxymethyl)-3,N4-ethenocytosine, a potential product resulting from glycidaldehyde reaction. Biochemistry. 2002; 41:2158–2165. [PubMed: 11841206]
15. Saparbaev M, Laval J. 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. Proceedings of the National Academy of Sciences of the United States of America. 1998; 95:8508–8513. [PubMed: 9671708]
16. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011; 333:1303–1307. [PubMed: 21817016]
17. Zhang P, Huang B, Xu X, Sessa WC. Ten-eleven translocation (Tet) and thymine DNA glycosylase (TDG), components of the demethylation pathway, are direct targets of miRNA-29a. Biochem Biophys Res Commun. 2013; 437:368–373. [PubMed: 23820384]
18. Haushalter KA, Stukenberg MW Todd, Kirschner MW, Verdine GL. Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. Current biology : CB. 1999; 9:174–185. [PubMed: 10074426]
19. Mi R, Dong L, Kaulgud T, Hackett KW, Dominy BN, Cao W. Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG. Journal of molecular biology. 2009; 385:761–778. [PubMed: 18835277]
20. Haas BJ, Sandigursky M, Tainer JA, Franklin WA, Cunningham RP. Purification and characterization of Thermotoga maritima endonuclease IV, a thermostable apurinic/apyrimidinic endonuclease and 3′-repair diesterase. Journal of bacteriology. 1999; 181:2834–2839. [PubMed: 10217775]
21. Xia B, Liu Y, Li W, Brice AR, Dominy BN, Cao W. Specificity and catalytic mechanism in family 5 uracil DNA glycosylase. The Journal of biological chemistry. 2014; 289:18413–18426. [PubMed: 24838246]
22. Lee HW, Dominy BN, Cao W. New Family of Deamination Repair Enzymes in Uracil-DNA Glycosylase Superfamily. J Biol Chem. 2011; 286:31282–31287. [PubMed: 21642431]
23. Pettersen HS, Sundheim O, Gilljam KM, Slupphaug G, Krokan HE, Kavli B. Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. Nucleic acids research. 2007; 35:3879–3892. [PubMed: 17537817]
24. Boorstein RJ, Cummings A Jr, Marenstein DR, Chan MK, Ma Y, Neubert TA, Brown SM, Teebor GW. Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1. The Journal of biological chemistry. 2001; 276:41991–41997. [PubMed: 11526119]
25. Kavli B, Sundheim O, Akbari M, Otterlei M, Nilsen H, Skorpen F, Aas PA, Hagen L, Krokan HE, Slupphaug G. hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. J Biol Chem. 2002; 277:39926–39936. [PubMed: 12161446]

26. Masaoka A, Matsubara M, Hasegawa R, Tanaka T, Kurisu S, Terato H, Ohyama Y, Karino N, Matsuda A, Ide H. Mammalian 5-formyluracil-DNA glycosylase. 2. Role of SMUG1 uracil-DNA glycosylase in repair of 5-formyluracil and other oxidized and deaminated base lesions. Biochemistry. 2003; 42:5003–5012. [PubMed: 12718543]
27. Matsubara M, Masaoka A, Tanaka T, Miyano T, Kato N, Terato H, Ohyama Y, Iwai S, Ide H. Mammalian 5-formyluracil-DNA glycosylase. 1. Identification and characterization of a novel activity that releases 5-formyluracil from DNA. Biochemistry. 2003; 42:4993–5002. [PubMed: 12718542]
28. Doseth B, Ekre C, Slupphaug G, Krokan HE, Kavli B. Strikingly different properties of uracil-DNA glycosylases UNG2 and SMUG1 may explain divergent roles in processing of genomic uracil. DNA repair. 2012; 11:587–593. [PubMed: 22483865]
29. Kemmerich K, Dingler FA, Rada C, Neuberger MS. Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung−/−Msh2−/− mice. Nucleic acids research. 2012; 40:6016–6025. [PubMed: 22447450]
30. Zhang Z, Shen J, Yang Y, Li J, Cao W, Xie W. Structural Basis of Substrate Specificity in Geobacter metallireducens SMUG1. ACS chemical biology. 2016; 11:1729–1736. [PubMed: 27071000]
31. Lucas-Lledo JI, Maddamsetti R, Lynch M. Phylogenomic analysis of the uracil-DNA glycosylase superfamily. Molecular biology and evolution. 2011; 28:1307–1317. [PubMed: 21135150]
32. Fisher CL, Pei GK. Modification of a PCR-based site-directed mutagenesis method. BioTechniques. 1997; 23:570–571, 574. [PubMed: 9343663]
33. Pang P, Yang Y, Li J, Wang Z, Cao W, Xie W. SMUG2 DNA glycosylase from Pedobacter heparinus as a new subfamily of the UDG superfamily. The Biochemical journal. 2017; 474:923–938. [PubMed: 28049757]
34. Xia B, Liu Y, Guevara J, Li J, Jilich C, Yang Y, Wang L, Dominy BN, Cao W. Correlated Mutation in the Evolution of Catalysis in Uracil DNA Glycosylase Superfamily. Sci Rep. 2017; 7:45978. [PubMed: 28397787]
35. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23:2947–2948. [PubMed: 17846036]
36. Pei J, Kim BH, Grishin NV. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic acids research. 2008; 36:2295–2300. [PubMed: 18287115]
37. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution. 2013; 30:2725–2729. [PubMed: 24132122]
38. Foster PL. Methods for determining spontaneous mutation rates. Methods in enzymology. 2006; 409:195–213. [PubMed: 16793403]
39. Hall BM, Ma CX, Liang P, Singh KK. Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinformatics. 2009; 25:1564–1565. [PubMed: 19369502]
40. Sarkar S, Ma WT, Sandri GH. On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. Genetica. 1992; 85:173–179. [PubMed: 1624139]
41. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER, Current protocols in bioinformatics / editoral board. Andreas D Baxevanis … [et al]. 2014; 47:5 6 1–32.
42. Simonetti FL, Teppa E, Chernomoretz A, Nielsen M, Marino Buslje C. MISTIC: Mutual information server to infer coevolution. Nucleic acids research. 2013; 41:W8–14. [PubMed: 23716641]
43. Otterlei M, Kavli B, Standal R, Skjelbred C, Bharati S, Krokan HE. Repair of chromosomal abasic sites in vivo involves at least three different repair pathways. EMBO J. 2000; 19:5542–5551. [PubMed: 11032821]
44. Mi R, Alford-Zappala M, Kow YW, Cunningham RP, Cao W. Human endonuclease V as a repair enzyme for DNA deamination. Mutat Res. 2012; 735:12–18. [PubMed: 22664237]
45. Dou H, Mitra S, Hazra TK. Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. J Biol Chem. 2003; 278:49679–49684. [PubMed: 14522990]

