DNA Repair 11 (2012) 714–725

Contents lists available at SciVerse ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

# Structural and biochemical studies of a plant formamidopyrimidine-DNA glycosylase reveal why eukaryotic Fpg glycosylases do not excise 8-oxoguanine

Stéphanie Duclos[a], Pierre Aller[a], Pawel Jaruga[b], Miral Dizdaroglu[b], Susan S. Wallace[a,**], Sylvie Doublié[a,*]

[a] Department of Microbiology and Molecular Genetics, The Markey Center for Molecular Genetics, University of Vermont, Stafford Hall, 95 Carrigan Drive, Burlington, VT 05405-0068, USA
[b] Biochemical Science Division, National Institute of Standards and Technology, Building 227/A243, Gaithersburg, MD 20899, USA

ARTICLE INFO

*Article history:*
Received 7 May 2012
Received in revised form 15 June 2012
Accepted 15 June 2012
Available online 11 July 2012

*Keywords:*
Base excision repair
DNA glycosylase
Abasic site
8-oxoguanine
Fpg
MutM

ABSTRACT

Formamidopyrimidine-DNA glycosylase (Fpg; MutM) is a DNA repair enzyme widely distributed in bacteria. Fpg recognizes and excises oxidatively modified purines, 4,6-diamino-5-formamidopyrimidine, 2,6-diamino-4-hydroxy-5-formamidopyrimidine and 8-oxoguanine (8-oxoG), with similar excision kinetics. It exhibits some lesser activity toward 8-oxoadenine. Fpg enzymes are also present in some plant and fungal species. The eukaryotic Fpg homologs exhibit little or no activity on DNA containing 8-oxoG, but they recognize and process its oxidation products, guanidinohydantoin (Gh) and spiroiminohydantoin (Sp). To date, several structures of bacterial Fpg enzymes unliganded or in complex with DNA containing a damaged base have been published but there is no structure of a eukaryotic Fpg. Here we describe the first crystal structure of a plant Fpg, *Arabidopsis thaliana* (AthFpg), unliganded and bound to DNA containing an abasic site analog, tetrahydrofuran (THF). Although AthFpg shares a common architecture with other Fpg glycosylases, it harbors a zincless finger, previously described in a subset of Nei enzymes, such as human NEIL1 and Mimivirus Nei1. Importantly the "αF–β9/10 loop" capping 8-oxoG in the active site of bacterial Fpg is very short in AthFpg. Deletion of a segment encompassing residues 213–229 in *Escherichia coli* Fpg (EcoFpg) and corresponding to the "αF–β9/10 loop" does not affect the recognition and removal of oxidatively damaged DNA base lesions, with the exception of 8-oxoG. Although the exact role of the loop remains to be further explored, it is now clear that this protein segment is specific to the processing of 8-oxoG.



## 1. Introduction

Oxidatively induced DNA damage is the unavoidable consequence of the effect of reactive oxygen species (ROS) on cellular DNA. ROS are common damaging agents predominantly generated by endogenous factors like respiration or photosynthesis. Unrepaired oxidatively induced DNA lesions can lead to severe consequences including apoptosis, mutations, and cancer. Hence, to maintain genomic stability, cells have developed mechanisms to detect and repair DNA damage. The base excision repair (BER) system, a highly conserved repair pathway, is responsible for the repair of the vast majority of oxidative lesions in living organisms across phyla [1–6]. Identification and removal of such lesions are catalyzed by lesion-specific DNA glycosylases that are grouped into two families, based on structural motifs: the "Helix-hairpin-Helix" (HhH) family, named after the HhH structural motif involved in DNA binding, and the Fpg/Nei family, named after the prototypical bacterial members formamidopyrimidine DNA glycosylase (Fpg) and endonuclease eight (Nei) [2,7–10]. Although the DNA glycosylases that specialize in the repair of oxidatively induced lesions fall into two distinct structural families, they all employ a common mechanism involving the extrusion of the target base into the active site pocket.

Among the wide variety of oxidative DNA modifications, one of the most common is the oxidation product of guanine, 8-oxoguanine (8-oxoG), which, if unrepaired, can lead to GC → TA transversion mutations [11]. In archaea, fungi and animals, the 8-oxoguanine glycosylase (Ogg) is responsible for the excision of 8-oxoG and belongs to the "HhH" family, whereas in bacteria, 8-oxoG is predominantly removed by Fpg. Remarkably, plants have an Ogg as well as variants of Fpg resulting from alternative splicing. Although the conservation of both Ogg and Fpg appears to be redundant, the characterization of Fpg1 from *Arabidopsis thaliana* (AthFpg1) clearly shows that 8-oxoG is a

* Corresponding author. Tel.: +1 802 656 9531.
** Corresponding author. Tel.: +1 802 656 2164.
*E-mail addresses:* swallace@uvm.edu (S.S. Wallace), sdoublie@uvm.edu (S. Doublié).

1568-7864/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dnarep.2012.06.004

poor substrate for the plant Fpg; however, it efficiently processes 4,6-diamino-5-formamidopyrimidine (FapyA), 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG) and the further oxidation products of 8-oxoG, guanidinohydantoin (Gh) and spiroiminodihydantoin (Sp1 and Sp2) [12,13]. Like other glycosylases recognizing oxidatively induced DNA lesions, AthFpg1 is bifunctional and therefore catalyzes the hydrolysis of the N-glycosylic bond between the damaged base and the deoxyribose (glycosylase activity) before cleaving the DNA backbone (lyase activity).

Here we report the crystal structures of a functional C-terminal deletion form of AthFpg1 (AthFpgΔ88) unliganded and in complex with duplex DNA containing a non-hydrolyzable analog of an abasic site, tetrahydrofuran (THF). The structure of AthFpgΔ88 is similar to that of other members of the Fpg/Nei family and, as expected from the absence of zinc-binding cysteine residues, it harbors a zincless finger as observed in human NEIL1 and Mimivirus Nei1 (MvNei1) [14,15]. However, the most striking difference between AthFpg and the bacterial Fpg proteins is the absence of the αF–β9/10 loop, previously shown to be required for 8-oxoG recognition [16]. We further show here that the αF–β9/10 loop is not only essential for removing 8-oxoG but it is specific for this lesion.

## 2. Material and methods

### 2.1. Protein expression and purification

Two C-terminal deletion constructs of AthFpg1 were designed based on a prediction of disordered regions (PONDR®, [17]) and an alignment of Fpg/Nei sequences. First, a deletion of 109 C-terminal residues was generated by PCR amplification using primers 5′-GGAATTCCATATGCCGGAGCTTCCAGAG-3′ and 5′-GCTTGTCGACGCATTTCCCATAAAGTTTCTGCAGTTCTGG-3′ and AthFpg1 in a pET28b vector, kindly provided by Dr. Terrence Murphy (University of California, Davis) and subsequently cloned into NdeI-SalI sites of the pET22b expression vector (Novagen). A hexa-histidine tag was added to the C-terminus of the gene by mutating the TGA stop codon into a TGC, coding for a cysteine, resulting in the insertion of a sequence of ten residues, CVDKLAAALE, between the coding sequence and the hexa-His tag. A second truncated protein, lacking the 88 C-terminal residues, was engineered after PCR amplification of the corresponding coding region of the AthFpg1 gene using primers 5′-GGAATTCCATATGCCGGAGCTTCCAGAG-3′ and 5′-CCGCTCGAGATCGTCTTCTTTGGGTTTCAC-3′. The amplified fragment was digested with NdeI and XhoI restriction enzymes and cloned into the NdeI-XhoI sites of the pET22b expression vector, adding a hexa-histidine tag at the C-terminus of the gene. The recombinant AthFpgΔ109 and AthFpgΔ88 were expressed as described previously [12]. Briefly, the proteins were expressed in Rosetta (DE3) pLysS *Escherichia coli* cells (Novagen), induced with 1 mM IPTG and grown overnight at 16 °C. The cell pellet was resuspended in 50 mM sodium phosphate pH 8.0, 150 mM NaCl, 10 mM imidazole pH 8.0, 10% (v/v) glycerol, 5 mM βME, 1 mM PMSF, and 10 mM benzamidine and sonicated. Cleared cell lysate was loaded on a HiTrap chelating HP column (GE healthcare) charged with nickel sulfate and the proteins were eluted with a linear gradient of imidazole (10–500 mM in 20 column volumes). The pooled fractions of protein were dialyzed in 20 mM HEPES-NaOH, pH 7.6, 150 mM NaCl, 10% (v/v) glycerol, 5 mM βME and loaded on a HiTrap SP FF column (GE healthcare). A linear NaCl gradient (150 mM–1 M in 20 CV) was used for elution and the protein fractions were pooled and dialyzed in a crystallization buffer (20 mM HEPES-NaOH pH 7.6, 150 mM NaCl, 1 mM DTT). The selenomethionyl variant of AthFpgΔ109 was prepared by inhibiting methionine biosynthesis [18] and purified as described above, except for the final storage buffer, which was 50 mM Tris–HCl pH 8.0, 100 mM NaCl, 10% (v/v) glycerol, and 1 mM DTT. Proteins were concentrated up to 30 mg/ml (Millipore Amicon Ultra -4), flash frozen and stored at −80 °C.

The QuikChange Lightning site-directed mutagenesis kit (Stratagene) was used to introduce the WI187-188IY double mutation into AthFpgΔ88 in the pET22 (Novagen) expression vector. The resulting variant was sequenced in its entirety. The AthFpgΔ88 WI187-188IY variant protein was overexpressed and purified as described above.

EcoFpg and EcoFpg IY170-171WI were overexpressed and purified as previously described [19]. The double mutant EcoFpg IY170-171WI was obtained using the QuikChange Lightning site-directed mutagenesis kit (Stratagene). The resulting variant was sequenced in its entirety.

The (Δ213–229) deletion construct of *E. coli* Fpg (EcoFpgΔ213–229) was generated using an overlapping-extension PCR strategy, which involves two sequential PCR steps [20]. Briefly, the two EcoFpg fragments flanking the region to be deleted are amplified via two separate PCR reactions. An overlapping region of eighteen nucleotides at the junction of the two PCR products was created using a chimeric primer in each PCR reaction. The two amplified fragments were subsequently used as a template in a second PCR reaction that is performed using primers framing the EcoFpg gene. The resulting amplified sequence was subcloned using NdeI and XhoI restriction sites of a pET-22b vector (Novagen). Overexpression and purification of EcoFpgΔ213–229 were performed according to the protocol used for wild-type EcoFpg.

Multiple sequence alignments were generated by COBALT ([21], included in the NCBI C++ toolkit) using all fungal and plant Fpg sequences identified by BLAST [22].

