


DR0022 from *Deinococcus radiodurans* is an acid uracil-DNA glycosylase

Jing Li, Ye Yang, Chenyan Chang and Weiguo Cao 

Department of Genetics and Biochemistry, Clemson University, Clemson, SC, USA

Keywords

deamination; DNA glycosylase; DNA repair; radiation; uracil

Correspondence

W. Cao, Department of Genetics and Biochemistry, Clemson University, Room 049 Life Sciences Facility, 190 Collings Street, Clemson, SC 29634, USA
Tel: +1 864 656 4176
E-mail: wgc@clemson.edu

(Received 11 March 2022, revised 8 April 2022, accepted 23 May 2022)

doi:10.1111/febs.16533

Uracil-DNA glycosylase (UDG) initiates base excision repair (BER) by removing damaged or modified nucleobases during DNA repair or mammalian demethylation. The UDG superfamily consists of at least six families with a variety of catalytic specificities and functions. *Deinococcus radiodurans*, an extreme radiation resistant bacterium, contains multiple members of UDG enzymes within its genome. The present study reveals that the putative protein, DR0022, is a uracil-DNA glycosylase that requires acidic conditions for its glycosylase activity, which is the first case of such an enzyme within the UDG superfamily. The key residues in the catalytic motifs are investigated by biochemical, enzyme kinetics, and *de novo* structural prediction, as well as molecular modeling analyses. The structural and catalytic roles of several distinct residues are discussed in light of predicted and modeled DR0022 glycosylase structures. The spontaneous mutation rate analysis performed in a *dr0022* deficient *D. radiodurans* strain indicated that the *dr0022* gene plays a role in mutation prevention. Furthermore, survival rate analysis in a *dr0022* deficient *D. radiodurans* strain demonstrated its role in stress resistance, including γ -irradiation. Additionally, the novel acid UDG activity in relationship to its *in vivo* roles is discussed. This work underscores the functional diversity in the UDG superfamily.

Introduction

In 1974, Uracil-DNA glycosylase (UDG) was discovered in *Escherichia coli* to be involved in removal of uracil, a deamination product generated from hydrolytic deamination and nitrosative stress-induced oxidative deamination of cytosine in DNA [1–3]. Over decades of investigation, at least six families of glycosylase enzymes have been found in the UDG superfamily. Although the prototypical *E. coli* enzyme is classified as a narrow specificity family 1 uracil *N*-glycosylase (UNG), UDGs in other families exhibit a variety of substrate specificities, from family 2 MUG/TDG, family 3 SMUG1, family 5 UDGb as broad specificity enzymes, to family 4 UDGa as narrow specificity enzymes, and to family 6 HDG as

hypoxanthine DNA glycosylase [4–11]. Within each UDG family, the specificity may also vary depending on the source of organisms [12–19]. A notable expansion of specificity with profound biological significance occurred in family 2 MUG/TDG, in which the human TDG becomes a DNA glycosylase involved in DNA demethylation by removing 5-formylcytosine and 5-carboxylcytosine derived from oxidation of 5-methylcytosine by ten-eleven translocation methylcytosine dioxygenase [15,16]. A rather unusual class of UDG enzymes were found originally in *Mycobacterium* that cross links with the DNA after the excision of uracil [20–23]. These results indicate that enzymes in the UDG superfamily are adaptable to acquire new

Abbreviations

BER, base excision repair; Dra, *Deinococcus radiodurans*; PDB, Protein Data Bank; Tth, *Thermus thermophilus*; UNG, uracil *N*-glycosylase; UDG, uracil-DNA glycosylase.

enzymatic functions. In addition to enzymes in the UDG superfamily, endonuclease V and some enzymes in the endonuclease III superfamily are capable of acting on uracil in DNA [24–27].

Deinococcus radiodurans (Dra) is an extremely radiation resistant bacterium initially found in canned meat [28]. The sequenced genome reveals the existence of a multitude of DNA repair enzymes. Four genes for enzymes (DR0689, DR1751, DR0022, DR0715) were found in the sequenced genome as putative UDG enzymes. Previous biochemical investigation determined that DR0689 and DR1751 were authentic family 1 UNG and authentic family 4 UDGA enzymes, respectively, with uracil-DNA glycosylase activity on both double-stranded and single-stranded DNA [29]. DR0715, on the other hand, only showed weak activity on G/U base pair and moderate activity on ethenocytosine-containing DNA [29]. Yet, the UDG activity from DR0715 was later confirmed by a structural and biochemical study [30]. Previous investigation on DR0022 did not find any detectable glycosylase activity, regardless of whether the protein was expressed in *E. coli* or by *in vitro* transcription–translation system [29]. In the present study, we report positive detection of uracil-DNA glycosylase activity in DR0022 and the investigation of its role in mutation reduction and stress resistance, and also discuss its potential catalytic mechanism.

Results

Screen for activity and biochemical analyses

Initially, we tested the DNA glycosylase activity using purified DR0022 protein under our routine assay conditions at pH 7.5 without salt using U, I (hypoxanthine), X (xanthine), and O (oxanine) as substrates (Fig. 1A–D). Consistent with the previous study [29], we failed to detect any glycosylase activity under these assay conditions. We then surmised that this unique UDG may exert its glycosylase activity under different assay conditions. To test whether the lack of activity was a result of the pH, we measured the glycosylase activity in buffers with pH ranging from 4–8 using the U-containing substrate. Unexpectedly, we were able to detect UDG activity around pH 5.5 (Fig. 1D,E). No UDG activity was found at neutral or alkaline pH conditions (Fig. 1D,E). The glycosylase activity was found with three mismatched uracil base pairs in the order G/U > C/U > T/U. No activity was found with the A/U base pair or the single-stranded uracil-containing substrate (Fig. 1D,E). Based on this novel finding, we went back to measure the glycosylase

activity on I, X, O, 5-dihydrouracil, 8-oxoG, thymine glycol, AP site, and 5-hydroxyuracil and 5-hydroxycytosine substrates at pH 5.5. We did not find any activity with these substrates (data not shown). However, we found glycosylase activity on hydroxymethyluracil (hmU) in three mismatched substrates in the order G/hmU > C/hmU > T/hmU. Again, no activity was found with the A/hmU base pair or the single-stranded hmU substrate (Fig. 1F). The acidic glycosylase activity was optimal at 0 mM salt and at 37 °C. To define the kinetic properties of the glycosylase activity, we performed time course analyses using the kinetics methodology used previously [10,31,32]. The K_m values were approximately 1000 nM for the G/U, G/hmU, C/hmU, and T/hmU base pairs, but at lower values for the C/U and T/U base pairs (Table 1). The k_2 value was 0.166 min⁻¹ with G/U, which was 3- to 18-fold higher than for other substrates (Table 1). Overall, the catalytic efficiency as defined by k_2/K_m was higher with uracil-containing substrates than that with hmU-containing substrates (Table 1).

In addition to DR0022, we found putative genes from other bacterial species that have similar motif architecture and grouped them together as DR0022-like family (Fig. 2). Among the different families in UDG superfamily, DR0022 is most similar to family 4 UDGA in the conservation of catalytic residues in motifs 1, 2, and 3 (Fig. 2). A notable difference is that motif 1 in DR0022 starts with LEAP rather than GEGP in *Thermus thermophilus* (Tth) UDGA (Fig. 2). Essentially, DR0022 and DR0022-like UDGs are an exception in the UDG superfamily by starting motif 1 with a leucine rather than a glycine. On the other hand, motifs 2 and 3 are well conserved with an initiating Asn and His, respectively (Fig. 2). To probe the functional role of these motifs in the catalysis of the acid uracil-DNA glycosylase, we constructed five single-point mutants in five positions (Fig. 2). As shown in Table 2, no glycosylase activity was detected with L43G, E44A in motif 1, N90A in motif 3, and H159A in motif 2. Interestingly, A50E mutant lost glycosylase activity on hmU but exhibited an elevated UDG activity. The higher level of UDG activity on G/U base pair was further investigated by kinetics analysis, which showed approximately 2-fold reduction of K_m and doubling of k_2 , resulting in an almost 5-fold increase in k_2/K_m (Table 1). Subsequently, we introduced L43G into the A50E mutant and found that the double mutant had a detectable but lower level of UDG activity compared to A50E mutant and the wild-type enzyme (Table 2). The implication of these results in

