Supplementary Materials and Methods and Figures

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, S. Korea and [3]Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, 560064, India

[*]*Correspondence to*: Umesh Varshney, Phone: +918022932686, Fax: +918023602697, E-mail: varshney@mcbl.iisc.ernet.in; uvarshney@gmail.com

Supplementary Materials and Methods

Cloning of *M. avium* UdgX (*Mav*UdgX) and *R. imtechensis* UdgX (*Rim*UdgX) and its purification: The ORFs of *Msm*UdgX homologs from *M. avium* (*Mav*UdgX, 666 bp) and *R. imtechensis* (*Rim*UdgX, 651 bp) were PCR amplified using *Mav*UdgX Fp (5′ CCCGCATATGGTAGCTATTACGCGA 3′) containing NdeI site and *Mav*UdgX Rp (5′ GCAGGAATTCTTGCCGCACAAGAA 3′) containing EcoRI site for *Mav*UdgX; and *Rim*UdgX Fp (5′ CGACGCATATGTCACGCGCCACGAA 3′) containing NdeI site and *Rim*UdgX Rp (5′ CCAGGAATTCGGCTGGTTGATCACAGG 3′) containing EcoRI site for *Rim*UdgX . PCR (50 µL) consisted 200 µM dNTPs, 20 pmol each of *Mav*UdgX Fp/*Rim*UdgX Fp and *Mav*UdgX Rp/*Rim*UdgX Rp primers and 1 U of Phusion DNA polymerase (NEB) and 10 µL supernatant (out of 50 µL) of *M. avium/R. imtechensis* single colonies boiled in water. PCR conditions included initial denaturation at 98 ^{o}C for 4 min followed by 30 cycles of incubations at 98 ^{o}C for 1 min, 63 ^{o}C for 30 s and 72 ^{o}C for 45 s and then a final extension at 72 ^{o}C for 10 min. The PCR products were cloned into pJET1.2 (MBI) vector, digested with NdeI and EcoRI and cloned into similarly digested pET14b vector generating pET14b*Mav*UdgX and pET14b*Rim*UdgX, the resulting plasmid was confirmed by DNA sequencing (Macrogen, S. Korea). N-terminally His-tagged *Mav*UdgX and *Rim*UdgX were purified using pET14b*Mav*UdgX and pET14b*Rim*UdgX constructs, respectively from *E. coli* Rosetta (DE3) strain. The purification protocol is similar to the one used to purify *Msm*UdgX.

Assay of *Msm*UdgX using different substrates, under different temperature, pH, *M. smegmatis* cell free extract and Ung protein: Standard UDG assays were done for *Msm*UdgX as previously mentioned using ^{32}P-end labelled substrates of different sizes or oligomer containing different base modifications, the reaction mixture was resolved on 15% SDS PAGE or 15% polyacrylamide (19:1) 8 M urea gels, respectively. For assays at different temperatures, *Msm*UdgX along with the reaction buffer without the substrate was kept at the

required temperatures (45 ^{o}C to 65 ^{o}C) for 10 min and then the substrate was added and incubated further at the same temperature for 20 min. For assays at different pH, UDG assays were done in buffers having different pH (pH of 4.5 to 10. For checking the effect of cell free extract, *Msm*UdgX was mixed with different concentration of *M. smegmatis* cell free extract and the same assay was followed, a reaction having cell free extract alone was also taken as a control. For assays with Ung protein, *Msm*UdgX was allowed to reacts with the substrate for 10 min, and then Ung was added to the reaction, the reaction was allowed to go further for another 10 min. The order of Ung and *Msm*UdgX was reversed or both were added simultaneously in another reaction.

Generation of R107S, H109S, SSAS and LD mutations in the KRRIH region of *Msm*UdgX: To generate *Msm*UdgX R107S, *Msm*UdgX Fp (5′ TGCATATGGCGGG TGCGCAAGAT 3′), and *Msm*UdgX RS Rp (5′ GTGGATCGATC GTTTGCCTCCCGCT GCCCGCGTG 3′) were used to amplify the N-terminal region of *Msm*UdgX; and *Msm*UdgX RS Fp (5′ CAAACGATCGATCCACAAGACCCCCAGTCGTACC 3′) and *Msm*UdgX Rp (5′ CAAGCTTGCAGATGGGCTCCATC 3′) to amplify the C-terminal region, using standard Pfu DNA polymerase based PCR with an annealing at 60 ^{o}C for 35 s and an extension of 1 min. Both the PCR products were mixed together and used as a template along with *Msm*UdgX Fp and *Msm*UdgX Rp primers for another round of PCR using Pfu DNA polymerase with PCR condition similar to that used for amplification of *Msm*UdgX gene (see main text). The amplicon was eluted and digested with NdeI/HindIII and cloned into similarly digested pET14b vector to generate pET14b*Msm*UdgX R107S. The construct was confirmed by PvuI (site introduced in the mutagenic primer, *Msm*UdgX RS Rp) digestion and DNA sequencing. To generate *Msm*UdgX H109S, *Msm*UdgX Fp and *Msm*UdgX HS Rp (5′ GGTCTTGGATATC CGTCGTTTGCCTCCCGCTGC 3′) having EcoRV site were used to amplify the N-terminal region of *Msm*UdgX and *Msm*UdgX HS Fp (5′ CGACGGAT

ATCCAAGACCCCCAGT CGTACCGAG 3′) and *Msm*UdgX Rp to amplify the C-terminal region. Both the PCR product were used as a template along with *Msm*UdgX Fp and *Msm*UdgX Rp primers for another round of PCR and cloned into pET14b vector as above for pET14b*Msm*UdgX H109S. The construct was confirmed by EcoRV digestion and DNA sequencing. To generate *Msm*UdgX SSAS, *Msm*UdgX Fp and *Msm*UdgX SSAS Rp (5′ CTTGCTAGCGCTGCTTTTGCCTCCCGCTGCCCGCGTGAACT 3′) having NheI site were used to amplify the N-terminal region of *Msm*UdgX and *Msm*UdgX HS Fp (5′ AAGCAGCGCTAGCAAGACCC CCAGTCGTACCGAGGTGGTG 3′) and *Msm*UdgX Rp to amplify the C-terminal region. Both the PCR products were used as a template along with *Msm*UdgX Fp and *Msm*UdgX Rp primers for another round of PCR and cloned into pET14b vector as done above to generate pET14b *Msm*UdgX SSAS. The plasmid was confirmed by NheI digestion and DNA sequencing. To generate *Msm*UdgX LD (loop deleted), *Msm*UdgX Fp and *Msm*UdgX LD2 Rp (5′ TGCAGGCCGCCGAACAATCGTACCCCCAGTCGTACC GAGGTG 3′) were used to amplify the N-terminal region of *Msm*UdgX and *Msm*UdgX HS Fp (5′ ACGATTGTTCGGCGGCCTGCACTTGACCGCGTTGGTGACGTA 3′) and *Msm*UdgX Rp to amplify the C-terminal region. Both the PCR product were used as a template along with *Msm*UdgX Fp and *Msm*UdgX Rp primers for another round of PCR and cloned into pET14b vector as done above to generate pET14b*Msm*UdgX LD. The plasmid was confirmed by DNA sequencing.

