Clemson University
TigerPrints

Publications | Genetics and Biochemistry

3-2017

SMUG2 DNA Glycosylase from Pedobacter heparinus as a New Subfamily in UDG Superfamily

Panjiao Pang
Sun Yat-sen University

Ye Yang
Clemson University

Jing Li
Clemson University

Zhong Wang
Sun Yat-sen University

Weiguo Cao
Clemson University, wgc@clemson.edu

See next page for additional authors

Follow this and additional works at: http://tigerprints.clemson.edu/gen_biochem_pubs

 Part of the Biochemistry, Biophysics, and Structural Biology Commons

Recommended Citation
Please use the publisher's recommended citation: http://www.biochemj.org/content/474/6/923.long

This Article is brought to you for free and open access by the Genetics and Biochemistry at TigerPrints. It has been accepted for inclusion in Publications by an authorized administrator of TigerPrints. For more information, please contact awesole@clemson.edu.

Authors
Panjiao Pang, Ye Yang, Jing Li, Zhong Wang, Weiguo Cao, and Wei Xie

This article is available at TigerPrints: http://tigerprints.clemson.edu/gen_biochem_pubs/19

SMUG2 DNA Glycosylase from *Pedobacter heparinus* as a New Subfamily in UDG Superfamily

Panjiao Pang[1], Ye Yang[2], Jing Li[2], Zhong Wang[1], Weiguo Cao[2], Wei Xie[3, *]

[1]School of Pharmaceutical Sciences, The Sun Yat-Sen University, 132 E. Circle Rd. University City, Guangzhou, Guangdong 510006, People's Republic of China.

[2]Department of Genetics and Biochemistry, Clemson University, 190 Collings Street, Clemson, SC 29634, United States

[3] State Key Laboratory for Biocontrol, School of Life Sciences, The Sun Yat-Sen University, 135 W. Xingang Rd., Guangzhou, Guangdong 510275, People's Republic of China.

*Correspondence should be addressed to:

Weiguo Cao: wgc@clemson.edu; Wei Xie: xiewei6@mail.sysu.edu.cn.

Running title: SMUG2 DNA Glycosylase from *Pedobacter heparinus*

Formatted: Not Superscript/ Subscript

Summary Statement

We discovered a glycoyslase named PheSMUG2 that removes uracil or hypoxanthine/xanthine from DNA. Phylogenetic analyses, biochemical and crystallographic studies indicated that PheSMUG2 belongs to a new subfamily, and they provide mechanistic insight into the molecular mechanism of the uracil DNA glycosylase superfamily.

Abstract

Base deamination is a common type of DNA damage that occurs in all organisms. DNA repair mechanisms are essential to maintain genome integrity, in which the base excision repair (BER) pathway plays a major role in the removal of base damage. In the BER pathway, the uracil DNA glycosylase superfamily is responsible for excising the deaminated bases from DNA and generates apurinic/apyrimidinic (AP) sites. Using bioinformatics tools, we identified a family 3 SMUG1-like DNA glycoyslase from *Pedobacter heparinus* (named as Phe SMUG2), which display catalytic activities towards DNA containing uracil or hypoxanthine/xanthine. Phylogenetic analyses show that SMUG2 enzymes are closely related to family 3 SMUG1s but belong to a distinct branch of the family. The high solution crystal structure of the apoenzyme reveals that the general fold of Phe SMUG2 resembles SMUG1s, yet with several distinct local structural differences. Mutational studies, coupled with structural modeling, identify several important amino acid residues for the glycosylase activity. Substitution of G65 with a tyrosine results in loss of all glycosylase activity. The crystal structure of the G65Y mutant suggests a potential misalignment at the active site due to the mutation. The relationship between the new subfamily and other families in UDG superfamily is discussed. This study provides new mechanistic insight into the molecular mechanism of the UDG superfamily.

Abbreviations

BER, base excision repair; AP, apurinic/apyrimidinic; UDG, uracil DNA glycosylase; fC, formyl cytosine; caC, carboxyl cytosine; MUG, mismatch-specific glycosylase; xSMUG1, *Xenopus laevis* single-strand-selective monofunctional UDG; hUDG, human UDG; Gme SMUG1, *Geobacter metallireducens* SMUG1; Phe SMUG2, *Pedobacter heparinus* Single-strand-selective monofunctional uracil-DNA-glycosylase 2; TDG, thymine DNA glycosylase; hTDG, human TDG; HDG, hypoxanthine DNA glycosylase; XDG, xanthine DNA glycosylase, SeMet, selenomethionine; SAD, single-wavelength anomalous dispersion.

Keywords

Deamination, DNA repair, SMUG, uracil-DNA glycosylase, enzyme evolution.

INTRODUCTION

DNA bases routinely undergo modifications like deamination, oxidation and alkylation, caused by endogenous and environmental agents [1-3]. For example, the deamination of adenine (A), and cytosine generate hypoxanthine (I) and uracil (U) respectively, while the deamination of guanine largely produces xanthine (X) [4, 5]. The deaminated base may lead to mutations and other defects to the living species if they are not corrected by DNA repair systems [6-11]. The frequently generated deamination product uracil within DNA is removed by uracil DNA glycosylase (UDG), which initiates the base excision repair (BER) pathway [12]. The UDG superfamily typically contains a four-strand β-sheet surrounded by α-helices. The catalytic mechanism involves activation of the leaving group, stabilization of the oxocarbenium ion, and positioning of a water molecule, which leads to hydrolysis of the *N*-glycosidic bond and generation of an apurinic/apyrimidinic (AP) site [13]. The superfamily is classified into six families based on their sequence homology, biochemical and structural similarities [14, 15]. Family 1 UDGs (also called UNGs), as represented by the *Escherichia coli* (*E. coli*), human, and herpes simplex virus 1 UDGs, are enzymes acting on uracil in both double-stranded (ds) and single-stranded (ss) DNA [14, 16]. Family 2 is composed of well-characterized human thymine DNA glycosylase (TDG) [17], *E. coli* mismatch-specific glycosylase (MUG) [18, 19], and fission yeast *Schizosaccharomyces pombe* TDG with broad substrate specificity [20]. Whereas human TDG (hTDG) was recognized as a part of demethylation system due to its DNA glycosylase activity on formyl cytosine (fC) and carboxyl cytosine (caC), the *E. coli* enzyme was initially found as a mismatch-specific UDG and later as a robust xanthine DNA glycosylase [21]. Family 3 enzymes are considered as a hybrid as they contain signature motif 1 of family 2 MUG/TDG and signature motif 2 of family 1 UNG,

as represented by African clawed frog *Xenopus laevis* single-strand-selective monofunctional UDG (xSMUG1) [22], *Geobacter metallireducens* SMUG1 (Gme SMUG1) and human SMUG1 [23]. Interestingly, not only do SMUG1s display UDG activities, they also exhibit xanthine DNA glycosylase (XDG) activities. Family 4 UDG enzymes are a group of prokaryotic iron–sulfur-containing enzymes that act on both ss- and ds-DNA containing uracil [24, 25]. Family 5 UDG enzymes are found in prokaryotic bacteria and archaea, only work on ds DNA [26, 27]. Lastly, family 6 is mainly made of HDG enzymes [28].

