Published online 28 March 2017

Nucleic Acids Research, 2017, Vol. 45, No. 10 ***5863–5876***
doi: 10.1093/nar/gkx209

Uracil DNA glycosylase (UDG) activities in *Bradyrhizobium diazoefficiens*: characterization of a new class of UDG with broad substrate specificity

Ullas Valiya Chembazhi[1,†], Vinod Vikas Patil[2,3,†], Shivjee Sah[1], Wayne Reeve[4], Ravi P. Tiwari[4], Euijeon Woo[2,3,*] and Umesh Varshney[1,4,5,*]

[1]Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560012, India, [2]Disease Target Structure Research Center, Korea Research Institute of Bioscience and Biotechnology, 125 Gwahak-Ro, Yuseon-Gu, Daejeon 34141, Republic of Korea, [3]Department of Bio-Analytical Science, University of Science and Technology, Daejeon 34113, Republic of Korea, [4]Centre for Rhizobium Studies, School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA 6150, Australia and [5]Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560064, India

Received August 15, 2016; Revised March 17, 2017; Editorial Decision March 17, 2017; Accepted March 27, 2017

ABSTRACT

Repair of uracils in DNA is initiated by uracil DNA glycosylases (UDGs). Family 1 UDGs (Ung) are the most efficient and ubiquitous proteins having an exquisite specificity for uracils in DNA. Ung are characterized by motifs A (GQDPY) and B (HPSPLS) sequences. We report a novel dimeric UDG, Blr0248 (*Bdi*Ung) from *Bradyrhizobium diazoefficiens*. Although *Bdi*Ung contains the motif A (GQDPA), it has low sequence identity to known UDGs. *Bdi*Ung prefers single stranded DNA and excises uracil, 5-hydroxymethyl-uracil or xanthine from it. *Bdi*Ung is impervious to inhibition by AP DNA, and Ugi protein that specifically inhibits family 1 UDGs. Crystal structure of *Bdi*Ung shows similarity with the family 4 UDGs in its overall fold but with family 1 UDGs in key active site residues. However, instead of a classical motif B, *Bdi*Ung has a uniquely extended protrusion explaining the lack of Ugi inhibition. Structural and mutational analyses of *Bdi*Ung have revealed the basis for the accommodation of diverse substrates into its substrate binding pocket. Phylogenetically, *Bdi*Ung belongs to a new UDG family. *Bradyrhizobium diazoefficiens* presents a unique scenario where the presence of at least four families of UDGs may compensate for the absence of an efficient family 1 homologue.

INTRODUCTION

Uracil (as dUMP) can be incorporated into DNA by DNA polymerases (from dUTP, forming an A:U pair) or arise by deamination of cytosines (resulting in G:U pairs) (1,2). If unrepaired, A:U base pairs can hinder recognition by DNA binding proteins (3,4) while G:U base pairs can cause G:C to A:T mutations in the subsequent rounds of DNA replication, posing a threat to genomic integrity. Enzymes of the uracil DNA glycosylase (UDG) superfamily initiate base excision repair to remove uracil residues (5–7). Five distinct families of UDGs have so far been established which differ in their primary amino acid sequences and substrate specificities. However, UDGs possess a conserved α/β structural fold and seemingly a common evolutionary origin (8,9). Ubiquitously found family 1 UDGs (Ung), exemplified by the well-studied 'master catalyst' *Eco*Ung, are the most efficient of UDGs and show an exquisite specificity for uracil in DNA (10). Conserved orthologues of family 1 are found in nearly all organisms studied, including many viruses (5,10). This UDG family is characterized by motif A (GQDPY) responsible for catalysis and motif B (HPSPLS) implicated in the stabilization of the enzyme substrate complex (11–14). Interestingly, phage PBS-1 or 2 that infects *Bacillus subtilis* encodes a protein inhibitor, Ugi that forms a physiologically irreversible non-covalent complex in 1:1 stoichiometry specifically with family 1 UDGs to allow occurrence of uracil in their genomes (15,16).

While *Eco*Ung was the first UDG to be identified and extensively studied, several UDGs have been reported since then (5,17–21). Family 2 UDGs referred to as Mug/TDG have GINPG as motif A and MPSSSAR as motif B. While Mug/TDG excise uracil from double stranded (ds) DNA

*To whom correspondence should be addressed. Tel: +91 80 22932686, Fax: +91 80 23602697; Email: varshney@mcbl.iisc.ernet.in
Correspondence may also be addressed to Euijeon Woo. Tel: +82 428798432, Fax: +82 428798596; Email: ejwoo@kribb.re.kr
†These authors contributed equally to this work as first authors.

(G:U and A:U pairs) with a low efficiency, they also excise thymine from T:G pairs (18,22–24). GMNPG and HPSPRN define motif A and motif B for family 3 UDGs (SMUG). SMUGs excise uracil from single stranded (ss) DNA with higher efficiency than from dsDNA (19,25). Family 4 and family 5 UDG members contain a 4Fe–4S cluster. Motif A and Motif B are represented by GE(A/G)PG and HPAAVL, respectively, for family 4, and by GLAPA and HPSPLN for family 5 (20,21,26,27). While family 4 UDGs act on ssDNA and dsDNA substrates, UDG activity of family 5 is enhanced on dsDNA substrates.

Members of Rhizobiaceae are an important group of soil bacteria that are recognized for their ability to fix atmospheric nitrogen through a symbiotic association with root nodules of leguminous plants. Formation and maintenance of root nodules in leguminous plants represents an interesting scenario of oxidative and nitrosative stress to invading rhizobial cells. To establish symbiotic associations with leguminous plants, rhizobia should efficiently overcome the initial defense response elicited by the plant. Similar to the response to pathogens, plants also respond to rhizobial infection through an oxidative burst, with enhanced production of superoxide and H_2O_2 (28–30). Further, high rates of respiration to support the process of nitrogen fixation along with autoxidation of leghemoglobins in root nodules can produce high levels of reactive oxygen species (ROS) (31–41). In addition to ROS, nodules also generate reactive nitrogen species (RNS) like nitric oxide (NO) and peroxynitrite ($ONOO^-$) (42,43).

Oxidative and nitrosative stress result in various kinds of lesions in DNA including deamination of cytosines, posing a threat to the genomic integrity of rhizobia (44,45). Interestingly, their genome contains high G+C content making them inherently more susceptible to cytosine deamination events that produce uracil in the genome which can lead to transition mutations. Hence, one could expect these organisms to have evolved sophisticated strategies of DNA repair to cope with their hostile environment. However, aspects of DNA repair in rhizobium are currently under explored. In the present study, we used a member of α-proteobacteria, *Bradyrhizobium diazoefficens* (previously *B. japonicum*) (46) as our model organism, which is well known for its ability to form nitrogen fixing root nodules through a symbiotic association with *Glycine max* (47,48) to explore aspects of uracil excision repair and report the presence of at least four UDGs in this organism with a distinct absence of a classical family 1 UDG. Furthermore, we demonstrate that in *Bradyrhizobium diazoefficens* there exists a unique base excision repair system catalyzed by a member of a new UDG family.

MATERIALS AND METHODS

Bacterial strains, plasmids, DNA oligomers, media and growth conditions

Bacterial strains and plasmids are listed in Supplementary Table S1. DNA oligomers are listed in Supplementary Table S2. *Escherichia coli* CC102 (49) was obtained from Coli Genetic Stock Centre (CGSC). *E. coli* strains were grown in Luria-Bertani broth (LB) or on LB containing 2% (w/v) agar (Difco, USA) at 37°C. *Bradyrhizobium diazoefficiens* was grown in TY medium (0.5% tryptone, 0.3% yeast extract, 6 mM $CaCl_2$) or on TY with 2% agar at 30°C. Ampicillin (Amp), kanamycin (Kan) and chloramphenicol (Cm) were added to media at a concentration of 100, 25 and 30 μg ml^{-1} respectively.

Cell lysate preparation

Saturated cultures were obtained in 2 ml media by incubation at 37°C overnight for *E. coli* strains and at 30°C for 5–6 days for *B. diazoefficiens* and then subcultured (1% for *E. coli* and 10% for *B. diazoefficiens*) into 2 ml media and incubated until the OD_{600} reached ∼0.6. Cells were pelleted, re-suspended in 200 μl TME (25 mM Tris–HCl pH 8.0, 2 mM β-mercaptoethanol, 1 mM Na_2EDTA), sonicated with 2 s alternate pulses for 20 s, six to seven times with 50–60 s intervals and centrifuged at 13 000 rpm for 1.5 h on table top centrifuge at 4°C. The supernatant was collected, mixed with an equal volume of 2× storage buffer (80% glycerol, 1 mM β-mercaptoethanol, 20 mM Tris–HCl pH 8.0 and 200 mM NaCl) and stored at –20°C. Protein concentration in cell extracts was estimated by the Bradford assay using BSA as standard (50).

Cloning of putative UDGs

Bdi-ung ORF was PCR amplified from genomic DNA with forward (*Bdi*Ung Fp, 5′-GCATTTCATATGCTCACAG AATTCG-3′) and reverse (*Bdi*Ung Rp, 5′-GCCCAAAGC TTCGGGTTTGCGGCCG-3′) primers containing NdeI and HindIII sites, respectively and Pfu DNA polymerase. The reaction was heated at 94°C for 4 min followed by 35 cycles of incubations at 94°C for 1 min, 50°C for 45 s and 70°C for 50 s. The PCR product (856 bp) was ligated into pTrcNdeIHis vector through NdeI and HindIII sites, such that a hexa-His tag is attached to the N-terminus of the protein. The clones were confirmed by DNA sequencing. Other putative UDG genes, including *bll3023*, *blr5068* and *blr6661,* were also similarly cloned into pTrcNdeIHis vector (with hexa-His tag, through NdeI and HindIII sites).

Purification of proteins

For protein overexpression and purification, plasmid pTrc*Bdi*Ung was introduced into *E. coli* Δ*ung* strain by transformation. The *E. coli* TG1Δ*ung*::*cm* harbouring pTrc*Bdi*Ung was inoculated into 10 ml LB containing Amp and Cm and grown overnight. The inoculum was diluted 1:100 (v/v) (1%) into 600 ml fresh medium and grown at 37°C to an OD_{600} of ∼0.6. The culture was induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and grown for another 4–5 h. Cells were harvested by centrifugation (7500 rpm, 4°C, 5 min), resuspended in buffer A (20 mM Tris–HCl pH 8.0, 500 mM NaCl, 10% glycerol v/v, 2 mM β-mercaptoethanol) containing 30 mM imidazole, sonicated and centrifuged at 29 000 rpm (Avanti™ J-30I, JA30.50 Ti) for 2 h 30 min at 4°C. The supernatant was loaded onto a 5 ml Ni-NTA column pre-equilibrated with buffer A, washed with 30 ml of buffer A and eluted with a linear gradient of 30 mM to 1 M imidazole in buffer A.

Aliquots from the fractions were analyzed on 15% SDS-PAGE, fractions containing pure protein were pooled, dialyzed against buffer A, then against storage buffer (20 mM Tris–HCl pH 8.0, 300 mM NaCl, 50% glycerol v/v, 2 mM β-mercaptoethanol), and stored at –20°C. Protein concentration was estimated by the Bradford assay using BSA as standard (50). Other UDGs were also purified using the same procedure, except that Bll3023 was purified from *E. coli* TG1, cultured in the presence of 0.01% $FeCl_3$.