Generation of F4, F1 and F4LD mutations in motif A along with the KRRIH region of *Msm*UdgX: To generate *Msm*UdgX Q53A (F4) mutant, the primers set *Msm*UdgX QA Fp 5′ GATGATCGGCGAGGCGCCCGGTGACAAAG 3′ and *Msm*UdgX QA Rp 5′ CTTTGTCA CCGGGCGCCTCGCCGATCATC 3′ were used. PCR amplification of left (150 bp) and right (512 bp) regions of the mutant amino acid was done using Pfu DNA polymerase with primers *Msm*UdgX Fp/*Msm*UdgX QA Rp and *Msm*UdgX QA Fp/*Msm*UdgX Rp respectively,

PCR conditions were, initial denaturation at 94 ^{o}C for 4 mins and 29 cycles of 94 ^{o}C for 1 min, 55 ^{o}C for 30 s, 70 ^{o}C for 30 s and 1 min both for the left and right regions, followed by final extension for 10 min. Both the PCR products were eluted and mixed together along with *Msm*UdgX Fp and *Msm*UdgX Rp for another round of PCR amplification. The PCR product was again eluted and cloned into pJET1.2 vector, the *Msm*UdgX QA was then mobilized into pET14b vector to get the final expression vector pET14b *Msm*UdgX Q53A which was confirmed by restriction digestion and DNA sequencing. To generate *Msm*UdgX F1 (GQDPY), the primer set *Msm*UdgX GQDPY Fp (5′ ATCATGATGATCGGCCAGGAT CCCTATGACAAAGAGGACCTG 3′) and *Msm*UdgX GQDPY Rp (5′ CAGGTCCTCTTT GTCATAGGGATCCTGGCCGATCATCATGAT 3′) was used to inverse PCR amplify the plasmid pET14b*Msm*UdgX using Pfu DNA polymerase with primers *Msm*UdgX GQDPY Fp and *Msm*UdgX GQDPY Rp. PCR conditions were, initial denaturation at 94 ^{o}C for 4 min and 24 cycles of 94 ^{o}C for 1 min, 55 ^{o}C for 30 s, 70 ^{o}C for 11 min, followed by final extension of 10 min. The PCR product was digested with DpnI restriction enzymes overnight and transformed into competent *E. coli* TG1. Plasmids were prepared from the transformants and confirmed by restriction digestion. The resultant plasmid pET14b*Msm*UdgX F1 was further confirmed by DNA sequencing. To generate *Msm*UdgX F4 (Q53A) LD, primer used are those used for generating *Msm*UdgX LD mutants but the template used was pET14b*Msm*UdgX F4.

Generation of H109Q, H109G and H109A, and C27S mutations: To mutate the H109 to other amino acids, a primer *Msm*UdgX NNN Fp (5′ CGACGCATCNNNAAGACCCCCAG TCGTACCGAG 3′) which has any combination of nucleotides i. e., NNN for the histidine codon and *Msm*UdgX NNN Rp (5′ GGTCTTNNNGATGCGTCGTTTGCCTCCCGCTGC 3′) were taken at a concentration of 30 pmoles for PCR amplification of pET14bUdgX H109S, PCR conditions include, initial denaturation at 94 ^{o}C for 4 min, and 18 cycles of 94

^{o}C for 1 min, 60.4 ^{o}C for 35 s, 70 ^{o}C for 11 min and final extension at 70 ^{o}C for 20 min using 1U of Pfu DNA polymerase. The PCR product was digested with DpnI overnight and transformed into *E. coli* TG1 competent cells. Plasmid (i.e., pET14b *Msm*UdgX NNN) was prepared from the transformants and screened by digestion with EcoRV/HindIII digestion. Positive clones were further confirmed by DNA sequencing. The method allowed us to obtain pET14b*Msm*UdgX H109Q, H109G and H109A mutants. To generate C27S mutant, primer set *Msm*UdgX CS Fp (5′ GCAGGCGAAAGCCGAGG CTGCGGGTTGTATCGC 3′) and *Msm*UdgX CS Rp (5′ ATACAACCCGCTGCCTCGG CAT TCGCCTGCCGC 3′) were used. The protocol followed was the same as mentioned above.

Expression analysis of *Msm*UdgX in *M. smegmatis*: Expression of *Msm*UdgX was analysed by immunoblotting using anti-UdgX antibodies raised in rabbit. Cell free extract of *M. smegmatis* WT, and *M. smegmatis* expressing *Msm*UdgX in a multicopy plasmid was prepared and 30 µg of the extract are resolved on 15% SDS-PAGE. The protein were transferred to a polyvinylidene difluoride (PVDF) membrane (Hybond-P, GE Healthcare) using transfer buffer [25 mM Tris-HCl (pH 8.8), 200 mM glycine, and 0.1% (w/v) SDS containing 15% (v/v) methanol] in a semidry transfer apparatus (Bio-Rad) for 20 min at 25 V. The membrane was treated with blocking solution containing 5% (w/v) skimmed milk powder in Tris-buffered saline–Tween [TBS-T; 20 mM Tris-HCl (pH 8), 150 mM NaCl, and 0.2% (v/v) Tween 20] for 2 h. Washed thrice with TBS-T buffer for 10 min each. Diluted (1:2000) anti-*Msm*UdgX antibody was added and the membrane allowed to rock for another 2 h. The membrane was then washed again thrice with TBS-T buffer for 10 min each followed by addition of 1:2000 diluted anti-IgG conjugated with alkaline phosphatase, rocked for 2 h and then washed again as before. The membrane was then equilibrated with 100 mM Tris-HCl (pH 9), and 100 µL each of 5-bromo-4-chloro-3'-indolyphosphate (BCIP, 5 mg/mL)

and nitro-blue tetrazolium chloride (NBT, 30 mg/mL) were then added along with 40 µL of 2 M $MgCl_2$ and allowed to rock until colour developed.

Mass spectrometric analysis of *Msm*UdgX and its complex with TTUTT oligomer: *Msm*UdgX (~5 µg, ~200 pmol) was allowed to bind 1 nmol of TTUTT oligomer in UDG buffer [50 mM Tris-HCl (pH 8.0), 1 mM Na_2EDTA, 1 mM DTT, 25 µg mL^{-1} of BSA] for 20 min at 37 ^oC. The reaction was then mixed with SDS loading dye; heated at 90 ^oC for 10 min and run on 12% SDS-PAGE, same amount of unbound protein was also loaded. Both the *Msm*UdgX and *Msm*UdgX bound with TTUTT (shifted band) were cut out, minced into 1 mm^3 pieces and transferred into a sterile microcentrifuge tube. The gel was washed with 500 µL of wash solution (50% acetonitrile, 50 mM ammonium bicarbonate) and incubated at room temperature for 15 min with gentle agitation (vortex mixer on lowest setting), the solution was removed and the washings were repeated twice more until the gel got decolourised. The gel was dehydrated with 100% acetonitrile for 5 min. Acetonitrile was removed and the gel was completely dried at room temperature for 10-20 min in a centrifugal evaporator. The gel piece was then rehydrated in 150 µL reduction solution (10 mM DTT, 100 mM ammonium bicarbonate) for 30 min at 56 ^oC. Reduction solution was removed with a pipette, 100 µL of alkylation solution (50 mM iodoacetamide, 100 mM ammonium bicarbonate) was added and incubated for 30 min in the dark at room temperature. The alkylation solution was removed with a pipette, 500 µL of wash solution was added and incubated at room temperature for 15 min with gentle agitation. The wash solution was discarded and the gel was dehydrated in 100 µL 100% acetonitrile for 5 min. The acetonitrile was discarded and the gel was completely dried at room temperature in a centrifugal evaporator. The gel was then rehydrated with a minimal volume of protease digestion solution (trypsin, 20 ng μL^{-1}) and kept overnight at 37 ^oC. The sample was spun down briefly by centrifugation and the supernatant (containing tryptic peptides) was transferred to sterile

centrifuge tube. A volume of 25-50 µL of extraction solution (60% acetonitrile, 0.1% TFA) was added to the gel pieces and agitated gently by vortexing at lowest setting. The sample was spun down by brief centrifugation. The supernatant was (containing additional tryptic peptides) was transferred to the previous tube. This was repeated one more time and the pooled extracted peptides were dried by centrifugal evaporation to near dryness. An aliquot (5 µL) of resuspension solution (50% acetonitrile, 0.1% TFA) was added to each tube and gently agitated on a vortex at lowest setting. The sample was spun down and 0.5 µL was mixed with 0.5 µL of α-cyano-4-hydroxycinnamic acid matrix (10 mg/mL in 50% acetonitrile, 0.1% TFA) and spotted on MALDI plate.