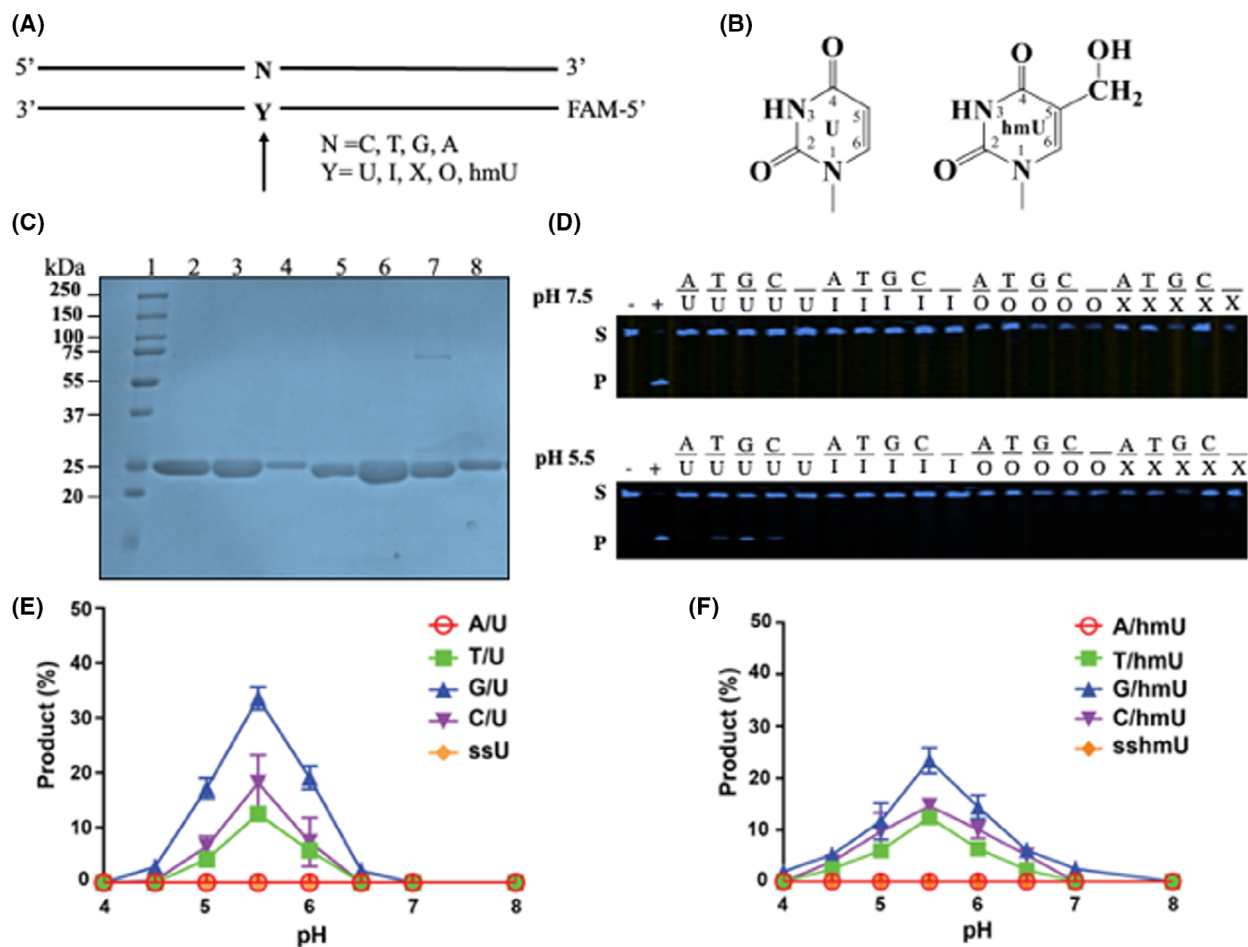


Fig. 1. Deaminated DNA glycosylase activity in DR0022 protein. (A) Scheme of oligodeoxyribonucleotide substrate. (B) Chemical structures of U and hmU. (C) SDS/PAGE (12%) analysis of purified wild-type and mutant DR0022 proteins. Lane 1, protein ladder; lane 2, DR0022-WT protein; lane 3, DR0022-L43G protein; lane 4, DR0022-E44A protein; lane 5, DR0022-A50E protein; lane 6, DR0022-N90A protein; lane 7, DR0022-H159A protein; lane 8, DR0022 L43G-A50E protein. (D) Electrophoresis analysis of glycosylase assay samples on 7 M-urea denaturing sequencing gel (10%). -, Negative control, no enzyme added to the reaction mixture; +, positive control, *E. coli* family 1 UNG was added to the reaction mixture (pH 7.5) in place of DR0022; S, substrate; P, product. (E, F) DNA glycosylase activity of DR0022 on U-containing substrate (E) and hmU-containing substrate (F). Cleavage reactions were performed as described in the [Materials and methods](#) with 100 nM wild-type DR0022 protein and 10 nM substrate. Data are the average of three independent experiments. Error bars indicate the SD, $n = 3$.

relationship to the catalytic mechanism is discussed in detail further below.

In vivo analysis

To understand the role of DR0022 mutation prevention *in vivo*, we deleted the *dr0022* gene from the Dra genome using double crossover recombination of a kanamycin resistance cassette into the genome [33]. The deletion of *dr0022* was confirmed by PCR analysis. We first examined the effect of Dra (*dr0022*⁻) on spontaneous mutation by measuring the mutation rates for the rifampicin resistance. In the presence of the *dr0022* gene, the mutation rate was 4.67×10^{-9} in

Dra (Fig. 3). Once the *dr0022* gene was deleted, the mutation rate jumped to 24.67×10^{-9} , which is more than a 5-fold increase over the wild-type strain. We then investigated the nature of base pair changes in the *rpoB* gene in Dra (*dr0022*⁻). As shown in Table 3, the majority of the base pair changes was A/T to T/A, followed by G/C to A/T, then by A/T to C/G. The A/T transversion (A/T to T/A, A/T to C/G) represented 82% of the substitutions.

Response to stresses

To understand the role of *dr0022* under various stress conditions, we measured survival rates of *dr0022*

Table 1. Kinetic constants of wild-type and A50E DR0022 glycosylase on U- and hmU-containing DNA substrates. Each kinetic constant was determined as described in the [Materials and methods](#). Data are the average of three independent experiments. NA, no activity detected under the assay conditions.

Protein	Substrates	K_m (nM)	k_2 (min ⁻¹)	k_2/K_m (min ⁻¹ nM ⁻¹)
WT	G/U	1257 ± 143	0.166 ± 0.007	1.32 × 10 ⁻⁴
	C/U	320 ± 41	0.052 ± 0.002	1.63 × 10 ⁻⁴
	T/U	396 ± 24	0.039 ± 0.006	9.85 × 10 ⁻⁵
	G/hmU	1080 ± 196	0.021 ± 0.001	1.94 × 10 ⁻⁵
	C/hmU	1640 ± 419	0.014 ± 0.002	8.53 × 10 ⁻⁶
	T/hmU	1034 ± 152	0.009 ± 0.001	8.70 × 10 ⁻⁶
A50E	G/U	667 ± 122	0.41 ± 0.02	6.15 × 10 ⁻⁴
	G/hmU	NA	NA	NA

deficient Dra strain treated with acidified NaNO₂, H₂O₂, UV and γ-irradiation. When treated with acidified NaNO₂, the survival rates of the *dr0022* deficient

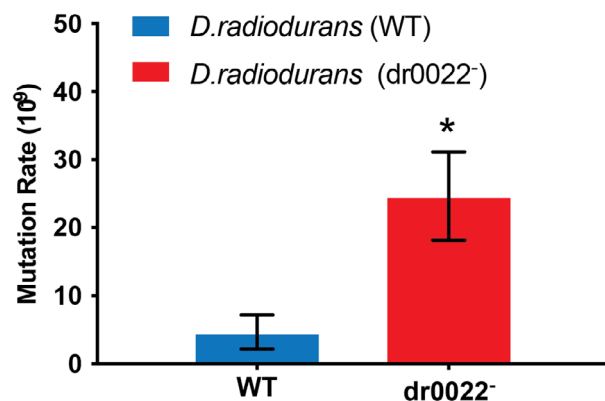
Dra strain and the wild-type strain were readily reduced at a NaNO₂ concentration of 10 mM (Fig. 4 A). At 30 mM, both the *dr0022* deficient and the wild-type strains completely lost survival. Notably, the survival rates of the wild-type strain were statistically lower than the *dr0022* deficient strain at concentrations of 10 and 20 mM (Fig. 4A). The decrease of survival rates was much slower when treated with H₂O₂, with a sharp drop at 10 mM and a slow decline up to 80 mM (Fig. 4B). The lower level of survival in the wild-type strain may suggest active removal of nucleobase lesions, which may introduce more AP sites and/or strand breaks and lead to increase lethality. The survival curves of both strains with UV-C treatment were almost identical, with a 50% reduction at 10 min and an almost 100% reduction at 20 min (Fig. 4C). When treated with γ-irradiation, the survival rates of both strains experienced similar levels of reduction up to 3 kGy (Fig. 4D). The responses of both strains then

		L43	E44	A50	N90	H159	
(DR0022-Like)	Dra	N--42-LEAPGPQAAQSRGGSGFISMDNDDH-22-NIVPWYV-GDD-59-HPSGQAL--34-C	▼	▼	▼	▼	(DR0022)
	Cph	N--65-LEAPGPKAV----ASGFVSRDNPDE-22-NIVPWYI-GTG-58-HPSPQFI-143-C					
	Ses	N--85-LESPGPASTATRAGSGLISLDNDDQ-22-NVVPWYL-----56-HPSPQSI--23-C					
	Mra	N--64-LETPGA-VLRT----GFVTRDSANG-22-NAVPLIHAEG-58-HPSPTYV--26-C					
	Dap	N--59-LESPGP-VVSR---TRFVSMNDPDG-22-NVVPWQM-SEN-54-HPSPQNF--34-C					
	Psy	N--60-LESPARTVSLPR----YVSRDNPDP-22-NLYPWLP-DLD-58-HPSPLSV--22-C					
Family 4 (UDGa)	Tth	N--39-GEFGPEEEDK-TGRP-FVGKAGQLL-17-NIVKCRPPQN--65-HPAYLLR--44-C					
	Pae	N--39-GEAPGASEDE-AGRP-FVGAAGQLL-17-NVVKCRPPNN--72-HPAAVLR--28-C					
	Tma	N--44-GEFGPEEEDK-TGRP-FVGRAGMLL-17-NVVKCRPPNN--66-HPSYLLR--25-C					
	Dra	N--60-GEFGPAEEDR-DGRP-FVQAGQLL-17-NVTKCRAPNN--71-HPAYLLR--49-C					(DR1751)
Nmu	N-123-GEFGPAQEDA-LGEP-FVQAGKLL-18-NIVKCRPPGN--64-HPAYLLR--26-C						
Family 1 (UNG)	Eco	N--61-GQDPYHGPQAHGLA-FSVRPGIAT-37-NTVLTVRAGQ--54-HPSPLSA--36-C					
	Dra	N--80-GQDPYHGNQAHGLS-FSVRPGVRV-37-NAVLTVRAGQ--55-HPSPSE--35-C					(DR0689)
	Mtu	N--65-GQDPYPTPGHAVGLS-FSVAPDVRP-37-NRVLTVRPSN--54-HPSPLSA--30-C					
	Xla	N-142-GQDPYHGNQAHGLC-FSVKPKVPP-37-NAVLTVRAHN--54-HPSPLSV--30-C					
	Hsa	N-142-GQDPYHGNQAHGLC-FSVQRVPP-37-NAVLTVRAHQ--54-HPSPLSV--30-C					
	HSV1	N--85-GQDPYHHPQAHGLA-FSVRANVPP-37-NTTLTVKRGK--53-HPSPLSK--28-C					
Family 2	Eco	N--15-GINPGLSSAG-TGFP-FAHPANRFW-29-KLVDRPTVQA--62-NPSGLSR--22-C					(MUG/TDG)
Family 3	Gme	N--55-GMNPWPWMAQTGVP-FGEVAVVTE-56-NYCPLFLTA--64-HPSPASP--21-C					(SMUG1)
Family 5	Tth	N--56-GLAPGAHGSNRTGRP-FTGDASGAF-30-AAVRCAPPKN--69-HVSRQNT--23-C					(UDGb)
Family 6	Mba	N--19-GSLPGDVSIR-KHQY-YGHPGNDFW-31-DVFKAGKREG--52-SSSGANR--16-C					(HDG)

