8452–8463 *Nucleic Acids Research, 2015, Vol. 43, No. 17* *Published online 24 August 2015*
doi: 10.1093/nar/gkv854

A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily

Pau Biak Sang[1], Thiruneelakantan Srinath[1], Aravind Goud Patil[1], Eui-Jeon Woo[2] and Umesh Varshney[1,3,*]

[1]Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore, 560012, India, [2]Functional Genomic Research Center, Korea Research Institute of Bioscience and Biotechnology, 125 Gwahakro, Yuseongu, Daejeon, South Korea and [3]Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, 560064, India

Received July 01, 2015; Revised August 10, 2015; Accepted August 11, 2015

ABSTRACT

Uracil DNA glycosylases (UDGs) are an important group of DNA repair enzymes, which pioneer the base excision repair pathway by recognizing and excising uracil from DNA. Based on two short conserved sequences (motifs A and B), UDGs have been classified into six families. Here we report a novel UDG, UdgX, from *Mycobacterium smegmatis* and other organisms. UdgX specifically recognizes uracil in DNA, forms a tight complex stable to sodium dodecyl sulphate, 2-mercaptoethanol, urea and heat treatment, and shows no detectable uracil excision. UdgX shares highest homology to family 4 UDGs possessing Fe-S cluster. UdgX possesses a conserved sequence, KRRIH, which forms a flexible loop playing an important role in its activity. Mutations of H in the KRRIH sequence to S, G, A or Q lead to gain of uracil excision activity in *Msm*UdgX, establishing it as a novel member of the UDG superfamily. Our observations suggest that UdgX marks the uracil-DNA for its repair by a RecA dependent process. Finally, we observed that the tight binding activity of UdgX is useful in detecting uracils in the genomes.

INTRODUCTION

Uracil DNA glycosylases (UDGs) recognize uracil, inadvertently present in DNA and initiate uracil excision repair pathway (1,2) by cleaving the N-glycosidic bond between the uracil and the deoxyribose sugar, releasing uracil and leaving behind an abasic site (AP-site) (3,4). The AP-site is then processed and restored to a canonical base by the subsequent actions of AP-endonuclease, dRPase, DNA polymerase and DNA ligase enzymes (5–7). Uracil DNA glycosylase from *E. coli* (*Eco*Ung), which is ubiquitously present in all organisms studied so far including many viruses, was the first DNA glycosylase to be discovered (1,8). Since then several novel UDGs have been discovered (9–13) leading to establishment of UDG superfamily (4). The last UDG discovered belonged to the 6th family, which represents hypoxanthine DNA glycosylase lacking uracil excision activity (9). Family 1 UDGs (Ung/UNG) are the most conserved and most extensively studied proteins of the UDG superfamily (8,14–17). Ung proteins, which are amongst the most efficient enzymes are characterized by motif A sequence, GQDPY involved in substrate catalysis, and motif B sequence, HPSPLS involved in stabilizing the enzyme substrate complex (3,18–20). Family 1 UDGs (Ung) are specifically inhibited by *B. subtilis* phage PBS-1 or 2 encoded proteinaceous inhibitor called Ugi (uracil DNA glycosylase inhibitor) by forming a physiologically irreversible non-covalent complex in 1:1 stoichiometry (21,22). Family 2 UDGs (Mug/TDG) are mismatch specific DNA glycosylases which excise thymine from T:G pair and possess GINPG and MPSSAR as motifs A and B sequences, respectively. Mug/TDG are specific for dsDNA and excise uracil form U:A and U:G pairs less efficiently (10,23–25). Family 3 UDGs (SMUG) possess motifs A and B sequences defined by GMNPG and HPSPRN, respectively. Although initially designated as single strand selective monofunctional uracil DNA glycosylase (SMUG), they were later shown to be active on double stranded substrate (11,26). SMUG proteins are mostly present in eukaryotes and a few eubacterial species (27). Family 4 and family 5 UDGs are 4Fe-4S cluster containing proteins mostly found in thermophilic bacteria and archaea but absent in eukaryotes. The motifs A and B sequences of family 4 UDGs are GE(A/G)PG and HPAAVL, respectively (12,28), whereas these sequences for the family 5 UDGs are GLAPA and HPSPLN, respectively. Family 5 UDGs have broad substrate specificity (13,29–30). Family 6 UDGs have motifs A and B sequences of GSLPG and SSSGAN, respectively (9). Although the UDGs from different families differ in their primary amino acid sequences, they possess the same α/β structural fold and seem to have a common evolutionary origin (31,32).

*To whom correspondence should be addressed. Tel: +91 8022932686; Fax: +91 8023602697; Email: varshney@mcbl.iisc.ernet.in

Earlier investigations on the mycobacterial UDGs from our laboratory showed the presence of family 1 (Ung) and family 5 (UdgB) UDGs (30). In *M. smegmatis*, Ung deficiency results in increased C to T mutations and retarded growth under acidified sodium nitrite stress (33). *Msm*Ung and *Msm*UdgB show a synergistic effect in preventing mutations and in growth under acidified sodium nitrite and hypoxic stress (34,35). Search for other UDGs in this class of organisms has now revealed the presence of yet another UDG with motifs A and B sequences of GEQPG and HPSSLL, respectively. This UDG (designated as *Msm*UdgX) does not fall into any of the already known UDG families. Homologs of *Msm*UdgX are also present in other species of mycobacteria such as *M. avium, M. haemophilum* but absent from *M. tuberculosis*. Other organisms possessing this protein are *Rhodococcus spp*, *Streptomyces coelicolor*, *Gordonia namibiense*, *Bradyrhizobium japonicum* and *Nocardia farcidia*. Alignment of UdgX with the other UDGs shows that it is related to family 4 UDGs. However, there is an extra stretch of amino acids, AAGGKRRIH, which is unique to this group of proteins. Within this sequence, the region KRRIH is conserved amongst all UdgX. We show that unlike other UDGs, UdgX has an unusual property of forming an extremely tight complex with uracil-DNA, which is further processed through a RecA dependent process.

MATERIALS AND METHODS

Bacterial strains, plasmids, media and growth conditions

Bacterial strains (36–38) and plasmids used are listed in Table 1. *E. coli* strains were grown in Luria-Bertani broth (LB) or LB containing 1.5% (w/v) agar (Difco, USA). Media were supplemented with ampicillin (Amp), kanamycin (Kan) and hygromycin (Hyg) as needed at 100 μg ml^{-1}, 25 μg ml^{-1} and 150 μg ml^{-1}, respectively, for *E. coli*. For *M. smegmatis* growth, Kan, Hyg and gentamycin (Gm) were supplemented at 50 μg ml^{-1}, 50 μg ml^{-1} and 5 μg ml^{-1}, respectively, when required. *M. avium* and *Rhodococcus imtechensis* were procured from IMTECH, Chandigarh, India. Knockout strains of *E. coli* were procured from the Coli Genetic Stock Center (CGSC).

Cloning of *Msm*UdgX and its purification

The open reading frame (ORF) of *Msm*UdgX (MSMEG_0265) was amplified by PCR using a forward primer, *Msm*UdgX Fp (5′ TGCATATGGCGGGTGCGCAAGAT 3′) and a reverse primer, *Msm*UdgX Rp (5′ CAAGCTTGCAGATGGGCTCCATC 3′) containing NdeI and HindIII sites (underlined), respectively. PCR (50 μl) consisted of ∼300 ng of *M. smegmatis* genomic DNA, 200 μM dNTPs, 20 pmol each of *Msm*UdgX Fp and *Msm*UdgX Rp primers and 1 U of DyNAzyme EXT DNA polymerase (Finnzyme, Finland). PCR conditions included initial denaturation at 94°C for 4 min followed by 29 cycles of incubations at 94°C for 1 min, 60°C for 30 s and 72°C for 1 min. The PCR product (662 bp) was blunt end ligated into pJET1.2 (MBI) vector to generate pJET*Msm*UdgX and confirmed by restriction digestion and DNA sequence analysis (Macrogen, S. Korea). The pJET*Msm*UdgX was digested with NdeI and HindIII. The released *Msm*UdgX ORF was cloned into similarly digested pET14b to generate pET14b*Msm*UdgX. To purify *Msm*UdgX, pET14b*Msm*UdgX which contains 6× His tag at the N-terminal was introduced into *E. coli* BL21 (DE3) or *E. coli* Rosetta (DE3) by transformation. Isolated colonies were inoculated into 50 ml LB with Amp and grown until saturation (or overnight). Inoculum (1%) was added into 3 L LB medium containing Amp and 0.01% $FeCl_3$, grown to OD_{600} of 0.6 at 37°C under shaking, supplemented with 0.5 mM IPTG and allowed to grow further for 2 h. Cells were harvested by centrifugation, suspended in buffer A [20 mM Tris-HCl (pH 8), 500 mM NaCl, 10% glycerol (v/v), 2 mM β-mercaptoethanol and 20 mM imidazole], lysed by sonication and centrifuged at 24 000 rpm (SW28 Ti, Beckman coulter) 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 20 ml of buffer A and eluted with a gradient of imidazole (20–1000 mM) in the same buffer. The fractions were analyzed on 15% SDS-PAGE. Fractions enriched for UdgX were pooled, loaded onto Superdex-G75 gel filtration column and eluted in buffer B [20 mM Tris-HCl (pH 8), 400 mM NaCl, 10% glycerol (v/v) and 2 mM β-mercaptoethanol]. The purity of UdgX was checked on 15% SDS-PAGE. Fractions containing pure UdgX were pooled, concentrated using a 10 kDa cutoff Centricon (Millipore) and estimated by Bradford's method using bovine serum albumin (BSA) as standard (39). The proteins were dialyzed against buffer A containing 50% glycerol and stored in -20°C.