46. Wibley JE, Waters TR, Haushalter K, Verdine GL, Pearl LH. Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. Mol Cell. 2003; 11:1647–1659. [PubMed: 12820976]

47. Matsubara M, Tanaka T, Terato H, Ohmae E, Izumi S, Katayanagi K, Ide H. Mutational analysis of the damage-recognition and catalytic mechanism of human SMUG1 DNA glycosylase. Nucleic acids research. 2004; 32:5291–5302. [PubMed: 15466595]

48. Drohat AC, Stivers JT. Escherichia coli uracil DNA glycosylase: NMR characterization of the short hydrogen bond from His187 to uracil O2. Biochemistry. 2000; 39:11865–11875. [PubMed: 11009598]

49. Barrett TE, Savva R, Barlow T, Brown T, Jiricny J, Pearl LH. Structure of a DNA base-excision product resembling a cisplatin inter-strand adduct. Nature structural biology. 1998; 5:697–701. [PubMed: 9699633]

50. Nilsen H, Haushalter KA, Robins P, Barnes DE, Verdine GL, Lindahl T. Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. EMBO J. 2001; 20:4278–4286. [PubMed: 11483530]

51. Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, Baquero F, Berche P, Bloecker H, Brandt P, Chakraborty T, Charbit A, Chetouani F, Couve E, Daruvar A de, Dehoux P, Domann E, Dominguez-Bernal G, Duchaud E, Durant L, Dussurget O, Entian KD, Fsihi H, Garcia-del Portillo F, Garrido P, Gautier L, Goebel W, Gomez-Lopez N, Hain T, Hauf J, Jackson D, Jones LM, Kaerst U, Kreft J, Kuhn M, Kunst F, Kurapkat G, Madueno E, Maitournam A, Vicente JM, Ng E, Nedjari H, Nordsiek G, Novella S, Pablos B de, Perez-Diaz JC, Purcell R, Remmel B, Rose M, Schlueter T, Simoes N, Tierrez A, Vazquez-Boland JA, Voss H, Wehland J, Cossart P. Comparative genomics of Listeria species. Science. 2001; 294:849–852. [PubMed: 11679669]

52. Ambjornsson T, Banik SK, Krichevsky O, Metzler R. Breathing dynamics in heteropolymer DNA. Biophys J. 2007; 92:2674–2684. [PubMed: 17237209]

53. Ambjornsson T, Banik SK, Krichevsky O, Metzler R. Sequence sensitivity of breathing dynamics in heteropolymer DNA. Phys Rev Lett. 2006; 97:128105. [PubMed: 17026004]

54. Frederico LA, Kunkel TA, Shaw BR. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry. 1990; 29:2532–2537. [PubMed: 2185829]

55. Lindahl T, Nyberg B. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. Biochemistry. 1974; 13:3405–3410. [PubMed: 4601435]

56. Fairchild TM, Foegeding PM. A proposed nonpathogenic biological indicator for thermal inactivation of Listeria monocytogenes. Applied and environmental microbiology. 1993; 59:1247–1250. [PubMed: 8489233]

57. Parikh SS, Walcher G, Jones GD, Slupphaug G, Krokan HE, Blackburn GM, Tainer JA. Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97:5083–5088. [PubMed: 10805771]

Highlights

- A DNA glycosylase from *Listeria innocua* represents a new branch in UDG superfamily.
- *L. innocua* SMUG1-like glycosylase is single-strand selective.
- A Ser-Ser doublet in motif 1 underlies a unique correlated evolution.
- This study underscores divergent paths in the evolution of DNA repair enzymes.