Fig. 2. Sequence alignment of DR0022-like enzymes with other UDG families. The sequence alignment was performed by CLUSTAL OMEGA (<http://www.clustal.org>) and adjusted manually. The GenBank accession number is provided after the species name. The sites in DR0022 that are subject to site-directed mutagenesis analysis are marked by an arrow. 0022-like: Dra (DR0022) *Deinococcus radiodurans* R1, [AAF09614](#); Cph, *Chlorobium phaeobacteroides* DSM 266, [ABL66232.1](#); Ses, *Saccharothrix espanaensis*, [WP_051075528.1](#); Mra, *Methylobacterium radiotolerans*, [WP_076728117.1](#); dap, *Deinococcus apachensis*, [WP_019587937.1](#); Psy, *pseudomonas syringae*, [WP_010924790.1](#). Family 4 (UDGa): Tth, *Thermus thermophilus* HB27, [YP_004341.1](#); Pae, *Pyrobaculum aerophilum* str. IM2, [NP_558739.1](#); Tma, *Thermotoga maritima* MSB8, [NP_228321.1](#); Dra (DR1751), *D. radiodurans* R1, [NP_295474](#); Nmu, *Nitrosospora multiformis*, [WP_074775154.1](#). Family 1 (UNG): Eco, *E. coli*, [NP_289138](#); Dra (DR0689), *D. radiodurans* R1, [NP_294412](#); Mtu, *Mycobacterium tuberculosis*, [CKL94140.1](#); Xla, *Xenopus laevis*, [OCU01661.1](#); Hsa, *Homo sapiens*, [NP_003353](#); HSV1, herpes simplex virus 1, 1UDI. Family 2 (MUG/TDG): Eco, *E. coli*, [POA9H1](#). Family 3 (SMUG1): Gme, *Geobacter metallireducens* GS-15, [YP_383069](#). Family 5 (UDGb): Tth, *T. thermophilus* HB8, [YP_144415.1](#). Family 6 (HDG): Mba, *Methanosarcina barkeri* str. Fusaro, [YP_304295.1](#).

Table 2. DNA glycosylase activity of DR0022 on U- and hmU-containing DNA. Experiments were performed as described as [Materials and methods](#). Data are the average of three independent experiments.

	A U	T U	G U	C U	U	A hmU	T hmU	G hmU	C hmU	hmU
Wild-type	0	12	34	18	0	0	12	23	15	0
L43G	0	0	0	0	0	0	0	0	0	0
E44A	0	0	0	0	0	0	0	0	0	0
A50E	0	28	46	26	0	0	0	0	0	0
N90A	0	0	0	0	0	0	0	0	0	0
H159A	0	0	0	0	0	0	0	0	0	0
L43G-A50E	0	6	16	11	0	0	0	0	0	0

**Fig. 3.** Rif resistance mutation rates from the wild-type and *dr0022* deficient *Deinococcus radiodurans* strains. *D. radiodurans* (WT) and *D. radiodurans* (*dr0022*⁻) cells (1×10^9) were plated on TGY plates containing $50 \mu\text{g}\cdot\text{mL}^{-1}$ rifampicin. Results are from three independent experiments. * $P < 0.05$ by *t*-test. Error bars indicate the SD, $n = 3$.

started to diverge at higher γ -irradiation dosages, with the survival rate of the *dr0022* deficient strain dropping to around 10%, whereas the wild-type strain reduced to around 20% at levels of 5 and 7 kGy, indicating that *dr0022* plays a role in radiation resistance (Fig. 4D).

Discussion

Dra, as an extremely radiation resistant and stress resistant microorganism, has an expanded DNA repair repertoire to process a wide variety of DNA lesions [34]. Previous studies have identified several uracil-DNA glycosylases and confirmed enzymatic activities in several of them [29]. The present study identifies DR0022 as an acid uracil- and hydroxymethyluracil-DNA glycosylase. To our knowledge, this is the first report of such a UDG enzyme in UDG superfamily.

To understand the structural differences between DR0022 and other UDG enzymes, we took advantage

Table 3. Distribution of mutations leading to Rif^r in *dr0022* deficient *Deinococcus radiodurans* strain.

Mutation site in <i>rpoB</i> (bp)	Amino acid change	Base pair change	Number
1274	D425V	AT-TA	39
1319	S440F	GC-AT	9
1304	H435P	AT-CG	2

of a state-of-the-art protein structure prediction method, ALPHAFOLD [35]. ALPHAFOLD is developed based on deep learning algorithm that demonstrates highly accurate domain structure and amino acid side chain predictions. To validate the accuracy of ALPHAFOLD, we predicted the structures of the Tth family 4 UDGA and the Tth family 5 UDGB using ALPHAFOLD at the Colab online server [36]. Superimposition demonstrated that the predicted structure was highly similar to the corresponding solved crystal structures, with a rmsd of 0.289 \AA [Protein Data Bank (PDB): 1UI0] and 0.375 \AA (PDB: 2DP6), respectively (Fig. 5). We then used ALPHAFOLD to predict the three-dimensional structure of DR0022 protein. The predicted DR0022 protein structure showed a typical structural fold, as seen in other UDG protein structures solved by X-ray crystallography with a predicted local distance difference test score above 90 (Fig. 6A), an indication of high confidence in the structural prediction. We then used SWISS-MODEL (<https://swissmodel.expasy.org>) to build a modeled structure based on known crystal structures. Analysis of sequence identity in SWISS-MODEL indicated that DR0022 had a relatively higher sequence identity with the Tth family 5 UDGB protein (PDB: 2DP6, 25% identity) and the Tth family 4 UDGA (PDB: 1UI0, 20% identity). We therefore built the DR0022 structure using Tth family 5 UDGB protein (PDB: 2DP6) and Tth family 4 UDGA protein (PDB: 1UI0) (Fig. 6B,C). The predicted DR0022 structure had a very similar structure to the modeled structures using 2DP6 and 1UI0 structures as the

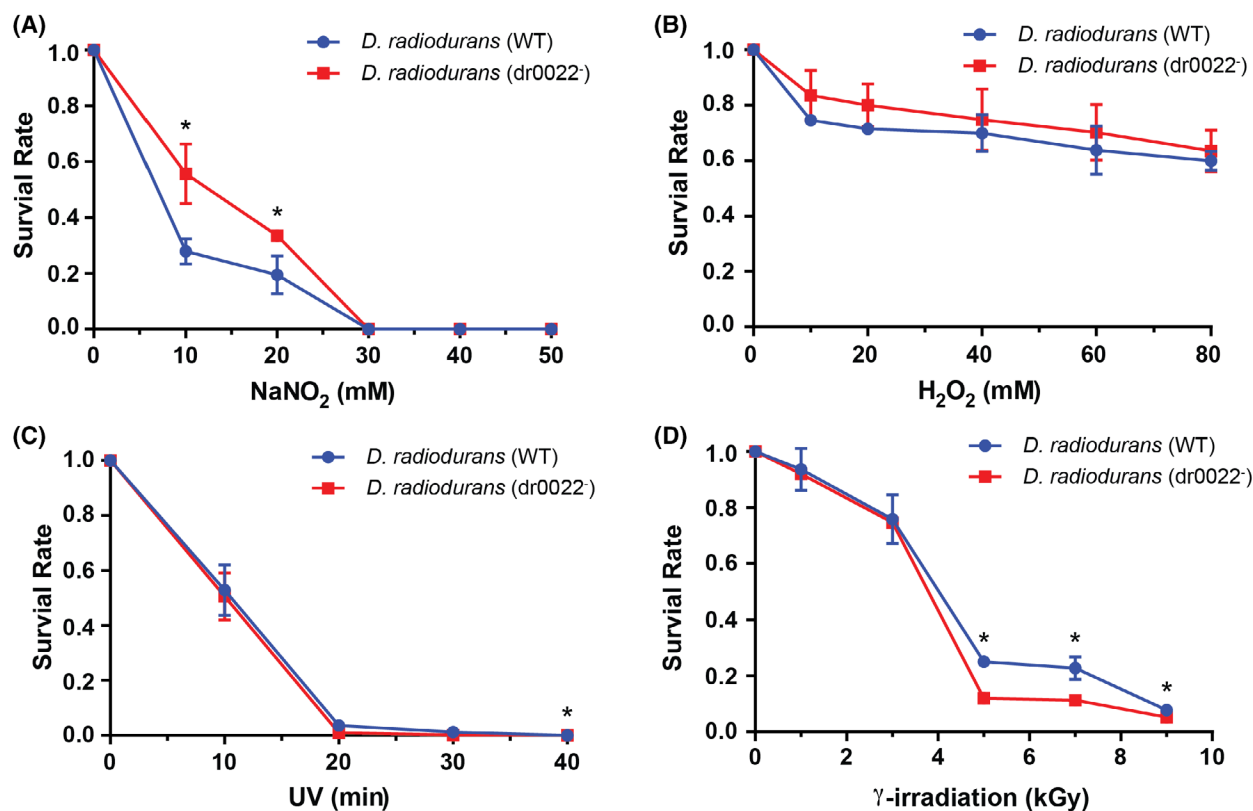


Fig. 4. Survival rates of the *dr0022* deficient strain under various stress. * $P < 0.05$ by *t*-test. (A) Survival rates of wild-type and *dr0022* deficient *Deinococcus radiodurans* strains under nitrosative stress. (B) Survival rates of wild-type and *dr0022* deficient *D. radiodurans* strains under H₂O₂ treatment. (C) Survival rates of wild-type and *dr0022* deficient *D. radiodurans* strains under UV radiation. (D) Survival rates of wild-type and *dr0022* deficient *D. radiodurans* strains under γ -irradiation. Error bars indicate the SD, $n = 3$.

template, with rmsd values of 1.517 and 1.115 Å, respectively. As a computational method to predict three-dimensional structures, ALPHAFOLD-based *de novo* structural prediction has been accessed and evaluated independently as powerful and accurate [37]. As a test case here, the predicted DR0022 structure appears to be highly consistent with the modeled structures.