Radiolabeling of substrates

DNA oligomers (10 pmol) were 5′ ^{32}P-end labeled using 10 μCi of [γ-^{32}P] ATP (6000 Ci/mmol) and T4 polynucleotide kinase and purified on Sephadex G-50 minicolumns (30). SSU9 which has U residue at the 9th position from the labeled end was used as ssDNA substrate. SSU9 was annealed with complementary oligomer with G residue opposite U to generate dsDNA substrate, SSU9:G.

Activity assays of *Msm*UdgX

Assays for *Msm*UdgX with SSU9 or SSU9:G were carried out in 10 μl reactions in UDG buffer (50 mM Tris-HCl (pH 8), 1 mM Na_2EDTA, 1 mM DTT, 25 μg/ml BSA) for 20 min at 37°C. The reactions were stopped by addition of 5 μl 0.2 N NaOH and heating at 90°C for 10 min, mixed with 5 μl formamide dye (80% formamide, 0.05% each of bromophenol blue and xylene cyanol FF, 10 mM NaOH, 2 mM Na_2EDTA), boiled for 5 min and 15 μl aliquots were analyzed on 15% polyacrylamide (19:1) 8 M urea gels. The gels were exposed to phosphor imager cassette to acquire the image. In another experiment, *Msm*UdgX (∼1 μg) was used with varying size of unlabeled uracil containing DNA oligomers (∼50 pmol), mixed with SDS loading dye, heated at 90°C for 10 min and resolved on 15% SDS-PAGE.

Table 1. List of Strains/Plasmids/Oligomers

Strain/plasmid	Details	Reference
E. coli BL21 (DE3)	F$^-$ *ompT hsdS$_B$(r_B^- m_B^-) gal dcm* (DE3)	Novagen
E. coli Rosetta (DE3)	F$^-$ *ompT hsdS$_B$(r_B^- m_B^-) gal dcm* (DE3) pRARE	Novagen
M. smegmatis mc^2 155	A high efficiency transformation strain of *M. smegmatis*	(36)
pET14b*Msm*UdgX	pET14b plasmid with *Msm*UdgX cloned in its NdeI/HindIII sites	This study
pET14b*Mav*UdgX	pET14b plasmid with *Mav*UdgX cloned in its NdeI/HindIII sites	This study
pET14b*Rim*UdgX	pET14b plasmid with *Rim*UdgX cloned in its NdeI/HindIII sites	This study
pMVUdgX	pMV261(hyg^R) plasmid with UdgX ORF cloned in its BamHI/HindIII site	This study
pMVUdgX350	pMV261(hyg^R) plasmid with UdgX ORF with 350 upstream region taking the native promoter along with it and cloned in XbaI/PstI replacing the hsp60 promoter present in pMV261	This study
SSU9	Substrate for UDG assay having uracil in the 9th position 5′CTCAAGTGUAGGCATGCAAGAGCT3′	(30)
SSU9:G	Oligomer complimentary to SSU9 with G opposite to uracil 5′CTTGCATGCCTGCACTTGAGTGCA 3′	(30)
DHU	5′ GGCTGCTAC(DHU)AGGCGAAGTG 3′	(30)
Hx	5′ GGCTGCTAC(Hx)AGGCGAAGTG 3′	(30)
HmU	5′CTCAAGTG(HmU)AGGCATGCAAGAGCT3′	This study
M. smegmatis ΔrecA	*M. smegmatis* mc2 155 strain where *recA* gene is disrupted with kan^R cassette	This study
E. coli MG1655	An *E. coli* K strain, F$^-$ λ$^-$*rph-1*	(37)
E. coli (WT) (BW22113)	F-, *Δ(araD-araB)567*, *ΔlacZ4787*(::rrnB-3), λ$^-$, *rph-1*, *Δ(rhaD-rhaB)568*, *hsdR514*	(38)
E. coli ΔrecA (BW26355)	BW22113 but *ΔrecA635::kan*	(38)
E. coli ΔuvrB (JW0762–2)	BW22113 but *ΔuvrB751::kan*	(38)
E. coli ΔrecB (JW2788–1)	BW22113 but *ΔrecB745::kan*	(38)
E. coli ΔruvA (JW1850–2)	BW22113 but *ΔruvA786::kan*	(38)
E. coli ΔdinB (JW0221–1)	BW22113 but *ΔdinB749::kan*	(38)
E. coli ΔumuDC (EJ120)	AB 1157 but *Δ (umuDC)595*::Chl	(53)
pTrcUgi	pTrc99a plasmid with Ugi ORF (EcoRV/BamHI end filled) cloned in EcoRI (end filled) site	This study
pTrcUdgX	pTrc99c plasmid with *Msm*UdgX ORF cloned in NcoI/HindIII site	This study
pTrcUdgX-Ugi	pTrcUdgX with Ugi (released from pTrcUgi EcoRV/HindIII endfilled) cloned in EcoRV site	This study
pMVUgi	pMV261 (Kan^R) with Ugi ORF (EcoRV/BamHI end filled) cloned in EcoRI(end filled) site	This study
pMVUdgX	pMV261 (Hyg^R)with *Msm*UdgX ORF cloned in BamHI/HindIII site	This study
pMVUdgX-Ugi	pMVUdgX with Ugi (released from pMVUgi digested with XbaI/NheI) cloned in XbaI site	This study

Uracil release assay

Substrate DNA containing [^{3}H]-uracil was prepared by PCR amplification of *Msm*UdgX ORF using 10 pmol each of *Msm*UdgX Fp and *Msm*UdgX Rp primers, Taq DNA polymerase (2U), 40 μM [^{3}H]-dUTP (GE healthcare, UK) and 200 μM each of dCTP, dATP and dGTP in a 30 μl volume. PCR conditions involved heating at 94°C for 4 min followed by 29 cycles of 94°C for 1 min, 60°C for 30 s and 72°C for 1 min and final extension at 72°C for 10 min. The PCR product was purified and used in the uracil release assays. Uracil release was determined by incubating 1 μg of the *Msm*UdgX (pre-incubated with 0.8 μg of Ugi or buffer alone for 15 min at room temperature followed by 30 min on ice) or Ung (control) with 21 000 c.p.m. of ^{3}H-labeled uracil DNA and incubated at 37°C for 30 min. At the end of the reaction, the samples were mixed with cold uracil and spotted on the CEL 300 PEI/UV254 (POLYGRAM, Macherey-Nagel, Germany) thin layer chromatography (TLC) plates, and developed with 0.2% formic acid (v/v) containing 0.55 M LiCl as the mobile phase at 4°C. The TLC plate was dried and exposed to phosphorimager cassette for 7 days and the cassette was scanned for image acquisition.

Homology modelling of *Msm*UdgX

The structure of *Msm*UdgX was modeled by Phyre2 server (40) using PDB_ID: 1UI0 as template which corresponds to crystal structure of a uracil-DNA glycosylase from *Thermus thermophilus* HB8. Both model and template structures superimposed well with low RMSD (0.11 Å).

Generation of mutations in UdgX, their purification and activity assays

PCR based methods (see supplementary material) were used to mutate the KRRIH and the motif A regions of UdgX. Mutant proteins were purified from *E. coli* BL21 (DE3) *ung*- strain (except for *Msm*UdgX F4, which was purified from *E. coli* Rosetta (DE3) strain) using Ni-NTA column chromatography as described for the wild type *Msm*UdgX.

Generation of constructs for Ugi, *Msm*UdgX and *Msm*UdgX-Ugi for expression in *E. coli* and *M. smegmatis*

Ugi ORF was PCR amplified from *B. subtilis* phage PBS-2 DNA (∼100 ng) with primers, Ugi Fp (5′ AGGAGGATCCTCAACATGACAAATTTATCT 3′) containing BamHI

site and Ugi Rp (5′ ATAGGGATATCCCTATACAC-TAATATTTATAC 3′) containing EcoRV site using Pfu DNA polymerase. PCR consisted of denaturation at 94°C for 5 min, followed by 30 cycles of heating at 94°C for 1 min, annealing at 53°C for 35 s and extension at 70°C for 1 min, and then a final extension at 70°C for 10 min. The amplicon was eluted and ligated into pJET1.2 to generate pJETUgi. (Note: The BamHI/EcoRV sites were generated originally for cloning Ugi in pcDNA3.1+ vector, which is not used in this study). pJETUgi was digested with BamHI, end-filled with Klenow DNA polymerase and then digested with EcoRV. The resulting product was cloned into the EcoRI digested and Klenow DNA polymerase end-filled pTrc99a and pMV261 (Kan^R) vectors generating pTrcUgi and pMVUgi (Kan^R), respectively, and confirmed by restriction digestion and DNA sequencing. To generate pTrcUdgX-Ugi, Ugi was released from pTrcUgi by digestion with EcoRV/HindIII, end-filled with Klenow DNA polymerase and cloned into EcoRV digested pTrcUdgX to generate pTrcUdgX-Ugi. To construct pMVUdgX-Ugi, Ugi was released from pMVUgi by XbaI/NheI digestion and cloned into the XbaI site of pMVUdgX (Hyg^R). Both pTrcUdgX-Ugi and pMVUdgX-Ugi were validated by XmnI digestion and DNA sequencing.