Many crystal structures have been determined to understand the catalytic mechanism of UDG, especially for the family 1 and 2 members. The first structure of UDG was from the herpes simplex virus, and the α/β fold revealed by the high-resolution structure allows one to propose a catalytic mechanism for the hydrolytic base excision (PDB 1UDG) [29]. The co-crystal structure of human UDG (hUDG) complexed with a damaged DNA substrate was subsequently determined. The enzyme employs a base-flipping mechanism, and the flipped uracil base interacts with the uracil-binding pocket using its phosphate, ribose and base (PDB 4SKN) [30]. Three residues are essential for catalysis: L272 is required for intercalation of the enzyme into the DNA major groove, and D145 (D88 in HSV UDG) is required for activating the active-site water for the nucleophilic attack while H268 (H210 in HSV UDG) is critical for catalysis as well. Therefore, UDG is considered to flip uracil nucleosides out of the DNA base stack using a ‘push-pull’ mechanism, in which the leucine penetrates into DNA (push), and complementary interactions from the uracil recognition pocket facilitate the final productive binding (pull). The structural studies on MUGs paralleled those on UDGs. The first MUG, *E. coli MUG* (EcMUG) structures in its DNA-free and DNA-bound forms were determined in

1996 (PDBs 1MUG, 1MWJ, 1MWI) [18, 31]. Despite low sequence identity with UDGs, the EcMUG structures are structurally and functionally similar to the former. However, subtle sequence and structural differences between the two types of enzymes gave rise to different substrate specificities. Interestingly, the catalytic residue aspartate (D64 in *E. coli* UNG) is replaced by an asparagine residue in family 2 MUG/TDG (N18 in EcMUG). It has been hypothesized that although N18 is unable to act as a general base like its aspartate counterpart in family 1 UNGs, the conserved asparagine residue can still bind and position a water molecule in a similar manner to the catalytic aspartate in family 1 UNGs [18]. Aside from the differences in motif 1, the motif 2 in family 1 UNG and family 2 MUG/TDG adopt a nearly identical conformation. However, the catalytically important histidine residue in motif 2, which is highly conserved in family 1 UNG and other families, is missing from motif 2 in family 2 MUG/TDG (Fig. 1A), underscoring their difference in catalytic mechanism.

Previously, we identified and characterized the family 3 SMUG1 enzyme from *Geobacter metallireducens* (Gme SMUG1) [23]. Gme SMUG1 not only possesses the UDG activity but also the XDG activity. As an enzyme with hybrid motifs, the conserved catalytic residue in motif 1 in SMUG1 is an asparagine (N58 in Gme SMUG1), which is found in family 2 but not in family 1. On the other hand, the invariable histidine (H210 in Gme SMUG1) in motif 2 in SMUG1s is not found in family 2. In this study, through bioinformatics search, we identified a putative DNA glycosylase (Genbank WP_ 012780920.1) from the bacterium *Pedobacter heparinus* that shares sequence similarities in motifs 1, 2 and 3 with family 3 SMUG1 and family 2 MUT/TDG enzymes (Fig. 1A). Different from typical family 3 SMUG1 enzymes but similar to some enzymes in family 2 MUG/TDG, this DNA glycosylase possesses not only UDG and XDG activity, but also

hypoxanthine DNA glycosylase (HDG) activity. Structurally, it is more similar to family 3 SMUG1 than to family 2 MUG/TDG. Mutational analysis reveals key catalytic residues responsible for the removal of pyrimidine damage and purine damage. This study establishes a new subfamily in UDG superfamily.

MATERIALS AND METHODS

Reagents, media and strains

All routine chemical reagents were purchased from Sigma Chemicals (St. Louis, MO), Fisher Scientific (Suwanee, GA), or VWR (Suwanee, GA), and all buffers were prepared in high quality deionized water from a Thermo Scientific Nanopure Water System (Suwanee, GA) having resistivity greater than 18.2 MΩ.cm. Restriction enzymes, Phusion High-Fidelity DNA polymerase and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). HisTrap FF and HiTrap SP FF columns were purchased from GE Healthcare Life Sciences (Piscataway, NJ). Single-stranded oligonucleotides were ordered from Integrated DNA Technologies Inc. (Coralville, IA), and all the synthetic oligonucleotides were purified by high-performance liquid chromatography. Hi-Di Formamide and GeneScan 500 LIZ dye Size Standard for ABI 3130xl were purchased from Applied Biosystems. Sonication buffer consisted of 20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 0.1% Triton X-100 and 40 mM imidazole with freshly added 1 mM dithiothreitol (DTT) and 0.15 mM phenylmethylsulfonyl fluoride (PMSF). Buffer A consisted of 20 mM Tris-HCl (pH 7.5), 300 mM NaCl and 40 mM imidazole. Buffer B consisted of 20 mM Tris-HCl (pH 7.5), 300 mM NaCl and 500 mM imidazole. Protein storage buffer consisted of 10 mM Tris-HCl (pH 7.5), 50mM KCl, 1 mM DTT, 1 mM EDTA, 50% Glycerol. The *Pedobacter heparinus* DSM 2366 strain was purchased from DSMZ (Germany).

Plasmid construction, cloning and expression of protein

The *P. heparinus* SMUG2 gene (GenBank accession number: WP_012780920.1) was amplified by PCR using the forward primer Phe_SMUG2_F (5'- GGG AAT TCC ATA TGA TGA CCT TTG CAG AC- 3'; the *Nde*I site is underlined) and the reverse primer Phe_SMUG2_R (5'- CCG CTC GAG TAT CGC AGA ATT GTT AAG -3'; the *XhoI* site is underlined). The PCR reaction mixture (25 μl) consisted of 8 ng of *P. heparinus* genomic DNA, 200 nM forward primer Phe_SMUG2_F and reverse primer Phe_SMUG2_R, 1 x Phusion PCR buffer (New England Biolabs), 200 μM each dNTP, and 1 unit of Phusion DNA polymerase. The PCR procedure included a pre-denaturation 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 pET-21a (+) were digested with *Nde*I and *Xho*I, 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 *P. heparinus* SMUG2 gene in the resulting plasmid (pET-21a(+) -Phe-SMUG2) was confirmed by DNA sequencing. The resulting plasmid with wild-type SMUG2 was used as the template plasmid for all other SMUG2 mutants. Amplification of the mutant DNA and *Dpn*I mediated site-directed mutagenesis procedures were modified as previously described [32] by using primers carrying the desired mutations. Briefly, PCR mixtures (25 μl) contained 10 ng of pET-21a (+)-Phe-SMUG2 as a template, 65 nM of each primer pair, 200 μM each dNTP, 1 x Phusion PCR polymerase buffer, and 1 unit of Phusion DNA polymerase. The PCR procedure included a pre-denaturation step

at 98°C for 2 min; 25 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 30 s, annealing at 55°C for 30 s and extension at 68°C for 5 min; and a final extension step at 68°C for 10 min. After treatment with 2 units of *Dpn*I for 1 h at 37°C, 5-µl PCR products were transformed into *E. coli* DH5α competent cells. Successful insertion and mutation in the resultant clones were confirmed by DNA sequencing.

To express the C-terminal His6-tagged wild-type and mutant Phe SMUG2, the recombinant plasmids were transformed into *E. coli* strain BL21 ($ung^{-}mug^{-}$) by electroporation. Induction, sonication and purification were carried out as previously described with modifications [28]. An overnight *E. coli* culture from a single colony transformed with recombinant plasmid was diluted 100-fold into LB medium (500 ml) supplemented with 100 µg/mL ampicillin and grown at 37°C with shaking at 250 rpm until the optical density at 600 nm reached about 0.4. After adding isopropyl-1-thio-α-d-galctopyranoside (IPTG) to a final concentration of 1 mM, the culture was grown at 22°C for an additional 18 h. The cells were collected by centrifugation at 5,000 rpm with JLA-81000 rotor at 4°C and washed once with pre-cooled sonication buffer.