Despite the similarity in motifs 2 and 3, the most obvious difference is that the first residue in motif 1 of DR0022 is a leucine rather than a glycine residue, which is universally seen in other families of UDG superfamily (Fig. 7). Interestingly, the substitution of Leu with Gly (L43G) rendered DR0022 inactive under both acidic and alkaline pH conditions, suggesting that the enzyme has adapted to the hydrophobic Leu residue in this position. The predicted and modeled DR0022 structures suggest that L43 may stack on the uracil ring to enhance or facilitate other interactions with uracil in the base binding pocket (Fig. 7). Interestingly, the acid UDG activity can be partially rescued by the L43G-A50E double mutant (Table 2),

suggesting the correlated nature of certain amino acid changes during evolution.

The base recognition pocket and the catalytic center bear similarities to other UDG enzymes (Fig. 8). In the predicted and modeled DR0022 structures, the backbone of E44 of motif 1 interacts with O2 of uracil as seen in family 4 UDGa [10]. H159 of motif 2 interacts with O2 of uracil to stabilize the departing uracil as seen in Tth UDGa [10]. N90 of motif 3 may engage in extensive interactions with the N3 and O4 moieties of uracil. These important roles in substrate recognition and catalysis are demonstrated by the lack of glycosylase activities in amino acid substitutions in these positions (Tables 1 and 2). These results indicate that DR0022 uses the same general catalytic mechanism as seen in family 4 UDGa to remove uracil or hmU in DNA [10]. An interesting observation is that the substitution of alanine with glutamate at A50 position (A50E) actually increased the acid UDG activity (Table 2), and kinetics analysis demonstrated that k_2/K_m value was almost 5-fold higher than the

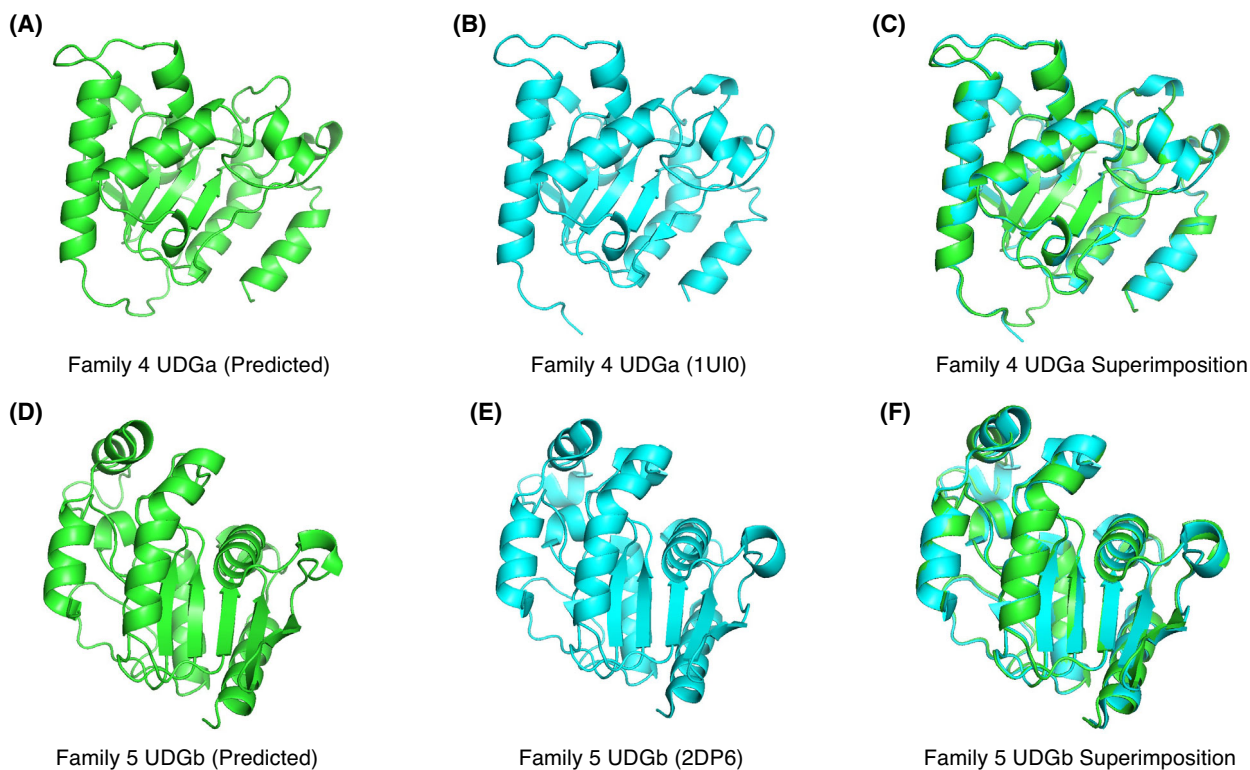


Fig. 5. Comparison of predicted Tth family 4 UDGa and family 5 UDGb structures with corresponding crystal structures. (A) ALPHAFOLD2 predicted structure of *Thermus thermophilus* family 4 UDGa protein. (B) Crystal structure of *T. thermophilus* family 4 UDGa (PDB: 1UI0). (C) Superimposition of predicted family 4 UDGa structure with corresponding crystal structure. (D) ALPHAFOLD2 predicted structure of *T. thermophilus* family 5 UDGb protein. (E) Crystal structure of *T. thermophilus* family 5 UDGb (PDB: 2DP6). (F) Superimposition of predicted family 5 UDGBs structure with corresponding crystal structure.

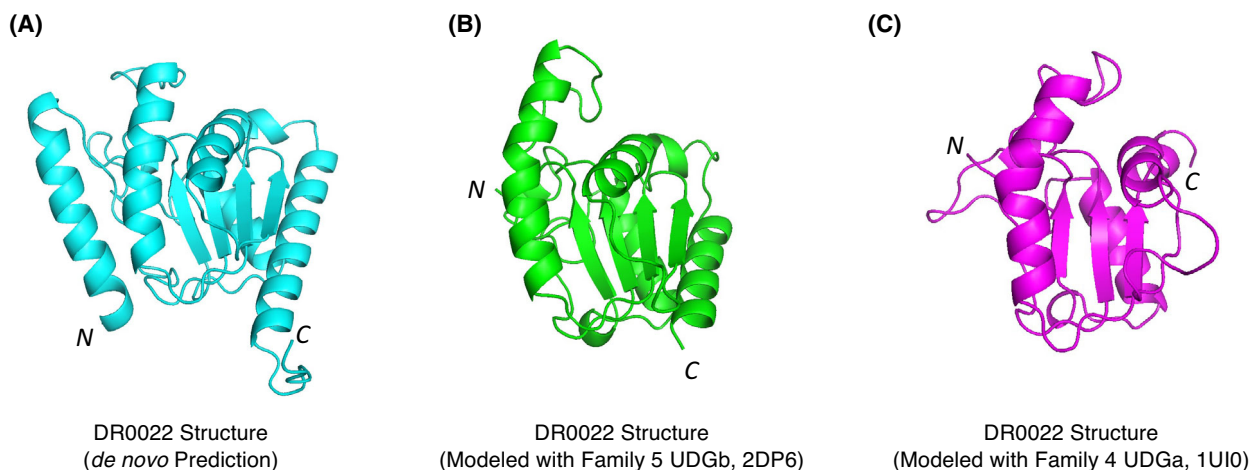


Fig. 6. Modeled and predicted DR0022 protein structures. (A) Structure of DR0022 protein constructed by *de novo* prediction using ALPHAFOLD2. (B) Structure of DR0022 protein modeled after *Thermus thermophilus* family 5 UDGb crystal structure (PDB: 2DP6) using SWISS-MODEL. (C) Structure of DR0022 protein modeled after *T. thermophilus* family 4 UDGa crystal structure (PDB: 1UI0) using SWISS-MODEL.