Cloning of *Msm*UdgX in pMV261 (Hyg^R) vector with *hsp60* promoter and its native promoter: *Msm*UdgX ORF was PCR amplified using *Msm*UdgX Fp (BamHI) (5′ CTGGAT CCGGCGGGTGCGCAAGAT 3′) and *Msm*UdgX Rp (5′ CAAGCTTGCAG ATGGGCTCCATC 3′) using Pfu DNA polymerase with initial denaturation at 94 ^oC for 4 min followed by 30 cycles of 94 ^oC for 1 min, 60 ^oC for 30 s, and 70 ^oC for 1 min. The PCR product was cloned into pJET1.2 vector, release by digestion with BamHI/HindIII and subcloned into similarly digested pMV261 (hyg^R) vector, thus generating a plasmid in which MsmUdgX is driven by hsp60 promoter present in pMV261 plasmid. The resulting plasmid was confirmed by restriction digestion and sequencing.

For cloning *Msm*UdgX with its native promoter, primers *Msm*UdgX 350up Fp (5′ CGCAGTCTAGAAATTGGACGGACCGG 3′) containing XbaI site which anneals at ~350 bp upstream of the start codon and *Msm*UdgX 100dn Rp (5′ GCCGCTGCAGCGACT CAACGAAGC 3′) containing PstI site which anneals ~100 bp downstream of the stop codon were used. Amplification was done using Pfu DNA polymerase based PCR with an initial denaturation at 94 oC for 4 min, and 30 rounds of denaturation at 94 oC for 1 min, annealing at a gradient of 62 oC for 35 s and an extension at 70 oC for 2 min 30 s followed by final

extionsion for 10 min. The PCR product was eluted and digested with XbaI/PstI and ligated to similarly digested pMV261 (hygR) generating pMVUdgX350, thus replacing the hsp60 promoter present in the plasmid. The resulting plasmid was confirmed by restriction digestion and sequencing.

Generation of *recA* knockout in *M. smegmatis*: The upstream region (left flank "LF") of *M.smegmatis recA* gene (MSMEG_2723) was PCR amplified using primers *Msm*RecA LF Fp (5' TGCGGCGGCCGCGCGGATC 3') which contains NotI site and anneals at ~900 bp upstream of the start codon, and *Msm*RecA LF Rp (5' TCGCGAATTCGAGGGCC TTTTCGC 3') which contains EcoRI and anneals ~100 bp downstream of the start codon. Similarly, the downstream region (right flank "RF") was amplified using *Msm*RecA RF Fp (5' GCTGAATTCGATGACGTCCTCC 3') which contains EcoRI site and anneals ~100 bp upstream of the stop codon and *Msm*RecA RF Rp (5' CGTACTAGTATCGCAGTTG GTGGT 3') which contains SpeI site and anneals ~900 bp downstream of the stop codon. In both the cases, PCR was done in a 25 µL volume containing 1U of Pfu DNA polymerase, 10 pmoles of each primer, 200 µM dNTP, ~200 ng genomic DNA and Pfu reaction buffer. After initial denaturation at 94 ^{o}C for 4 min, 30 cycles of reaction were done at 94 ^{o}C for 1 min, 60 ^{o}C for 35 s, 70 ^{o}C for 2 min and final extension at 70 ^{o}C for 10 min. The right flank (RF) was cloned into pJET1.2 vector. A positive clone was then digested with NotI/EcoRI (as there is one NotI site in pJET1.2 vector), and used to clone the left flank (LF) amplicon digested with NotI/EcoRI and generate pJETrecA LF-RF. This construct was confirmed by BamHI digestion. Subsequently, kanamycin cassette was released from pUC4k by digestion with EcoRI (as ~1.2 kb fragment), eluted and ligated to EcoRI digested pJETrecA LF-RF generating pJETrecA LF-kan-RF, this was again confirm by HindIII digestion. This *recA* knockout construct was then release by digestion with NotI/SpeI and ligated to similarly digested pPR27 vector. This final construct pPR*MsmrecA::kan* was then electroporated into

M. smegmatis electrocompetent cells and plated in LB-Tween agar plate containing kanamycin and incubated at 30 ^{o}C. Two transformants were inoculated into LB-Tween containing kanamycin and incubated at 30 ^{o}C till early saturation. An aliquot (100 µL) of this culture was then plated onto LB-Tween agar plate containing 10% sucrose and kanamycin, and incubated at 39 ^{o}C. The colonies which appeared were then patched on another plate having the same composition and incubated at 39 ^{o}C. The same colonies were also patched on LB-Tween agar plate containing gentamycin. The colonies which do not grow on the gentamycin plate were considered positive clones. The putative positive knockout clones are screened for UV sensitivity by exposing to UV light of 50 J/cm^2 and further confirmed by PCR amplification using primers flanking the recA gene locus as seen by the presence of 1.4 kb PCR product as compare to 1.1 kb in the wildtype.

Generation of *Msm*UdgX, Ugi co-expression construct: To understand the importance of *Msm*UdgX, it was desirable to generate a strain having excess uracil in its genome, so that in such a background overexpression of *Msm*UdgX might result in *Msm*UdgX binding to uracils and the cells defective in repairing the complex will show defective growth. This was done by over expressing Ugi (a proteinaceous inhibitor of Family 1 UDGs) and *Msm*UdgX from the same plasmid construct. For co-expression of these in *E. coli*, we used pTrc99c vector where both the *Msm*UdgX and Ugi were clone under trc promoter and for *M. smegmatis* we used pMV261 to clone both the genes under the *hsp60* promoter.

Detection of uracil in genomic DNA using *Msm*UdgX: Genomic DNA (gDNA) from *E. coli* wild type and RZ1032 (*ung-, dut-*) strains at different concentrations was diluted in 100 µL of TE buffer [10 mM Tris-HCl (pH 8), 1 mM Na_2EDTA] and blotted onto a positively charged nylon membrane (Hybond-XL, GE Healthcare) using a dotblot apparatus (BioRad). The gDNA was crosslink to the membrane using UV light. The membrane was then kept in a blocking solution i.e., 5% (w/v) skimmed milk powder made in TBS-T [20 mM Tris-HCl (pH

8), 150 mM NaCl and 0.2% Tween 20] buffer for 2 h. Washed thrice with TBS-T buffer for 10 min each. 2 µg of *Msm*UdgX protein was then added to the membrane soaked in 10 mL of blocking solution and allowed to rock for 2 h. Washed thrice with TBS-T buffer for 10 min each and then 1:2000 diluted anti-*Msm*UdgX antibody (generated in rabbit) was added and allowed to rock for another 2 h. The membrane was then again washed thrice with TBS-T buffer for 10 min each followed by addition of 1:2000 diluted anti-IgG conjugated with HRP, rock for 2 h and then washed as before. HRP substrate (luminol) and H_2O_2 were added in 1:1 ratio, and then the chemiluminesence was captured using a ChemiDoc.