To purify the Phe SMUG2 protein, the cell pellet from a 500-ml culture was suspended in 7 ml sonication buffer and sonicated at output 5 for 3 x 1 min with 5 min rest on ice between intervals using Qsonica model Q125. The sonicated solution was clarified by centrifugation at 12,000 rpm with JLA-16.250 rotor at 4°C for 20 minutes. The supernatant was transferred into a fresh tube and loaded onto a 1 ml HisTrap FF column. The column was washed with Buffer A for 20 ml. The bound protein in the column was eluted with a linear gradient of 0-100 % Buffer B. Fractions containing a

SMUG2 protein were identified by SDS-PAGE, pooled, diluted to 3-fold and applied to a 1-ml HiTrap SP FF column, which was pre-equilibrated with buffer A without imidazole (20 mM Tris-HCl (pH 7.5), 300 mM NaCl). By linear gradient elution (100 mM-1000 mM NaCl), fractions containing SUMG2 protein were identified by SDS-PAGE, pooled, concentrated and exchange to storage buffer through Microcon YM 10 (Millipore). The protein concentration was quantified by Bradford method using bovine serum albumin as a standard. The SMUG2 protein was stored in aliquots at -20°C.

Oligodeoxynucleotide substrates

The oligonucleotides used for DNA glycosylase activity assay were prepared as previously described [23]. Briefly, the lesion-containing strand (10 μM) was mixed with the unlabeled strand in 1.5-fold molar excess, incubated at 85 °C for 3 min, and allowed to form duplex DNA substrates at room temperature for more than 30 min.

DNA glycosylase activity assay

DNA glycosylase cleavage assays for Phe SMUG2 proteins were performed at 37°C for 1 h in a 10-μl reaction mixture containing 10 nM oligonucleotide substrate, 250 nM glycosylase protein unless noted otherwise, 20 mM Tris-HCl (pH 7.5), 50 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 M NaOH. To quantify the cleavage products and remaining substrates, the reaction mixtures (2 μl) were mixed with 7.8-μl Hi-Di formamide and 0.2-μl GeneScan 500 LI Size Standard (Life Technologies) and analyzed by Applied Biosystems 3130xl sequencer with a fragment analysis module. Cleavage products and remaining substrates were quantified by the GeneMapper software.

Kinetics measurements were carried out under SMUG2 in excess condition, and the reactions were performed with 50-nM oligonucleotide substrate and indicated

concentrations of the glycosylase. Samples were withdrawn at indicated time points. All experiments were repeated for at least three times. The rate constants were determined by curve fitting using the integrated first-order rate equation, $P = P_{max}(1 - e^{-v_0 t})$, where P is the product yield, P_{max} is the maximal yield, t is time and v_0 is the initial rate. The v_0 was plotted as function of enzyme concentration, and data were fitting to the equation $v_0 = v_{max}\,[E]/(K_m + [E])$ [33].

Phylogenetic analysis

A total of 28 glycosylase protein sequences belonging to various UDG families were retrieved from GeneBank and aligned by sequence alignment program ClustalX2 [34]program and a structure-based alignment program PROMALS3D [35]. Subsequently, the resulting alignment was refined manually based on comparison of results. The phylogenetic tree was generated with Neighbor-Joining method within MEGA v6.0 software package [36] to infer the evolutionary history between different UDG families.

Expression and purification of selenomethionine (SeMet)-labeled protein

A 20-mL overnight Luria-Bertani culture of the WT Phe SMUG2/pET-21a (+) containing 50-μg/ml ampicillin was prepared. The next day the cells were resuspended with 10-mL M9 medium after centrifugation and were transferred to the 1L M9 salt medium containing 50-μg/ml ampicillin. When the OD at 600 nm reached 0.5, an amino acid mixture containing 100 mg lysine, phenylalanine, threonine, 50 mg isoleucine, lecucine, valine, and 60 mg SeMet was added. Then the growth temperature was reduced to 30ºC. The culture was induced with 0.5 mM IPTG and kept shaking for 16~18 hours. The cells were harvested by centrifugation at 4,000 RPM for 20 min, resuspended in pre-chilled nickel-nitrilotriacetic acid (Ni-NTA) buffer A containing 20 mM Trsi-HCl pH 8.0, 250

mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol (β-ME), 1 mM phenylmethylsulfonyl fluoride (PMSF). The cells were disrupted by ultrasonication and the supernatant was obtained by centrifugation at 14,000 rpm for 1 h at 4°C. The supernatant was then applied onto Ni-NTA affinity resin (Qiagen) equilibrated with Ni-NTA buffer A. The target protein was eluted with Ni-NTA buffer B (20 mM Tris-HCl (pH 8.0), 250 mM NaCl, 250 mM imidazole, 10 mM β-ME and 1 mM PMSF). The target protein fractions were pooled and dialyzed in a buffer consisting of 20 mM Tris-HCl (pH 8.0), 250 mM NaCl, and 2 mM DTT. The dialyzed protein was applied onto a HiTrap Heparin HP column (GE Healthcare) equilibrated with Heparin HP buffer A (20 mM Tris-HCl (pH 8.0), 50 mM NaCl, 2 mM DTT), and the Phe SMUG2 protein was eluted with 20 mM Tris-HCl (pH 8.0), 2 mM DTT with a NaCl concentration gradient from 0.05-1 M. The final protein were concentrated to 8 mg/ml by a Millipore centrifugal filter (molecular-weight cutoff 10 kDa) and stored at -80°C.

Crystallization, data collection and structure determination

The initial screens for Phe SMUG2 crystals were manually set up using the sitting-drop vapor-diffusion method, with crystal screens I and II and the index screen (Hampton Research, CA, USA). The sample was mixed with the well solution in 1:1 ratio (v/v). The crystals were obtained from 27% PEG 3350，0.25 M NaCl, 0.1 M Tris-HCl pH 8.5/0.1 M HEPES pH 8.0. The SeMet-labeled wild-type protein and the G65Y mutant were crystallized under the same condition, with the SeMet-labeled sample being kept in a reducing environment throughout.

All crystals were flash frozen in liquid nitrogen after being soaked in a cryoprotectant, containing all the reservoir solution components supplemented with 20%

glycerol (v/v). X-ray diffraction data were collected using beamline 17U1 (BL17U1) at the Shanghai Synchrotron Radiation Facility (SSRF, Shanghai, P. R. China). A complete dataset for the native and SeMet-labeled crystals (single-wavelength dataset collected at 0.979 Å, the peak position of SeMet anomalous wavelength) each were collected, with an oscillation angle 1°. The diffraction images were processed using HKL2000 [37]. The SeMet-MUG structure was determined with single-wavelength anomalous dispersion (SAD) phasing using Phaser [38], Cell content analysis indicated that there was one molecule in the asymmetric unit. The model-building and refinement were carried out by Coot and Phenix.refinement [39, 40]. The higher resolution native data was further built and completely traced based on the initial model from SAD phasing. The final model was validated by Molprobity [41]. The structure of the G65Y mutant was obtained using molecular replacement with the WT structure as the search model. All the data collection and refinement statistics are presented in Table 1.