### 2.2. Substrates

The 35-mer oligodeoxynucleotides used for the glycosylase and lyase activity assays were purchased from Midland Certified Reagent Co (Midland, TX) and purified by urea PAGE. The sequence of the damage-containing strand was 5′-TGTCAATAGCAAG**(X)**GGAGAAGTCAATCGTGAGTCT-3′, where X was either 8-oxoG or a uracil, which was used to create an apurinic/apyrimidic site (AP site). The complementary oligonucleotide sequence was the following: 5′-AGACTCACGATTGACTTCTCC(C)CTTGCTATTGACA-3′, in which (C) denotes the cytosine opposite the damaged base. Guanidinohydantoin (Gh), spiroiminodihydantoin 1 (Sp1) and 2,6-diamino-4-hydroxy-5*N*-methyl-formamidopyrimidine (MeFapyG) were synthesized as described previously [23,24] in the following sequence context: 5′-TGTTCATCATGCGTC**(Y)**TCGGTATATCCCAT-3′, Y being either Gh or Sp1, and 5′-TCATCATGCGTC-**(MeFapyG)**TCGGTATATCC-3′. The complementary strands for Gh/Sp1- and MeFapyG-containing DNA were 5′-ATGGGATATACCGA(C)GACGCATGATGAACA-3′ and 5′-GGATATACCGA(C)GACGCATGATGA-3′, respectively. The end labeling of substrates was carried out on 1 pmole of each damage-containing strand using T4 polynucleotide kinase (New England Biolabs, Beverly, MA) in presence of [α-$^{32}$P] dATP, for 30 min at 37 °C. The phosphorylation reaction was terminated by addition of 1 mM EDTA and heat inactivation. The end-labeled DNA was separated from the [α-$^{32}$P] dATP by ethanol precipitation and diluted in 9 pmoles of the appropriate non-labeled damage-containing oligodeoxynucleotide and 10 pmoles of the complementary oligodeoxynucleotide to a final concentration of 250 nM in10 mM Tris–HCl (pH 8.0) and 50 mM NaCl.

In order to create an AP site, a double-stranded uracil-containing oligodeoxynucleotide was treated with 2 units of uracil DNA Glycosylase (New England Biolabs) for 30 min at 37 °C.

### 2.3. Glycosylase/lyase activity assays

Glycosylase/lyase assays were performed under single-turnover conditions with 10 nM lesion of an AP:G substrate and 50 nM active enzyme in 20 mM Tris–HCl, 75 mM NaCl, 1 mM EDTA (pH 8.0) with 100 μg/ml BSA at 37 °C for EcoFpg and its variants (EcoFpg IY170-171WI and EcoFpgΔ213–229) or 10 mM Bis Tris Propane-HCl, 10 mM $MgCl_2$, 1 mM DTT (pH 7.0) [New England Biolabs NEBuffer 1] with 100 μg/ml BSA at room temperature for AthFpgΔ88 and AthFpgΔ88 WI187-188IY. At 1, 5, 10 and 15 min incubation, 10 μl aliquots were taken out and the reaction stopped by adding an equal volume of formamide stop loading buffer (98% formamide, 5 mM EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue). Lyase activity assays using 10 nM abasic site and 50 nM EcoFpg or EcoFpgΔ213–229 were stopped after 30, 60, 150 and 300 s incubation. The reaction products were separated from the uncleaved substrates with a 12% (w/v) denaturing polyacrylamide gel and quantified with an isotope imaging system (Molecular Imaging System, Bio-Rad).

### 2.4. Single-turnover kinetics

A rapid-quench flow instrument (RQF-3, KinTek, State College, PA) was used to determine the $k_{obs}$ of AthFpgΔ88 and EcoFpg, as well as their respective variants AthFpgΔ88 WI187-188IY and EcoFpg IY170-171WI, on an AP:G oligodeoxynucleotide substrate. Each reaction was carried out using 30 nM of active enzyme in presence of 3 nM of substrate in the appropriate reaction buffer. Reactions were terminated at selected time points (0.006, 0.025, 0.05, 0.1, 0.2, 0.5, 1, 2 and 5 s) with formamide stop loading buffer supplemented with 0.2 N NaOH. Kinetic parameters were determined using Prism 4 (GraphPad Software, Inc., La Jolla CA).

### 2.5. Analysis by gas chromatography/isotope dilution mass spectrometry (GC/MS)

DNA base lesions were introduced in calf thymus DNA via γ-irradiation with 40 Gray in $N_2O$-saturated phosphate buffer solution as previously described [25,26]. 50 μg of γ-irradiated DNA was incubated with 1 μg of active wild-type EcoFpg or EcoFpgΔ213–229 variant in 50 mM phosphate buffer (pH 7.4) containing 100 mM KCl, 1 mM EDTA and 0.1 mM DTT at 37 °C for 1 h. The released damaged bases were identified and quantified by GC/MS analysis, performed as previously described [25–27]. Data were collected from three independent experiments using three independently prepared DNA substrates.

### 2.6. DNA substrates for crystallography

The oligodeoxynucleotides used in this study were purchased from Operon Technology and purified by PAGE. A 16 mer tetrahydrofuran-containing oligodeoxynucleotide 5′-AGCGTCCAXGTCTACC-3′, where X is THF, and a 15 mer complementary oligodeoxynucleotide 5′-GCAGGTGCAGATGGT-3′ were annealed in a 1:1 ratio to generate a sticky-ended DNA duplex. The protein/DNA complex was produced by mixing 3 mg/mL (88 μM) of AthFpgΔ88 with a 1.1 molar excess of DNA duplex.

### 2.7. Crystallization

All crystals were obtained by the hanging-drop vapor diffusion method. The AthFpgΔ109 protein was crystallized at 12 °C by mixing 1 μl of protein solution at 15 mg/ml with 1 μl of reservoir solution (24% (w/v) PEG 3350, 175 mM $MgCl_2$, and 100 mM Tris–HCl, pH 8.0). In order to improve the quality of the crystals, microseeding was performed using similar crystallization conditions with PEG 3350 reduced to 20%. The concentrations of PEG 3350 and glycerol in the drop were adjusted to 25% and 5%, respectively for cryoprotection. The crystals were harvested after a 5 min equilibration and flash-cooled in liquid nitrogen. Crystals of AthFpgΔ88 were obtained at 12 °C by mixing 1 μl of protein solution at 15 mg/ml and 1 μl of reservoir solution (32% 15/4EO/OH, 150 mM $(NH_4)_2SO_4$, 50 mM Bis–Tris pH 6.5) optimized from condition 57 of the Index HT screen (Hampton Research, Aliso Viejo, CA). The crystals were harvested directly from the drop and flash-cooled in liquid nitrogen. The AthFpgΔ88 in complex with THF-containing DNA crystallized at 4 °C after mixing 0.5 μl of protein/DNA complex solution with 0.5 μl of reservoir solution (20% (v/v) PEG 2000 MME, 100 mM KBr, and 100 mM sodium acetate pH 5.0). Glycerol was added to the drop to a final concentration of 15% (v/v) after which the crystals were harvested and flash cooled in liquid nitrogen.

### 2.8. Data collection

The selenomethionyl variant of AthFpgΔ109 protein crystallized in a primitive tetragonal space group ($P4_12_12$ or $P4_32_12$) and each unit cell contains 2 molecules per asymmetric unit. Multiple-wavelength anomalous diffraction (MAD) data were collected at three wavelengths corresponding to the peak, inflection, and high-energy remote wavelengths of the K-edge of selenium at beamline X29 (National Synchrotron Light Source, Brookhaven). The resulting 2.35 Å dataset was integrated with DENZO and scaled with SCALEPACK (Otwinowski and Minor, 1997). Data from a crystal of AthFpgΔ88 were collected to 1.7 Å resolution on our laboratory MAR345 image-plate detector (MarResearch, Hamburg) using a rotating anode RU-200 X-ray generator (Rigaku) equipped with Xenocs mirrors. The resulting primitive orthorhombic ($P2_12_12_1$) dataset was processed with HKL2000 [28] and each asymmetric unit comprises one molecule. AthFpgΔ88 in complex with THF-containing DNA was crystallized in a primitive orthorhombic space group ($P2_12_12_1$) with 2 molecules per asymmetric unit. Diffraction data were collected to 2.8 Å at a wavelength of 1.54179 Å at the advanced light source (ALS) synchrotron by Reciprocal Space Consulting, LLC (beamline 5.0.3), integrated with Mosflm [29] and reduced with SCALA [30]. Data collection statistics are summarized in Table 1.

### 2.9. Phasing and refinement

Two out of four selenium sites were located by the program SOLVE [31] and the program AutoSHARP [32] identified the two remaining selenium sites. AutoSHARP refined the selenium positions and the space group was found to be $P4_12_12$ and not the enantiomorphic $P4_32_12$, based on map quality and continuity. The program RESOLVE [33] used the resulting phasing information for density modification and iterative model building resulting in 65% of the model being built. The remaining residues were built into the electron density map using the program COOT [34]. The final model was obtained through iterative cycles of refinement including energy minimization and B-factor refinement using CNS [35], alternating with manual model building in COOT. Water molecules were built in the residual electron density map using CNS and manually checked with COOT. Two molecules of glycerol from the crystallization buffer are positioned next to the N-terminal proline and next to the β4–β5 loop of molecule A.

The structure of AthFpgΔ88 was solved by molecular replacement with PHASER [36] using the AthFpgΔ109 coordinates devoid of all non-protein atoms as a starting model. Rigid-body minimization followed by energy minimization and B-factor refinement cycles were performed using CNS 1.2 [37]. Water molecules were positioned in Fo-Fc maps using CNS and COOT. One molecule of