Sang et. al. Fig. S1

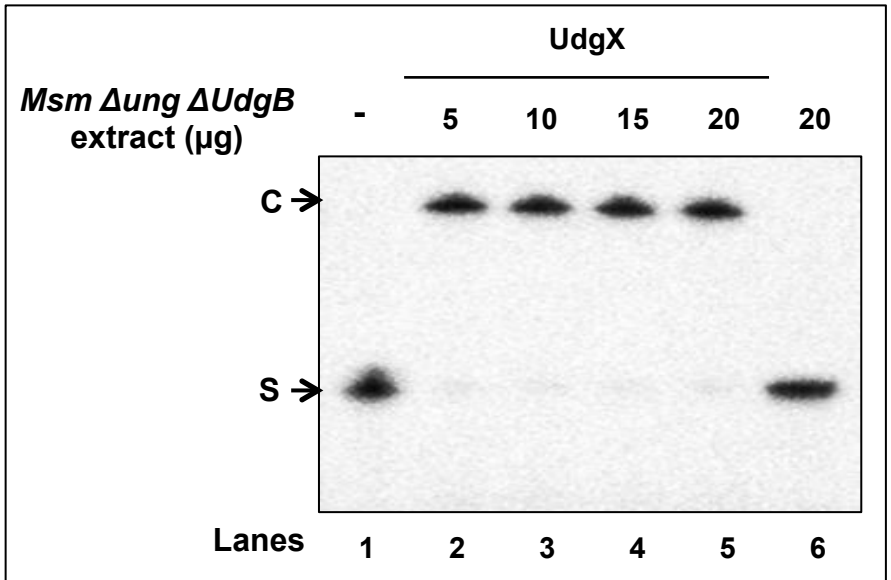

Figure S1. *Msm*UdgX activity assay in the presence of *M. smegmatis Δung ΔudgB* cell free extracts using 5' ^{32}P-labelled SSU9 substrate. ~100 ng of *Msm*UdgX was mixed with the indicated amounts of *M. smegmatis ΔungΔudgB*, cell free extracts, the mixture was used in a standard UDG activity assay. Cell free extract without *Msm*UdgX (lane 6) was taken as a control. Reactions were resolved on 8 M Urea-PAGE (15%) and analyzed by phosphor imaging. The absence of any product band indicated that *Msm*UdgX even in the precence of cellular proteins UdgX did not result in product release.

Sang et. al. Fig. S2

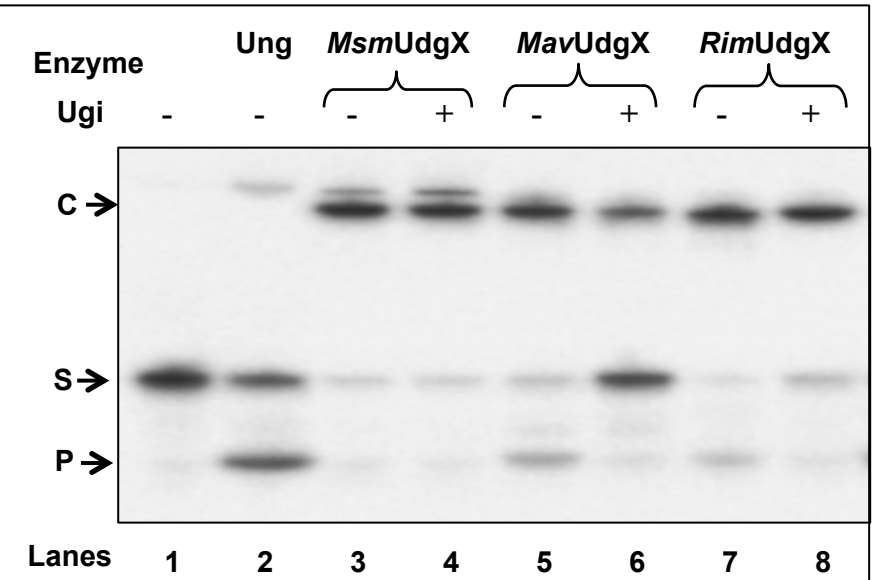

Figure S2. Assay with UdgX homologs from *M. avium* and *R. imtechensis* using 5' ^{32}P-labelled SSU9 substrate. *Mav*UdgX and *Rim*UdgX were partially purified from *E. coli* BL21 (DE3) strain, ~100 ng of each protein were used in a standard UDG activity assay, the assay was done in the presence and absence of Ugi to distinguish from any Ung activity coming for the contaminating *Eco*Ung. Reactions were resolved on 8 M Urea-PAGE (15%) and analyzed by phosphor imaging. The formation of complex by *Mav*UdgX and *Rim*UdgX indicates that this binding activity is a general property of UdgX proteins.

Sang et. al. Fig. S3

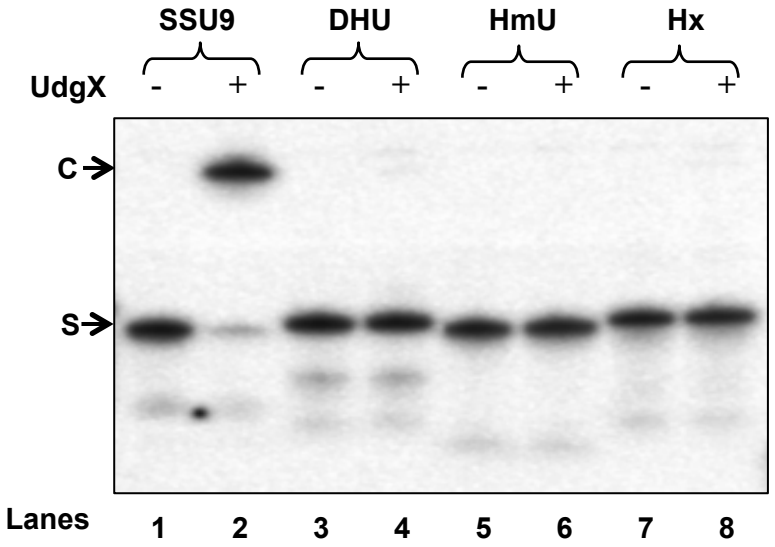

Figure S3. Assay of *Msm*UdgX with DNA oligomers (5' ^{32}P-labelled) harboring modified bases, dihydroxyuracil (DHU), hydroxymethyluracil (HmU) or hypoxanthine (Hx), and SSU9 substrate. ~100 ng of *Msm*UdgX protein was used to check for its activity on DNAs harboring modified bases in a standard UDG activity assay. Reactions were resolved on 8 M Urea-PAGE (15%) and analyzed by phosphor imaging. The absence of any product/complex bands (compared to no UdgX control) indicates that *Msm*UdgX is highly specific for uracil in DNA.

Sang et. al. Fig. S4

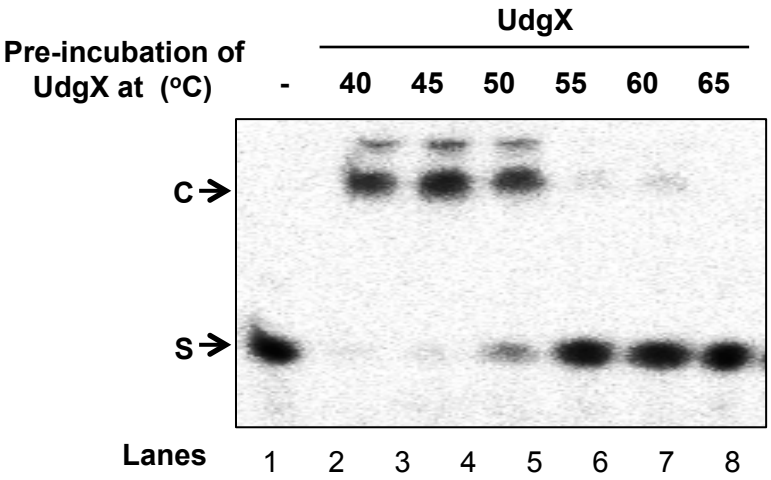

Figure S4. Activity assay of *Msm*UdgX at different temperature using 5' ^{32}P-labelled SSU9 substrate. ~100 ng of *Msm*UdgX was pre-incubated at the desired temperature for 10 min, substrate was added and incubated again at the same temperature for another 20 min. Reactions were resolved on 8 M urea-PAGE (15%) and analyzed by phosphor imaging. UdgX binding activity was observed until 50 °C which indicated that UdgX is thermostable till 50 °C.