Growth analysis of *E. coli* lacking different DNA repair and SOS response genes

Growth analyses of *E. coli* (3–5 replicates) harboring pTrc99c or pTrcUdgX-Ugi were performed in 100 well microtiter plates. Aliquots (200 μl) of 10^{-2} dilutions of saturated cultures in LB containing Amp and 0.5 mM IPTG were taken in the absence or presence of 1 to 2.5 mM H_2O_2 or 1 to 2.5 mM $NaNO_2$. For growth in $NaNO_2$, the medium was adjusted to a pH of 5.5. Growth was monitored as OD_{600} for 24 h at 37°C using Bioscreen C kinetic growth reader under constant shaking. Growth curves were prepared from the growth of three independent colonies for each strain and the mean ± SD were plotted.

Growth analysis of *M. smegmatis* Δ*recA* strain

Isolated colonies of *M. smegmatis* (wild type and *ΔrecA*) harboring pMV261 or pMV261UdgX-Ugi were inoculated in LB-Tween containing Hyg till saturation. Aliquots (200 μl) of 10^{-2} dilution of the culture were then inoculated into microtiter plates supplemented with 0.5% (w/v) BSA, Hyg and 0, or 3 mM H_2O_2 and 1 mM $NaNO_2$. pH of the medium for the $NaNO_2$ treated culture was adjusted to 5.5. The culture growth was monitored by measuring OD_{600} at regular intervals using Bioscreen C kinetic growth reader. Growth curves were prepared from three independent colonies for each strain and the mean ± SD values were plotted.

RESULTS

Identification of *Msm*UdgX and its homologs

BLAST analysis of *M. smegmatis* genome using family 4 UDG sequence (*Tth*UdgA, TTHA1149) as query identified a family 4 like protein (MSMEG_0265) which we designated as *Msm*UdgX. Subsequent BLAST analysis using *Msm*UdgX as query sequence to search for its homologs revealed the presence of UdgX in other mycobacteria like *M. avium* complex, *M. haemophilum, M. chubense* etc., other actinobacteria, e. g. *Streptomyces coelicolor, Rhodococcus spp, Nocardia farcinica, Gordonia naminbiensis* etc.), and other organisms such as *Xanthomonas axonopodis*, *Thiobacillus denitrificans, Rhizobium leguminosarum, Bradyrhizobium japonicum* etc. Homologs of UdgX are absent from archaea and eukaryotes. The protein is also annotated as SPO1 DNA polymerase or DNA polymerase because of its homology with the N-terminal of *Bacillus* phage SPO1 DNA polymerase. Multiple sequence alignment of *Msm*UdgX and its homologs from different organisms shows the presence of many conserved amino acids. The motif A, GEQPG and motif B, HPS(S/A)(L/I)L, sequences are well conserved across different organisms (Figure 1A).

Purification of *Msm*UdgX and UV-VIS scan

Our attempts to purify *Msm*UdgX from *E. coli ung*- strain were unsuccessful. In fact, we did not even succeed in transforming *E. coli ung*- strain with the expression constructs. Hence, an N-terminally His-tagged *Msm*UdgX was purified to near homogeneity from *E. coli* BL21 (DE3) or *E. coli* Rosetta (DE3) harboring pET14b*Msm*UdgX (Figure 1B). The purified protein was brown in color (Figure 1C (ii) inset) indicating it to be an 4Fe-4S cluster protein like the family 4 and family 5 (Figure 1C (iii) inset) UDGs. UV-VIS scan of UdgX showed a peak at 415 nm (besides the expected peak at 280 nm) which is also indicative of the presence of Fe-S cluster. UdgB (a known Fe-S cluster protein) but not Ung (a non Fe-S cluster protein) showed a similar peak at 415 nm (compare Figure 1C (ii) with (i) and (iii)).

*Msm*UdgX binds specifically and tightly to uracil-DNA but does not excise uracil

We checked *Msm*UdgX activity on uracil containing ss-DNA and dsDNA substrates (SSU9 and SSU9:G, respectively). Analysis of the reaction products on 8 M urea polyacrylamide gels revealed a slower migrating band, 'C' instead of a faster migrating product, 'P' band when compared with the substrate, 'S' band (Figure 2A (ii), compare lanes 4, 5 and 8 with 2 and 7). Addition of Ugi in reactions inhibited Ung activity (compare lanes 2 and 3). However, the complex formation by *Msm*UdgX remained unchanged (lane 5). Thus, unlike other UDGs, *Msm*UdgX did not appear to excise uracil but formed a stable complex ((Figure 2(A) panel (i)). Proteinase K digestion following *Msm*UdgX reaction still revealed a slightly shifted band ('S*') compared to the uracil-DNA ('S') indicating the presence of an attached peptide (Figure 2B, compare lane 1 with 3). To check for the requirement of any co-factors/partners for activity, we mixed *M. smegmatis ΔungΔudgB* cell free extract with *Msm*UdgX. However, the binding property of *Msm*UdgX remained unchanged (Supplementary Figure S1).

As complex formation has so far not been observed for any UDG superfamily proteins, we cloned, partially purified and checked the activities of UdgX proteins from *M.*

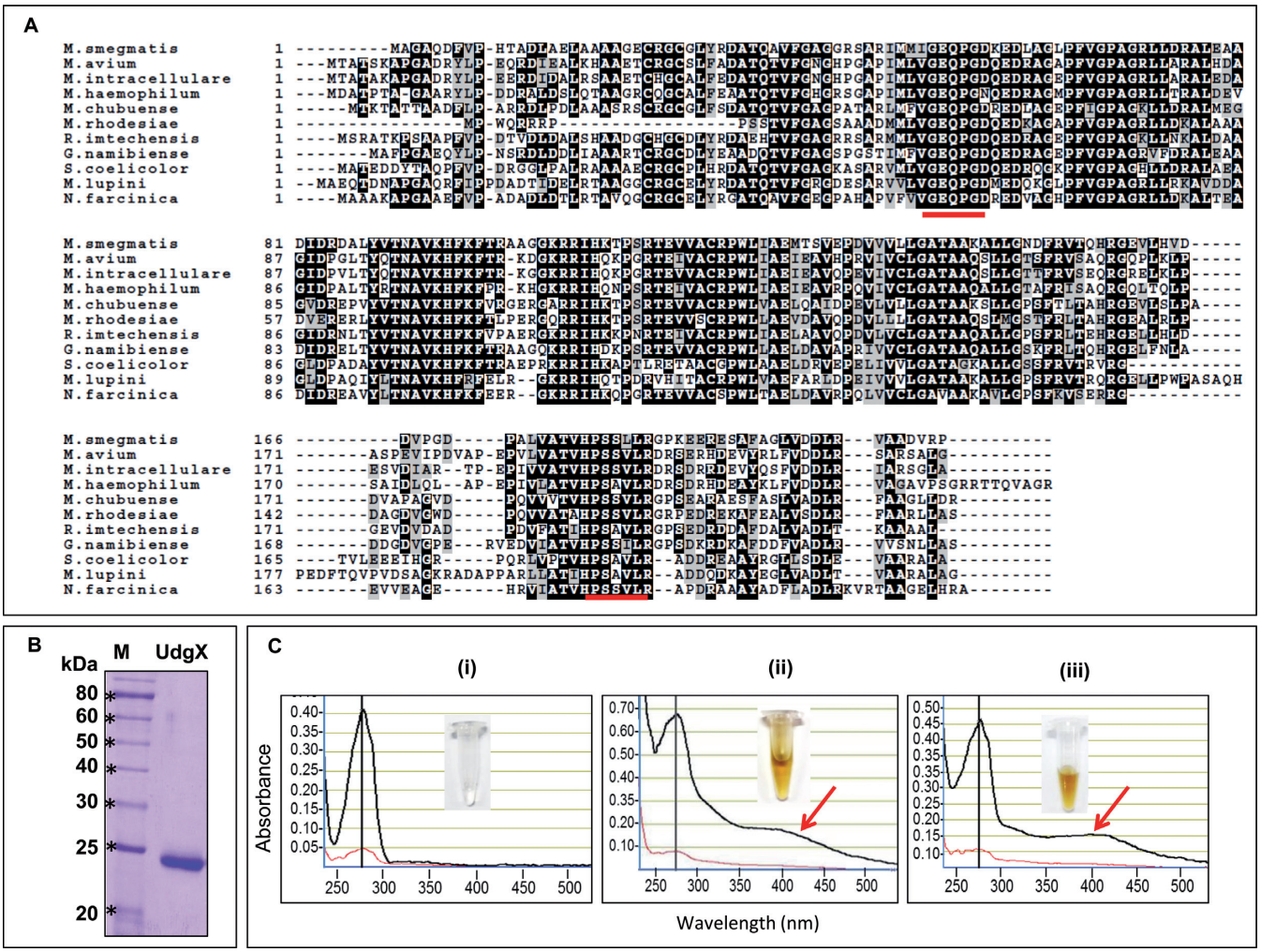

Figure 1. Multiple sequence alignment of *Msm*UdgX with its homologs, purification of *Msm*UdgX and its UV-VIS scan. (**A**) Multiple sequence alignment of *Msm*UdgX and its homologs from other bacteria using Clustal W and Boxshade server. Identical residues are shown in black whereas the similar residues are shown in light gray boxes. The motif A and motif B sequences have been underlined. NCBI reference sequence ID are: *M. smegmatis*, WP_011726794.1; *M. avium*, ZP_05218036.1; *M. intracellulare*, ZP_05227544.1; *M. haemophilum*, ZP_21917190.1; *M. chubuense*, YP_006453151.1; M. rhodesiae, WP_014212855.1; *R. imtechensis*, ZP_10004147.1; *G. namibiense*, ZP_10960189.1; *S. coelicolor*, NP_628659.1; *M. lupini*, ZP_21028631.1; and *N. farcinica*, YP_118940.1. (**B**) 12% SDS-PAGE analysis of *Msm*UdgX showing its purity. (**C**) UV-VIS scan of, (**i**) *Msm*Ung (**ii**) *Msm*UdgX, (**iii**) *Mtu*UdgB. The black and red lines are absorbances at 1 mm and 0.1 mm, respectively. The presence of additional absorbance peaks at 415 nm (arrow) in *Msm*UdgX and *Mtu*UdgB are indicative of Fe-S cluster. Insets show the colors of the protein stocks.