**Table 1**
Data collection and refinement statistics.

| X-rays | Se-AthFpgΔ109 | | | AthFpgΔ88 | AthFpgΔ88/THF complex |
|---|---|---|---|---|---|
| | Se-Peak | Se-Inflection | Se-Remote | | |
| | Brookhaven (NSLS X29) | | | Home source | Berkeley (ALS) |
| Wavelength (Å) | 0.9792 | 0.9793 | 0.95 | 1.54179 | 1.000 |
| Space group | $P4_12_12$ | | | $P2_12_12_1$ | $P2_12_12_1$ |
| Unit-cell parameters (Å,°) | $a=b=93.054$, $c=157.7$, $\alpha=\beta=\gamma=90$ | | | $a=50.688$, $b=62.448$, $c=98.016$, $\alpha=\beta=\gamma=90$ | $a=46.4$, $b=98.57$, $c=181.48$, $\alpha=\beta=\gamma=90$ |
| Molecules per asymmetric unit | 2 | | | 1 | 2 |
| Data collection statistics | | | | | |
| Resolution (Å) | 50–2.33 (2.41–2.33) | 50–2.33 (2.41–2.33) | 50–2.26 (2.34–2.26) | 20–1.7 (1.76–1.7) | 20–2.8 (2.95–2.8) |
| Measured reflections | 410937 | 4129889 | 466028 | 277763 | 133571 |
| Unique reflections | 30578 | 30599 | 33650 | 34647 | 21150 |
| Redundancy | 13.5 (13.2) | 13.6 (13.1) | 13.9 (12.7) | 8.5 (4.9) | 6.3 (5.9) |
| Completeness (%) | 99.4 (95.1) | 99.5 (94.9) | 99.4 (93.6) | 94.7 (82.7) | 99.2 (97.6) |
| $R_{merge}$ (%) | 9 (41.4) | 11.2 (63) | 9.4 (74.8) | 7.9 (44.5) | 9.4 (53.8) |
| Overall I/σ(I) | 22.66 (9.2) | 21.218 (6.9) | 22.86 (5.49) | 21.06 (2.75) | 11.5 (2.6) |
| MAD phasing statistics | | | | | |
| Selenium sites | 4 | | | | |
| $R_{cullis}$ | 0.926 | 0.979 | 0.967 | | |
| Phasing power | 0.591 | 0.283 | 0.563 | | |
| Overall figure of merit | 0.281/0.794 | | | | |
| Refinement statistics | | | | | |
| $R_{work}$ (%) | 22.23 | | | 20.97 | 22.50 |
| $R_{free}$ (%) | 27.32 | | | 21.88 | 28.69 |
| r.m.s.d bond distance (Å) | 0.0065 | | | 0.0050 | 0.0069 |
| r.m.s.d. bond angle (°) | 1.30 | | | 1.32 | 1.29 |
| *B*-factor (Å$^2$) | | | | | |
| Wilson *B*-factor (Å) | 38.6 | 40.5 | 45.1 | 29.8 | 82.1 |
| Protein (A/B) | 28.48/55.47 | | | 27.31 | 62.53/67.88 |
| Water | 34.76 | | | 34.25 | 59.75 |
| DNA (A/B) | | | | | 85.76/81.43 |
| Ramachandran plot | | | | | |
| Most favored (%) | 86.9 | | | 92.5 | 84.6 |
| Additional allowed (%) | 13.1 | | | 7.5 | 15.4 |
| Generously allowed (%) | 0 | | | 0 | 0 |
| Disallowed (%) | 0 | | | 0 | 0 |
| r.m.s.d. vs. AthFpgΔ109 (Å) | | | | 0.54 | 0.88 |

glycerol from the cryoprotectant buffer is located next to the N-terminal proline. Residues Glu 6, Thr 51, Ser 141, Met 155, Glu 162, Gln 195, Lys 209, Ser 218, and Asp 260 appear to adopt alternate conformations.

The structure of AthFpgΔ88 in complex with THF-containing DNA was solved by molecular replacement with PHASER using the unliganded AthFpgΔ88 coordinates devoid of all non-protein atoms as a starting model. Rigid-body minimization, refinement procedure, and water molecule addition were performed using CNS, as described above for the unliganded AthFpgΔ88 structure. A final round of refinement was performed with REFMAC5 using the TLS option [38]. The quality of the models for all three structures was evaluated using PROCHECK [39]. All non-glycine residues are found in the most favored and additionally allowed regions of the Ramachandran plot. RMS deviations were calculated with SSM Superposition [40] and LSQ Superpose [34]. DNA bend angles were calculated with CURVES, [41]. All structural figures were made with The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC [42].

### 2.10. Protein data bank accession codes

Atomic coordinates and structure factor amplitudes have been deposited with the Protein data bank (http://www.pdb.org) and are accessible under accession codes 3TWK (AthFpgΔ109), 3TWL (AthFpgΔ88) and 3TWM (AthFpgΔ88 in complex with THF-containing DNA).

## 3. Results

### 3.1. Structure determination

We were interested in solving the crystal structure of *A. thaliana* Fpg1 (AthFpg1) because of its unusually low activity towards 8-oxoG, a preferred substrate for other Fpg enzymes [12,43]. Analysis of the protein sequence indicated that the C-terminal region of AthFpg1 was likely to be disordered [17], and would likely represent an impediment to crystallization. A segment encompassing 109 C-terminal residues of AthFpg1 was deleted based on published work of Ohtsubo and colleagues [44], who showed that these residues are dispensable for enzymatic activity. The structure of AthFpgΔ109 was solved by multiple-wavelength anomalous diffraction at a resolution of 2.35 Å, using a selenomethionyl variant of the protein [45]. The resulting model was refined to an $R_{free}$ of 0.273 and an $R_{Work}$ of 0.222. The primitive tetragonal crystal comprises two monomers of AthFpgΔ109 per asymmetric unit, which happen to be covalently linked via two cysteines (one in each monomer) located in a ten amino-acid segment introduced by cloning between the C-terminus of the truncated gene and the hexa-His tag. The serendipitous disulfide bridge may be responsible for the stabilization of AthFpgΔ109 in solution and may have facilitated crystallization. Monomer B, which comprises residues 2–40 and 50–296, is less ordered than monomer A (residues 2–293), with a higher average B-factor (28.48/55.47 Å$^2$ for molecule A vs. molecule B).

Although the truncated form of AthFpg1 was reported to be active [44], it is not as active as the wild-type enzyme (data not shown) and we were unable to obtain crystals in complex with a DNA substrate. In order to produce a soluble and active enzyme, we designed a new construct with a shorter truncation of the C-terminus, based on sequence similarity with human NEIL1 (hNEIL1) [14]. For hNEIL1, the expression of a soluble protein is highly compromised when the entire disordered region (C-terminal 106 amino acids) is deleted whereas deletion of a shorter fragment (C-terminal 56 amino acids) has no deleterious impact on the solubility or the activity of the enzyme [46]. Although these disordered regions do not show any sequence homology, in both cases the twenty to thirty residues following the zincless finger are mostly positively charged with a high arginine and lysine content. Therefore, we surmised that this sequence might be important for the stability of the protein and/or DNA binding and designed a longer construct, this time lacking 88 C-terminal residues and hexa-His tagged at the C-terminus. The resulting protein, AthFpgΔ88, crystallized in a primitive orthorhombic space group ($P2_12_12_1$) with one monomer per asymmetric unit and its structure was solved by molecular replacement using the previously solved structure of AthFpgΔ109 as a starting model. The final model, comprising residues 2–86, 101–261 and 268–286, was refined to an $R_{free}$ of 0.219 and $R_{Work}$ of 0.210 at 1.7 Å resolution.

Crystals of AthFpgΔ88 in complex with a DNA containing a stable abasic site analog (tetrahydrofuran, THF) were obtained with a sticky-ended DNA duplex composed of a 16-mer THF-containing DNA strand annealed with a 15-mer complementary DNA strand, with G opposite THF. We determined the structure of this complex by molecular replacement using the structure of the AthFpgΔ88 as a search model. The asymmetric unit of the crystal is composed of two AthFpgΔ88 monomers (residues 2–278) each bound to DNA. The final model was refined to an $R_{free}$ of 0.225 and $R_{Work}$ of 0.287 at 2.8 Å resolution.

### 3.2. Overall structure of unliganded AthFpg

The crystallization of AthFpgΔ109 in a primitive tetragonal space group provided two structures of the unliganded protein, linked together by a disulfide bridge. A third structure of the unliganded protein was obtained from a primitive orthorhombic crystal of AthFpgΔ88. The three structures are nearly identical, as indicated by a low root-mean-square deviation (r.m.s.d.) of 0.5–0.7 Å (calculated on the first 280Cα atoms) (See Table 1). Each monomer measures approximately 60 × 38 × 30 Å and shares a common fold with Fpg/Nei enzymes of known structure. AthFpgΔ88 (residues 2–278) consists of two compact domains linked by a flexible linker (Fig. 1). The N-terminal domain is composed of two α-helices flanking a two-layered β-sandwich, each layer comprising four antiparallel β-strands. The C-terminal domain comprises four α-helices, two of which form the helix-two-turn-helix (H2TH) motif, and two antiparallel β-strands involved in a structural motif mimicking an antiparallel β-hairpin zinc finger (see below). The twenty-four C-terminal residues of the AthFpgΔ88 construct, including the C-terminal hexa-His tag, were not built because of disorder beyond residue 286. In contrast, the C-terminus of the AthFpgΔ109 form is well ordered (Supplemental Figure 1A) and the additional sixteen residues, including the C-terminal hexa-His tag, are organized in an α-helix. The serendipitous intermolecular disulfide bridge at position 281 of each monomer appears to be a key factor in the stability and order of the C-terminal region of the AthFpgΔ109 construct.

The absence of canonical zinc-binding cysteine residues in the region following the H2TH motif suggested that AthFpg1 does not have a DNA-binding zinc finger motif. Indeed, no coordination of a zinc atom was observed in the AthFpg structure but it does harbor

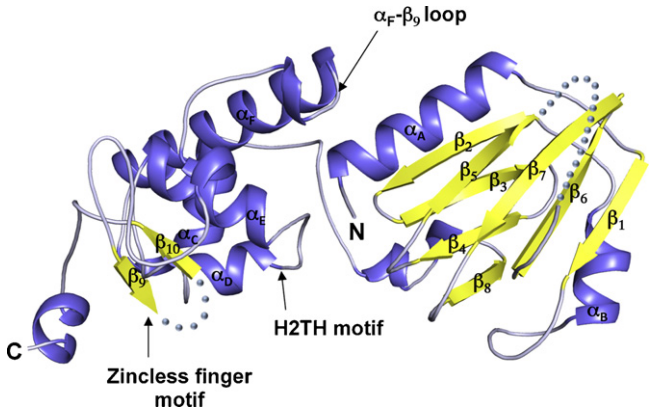


**Fig. 1.** Ribbon diagram of unliganded AthFpgΔ88. Helices are shown in blue and β-strands in yellow. Spheres represent missing residues. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

a structural motif, formed by two antiparallel β-strands (β9 and β10), mimicking an antiparallel β-hairpin zinc finger, as described for hNEIL1 and MvNei1 [14] [15]. The two antiparallel β-strands of the motif superimpose well onto those of the zinc finger motif observed in *Bacillus stearothermophilus* Fpg (BstFpg) [47] as well as those of the zincless finger motifs of hNEIL1 and MvNei1 (Supplemental Figure 1B). Moreover, this zincless finger motif harbors an Arg residue (Arg 267), which is conserved among Fpg/Nei family members (Arg 277 in hNEIL1 and MvNei1, and Arg 264 in BstFpg) and interacts with the DNA backbone [14,15,47]. The loops flanking the two β strands are stabilized by a network of seven main-chain H-bond interactions involving residues 252, 254, 256 and 257 in the loop preceding β-strand 9 and residues 272, 274–276 in the loop following β-strand 10.

Interestingly, the loop connecting β strands 9 and 10 of the zincless finger motif of the unliganded form of AthFpgΔ88 is disordered whereas the whole zincless finger motif of the unliganded AthFpgΔ109 form is ordered, thanks to the proximity of the C-terminal domain of the other monomer (Fig. 1 and S1). The loop is also ordered in the furan-bound form of AthFpgΔ88 (Fig. 2A) (See below).