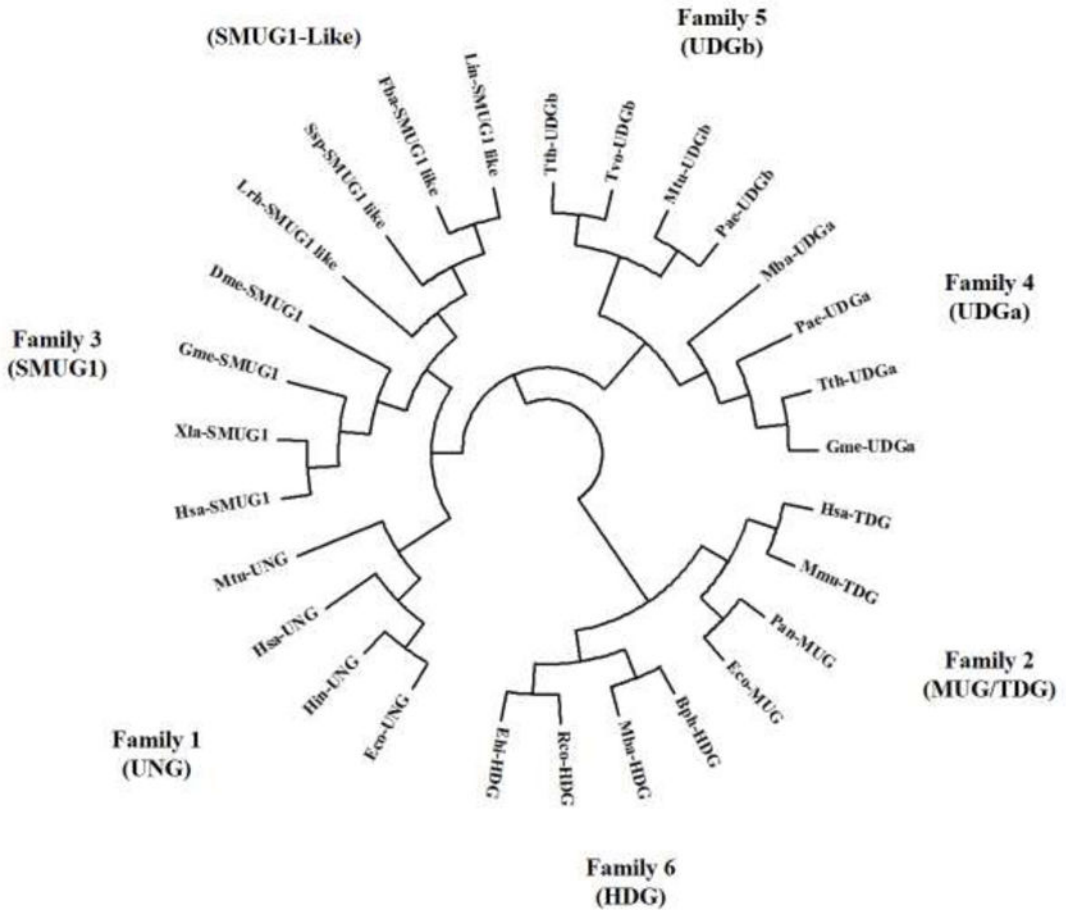

Figure 1. Phylogenetic analysis of UDG superfamily

The phylogenetic analysis was performed using the neighbor-joining method in MEGA 6. GenBank accession numbers are shown after the species names. SMUG1-like glycosylase: Lin, *Listeria innocua* WP_010991469.1; Lrh, *Lactobacillus rhamnosus* WP_014569641.1; Fba, *Firmicutes bacterium CAG:822* WP_021872912.1; Ssp, *Streptomyces sp. NRRL WC-3626* WP_030213602.1. Family 1 (UDG): Eco, *Escherichia coli*, NP_289138; Hin, *Haemophilus influenzae* KR494, YP_008544610.1; Mtu, *Mycobacterium tuberculosis*, WP_003908950.1; Hsa, *Homo sapiens*, NP_003353. Family 2 (MUG/TDG): Eco, *Escherichia coli*, P0A9H1; Pan, *Pantoea ananatis LMG 20103*, ADD78558.1; Hsa, *Homo sapiens*, NP_003202; Mmu, *Mus musculus*, XP_003945901.1. Family 3 (SMUG1): Gme, *Geobacter metallireducens* GS-15,YP_383069; Hsa, *Homo sapiens*, NP_055126; Xla, *Xenopus laevis* AAD17300; Dme, *Drosophila melanogaster*, NP_650609.1. Family 4 (UDGa): Tth, *Thermus thermophilus* HB27, YP_004341.1; Pae, *Pyrobaculum aerophilum str*. IM2, NP_558739.1; Gme, *Geobacter metallireducens* GS-15, YP_006721625.1; Mba, *Methanosarcina barkeri* str. Fusaro, YP_305330.1. Family 5 (UDGb): Tth, *Thermus thermophilus* HB8, YP_144415.1; Pae, *Pyrobaculum aerophilum* str. IM2, NP_559226; Tvo, *Thermoplasma volcanium* GSS1, NP_111346.1; Mtu, *Mycobacterium tuberculosis* H37Rv, P64785 (Rv1259). Family 6 (HDG): Bph, *Burkholderia phymatum* STM815, YP_001858334.1; Mba, *Methanosarcina* barkeri str. Fusaro, YP_304295.1; Rco, *Ricinus communis*, XP_002536323.1; Ehi, *Entamoeba histolytica* HM-1:IMSS, XP_655177.1.