DNA oligomers and their radiolabelling

The 24 nucleotide long ssU9 (ctcaagtgUaggcatgcaagagct) and the 37 nucleotide long tetra-loop hairpin GU9 (ctcaagtgUaggcatgcttttgcatgcctgcacttga) containing uracil in the stem region, were used as ssDNA and dsDNA substrates, respectively. Other DNA oligomers are listed in Supplementary Table S2. DNA oligomers (10 pmol) were 5′ ^{32}P-end labelled using 10 μCi of [γ-^{32}P] ATP (6000 Ci/mmol) and T4 polynucleotide kinase and purified on Sephadex G-50 minicolumns (1 ml). To generate duplexes (HmU and ssX9 with ssU9-Compl-G, Compl-T with G-oligo, and I with Compl-G), 5′ ^{32}P- end labelled DNA oligomers (with modified bases) were annealed with complementary oligomers in approximately equimolar ratios in 40 mM Tris–HCl pH 7.5, 20 mM $MgCl_2$ and 50 mM NaCl, heated at 90°C for 5 min and then slowly cooled to room temperature.

Base excision and inhibition assays

Appropriate amounts of cell extract/pure protein were incubated with 5′ ^{32}P-end labelled DNA oligomers (∼10000 cpm; ∼10 nM concentration) in 10 μl reaction volumes in UDG buffer (50 mM Tris–HCl pH 8.0, 1 mM Na_2EDTA, 1 mM DTT and 25 μg ml^{-1} of BSA) at 37°C for the indicated times. The reactions were stopped by the addition of 8 μl each of 0.2 N NaOH and formamide dye (80% formamide, 0.05% each of bromophenol blue and xylene cyanol FF, 10 mM NaOH and 2 mM Na_2EDTA) and boiled at 90°C for 15 min. Aliquots (20 μl) were analyzed on 15% polyacrylamide (19:1) 8 M urea gels, exposed to phosporimaging screens and imaged using Fujifilm analyzer. To see if *Bdi*Ung is inhibited by Ugi, *Bdi*Ung (1 μg, ∼32 pmol of monomer) was pre-incubated with excess Ugi (1 μg; ∼100 pmol) in 4 μl volume for 30 min before the reaction.

Gel filtration chromatography

Proteins were subjected to gel filtration chromatography using Superdex 75 (GE Healthcare) column equilibrated with a buffer containing 20 mM Tris–HCl pH 8.0, 150 mM NaCl, 2 mM β-mercaptoethanol, 10% glycerol. Elution was performed with the same buffer at a flow rate of 1 ml/min and the absorbance at 280 nm was recorded to obtain the elution profile.

EMSA for checking Ugi binding

Ugi (∼8 μg, ∼850 pmol) was incubated with *Bdi*Ung (5 μg, ∼160 pmol of monomer) or *Eco*Ung (2 μg, ∼80 pmol) in 25 μl volumes in UDG buffer at room temperature for 20 min and then stored on ice (4°C) for 15 min. Loading dye (5 μl) containing 30% (v/v) glycerol, 0.25% (w/v) bromophenol blue and 0.25% (w/v) xylene cyanol FF was added, and the samples were analyzed on 15% native PAGE (29:1 crosslinking, pH 6.8). After electrophoresis, the gel was stained using Coomassie brilliant blue.

Impact of uracil and AP DNA on activity of *Bdi*Ung

To check the effect of free uracil and AP DNA on uracil excision enzymes, varying concentrations of free uracil or unlabelled ssF9 (ctcaagtgFaggcatgcaagagct, where F is tetrahydrofuran, a stable mimic of AP site provided by Midland Certified Reagent Co.) were pre-incubated with 100 pg (∼4 fmol) *Eco*Ung or 10 ng (∼0.32 pmol monomer) *Bdi*Ung at room temperature followed by incubation on ice for 15 min. The reaction mixtures were followed by addition of 5′ end ^{32}P-labelled ssU9 (∼0.1 pmol) and incubated at 37°C for 30 min for uracil excision.

Generation of site directed mutation in *Bdi*Ung, and its characterization

PCR based method (see supplementary material) was used to generate A59Y mutant of *Bdi*Ung. Mutant protein was purified from *E. coli* TG1Δ*ung::cm* strain using Ni-NTA column chromatography as described for the wild type *Bdi*Ung.

Generation of *ung* knockout in CC102 strain

A P1 phage lysate generated using *E. coli* MG1655Δ*ung*::*kan* was used to transduce the Δ*ung*::*kan* allele into *E. coli* strain CC102. Transfer of the Δ*ung::kan* allele into the transductants was verified by diagnostic PCR (with *Eco*Ung-up-Fp and *Eco*Ung-dn-Rp primers which flank the *ung* locus) and uracil excision assays (Supplementary Figure S1).

Assay for reversion of Lac$^-$ to Lac$^+$ in *E. coli*

E. coli CC102 with a mutation in *lacZ* was used to assay for reversion to Lac$^+$. Isolated colonies of this strain and its Δ*ung*::*kan* derivative were transformed with pTrc99c or *Eco*Ung/*Bdi*Ung expression constructs. Cells were inoculated in replicates of ten in 2 ml LB supplemented with Amp and grown for 24 h with shaking at 37°C. Serial dilutions of saturated cultures were made and 50 μl of 10^{-5} dilution was plated onto M9 minimal media plates containing 0.2% glucose for viable counts. The saturated culture (1.5 ml) was spun down and plated onto M9 minimal media containing 0.2% lactose for counts of Lac$^+$ revertants. The plates were incubated at 37°C until colonies appeared (24 h for glucose plates and 72 h for lactose plates). The reversion frequency was obtained by dividing the number of revertants (per ml culture) by the number of bacteria plated (viable count).

Crystallization and refinement

The purified *Bdi*Ung was dialyzed against 20 mM Tris–HCl, 50 mM NaCl, 5 mM β-mercaptoethanol, pH 8.5 and

concentrated to 10 mg/ml before freezing in liquid nitrogen and storage at -70°C. Crystals of *Bdi*Ung were generated using sitting-drop vapor-diffusion in 20% PEG3350, 200 mM sodium citrate, 100 mM sodium citrate/citric acid pH 4.0 with a protein:precipitant ratio of 1:1. *Bdi*Ung was co-crystallized with ligands (uracil, and xanthine) using 9:1 ratio (*Bdi*Ung:saturated solution of uracil or xanthine) with an incubation period of 15 min before crystallization setup. Before data collection, the crystals were cryocooled to 93 K using a cryoprotectant consisting of mother liquor with 30% glycerol. The native *Bdi*Ung crystal diffracted at 2 Å, selenomethionine-labelled *Bdi*Ung crystal at 2.3 Å, *Bdi*Ung-uracil crystal at 2.9 Å and *Bdi*Ung-xanthine crystal at 2.8 Å. The diffraction data were processed and scaled using HKL2000 (53). Structure refinement and model building were performed using the programs COOT (54) and PHENIX (55). The structure was visualized using Pymol software (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

RESULTS

Identification of UDG homologues and Blr0248 (*Bdi*Ung) as putative UDGs

Analysis of *B. diazoefficiens* USDA110 genome through blast searches indicated the absence of a conserved family 1 homologue in this organism. Further analysis based on motif based BLAST (NCBI) searches revealed many putative UDGs (Table 1), among which most could be classified into established families of UDGs: *blr6661* (family 2 MUG), *bll2523* (family 4 UdgA), *blr5068* (family 5 UdgB) and *bll3023* (family 4 UdgX). Of particular interest among the identified ORFs was *blr0248* which codes for a 272 amino acid long protein Q89XRO_BRAJA containing a putative motif A (GQDPA) similar to family 1 UDGs (referred to as *Bdi*Ung). However, *Bdi*Ung lacked the motif B of family 1 UDGs and showed low sequence similarity to established UDG families (Supplementary Figure S2). Homology searches through BLAST revealed several homologues of *Bdi*Ung from many organisms (Figure 1), which form the putative members of a new family of UDGs. Homologues of *Bdi*Ung were found in various lineages of bacteria including actinobacteria (*Streptacidiphilus rugosus*, *Kitasatospora azatica*, *Intrasporangium calvum*, *Streptomyces xylophagus*, *Frankia* sp. EuI1c and *Janibacter* sp. HTCC2649) and β-proteobacteria (*Burkholderia ubonensis*, *Chromobacterium haemolyticum*, *Nitrosospira* sp. NpAV). Evolutionary analysis based on neighbor joining method (Supplementary Figure S3) showed that these bacteria possess a new family of UDGs related to family 1.

Cloning and purification of putative UDGs

We cloned ORFs of *blr0248, blr6661, blr5068* and *bll3023* from *B. diazoefficiens* USDA110 into the pTrcNdeIHis expression vector along with a 20 amino acid long presequence containing a hexa-His tag. All UDGs other than Bll3023 were purified from an *E. coli* TG1 *Δung*::*cm* strain to ensure that the UDG preparations were free from *Eco*Ung contamination (Supplementary Figure S4). Since UdgX proteins are toxic in *ung*$^{-}$ strains (56), Bll3023 was purified from *E. coli* TG1. As shown in Supplementary Figure S5, Bll3023, Blr6661, Blr5068 possessed uracil excision activities, which corresponded to their known classes of proteins, and were not pursued any further.

*Bdi*Ung: a novel Ugi insensitive UDG with broad substrate specificity

The activity of purified *Bdi*Ung was analyzed on uracil containing ssDNA and dsDNA oligomers (ssU9 and GU9) and was found to be active on both the substrates (Figure 2A and B, compare lanes 2 with 1). *Bdi*Ung reactions with/without preincubation with Ugi showed product bands of similar intensity (compare lanes 2 and 3). However, under the same conditions, preincubation of *Eco*Ung with Ugi completely inhibited uracil excision activity (compare lanes 5 with 4 and 1). These observations suggest that Ugi does not inhibit *Bdi*Ung to any detectable levels and reveals a fundamental difference of *Bdi*Ung to family 1 UDG members.

Although the activity of *Bdi*Ung was not inhibited by Ugi, it was still possible that it formed a complex with Ugi (like family 1 UDGs). Therefore, electrophoretic mobility shift assays were carried out using native PAGE after pre-incubation of *Bdi*Ung or *Eco*Ung with Ugi (Figure 2C). As a control, *Eco*Ung–Ugi complex formed *in vivo* was also loaded (lane 4). While pre-incubation of *Eco*Ung with Ugi resulted in a shift in the migration of *Eco*Ung in native PAGE (compare lane 2 with lanes 1 and 3), *Bdi*Ung did not show any shift in its migration (nor a new band) upon pre-incubation with Ugi (compare lanes 5 and 6) suggesting the absence of complex formation. The *Eco*Ung-Ugi (lane 2) complex migrated to the same position for purified *Eco*Ung–Ugi complex (lane 4) formed *in vivo* (57).

Even though the bioinformatics and biochemical analyses with *Bdi*Ung suggested that *B. diazoefficiens* lacks a classical family 1 UDG, to rule out other unidentified proteins as family 1 UDG (Ung), we decided to carryout UDG assays with cell free extracts prepared from *B. diazoefficiens* (Figure 2D). As expected, the assays revealed UDG activity (compare lane 5 with 1). However, the activity was not diminished by Ugi (compare lane 5 with lanes 6 and 7). As a control, UDG activity of *E. coli* cell free extracts was inhibited by Ugi (compare lane 2 with lanes 3 and 4). These observations suggest that *B. diazoefficiens* lacks Ugi sensitive UDG activity.