Sang et. al. Fig. S5

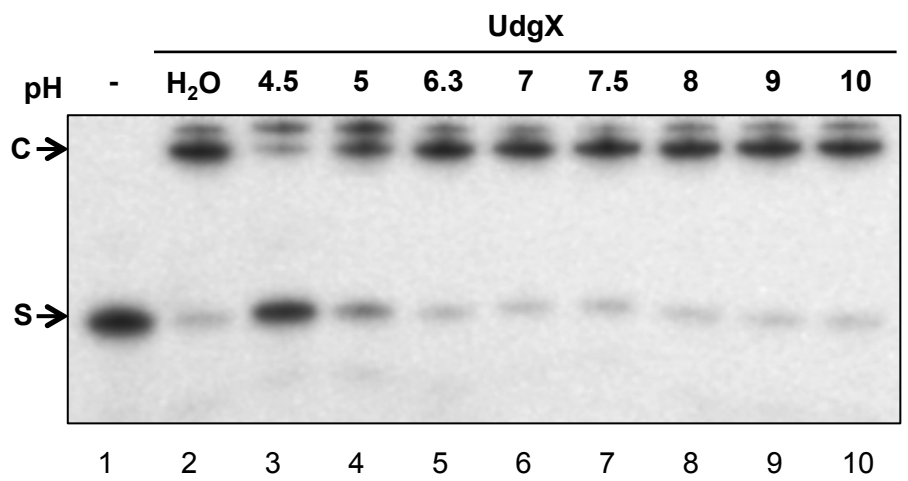

Figure S5. Activity assay of *Msm*UdgX using 5' ^{32}P-labelled SSU9 in buffer of varying pH. ~100 ng of *Msm*UdgX was used for an UDG activity assay in buffers having different pH from 4.5 to 10. Un-buffered H_2O was used as control. Reactions were resolved on 8 M urea-PAGE (15%) and analyzed by phosphor imaging. UdgX binding activity was minimal at lower pH but complex formation was observed in all pH above 5

Sang et. al. Fig. S6

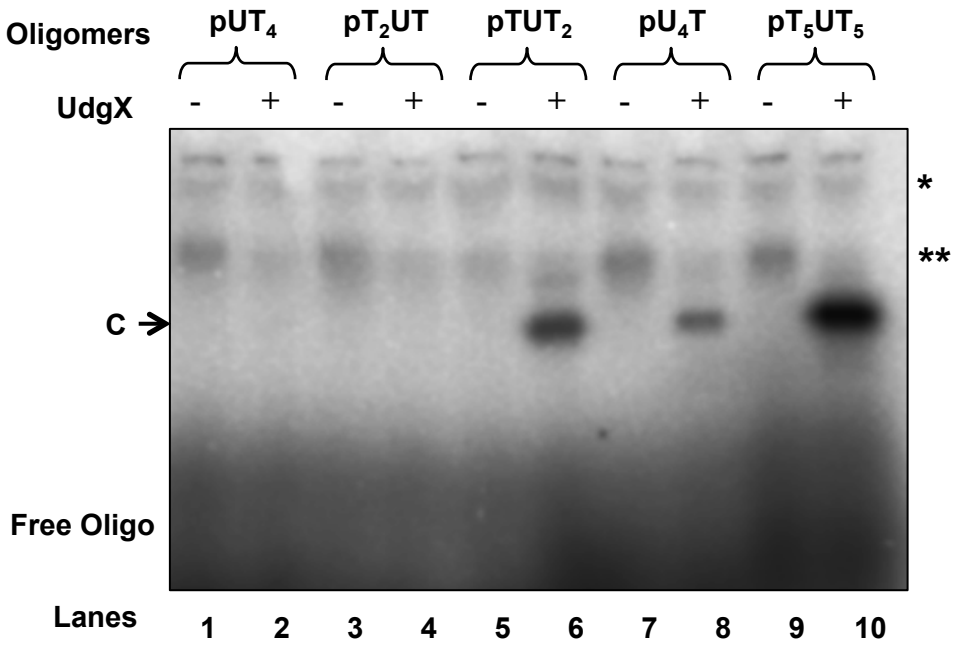

Figure S6. *Msm*UdgX binding assay with different size DNA oligomers. 5' ^{32}P-labelled DNA oligomers of different sizes and having uracil at different positions (as indicated) were used as substrates for *Msm*UdgX binding. The reaction mix was resolved on SDS PAGE (15%) and analyzed by phosphor imaging. Bound oligomers are seen as complex (with slower mobility, marked 'C') whereas the free oligomers migrated to the bottom of the gel. Band marked * and ** are background bands present in all lanes. The minimum oligomer size for binding UdgX is pTUTT.

Sang et. al. Fig. S7

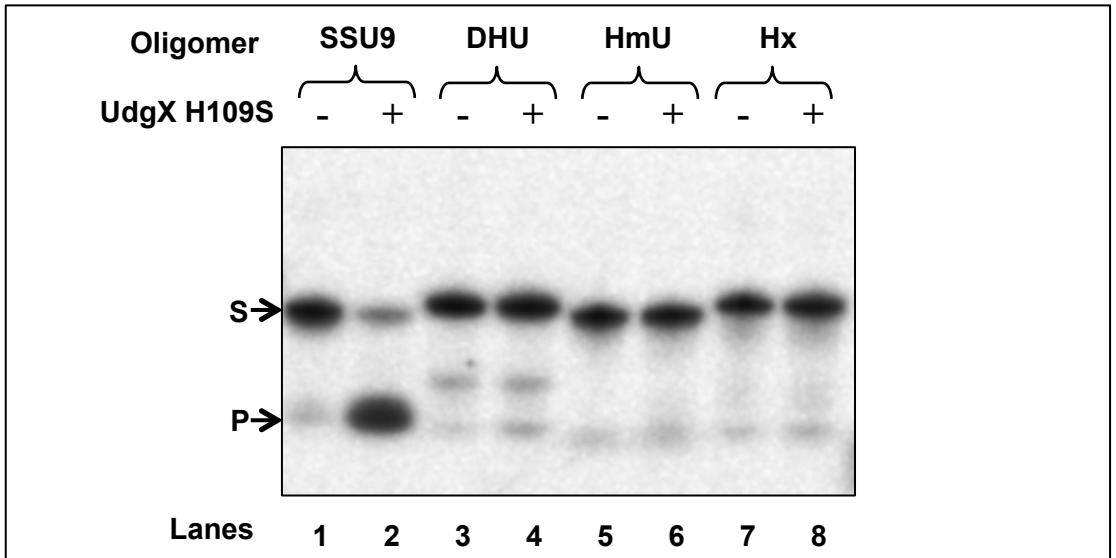

Figure S7. Activity assay of *Msm*UdgX H109S on 5' ^{32}P-labelled DNA oligomers containing dihydroxyuracil (DHU), hydroxymethyluracil (HmU) or hypoxanthine (Hx), and SSU9 as a substrate. ~100 ng of *Msm*UdgX H109S protein was used to check its activity in a standard UDG activity assay. Reactions were resolved on 8 M Urea-PAGE (15%) and analyzed by phosphor imaging. The absence of any product or complex band with the oligomers having DHU, HmU or Hx indicated that *Msm*UdgX H109S is highly specific for uracil in DNA (compare '-' UdgX H109S lanes with the '+' UdgX H109S lanes).