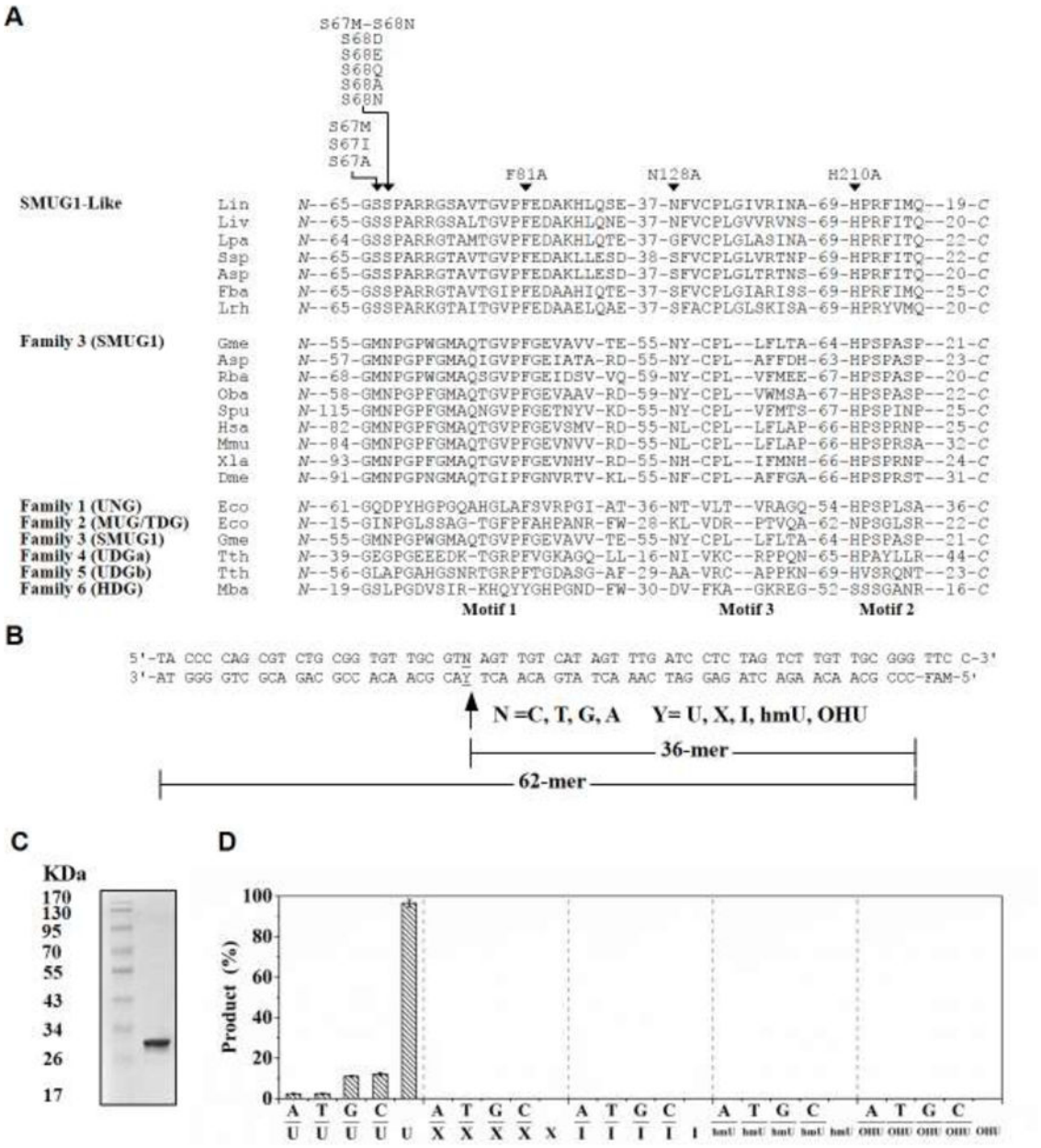

Figure 2. Multiple sequence alignment and DNA glycosylase activity of Lin SMUG1-like glycoyslase