To check if *Bdi*Ung is inhibited by either of the reaction products (uracil, or AP DNA) we employed a uracil excision assay with 5′ end ^{32}P-labelled ssU9, in the presence of varying concentrations of uracil or a tetrahydrofuran-containing DNA (ssF9), a stable mimic of AP DNA (Figure 3). While free uracil caused a 5-fold reduction in *Bdi*Ung activity at 10 mM concentration (panels i and ii), ssF9, in spite of its high concentrations (of up to 200 μM) did not significantly inhibit *Bdi*Ung (panels iii and iv). As a control, both the products inhibited uracil excision activity of *Eco*Ung (panels i–iv).

Substrate specificity of *Bdi*Ung

To investigate if *Bdi*Ung is highly specific towards uracil in DNA (like family 1 UDGs, Ung) or relaxed in substrate

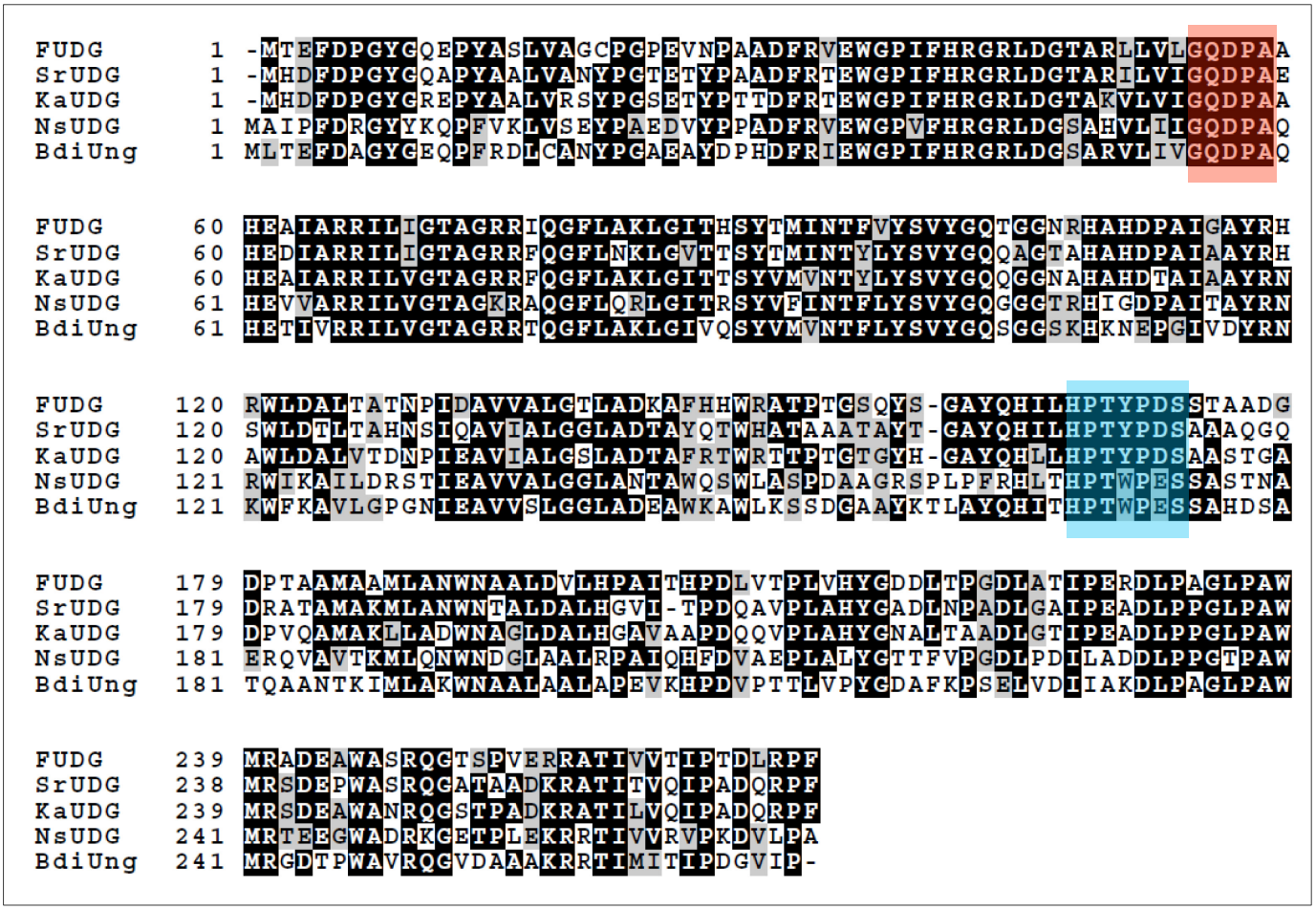

Figure 1. Multiple sequence alignment of *Bdi*Ung with putative homologues. The protein sequence of *Bdi*Ung was aligned against the protein sequences of *F*UDG (from *Frankia* sp. EuI1c), *Sr*UDG (from *Streptacidiphilus rugosus*), *Ka*UDG (from *Kitasatospora azatica*) and *Ns*Ung (from *Nitrosospira* sp. NpAV) using clustal omega. The schematic representation was obtained by using BOXSHADE (www.ch.embnet.org). Identical residues are shaded in black while similar residues in grey. Putative motif A and B are additionally shaded in red and blue, respectively.

Table 1. Putative UDGs identified from *Bradyrhizobium diazoefficiens* through genome analysis

Gene	Protein	Motif
blr0248	Q89XRO_BRAJA	GQDPA
bll2523	Q89S82_BRAJA	UDG F4 TTUDGA like
bll3023	Q89QV1_BRAJA	UDG F4 TTUDGA like
blr5068	Q89K49_BRAJA	UDG F5 TTUDGB like
blr6661	Q89FNB_BRAJA	UDG F2 MUG

choice like UDGs belonging to some other families, we used 5′ end ^{32}P-labelled DNA oligomers containing different base modifications. *Bdi*Ung utilized DNA oligomers containing xanthine (X) (Figure 4A, compare lane 8 with 7) and a uracil analogue 5-hydroxymethyluracil (HmU) (Figure 4A, compare lane 10 with 9) in ssDNA as substrates. However, it did not utilize DNA oligomer containing inosine (I) in ssDNA or dsDNA contexts to any detectable levels (Figure 4A and B). Thymine DNA glycosylase activity was also not detected (over the control reaction) using a G:T pair containing dsDNA (Figure 4B, lanes 3 and 4).

To determine the minimum substrate from which *Bdi*Ung could excise uracil, we used 5' end ^{32}P-labelled DNA oligomers having a different number of bases on the 5' and 3' sides of the uracil (Figure 5A). *Eco*Ung whose activity on these substrates has been well characterized (51) was used as control. With reference to *Eco*Ung, *Bdi*Ung activity was detectable on oligomers having one phosphorylated nucleotide 5' of the uracil (compare lane 9 with 8) and the two nucleotides 3' of it (compare lane 3 with 2). However, excision of uracil from the 5' terminal or the penultimate position (from the 3' end) was undetectable. Thus, the minimum size substrate required for detectable excision of uracil by *Bdi*Ung is pNUNN (Figure 5A).

Genomes with a higher G+C content are inherently prone to forming stable intra-molecular structures such as the hairpin loops due to various topological changes during physiological processes like transcription and replica-

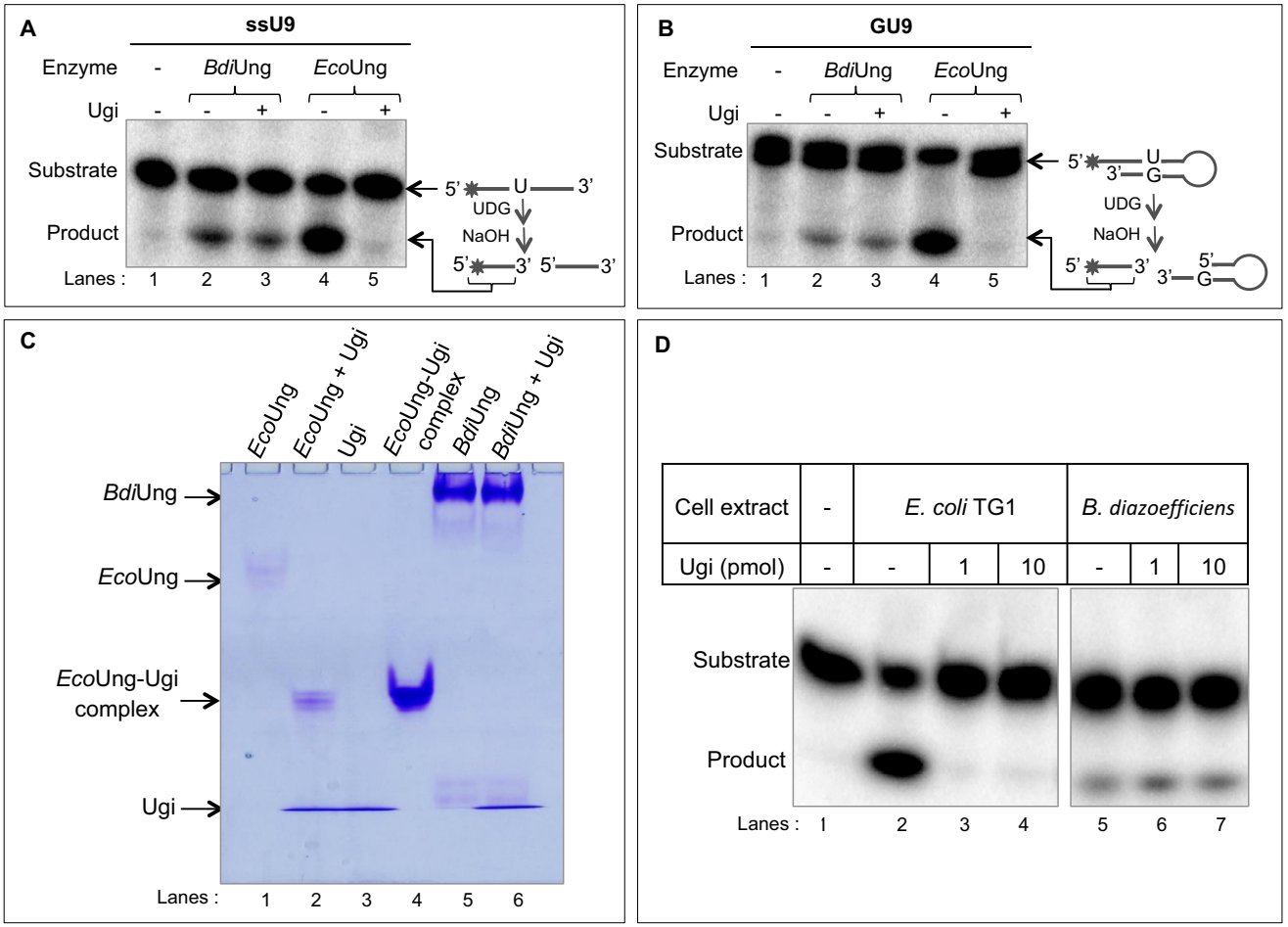

Figure 2. Uracil excision assays with single stranded and double stranded substrates. Assays were carried out at 37°C for 30 min with 1 μg of *Bdi*Ung (~32 pmol monomer) or *Eco*Ung (~4 pmol) with (+) or without (–) pre-incubation with 1 μg of Ugi (~100 pmol) as described in Materials and Methods using ~0.1 pmol of ssU9 (**A**) or GU9 (**B**) DNAs, and resolved on 8 M urea PAGE (15%). (**C**) Electrophoretic mobility shift assay to analyze complex formation of *Bdi*Ung with Ugi. After incubation of enzyme with/without Ugi, reaction products were resolved on 15% native PAGE (29:1 crosslinking, pH 6.8). While complex *Eco*Ung-Ugi complex formed is found to migrate adjacent to pure complex control, no band indicative of *Bdi*Ung-Ugi complex could be seen. (**D**) Uracil excision assay on ssU9 with 2 μg of *B. diazoefficiens* cell free extracts with/without pre-incubation with varying amounts of Ugi. *E. coli* TG1 cell free extract (2 μg) was used as a control. Reaction products were resolved on 8 M urea PAGE (15%).