avium and *R. imtechensis*. Even these proteins (*Mav*UdgX and *Rim*UdgX) revealed the uracil-DNA binding property (Supplementary Figure S2). Since the UdgX proteins were purified from *E. coli ung*$^+$ strain, as a control, Ugi was also included in the reactions. The presence of Ugi diminished (Supplementary Figure S2, lanes 6 and 8) the product bands, 'P' observed in lanes 5 and 7, suggesting them due to the contaminating *Eco*Ung in these preparations.

UDGs excise uracil from the DNA and generate AP-site. To check if *Msm*UdgX bound uracil-DNA or AP-DNA (after uracil excision), we investigated uracil release from [^{3}H]-uracil-DNA (Figure 2C). Uracil was released by Ung but not by *Msm*UdgX (compare lanes 2 and 4). As expected, uracil excision by Ung was inhibited by Ugi (lane 3). These data suggest that *Msm*UdgX binds uracil-DNA without excising the base to any detectable level.

We then analyzed formation of *Msm*UdgX complexes with uracil-DNAs of varying sizes on SDS-PAGE (following addition of the SDS sample loading dye and heating at 90°C for 10 min). Extent of band shifts (due to complex formation) corresponded to the sizes of uracil-DNA (Figure 2D). This analysis also indicated that the *Msm*UdgX and uracil-DNA complexes are resistant to β-mercaptoethanol, SDS and heat. Further, once *Msm*UdgX binds to uracil-DNA, Ung is unable to act on it (Figure 2E, lane 4). However, when *Msm*Ung and *Msm*UdgX were added simultaneously, or when *Msm*Ung was added first, uracil excision but no complex formation was seen (Figure 2E, lanes 5 and 6) indicating that *Msm*UdgX recognized uracil-DNA less efficiently than *Msm*Ung. Importantly, it suggested that *Msm*UdgX did not bind AP-DNA. Also, *Msm*UdgX neither excised nor bound DNAs with other modified bases like dihydroxyuracil, hydroxymethy-

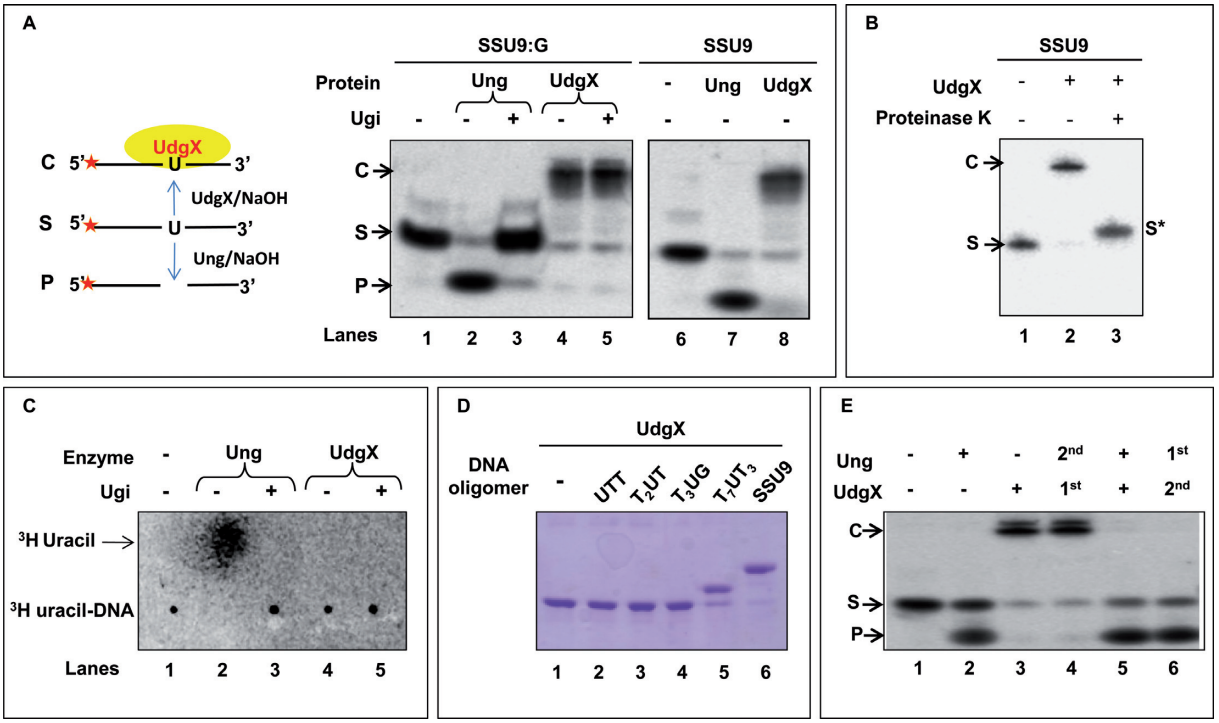

Figure 2. Activity assay of *Msm*UdgX. (**A**) Assay of *Msm*UdgX (100 ng) on 5′ ^{32}P-end labeled SSU9:G and SSU9 substrate, 'S' (∼10,000 c.p.m). Family 1 UDG (Ung) was used as control, Ugi was added as indicated. Reactions were resolved on 8 M Urea PAGE (15%) and analyzed by phosphor imaging. The presence of uracil excision activity can be seen as a fast migrating species forming a product 'P' whereas uracil binding activity form a slow migrating complex 'C' [panel (i) on left shows a diagrammatic sketch whereas the panel on right (ii) shows the results of the experiment]. (**B**) Proteinase K (1.2 U) digestion of the complex, 'C' observed in A, and re-resolved on 8 M Urea PAGE (15%) and analyzed by phosphor imaging. (**C**) Uracil release assay using ^{3}H-uracil containing DNA substrate. *Msm*UdgX (∼1 μg) or Ung (100 ng) was allowed to react with ^{3}H-uracil DNA substrate (∼21 000 c.p.m.), the samples were mixed with unlabeled uracil and spotted on the CEL 300 PEI/UV254, resolved using a mobile phase of 0.2% formic acid (v/v) containing 0.55 M LiCl at 4°C and analyzed by phosphor imaging. Released uracil is seen as a fast migrating species. (**D**) Binding of *Msm*UdgX (∼1 μg) with unlabeled uracil containing DNA oligomers (50 pmol) of different sizes resolved on 15% SDS PAGE and visualized by Coomassie blue staining. (**E**) Activity assay of *Msm*UdgX (4 pmol) in the presence of *Msm*Ung (4 pmol) using 5′ ^{32}P-labeled SSU9:G. Reactions were resolved on 8 M urea-PAGE (15%) and analyzed by phosphor imaging. The 1st and 2nd represents the order of addition, and '+' in both *Msm*Ung and *Msm*UdgX indicates that both the protein are added simultaneously.