A. Multiple sequence alignment of SMUG1-like and SMUG1 enzymes. GenBank accession numbers are shown after the species names. SMUG-like glycosylase: Lin, *Listeria innocua* WP_010991469.1; Liv, *Listeria ivanovii* WP_025279932.1; Lpa, *Lactobacillus paracasei* WP_016381167.1; Ssp, *Streptomyces sp.* NRRL WC-3626 WP_030213602.1; Asp, Amycolatopsis sp. WP_037336521; Fba, *Firmicutes bacterium CAG:822* WP_021872912.1; Lrh, *Lactobacillus rhamnosus* WP_014569641.1; Family 3 (SMUG1): Gme, *Geobacter metallireducens* GS-15, YP_383069; Asp, *Azoarcus sp.* BH72, YP_935478; Rba, *Rhodopirellula baltica SH 1*, NP_869403; Oba, *Opitutaceae bacterium* TAV2, ZP_02013615.1; Spu, *Strongylocentrotus purpuratus*, XP_782746.1; Hsa, *Homo sapiens*,

NP_055126; Mmu, *Mus musculus*, NP_082161; Xla, *Xenopus laevis*, AAD17300; Dme, *Drosophila melanogaster*, NP_650609.1; Family 1 (UNG): Eco, *Escherichia coli*, NP_289138; Family 2 (MUG/TDG): Eco, *Escherichia coli*, P0A9H1. Family 3 (SMUG1): Gme, *Geobacter metallireducens* GS-15, YP_383069. Family 4 (UDGa): Tth, *Thermus thermophilus* HB27, YP_004341.1. Family 5 (UDGb): Tth, *Thermus thermophilus* HB8, YP_144415.1; Family 6 (HDG): Mba*, Methanosarcina barkeri* str. Fusaro, YP_304295.1. **B**. Sequences of deoxyoligonucleotide substrates. U, uracil; X, xanthine; I, hypoxanthine; hmU, 5-hydroxymethyluracil; OHU, 5-hydroxyuracil. FAM, 6-FAM fluorophore. **C**. Purified Lin SMUG1-like glycosylase protein as shown in 12% SDS-PAGE gel. **D**. DNA glycosylase activity of wt Lin SMUG1-like glycosylase on U-, X-, I-, hmU- and OHU-containing substrates. Cleavage reactions were performed as described in Material and Methods with 20 μM wt Lin SMUG1-like protein and 10 nM substrate and incubated for 2 hours. Data are the averages of three independent experiments.