Sang et. al. Fig. S8

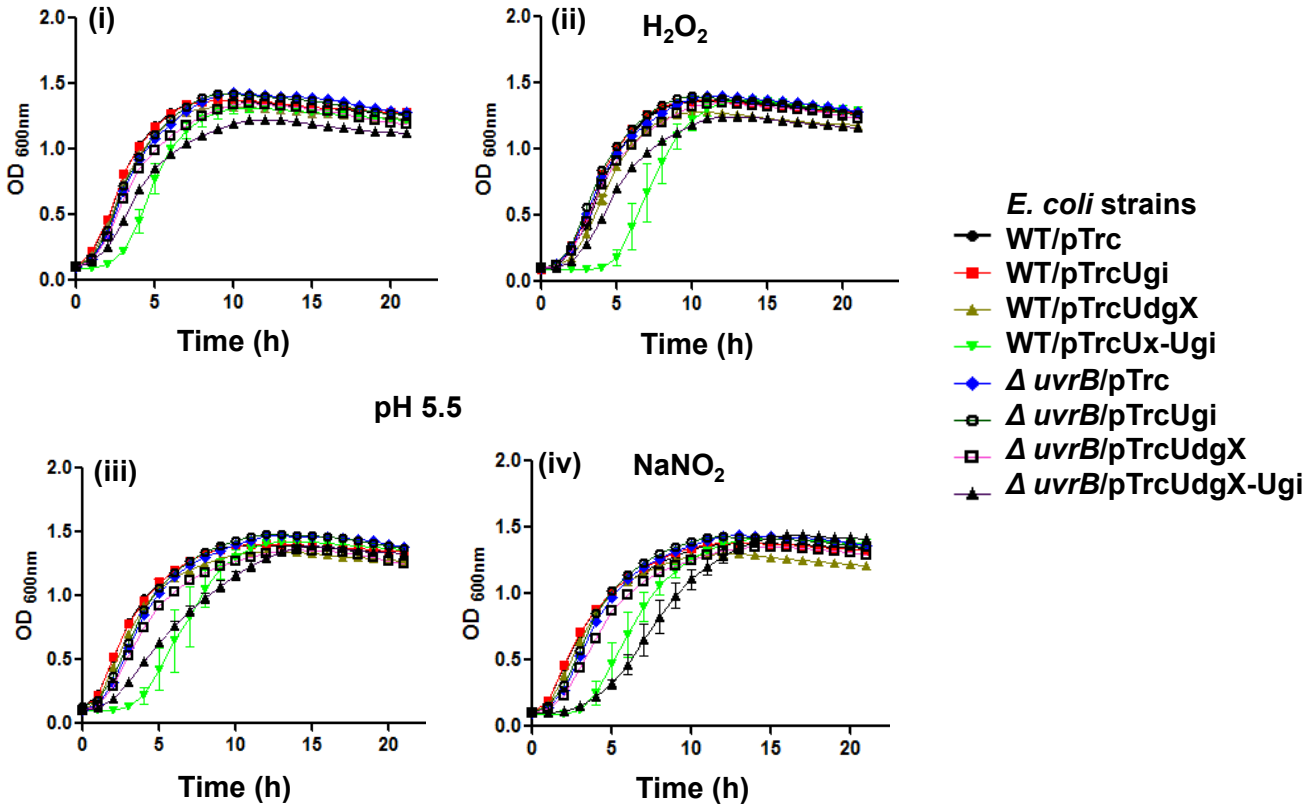

Figure S8. Growth curve analysis of *E. coli* WT and its *ΔuvrB* derivative harbouring different plasmids in the absence or presence of H_2O_2 or $NaNO_2$. *E. coli* strains (3 to 5 replicates) were used for growth curve analysis in 100 well microtiter plates. Aliquots (200 µL) of 10^{-2} dilutions of saturated cultures in LB containing Amp and 0.5 mM IPTG were taken in the absence or presence of 2 mM H_2O_2 or 1 mM $NaNO_2$. For growth in NaNO2, 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.

Sang et. al. Fig. 9

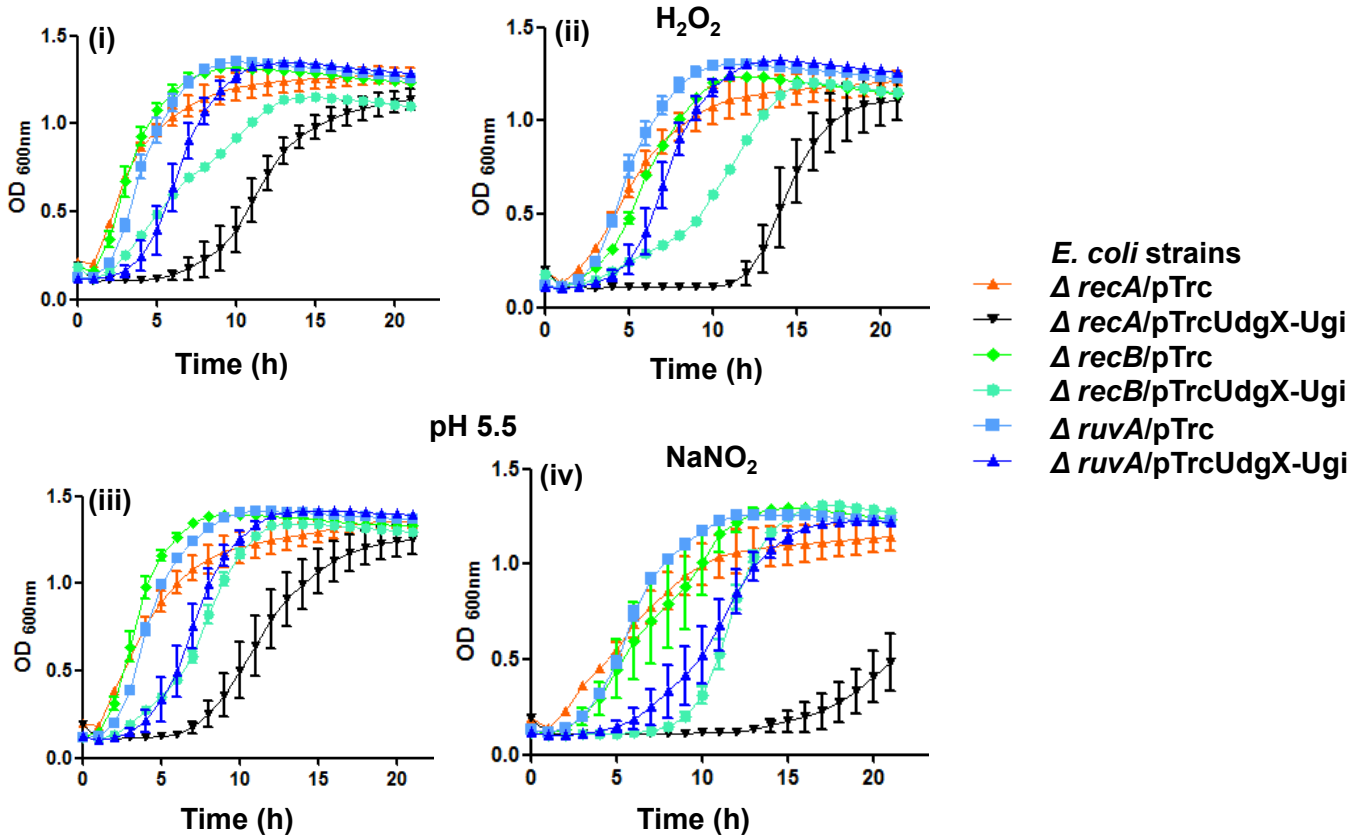

Figure S9. Growth curve of *E. coli* and its *ΔrecA, ΔrecB and ΔruvB* derivatives harbouring different plasmid in the absence or presence of H_2O_2 or $NaNO_2$. *E. coli* strains (3 to 5 replicates) were used for growth curve analysis in 100 well microtiter plates. Aliquots (200 µL) of 10^{-2} dilutions of saturated cultures in LB containing Amp and 0.5 mM IPTG were taken in the absence or presence of 2.5 mM H_2O_2 or 2.5 mM $NaNO_2$. For growth in NaNO2, 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.