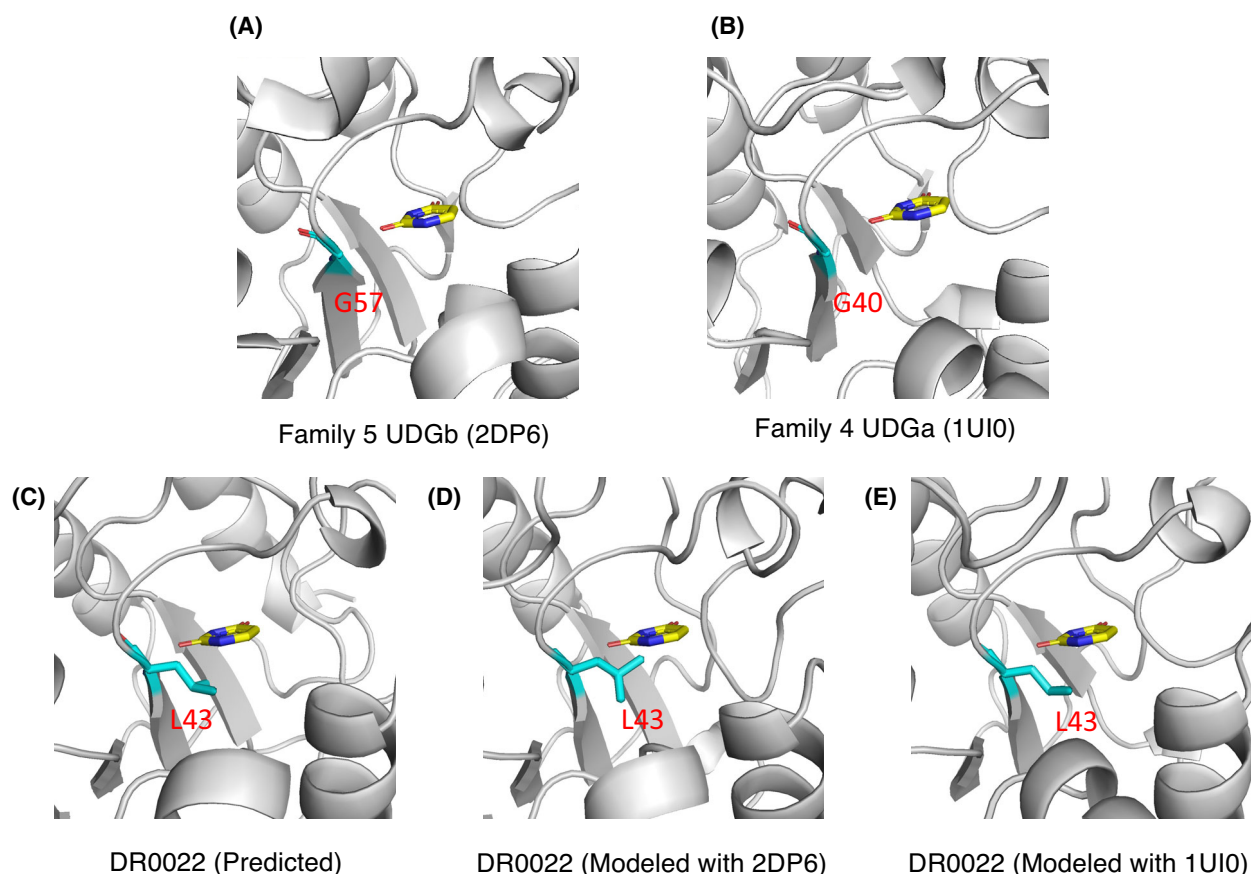


Fig. 7. Close-up view of L43 in predicted and modeled DR0022 structures. Uracil is shown as licorice. (A) Crystal structure of *Thermus thermophilus* family 5 UDGb (PDB: 2DP6). G57 is shown as licorice. (B) Crystal structure of *T. thermophilus* family 4 UDGa (PDB: 1UI0). G40 is shown as licorice. (C) Predicted structure of DR0022 protein using ALPHAFOLD2. L43 is shown as licorice. (D) Modeled structure of DR0022 protein with 2DP6 as the template using SWISS-MODEL. L43 is shown as licorice. (E) Modeled structure of DR0022 protein with 1UI0 as the template using SWISS-MODEL. L43 is shown as licorice.

wild-type enzyme (Table 1). Previously, we compared the structural difference between the Tth UDGa and the *E. coli* family 1 UNG and found that E47 in Tth UDGa can block the entry of thymine into the recognition pocket [10]. The predicted and modeled DR0022 structures indicate that A50E can play the same role in DR0022, which explains the elimination of hmU activity in this mutant. As shown in Fig. 9, the Tth family 5 UDGb has a glycine (G64) and the Tth family 4 UDGb has a glutamate (E47) in the corresponding positions. In comparison with the wild-type structures, the A50E mutant adopts a conformation that is similar to family 4 UDGa. Accordingly, it is likely that A50E may now favorably interact with the uracil base, thus enhancing the UDG activity.

The absolute requirement for an acidic environment allowing DR0022 to act as a UDG is unprecedented. We considered two scenarios to explain this phenomenon. The first scenario relates to the substrate.

Uracil is a lactam with two intracyclic amide bonds, which are subject to tautomerization. The lactam form of uracil predominates under normal neutral pH [38]. However, the lactim form of uracil increases under acidic pH conditions. In the Tth UDGa structure, N80, as the first residue in motif 3 forms a bidentate hydrogen bond with N3 and O4 of uracil in the lactam form [10,39]. It is interesting to note that N90 of motif 3 in DR0022 has different conformations. In the predicted structure of DR0022 protein, N90 adopts a conformation that facilitates the formation of a bidentate hydrogen bond with N3 and O4 of uracil in the lactim form (Fig. 8C,F). Our previous study already underscores the important role of K68N of *E. coli* MUG in substrate recognition and catalysis [5]. Interestingly, the sequencing analysis of Rif^R in *dr0022* deficient Dra strain revealed that the A/T to T/A transversion was the predominant amino acid substitution (Table 3). This led us to speculate that DR0022 may recognize a

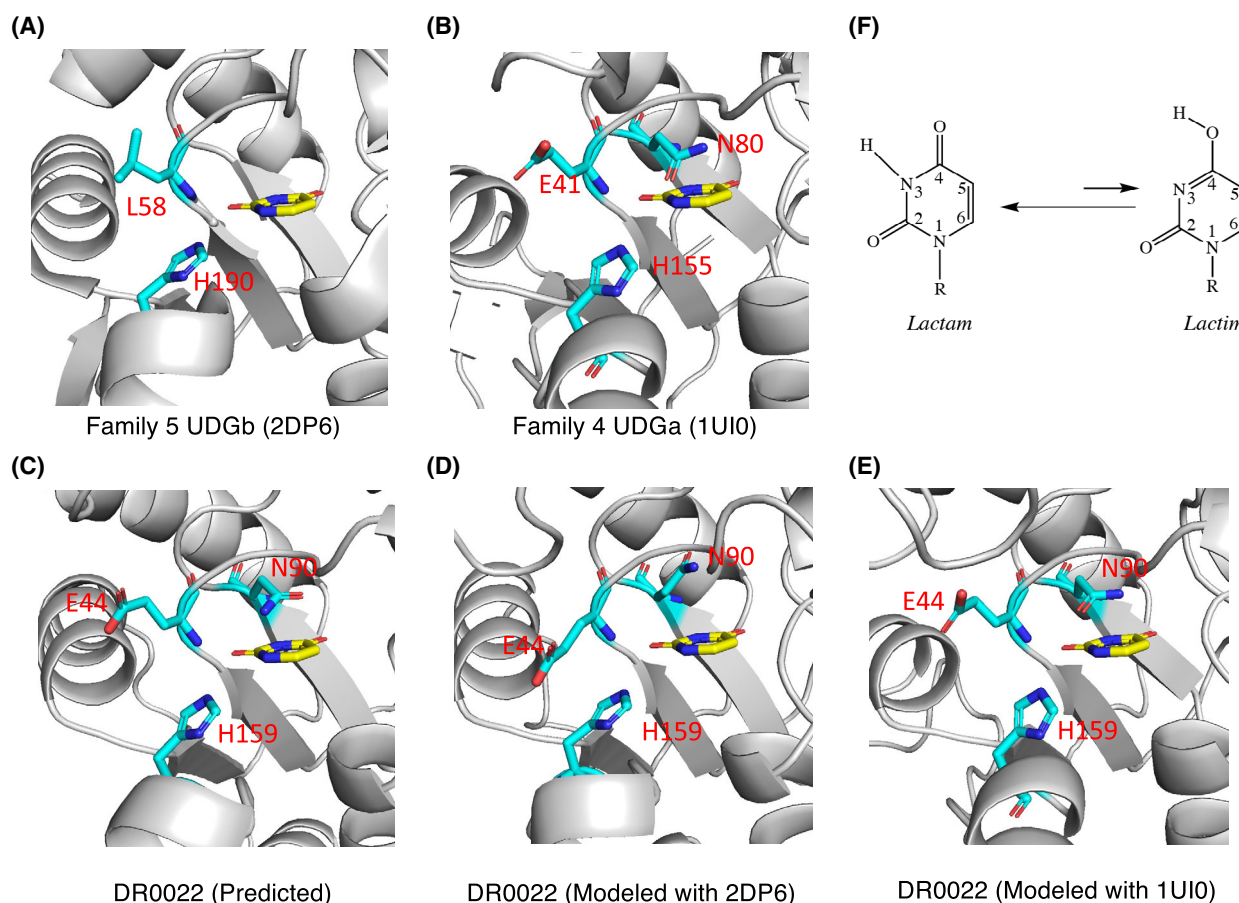


Fig. 8. Base recognition pocket and catalytic center in DR0022 according to predicted and modeled structures. Uracil is shown as licorice. (A) Crystal structure of *Thermus thermophilus* family 5 UDGb (PDB: 2DP6). L58 and H190 are shown as licorice. (B) Crystal structure of *T. thermophilus* family 4 UDGa (PDB: 1UI0). E41, N80, and H155 are shown as licorice. (C) Predicted structure of DR0022 protein using ALPHAFOLD2. E44, N90 and H159 are shown as licorice. (D) Modeled structure of DR0022 protein with 2DP6 as the template using SWISS-MODEL. E44, N90, and H159 are shown as licorice. (E) Modeled structure of DR0022 protein with 1UI0 as the template using SWISS-MODEL. E44, N90, and H159 are shown as licorice. (F) Lactam-lactim shift of uracil base.

thymine derivative (T^{*}) that bears some structural similarity to the uracil in the lactim form (Fig. 8F). The thymine derivative may mimic an adenine base to pair with T during DNA replication. As a result, a A/T base pair is converted to a T/A base pair. In the second scenario, we assume the uracil still maintains its lactam form. However, the interaction with the uracil is not possible until one of the key catalytic residues in the active site is protonated at acidic pH conditions.