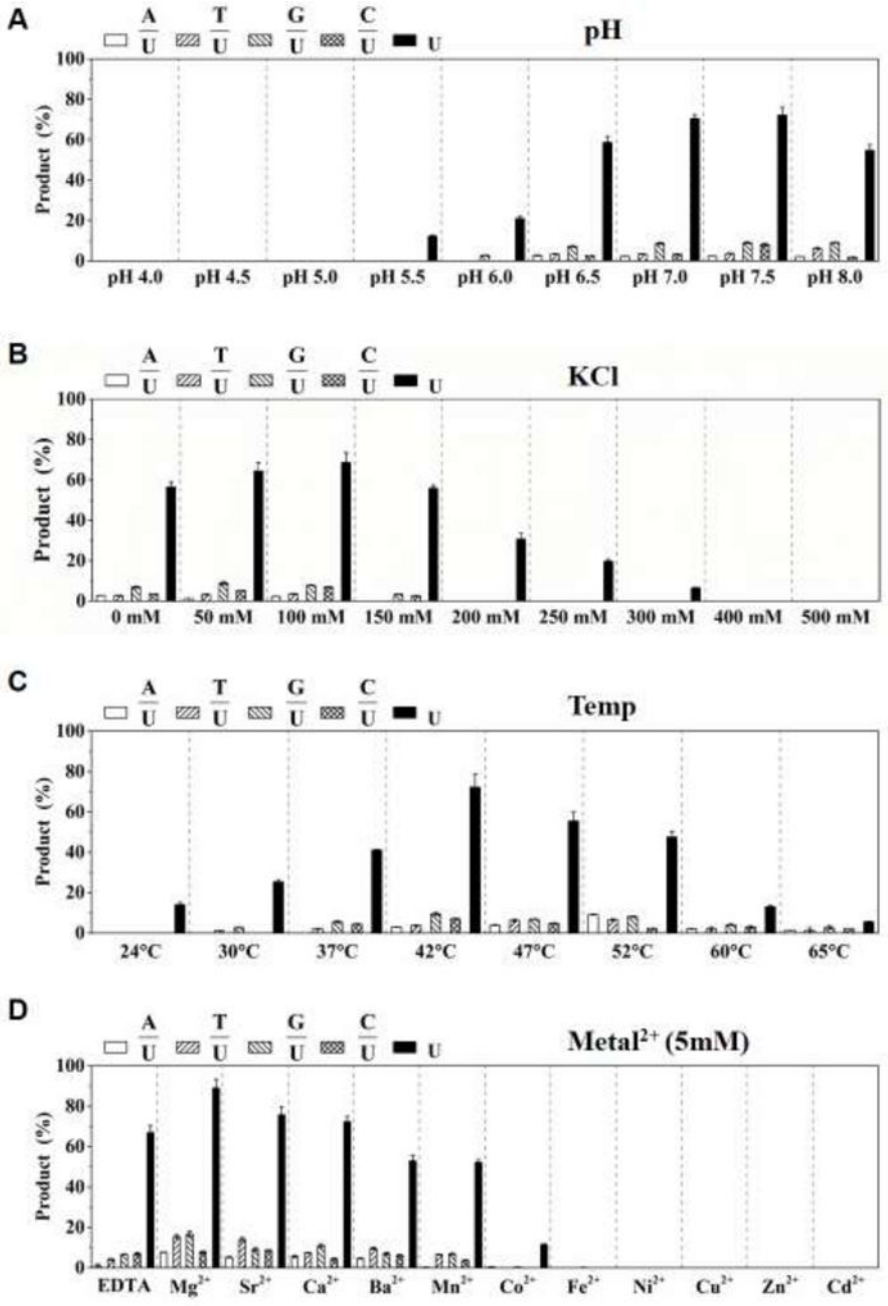

Figure 3. Effects of assay conditions on UDG activity of Lin SMUG1-like glycosylase

DNA glycosylase assays were performed as described in Material and Methods with 5 μM wt Lin SMUG1-like glycosylase and 10 nM DNA substrate and incubated for 1 hour. Data are the averages of three independent experiments. **A**. pH effect. The Tris-HCl buffer was replaced with 20 mM citrate-phosphate universal buffer at the indicated pH value. 4.0–8.0. **B**. salt effect. KCl was added to the reaction mixtures in the range of 0 to 500 mM. **C**. Temperature effect. The reaction mixtures were incubated at indicated temperatures before quenching. **D**. Divalent metal ion effect. Divalent metal ions in the final concentration of 5 mM were supplemented to the reaction mixtures.

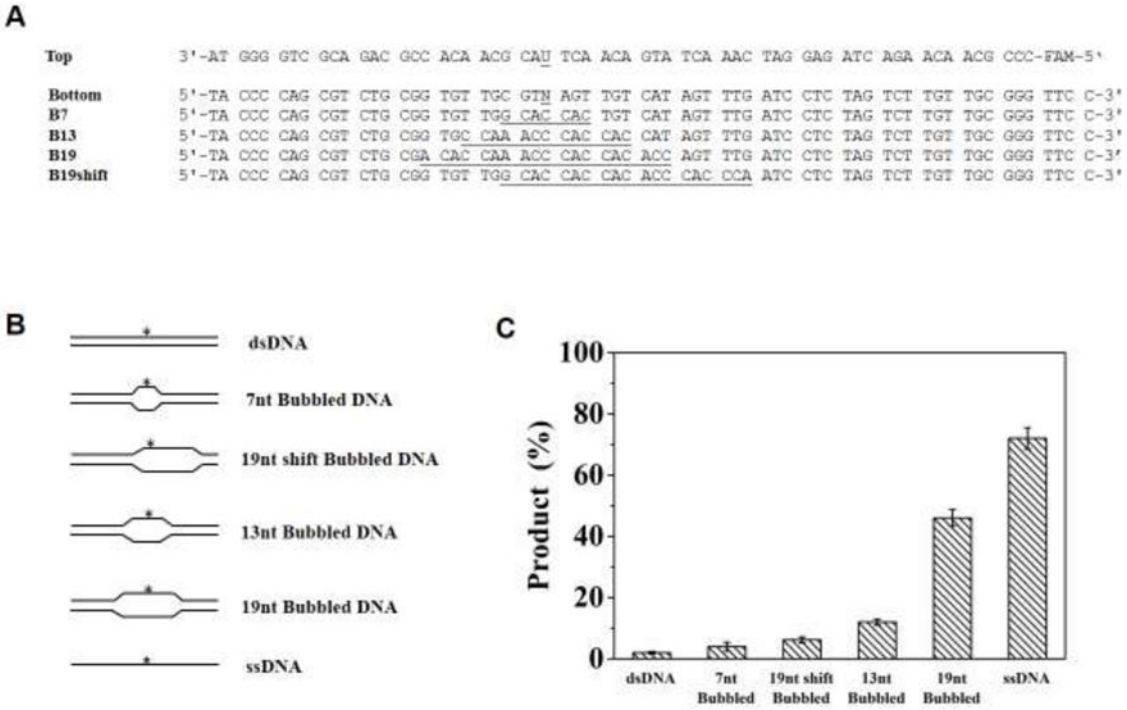

Figure 4. UDG activities of Lin SMUG1-like glycosylase on bubble-containing DNA substrates

A. Sequence of bubble substrates. **B**. Schematic illustration of bubble substrates. **C**. UDG activity of Lin SMUG1-like glycosylase on bubble-containing DNA substrates. Cleavage reactions were performed as described in Material and Methods with 5 μM Lin SMUG1-like and 10 nM substrate and incubated for 1 hour. Data are the average of three independent experiments.