### 3.3. AthFpg complex with DNA containing an abasic site analog

The crystals of AthFpgΔ88 in complex with THF contain two nearly identical complexes, A and B, per asymmetric unit and the following description applies to complex A (See Table 1). The superposition of the AthFpgΔ88-THF complex structures onto the unliganded AthFpgΔ88 structure shows that no conformational change occurs after DNA binding (r.m.s.d. of Cα atoms are 0.67 and 0.62 Å), in contrast with the open and closed conformations described for the free and liganded forms of the EcoNei [48,49]. In that respect AthFpg is similar to other Fpg/Nei glycosylases (LlaFpg; [50] *B. stearothermophilus* Fpg; [51]), hNEIL1 [14] and MvNei1 [15] [52], which do not undergo a conformational change upon binding DNA. Although the free and bound forms of AthFpgΔ88 are observed in a "closed" conformation, binding to THF-containing DNA does induce small-scale changes involving few amino acid side chains in the vicinity of the lesion (see below).

In the AthFpgΔ88-DNA complex, the DNA lies in the cleft defined by the N- and C-terminal domains of the enzyme, along an axis roughly perpendicular to the long axis of AthFpgΔ88 (Fig. 2A). The DNA duplex is mostly B-form, except in the vicinity of the abasic site analog, which lies in an extra-helical conformation. As a result of the extensive contacts between AthFpgΔ88 residues and the THF-containing DNA strand ($P^1$ to $P^{-2}$) (Fig. 2B), the DNA is

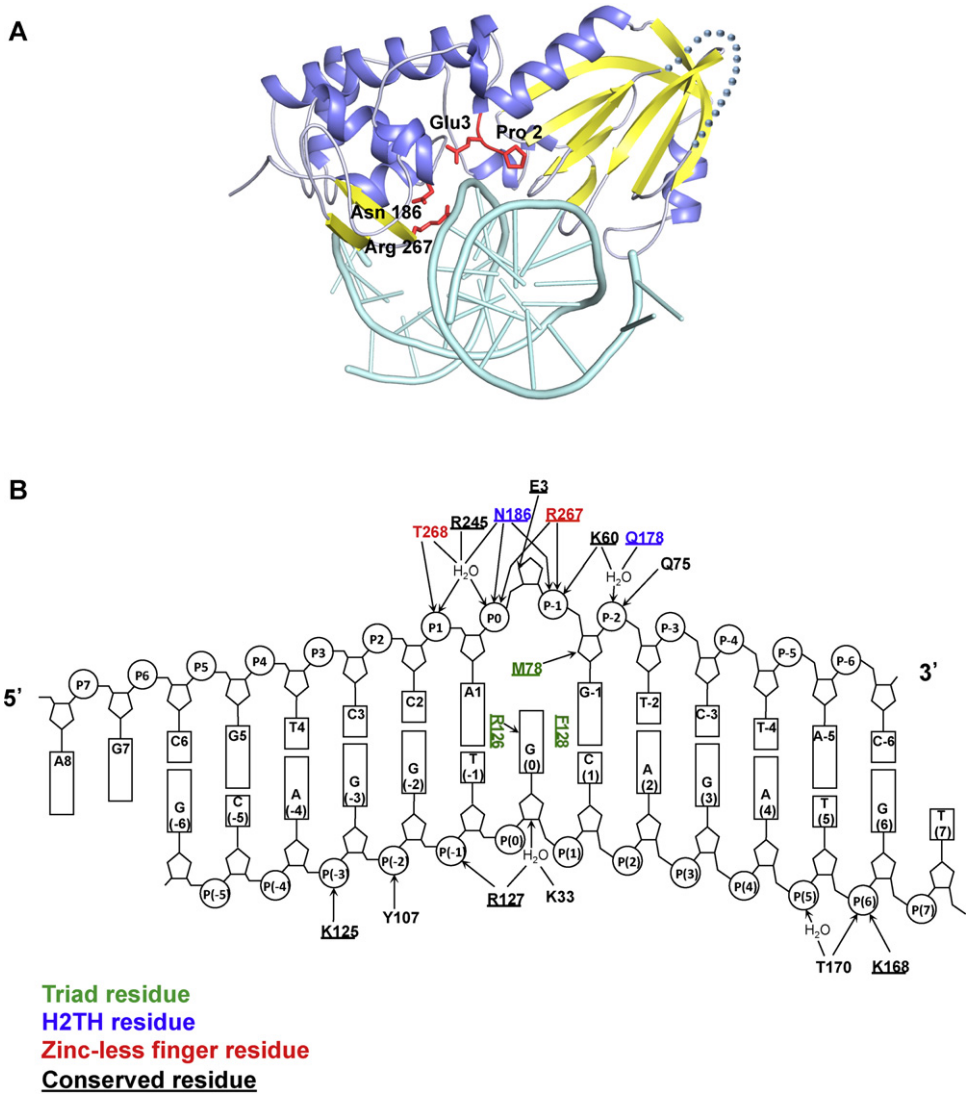


**Fig. 2.** (A) Ribbon diagram of the AthFpgΔ88/THF complex. The strictly conserved Arg (Arg 267) of the zincless finger, the conserved Asn (Asn 186) of the H2TH motif, the catalytic proline (Pro 2) and glutamate (Glu 3) are highlighted in red. Spheres indicate missing residues. (B) Schematic representation of AthFpgΔ88-DNA interactions. Position 0 and (0) are assigned to THF and its opposite base, respectively, and numbering is positive towards the 5′-end. The arrows represent hydrogen bonds and point towards the acceptor. $H_2O$ indicates water-mediated interactions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

kinked at the lesion site by an angle of ∼50° [41]. This observation is consistent with the 40° DNA bend previously reported for the complex of MvNei1 with THF-containing DNA [15] but not as pronounced as the 70° DNA bend measured for the same abasic site analog in complex with *Lactococcus lactis* Fpg [53].

As mentioned above, no conformational change was observed between the liganded and unliganded forms of AthFpgΔ88. However, DNA binding induces movements from amino acid side chains accommodating the substrate in the binding pocket or participating in DNA binding. The superposition of the bound form of AthFpgΔ88 and monomer A of AthFpgΔ109 (the only unliganded structure with an ordered zincless finger motif) shows that the tip of the zincless finger motif moves by about 5 Å upon binding DNA (Supplemental Figure 1A). A movement of the same magnitude was observed in MvNei1 when comparing the unliganded enzyme with a complex bound to furan-containing DNA [15].

The H2TH and zincless finger motifs, signature motifs of the Fpg/Nei family, are involved in DNA binding [49,50] along with residues distributed on three loops (β1–α3, β2–β3, and β4–β5) of the N-terminal domain (Fig. 2B). The majority of the residues interacting with the phosphates of the lesion-containing DNA strand (Fig. 2B) are conserved in the Fpg/Nei family. The conserved Arg 267 of the zincless finger motif participates in a hydrogen bond with phosphates $P^{-1}$ and $P^0$ in a manner similar to the analogous arginine in other Fpg/Nei enzymes with a canonical zinc finger motif [47,49,50,54] or zincless finger motif (MvNei1; [15]). The adjacent residue, Thr 268, interacts with phosphate $P^1$. The H2TH motif also plays a key role in DNA binding, especially the conserved Asn 186 (Asn 168 in EcoFpg), which is involved in hydrogen bonds with $P^{-1}$ and $P^0$ via its main chain and side chain, respectively. Lys 60 is a conserved residue from the N-terminal domain (Lys 57 in EcoFpg) involved in DNA binding as well as catalysis [47]. Site-directed mutagenesis of EcoFpg and hNEIL1 showed that this conserved lysine is essential for glycosylase activity but not lyase activity [55,56]. In AthFpgΔ88, the side chain of Lys 60 hydrogen bonds with phosphate $P^{-1}$ and interacts with phosphate $P^{-2}$ via a water molecule. On the 3′ side of the lesion, the side chains of three residues, Asn 186 of the H2TH, Arg 245 and Thr 268 of the zincless finger motif participate in water-mediated interactions with phosphates $P^0$ and $P^1$.

Unlike the lesion-containing DNA strand, in which only the four phosphates flanking the abasic site ($P^1$ to $P^{-2}$) interact with AthFpg,

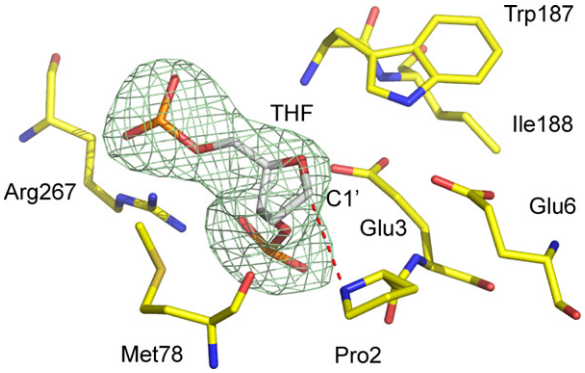


**Fig. 3.** Close-up view of the everted THF. AthFpgΔ88 residues are shown in yellow and the DNA in gray. A red dashed line indicates the distance between the amino group of Pro 2 and C1′ of THF. A simulated annealing omit map (green mesh) is contoured at 3σ. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

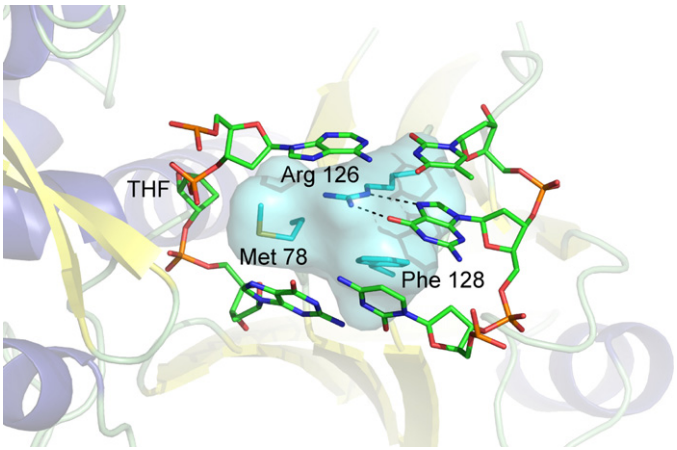


**Fig. 4.** Close-up view of the "void filling" triad. The residues forming the intercalation triad (Met 78, Arg 126 and Phe 128) are shown in cyan and the DNA is shown in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the contacts with the complementary DNA strand encompass a larger segment of the DNA molecule, from $P^{(-3)}$ to $P^{(6)}$ (Fig. 2B). Those contacts mainly involve residues distributed along two loops from the N-terminal domain (β1–αB loop and β7–β8 loop) and one loop from the C-terminal domain (αE–αF loop). The side chains of the N-terminal domain residues (Lys 33, Tyr 107, Lys 125, and Arg 127) participate in hydrogen bond and salt bridge interactions with phosphates $P^{(-3)}$ to $P^{(0)}$. On the 5′ side of the estranged base, phosphate $P^{(6)}$ is involved in hydrogen bonding with main-chain atoms of Lys 168 and Thr 170 as well as the side chain of Thr 170.