Gel mobility shift assay

DNA binding affinity of Phe SMUG2 protein was measured according to a previously described method with modifications [42]. Indicated concentrations of Phe SMUG2 protein, from 0 nM to 4000 nM, was incubated with 5 nM DNA substrates, 20 mM Tris-HCl (pH 7.5), 50 mM KCl, 5 mM EDTA and 1 mM DTT. The binding mixtures were analyzed by 12% polyacrylamide gel electrophoresis and visualized by Typhoon 7000 imager (GE Healthcare). The concentration of free DNA is given as follow:

$$DNA_{free} = \frac{-SMUG2_{total}+DNA_{free}-K_d+\sqrt{(SMUG2_{total}-DNA_{total}+K_d)^2+4K_d}}{2} \quad (1)$$

where $SMUG2_{total}$ is the initial total Phe SMUG2 protein concentration, DNA_{total} is the initial concentration of DNA substrate. Dissociation constant (K_d) of Phe SMUG2 was determined by nonlinear curve fitting with Equation 1.

RESULTS

Identification and glycosylase activity of Phe SMUG2

In BLAST search of uracil DNA glycosylase, a putative DNA glycosylase gene from the nonspore-forming, gram-negative bacterium *Pedobacter heparinus* caught our attention. The open reading frame of this protein shares 24% amino acid sequence similarity to family 3 Gme SMUG1 and 22% to family 2 EcMUG, respectively. Its motif 1 is similar to family 2 MUG/TDG as it contains the GINPG sequence (Fig. 1A). On the other hand, motif 2 is more family 3 SMUG1-like as both harbor the conserved histidine residue. Motif 3 shares a CPLXF sequence (X is any amino acid) in the middle with family 3 SMUG1 but lacks the conserved asparagine residue in the first position. In family 2 MUG/TDG, this position is also not conserved (Fig. 1A). Using this ORF to BLAST Genbank, we found putative genes containing the same type of motifs in firmicutes, spirochaetes, chlorobium, bacteroidetes, cyanobacteria and proteobacteria. No homologs were found in archaea but one ORF was found in the eukaryotic organism *Ancylostoma Ceylanicum* (roundworm, Genbank accession no. EPB65414.1). To better define the relationship between these putative DNA glycosylases and other known families in UDG superfamily, we conducted a phylogenetic analysis including all six previously known families. This and other similar genes were placed in a clade along with the known family 3 SMUG1 (Fig. 1B). As such, we named them as a family 3 SMUG2.

To understand its functions, we cloned and expressed Phe SMUG2, and analyzed its UDG activities using fluorescently labeled oligonucleotide substrates containing all three deaminated bases, hypoxanthine (I), uracil (U), and xanthine (X) (Fig. 2A). Under the assay conditions, Phe SMUG2 exhibited activity in excising all uracil-containing substrates except for single-stranded uracil-containing DNAs (Fig. 2B). The enzyme also showed glycosylase activity on hypoxanthine- and xanthine-containing DNA. Similar to its UDG activity, Phe SMUG2 was not active towards ss-DNA containing hypoxanthine or xanthine. To quantitatively measure the glycosylase activity, we determined the kinetic parameters. Because UDG enzymes with the exception of family 1 UNGs are slow enzymes that require enzyme in excess for the detection of glycosylase activity [43], we determined K_m and k_2 values using the methods as previously described (Fig. 3) [33, 44]. The K_m values of Phe SMUG2 varied from 0.51 μM and 0.61 μM for G/U and C/X to 1.08 μM for A/U and 2.90 μM for T/I (Table 2). On the other hand, the k_2 values were highest with G/U and lowest with T/I (Table 2). As judged by k_2/K_m values, Phe SMUG2 was most efficient on and specific for G/U base pairs, intermediate on A/U and C/X base pairs and least efficient on T/I base pairs (Table 2).

Mutational analysis of Phe SMUG2

The UDG superfamily members all contain three motifs that are involved in base recognition and catalysis. As mentioned above, the three motifs in Phe SMUG2 show unique hybrid features in comparison with other UDG families. To explore the relationship between the key residues and their catalytic properties, we created the I62M, N63A, G65Y, S124N and H205S mutants through site-directed mutagenesis, and characterized their kinetic constants using various DNA substrates (Table 2). I62 is in the position equivalent to I17 in EcMUG, and its main chain points to O2 of uracil [18, 31].

I62M reduced k_2/K_m for A/U and G/U substrates by 604-fold and 516-fold, for T/I by 98-fold and for C/X by 11-fold, respectively. N63 is in the position equivalent to N18 in family 2 EcMUG and N58 in family 3 Gme SMUG1, which is assumed to position a water molecule to attack the glycosidic bond [18, 45]. N63A reduced k_2/K_m for A/U and G/U substrates by 21-fold and 13-fold, for T/I by 11-fold and for C/X by 52-fold, respectively. G65 is in a highly conserved position as seen in families 2 and 3 enzymes (Fig. 1A). However, this position in family 1 UNG is occupied by a tyrosine residue. The G65Y mutant lost all glycosylase activity beyond the lower detection limit under our assay conditions (data not shown). S124 is the first position in motif 3, which is highly conserved in families 3 and 1 (Fig. 1A). When S124 is substituted by an asparagine residue, the k_2/K_m values were increased 71-fold and 298-fold for A/U and G/U, 8.6-fold for T/I, and reduced 141-fold for C/X, respectively. More interestingly, S124N gained weak catalytic activity on both single-stranded uracil-containing substrate and the G/T base pair, while the wild type enzyme and other mutants did not show this type of activity. H205 is in the first position in motif 2, which is highly conserved in several families except for family 2 MUG/TDG (Fig. 1A). When H205 was substituted by a serine residue, it exhibited immense effect on UDG activity as H205S reduced its activity on A/U base pair to beyond detection and reduced the k_2/K_m value for G/U by almost 10000-fold (Table 2). On the other hand, H205S had little effect on HDG and XDG activities. The implication of mutational analysis will be discussed later.

Crystal structures of the WT and G65Y proteins

In order to fully study Phe SMUG2, we solved the crystal structures of the apoenzyme, and the G65Y mutant. The full-length wild type enzyme contains 244 residues, including the C terminal his6-tag. The space group belongs to the $P4_12_12$ space group and the native

crystal diffracted to a high resolution of 1.8 Å with a completeness of 99.9% (Table 1). However, molecular replacement using several homologous structures failed to produce a plausible solution. Therefore, we crystallized SeMet-labeled protein and carried out experimental phasing with single-wavelength anomalous dispersion (SAD) phasing, using a 2.7-Å dataset collected at the peak wavelength for SeMet (0.9793 Å). After the determination of the SeMet sites, density modification was carried out until an interpretable map was obtained. The high-resolution native dataset was merged in to improve the map and the full-length protein model was manually built. Each asymmetric unit contains only one subunit, with the refined model containing 234 visible residues. Similar to Gme SMUG1, the protein forms a monomer in solution, as indicated by the size-exclusion chromatography (data not shown). The finished model also contains 176 water molecules (Table 1). All the residues are in good geometry with no residues falling in the outlier region. The protein displays a spherical shape, with a central four-strand β-sheet surrounded by 11 α-helices (Fig. 4A).