tion. The hairpin like structures formed on single stranded regions may lead to the appearance of unpaired C residues highly prone to deamination. In addition, in the loop structures such as the tetraloops, uracils may be fixed into extrahelical (flipped out) positions (52,58). Earlier studies have shown inefficient excision of uracils from the tetraloops by *Eco*Ung (59). However, the uracil excision activity of Ung from *Mycobacterium smegmatis*, a G+C rich organism (like *B. diazoefficiens*) was found to be more efficient than that of *Eco*Ung (60). Thus, we checked the uracil excision activity of *Bdi*Ung on four hairpin DNA structures that contained uracil at different positions of a tetraloop hairpin (Figure 5B). *Bdi*Ung excised uracil from all the loop positions. However, unlike *Eco*Ung, which showed a significant decrease in excision of uracil from the U1-hairpin, U2-hairpin and U3-hairpin substrates when compared with the excision of uracil from ssDNA (compare lanes 5, 8 and 11 with lane 2), *Bdi*Ung was not as severely compromised on the same hairpin substrates when compared with its activity on ssDNA (compare lanes 6, 9, 12 with lane 3). Excision of uracil by *Bdi*Ung from U4-hairpin was as efficient as its excision of ssDNA (compare lane 15 with 3).

Overall structure of *Bdi*Ung

The structure of *Bdi*Ung was determined using the single anomalous dispersion (SAD) method. The detailed statistics for the structure are described in Table 2. In agreement with the gel filtration chromatography results (Supplementary Figure S6), the overall structure is a dimer with a typical α/β/α sandwich fold of UDG proteins for each monomer such that a central β-sheet is layered by α-helixes on both sides (Figure 6A and B). Using the Dali server, the nearest structural homologue occurred with a family 4 UDG from *Sulfolobus tokadaii* (sequence identity, 12%, Supplementary Figure S7) with a *Z*-score of 14.4 and a 1.1 Å rmsd over 31 Cα atoms (61). However, unlike family 4 UDGs, *Bdi*Ung lacks the conserved cysteine residues for metal chelation and hence does not contain an iron-sulfur cluster [4Fe–S] in the structure, which is a charac-

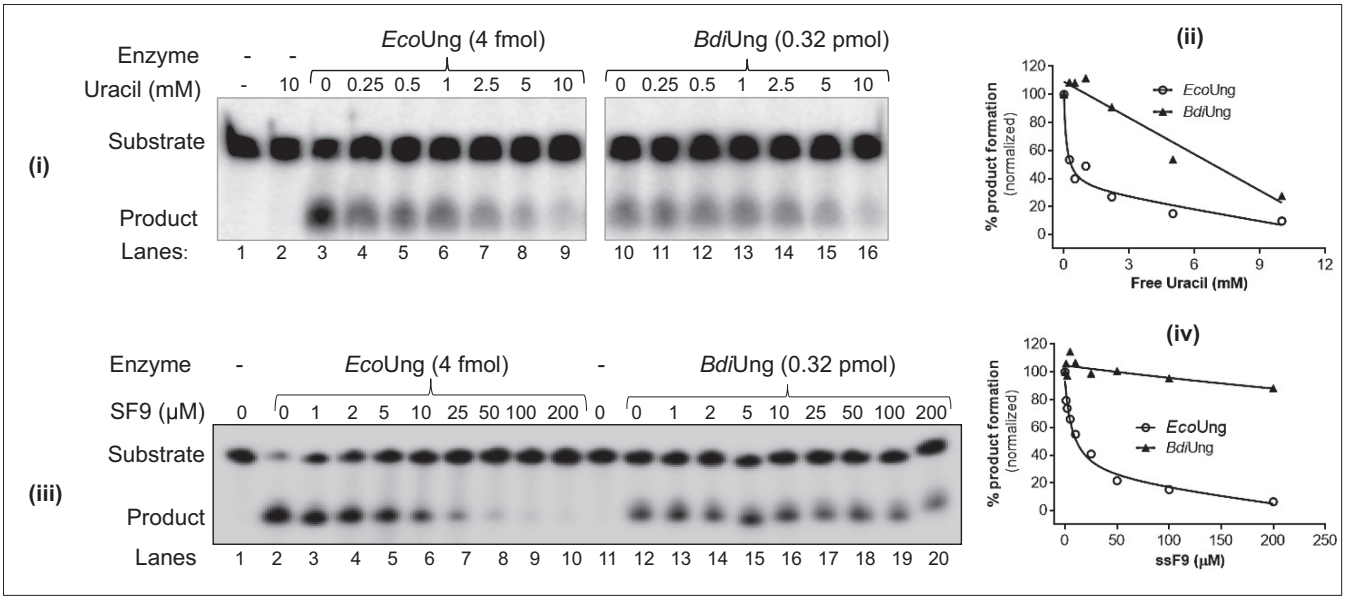

Figure 3. Uracil excision assays for product inhibition. Uracil excision assay using ∼0.1 pmol 5′ end ^{32}P-labelled ssU9 substrate with the indicated amounts of *Eco*Ung or *Bdi*Ung in the presence of varying concentrations of uracil (i and ii), and ssF9 (a stable AP DNA mimic harbouring tetrahydrofuran at position 9) (iii and iv) were carried out as described in Materials and Methods. Reactions were resolved on 8M urea PAGE (15%) and imaged. The estimated values (%) of substrate (S) to product (P) conversion [P/(S + P) × 100] are plotted against the concentrations of the inhibitor in the reaction. The data were fitted with the one site-total of saturation binding equations ($Y = B_{max} * X/(K_d + X)$ + NS*X + background) (where Y is specific binding/activity, B_{max} is the maximum specific binding/activity in the same units as Y, K_d is the equilibrium binding constant in the same units as X, NS is the slope of nonspecific binding in Y unit divided by X unit and background is the amount of nonspecific binding) using GraphPad Prism software. Amounts of product formed in control (without free uracil/AP DNA) reaction in panels i and iii were scaled to 100%, and the remaining activities shown relative to this reference to generate plots in panels ii and iv.

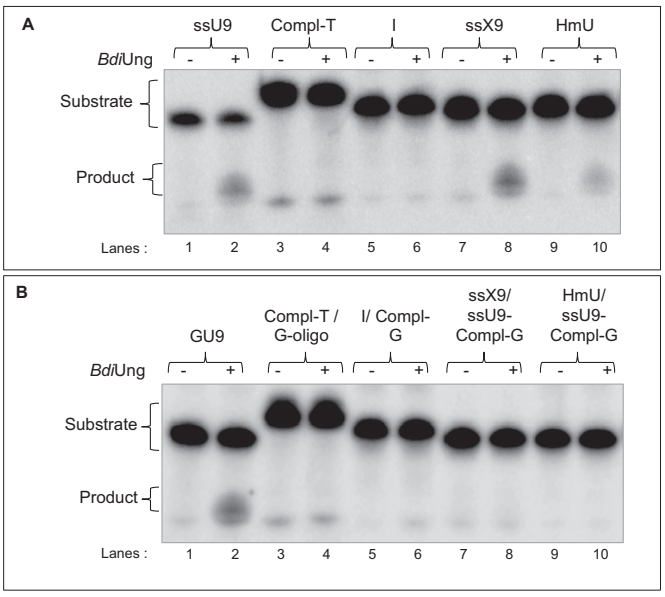

Figure 4. *Bdi*Ung shows broad substrate specificity. Excision assays using *Bdi*Ung (1 μg, ∼32 pmol monomer) with (**A**) ssDNA, and (**B**) dsDNA containing modified bases (∼0.1 pmol each); inosine (I), xanthine (X) and 5-hydroxymethyluracil (HmU) were carried out at 37°C for 30 min as described in Materials and Methods. The reactions were resolved on 8 M urea PAGE (15%) and imaged.

teristic feature of family 4 UDGs. The superposition with family 1 Ung shows the conservation of the scaffold core domain with four β-strands in the center and α-helices on both sides, with 7.6 Å rmsd over 70 Cα atoms (Figure 6C). The C-terminal region extends to constitute a unique hairpin like β-strands β5β6 (residues A248–R250, T261–T265) in each monomer that associate to form a β-sheet in the dimer (Figure 6A). The dimer association of *Bdi*Ung constituted by both the N-terminal and the C-terminal of each monomer yields 1781 Å^2 of buried interface, 13.4% of the total surface area. The N terminal is mostly involved in hydrophobic interactions with residues (M1, L2, T3, F5, E11, R15, A19, N20, Y21) whereas the C terminal interface (255–272) is stabilized by hydrogen bonds of the antiparallel β-sheet (R260, I262, I264, I266, D268) with additional charged interaction between D268 and R260 of the adjacent monomers (Supplementary Figure S8). The surface analysis of *Bdi*Ung shows a narrow groove with predominately positively charged residues (69-75, 165–168 and 190–194), a probable site for the binding of negatively charged phosphate backbone of ssDNA while the additional positively charged region (146–151) could be utilized for binding of separated strand after flipping of uracil towards the ligand binding pocket (Supplementary Figure S9).

Motif analysis

We analyzed *Bdi*Ung structure for motifs common in UDG family proteins (Figure 7A, B). The motif 'GQDPA' in *Bdi*Ung occupies a position analogous to 'GQDPY' motif in family 1 UDG with respect to the active site and hence was identified as the 'water-activating' loop or motif A. The motif of 'PPS' known to be involved in 'pinching' target

Table 2. Data collection and structure refinement statistics

Ligand	*Bdi*Ung native	Selenomethionine-labelled *Bdi*Ung	Uracil bound *Bdi*Ung	Xanthine bound *Bdi*Ung
X-ray source	7A, PAL	7A, PAL	Home source	5C, PAL
Space group	$P2_12_12_1$	$P2_12_12_1$	$C121$	$P2_12_12_1$
Unit cell dimension				
a, b, c (Å)	70.65, 90.03, 255.67	70.95, 90.06, 256.26	209.53, 89.63,143.78	70.02, 89.90, 255.34
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 96.20, 90.00	90.00, 90.00, 90.00
Resolution (Å)	1.95	2.28	2.86	2.80
R_{sym} (%)[a,b]	13.2 (56.7)	6.5 (60.8)	14.0 (33.3)	14.0 (52.8)
$I/\sigma(I)$	13.34 (1.78)	21.39 (9.41)	09.67 (3.43)	15.13 (4.59)
Completeness (%)	98.71	91.73	83.20	99.91
Redundancy	4.0 (3.2)	7.0 (6.5)	2.7 (3)	7.0 (6.3)
Refinement				
Resolution (Å)	1.95	2.28	2.86	2.80
Unique reflection	117958	70019	50734	40422
R_{work}/R_{free} (%)[c]	24.98/28.93	18.20/21.08	25.25/30.21	22.47/24.67
rmsd				
Bond lengths (Å)/angles (°)	0.009/1.30	0.010/1.22	0.009/1.00	0.015/1.00
Average B values (Å^2)	47.40	33.20	24.80	41.10
Ramachandran plot (%)				
Favoured region	96.60	96.50	95.00	96.50
Allowed region	3.40	3.50	5.00	3.50
PDB code	5GN2	5GN3	5GNW	5GRK

[a]Numbers in parentheses are statistics from the highest-resolution shell.
[b]Rsym=∑|Iobs−Iavg|/Iobs, where Iobs is the observed intensity of individual reflection and Iavg is the average over symmetry equivalents.
[c]Rwork=∑||Fo|−|Fc||/∑|Fo|, where Fo and Fc are the observed and calculated structure factor amplitudes, respectively. Rfree was calculated using 5% of the data.