### 3.4. Interactions with THF and the opposite base

Although the DNA described in this study lacks a damaged base, we clearly observe the abasic site mimic in an extra-helical conformation (Fig. 3). The binding site surrounding THF is composed of the catalytic residues, Pro 2 and Glu 3, and three residues within van der Waals distance of the THF, Glu 6, Met 78, and Trp 188. No significant change is observed in the active site when it accommodates THF, with the exception of a shift of Pro 2 of about 0.8 Å, which positions the α-amino group of the proline within van der Waals distance (3.6 Å) of the C1′ of THF. The shift in the orientation of Pro 2 positions the α-amino group of Pro 2 for nucleophilic attack on the C1′ atom of THF, as described previously for Fpg and Nei [47,49]. The distances between THF and the catalytic residues are similar to those reported for the structure of MvNei1 protein in complex with furan [15] but shorter than those observed in the structure of LlaFpg in complex with the same abasic site analog [53]. The DNA backbone torsion angles at the abasic site of the AthFpgΔ88/THF complex are comparable to those observed for the MvNei1/THF complex (Supplemental Table 1).

We observe that, beside Pro 2, the side chains of three residues located close to the lesion site, Met 78, Arg 126 and Phe 128, shift after eversion of the lesion (Fig. 4). Arg 126 and Met 78 occupy the space left vacant in the DNA helix after eversion of THF. Met 78 lies within van der Waals distance of the extruded deoxyribose of the THF lesion. These three amino acids (Met 78, Arg 126 and Phe 128) are strictly conserved among Fpg family members (Met 74, Arg 109, and Phe 111 in EcoFpg) and are described as "void-filling" residues [54]. In EcoNei, three successive residues (Gln 69, Leu 70, and Tyr 71) play an equivalent intercalating role [49]. In the AthFpgΔ88/THF complex, Phe 128 is wedged between the orphaned base (guanine $G^{(0)}$) and cytosine $C^{(1)}$ on the 5′ side. Phe 114 of BstFpg was found in a similar position when complexed with undamaged DNA and was postulated to be a damage-sensing amino acid residue [57]. This hypothesis was supported by a recent single-molecule study in which the EcoFpg F111A variant was shown to lose the search mode exhibited by wild-type EcoFpg [58]. The side chain of the second "void-filling" residue, Arg 126, completes the stabilization by interacting with G via its Hoogsteen face (O6 and N7). The estranged G adopts a *syn* rotamer as observed in the BstFpg/rAb:G complex [47]. In addition, the guanidinium group of Arg 126 stacks against the 5′ neighbor (adenine $A^{1}$) of the damaged base. This arginine was proposed to be the primary molecular determinant for the specificity of Fpg for the base opposite the lesion [47,50].

### 3.5. The αF–β9/10 loop

Another structural characteristic of Fpg enzymes is the presence of a "recognition" loop, also named "αF–β9/10" loop [50,51], or the 8-oxoG capping loop [16]. It was previously postulated that this loop might be involved in the recognition of the lesion as well as the accommodation of the damaged base in the binding pocket [50,51,59]. In the structures of bacterial Fpg unliganded or in complex with an extra-helical damaged base, this loop is ordered and caps the binding site [51,59,60]. However, in the structures of Fpg in complex with abasic site analogs or covalently trapped with DNA, this loop appears to be disordered [47,50,53,54]. Although the αF–β9 loop is not required by BstFpg to catalyze the *β*-lyase reaction on abasic sites, it is essential for the excision of the 8-oxoG [16]. Although no significant sequence homology was found among bacterial Fpg proteins in the loop region, it seems to play the same role in all these enzymes by closing the active site and holding the substrate in its binding pocket. For AthFpg1, the sequence alignment predicted a much shorter loop, which was confirmed by the AthFpgΔ88 structures (Fig. 5). This results in a shallow and open recognition pocket, making it unclear how this enzyme recognizes its substrates.

In order to assess the importance of the loop with regard to substrate specificity, a 17 amino-acid region (residues 213–229) corresponding to the αF–β9 loop was deleted in EcoFpg. The glycosylase activity of this deletion variant (EcoFpgΔ213–229) was tested using four different substrates with C as the opposite base: 8-oxoG, MethylFapyG, guanidinohydantoin (Gh), and spiroiminodihydantoin (Sp1), along with dihydrouracil (DHU) opposite G (Fig. 6A and Supplemental Figure 2). The activity patterns confirm that the deletion variant fails to remove 8-oxoG but retains an efficient lyase activity (Fig. 6B), as shown by Verdine and colleagues [16]. However, our work further establishes that deletion

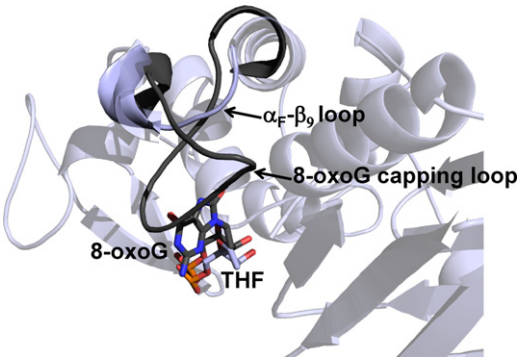


**Fig. 5.** Close-up view of the recognition loop. Superposition of the AthFpgΔ88/THF complex (light blue) with the αF–β10 loop of the BstFpg/8-oxoG complex (dark gray, PDB code 1R2Y, [51]). THF is shown in light blue and 8-oxoG in dark gray. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

of the loop does not affect the recognition and excision of other target lesions such as MeFapyG, Gh, Sp1, or DHU (Fig. 6A and Supplemental Figure 2). In fact the activity of the deletion variant on these lesions is the same as wild-type EcoFpg. In addition, a "pooled assay" was performed using both wild-type EcoFpg and the EcoFpgΔ213–229 variant in which the products released from γ-irradiated calf thymus DNA containing multiple lesions were quantified using gas chromatography/mass spectrometry (GC/MS). The results clearly show that the variant exhibits a substrate recognition pattern comparable to that of wild-type EcoFpg and efficiently excises FapyG, FapyA and 8-oxoA, with the only major difference being that it does not release 8-oxoG from the DNA (Fig. 6C).

### 3.6. The H2TH motif

As described earlier, conserved residues of the H2TH motif contact the backbone of the DNA strand containing the damaged base via hydrogen bonding interactions. In prokaryotic Fpg proteins, a conserved asparagine in the H2TH motif is followed by an isoleucine and a tyrosine (Ile 169/Tyr 170 in EcoFpg and Ile 175/Tyr 176 in BstFpg) (Supplemental Figure 3). This tyrosine is engaged in hydrogen bonding with the side chain of a conserved Ser/Thr in the αF–β9 loop (Thr 214 in EcoFpg and Ser 220 in BstFpg). Interestingly, in Nei enzymes the tyrosine of the H2TH motif switched positions with the isoleucine (Sequence is NYI compared to NIY in bacterial Fpg). The tyrosine in EcoNei, hNEIL1, and MvNei1 (Tyr 169, Tyr 177, and Tyr 174, respectively) does not contact any residues from the loop. Instead it interacts with Glu 6 via one hydrogen bond [14,15,49,52]. In plant and fungal Fpg proteins, a tryptophan replaces the tyrosine in the Nei motif (Sequence is thus NWI). In AthFpgΔ88 Glu 6 is similarly stabilized via a hydrogen bond with Trp 187 [distances for W187 (Nε1)-E5 (Oε2) are 3.6 Å and 3 Å with and without DNA, respectively]. Since bacterial and eukaryotic Fpg enzymes differ substantially in their substrate specificity and we observed differences in the residues immediately following the conserved asparagine in the H2TH motif, we set out to investigate if Trp-Ile vs. Ile-Tyr played a role in substrate recognition. We substituted the AthFpg Trp-Ile motif with the EcoFpg Ile-Tyr motif (WI187-188IY mutation) and made the reciprocal mutation in EcoFpg (IY169-170WI mutation).

Kinetic parameters ($k_{obs}$) for AthFpgΔ88 and EcoFpg and their respective variants (AthFpgΔ88 WI187-188IY and EcoFpg IY169-170WI) were first determined with an AP site opposite G (Table 2). All of the single-turnover kinetics fit a double exponential, likely a consequence of the fact that naturally-occuring AP sites can adopt open and closed ring conformations in solution. The kinetic rates

**Table 2**
Kinetic rates.

| Enzyme | $k_1$ (min$^{-1}$) | $k_2$ (min$^{-1}$) | $A_1$ | $A_2$ |
|---|---|---|---|---|
| AthFpgΔ88 | 2877 | 90 | 2.83 | 0.19 |
| AthFpgΔ88 WI187/188IY | 1266 | 46 | 2.68 | 0.24 |
| EcoFpg | 1061 | 124 | 1.46 | 1.45 |
| EcoFpg IY170/171WI | 457 | 88 | 1.05 | 1.73 |

Kinetic rates ($k_{obs}$) per minute of AthFpgΔ88, EcoFpg and their respective mutants, AthFpgΔ88 WI187IY and EcoFpg IY169WI with AP:G. $k_1$ and $k_2$ values are determined by single-turnover experiments and fit a double-exponential kinetics. The amplitudes for the fast and slow phases ($A_1$ and $A_2$) are also reported (in arbitrary units).

associated with the lyase activity of the variants were moderately affected by the mutation (2-fold decrease).

Each wild-type and variant enzyme was assayed for glycosylase activity using four different lesions (8-oxoG:C, MeFapyG:C, Gh:C, and Sp1:C) (Fig. 7). Surprisingly, AthFpgΔ88 WI187-188IY showed no significant decrease in activity compared to AthFpgΔ88 (Fig. 7A). In contrast, the reciprocal mutation in EcoFpg negatively impacted glycosylase activity on MeFapyG (Fig. 7B). The glycosylase activity of EcoFpg IY169-170WI with the other three substrates (8-oxoG, Gh, and Sp1) was only slightly decreased compared to wild-type EcoFpg.