A DALI search retrieves several structural homologs for Phe SMUG2, with xSMUG1 and Gme SMUG1 on top of the list (PDBs 1OE6, 1OE5 5H98, and 5H93) (Table 3). While the central β-sheets align well among these proteins, the surrounding helices are generally off by 3.0~5.0 Å in position (Fig. 4B), which explains the failure to obtain a solution using the homologous structures. The RMSD is 2.5 Å over 196 Cα atoms between Phe SMUG2 and Gme SMUG1, indicating divergent evolution between two types of SMUG families. Of note, four regions in Phe SMUG2 display relatively large structure variations. The first region is the N-terminus, where α1 is significantly shorter than its counterparts. The second region is the C86-L94 fragment, which forms a much shorter loop, as compared to the long loops formed in SMUG1 (Fig. 4B). Third, the

G129-Y142 region forms a pair of long antiparallel β–hairpin whereas SMUG1 only forms shorter loops. Last but not least, the P206-K213 fragment, which contains motif 2, forms a short helix while the corresponding region in SMUG1s lacks obvious secondary structural elements. These differences are generally located in the regions where sequence deletion or insertions are observed (Fig. 4C). Despite these differences, the general fold of the enzymes is preserved well. Interestingly, at motif 1, a major site that distinguishes the two types SMUG1s from each other by sequence, the local structures superimpose quite well.

To understand the inactivity of the G65Y mutant, we determined its structure at 2.25 Å. The structure of the mutant is nearly completely identical to that of WT, with a RMSD 0.11Å over 223 Cαs (Fig. 5A). The only evident exception is that at position 65, the tyrosine points up to the entrance of the substrate-access channel (Fig. 5B). Another difference is the antiparallel β3-β4 hairpin structure in region 3 is much shorter in the mutant. The structural consequences of this mutation will be discussed later.

DISSCUSSION

Enzymes in UDG superfamily are ubiquitous in three domains of life, bacteria, archaea and eukaryotes. Even though six families have been found in nature, new families with unique properties are still emerging. This study focuses on a new class of enzymes that has a unique combination of the three motifs. Because of their importance in base recognition and catalysis, the sequence and structural properties of these motifs have great impact on the enzymatic activity of UDG enzymes. Phylogenetically, family 3 SMUG2 is more similar to family 3 SMUG1 than to other families. In terms of glycosylase activity, it is similar to SMUG1 as both classes of enzymes contain UDG and XDG activity. However, Phe SMUG2 does not seem to act on single-stranded substrates

while Gme SMUG1 can [23]. In addition, Phe SMUG2 contains HDG activity but Gme SMUG1 does not. Structurally, Phe SMUG2 is also more closely related to Gme SMUG1 than other UDG families. The overall glycosylase fold is conserved in SMUGs with all the secondary elements in both structures nearly identical. In addition, the central four-strand β-sheet superimpose well (Fig. 4B). However, the backbone traces of the surrounding helices display minor movements (generally 3-5 Å). In addition, there are distinct differences in four regions, which may confer substrate preferences or activity differences to the enzymes.

To probe the amino acids that are either notably different between SMUG2 and SMUG1 or presumably important for its catalytic function, we selected several positions within these motifs for mutational study. In motif 1, the SMUG2 enzymes contain a family 2 like GINPG sequence (Fig. 1A). The I62M mutant, which converts this segment to family 3 SMUG1 like, has great impact on UDG activity (Table 2). These results suggest that the SMUG2 enzymes have adapted to having a hydrophobic isoleucine in this segment (Fig. 6A). On the other hand, the long side chain of M57 appears to be a better fit for the active site of SMUG1 enzymes (Fig. 6B). N63 is a conserved residue in both family 2 and family 3 enzymes, presumably by positioning a water molecule for attacking the glycosidic bond. The N63A mutation causes a 10- to 50-fold decrease in k_2/K_m value (Table 2). These results suggested that the catalytic role of N63 is also important for SMUG2 enzymes. The substantially increased UDG activity seen in Phe SMUG2 S124N mutant is in keeping with our previous analysis of the EcMUG-K68N mutant [46]. The K68N mutation allows the EcMUG to gain catalytic activity on A/U base pairs and to increase UDG activity overall. Here, the S124N mutation allows Phe SMUG2 to have a similar effect. The difference is that S124N even gain catalytic activity

on single-stranded uracil-containing DNA and G/T base pairs. Similar to EcMUG-K68N, it is likely that the asparagine residue uses its side chain to form bidentate hydrogen bonds with N3 and O4 of uracil (Fig. 6C). Likewise, the similar bidentate hydrogen bonds mediated by N136 in Gme SMUG1 enhance its UDG activity (Fig. 6B-C). The first histidine residue in motif 2 plays a critical role in UDG activity of Gme SMUG1 by forming a short-distance hydrogen bond with O2 of uracil [47] (Fig. 6B). The dramatic reduction in UDG activity suggests that H205 is likely to play a similar role in Phe SMUG2.

The G65Y mutant loses DNA glycosylase activity on all the tested substrates under the assay conditions. To better understand the mutational effects of G65Y, we generated a model of the Phe SMUG2-DNA complex by superimposing apo Phe SMUG2 onto DNA-bound hTDG (PDB 4Z47) [48]. The structural alignment of Phe SMUG2 and hTDG is shown in Fig. 5C. Similar to the superimposition pattern with SMUG1s, the structural variations at the same regions are seen, and the idiosyncratic β3-β4 hairpin insertion in Phe SMUG2 reaches out to the solvent (Fig. 4B-C). However, the core catalytic domain is conserved and these structural differences do not constitute any obstacles to the binding of DNA substrates from a structural point of view. According to the modeled structure, the bulky side chain of Y65 is very close to the flipped out ribose at the AP site (Fig. 5D). We reason that upon the binding of the DNA substrate, some key catalytic residues may be misaligned due to the local structural distortion caused by Y65 and catalysis is thus inhibited. Even so, the DNA-binding pattern is preserved in the G65Y mutant due to the general structural resemblance between the mutant and WT. Indeed, the G65Y mutant still maintains specific binding affinity to DNA containing

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

deaminated bases as determined by gel mobility shift analysis, with K_d values ranging from 108 nM for G/U, 143 nM for C/I and 11 nM for C/X (Fig. 7).

In addition, the base opposite to the uracil is likely to make contacts with the motif-2 containing region, which is a helix (P206-K213, α10) (Fig. 5E). This helix (as compared to a loop region in Gme SMUG1) inserts into the base stack, and the rigidity conferred by the helix may bring about catalytic differences during glycosylase reactions. Consequently, it may play a role in disrupting the base pair and may facilitate the base-flipping process. As a matter of fact, the helix is very close to the minor groove, and it is anticipated to undergo structural arrangements to avoid steric clashes.

Family 1 UNG has a conserved tyrosine residue in the equivalent position (Y66 in *E. coli* UNG), but the UNG enzymes are extremely robust in the removal of uracil. To compare the structural difference between SMUG2 and family 1 UNG, we further superimposed the *E. coli* UNG structure (PDB 1UDG) [29] with that of G65Y (Fig. 5F). It is clear that the orientation of this residue is different (compare Fig. 5D with Fig. 5G). The corresponding tyrosine Y66 tilts in such a way that it would not interfere with the catalytic function as G65Y in Phe SMUG2 does. Instead, Y66 in *E. coli* UNG likely blocks the interactions with thymine due to steric hindrance with the C5 methyl group.

In summary, this study presents a biochemical and structural analysis of a new subfamily of enzymes in UDG superfamily. Family 3 SMUG2 as represented by the homolog from *P. heparinus* is a new subfamily of enzymes with UDG, HDG and XDG activities. As evidenced from the comparison of the SMUG2 and SMUG1 subfamilies, while the overall structure and catalytic residues are conserved, the different combination of the three motifs and subtle sequence variations underlie the differences in specificity and catalytic efficiency. Globally, our structural superposition shows that although both

subfamilies are structurally conserved in terms their main chain traces as well as the orientation of the side, Phe SMUG2 is different from Gme SMUG1 as insertions/deletions occur in several regions between the two enzymes. The contribution of these three structurally distinct regions to enzyme activity requires further investigation. As clear from the mutational analysis presented in this work, amino acid substitutions in the active site can have a major impact on specificity and catalytic efficiency. Overall, this study demonstrates a new evolutionary path in the divergence of UDG superfamily.