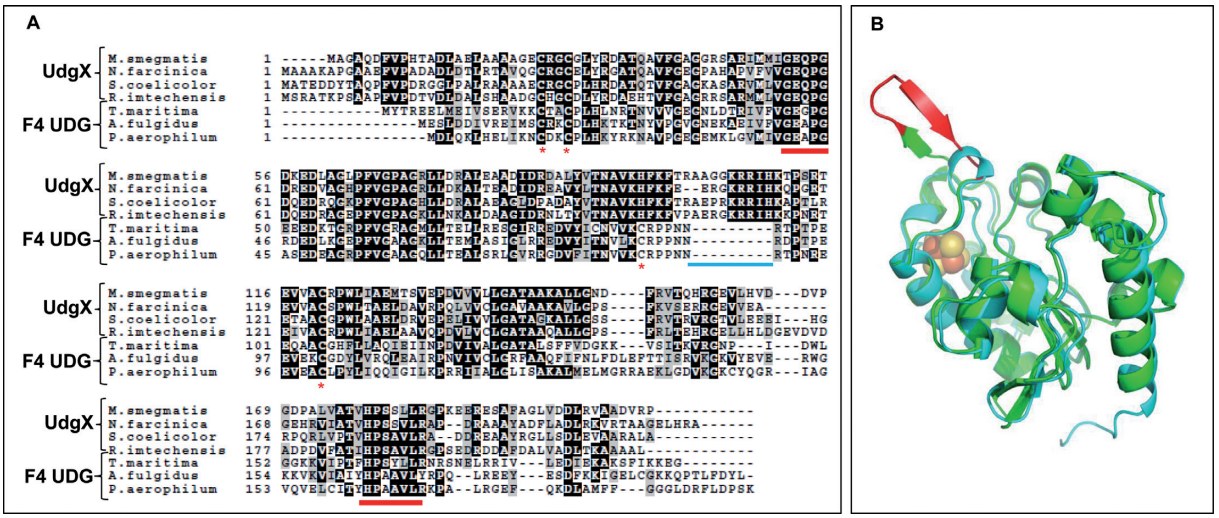

Figure 3. Alignment of UdgX with Family 4 UDGs and its homology modeling. (**A**) Multiple sequence alignment of *Msm*UdgX proteins with Family 4 UDGs from other bacteria. The motif A and B are indicated by red underline, cysteines involved in Fe-S cluster formation are indicated as red asterisk, and the extra stretch of amino acids in *Msm*UdgX is indicated by blue line. (**B**) Homology modeling of *Msm*UdgX structure using *Tth*UdgA (pdb:1UI0) as a template. Shown in green is *Msm*UdgX and the cyan is *Tth*UdgA. The extra stretch of amino acids (corresponding to AAGGKRRIH) in *Msm*UdgX is shown in red. The [Fe–S] cluster of *Tth*UdgA is shown as spheres.

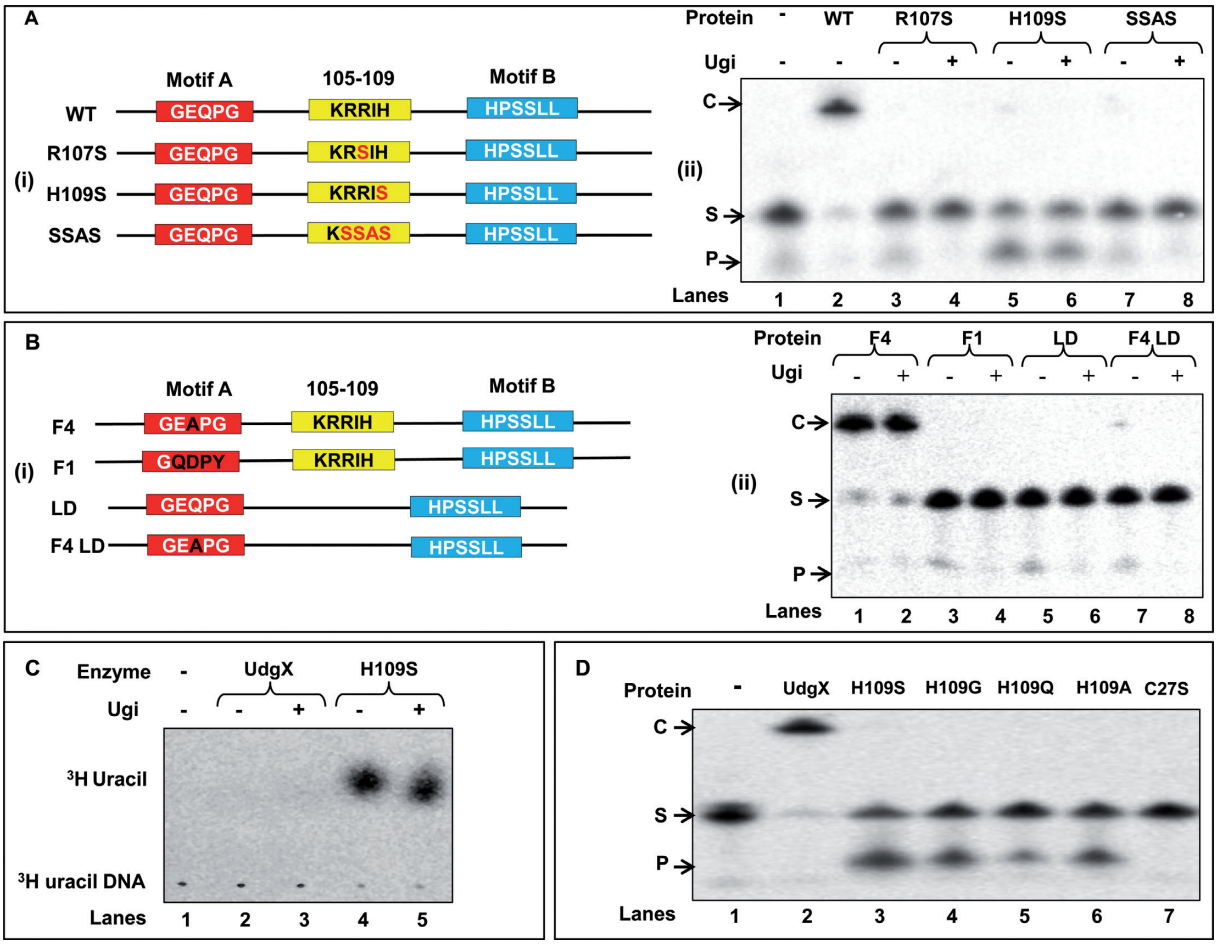

Figure 4. Site directed mutagenesis of the KRRIH region of *Msm*UdgX and its activity assay. (**A** (**i**) and **B** (**i**)) Line diagrams of *Msm*UdgX and its mutants highlighting the regions targeted by site directed mutagenesis. Motif A, motif B and the KRRIH loop region are shown in red, blue and yellow boxes, respectively. (**A** (**ii**)) Uracil excision/binding assays with *Msm*UdgX (WT) and its R107S, H109S and SSAS mutant derivatives. ∼200 ng of each protein was reacted with 5′ ^{32}P-labeled SSU9, the reaction mix was resolved on 8 M urea-PAGE (15%) and analyzed by phosphor imaging. The presence of uracil excision activity is seen by the fast migrating product band, 'P' whereas the uracil-DNA binding activity by the slow migrating complex, 'C'. (**B** (**ii**)) Uracil excision/binding assays with *Msm*UdgX (WT) and its F4, F1, LD and F4 LD mutant derivatives (∼200 ng each) using SSU9. (**C**) Uracil release assay of *Msm*UdgX H109S (∼200 ng) using ^{3}H-uracil containing DNA (21 000 c.p.m) as substrate. Released uracil was separated by TLC (CEL 300 PEI/UV254) and seen as a fast migrating species upon phosphor imaging (the radioactive spot comigrated with the cold uracil spot detected by UV shadowing). (**D**) Uracil excision/binding assay with *Msm*UdgX and its H109S, H109G, H109Q, H109A and C27S derivatives. ∼200 ng of each protein was used in a standard UDG assay with SSU9 substrate.

luracil or hypoxathine (Supplementary Figure S3). Also, *Msm*UdgX was active till a temperature of 50°C (Supplementary Figure S4), and in a wide range of pH above 4.5 (Supplementary Figure S5).

Minimum substrate requirement for binding of *Msm*UdgX is pNUNN

Analysis on SDS-PAGE did not allow us to resolve *Msm*UdgX complexes with lesser than undecameric uracil-DNA (Figure 2D). To determine the minimum size substrate for binding *Msm*UdgX, we used 5′ ^{32}P-labeled DNA oligomers having different numbers of bases on either side of the uracil. *Msm*UdgX could bind the oligomers having one 5′ phosphorylated nucleotide 5′ to uracil and two nucleotides 3′ of it (Supplementary Figure S6, lane 6). Whereas, the 5′ terminally penultimately located uracil-DNA were not bound by the protein (Supplementary Figure S6, lanes 2 and 4). Our observations suggest a pNUNN as the minimal substrate for *Msm*UdgX.

Multiple sequence alignment and homology modeling of *Msm*UdgX

Multiple sequence alignment showed that *Msm*UdgX aligned with the family 4 UDGs with ∼60% sequence identity. However, a stretch of extra amino acids, AAGGKRRIH was detected in *Msm*UdgX (Figure 3A). A subset of this region, KRRIH is conserved in UdgX. Further, motif A of UdgX possesses Q in place of A in family 4 UDGs. Homology modeling of *Msm*UdgX using family 4 UDG (*Tth*UdgA, pdb:1UI0) template showed that the AAGGKRRIH sequence in *Msm*UdgX forms an outloop near the active site of the protein (Figure 3B).