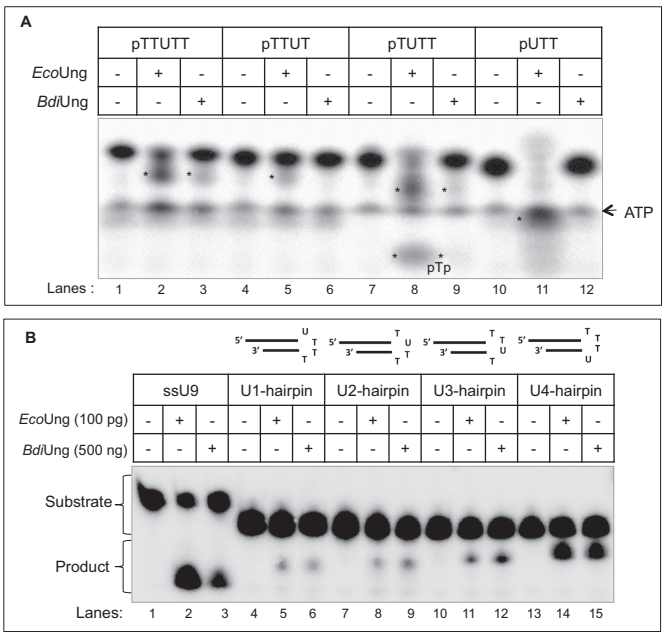

Figure 5. Excision of uracil from different structural contexts. Uracil excision assays with DNA oligomers (∼0.1 pmol each) containing a different number of nucleotides on either side of uracil (**A**) or with the hairpin DNA oligomers containing uracil in different positions in tetraloop (**B**) were carried out at 37°C as described in Materials and Methods. (A) Reactions were carried out with 1 μg of either protein (∼40 pmol *Eco*Ung or ∼32 pmol *Bdi*Ung monomer) for 30 min. More than one product band is observed in some reactions because incomplete alkaline cleavage of AP site (first reaction leaves a semialdehyde residue on 3′ end of 5′ product and the second reaction removes it, producing a 3′ phosphate on the 5′ product; products are marked with asterisks). (B) Reactions were performed for 30 min with 100 pg (4 fmol) and 500 ng (∼16 pmol monomer) of *Eco*Ung and *Bdi*Ung, respectively. Reaction products were resolved on 8 M urea PAGE (15%).

DNA is substituted by VGT residues in *Bdi*Ung and positioned in a groove that appears to bind DNA. Residues 90–94, YVMVN, constitute the fourth strand of the core β-sheet corresponding to VLLLN motif in family 1 whereas GS motif involved in 'holding' DNA substrate is substituted by GG residues in *Bdi*Ung. Interestingly, the motif B of HPSPLS residues, commonly called the Leu-intercalation loop, shows a significantly different shape and composition in *Bdi*Ung. This motif is substituted by residues 168–178 in *Bdi*Ung and forms an irregularly extended helical geometry that connects further to a long helix α6 at the C-terminal region. Due to the significantly extended protruding architecture of this region (Figure 7B(ii)), the uracil binding pocket is located distant from the tip of this region and no residues previously known for intercalation function such as leucine, arginine or tyrosine were identified in this motif. The higher B-factor of this region, 33.1 compared to 25.3, suggests the flexible characteristic of this region. This unique architecture is not observed in any other UDG family known to date and could be a signature motif in this family of *Bdi*Ung. The leucine intercalation loop in family 1 UDGs provides the major binding surface for the inhibitor protein Ugi, which enables docking to the minor-groove in DNA. The altered geometry and surface potential of this region in *Bdi*Ung also explains the absence of the binding affinity to the Ugi molecule (Supplementary Figure S10).

Active site

The uracil binding pocket in the active site is lined with residues such as Q56, P58, A59, E62, L69, G71, A73, M92 and N94 (Figure 8A, Supplementary Figure S12). Residues surrounding the O2 and O4 of uracil moieties are conserved similarly to family 1 UDG members. Essential residues for

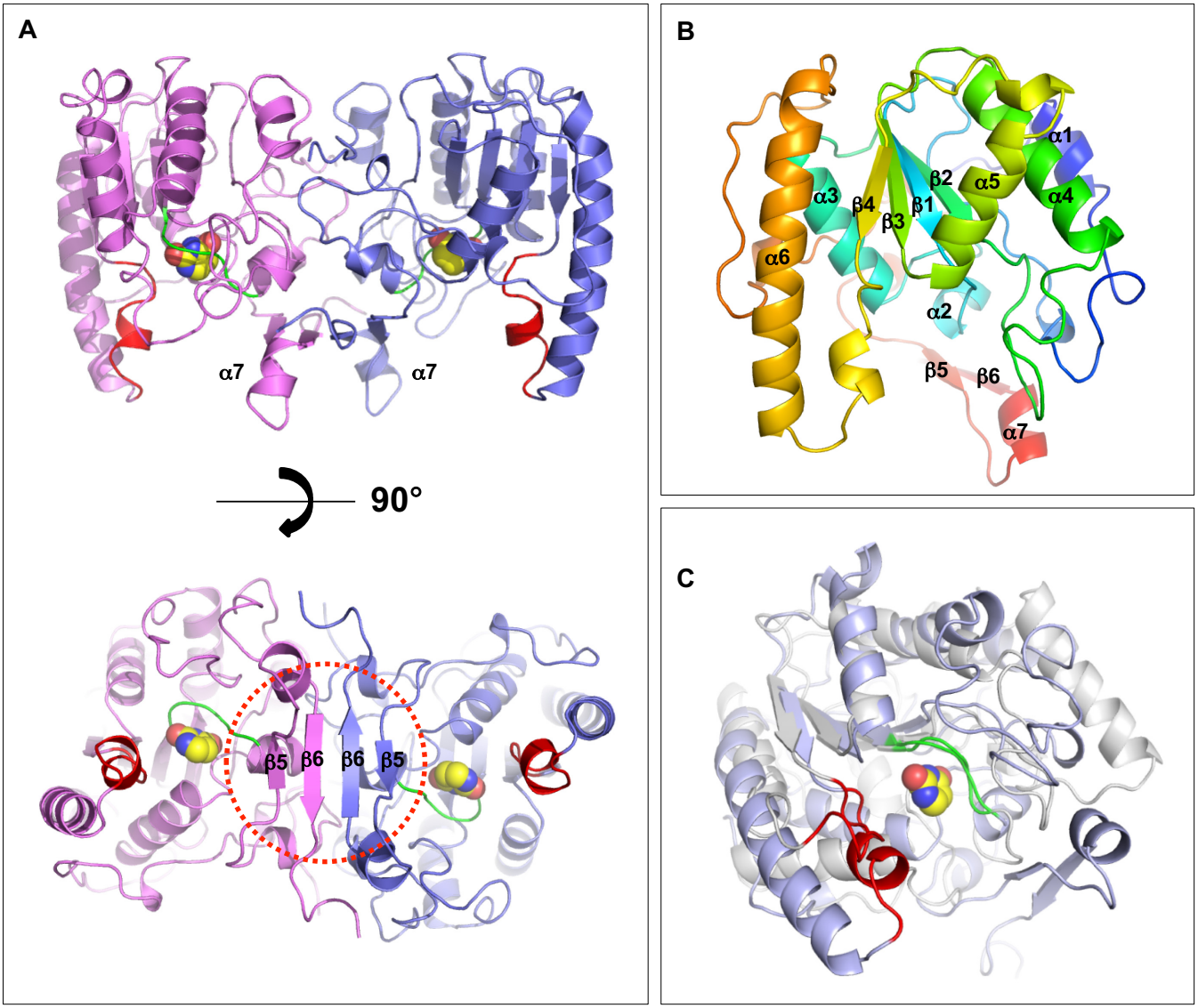

Figure 6. Overall structure of *Bdi*Ung. (**A**) A dimer of *Bdi*Ung. The molecule (top) is rotated perpendicular to highlight the formation of β-sheet in the dimeric arrangement. The corresponding regions of motifs A and B in family 1 UDG are colored in green and red, respectively, with the bound uracil molecule depicted in spheres. (**B**) A monomer of *Bdi*Ung with the typical α/β/α sandwich fold of UDG proteins. Rainbow color scheme from blue (N terminus) to red (C terminus) is shown. (**C**) Overlap of *Bdi*Ung monomer (blue) with the *Eco*Ung (PDB: IEUI) (white). Motif A (green) is well superposed whereas the corresponding region of motif B (red) is shaped significantly different in *Bdi*Ung. The bound uracil of *Bdi*Ung is depicted in spheres.

catalysis in family 1 UDGs such as H187 and D64 are also conserved in *Bdi*Ung (represented by H168 and D57), suggesting a catalytic mechanism for uracil cleavage similar to family 1 UDGs. The invariant aromatic residue that provides stable π-π stacking interaction with uracil ring in most UDGs, such as F62 in Ung, is absent and replaced by M92. Interestingly, the region surrounding uracil position 5 is differently shaped with a significantly larger cavity. The presence of A59 instead of a bulky residue (Y66 in *Eco*Ung) in motif A provides sufficient cavity space in the active site. Analysis of the uracil binding pocket between the apo and the uracil bound form reveals no significant conformational changes in the cavity region indicating that a structural change may not be required for the recognition of the uracil base by *Bdi*Ung. Given that the L191 of *Eco*Ung forms stable interactions with residues in the hydrophobic pocket of Ugi, the protruding loop of *Bdi*Ung, mainly constituted of polar residues such as E173, S174, S175, H177, D178, and S179, would disrupt any stable interaction with hydrophobic pocket of Ugi (Supplementary Figure S10). Similarly, a dimeric Ung protein from vaccinia virus, that has been reported to be resistant to Ugi inhibition, contains polar Arg at the position corresponding to L191 of *Eco*Ung affecting the stable hydrophobic interactions with Ugi (Supplementary Figure S11) (62).