Sang et. al. Fig. S10

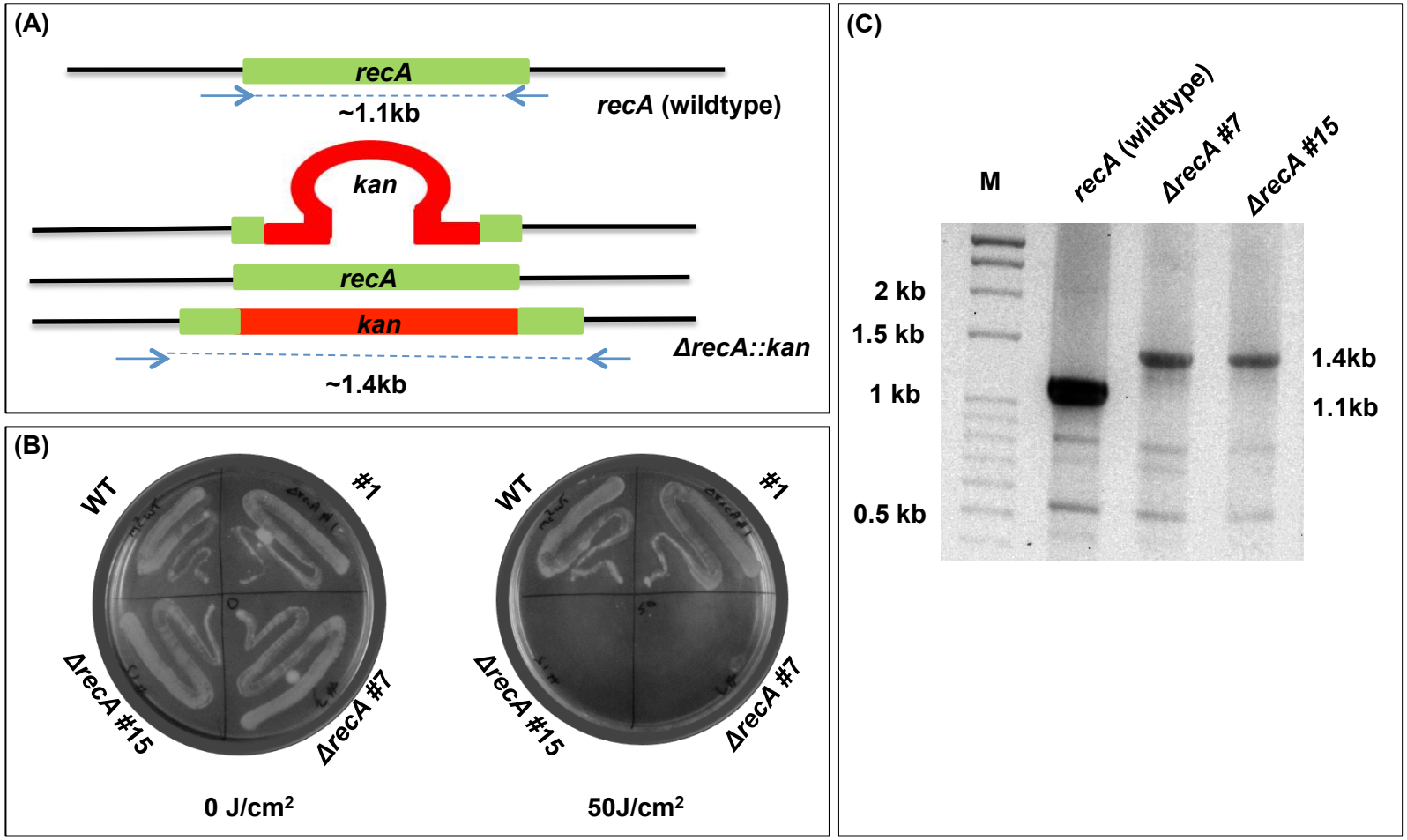

Figure S10. Generation of *M. smegmatis recA* knockout and its confirmation. Details of the Δ*recA*::*kan* strain generation are described in the supplementary text. Description of the various panels is as follows. **(A)** Schematics of *recA* gene and the *ΔrecA::kan* loci showing the primer binding sites used for knockout confirmation and the expected amplicon sizes. **(B)** UV sensitivity assay for analysis of the putative *M. smegmatis ΔrecA* strain. Isolates 7 and 15 show UV sensitivity suggesting them to be the putative Δ*recA*::*kan* strains. **(C)** PCR amplification of *recA* (WT) and *ΔrecA::kan* strains (#7 and #15) showing the wild type (1.1 kb) and the knockout (1.4 kb) amplicons, respectively.