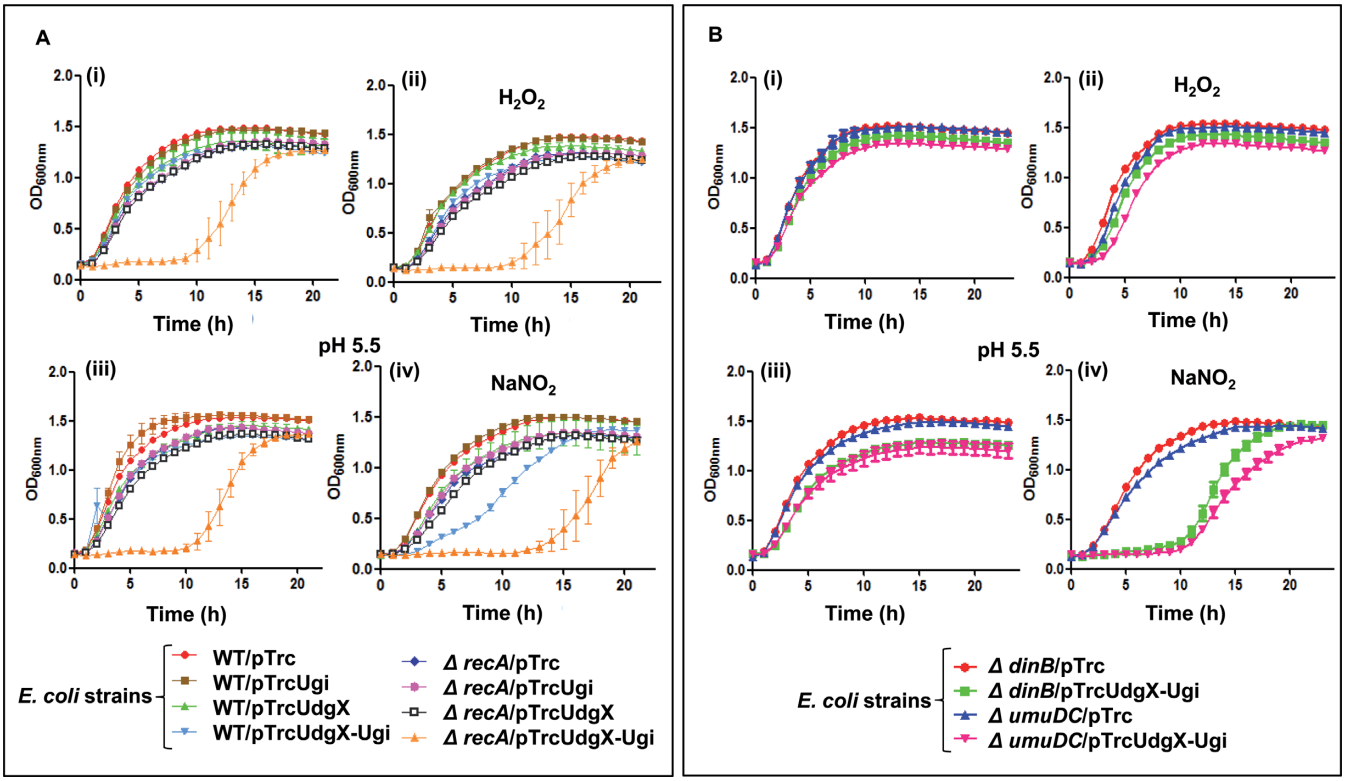

Figure 5. Growth curve analysis of *E. coli* WT and its different knockout strains derivative harboring various plasmids. (**A**) Growth curve of *E. coli* and its *ΔrecA* derivatives in the absence and presence of 1 mM H_2O_2 or 1.5 mM $NaNO_2$ (**B**) Growth curve analysis of *E. coli* and its *ΔdinB* and *ΔumuDC* derivatives in the absence or presence of 2.5 mM H_2O_2 or 2.5 mM $NaNO_2$. *E. coli* strains (3–5 replicates) were used for growth curve analysis in 100 well microtiter plates. Aliquots (200 μl) of 10^{-2} dilutions of saturated cultures in LB containing Amp and 0.5 mM IPTG were taken in the absence or presence of H_2O_2 or $NaNO_2$. For growth in $NaNO_2$, the medium was adjusted to a pH of 5.5. Growth was monitored as OD_{600} at 37°C for 24 h using Bioscreen C kinetic growth reader under constant shaking. Growth curves were prepared from the growth of three independent colonies for each strain and the mean ± SD were plotted.

Generation of mutant *Msm*UdgX and activity assay

We generated mutations in the KRRIH and motif A regions of *Msm*UdgX (Figure 4A and B). Unlike the wild type *Msm*UdgX expression constructs, mutant constructs (except for the *Msm*UdgX F4 (Q53A)) yielded transformants in the *ung*- strains. Such a loss of toxicity suggested that either the mutants lost the uracil-DNA binding activity or gained uracil excision activity. *Msm*UdgX R107S, *Msm*UdgX H109S, *Msm*UdgX SSAS (RRIH is replaced with SSAS) and *Msm*UdgX LD (ΔAAGGKRRIH, with region around the loop made similar to the family 4 UDGs) were made because of the proximity of these residues to the active site pocket and their conservation in UdgX proteins. Mutation of Q to A in the motif A in *Msm*UdgX Q53A (F4) makes it exactly similar to the motif A of family 4 UDGs; *Msm*UdgX GQDPY (F1) has the motif A, GEQPG changed to GQDPY (motif A of the family 1 UDGs). In *Msm*UdgX F4 LD mutant, motif A is similar to family 4 UDGs. Additionally, in the context of the family 4 UDG sequences, it lacks the AAGGKRRIH sequence (Figure 4A (i) and B (i)).

Assays with mutant proteins show that *Msm*UdgX H109S gained the uracil excision activity insensitive to the presence of Ugi (Figure 4A (ii), lanes 5 and 6), whereas *Msm*UdgX R107S, *Msm*UdgX SSAS, *Msm*UdgX LD, *Msm*UdgX F1, *Msm*UdgX F4 LD mutants lost even the complex formation activities without a gain of uracil excision activity (Figure 4A (ii) and B (ii)). These observations indicate the importance of the KRRIH loop in the function of UdgX. *Msm*UdgX Q53A (F4) retains the uracil-DNA binding activity indicating that Q53 is not important in the binding activity of *Msm*UdgX (Figure 4B (ii), lanes 1 and 2). *Msm*UdgX F4 LD where the loop is deleted from *Msm*UdgX Q53A loses the complex formation activity without any gain of the uracil excision activity again indicating the functional importance of the loop (Figure 4B (ii), lanes 7 and 8). A direct assay for uracil excision also showed that *Msm*UdgX H109S released uracil irrespective of the presence of Ugi (Figure 4C, lanes 4 and 5). Mutation of H109 to G, Q and A also led to the gain of uracil excision activity (Figure 4D lane 4, 5 and 6). These observations of gain of UDG activity by the diverse side chains at the position 109 emphasize the importance of H109 in tight binding of *Msm*UdgX to uracil-DNA. *Msm*UdgX H109S had no activity on dihydroxyuracil, hydroxymethyluracil and hypoxanthine in DNA (Supplementary Figure S7). Mutation of C27 involved in the formation of Fe-S cluster, inactivated UdgX (Figure 4D, lane 7).

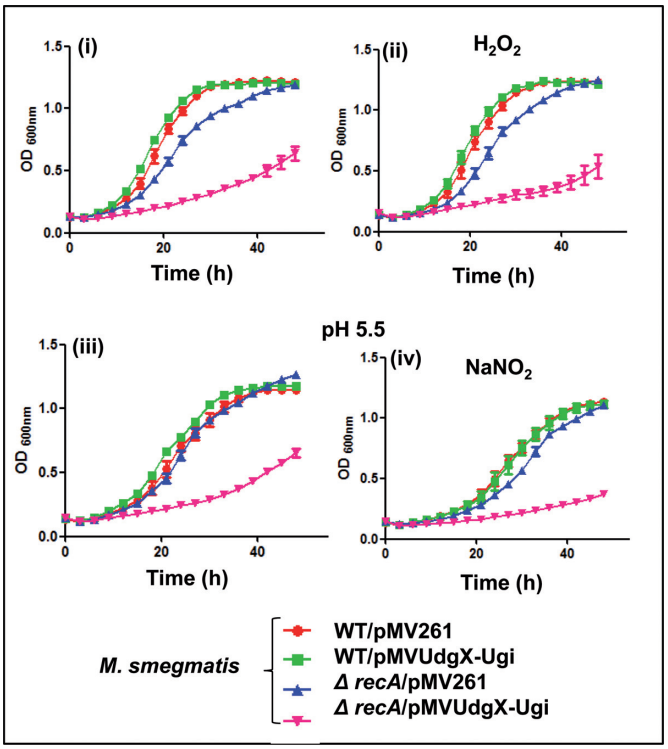

Figure 6. Growth curve analysis of *M. smegmatis* wild type (WT) and *ΔrecA* strains harboring different plasmids in the absence and presence of 2 mM H_2O_2 or 1 mM $NaNO_2$. *M. smegmatis* strains (3–5 replicates) were used for growth curve analysis in 100 well microtiter plates. Aliquots (200 μL) of 10^{-2} dilutions of saturated cultures in LB-Tween supplemented with Hyg and 0.5% (w/v) BSA were taken in the absence or presence of H_2O_2 or $NaNO_2$. For growth in $NaNO_2$, the medium was adjusted to a pH of 5.5. Growth was monitored as OD_{600} at 37°C for 50 h using Bioscreen C kinetic growth reader under constant shaking. Growth curves were prepared from the growth of three independent colonies for each strain and the mean ±SD were plotted.