## 4. Discussion

### 4.1. The structure of AthFpg is similar to that of bacterial Fpg or Nei glycosylases

The AthFpg amino acid sequence suggested that this enzyme would share substantial features with bacterial Fpg proteins such as the conserved catalytic residues (Pro 2 and Glu 3 and Lys 60, corresponding to Lys 57 in EcoFpg), the signature helix-two-turns-helix (H2TH) motif and the three "void-filling" residues (Met 78, Arg 126, and Phe 128) [12]. Indeed, the AthFpgΔ88 structures described in this study exhibit strong structural similarities with members of the Fpg/Nei family, such as TthFpg (PDB ID code 1EE8, rms deviation = 1.84 Å, 28.5% sequence identity [60]) and hNEIL1 (PDB ID code 1TDH, rms deviation = 2.06 Å, 27.5% sequence identity [14]). The structures of AthFpgΔ88 alone and in complex with the DNA substrate revealed that this enzyme does not undergo any large-scale conformational change upon DNA binding. Importantly, in contrast to the bacterial Fpg, AthFpg does not recognize 8-oxoG and prefers its oxidation products Gh, Sp1 and Sp2 as well as formamidopyrimidines, which makes the AthFpg substrate specificity closer to mammalian NEIL1 than to EcoFpg [12,13]. Furthermore, in the AthFpgΔ88 structures, the zinc-finger signature motif of bacterial Fpg is replaced with a zincless finger motif, as observed in NEIL1 and MvNei1 enzymes [14,15,52]. However, phylogenic studies showed that fungal and plant sequences of the Fpg/Nei family (including AthFpg) form a distinct clade closer to Fpg than Nei [12].

### 4.2. AthFpg harbors a truncated αF–β9 loop

As implied by the Fpg/Nei sequence alignment and illustrated by the AthFpgΔ88 structures, AthFpg harbors a very short αF–β9 loop, leaving the active site open and shallow. A recent study showed that the deletion or mutation of this loop is deleterious for the excision of 8-oxoG by bacterial BstFpg, but does not affect the lyase activity [16]. The authors described the damaged base detection as an intrahelical interrogation process, resulting in the complete extrusion of the damaged base from the DNA duplex. The encounter of Fpg with an intrahelical 8-oxoG lesion triggers the distortion and bending of the DNA, promoting the extrusion of the damaged base into the Fpg active site. These structural modifications of the

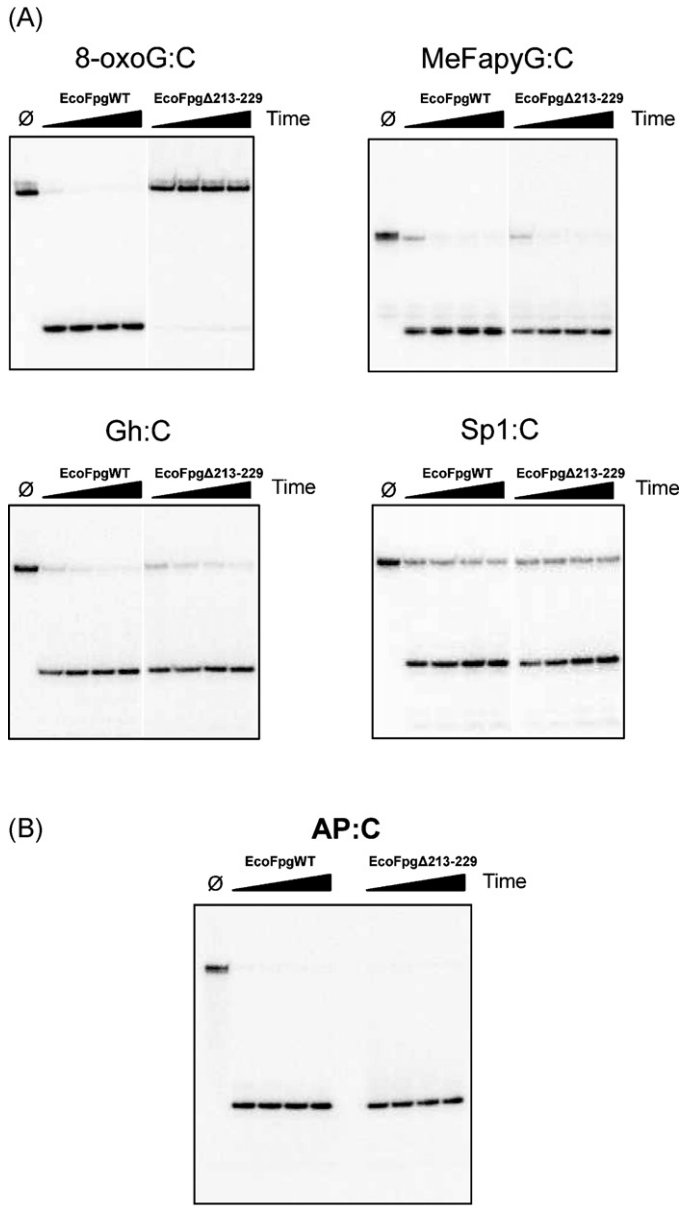


**Fig. 6.** Glycosylase/lyase activity assays and activity profile on γ-irradiated DNA of wild-type EcoFpg and EcoFpgΔ213–229. (A) The glycosylase assay was performed by incubating 10 nM of double-stranded substrate containing 8-oxoG:C, MeFapy:C, Gh:C, and Sp1:C with 50 nM of EcoFpgWT and EcoFpgΔ213–229. (B) The lyase assay was performed with a double-stranded substrate containing an abasic site opposite C. Each glycosylase reaction was incubated for 1, 5, 10, and 15 min at 37 °C, whereas lyase reactions were incubated for 0.5, 1, 2.5, and 5 min. (C) Activity profile of wild-type EcoFpg and EcoFpgΔ213–229 on γ-irradiated DNA. GC/MS analysis was used to identify and quantify the amount of damaged DNA bases released by each enzyme. The enzyme:substrate ratio was the same for each experiment.

DNA backbone coincide with the invasion of the double helix at the target site by the intercalating residues (Met 77, Arg 112, and Phe 114 in BstFpg). As a consequence of the full insertion of the lesion base in the catalytic pocket, the αF–β9 loop becomes ordered and engages in multiple main-chain contacts with 8-oxoG [61]. In addition to the initial intrahelical recognition of the damaged base, the excision process requires the ordering of the αF–β9 loop. A multiple-sequence alignment suggests that eukaryotic Fpg proteins have a very short loop, implying that these enzymes process their target lesions using a different mechanism.

### *4.3. Mutations in the H2TH motif do not significantly affect damage recognition*

In an attempt to identify potential residues in AthFpg responsible for its substrate specificity, we mutated residues in the active site. Based on the structure of AthFpgΔ88 in complex with THF, we targeted two residues of the H2TH motif, tryptophan 187 and isoleucine 188, located in the vicinity of THF and which differ from the conserved residues at the analogous positions of Nei and bacterial Fpg proteins. We observed that these two residues vary from one Fpg/Nei clade to another but are conserved within each clade. Fungal and plant Fpgs display an unusual conservation of a tryptophan immediately followed by an isoleucine (WI), at the positions corresponding to a tyrosine and an isoleucine in Nei (YI), or an isoleucine and a tyrosine (IY) in bacterial Fpg. We hypothesized that these differences may be of importance for substrate specificity, which varies among the different Fpg/Nei clades. Thus, we mutated Trp 187 and Ile 188 in AthFpgΔ88 to the corresponding Ile and Tyr in bacterial Fpg. Unexpectedly, the glycosylase activity assays performed on both AthFpgΔ88 and AthFpgΔ88 WI187-188IY showed that excision of Gh, Sp1 and MeFapyG opposite C is not altered by the mutation of the Trp-Ile motif (Fig. 7). Similarly, the reverse mutation in EcoFpg, EcoFpg IY170-171WI, does not significantly

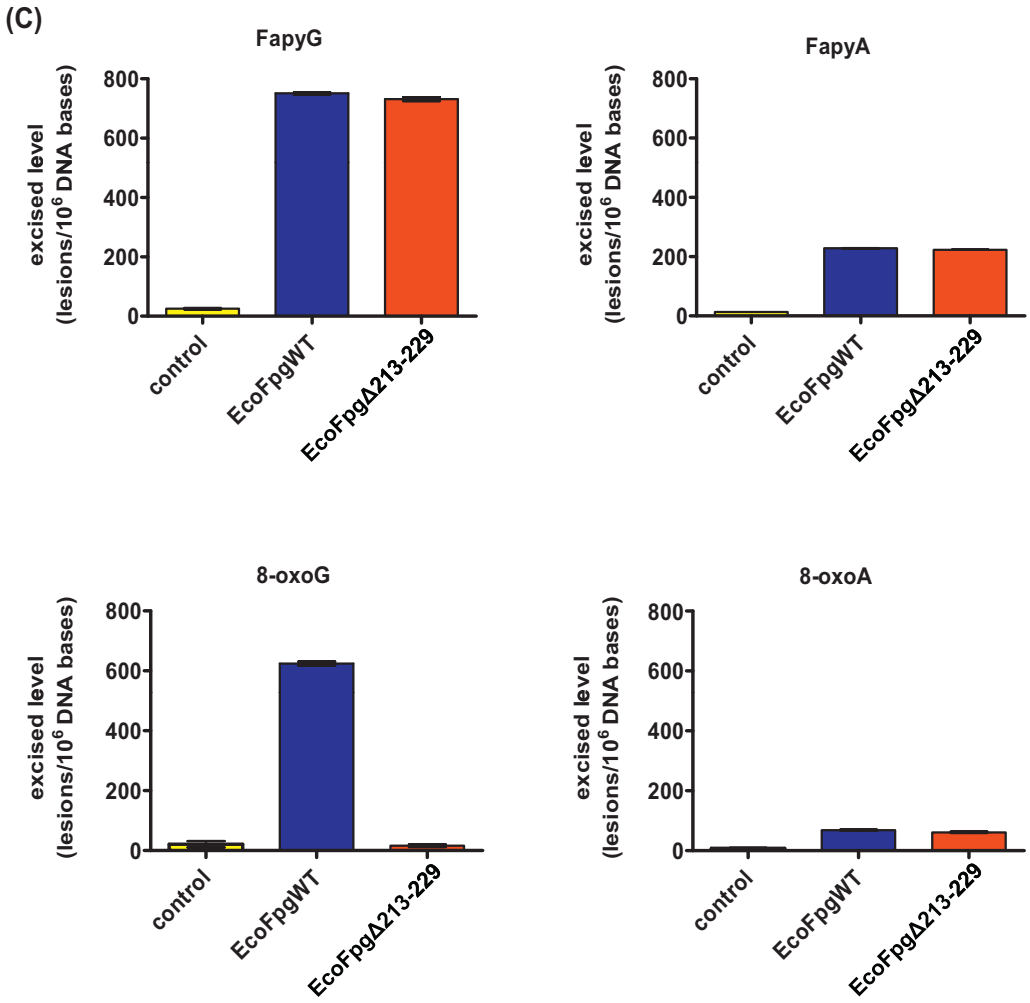


**Fig. 6.** (*Continued*).

reduce the glycosylase activity on Gh, Sp1 and MeFapyG. Moreover, similar to wild-type AthFpgΔ88, the WI187-188IY variant is still unable to process 8-oxoG opposite C. EcoFpg IY170-171WI, on the other hand, conserves its ability to remove 8-oxoG:C, suggesting that the Trp-Ile and Ile-Tyr motifs of AthFpg and EcoFpg are not involved in damage recognition or catalysis. The first residue of the motif play a part in stabilizing the everted lesion, as shown recently for Tyr 174 of the Tyr-Ile motif in MvNei1 [52]. In the crystal structures of MvNei1 bound to DNA containing thymine glycol or 5-hydroxyuracil the plane of the tyrosine is perpendicular to that of the lesion.