ACKNOWLEDGEMENTS

We thank Shanghai Synchrotron Radiation Facility (SSRF) for access to the BL17U1 beamlines, and Dr. Defeng Li for his help on structure determination.

FUNDING

This work was supported by Fundamental Research Funds for the Central Universities 16lgjc76, the Science and Technology Program of Guangzhou 201504010025, Foundation of State Key Laboratory for Biocontrol~~Key Laboratory of Gene Engineering of the Ministry of Education~~ 201~~65~~02, and Guangdong Innovative Research Team Program NO. 2011Y038. This work was also supported in part by the National Institutes of Health (GM090141 to W.C.).

AUTHOR CONTRIBUTIONS

Wei Xie and Weiguo Cao conceived and designed research; Panjiao Pang, Ye Yang and Jing Li performed research; Wei Xie, Weiguo Cao and Zhong Wang analyzed data and wrote the paper. All authors reviewed the manuscript.

ACESSION CODES

Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 5H0J and 5H0K.

ADDITIONAL INFORMATION

The authors declare no competing financial interests.

REFERENCES

1 Lindahl, T. (1993) Instability and decay of the primary structure of DNA. Nature. **362**, 709-715.

2 Spencer, J. P., Whiteman, M., Jenner, A. and Halliwell, B. (2000) Nitrite-induced deamination and hypochlorite-induced oxidation of DNA in intact human respiratory tract epithelial cells. Free Radic Biol Med. **28**, 1039-1050.

3 Samar B, Jennifer LC, Jacquin C, John SW and Tannenbaum, S. R. (1999) The chemistry of DNA damage from nitric oxide and peroxynitrite. Mutat Res. **424**, 37-39

4 Suzuki T, Yamaoka R, Nishi M, Ide H and M., K. (1996) Isolation and Characterization of a Novel Product, 2‘-Deoxyoxanosine, from 2‘-Deoxyguanosine, Oligodeoxynucleotide, and Calf Thymus DNA Treated by Nitrous Acid and Nitric Oxide. J Am Chem Soc. **118**, 2515-2516

5 Lucas, L. T., Gatehouse, D. and Shuker, D. E. (1999) Efficient nitroso group transfer from N-nitrosoindoles to nucleotides and 2'-deoxyguanosine at physiological pH. A new pathway for N-nitrosocompounds to exert genotoxicity. J Biol Chem. **274**, 18319-18326

6 Duncan, B. K. and Miller, J. H. (1980) Mutagenic deamination of cytosine residues in DNA. Nature. **287**, 560-561

7 Coulondre, C., Miller, J. H., Farabaugh, P. J. and Gilbert, W. (1978) Molecular basis of base substitution hotspots in Escherichia coli. Nature. **274**, 775-780

8 Gerald EW, Timothy RO and T., J. (2003) Stability, Miscoding Potential, and Repair of 2′-Deoxyxanthosine in DNA: Implications for Nitric Oxide-Induced Mutagenesis. Biochemistry. **42**, 3608-3616

9 Kamiya, H., Sakaguchi, T., Murata, N., Fujimuro, M., Miura, H., Ishikawa, H., Shimizu, M., Inoue, H., Nishimura, S., Matsukage, A. and et al. (1992) In vitro replication study of modified bases in ras sequences. Chem Pharm Bull **40**, 2792-2795

10 Wink, D. A., Kasprzak, K. S., Maragos, C. M., Elespuru, R. K., Misra, M., Dunams, T. M., Cebula, T. A., Koch, W. H., Andrews, A. W., Allen, J. S. and et al. (1991) DNA deaminating ability and genotoxicity of nitric oxide and its progenitors. Science. **254**, 1001-1003

11 Suzuki T, Yoshida M, Yamada M, Ide H, Kobayashi M, Kanaori K, Tajima K and KeisukeT. (1998) Misincorporation of 2′-Deoxyoxanosine 5′-Triphosphate by DNA Polymerases and Its Implication for Mutagenesis. Biochemistry. **37**, 11592-11598

12 Krokan, H. E., Drablos, F. and Slupphaug, G. (2002) Uracil in DNA--occurrence, consequences and repair. Oncogene. **21**, 8935-8948

13 Clifford, D., Andrew, S. A., Geir, S., Bodil, K., Ingrun, A., Hans, E. K. and John, A. T. (1995) Crystal Structure and Mutational Analysis of Human UraciI-DNA Glycosylase: Structural Basis for Specificity and Catalysis. Cell. **80**, 869-878

14 Huffman, J. L., Sundheim, O. and Tainer, J. A. (2005) DNA base damage recognition and removal: new twists and grooves. Mutat Res. **577**, 55-76

15 Pearl, L. H. (2000) Structure and function in the uracil-DNA glycosylase superfamily. Mutat Res. **460**, 165-181

16 Lindahl, T. (1974) 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. **71**, 3649-3653

17 Cortazar, D., Kunz, C., Saito, Y., Steinacher, R. and Schar, P. (2007) The enigmatic thymine DNA glycosylase. DNA repair. **6**, 489-504

18 Tracey, E. B., Renos, S., George, P., Tom, B. A., Tom, B. R., Josef, J. R. and Laurence, H. P. (1998) Crystal Structure of a G:T/U Mismatch-Specific DNA Glycosylase: Mismatch Recognition by Complementary-Strand Interactions. Cell. **92**, 117-129

19 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

20 Dong, L., Mi, R., Glass, R. A., Barry, J. N. and Cao, W. (2008) Repair of deaminated base damage by Schizosaccharomyces pombe thymine DNA glycosylase. DNA repair. **7**, 1962-1972
21 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
22 Pettersen, H. S., Sundheim, O., Gilljam, K. M., Slupphaug, G., Krokan, H. E. and Kavli, B. (2007) Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. Nucleic acids research. **35**, 3879-3892
23 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. Journal of molecular biology. **385**, 761-778
24 Hoseki, J., Okamoto, A., Masui, R., Shibata, T., Inoue, Y., Yokoyama, S. and Kuramitsu, S. (2003) Crystal structure of a family 4 uracil-DNA glycosylase from Thermus thermophilus HB8. Journal of molecular biology. **333**, 515-526
25 Hinks, J. A., Evans, M. C., De Miguel, Y., Sartori, A. A., Jiricny, J. and Pearl, L. H. (2002) An iron-sulfur cluster in the family 4 uracil-DNA glycosylases. J Biol Chem. **277**, 16936-16940
26 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
27 Kosaka, H., Hoseki, J., Nakagawa, N., Kuramitsu, S. and Masui, R. (2007) Crystal structure of family 5 uracil-DNA glycosylase bound to DNA. Journal of molecular biology. **373**, 839-850
28 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
29 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
30 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
31 Tracey EB, Orlando DS, Renos S, Tom B, Josef J, Gregory LV and HPearl., L. (1999) Crystal structure of a thwarted mismatch glycosylase DNA repair complex. EMBO J. **18**, 6599-6609
32 Fisher, C. L. and Pei, G. K. (1997) Modification of a PCR-based site-directed mutagenesis method. BioTechniques. **23**, 570-571, 574
33 King, K., Benkovic, S. J. and Modrich, P. (1989) Glu-111 is required for activation of the DNA cleavage center of EcoRI endonuclease. J Biol Chem. **264**, 11807-11815
34 Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. Bioinformatics. **23**, 2947-2948
35 Pei, J., Kim, B. H. and Grishin, N. V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic acids research. **36**, 2295-2300
36 Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution. **30**, 2725-2729
37 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276**, 307-326
38 Read, R. J. and McCoy, A. J. (2011) Using SAD data in Phaser. Acta Crystallogr D Biol Crystallogr. **67**, 338-344
39 Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr. **68**, 352-367