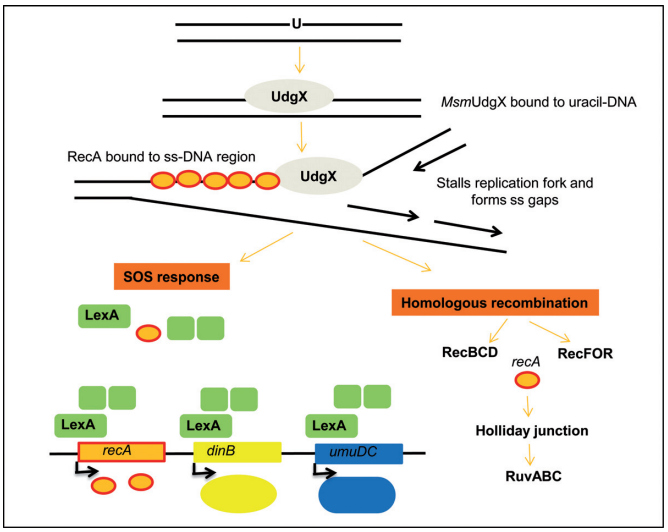

Figure 7. Proposed pathway for repair of uracil-*Msm*UdgX complexes. Binding of *Msm*UdgX to uracil would stall replication which could in turn lead to generation of single stranded gaps or double stranded breaks. Activation of RecA protein upon binding to ss DNA would induce the SOS response through cleavage of LexA. RecBCD proteins would also repair the complex through homologous recombination repair pathway.

Effect of co-expression of *Msm*UdgX and Ugi in *E. coli* deficient in nucleotide excision repair or homologous recombination repair

The observation that *Msm*UdgX forms an extremely tight complex without excising uracil presents an unusual physiological consequence of interfering with the DNA transaction processes. Thus, to understand a possible involvement of nucleotide excision repair (NER) or recombinational repair in the resolution of the complex (41,42), we developed an assay system by enhancing the presence of uracil in the genome by expression of Ugi (to inactivate the major UDG activity of Ung), overexpression of *Msm*UdgX to allow excessive formation of *Msm*UdgX and uracil-DNA complexes in the genome. If not repaired such complexes would impair cellular growth, the severity of which would enhance in the absence of the downstream repair proteins/pathways. We observed that cells lacking UvrB (NER pathway) revealed no significant growth defect when *Msm*UdgX and Ugi are co-expressed (although a similar minor growth retardation is observed in both the wild type and the mutant strains (Supplementary Figure S8)). However, a significant growth retardation occurred in the strain lacking RecA where *Msm*UdgX and Ugi were overexpressed (Figure 5A). The growth defect was enhanced under acidified nitrite stress (Figure 5A). This observation indicates that RecA is important for the downstream repair of the complex of *Msm*UdgX and uracil-DNA. Since RecA has a role in recombination as well as the SOS response, we checked for the growth of *E. coli* lacking *recB, ruvC* (homologous recombination pathway genes) and *dinB, umuDC* (SOS response genes) (Supplementary Figure S9 and Figure 5B). In these strains also, co-expression of *Msm*UdgX and Ugi resulted in a growth defect.

*Msm*UdgX and RecA mediate repair in *M. smegmatis*

To investigate the importance of RecA mediated pathway in mycobacteria, we generated a knockout of *recA* in *M. smegmatis* (Supplementary Figure S10 A, B and C), and used it to study the impact of co-expressed *Msm*UdgX and Ugi using a multicopy pMV261UdgX-Ugi (Figure 6). We observed that *M. smegmatis* also showed a phenotype similar to the one observed with *E. coli* model further suggesting that the repair of *Msm*UdgX uracil DNA complex is carried out by a RecA dependent pathway.

DISCUSSION

All known UDGs possess base excision activity regardless of the base they recognize. Here, we have identified a UDG superfamily protein from *M. smegmatis* (*Msm*UdgX) which forms a highly specific and extremely tight complex with uracil-DNA without excising the base. This unique binding property of *Msm*UdgX is common to *M. avium* and *R. imtechensis* UdgX proteins. Although binding to uracil base has been reported for archaeal family B polymerase (43,44), this is the first report from the uracil DNA glycosylase family which binds very tightly to the uracil containing DNA. UdgX proteins show highest homology to family 4 UDGs not only in the motifs A and B but also in their overall sequence and the presence of the conserved C residues known

to form Fe-S cluster. Although one of the C in UdgX is replaced with H, it still co-ordinates the Fe-S cluster seen in many proteins (45,46). However, such a replacement of C with H in UdgX might be the reason why it is not as thermostable as the family 4 UDGs (47). An amino acids stretch of AAGGKRRIH is unique to UdgX and its homologs, and the subset sequence of KRRIH within this is highly conserved in all UdgX. Activity assays under different conditions of temperatures, pH and even the presence of *M. smegmatis* cell free extracts did not result in any detectable uracil excision. However, the binding activity was maintained in all the conditions tested. The tight binding property might be important under conditions that may lead to generation of excessive uracil in DNA wherein UdgX binding may protect the genome from further replication or transcription to avoid mutations in the daughter strands or the translated proteins, respectively. Also, an excision of uracils which are closely placed can cause double stranded DNA breaks which would be detrimental to the cell (48). In fact, in *M. smegmatis*, Ung is known to be down-regulated during hypoxic growth (49). Under these growth conditions, overexpression of *Msm*Ung using a hypoxia specific promoter leads to decrease in survival of the bacteria (49). As binding of *Msm*UdgX to uracil-DNA prevents uracil excision, UdgX when present might have a protective role for the genomes under the physiological conditions that result in occurrence of excessive uracil in DNA.

To understand the nature of the complex, we carried out mass spectrometric analysis of free and DNA (TTUTT) bound *Msm*UdgX (Supplementary Figure S11A). The two extra peaks of m/z of 838.43 and 994.53 in the spectrum (Supplementary Figure S11B) of the complex corresponded to IHKTPSR and RIHKTPSR peptides (based on the MS/MS analysis, Supplementary Figure S11C) that overlapped with the sequence of the KRRIH region, indicating a functional role for this region. However, the exact nature of complex formation between *Msm*UdgX and the uracil-DNA or the amino acids involved in the complex formation remains unclear. The crystallographic analyses of *Msm*UdgX in complex with uracil-DNA might be crucial in providing an insight on the nature of the complex formation, as well in the mechanism of catalysis of uracil excision by other UDGs.

We observed that mutation of H109 of the ^{105}KRRIH^{109} loop to S (or other residues, Figure 4A) led to the gain of uracil excision activity by *Msm*UdgX, whereas changing the R107 to S, 'RRIH' to 'SSAS' or deleting the loop altogether led to loss of even the binding activity. On the other hand, mutation of Q in the motif A (GEQPG) to A to mimick motif A of the family 4 UDGs, does not affect its complex formation activity. However, changing it to GQDPY to mimick motif A of family 1 UDGs led to the loss of its complex formation property without any gain of uracil excision activity. The observation with the *Msm*UdgX F4 indicates that Q of motif A is not important in complex formation. The non-functionality of *Msm*UdgX F1 (GQDPY) might be due to the drastic amino changes that have been made in the motif A. Although the family 1 and family 4 UDGs share poor sequence homology, their core structure and the mechanism of substrate recognition seems similar (47). In fact introduction of a polar amino acid residue in the motif A of family 5 UDG also decreases the enzymatic activity (13).

Expression of *Msm*UdgX in the laboratory grown *M. smegmatis* was not detectable in our assays (Supplementary Figures S12 and S13). However, we could detect the protein when a multicopy plasmid (pMVUdgX350) was used for *Msm*UdgX expression under its native promoter (S13) suggesting that the *udgX* is a functional gene but its expression levels in laboratory grown *M. smegmatis* are low. As a control, introduction of pMVUdgX (where UdgX expression is under *hsp60* promoter) in *M. smegmatis*, as expected, resulted in *Msm*UdgX expression and complex formation (Supplementary Figures S12 and S13).

Formation of protein-DNA complexes has been reported in the context of DNA transactions caused by various endogenous and exogenous agents (50,51). These complexes are processed by DNA repair mechanisms involving NER or homologous recombination (HR) (41,52). In our analyses, strains lacking RecA were the most affected under the conditions of excessive *Msm*UdgX uracil-DNA complex formation suggesting that RecA is involved in their repair. RecA is known to have multiple roles, for example, the regulation of SOS response as well as in HR. Furthermore, deficiencies of *recB* (HR), *ruvA* (HR), *dinB* (SOS) and *umuDC* (SOS) but not *uvrB* (NER, SOS) also adversely impacted the repair. Thus, RecA might play an important role in the repair of *Msm*UdgX uracil-DNA complex.

As presence of DNA glycosylases could generate double strand DNA breaks (which are substrates for the RecBCD pathway), inactivation of Ung or Fpg DNA glycosylases has been observed to enhance survival of RecBCD deficient cells (48,53). We propose that binding of *Msm*UdgX to uracil in DNA might stall replication fork which might lead to generation of single stranded gap or double strand DNA breaks. Such lesions would invite RecA to bind to the single stranded regions which in turn could lead to cleavage of LexA and induction of the SOS responsive genes like *dinB* and *umuDC* (Figure 7). In fact, RecA is already known to play important role in the recombination repair of many stalled replication fork (54). Thus, the *Msm*UdgX uracil-DNA complex may be repaired by homologous recombination like other protein DNA crosslinks (42).