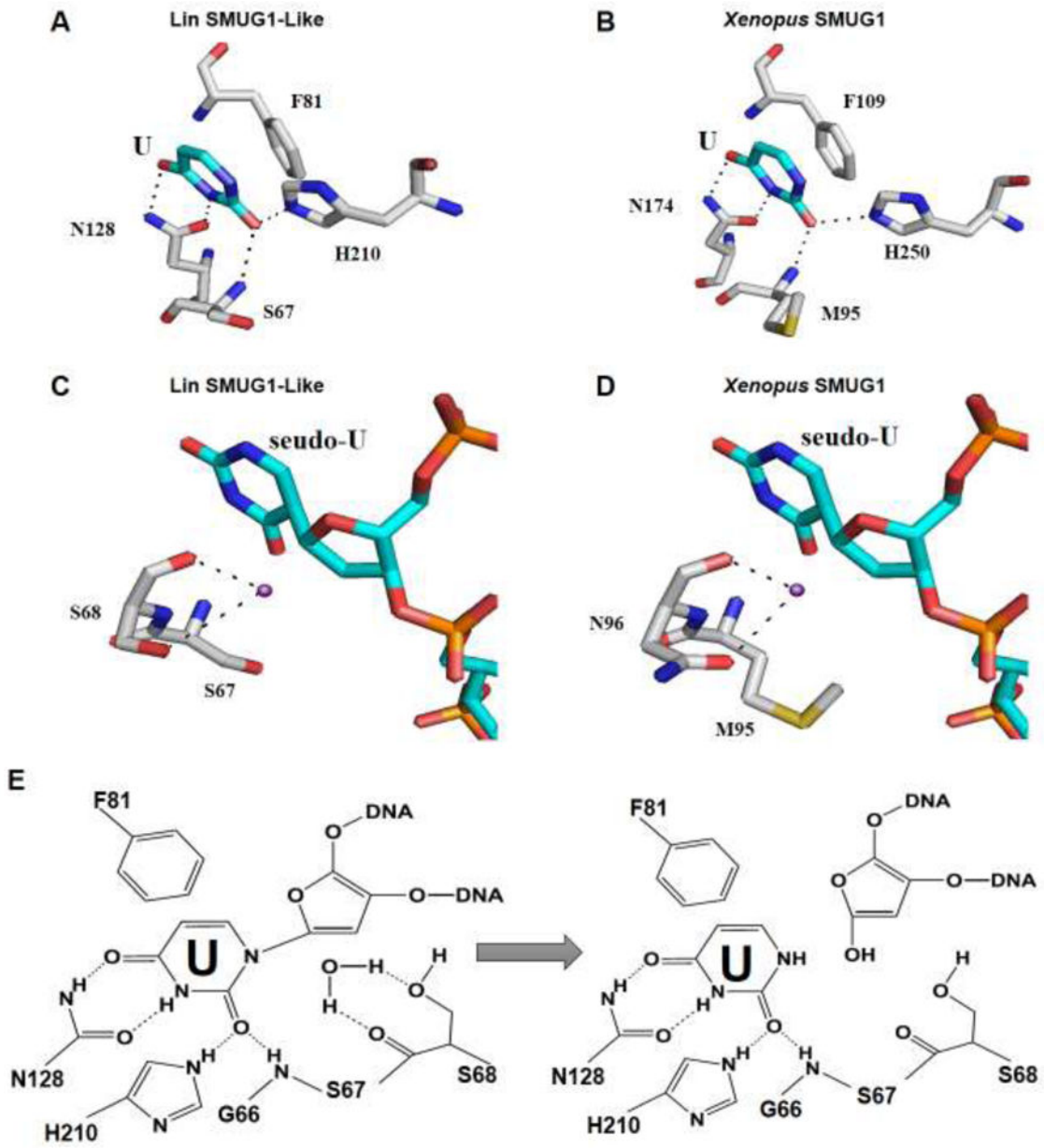

Figure 5. Interactions between Lin SMUG1-like glycosylase and uracil-containing DNA

Uracil is shown in cyan. Hydrogen bonds are shown as dash lines. **A**. Interactions between Lin SMUG1-like glycosylase and uracil base. Lin SMUG1-like glycosylase structure was modeled based on *Xenopus* SMUG1 structure (PDB code 1OE5). **B**. Interactions between *Xenopus* SMUG1 glycosylase (PDB code 1OE5) and uracil base. **C**. Modeled interactions between S67 and S68 of Lin SMUG1-like glycosylase and uracil base. A water molecule is shown as a purple circle. The DNA with seudo-U base is taken from PDB 1EMH [57]. **D**. Modeled interactions between M95 and N96 of *Xenopus* SMUG1 glycosylase and uracil base. A water molecule is shown as a purple circle. The DNA with seudo-U base is taken from PDB 1EMH. **E**. Hypothetical catalytic mechanism of Lin SMUG1-mediated hydrolysis of the N-glycosidic bond.