In addition to the extreme resistance to radiation, Dra is reported to be highly resistant to DNA damaging agents [40]. Most notably, the previous study found that Dra was 62-fold more resistant to nitrous acid treatment, 55-fold more resistant to radiation treatment, and 33-fold more resistant to UV-C treatment than *E. coli* [40]. We determined the survival rates of *dr0022* deficient Dra strain aiming to

understand the contribution of *dr0022* to resistance to various stresses (Fig. 4). Among the four stress conditions tested, DR0022 appears to contribute to acidified nitrous acid treatment and γ -irradiation (Fig. 4). Acidified nitrous acid can generate nitrating agent, which can deaminate nucleobases. As a uracil-DNA glycosylase, DR0022 would be expected to remove the cytosine deamination product, uracil, during the repair process. Continuous removal of uracil will generate large number of AP sites, which may cause cytotoxicity. As such, the *dr0022*⁻ strain appeared more resistant to acidified nitrous acid treatment than the wild-type strain.

A previous study found that the expression of *dr0022* was increased by over 3-fold when Dra was treated with a high dosage of γ -irradiation [41]. In the survival assay, we found a statistically significant

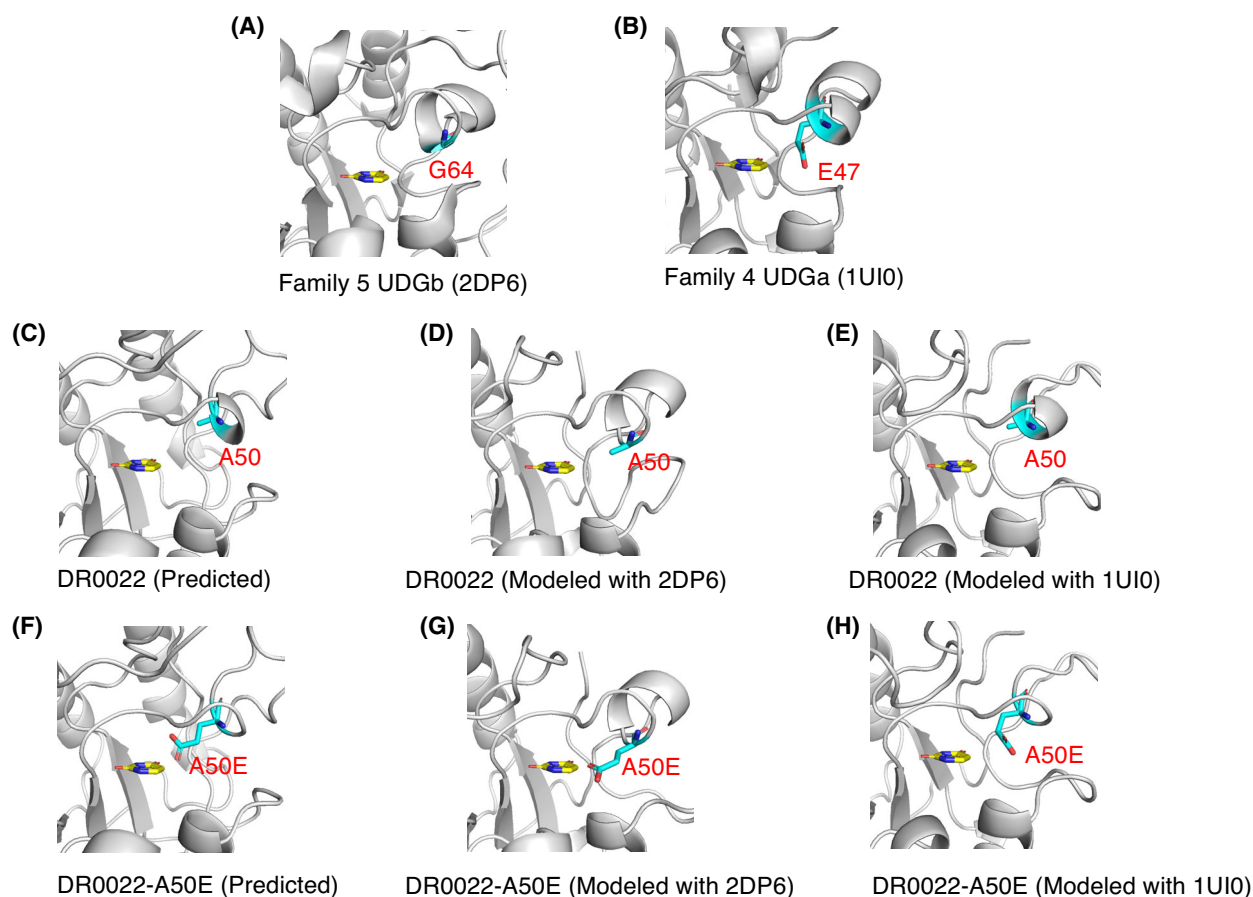


Fig. 9. Close-up view of A50 and A50E in predicted and modeled DR0022 structures. Uracil is shown as licorice. (A) Crystal structure of *Thermus thermophilus* family 5 UDGb (PDB: 2DP6). G64 is shown as licorice. (B) Crystal structure of *T. thermophilus* family 4 UDGa (PDB: 1UI0). E47 is shown as licorice. (C) Predicted structure of DR0022 protein using ALPHAFOLD2. A50 is shown as licorice. (D) Modeled structure of DR0022 protein with 2DP6 as the template using SWISS-MODEL. A50 is shown as licorice. (E) Modeled structure of DR0022 protein with 1UI0 as the template using SWISS-MODEL. A50 is shown as licorice. (F) Predicted structure of DR0022(A50E) protein using ALPHAFOLD2. A50E is shown as licorice. (G) Modeled structure of DR0022(A50E) protein with 2DP6 as the template using SWISS-MODEL. A50E is shown as licorice. (H) Modeled structure of DR0022(A50E) protein with 1UI0 as the template using SWISS-MODEL. A50E is shown as licorice.

difference at a high dose of γ -irradiation between the wild-type strain and the *dr0022* deficient strain (Fig. 4 D). These results suggest that DR0022 can play a role in repairing DNA damage after radiation. In conjunction with the analysis above, we speculate that DR0022 may recognize a base lesion generated after a high dosage of irradiation that bears some structural similarity to uracil in its lactim form. By removing this cytotoxic lesion, DR0022 enhances the survival of *Dra* after γ -irradiation.

In summary, the present study reports the finding of a new acid uracil-DNA glycosylase in *Dra*. The biochemical and mutational analyses define the base lesion pocket and suggest a catalytic mechanism. The *in vitro* and *in vivo* investigation on mutation prevention raises the prospect that the acidic environment

required for the glycosylase activity is related to its role in removing a base lesion that could generate a transversion mutation. The involvement of this glycosylase in radiation resistance is intriguing. More studies are needed to further decipher the fascinating evolution of the UDG superfamily to fulfill a variety of biological functions.

Materials and methods

Reagents, media, and strains

All routine chemical reagents were purchased from Sigma Chemicals (St Louis, MO, USA), Fisher Scientific (Suwanee, GA, USA), or VWR (Suwanee, GA, USA). Restriction enzymes, Phusion DNA polymerase, and T4 DNA

ligase were purchased from New England Biolabs (Beverly, MA, USA). BSA and dNTPs were purchased from Promega (Madison, WI, USA). The gel DNA recovery kit was purchased from Zymo Research (Irvine, CA, USA). Oligodeoxyribonucleotides were ordered from Integrated DNA Technologies Inc. (Coralville, IA, USA) and Eurofins Genomics (Huntsville, AL, USA). The LB (Miller) medium was prepared according to standard recipes. Hi-Di™ Formamide and GeneScan™ 500 LIZ™ dye Size Standard for ABI3130xl were purchased from Applied Biosystems (Waltham, MA, USA). The sonication buffer consisted of 20 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0), 50 mM NaCl, 2.5 mM dithiothreitol and 0.15 mM phenylmethanesulfonyl fluoride. The TE buffer consisted of 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA.

Cloning, site-directed mutagenesis, expression, and purification of DR0022

The *dr0022* gene from Dra R1 (GenBank accession number: [AE_000513.1](#)) was amplified by PCR using the forward primer DR0022 F (5'-GCTCTAGACCATATGTTGCCGACCACGTC^{CG}-3'; the *NdeI* site is underlined) and the reverse primer DR0022 R (5'-CCGCTCGAGGGAAGACCAACTGGG^{CCG}-3'; the *XhoI* site is underlined). The PCR reaction mixture (50 µL) consisted of 20 ng of Dra R1 genomic DNA, 500 nM forward and reverse primers, 1 × Phusion DNA polymerase buffer, 200 µM each dNTP and 0.2 unit of Phusion DNA polymerase (New England Biolabs). The PCR procedure included a pre-denaturation step at 98 °C for 30 s; 30 cycles of three-step amplification with each cycle consisting of denaturation at 98 °C for 15 s, annealing at 65 °C for 20 s, and extension at 72 °C for 30 s; and a final extension step at 72 °C for 3 min. The PCR product was purified by the gel DNA recovery kit (Zymo Research). The purified PCR product and plasmid pET21a were digested by *NdeI* and *XhoI*, purified by the gel DNA recovery kit, and ligated in accordance with the manufacturer's instructions. The ligation mixture was transformed into *E. coli* strain HB101 competent cells by electroporation. The sequence of the *dr0022* gene in the resulting plasmid (pET21a-DR0022) was confirmed by DNA sequencing.