Finally, could there be a practical utility of the fact that *Msm*UdgX binds specifically to uracil in the DNA? We considered that *Msm*UdgX would have the potential to probe for the detection of uracil in genomic DNA. To test this, we spotted different amounts of *E. coli* wild type and *ung-dut-* genomic DNA on a membrane and probed with *Msm*UdgX, which was then detected with anti-UdgX antibodies. As shown in Supplementary Figure S14, we can detect the presence for uracil in the genomic DNA from *E. coli ung-dut-* strain but not from the wild type strain (Supplementary Figure S14 compare lanes 2, 3 and 4 with 1). This analysis can be further extended to tissue and cell lines to detect the presence of uracil. Although uracil detection methods using aldehyde reactive probes or comet assays are already well established, its detection by using *Msm*UdgX would be more specific and straightforward.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank our laboratory colleagues for their suggestions on the manuscript. U.V. is a J. C. Bose fellow of the Department of Science and Technology (DST), New Delhi. P.B.S. was a senior research fellow of Council of Scientific and Industrial Research, New Delhi.

FUNDING

Department of Biotechnology (DBT), New Delhi; Council of Scientific and Industrial Research, New Delhi. Funding for open access charge: Government of India.
Conflict of interest statement. None declared.

REFERENCES

1. 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.
2. Krokan,H.E., Drablos,F. and Slupphaug,G. (2002) Uracil in DNA–occurrence, consequences and repair. *Oncogene*, **21**, 8935–8948.
3. 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.
4. Pearl,L.H. (2000) Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res.*, **460**, 165–181.
5. David,S.S. and Williams,S.D. (1998) Chemistry of Glycosylases and Endonucleases Involved in Base-Excision Repair. *Chem. Rev.*, **98**, 1221–1262.
6. Zharkov,D.O. (2008) Base excision DNA repair. *Cell. Mol. Life. Sci.*, **65**, 1544–1565.
7. Seeberg,E., Eide,L. and Bjoras,M. (1995) The base excision repair pathway. *Trends Biochem. Sci.*, **20**, 391–397.
8. 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.
9. 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.
10. Neddermann,P. and Jiricny,J. (1993) The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *J. Biol. Chem.*, **268**, 21218–21224.
11. Haushalter,K.A., Todd Stukenberg,M.W., Kirschner,M.W. and Verdine,G.L. (1999) Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Curr. Biol.*, **9**, 174–185.
12. 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.*, **9**, 531–534.
13. 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.
14. Purnapatre,K. and Varshney,U. (1998) Uracil DNA glycosylase from Mycobacterium smegmatis and its distinct biochemical properties. *Eur. J. Biochem.*, **256**, 580–588.
15. Krokan,H.E., Otterlei,M., Nilsen,H., Kavli,B., Skorpen,F., Andersen,S., Skjelbred,C., Akbari,M., Aas,P.A. and Slupphaug,G. (2001) Properties and functions of human uracil-DNA glycosylase from the UNG gene. *Prog. Nucleic Acid Res. Mol. Biol.*, **68**, 365–386.
16. 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.
17. Varshney,U., Hutcheon,T. and van de Sande,J.H. (1988) Sequence analysis, expression, and conservation of Escherichia coli uracil DNA glycosylase and its gene (ung). *J. Biol. Chem.*, **263**, 7776–7784.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. Gallinari,P. and Jiricny,J. (1996) A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature*, **383**, 735–738.
24. 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.
25. 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.
26. 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.
27. Pettersen,H.S., Sundheim,O., Gilljam,K.M., Slupphaug,G., Krokan,H.E. and Kavli,B. (2007) Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic Acids Res.*, **35**, 3879–3892.
28. 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.
29. 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.
30. Srinath,T., Bharti,S.K. and Varshney,U. (2007) Substrate specificities and functional characterization of a thermo-tolerant uracil DNA glycosylase (UdgB) from Mycobacterium tuberculosis. *DNA Repair (Amst)*, **6**, 1517–1528.
31. Aravind,L. and Koonin,E.V. (2000) The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates. *Genome Biol.*, **1**, RESEARCH0007.
32. Lucas-Lledo,J.I., Maddamsetti,R. and Lynch,M. (2011) Phylogenomic analysis of the uracil-DNA glycosylase superfamily. *Mol. Biol. Evol.*, **28**, 1307–1317.
33. Venkatesh,J., Kumar,P., Krishna,P.S., Manjunath,R. and Varshney,U. (2003) Importance of uracil DNA glycosylase in Pseudomonas aeruginosa and Mycobacterium smegmatis, G+C-rich bacteria, in mutation prevention, tolerance to acidified nitrite, and endurance in mouse macrophages. *J. Biol. Chem.*, **278**, 24350–24358.
34. Malshetty,V.S., Jain,R., Srinath,T., Kurthkoti,K. and Varshney,U. (2010) Synergistic effects of UdgB and Ung in mutation prevention and protection against commonly encountered DNA damaging agents in Mycobacterium smegmatis. *Microbiology*, **156**, 940–949.
35. Rex,K., Kurthkoti,K. and Varshney,U. (2013) Hypersensitivity of hypoxia grown Mycobacterium smegmatis to DNA damaging agents: implications of the DNA repair deficiencies in attenuation of mycobacteria. *Mech. Ageing Dev.*, **134**, 516–522.
36. Snapper,S.B., Melton,R.E., Mustafa,S., Kieser,T. and Jacobs,W.R. Jr (1990) Isolation and characterization of efficient plasmid

transformation mutants of Mycobacterium smegmatis. *Mol. Microbiol.*, **4**, 1911–1919.
37. Blattner,F.R., Plunkett,G. 3rd, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F. *et al.* (1997) The complete genome sequence of Escherichia coli K-12. *Science*, **277**, 1453–1462.
38. Baba,T., Ara,T., Hasegawa,M., Takai,Y., Okumura,Y., Baba,M., Datsenko,K.A., Tomita,M., Wanner,B.L. and Mori,H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006 0008.
39. 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.
40. Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
41. Nakano,T., Morishita,S., Katafuchi,A., Matsubara,M., Horikawa,Y., Terato,H., Salem,A.M., Izumi,S., Pack,S.P., Makino,K. *et al.* (2007) Nucleotide excision repair and homologous recombination systems commit differentially to the repair of DNA-protein crosslinks. *Mol. Cell*, **28**, 147–158.
42. Nakano,T., Katafuchi,A., Matsubara,M., Terato,H., Tsuboi,T., Masuda,T., Tatsumoto,T., Pack,S.P., Makino,K., Croteau,D.L. *et al.* (2009) Homologous recombination but not nucleotide excision repair plays a pivotal role in tolerance of DNA-protein cross-links in mammalian cells. *J. Biol. Chem.*, **284**, 27065–27076.
43. Connolly,B.A., Fogg,M.J., Shuttleworth,G. and Wilson,B.T. (2003) Uracil recognition by archaeal family B DNA polymerases. *Biochem. Soc. Trans.*, **31**, 699–702.
44. Fogg,M.J., Pearl,L.H. and Connolly,B.A. (2002) Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol.*, **9**, 922–927.
45. Peters,J.W., Lanzilotta,W.N., Lemon,B.J. and Seefeldt,L.C. (1998) X-ray crystal structure of the Fe-only hydrogenase (CpI) from Clostridium pasteurianum to 1.8 angstrom resolution. *Science*, **282**, 1853–1858.
46. Messick,T.E., Chmiel,N.H., Golinelli,M.P., Langer,M.R., Joshua-Tor,L. and David,S.S. (2002) Noncysteinyl coordination to the [4Fe-4S]2+ cluster of the DNA repair adenine glycosylase MutY introduced via site-directed mutagenesis. Structural characterization of an unusual histidinyl-coordinated cluster. *Biochemistry*, **41**, 3931–3942.
47. Hoseki,J., Okamoto,A., Masui,R., Shibata,T., Inoue,Y., Yokoyama,S. and Kuramitsu,S. (2003) Crystal structure of a family 4 uracil-DNA glycosylase from Thermus thermophilus HB8. *J. Mol. Biol.*, **333**, 515–526.
48. Dianov,G.L., Timchenko,T.V., Sinitsina,O.I., Kuzminov,A.V., Medvedev,O.A. and Salganik,R.I. (1991) Repair of uracil residues closely spaced on the opposite strands of plasmid DNA results in double-strand break and deletion formation. *Mol. Gen. Genet.*, **225**, 448–452.
49. Kurthkoti,K. and Varshney,U. (2010) Detrimental effects of hypoxia-specific expression of uracil DNA glycosylase (Ung) in Mycobacterium smegmatis. *J. Bacteriol.*, **192**, 6439–6446.
50. Barker,S., Weinfeld,M. and Murray,D. (2005) DNA-protein crosslinks: their induction, repair, and biological consequences. *Mutat. Res.*, **589**, 111–135.
51. Verdine,G.L. and Norman,D.P. (2003) Covalent trapping of protein-DNA complexes. *Annu. Rev. Biochem.*, **72**, 337–366.
52. Minko,I.G., Zou,Y. and Lloyd,R.S. (2002) Incision of DNA-protein crosslinks by UvrABC nuclease suggests a potential repair pathway involving nucleotide excision repair. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 1905–1909.
53. Spek,E.J., Vuong,L.N., Matsuguchi,T., Marinus,M.G. and Engelward,B.P. (2002) Nitric oxide-induced homologous recombination in Escherichia coli is promoted by DNA glycosylases. *J. Bacteriol.*, **184**, 3501–3507.
54. Lusetti,S.L. and Cox,M.M. (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annu. Rev. Biochem.*, **71**, 71–100.