40 Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr. **66**, 486-501
41 Chen, V. B., Arendall-III, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. **66**, 12-21
42 Yang, Y., Kucukkal, T. G., Li, J., Alexov, E. and Cao, W. (2016) Binding Analysis of Methyl-CpG Binding Domain of MeCP2 and Rett Syndrome Mutations. ACS chemical biology. Accepted.
43 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
44 Vermote, C. L. and Halford, S. E. (1992) EcoRV restriction endonuclease: communication between catalytic metal ions and DNA recognition. Biochemistry. **31**, 6082-6089
45 Zhang, Z., Shen, J., Yang, Y., Li, J., Cao, W. and Xie, W. (2016) Structural Basis of Substrate Specificity in Geobacter metallireducens SMUG1. ACS chemical biology. **11**, 1729-1736
46 Lee, D. H., Liu, Y., Lee, H. W., Xia, B., Brice, A. R., Park, S. H., Balduf, H., Dominy, B. N. and Cao, W. (2015) A structural determinant in the uracil DNA glycosylase superfamily for the removal of uracil from adenine/uracil base pairs. Nucleic acids research. **43**, 1081-1089
47 Drohat, A. C. and Stivers, J. T. (2000) Escherichia coli uracil DNA glycosylase: NMR characterization of the short hydrogen bond from His187 to uracil O2. Biochemistry. **39**, 11865-11875
48 Malik, S. S., Coey, C. T., Varney, K. M., Pozharski, E. and Drohat, A. C. (2015) Thymine DNA glycosylase exhibits negligible affinity for nucleobases that it removes from DNA. Nucleic Acids Res. **43**, 9541-9552

FIGURE LEGENDS

Figure 1. Sequence alignment and phylogenetic analysis of Phe SMUG2. **(A).** Sequence alignment of Phe SMUG2 with known glycosylase in UDG superfamily. GenBank accession numbers are shown after the species names. Phe SMUG2: Phe, *Pedobacter heparinus* DSM 2366, WP_012780920.1; Cth, *Chloroherpeton thalassium*, WP_012500224.1; Csp, *Chlorobium* sp. GBChlB, KER10671.1; Ace, *Arenibacter certesii*, WP_026813179.1. Family 3 (SMUG1): Gme, *G. metallireducens* GS-15, YP_383069, Rba, *R. baltica* SH 1, NP_869403, Hsa, *Homo sapiens*, NP_055126, Xla, *X. laevis*, AAD17300. Familiy 2 (TDG/MUG): Eco, *E. coli*, P0A9H1; *Hsa*, *H. sapiens*, NP_003202; Gga, *Gallus gallus*, NP_990081.1; Pan, *Pantoea ananatis LMG 20103*, ADD78558.1. Family 1 (UNG): *Eco*, *E. coli*, NP_289138, Family 4 (UDGa): *Tth*, *T. thermophilus* HB27, YP_004341.1. Family 5 (UDGb): *Tth*, *T. thermophilus* HB8, YP_144415.1. Family 6 (HDG): *Mba*, *Methanosarcina barkeri str. Fusaro*, YP_304295.1. (B). Phylogenetic tree of the UDG superfamily. The phylogenetic analysis was performed using the neighbor-joining method in MEGA 6. GenBank accession numbers are shown after the species names. SMUG2: Phe, *Pedobacter heparinus* DSM 2366, WP_012780920.1; Cth, *Chloroherpeton thalassium*, WP_012500224.1; Csp, *Chlorobium* sp. GBChlB, KER10671.1; Ace, *Arenibacter certesii*, WP_026813179.1. Family 3 (SMUG1): Gme, *G. metallireducens* GS-15, YP_383069, Rba, *R. baltica* SH 1, NP_869403, Hsa, *Homo sapiens*, NP_055126, Xla, *X. laevis*, AAD17300. Familiy 2 (TDG/MUG): *Hsa*, *H. sapiens*, NP_003202; Gga, *Gallus gallus*, NP_990081.1; Eco, *E. coli*, P0A9H1, Pan, *Pantoea ananatis LMG 20103*, ADD78558.1. Family 1 (UDG): *Eco*, *E. coli*, NP_289138; *Hin*, *Haemophilus influenzae* KR494, YP_008544610.1; *Mtu*, *Mycobacterium tuberculosis*, WP_003908950.1; *Hsa*, *H. sapiens*, NP_003353. Family 4

(UDGa): *Tth, T. thermophilus* HB27, YP_004341.1. ; *Pae, P. 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, T. thermophilus* HB8, YP_144415.1; *Pae, P. aerophilum* str. IM2, NP_559226; *Tvo, Thermoplasma volcanium* GSS1, NP_111346.1; *Mtu, M. 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. DNA glycosylase activity of wild type Phe SMUG2. (A). The chemical structures of DNA deaminated bases (uracil (U)-, hypoxanthine- (I) and xanthine (X)) and the oligodeoxyribonucleotide substrate **s**equences containing these bases. **(B).** DNA glycosylase activity of wt Phe SMUG2 on U-, X-, and I-containing substrates. Cleavage reactions were performed as described in Materials and Methods with 250 nM wt Phe SMUG2 protein and 10 nM substrate and incubated at 37°C for 60 min. Data are the average of three independent experiments.

Figure 3. Representative fitting curves for kinetics analysis. To determine the kinetics parameters of WT SMUG2 and the mutant enzymes with the G/U, T/I, C/X, A/U, ssU and G/T substrates, the reactions were carried out as described in Materials and Methods. Samples were withdrawn at indicated time points. **(A).** Wild type SMUG2 with the G/U substrate. **(B).** Wild type SMUG2 with the T/I substrate. **(C).** Wild type SMUG2 with the C/X substrate. **(D).** SMUG2-S124N with the G/T substrate.

Figure 4. The overall structure of apo-Phe SMUG2 and comparison to SMUG1s. (**A**). The overall structure of apo-Phe SMUG2. The helices, strands and loops are shown in red, yellow and green respectively. The N- and C-termini are labeled. (**B**). Superposition with

SMUG1s in wall-eyed stereo view. Phe SMUG2: pale green; xSMUG1: yellow orange; Gme SMUG1: pale cyan. The four regions of significant differences in Phe SMUG2 are colored red and indicated by the blue ovals. **(C).** The sequence alignment of Phe SMUG2 with two SMUG1s, with the secondary structure shown on top of the sequence. The four different regions are boxed.

Figure 5. The structure of Phe SMUG2-G65Y and the possible structural perturbation of the mutation. (A). The overall structure of the G65Y mutant. **(B).** The structural overlay of the G65Y mutant and the WT enzyme. The blue oval indicates the β3-β4 hairpin with large local structure differences. **(C)**. A structural comparison of Phe SMUG2-G65Y to human TDG structure by superposition of Phe SMUG2-G65Y (5H0K) onto human TDG in its DNA-bound form (PDB 4Z47). (**D**). A model of the G65Y-dsDNA complex. The coordinates of dsDNA are taken from the DNA substrate in complex with human TDG, which contains an AP site (PDB 4Z47). The red circle indicates the structural incompatibility between Y65 and the ribose, while the red arrow indicates the helix inserted into the base stacks (α10). (**E**). View of (a) from a 90-degree angle. (**F**). A model of the EcUNG-dsDNA complex, where Y66 tilts to another direction and accommodates dsDNA substrates (EcUNG coordinates from PDB 2EUG).