### 4.4. The αF–β9 loop is essential and specific for 8-oxoG processing

As mentioned above, fungal and plant Fpg proteins do not have the long αF–β9 loop seen in their bacterial counterparts closing the active site once the damaged base is flipped out [51,60,62]. This raises the question of how the eukaryotic Fpg glycosylases identify their target lesions and discriminate among different substrates. In order to understand the role of this loop, we deleted the entire region (residues 213–229) in EcoFpg and assayed the resulting deletion variant for glycosylase activity using double-stranded DNA containing 8-oxoG, Gh, Sp1, or MeFapyG opposite C and DHU opposite G. As shown in a previous study using an analogous deletion variant of *B. stearothermophilus* (BstFpg) [16], excision of 8-oxoG is entirely dependent on the presence of the loop whereas abasic site processing is not affected. Our data (Fig. 6 and Supplemental Figure 2) unambiguously confirm that EcoFpgΔ213–229 is unable to remove 8-oxoG opposite C but conserves its lyase activity. Moreover, glycosylase activity assays of EcoFpgΔ213–229 with four other substrates, including MeFapyG, DHU and the oxidation products of 8-oxoG, Gh and Sp1, clearly show that deleting the loop has virtually no impact on the excision efficiency of these substrates (Fig. 6A and Supplemental Figure 2). The activity pattern of wild-type EcoFpg and EcoFpgΔ213–229 was also examined on γ-irradiated DNA containing multiple lesions. The released lesions detected by GC/MS were qualitatively and quantitatively the same for both enzymes, except that no 8-oxoG was detected when the EcoFpg deletion variant was incubated with the substrate (Fig. 6C). These results, in agreement with the activity assays performed on EcoFpgΔ213–229, show that the loop is absolutely required for the excision of 8-oxoG, but not for other substrates. Although it is clear that this loop is specific for 8-oxoG, whether it is involved in the recognition or the stabilization of 8-oxoG in the binding pocket remains unclear. Considering that the "8-oxoG capping loop" is crucial for the removal of 8-oxoG by bacterial Fpgs, its absence in fungal and plant Fpgs is consistent with the fact that 8-oxoG is not a good substrate for eukaryotic Fpgs.

### 4.5. Concluding remarks

Here we present the first crystal structures of a eukaryotic Fpg, AthFpg, unliganded and in complex with THF-containing DNA. As predicted by an earlier phylogenetic study and confirmed by the

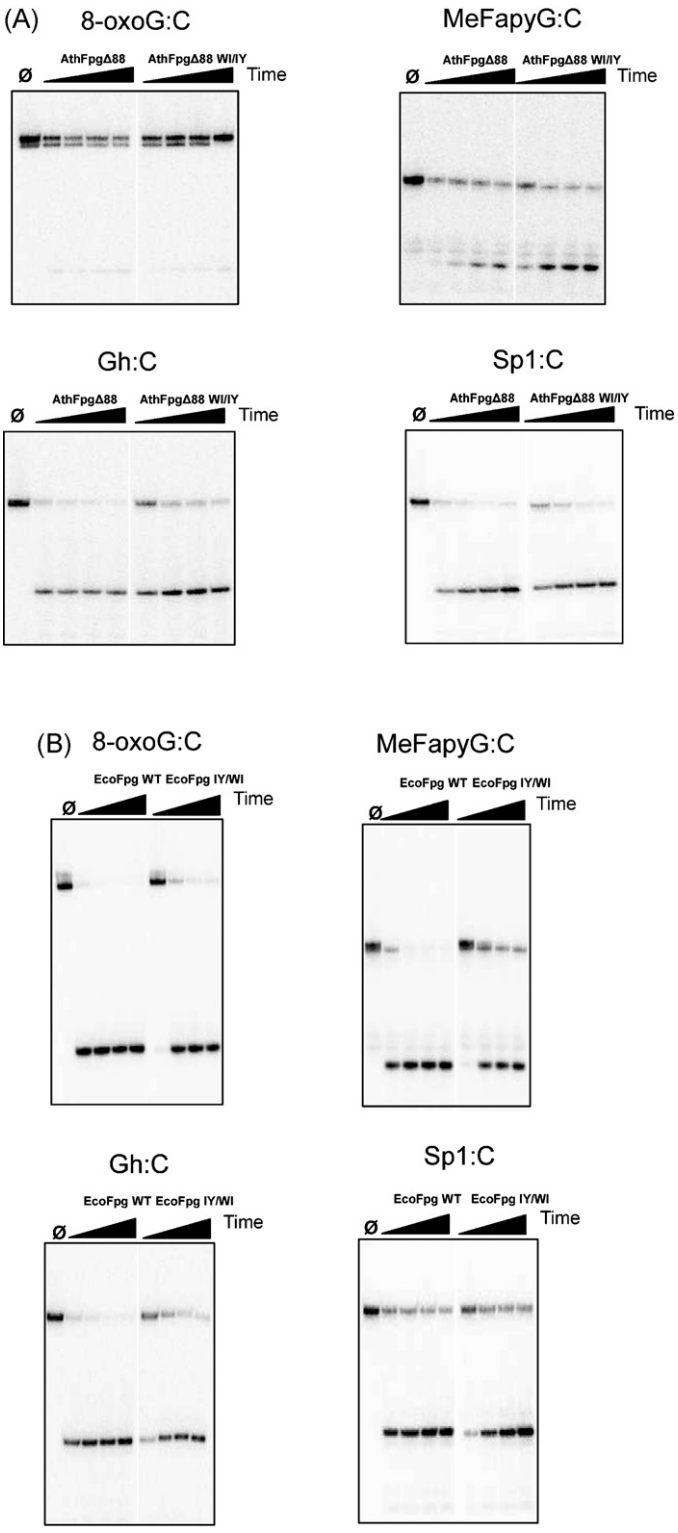


**Fig. 7.** DNA glycosylase activity of AthFpgΔ88 WI187-188IY and EcoFpg IY170-171WI. Excess active enzyme (50 nM) was incubated with double-stranded substrate (10 nM) containing 8-oxoG, MeFapyG, Gh, or Sp1 opposite C. Each glycosylase reaction was incubated for 1, 5, 10, and 15 min at 37 °C. AthFpgΔ88 and wild-type EcoFpg were used in the same condition to allow the comparison.

present structures, AthFpg exhibits a fold similar to that of bacterial Fpg proteins and harbors the structural features required for specific DNA binding and glycosylase catalytic activity [12]. Despite a common fold, AthFpg and bacterial Fpg proteins do not share the same substrate preference. Although we did not identify residues in the active site of AthFpg or EcoFpg responsible for their distinct substrate specificities, we did establish that the 8-oxoG capping loop in EcoFpg is specific for 8-oxoG. The question then becomes whether the loop participates in the recognition of 8-oxoG or if it stabilizes the everted lesion in the binding pocket. Recent data from the David and Kool laboratories using Fpg and nonpolar isosteres of 8-oxoG showed that in this enzyme system hydrogen bonds between the damaged base and the enzyme are not needed for efficient repair of the lesion [63], highlighting the role of facile extrusion from the DNA double helix for proper selection of the damaged base.

## Conflict of interest

None

## Acknowledgements

We thank April Averill and Lauren M. Harvey for help with protein purification, Dr. Alexei Soares for data collection at beamline X29 at the National Synchrotron Light Source (Upton, NY) and Dr. Virginia Rath (Reciprocal Space Consulting, LLC) for collecting a data set at beamline 5.0.3 at the Advanced Light Source (Berkeley, CA). We thank Dr. Carmelo Rizzo (Vanderbilt University) for providing us with oligodeoxynucleotides containing MethylFapyG, and Dr. Cynthia Burrows (University of Utah) for the oligodeoxynucleotides containing guanidinohydantoin and spiroiminodihydantoin. This research was supported by National Institutes of Health grant P01 CA098993 awarded by the National Cancer Institute. Financial support of National Synchrotron Light Source comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the U.S. Department of Energy, and from the National Center for Research Resources of NIH. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH.

Certain commercial equipment or materials are identified in this paper in order to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.dnarep.2012.06.004.

## References

[1] S.S. David, V.L. O'Shea, S. Kundu, Base-excision repair of oxidative DNA damage, Nature 447 (2007) 941–950.
[2] J.C. Fromme, A. Banerjee, G.L. Verdine, DNA glycosylase recognition and catalysis, Curr. Opin. Struct. Biol. 14 (2004) 43–49.
[3] H.E. Krokan, H. Nilsen, F. Skorpen, M. Otterlei, G. Slupphaug, Base excision repair of DNA in mammalian cells, FEBS Lett. 476 (2000) 73–77.
[4] A.K. McCullough, M.L. Dodson, R.S. Lloyd, Initiation of base excision repair: glycosylase mechanisms and structures, Annu. Rev. Biochem. 68 (1999) 255–285.
[5] S. Mitra, T. Izumi, I. Boldogh, K.K. Bhakat, J.W. Hill, T.K. Hazra, Choreography of oxidative damage repair in mammalian genomes, Free Radical Biol. Med. 33 (2002) 15–28.
[6] S.S. Wallace, D.L. Murphy, J.B. Sweasy, Base excision repair and cancer, Cancer Lett. (2012).
[7] S. Duclos, S. Doublié, S.S. Wallace, Consequences and repair of oxidative DNA damage, in: The Cellular Response to the Genotoxic Insult: The Question of Threshold for Genotoxic Carcinogens, The Royal Society of Chemistry, 2012, pp. 115–159.
[8] H.E. Krokan, R. Standal, G. Slupphaug, DNA glycosylases in the base excision repair of DNA, Biochem. J. 325 (Pt 1) (1997) 1–16.