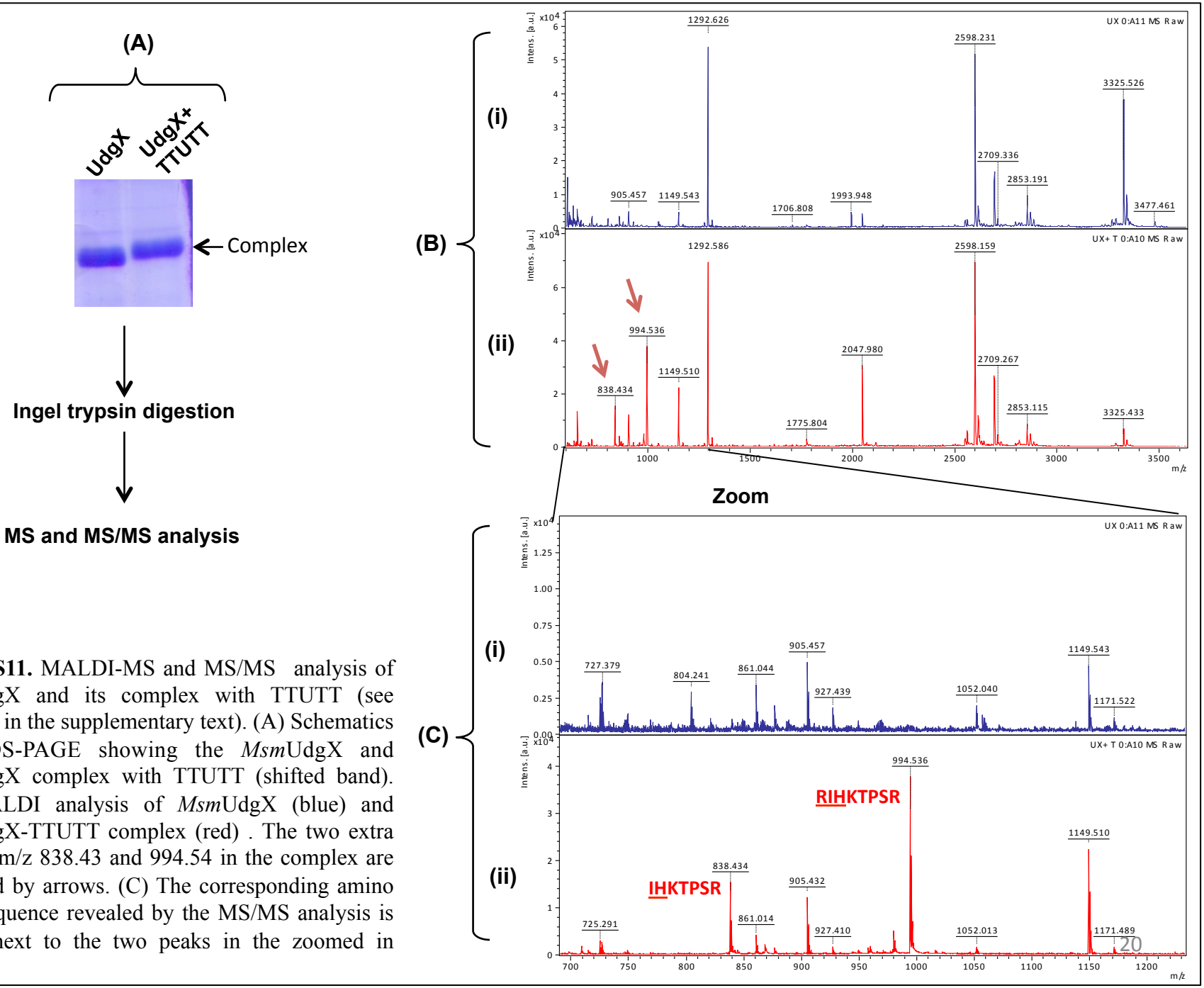

Figure S11. MALDI-MS and MS/MS analysis of *Msm*UdgX and its complex with TTUTT (see methods in the supplementary text). (A) Schematics and SDS-PAGE showing the *Msm*UdgX and *Msm*UdgX complex with TTUTT (shifted band). (B) MALDI analysis of *Msm*UdgX (blue) and *Msm*UdgX-TTUTT complex (red) . The two extra peak of m/z 838.43 and 994.54 in the complex are indicated by arrows. (C) The corresponding amino acids sequence revealed by the MS/MS analysis is shown next to the two peaks in the zoomed in image.

Sang et. al. Fig. S12

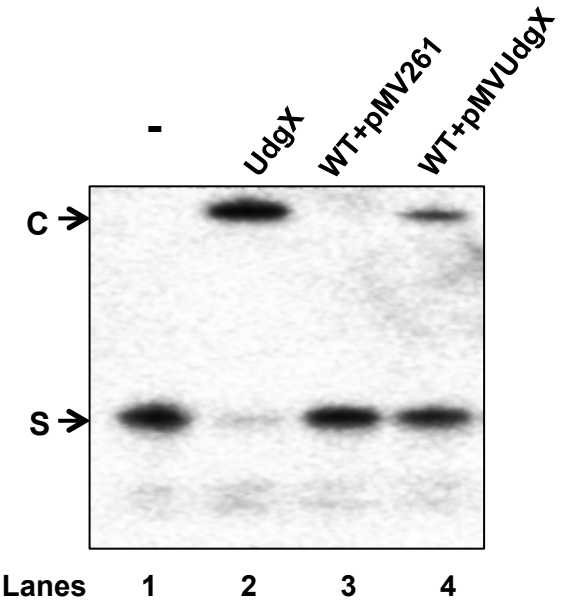

Figure S12. Activity assay of *M. smegmatis* cell free extract harbouring different plasmids (~5 µg each, pre-incubated with ~0.8 µg Ugi), using 5' ^{32}P-labelled SSU9 substrate. The reactions were resolved on 8 M Urea-PAGE (15%) and analyzed by phosphor imaging. Complex was observed with pure *Msm*UdgX protein (lane 2) as well as extract expressing *Msm*UdgX in pMV261 vector (lane 4). The observations suggest that under the laboratory grown conditions, the physiological levels of UdgX are undetectable by the complex formation assay.

Sang et. al. Fig. S13

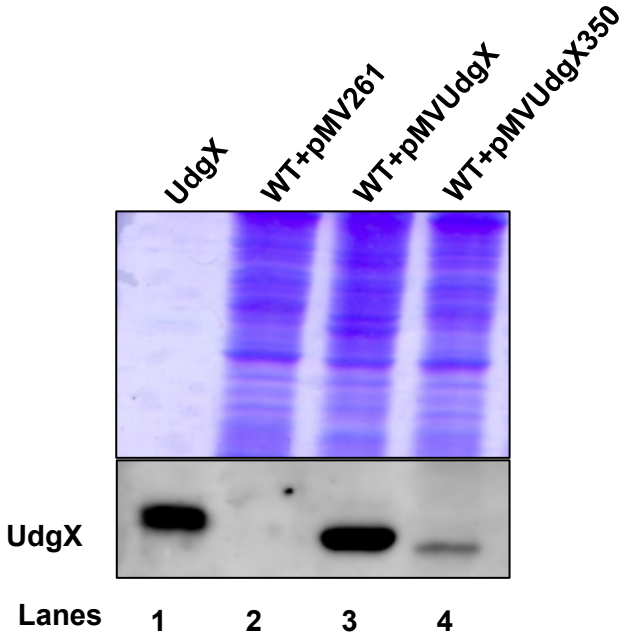

Figure S13. Immunoblotting of *M. smegmatis* cell free extracts harbouring different plasmid. Cell free extract (~30 µg) and 6X His-tagged *Msm*UdgX (100 ng) were resolved on SDS PAGE (15 %), transferred to a polyvinylidene difluoride (PVDF) membrane and probe with 1:2000 diluted anti-UdgX primary antibody followed by 1:2000 dilution of anti-IgG conjugated with HRP. HRP substrate (luminol) and H_2O_2 were added in 1:1 ratio. Chemiluminesence was captured using a ChemiDoc (Note: pMVUdgX has the *udgX* gene expressed under *hsp60* promoter and pMVUdgX350 *udgX* gene expressed under its native promoter). Marker UdgX migrates slower as it has an extra sequence of 6His on its N-terminal. Commassie blue stained gel is for the loading control.

Sang et. al. Fig. S14

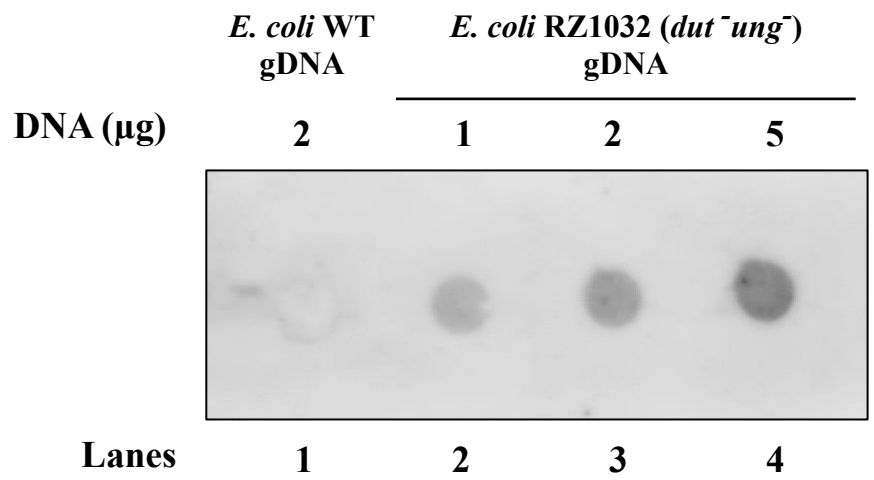

Figure S14. Detection of uracil from genomic DNA (gDNA) of *E. coli* using *Msm*UdgX as a probe. Different concentrated of genomic DNA was spotted onto a Nylon membrane, probed with *Msm*UdgX and detected by immunoblotting using anti-*Msm*UdgX antibodies (see supplementary text for materials and methods). Signal is seen for DNA samples from *E. coli* RZ1032 (known to harbor uracil in DNA) but not in *E. coli* wild type for *ung* and *dut*.