Mutational analysis of *Bdi*Ung

In spite of the very low homology with family 1 UDGs, the crystal structure of UDGs revealed a conservation of most of the key residues implicated in catalysis by family 1

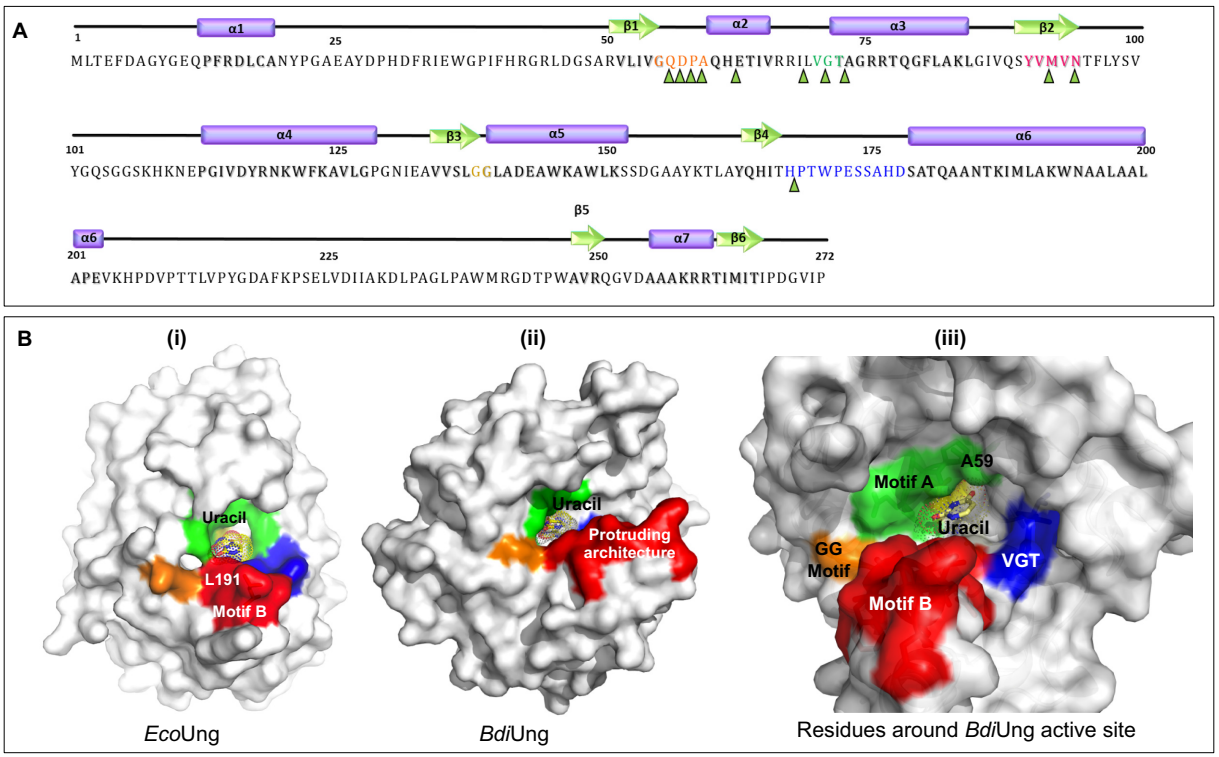

Figure 7. Primary sequence and motifs of *Bdi*Ung. (**A**) Secondary structure elements and motifs are represented along with the primary sequence. Alpha helices are represented as purple cylinders while beta strands are shown as green arrows and corresponding regions in the primary sequence are shaded. Sequences are colored as follows: motif A: orange, PPS motif: green, VLLLN motif: pink, GS motif: yellowish green, sequence forming protruding structure (motif B): blue. Residues constituting ligand binding cavity are marked by green triangles. The nomenclatures of the motifs are done with respect to family 1 UDGs. (**B**) Surface diagrams of *Eco*Ung (PDB: IEUI) and *Bdi*Ung generated using Pymol. (i) The conserved motifs in Ung are colored (motif A; green, motif PPS; blue, motif VLLLN; yellow, motif GS; orange and motif B; red). (ii) The corresponding regions of *Bdi*Ung to Ung have the same color. (iii) The residues surrounding the uracil binding pocket are shown in the surface diagram.

UDGs (Figure 8A ii). The presence of Y66 in *Eco*Ung allows uracil (with no bulky substitutions at position 5 of the pyrimidine ring) but not the 5-hydroxymethyluracil (with a bulky substitution at position 5 of the pyrimidine ring), or the xanthine (a purine) in the active site pocket. Thus, to test the importance of A59 in *Bdi*Ung in conferring broad substrate specificity of excising uracil, xanthine, and 5-hydroxymethyluracil, we generated the A59Y mutant and purified the mutant protein from the TG1*Δung::cm* strain (Supplementary Figure S4B). Interestingly, the A59Y mutation led to a decrease in its activity on xanthine, and 5-hydroxymethyluracil containing ssDNAs but not on uracil containing ssDNA (Figure 9, compare lane 9 with 8; 6 with 5; and 3 with 2).

*Bdi*Ung does not rescue *ung* deficiency in *E. coli*

A LacZ reversion assay can be used to detect UDG deficiency using an *E. coli* CC102 Lac$^-$ background. This strain possesses a mutation at the 461st codon in its *lacZ* gene converting the codon GAG to GGG (E461G) which is crucial for LacZ activity. The absence of UDG will enable C to T mutations to occur, enabling a reversion of the central G in codon 461 to A. As expected, a deficiency of *Eco*Ung (in CC102Δ*ung::kan* strain) resulted in an increased Lac$^-$ to Lac$^+$ phenotype and introduction of *Eco*Ung (pTrc*Eco*Ung) decreased the reversion frequency efficiently (Figure 10). However, under the same conditions, *Bdi*Ung did not significantly change Lac$^-$ to Lac$^+$ reversion. This observation suggests that *Bdi*Ung does not substitute for the deficiency of *Eco*Ung in *E. coli* in reducing C to T mutations.

DISCUSSION

Uracil DNA glycosylases are essential to prevent incorporation of uracil in DNA. The high G+C content in the root nodule bacterium *B. diazoefficiens* could predispose this organism to the accumulation of uracils in DNA, especially in cells exposed to reactive oxygen and nitrogen species collectively referred to as ROS/RNS. Both ROS (31–41) and RNS species (such as NO and $ONOO^-$) (42,43) are produced in nitrogen fixing nodule environments. *B. diazoefficiens* contains five different UDGs (four belonging to families 2, 4 and 5) with a distinct absence of the classical family 1 UDG, which were thought to be ubiquitously present in all organisms since they provide one of the most efficient DNA repair mechanisms.

Interestingly, in the case of *Drosophila* also, while the earlier studies have shown the presence of UDG activity (63), the genome sequencing has now revealed lack of the classical family 1 UDG (UNG) protein (64). Thus, the presence of weak UDG activity in these organisms may be attributed to the other members of the UDG family (64). In fact, re-

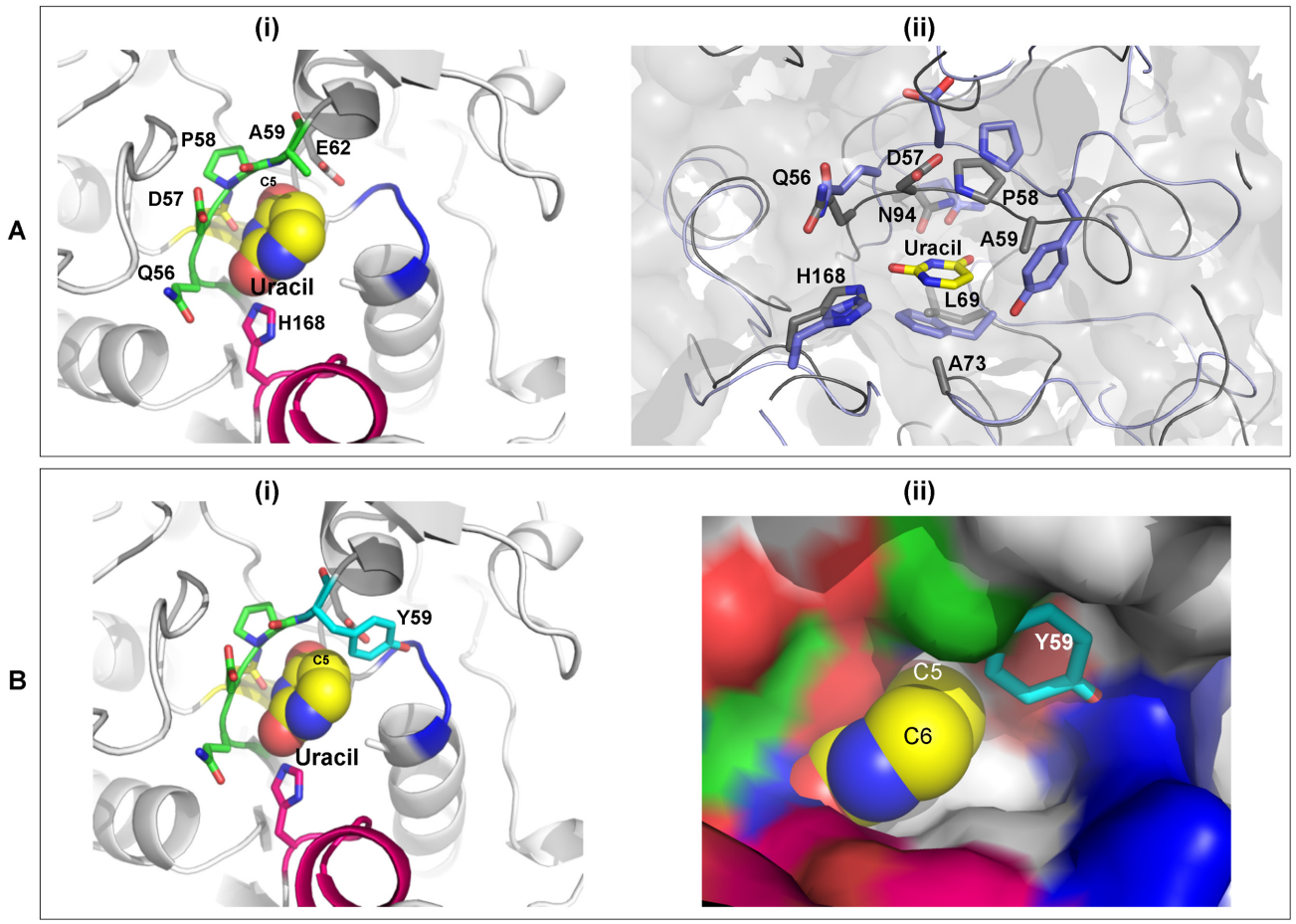

Figure 8. Active site of *Bdi*Ung. (**A**) (i) The active site of *Bdi*Ung is shown along with uracil depicted as spheres. Motif A and H168 are highlighted. (ii) The ligand binding cavity of *Bdi*Ung and *E. coli* Ung (PDB Id: 1EUI) show high similarity. Overlap of *Bdi*Ung (grey) and *E. coli* Ung (blue) active site. The residues are labelled for *Bdi*Ung. (**B**) Model structures for mutation A59Y (i) *Bdi*Ung active site showing the A59Y mutation. (ii) Surface diagram of *Bdi*Ung showing the steric hindrance to substituents at C5 of uracil molecule caused by the A59Y mutation.

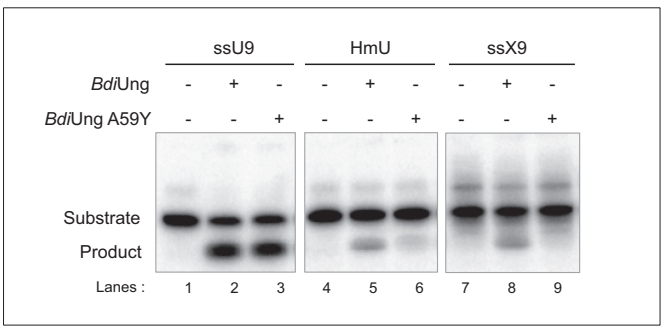

Figure 9. Activities of *Bdi*Ung A59Y mutant. The excision assays were carried out with oligomers (~0.1 pmol each) containing different lesions (uracil in ssU9, xanthine in ssX9 and 5-hydroxymethyluracil in HmU) using *Bdi*Ung and *Bdi*Ung A59Y mutant (each 1 μg, ~32 pmol monomer) at 37°C for 30 min as described in Materials and Methods. Products were resolved on 8 M urea PAGE (15%).