The resulting plasmid with wild-type *dr0022* was used as the template plasmid for all other DR0022 mutants. Amplification of the mutant DNA and DpnI mediated site-directed mutagenesis procedures were modified using primers carrying the desired mutations, as described previously [42]. Briefly, PCR mixtures (25 µL) contained 10 ng of pET-21a (+)-DR0022 as a template, 65 nM each primer pair, 200 µM each dNTP, 1 × 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 a 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 DpnI 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. The pET21a-DR0022 wild-type and mutants were transformed into *E. coli* strain BH214 (*ung*⁻, *mug*⁻) by the standard protocol to express the C-terminal His-6-tagged DR0022 protein. Induction, sonication, and purification were carried out as described previously [7]. Briefly, bacterial cells containing pET21a-DR0022 wild-type or mutant plasmid from a 500-mL culture grown to late exponential phase were harvested by centrifugation at 3993 g for 10 min. The cell pellet was suspended in 7 mL of lysis buffer [20 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0), 50 mM NaCl, 0.1 mM dithiothreitol, and 0.15 mM phenylmethanesulfonyl fluoride] and followed by sonication at output 5 for 3 × 1 min with a 5-min rest on ice between intervals using a Sonifier Cell Disruptor 350 (Branson, Brookfield, CT, USA). The yields from protein expression ranged around from 2–5 mg·500 mL⁻¹. The lysate was clarified by centrifugation at 21 000 g for 20 min and filtered through a 25-mm GD/X syringe filter (Whatman, Little Chalfont, UK). The supernatant was transferred into a fresh tube and loaded onto a 1-mL HiTrap chelating column (GE Healthcare, Chicago, IL, USA), followed by a 1-mL HiTrap SP column (GE Healthcare). The SDS/PAGE analysis of the purified proteins is shown in Fig. 1C. The DR0022 protein was stored in aliquots at -80 °C. Prior to use, the protein was diluted in an equal volume of 2 × storage buffer (20 mM Tris-HCl (pH 8.0), 2 mM dithiothreitol, 2 mM EDTA, 400 µg·mL⁻¹ BSA, and 100% glycerol).

DNA glycosylase activity assay

The DNA glycosylase cleavage assays for DR0022 were performed at optimized temperature 37 °C for 60 min in a 10-µL reaction mixture containing 10 nM oligonucleotide substrate [43], an indicated amount of glycosylase, 20 mM citric acid and phosphate buffer with the indicated pH (4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0) [44], 1 mM dithiothreitol, and 1 mM EDTA. The resulting abasic sites were cleaved by incubation at 95 °C for 5 min after adding 1 µL of 1 N NaOH. The reaction mixtures (2 µL) were mixed with 7.8 µL of Hi-Di formamide and 0.2 µL of GeneScan 500 LI Size Standard (Life Technologies, Carlsbad, CA, USA) and analyzed using a Applied Biosystems 3130xl sequencer with a fragment analysis module. Cleavage products and remaining substrates were quantified using GENEMAPPER (Applied Biosystems).

Enzyme kinetics analysis

The uracil-DNA glycosylase assay was performed under optimized reaction conditions at 37 °C in a 10-µL reaction

mixture containing 10 nM oligonucleotide substrate, an indicated excess amount of DR0022 enzyme, 20 mM citric acid and phosphate buffer, pH 5.5, 1 mM dithiothreitol, and 1 mM EDTA. The enzyme concentration ranged from 50 to 4000 nM. Samples were collected at time points from 10, 20, 30, 40, 50, and 60 min. The apparent rate constants for each concentration were determined by curve fitting using the integrated first-order rate:

$$P = P_{\max}(1 - e^{-k_{\text{obs}}t}) \quad (1)$$

where P is the product yield, P_{\max} is the maximal yield, t is time, and k is the apparent rate constant.

The kinetic parameters k_2 and K_m were obtained from plots of k_{obs} against the total enzyme concentration ($[E_0]$) using a standard hyperbolic kinetic expression [10,31,32]:

$$k_{\text{obs}} = \frac{k_2 [E_0]}{K_m + [E_0]} \quad (2)$$

Construction of the *dr0022* deficient strain

The *dr0022* deletion mutant was constructed by double crossover recombination of a kanamycin resistance cassette into the genome, with some modifications of the method as described previously [33]. Briefly, all primers were designed using the genome sequence information of Dra R1 available on NCBI (<https://www.ncbi.nlm.nih.gov>). The upstream fragment was amplified by PCR, with the primer set DR0022 up *Xba*I F (5'-GCTCTAGACCCAGAATGCCGAGGGTGGC-3'; the *Xba*I site is underlined) and DR0022 up *Bam*HI R (5'-CGCGGATCCAGTGAGCTAGGAGCGGCGCT-3'; the *Bam*HI site is underlined). Similarly, the downstream fragment was PCR-amplified with the primer set DR0022 down *Hind*III F (5'-CCCAAGCTTCGTTTACTAGG GCATTTGAC-3'; the *Hind*III site is underlined) and DR0022 down *Xho*I R (5'-CCGCTCGAGGGCGATACCCTGCCCGCAT-3'; the *Xho*I site is underlined). The upstream and downstream fragments were digested with *Bam*HI and *Hind*III, respectively, and ligated to the *Bam*HI-*Hind*III predigested fragment of the kanamycin resistance cassette obtained from a shuttle plasmid pRADK bearing the groEL promoter and the kanamycin resistance gene [33]. The ligation product was used as a template for PCR amplification of the full-length PCR product, with the primer set DR0022 up *Xba*I F and DR0022 down *Xho*I R. The resulting PCR product was digested by *Xba*I and *Xho*I and ligated into the pET21a vector that had been digested with the same enzymes to yield a pET21a-DR0022 deletion, which was then transformed into Dra R1. Mutant strains were selected on TGY agar plates supplemented with kanamycin (25 $\mu\text{g}\cdot\text{mL}^{-1}$). The *dr0022* gene mutant was designated as Dra (*dr0022*⁻).

Spontaneous mutation rate

The spontaneous mutation rates of Dra deficient mutants were tested using the rpoB/Rif^r system [45]. Briefly, the freshly growing Dra cells were diluted and tittered by plating on TGY plates. Spontaneous Rif^r mutants were obtained by inoculating a certain number (2×10^9) of Dra cells on TGY plates containing 50 $\mu\text{g}\cdot\text{mL}^{-1}$ rifampicin (Sigma Chemicals). The Rif^r mutation rate (f) was determined by the number of clones with rifampicin-resistance divided by the total number of Dra cells plated on TGY plates supplemented with 50 $\mu\text{g}\cdot\text{mL}^{-1}$ rifampicin.

Sequencing the *rpoB* gene for mutations

Using genomic DNA as a template, the *rpoB* gene was amplified with the following two primer pairs for direct sequencing. Primer pair 1: 5'-AAACTGTGCCGATGGTGGAC-3' (5' position 1058) and 5'-TAGCTCACGCGGCCATTAC-3' (5' position 1945). Primer pair 2: 5'-TCTTTCCCATCGACGAGTCC-3' (5' position 173) and 5'-CACGATGGGGCGGTTGTT-3' (5' position 1224). The PCR reaction included 1 \times Phusion HF buffer (Thermo Scientific, Waltham, MA, USA), 50 pmol each PCR primer, 40 nmol dNTP, 3% dimethylsulfoxide, 0.5 units of Phusion DNA polymerase (Thermo Scientific), 10 ng of genomic DNA, and double-distilled H₂O. The DNA was denatured at 95° for 4 min, amplified for 30 cycles of 95° for 30 s, 57° for 30 s, and 72° for 1 min and extended for 7 min at 72°. PCR products were purified with the Gel DNA recovery kit (Zymo Research) and sequenced with one of two primers, 5'-CATGCTGCTCGGCAACCC-3' (5' position 1221) or 5'-TGATTCACAAAGACACTGGCGT-3' (5' position 323), respectively.

Survival rates of Dra (*dr0022*⁻) under different stresses

To induce the nitrosative stress, Dra (*wt*) and Dra (*dr0022*⁻) cells were treated at the indicated concentration of NaNO₂ (pH 5.0) for 20 min. Next, cells were washed twice with TGY medium and plated on the TGY plates. To induce H₂O₂ stress, Dra (*wt*) and Dra (*dr0022*⁻) cells were treated with the indicated concentration of H₂O₂ for 20 min. Then, the cells were washed twice with TGY medium and cells were plated on the TGY plates. To induce UV stress, Dra (*wt*) and Dra (*dr0022*⁻) cells were treated under UV-C light for the indicated time. Then, the cells were washed twice with TGY medium and plated on the TGY plates. The irradiation with γ -rays assay was performed as described previously [46]. Briefly, cell suspensions of Dra (wild-type) and Dra (*dr0022*⁻) strains were irradiated at room temperature for 24 h with ¹³⁷Cs γ -rays at various distances from the source, which correspond to various doses (0, 1, 2, 3, 4, 5, 6, 7, and 8 kGy). After

irradiation, cells were plated on TGY plates. All of the plates from the four assays were incubated at 30 °C for 3 days prior to colony enumeration. Survival rates are calculated as a percentage of the number of colonies obtained with untreated cells. All results are the average of three independent experiments.

De novo prediction and modeling of DR0022 protein structures

The structures of the wild-type and mutant DR0022 proteins were predicted by ALPHAFOLD2 at the Colab server without templates or homologous structures [36]. The predicted local distance difference test scores calculated by Colab-ALPHAFOLD were used to evaluate the confidence in structural prediction. We also modeled the DR0022 wild-type and mutant protein structures using two other known uracil-DNA glycosylase structures as a template (PDB: 2DP6 and 1UI0) in SWISS-MODEL [47]. The predicted and modeled structures were analyzed with PYMOL (The PyMOL Molecular Graphics System, version 2.0; Schrödinger, LLC, New York, NY, USA).