Figure 6. The comparison of possible recognition modes by different SMUGs. The critical residues for recognition and catalysis are labeled and shown in ball-and-sticks model. The substrate uracil is modeled in by superposition of each individual protein structure with the uracil-containing xSMUG1 structure (PDB 1OE5). The side chain orientation of S124N is modeled by COOT. The possible hydrogen bonds formed by the histidine and asparagine residues are shown by the red dashed lines. (**A**). The binding

mode of uracil to Phe SMUG2-WT; (B) to Gme SMUG1-WT; (C) to Phe SMUG2-S124N.

Figure 7. Representative images of binding analysis of Phe SMUG2 G65Y mutant with damaged DNA substrates. Binding reactions were performed as described in Materials and Methods with 5 nM DNA and indicated amounts of Phe SMUG2 G65Y protein. Protein concentrations used were 0 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 300 nM, 500 nM, 1000 nM, 2000 nM and 4000 nM. (**A)**. Phe SMUG2 G65Y with G/U-containing DNA. (**B).** Phe SMUG2 G65Y with T/I-containing DNA. (**C).** Phe SMUG2 G65Y with C/X-containing DNA. (**D).** Phe SMUG2 G65Y with undamaged DNA as a control.

Table 1. Data collection and refinement statistics.

	WT (5H0J)	G65Y (5H0K)	SeMet
Data collection		SSRF BL17U1	
Wavelength	0.9791	0.9791	0.9793
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a, b, c (Å)	56.51, 56.51, 144.24	56.71, 56.71, 144.54	58.22, 58.22, 144.48
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50.0-1.80 (1.86-1.80)[a]	50.0-2.25 (2.33-2.25)	50.0-2.70 (2.80-2.70)
R_{merge}[b] (%)	0.109 (0.633)	0.161 (0.639)	0.162 (0.884)
$I/\sigma_{(I)}$	26.3 (4.7)	19.7 (8.5)	53.9 (11.1)
Completeness (%)	99.9 (100)	99.9 (99.8)	99.7 (100)
Redundancy	13.8 (14.4)	8.6 (8.7)	12.0 (12.0)
Refinement			
Resolution (Å)	39.97-1.80 (1.88-1.80)	38.65-2.25 (2.47-2.25)	
No. reflections	22437	11928	
R_{work}[c]/R_{free}[d]	0.180/ 0.197	0.201/0.246	
No. atoms			
Protein	1876	1900	
Water	176	147	
B-factors (Å^2)			
Protein	27.30	22.22	
Water	36.11	27.22	
R.m.s deviations			
Bond lengths (Å)	0.006	0.006	
Bond angles (°)	0.75	0.95	
Ramachandran favored (%)	98.3	98.3	
Allowed (%)	1.7	1.7	
Outliers (%)	0	0	

[a]Values in parentheses are for the highest-resolution shell. [b]$R_{merge} = \Sigma |(I - < I >)|/\sigma(I)$, where I is the observed intensity. [c]$R_{work} = \Sigma_{hkl} ||Fo| - |Fc||/ \Sigma_{hkl} |Fo|$, calculated from working data set. [d]R_{free} is calculated from 5.0% of data randomly chosen and not included in refinement.

Table 2. Kinetic constants of Phe SMUG2 DNA glycosylase[a]

Enzyme	Substrate	K_m (μM)	k_2 (min^{-1})	k_2/K_m (min^{-1} μM^{-1})
WT	A/U	1.08 ± 0.13	1.57 ± 0.08	1.45
	G/U	0.51 ± 0.08	34.20 ± 1.68	67.1
	T/I	2.90 ± 0.32	0.21 ± 0.02	0.07
	C/X	0.61 ± 0.07	3.53 ± 0.11	5.79
I62M	A/U	0.17 ± 0.02	0.00040 ± 0.00003	0.0024
	G/U	0.21 ± 0.09	0.027 ± 0.005	0.13
	T/I	0.28 ± 0.06	0.0002 ± 0.0001	0.00071
	C/X	0.20 ± 0.05	0.101 ± 0.021	0.51
N63A	A/U	0.56 ± 0.09	0.038 ± 0.003	0.068
	G/U	0.36 ± 0.08	1.92 ± 0.12	5.33
	T/I	0.24 ± 0.03	0.0015 ± 0.0001	0.0063
	C/X	0.34 ± 0.08	0.038 ± 0.003	0.11
S124N	A/U	0.32 ± 0.13	32.50 ± 3.52	101.56
	G/U	0.25 ± 0.09	287.5 ± 9.3	1150.00
	ssU	0.68 ± 0.19	0.056 ± 0.011	0.082
	T/I	0.91 ± 0.13	0.055 ± 0.003	0.060
	C/X	0.66 ± 0.03	0.027 ± 0.004	0.041
	G/T	0.92 ± 0.16	0.011 ± 0.002	0.012
H205S	A/U	N.A	N.A	N.A
	G/U	0.61 ± 0.15	0.0043 ± 0.0005	0.0070
	T/I	1.47 ± 0.32	0.17 ± 0.01	0.12
	C/X	0.36 ± 0.05	2.14 ± 0.09	5.94

[a.] The reactions were performed as described in Materials and Methods under DNA glycosylase activity assay. Data are an average of at least three independent experiments.

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

Table 3. The structural homologs of Phe SMUG2 retrieved by the Dali search, with decreasing structural similarities. N_{align} represents aligned residues, while $\%_{seq}$ and $\%_{sse}$ represent the sequence coverage and secondary structure identity in percentages.

	Q	P	Z	RMSD	Nalign	Ng	%seq	%sse	Match	%sse	Nres	PDB entry titles
1	0.39	6.8	8.5	2.54	196	11	17	76	1oe6:B	72	245	Xenopus smug1, an anti-mutator uracil-DAN glycosylase
2	0.37	6.0	7.9	2.71	184	16	14	82	5h98:A	82	215	Crystal structure of *Geobacter metallireducens* smug1
3	0.36	7.3	8.7	2.60	189	12	17	82	1oe5:A	74	243	Xenopus smug1, an anti-mutator uracil-DAN glycosylase
4	0.36	7.2	8.7	2.59	189	13	17	76	1oe6:A	76	244	Xenopus smug1, an anti-mutator uracil-DAN glycosylase
5	0.36	4.9	7.9	2.57	188	14	18	71	1oe5:B	71	245	Xenopus smug1, an anti-mutator uracil-DAN glycosylase
6	0.36	5.1	7.5	2.84	193	14	15	82	5h93:D	74	236	Crystal structure of *Geobacter metallireducens* smug1
7	0.35	5.3	7.7	2.83	189	11	15	76	5h99:A	72	228	Crystal structure of *Geobacter metallireducens* smug1 mutant N58D
8	0.35	5.1	7.6	2.96	179	13	16	76	5h98:B	81	197	Crystal structure of *Geobacter metallireducens* smug1
9	0.35	6.1	8.0	2.84	193	12	14	76	5h9i:A	72	240	Crystal structure of *Geobacter metallireducens* smug1 with Xanthine
10	0.34	5.5	7.8	2.74	184	13	14	76	5h99:B	72	229	Crystal structure of *Geobacter metallireducens* smug1 mutant N58D

Formatted Table