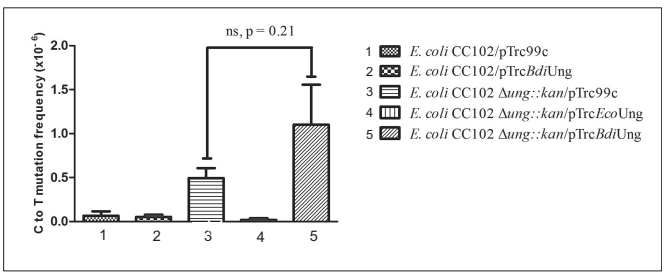

Figure 10. Mutation frequency (Lac$^+$ reversion) of *E. coli* CC102 strain and its Δ*ung*::*kan* derivatives harboring plasmid pTrc99c or its derivatives containing *Eco*Ung or *Bdi*Ung. Mutation frequency values are represented as mean ± SD from 10 independent colonies (replicates), calculated as the ratio of a number of colonies that appeared on minimal lactose plates compared to minimal glucose plates (per 1 ml culture). In Δ*ung*::*kan* strain, provision of *Bdi*Ung did not significantly change the C to T mutation frequency (comparing 3 and 5, P value = 0.21).

cent findings have shown that the lack of the highly efficient family 1 UDG (UNG) in Drosophila and the Holometabola in general, could be for an important physiological purpose (65). Developmental patterns in these organisms show strong correlation between the presence of uracil in their genomes and the lowered level of expression of dUTPase, an enzyme that hydrolyses dUTP to dUMP to avoid its incorporation in DNA. Thus, the lack of family 1 UDG (UNG)

activity in these organisms is physiologically relevant in allowing occurrence of uracil in DNA (66). Whether the lack of highly efficient family 1 UDG in *B. diazoefficiens* contributes to its specialized physiology during nodulation or otherwise is not known.

To study the role of UDGs in *B. diazoefficiens*, we have identified and characterized a novel uracil DNA glycosylase from *B. diazoefficiens* (Q89XRO_BRAJA, referred to as *Bdi*Ung, encoded by *blr0248*) containing a motif (GQDPA) similar to the active site motif A (GQDPY) of family 1 UDGs. *Bdi*Ung was found to be active on both ssDNA and dsDNA containing uracil with pNUNN as the minimum substrate required for detectable activity. *Bdi*Ung did not interact with Ugi and was therefore resistant to Ugi mediated inhibition in uracil excision. Further, unlike family 1 UDGs, the *Bdi*Ung showed broader substrate specificity and excised 5-hydroxymethyl uracil and xanthine present in ssDNA. The co-crystal structures of *Bdi*Ung with uracil, and xanthine provide a rationale for the relaxed substrate specificity (Supplementary Figure S12). While *Bdi*Ung was inhibited by free uracil, AP DNA (ssF9) did not inhibit its activity to any significant level even at 200 μM concentration (Figure 3).

Substrate specificity of *Eco*Ung is due to the well-tailored active site pocket made to specifically fit uracil (13). The efficient recognition of uracil is proposed to occur through a 'pinch, push, plug, and pull' mechanism (68–70). Initial backbone phosphate contacts pinch the DNA backbone. Then a conserved Leu (L191 in *Eco*Ung) acts as a wedge/plug to retain the flipped out uracil into the active site of the enzyme (67,68). Many specific hydrogen bonds and stacking interactions stabilize the uracil in extra-helical conformation within the active site (pull) (70). The side chain of Tyr in the GQDPY motif is in van der Waals' interaction with C5 of uracil, thus blocking thymine (with a methyl group on C5) and bulky purine rings (11,13,71).

While *Bdi*Ung exhibited no significant sequence similarity with any of the established families of UDGs, the crystal structure of the protein revealed that the overall fold is similar to family 4 UDGs in the absence of the characteristic 4Fe–S cluster and that most of the key amino-acid residues implicated in *Eco*Ung catalysis of family 1 UDG are present in the active site. Structural alignment of *Bdi*Ung with *Eco*Ung showed that the GQDPA motif of *Bdi*Ung was located in a position analogous to motif A of *Eco*Ung which is involved in activation of a water molecule that performs a nucleophilic attack on C1' of ribose during catalysis. *Bdi*Ung also contains Asn and His residues similar to those in *Eco*Ung, which form specific contacts with a uracil residue in the active site. In spite of the drastically different architecture of the active site periphery compared to *Eco*Ung, the similarity observed with respect to the key active site residues implies *Bdi*Ung catalysis to be similar to that of *Eco*Ung. But the substitution of the classical Leu intercalation loop by the extended protruding architecture with no apparent functional residue similar to L191 of *Eco*Ung (implicated in pushing the uracil/plugging the duplex void) could be the reason for its reduced activity on ds-DNA as well as the absence of its complex formation with Ugi (68,69).

A conserved Tyr in the motif A of Ung proteins prevents entry of bulky bases and position 5- substituted uracil into the active site of the family 1, Ung proteins. In *Bdi*Ung the Tyr residue of motif A is substituted by Ala explaining its broad substrate specificity. Consistent with this observation, *Bdi*Ung A59Y mutant revealed a much decreased activity on 5-hydroxymethyl uracil and xanthine containing ssDNAs (Figure 9). Still, the active site seems carefully designed to exclude any of the normal bases. Most interestingly, while all the UDGs studied so far show a conserved aromatic residue (Phe) in the active site that provides stability to the flipped out uracil in the active site, no such supporting aromatic residue was seen in the active site of *Bdi*Ung.

As revealed by the crystal structure and biochemical analyses, *Bdi*Ung is a dimeric protein. UNG from vaccinia virus was also characterized as a dimer (72). However, the dimeric interfaces of the two enzymes are significantly different, in that the *Bdi*Ung dimerizes primarily by the C-terminal extended hairpin strand with each active site of the monomer, facing the same side of the dimer, whereas the vaccinia virus UNG dimerizes through the main α/β/α sandwich fold with two active sites positioned opposite (Supplementary Figure S13). The functional significance of the dimerization is not clear at this stage and may need further investigation.

From the analysis of the impact of *Bdi*Ung in the mutation frequency, using *ung*$^-$ strains of *E. coli* CC102 (monitoring Lac$^+$ reversion frequency), we found that *Bdi*Ung cannot carry out a physiological function in rescuing C to T mutations similar to that of *Eco*Ung in *E. coli*. We attribute this to the poor efficiency of uracil excision from dsDNA and also to the fact that these studies were carried out in a heterologous system with a relatively much higher growth rate. Besides, the much higher efficiency with which *Bdi*Ung acts on ssDNA compared to dsDNA and the higher relative efficiencies of uracil excision from loop regions of DNA tells us that it is a single strand specific UDG. This could be further supported by the surface analysis of *Bdi*Ung. The unique protruding loop yields a narrow DNA binding groove at the uracil binding pocket, a suitable architecture for ssDNA substrate interaction (Supplementary Figure S9).

Interestingly, most organisms with homologues of *Bdi*Ung possessed genomes with high G+C contents (Supplementary Table S3). Organisms having G+C rich genomes are intrinsically prone to accumulating hairpin-loops in their genome. The relatively less compromising efficiency of *Bdi*Ung on uracil found in loop regions of DNA (as compared to ssU9), along with the presence of its homologues in high G+C containing organisms might be pointing to a physiological function of this family of UDGs.

In addition to *Bdi*Ung, we have identified and characterized three other UDGs from *B. diazoefficiens*. Besides a family 2 homologue (Blr6661) and a family 5 homologue (Blr5068), the organism also contains a homologue of the UdgX protein, which we recently identified as a novel uracil–DNA binding protein in *M. smegmatis* (56). Blr5068 was found to efficiently excise uracil from ssDNA and ds-DNA. Further, *bll2523* was identified as a gene coding for a putative family 4 homologue. The unusually high number of

UDGs and the presence of novel UDGs like *Bdi*Ung might compensate for the lack of a family 1 UDG homologue in this slow growing organism.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We acknowledge the DBT-IISc partnership programme, University Grants Commission, New Delhi for the Centre of Advanced Studies, the DST-FIST level II infrastructure supports and the KRIBB Research Initiative Program to carry out this work. UV is a J.C. Bose Fellow of the Department of Science and Technology (DST), New Delhi. S. Sah is a senior research associate supported by the Council of Scientific and Industrial Research, New Delhi.

FUNDING

Department of Biotechnology (DBT), New Delhi, and facilitated by the Distinguished Collaborator Award of the Murdoch University, Australia (to U.V. and R.P.T.); Murdoch University small grant scheme; National Research Fund [NRF-2015R1A2A2A03006970 to E.J.W.], Korea. Funding for open access charge: Research grants (to U.V.).
Conflict of interest statement. None declared.