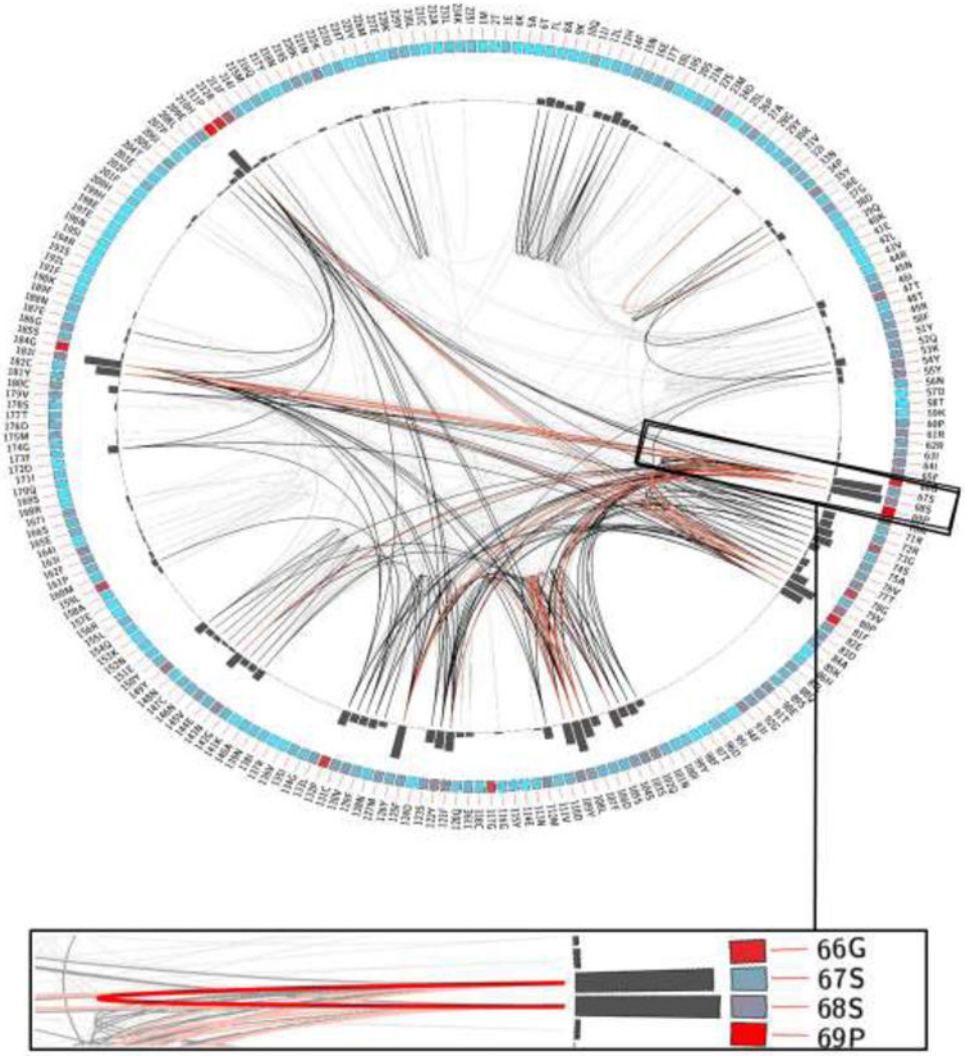

Figure 6. Circos diagram of Lin SMUG1-like glycosylase

The circular representation of the multiple sequence alignment using Lin SMUG1-like glycosylase as the reference sequence. The diagram contains the amino acid residue positions and residue identities mapped to the reference sequence. The square box below each residue represents the level of conservation ranging from red (highly conserved) to blue (less conserved). The bars in the histogram represent the co-evolutionary correlations from the mutual information analysis with a value higher than 6.5 [42]. The connecting lines between residue pairs follow a color scheme for ranking correlation between positions in the multiple sequence alignment where red indicates the top 5%, black between 95% and 70% and gray the remaining interactions.

Table 1

Kinetic constants of Lin SMUG1-like WT and mutant proteins for excision of uracil[a]

Lin SMUG1	Substrate	k_2/K_m (min^{-1} nM^{-1})
WT	ss U	3.5×10^{-6}
	19 bubble	1.5×10^{-6}
S67M	ss U	N.A.[b]
	19 bubble	N.A.
S68N	ss U	6.4×10^{-8}
	19 bubble	N.A.
S67M-S68N	ss U	1.4×10^{-7}
	19 bubble	N.A.

[a] Kinetic analysis was carried out as described in Material and Methods. Data are the averages of three independent experiments.

[b] N.A. No activity detected under the assay conditions.

Table 2

Antimutator effect of the Lin SMUG1-like glycosylase[a]

Plasmid	pBS-UNG	pBS-WT	pBS-S67M	pBS
Mutation Frequency ($Rif^R/10^9$ viable cells)[b]	32	464	939	950
Mutation Frequency ($Rif^R/10^9$ viable cells per generation)[c]	24	146	247	254

[a] pBS-UNG, pBluescript containing *E. coli ung* gene; pBS-WT, pBluescript containing Lin SMUG1-like glycosylase gene; pBS-S67M, pBluescript containing the S67M mutant of the Lin SMUG1-like glycosylase gene; pBS, pBluescript alone.

[a] The average mutation frequency was determined from at least 3 independent experiments.

[b] Each mutation frequency was determined from 48–60 independent cultures.