Acknowledgements

This project was supported in part by the National Institutes of Health (GM090141). We thank the Dr John Battista lab at Louisiana State University for providing *Deinococcus radiodurans* R1 strain and genomic DNA. We express our gratitude to Dr Liangyan Wang and other members of the Dr Yuejin Hua lab at Zhejiang University for providing *D. radiodurans* related materials and for help with performing the γ -irradiation experiments. We also thank members of the Cao lab for assistance and discussions, as well as Jenna Perry for editorial assistance.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

WC conceived and designed the project. JL designed and performed experiments. JL and WC analyzed data. YY participated in mutation rate experiments. CC assisted with structure prediction and modeling. WC and JL wrote the manuscript.

Data availability statement

All data generated or analyzed during this study are included in the published article.

References

- 1 Lindahl T. Instability and decay of the primary structure of DNA. *Nature*. 1993;**362**:709–15.
- 2 Burney S, Caulfield JL, Niles JC, Wishnok JS, Tannenbaum SR. The chemistry of DNA damage from nitric oxide and peroxyxynitrite. *Mutat Res*. 1999;**424**:37–49.
- 3 Wink DA, Mitchell JB. Chemical biology of nitric oxide: insights into regulatory, cytotoxic, and cytoprotective mechanisms of nitric oxide. *Free Radic Biol Med*. 1998;**25**:434–56.
- 4 Parikh SS, Putnam CD, Tainer JA. Lessons learned from structural results on uracil-DNA glycosylase. *Mutat Res*. 2000;**460**:183–99.
- 5 Lee DH, Liu Y, Lee HW, Xia B, Brice AR, Park SH, et al. A structural determinant in the uracil DNA glycosylase superfamily for the removal of uracil from adenine/uracil base pairs. *Nucleic Acids Res*. 2015;**43**:1081–9.
- 6 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. *J Biol Chem*. 2010;**285**:41483–90.
- 7 Lee HW, Dominy BN, Cao W. New family of deamination repair enzymes in uracil-DNA glycosylase superfamily. *J Biol Chem*. 2011;**286**:31282–7.
- 8 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. *J Mol Biol*. 2009;**385**:761–78.
- 9 Pearl LH. Structure and function in the uracil-DNA glycosylase superfamily. *Mutat Res*. 2000;**460**:165–81.
- 10 Xia B, Liu Y, Guevara J, Li J, Jilich C, Yang Y, et al. Correlated mutation in the evolution of catalysis in uracil DNA glycosylase superfamily. *Sci Rep*. 2017;**7**:45978.
- 11 Xia B, Liu Y, Li W, Brice AR, Dominy BN, Cao W. Specificity and catalytic mechanism in family 5 uracil DNA glycosylase. *J Biol Chem*. 2014;**289**:18413–26.
- 12 Hardeland U, Bentele M, Jiricny J, Schar P. The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs. *Nucleic Acids Res*. 2003;**31**:2261–71.
- 13 Hardeland U, Bentele M, Lettieri T, Steinacher R, Jiricny J, Schar P. Thymine DNA glycosylase. *Prog Nucleic Acid Res Mol Biol*. 2001;**68**:235–53.
- 14 Hashimoto H, Zhang X, Cheng X. Activity and crystal structure of human thymine DNA glycosylase mutant N140A with 5-carboxylcytosine DNA at low pH. *DNA Repair*. 2013;**12**:535–40.
- 15 He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its

- excision by TDG in mammalian DNA. *Science*. 2011;**333**:1303–7.
- 16 Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem*. 2011;**286**:35334–8.
- 17 Dong L, Mi R, Glass RA, Barry JN, Cao W. Repair of deaminated base damage by *Schizosaccharomyces pombe* thymine DNA glycosylase. *DNA Repair*. 2008;**7**:1962–72.
- 18 Lutsenko E, Bhagwat AS. The role of the *Escherichia coli* mug protein in the removal of uracil and 3,N(4)-ethenocytosine from DNA. *J Biol Chem*. 1999;**274**:31034–8.
- 19 Saparbaev M, Langouet S, Privezentzev CV, Guengerich FP, Cai H, Elder RH, et al. 1,N(2)-ethenoguanine, a mutagenic DNA adduct, is a primary substrate of *Escherichia coli* mismatch-specific uracil-DNA glycosylase and human alkylpurine-DNA-N-glycosylase. *J Biol Chem*. 2002;**277**:26987–93.
- 20 Ahn WC, Aroli S, Kim JH, Moon JH, Lee GS, Lee MH, et al. Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision. *Nat Chem Biol*. 2019;**15**:607–14.
- 21 Sang PB, Srinath T, Patil AG, Woo EJ, Varshney U. A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily. *Nucleic Acids Res*. 2015;**43**:8452–63.
- 22 Tu J, Chen R, Yang Y, Cao W, Xie W. Suicide inactivation of the uracil DNA glycosylase UdgX by covalent complex formation. *Nat Chem Biol*. 2019;**15**:615–22.
- 23 Jia Q, Zeng H, Tu J, Sun L, Cao W, Xie W. Structural insights into an MsmUdgX mutant capable of both crosslinking and uracil excision capability. *DNA Repair*. 2021;**97**:103008.
- 24 Cao W. Endonuclease V: an unusual enzyme for repair of DNA deamination. *Cell Mol Life Sci*. 2013;**70**:3145–56.
- 25 Yang Y, Park SH, Alford-Zappala M, Lee HW, Li J, Cunningham RP, et al. Role of endonuclease III enzymes in uracil repair. *Mutat Res*. 2019;**813**:20–30.
- 26 Im EK, Han YS, Chung JH. Functional changes in a novel uracil-DNA glycosylase determined by mutational analyses. *Mikrobiologiia*. 2008;**77**:644–50.
- 27 Chung JH, Im EK, Park HY, Kwon JH, Lee S, Oh J, et al. A novel uracil-DNA glycosylase family related to the helix-hairpin-helix DNA glycosylase superfamily. *Nucleic Acids Res*. 2003;**31**:2045–55.
- 28 Cox MM, Battista JR. *Deinococcus radiodurans* – the consummate survivor. *Nat Rev Microbiol*. 2005;**3**:882–92.
- 29 Sandigursky M, Sandigursky S, Sonati P, Daly MJ, Franklin WA. Multiple uracil-DNA glycosylase activities in *Deinococcus radiodurans*. *DNA Repair*. 2004;**3**:163–9.
- 30 Moe E, Leiros I, Smalas AO, McSweeney S. The crystal structure of mismatch-specific uracil-DNA glycosylase (MUG) from *Deinococcus radiodurans* reveals a novel catalytic residue and broad substrate specificity. *J Biol Chem*. 2006;**281**:569–77.
- 31 King K, Benkovic SJ, Modrich P. Glu-111 is required for activation of the DNA cleavage center of EcoRI endonuclease. *J Biol Chem*. 1989;**264**:11807–15.
- 32 Vermote CL, Halford SE. EcoRV restriction endonuclease: communication between catalytic metal ions and DNA recognition. *Biochemistry*. 1992;**31**:6082–9.
- 33 Gao GJ, Lu HM, Huang LF, Hua YJ. Construction of DNA damage response gene pprI function-deficient and function-complementary mutants in *Deinococcus radiodurans*. *Chin Sci Bull*. 2005;**50**:311–6.
- 34 Makarova KS, Aravind L, Wolf YI, Tatusov RL, Minton KW, Koonin EV, et al. Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. *Microbiol Mol Biol Rev*. 2001;**65**:44–79.
- 35 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;**596**:583–9.
- 36 Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold – making protein folding accessible to all. *bioRxiv*. 2021. <https://doi.org/10.1101/2021.08.15.456425>.
- 37 Akdel M, Pires DEV, Pardo EP, Jänes J r, Zalevsky AO, Mészáros B I, et al. A structural biology community assessment of AlphaFold 2 applications. *bioRxiv*. 2021. <https://doi.org/10.1101/2021.09.26.461876>.
- 38 Nelson D, Cox MM. *Lehninger principles of biochemistry*. New York, NY: Macmillan Learning; 2004.
- 39 Hoseki J, Okamoto A, Masui R, Shibata T, Inoue Y, Yokoyama S, et al. Crystal structure of a family 4 uracil-DNA glycosylase from *Thermus thermophilus* HB8. *J Mol Biol*. 2003;**333**:515–26.
- 40 Sweet DM, Moseley BE. The resistance of micrococcus radiodurans to killing and mutation by agents which damage DNA. *Mutat Res*. 1976;**34**:175–86.
- 41 Liu Y, Zhou J, Omelchenko MV, Beliaev AS, Venkateswaran A, Stair J, et al. Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. *Proc Natl Acad Sci USA*. 2003;**100**:4191–6.
- 42 Fisher CL, Pei GK. Modification of a PCR-based site-directed mutagenesis method. *Biotechniques*. 1997;**23**:570–1. 574.
- 43 Li J, Yang Y, Guevara J, Wang L, Cao W. Identification of a prototypical single-stranded uracil DNA glycosylase from listeria innocua. *DNA Repair*. 2017;**57**:107–15.

- 44 McIlvaine TC. A buffer solution for colorimetric comparison. *J Biol Chem.* 1921;**49**:183–6.
- 45 Kim M, Wolff E, Huang T, Garibyan L, Earl AM, Battista JR, et al. Developing a genetic system in *Deinococcus radiodurans* for analyzing mutations. *Genetics.* 2004;**166**:661–8.
- 46 Hua X, Xu X, Li M, Wang C, Tian B, Hua Y. Three nth homologs are all required for efficient repair of spontaneous DNA damage in *Deinococcus radiodurans*. *Extremophiles.* 2012;**16**:477–84.
- 47 Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;**46**:W296–303.