REFERENCES

1. Lindahl,T. (1982) DNA repair enzymes. *Annu. Rev. Biochem.*, **51**, 61–87.
2. Friedberg,E.C., Walker,G.C. and Siede,W. (1995) *DNA Repair and Mutagenesis*. ASM Press.
3. Jancso,A., Botfield,M.C., Sowers,L.C. and Weiss,M.A. (1994) An altered-specificity mutation in a human POU domain demonstrates functional analogy between the POU-specific subdomain and phage lambda repressor. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 3887–3891.
4. Rogstad,D.K., Liu,P., Burdzy,A., Lin,S.S. and Sowers,L.C. (2002) Endogenous DNA lesions can inhibit the binding of the AP-1 (c-Jun) transcription factor. *Biochemistry*, **41**, 8093–8102.
5. Lindahl,T. (1974) An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 3649–3653.
6. Krokan,H.E., Drablos,F. and Slupphaug,G. (2002) Uracil in DNA–occurrence, consequences and repair. *Oncogene*, **21**, 8935–8948.
7. Pearl,L.H. (2000) Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res.*, **460**, 165–181.
8. Aravind,L. and Koonin,E.V. (2000) The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates. *Genome Biol.*, **1**, RESEARCH0007.
9. Lucas-Lledo,J.I., Maddamsetti,R. and Lynch,M. (2011) Phylogenomic analysis of the uracil-DNA glycosylase superfamily. *Mol. Biol. Evol.*, **28**, 1307–1317.
10. Lindahl,T., Ljungquist,S., Siegert,W., Nyberg,B. and Sperens,B. (1977) DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli. *J. Biol. Chem.*, **252**, 3286–3294.
11. 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.
12. Handa,P., Acharya,N. and Varshney,U. (2002) Effects of mutations at tyrosine 66 and asparagine 123 in the active site pocket of Escherichia coli uracil DNA glycosylase on uracil excision from synthetic DNA oligomers: evidence for the occurrence of long-range interactions between the enzyme and substrate. *Nucleic Acids Res.*, **30**, 3086–3095.
13. Mol,C.D., Arvai,A.S., Slupphaug,G., Kavli,B., Alseth,I., Krokan,H.E. and Tainer,J.A. (1995) Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell*, **80**, 869–878.
14. Parikh,S.S., Putnam,C.D. and Tainer,J.A. (2000) Lessons learned from structural results on uracil-DNA glycosylase. *Mutat. Res.*, **460**, 183–199.
15. Cone,R., Bonura,T. and Friedberg,E.C. (1980) Inhibitor of uracil-DNA glycosylase induced by bacteriophage PBS2. Purification and preliminary characterization. *J. Biol. Chem.*, **255**, 10354–10358.
16. Ravishankar,R., Bidya Sagar,M., Roy,S., Purnapatre,K., Handa,P., Varshney,U. and Vijayan,M. (1998) X-ray analysis of a complex of Escherichia coli uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG. *Nucleic Acids Res.*, **26**, 4880–4887.
17. 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.
18. Neddermann,P. and Jiricny,J. (1993) The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *J. Biol. Chem.*, **268**, 21218–21224.
19. Haushalter,K.A., Todd Stukenberg,M.W., Kirschner,M.W. and Verdine,G.L. (1999) Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Curr. Biol.: CB*, **9**, 174–185.
20. Sandigursky,M. and Franklin,W.A. (1999) Thermostable uracil-DNA glycosylase from Thermotoga maritima a member of a novel class of DNA repair enzymes. *Curr. Biol.: CB*, **9**, 531–534.
21. 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.
22. Gallinari,P. and Jiricny,J. (1996) A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature*, **383**, 735–738.
23. Neddermann,P. and Jiricny,J. (1994) Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 1642–1646.
24. Moe,E., Leiros,I., Smalas,A.O. and McSweeney,S. (2006) 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.*, **281**, 569–577.
25. Wibley,J.E., Waters,T.R., Haushalter,K., Verdine,G.L. and Pearl,L.H. (2003) Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. *Mol. Cell*, **11**, 1647–1659.
26. 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.
27. Sartori,A.A., Schar,P., Fitz-Gibbon,S., Miller,J.H. and Jiricny,J. (2001) Biochemical characterization of uracil processing activities in the hyperthermophilic archaeon Pyrobaculum aerophilum. *J. Biol. Chem.*, **276**, 29979–29986.
28. Santos,R., Herouart,D., Sigaud,S., Touati,D. and Puppo,A. (2001) Oxidative burst in alfalfa-Sinorhizobium meliloti symbiotic interaction. *Mol. Plant-Microbe Interact.: MPMI*, **14**, 86–89.
29. Ramu,S.K., Peng,H.M. and Cook,D.R. (2002) Nod factor induction of reactive oxygen species production is correlated with expression of the early nodulin gene rip1 in Medicago truncatula. *Mol. Plant-Microbe Interact.: MPMI*, **15**, 522–528.
30. D'Haeze,W. and Holsters,M. (2002) Nod factor structures, responses, and perception during initiation of nodule development. *Glycobiology*, **12**, 79R–105R.
31. Becana,M. and Klucas,R.V. (1992) Transition metals in legume root nodules: iron-dependent free radical production increases during nodule senescence. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 8958–8962.
32. Becana,M. and Klucas,R.V. (1992) Oxidation and reduction of leghemoglobin in root nodules of leguminous plants. *Plant Physiol.*, **98**, 1217–1221.
33. Becana,M. and Rodríguez-Barrueco,C. (1989) Protective mechanisms of nitrogenase against oxygen excess and partially-reduced oxygen intermediates. *Physiol. Plantarum*, **75**, 429–438.
34. Becana,M., Dalton,D.A., Moran,J.F., Iturbe-Ormaetxe,I., Matamoros,M.A. and Rubio,M.C. (2000) Reactive oxygen species

and antioxidants in legume nodules. *Physiol. Plantarum*, **109**, 372–381.
35. Puppo,A. and Halliwell,B. (1988) Generation of hydroxyl radicals by soybean nodule leghaemoglobin. *Planta*, **173**, 405–410.
36. Becana,M., Moran,J.F. and Iturbe-Ormaetxe,I. (1998) Iron-dependent oxygen free radical generation in plants subjected to environmental stress: toxicity and antioxidant protection. *Plant Soil*, **201**, 137–147.
37. Fridovich,I. (1976) Superoxide dismutases: studies of structure and mechanism. *Adv. Exp. Med. Biol.*, **74**, 530–539.
38. Moreau,S., Davies,M.J., Mathieu,C., Herouart,D. and Puppo,A. (1996) Leghemoglobin-derived radicals. Evidence for multiple protein-derived radicals and the initiation of peribacteroid membrane damage. *J. Biol. Chem.*, **271**, 32557–32562.
39. Puppo,A. and Halliwell,B. (1988) Formation of hydroxyl radicals from hydrogen peroxide in the presence of iron. Is haemoglobin a biological Fenton reagent? *Biochem. J.*, **249**, 185–190.
40. Puppo,A., Rigaud,J. and Job,D. (1981) Role of superoxide anion in leghemoglobin autoxidation. *Plant Sci. Lett.*, **22**, 353–360.
41. Schneider,K. and Schlegel,H.G. (1981) Production of superoxide radicals by soluble hydrogenase from Alcaligenes eutrophus H16. *Biochem. J.*, **193**, 99–107.
42. Becana,M., Matamoros,M.A., Udvardi,M. and Dalton,D.A. (2010) Recent insights into antioxidant defenses of legume root nodules. *New Phytologist*, **188**, 960–976.
43. Baudouin,E., Pieuchot,L., Engler,G., Pauly,N. and Puppo,A. (2006) Nitric Oxide Is Formed in Medicago truncatula-Sinorhizobium meliloti Functional Nodules. *Mol. Plant-Microbe Interact.*, **19**, 970–975.
44. Halliwell,B. and Gutteridge,J.M.C. (1999) *Free Radicals in Biology and Medicine*. Oxford University Press.
45. Kreutzer,D.A. and Essigmann,J.M. (1998) Oxidized, deaminated cytosines are a source of C –>T transitions in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 3578–3582.
46. Delamuta,J.R., Ribeiro,R.A., Ormeno-Orrillo,E., Melo,I.S., Martinez-Romero,E. and Hungria,M. (2013) Polyphasic evidence supporting the reclassification of Bradyrhizobium japonicum group Ia strains as Bradyrhizobium diazoefficiens sp. nov. *Int. J. Syst. Evol. Microbiol.*, **63**, 3342–3351.
47. Mathis,J.N., McMillin,D.E., Champion,R.A. and Hunt,P.G. (1997) Genetic variation in two cultures of Bradyrhizobium japonicum 110 differing in their ability to impart drought tolerance to soybean. *Curr. Microbiol.*, **35**, 363–366.
48. Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T., Sasamoto,S., Watanabe,A., Idesawa,K., Iriguchi,M., Kawashima,K. *et al.* (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110 (supplement). *DNA Res.*, **9**, 225–256.
49. Cupples,C.G. and Miller,J.H. (1989) A set of lacZ mutations in Escherichia coli that allow rapid detection of each of the six base substitutions. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 5345–5349.
50. Sedmak,J.J. and Grossberg,S.E. (1977) A rapid, sensitive, and versatile assay for protein using Coomassie brilliant blue G250. *Anal. Biochem.*, **79**, 544–552.
51. Varshney,U. and van de Sande,J.H. (1991) Specificities and kinetics of uracil excision from uracil-containing DNA oligomers by Escherichia coli uracil DNA glycosylase. *Biochemistry*, **30**, 4055–4061.
52. Kumar,N.V. and Varshney,U. (1997) Contrasting effects of single stranded DNA binding protein on the activity of uracil DNA glycosylase from Escherichia coli towards different DNA substrates. *Nucleic Acids Res.*, **25**, 2336–2343.
53. Otwinowski,Z. and Minor,W. (1997) *Methods in Enzymology*. Academic Press, Vol. **276**, pp. 307–326.
54. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*, **60**, 2126–2132.
55. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.-W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D*, **66**, 213–221.
56. Sang,P.B., Srinath,T., Patil,A.G., Woo,E.J. and Varshney,U. (2015) A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily. *Nucleic Acids Res.*, **43**, 8452–8463.
57. Roy,S., Purnapatre,K., Handa,P., Boyanapalli,M. and Varshney,U. (1998) Use of a coupled transcriptional system for consistent overexpression and purification of UDG-Ugi complex and Ugi from Escherichia coli. *Protein Express. Purif.*, **13**, 155–162.
58. Ghosh,M., Kumar,N.V., Varshney,U. and Chary,K.V. (1999) Structural characterisation of a uracil containing hairpin DNA by NMR and molecular dynamics. *Nucleic Acids Res.*, **27**, 3938–3944.
59. Kumar,N.V. and Varshney,U. (1994) Inefficient excision of uracil from loop regions of DNA oligomers by E. coli uracil DNA glycosylase. *Nucleic Acids Res.*, **22**, 3737–3741.
60. Purnapatre,K. and Varshney,U. (1998) Uracil DNA glycosylase from Mycobacterium smegmatis and its distinct biochemical properties. *Eur. J. Biochem./FEBS*, **256**, 580–588.
61. Holm,L. and Rosenström,P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**, W545–W549.
62. Schormann,N., Grigorian,A., Samal,A., Krishnan,R., DeLucas,L. and Chattopadhyay,D. (2007) Crystal structure of vaccinia virus uracil-DNA glycosylase reveals dimeric assembly. *BMC Struct. Biol.*, **7**, 45.
63. Morgan,A.R. and Chlebek,J. (1989) Uracil-DNA glycosylase in insects. Drosophila and the locust. *J. Biol. Chem.*, **264**, 9911–9914.
64. Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F. *et al.* (2000) The genome sequence of Drosophila melanogaster. *Science*, **287**, 2185–2195.
65. Muha,V., Horvath,A., Bekesi,A., Pukancsik,M., Hodoscsek,B., Merenyi,G., Rona,G., Batki,J., Kiss,I., Jankovics,F. *et al.* (2012) Uracil-containing DNA in Drosophila: stability, stage-specific accumulation, and developmental involvement. *PLoS Genet.*, **8**, e1002738.
66. Horvath,A., Bekesi,A., Muha,V., Erdelyi,M. and Vertessy,B.G. (2013) Expanding the DNA alphabet in the fruit fly: uracil enrichment in genomic DNA. *Fly*, **7**, 23–27.
67. Handa,P., Roy,S. and Varshney,U. (2001) The role of leucine 191 of Escherichia coli uracil DNA glycosylase in the formation of a highly stable complex with the substrate mimic, ugi, and in uracil excision from the synthetic substrates. *J. Biol. Chem.*, **276**, 17324–17331.
68. Jiang,Y.L., Kwon,K. and Stivers,J.T. (2001) Turning on uracil-DNA glycosylase using a pyrene nucleotide switch. *J. Biol. Chem.*, **276**, 42347–42354.
69. Parikh,S.S., Mol,C.D., Slupphaug,G., Bharati,S., Krokan,H.E. and Tainer,J.A. (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.*, **17**, 5214–5226.
70. 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.
71. Kavli,B., Slupphaug,G., Mol,C.D., Arvai,A.S., Peterson,S.B., Tainer,J.A. and Krokan,H.E. (1996) Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *EMBO J.*, **15**, 3442–3447.
72. Burmeister,W.P., Tarbouriech,N., Fender,P., Contesto-Richefeu,C., Peyrefitte,C.N. and Iseni,F. (2015) Crystal structure of the vaccinia virus uracil-DNA glycosylase in complex with DNA. *J. Biol. Chem.*, **290**, 17923–17934.