[9] A. Prakash, S. Doublié, S.S. Wallace, The Fpg/Nei family of DNA glycosylases: substrates, structures and search for damage, in: P. Doetsch (Ed.), Progress in Molecular Biology and Translational Science, Academic Press, 2012.
[10] D.O. Zharkov, G. Shoham, A.P. Grollman, Structural characterization of the Fpg family of DNA glycosylases, DNA Repair (Amst.) 2 (2003) 839–862.
[11] A.P. Grollman, M. Moriya, Mutagenesis by 8-oxoguanine: an enemy within, Trends Genet. 9 (1993) 246–249.
[12] S.D. Kathe, R. Barrantes-Reynolds, P. Jaruga, M.R. Newton, C.J. Burrows, V. Bandaru, M. Dizdaroglu, J.P. Bond, S.S. Wallace, Plant and fungal Fpg homologs are formamidopyrimidine DNA glycosylases but not 8-oxoguanine DNA glycosylases, DNA Repair (Amst.) 8 (2009) 643–653.
[13] N. Krishnamurthy, X. Zhao, C.J. Burrows, S.S. David, Superior removal of hydantoin lesions relative to other oxidized bases by the human DNA glycosylase hNEIL1, Biochemistry 47 (2008) 7137–7146.
[14] S. Doublié, V. Bandaru, J.P. Bond, S.S. Wallace, The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 10284–10289.
[15] K. Imamura, S.S. Wallace, S. Doublié, Structural characterization of a viral NEIL1 ortholog unliganded and bound to abasic site-containing DNA, J. Biol. Chem. 284 (2009) 26174–26183.
[16] Y. Qi, M.C. Spong, K. Nam, A. Banerjee, S. Jiralerspong, M. Karplus, G.L. Verdine, Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme, Nature 462 (2009) 762–766.
[17] X. Li, P. Romero, M. Rani, A.K. Dunker, Z. Obradovic, Predicting protein disorder for N-, C-, and internal regions, Genome Inform. Ser. Workshop Genome Inform. 10 (1999) 30–40.
[18] S. Doublié, Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems, Methods Mol. Biol. 363 (2007) 91–108.
[19] V. Bandaru, J.O. Blaisdell, S.S. Wallace, Oxidative DNA glycosylases: recipes from cloning to characterization, Methods Enzymol. 408 (2006) 15–33.
[20] J. Lee, H.J. Lee, M.K. Shin, W.S. Ryu, Versatile PCR-mediated insertion or deletion mutagenesis, Biotechniques 36 (2004) 398–400.
[21] J.S. Papadopoulos, R. Agarwala, COBALT: constraint-based alignment tool for multiple protein sequences, Bioinformatics 23 (2007) 1073–1079.
[22] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[23] P.P. Christov, K.L. Brown, I.D. Kozekov, M.P. Stone, T.M. Harris, C.J. Rizzo, Site-specific synthesis and characterization of oligonucleotides containing an N6-(2-deoxy-D-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-N-me thylformamidopyrimidine lesion, the ring-opened product from N7-methylation of deoxyguanosine, Chem. Res. Toxicol. 21 (2008) 2324–2333.
[24] O. Kornyushyna, A.M. Berges, J.G. Muller, C.J. Burrows, In vitro nucleotide misinsertion opposite the oxidized guanosine lesions spiroiminodihydantoin and guanidinohydantoin and DNA synthesis past the lesions using *Escherichia coli* DNA polymerase I (Klenow fragment), Biochemistry 41 (2002) 15304–15314.
[25] M. Dizdaroglu, C. Bauche, H. Rodriguez, J. Laval, Novel substrates of *Escherichia coli* nth protein and its kinetics for excision of modified bases from DNA damaged by free radicals, Biochemistry 39 (2000) 5586–5592.
[26] M. Dizdaroglu, Chemical determination of oxidative DNA damage by gas chromatography-mass spectrometry, Methods Enzymol. 234 (1994) 3–16.
[27] M. Liu, V. Bandaru, J.P. Bond, P. Jaruga, X. Zhao, P.P. Christov, C.J. Burrows, C.J. Rizzo, M. Dizdaroglu, S.S. Wallace, The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 4925–4930.
[28] Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode, Methods Enzymol. 276 (1997) 307–326.
[29] A.G. Leslie, The integration of macromolecular diffraction data, Acta Crystallogr. D: Biol. Crystallogr. 62 (2006) 48–57.
[30] K. Diederichs, P.A. Karplus, Improved R-factors for diffraction data analysis in macromolecular crystallography, Nat. Struct. Biol. 4 (1997) 269–275.
[31] T.C. Terwilliger, J. Berendzen, Automated MAD and MIR structure solution, Acta Crystallogr. D: Biol. Crystallogr. 55 (1999) 849–861.
[32] C. Vonrhein, E. Blanc, P. Roversi, G. Bricogne, Automated structure solution with autoSHARP, Methods Mol. Biol. 364 (2007) 215–230.
[33] T.C. Terwilliger, R.W. Grosse-Kunstleve, P.V. Afonine, N.W. Moriarty, P.H. Zwart, L.W. Hung, R.J. Read, P.D. Adams, Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, Acta Crystallogr. D: Biol. Crystallogr. 64 (2008) 61–69.
[34] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr. D: Biol. Crystallogr. 60 (2004) 2126–2132.
[35] A.T. Brunger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography & NMR system: a new software suite for macromolecular structure determination, Acta Crystallogr. D: Biol. Crystallogr. 54 (1998) 905–921.
[36] A.J. McCoy, L.C. Storoni, R.J. Read, Simple algorithm for a maximum-likelihood SAD function, Acta Crystallogr. D: Biol. Crystallogr. 60 (2004) 1220–1228.
[37] A.T. Brunger, Version 1.2 of the crystallography and NMR system, Nat. Protoc. 2 (2007) 2728–2733.
[38] M.D. Winn, M.N. Isupov, G.N. Murshudov, Use of TLS parameters to model anisotropic displacements in macromolecular refinement, Acta Crystallogr. D: Biol. Crystallogr. 57 (2001) 122–133.
[39] R.A. Laskowski, D.S. Moss, J.M. Thornton, Main-chain bond lengths and bond angles in protein structures, J. Mol. Biol. 231 (1993) 1049–1067.
[40] E. Krissinel, K. Henrick, Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, Acta Crystallogr. D: Biol. Crystallogr. 60 (2004) 2256–2268.
[41] R. Lavery, H. Sklenar, Defining the structure of irregular nucleic acids: conventions and principles, J. Biomol. Struct. Dyn. 6 (1989) 655–667.
[42] Schrodinger, LLC, The PyMOL Molecular Graphics System, Version 1.3r1, in, 2010.
[43] J. Tchou, V. Bodepudi, S. Shibutani, I. Antoshechkin, J. Miller, A.P. Grollman, F. Johnson, Substrate specificity of Fpg protein. Recognition and cleavage of oxidatively damaged DNA, J. Biol. Chem. 269 (1994) 15318–15324.
[44] T. Ohtsubo, O. Matsuda, K. Iba, I. Terashima, M. Sekiguchi, Y. Nakabeppu, Molecular cloning of AtMMH, an *Arabidopsis thaliana* ortholog of the *Escherichia coli* mutM gene, and analysis of functional domains of its product, Mol. Gen. Genet. 259 (1998) 577–590.
[45] S. Doublié, Macromolecular Crystallography Protocols, Humana Press, Totowa, N.J., 2007.
[46] V. Bandaru, W. Cooper, S.S. Wallace, S. Doublié, Overproduction, crystallization and preliminary crystallographic analysis of a novel human DNA-repair enzyme that recognizes oxidative DNA damage, Acta Crystallogr. D: Biol. Crystallogr. 60 (2004) 1142–1144.
[47] J.C. Fromme, G.L. Verdine, Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM, Nat. Struct. Biol. 9 (2002) 544–552.
[48] G. Golan, D.O. Zharkov, H. Feinberg, A.S. Fernandes, E.I. Zaika, J.H. Kycia, A.P. Grollman, G. Shoham, Structure of the uncomplexed DNA repair enzyme endonuclease VIII indicates significant interdomain flexibility, Nucleic Acids Res. 33 (2005) 5006–5016.
[49] D.O. Zharkov, G. Golan, R. Gilboa, A.S. Fernandes, S.E. Gerchman, J.H. Kycia, R.A. Rieger, A.P. Grollman, G. Shoham, Structural analysis of an *Escherichia coli* endonuclease VIII covalent reaction intermediate, EMBO J. 21 (2002) 789–800.
[50] L. Serre, K. Pereira de Jesus, S. Boiteux, C. Zelwer, B. Castaing, Crystal structure of the *Lactococcus lactis* formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA, EMBO J. 21 (2002) 2854–2865.
[51] J.C. Fromme, G.L. Verdine, DNA lesion recognition by the bacterial repair enzyme MutM, J. Biol. Chem. 278 (2003) 51543–51548.
[52] K. Imamura, A. Averill, S.S. Wallace, S. Doublié, Structural characterization of viral ortholog of human DNA glycosylase NEIL1 bound to thymine glycol or 5-hydroxyuracil-containing DNA, J. Biol. Chem. 287 (2012) 4288–4298.
[53] K. Pereira de Jesus, L. Serre, C. Zelwer, B. Castaing, Structural insights into abasic site for Fpg specific binding and catalysis: comparative high-resolution crystallographic studies of Fpg bound to various models of abasic site analogues-containing DNA, Nucleic Acids Res. 33 (2005) 5936–5944.
[54] R. Gilboa, D.O. Zharkov, G. Golan, A.S. Fernandes, S.E. Gerchman, E. Matz, J.H. Kycia, A.P. Grollman, G. Shoham, Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA, J. Biol. Chem. 277 (2002) 19811–19816.
[55] H. Dou, S. Mitra, T.K. Hazra, Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2, J. Biol. Chem. 278 (2003) 49679–49684.
[56] O.M. Sidorkina, J. Laval, Role of lysine-57 in the catalytic activities of *Escherichia coli* formamidopyrimidine-DNA glycosylase (Fpg protein), Nucleic Acids Res. 26 (1998) 5351–5357.
[57] A. Banerjee, W.L. Santos, G.L. Verdine, Structure of a DNA glycosylase searching for lesions, Science 311 (2006) 1153–1157.
[58] A.R. Dunn, N.M. Kad, S.R. Nelson, D.M. Warshaw, S.S. Wallace, Single Qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA, Nucleic Acids Res. 39 (2011) 7487–7498.
[59] F. Coste, M. Ober, Y.V. Le Bihan, M.A. Izquierdo, N. Hervouet, H. Mueller, T. Carell, B. Castaing, Bacterial base excision repair enzyme Fpg recognizes bulky N7-substituted-FapydG lesion via unproductive binding mode, Chem. Biol. 15 (2008) 706–717.
[60] M. Sugahara, T. Mikawa, T. Kumasaka, M. Yamamoto, R. Kato, K. Fukuyama, Y. Inoue, S. Kuramitsu, Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, Thermus thermophilus HB8, EMBO J. 19 (2000) 3857–3869.
[61] Y. Qi, M.C. Spong, K. Nam, M. Karplus, G.L. Verdine, Entrapment and structure of an extrahelical guanine attempting to enter the active site of a bacterial DNA glycosylase, MutM, J. Biol. Chem. 285 (2010) 1468–1478.
[62] F. Coste, M. Ober, T. Carell, S. Boiteux, C. Zelwer, B. Castaing, Structural basis for the recognition of the FapydG lesion (2,6-diamino-4-hydroxy-5-formamidopyrimidine) by formamidopyrimidine-DNA glycosylase, J. Biol. Chem. 279 (2004) 44074–44083.
[63] P.L. McKibbin, A. Kobori, Y. Taniguchi, E.T. Kool, S.S. David, Surprising repair activities of nonpolar analogs of 8-oxoG expose features of recognition and catalysis by base excision repair glycosylases, J. Am. Chem. Soc. 134 (2012) 1653–1661.