Nucleic Acids Research, 2023, **51**, 12508–12521
https://doi.org/10.1093/nar/gkad1075
Advance access publication date: 16 November 2023
Structural Biology

Structural and biochemical insights into NEIL2's preference for abasic sites

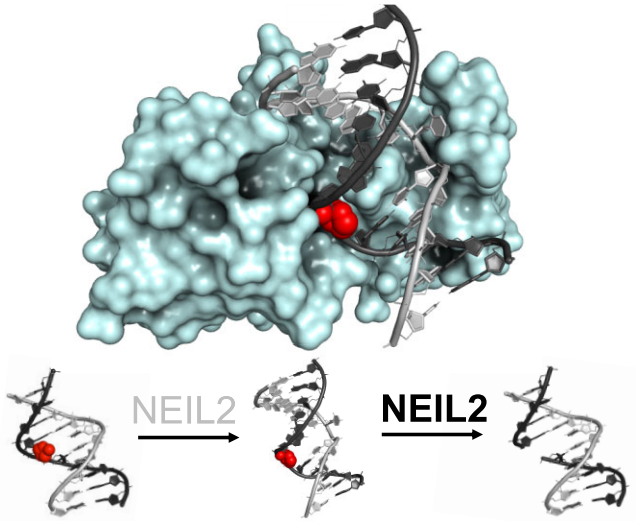
Brian E. Eckenroth [1,*], Joshua D. Bumgarner[2], Olivia Matsumoto-Elliott[2], Sheila S. David [2] and Sylvie Doublié [1]

[1]Department of Microbiology and Molecular Genetics, University of Vermont, Stafford Hall, 95 Carrigan Drive, Burlington, VT 05405, USA
[2]Department of Chemistry and Graduate Program in Chemistry and Chemical Biology, University of California Davis, Davis, CA 95616, USA
[*]To whom correspondence should be addressed. Tel: +1 802 656 9533; Fax: +1 802 656 0881; Email: Brian.Eckenroth@uvm.edu

Abstract

Cellular DNA is subject to damage from a multitude of sources and repair or bypass of sites of damage utilize an array of context or cell cycle dependent systems. The recognition and removal of oxidatively damaged bases is the task of DNA glycosylases from the base excision repair pathway utilizing two structural families that excise base lesions in a wide range of DNA contexts including duplex, single-stranded and bubble structures arising during transcription. The mammalian NEIL2 glycosylase of the Fpg/Nei family excises lesions from each of these DNA contexts favoring the latter two with a preference for oxidized cytosine products and abasic sites. We have determined the first liganded crystal structure of mammalian NEIL2 in complex with an abasic site analog containing DNA duplex at 2.08 Å resolution. Comparison to the unliganded structure revealed a large interdomain conformational shift upon binding the DNA substrate accompanied by local conformational changes in the C-terminal domain zinc finger and N-terminal domain void-filling loop necessary to position the enzyme on the DNA. The detailed biochemical analysis of NEIL2 with an array of oxidized base lesions indicates a significant preference for its lyase activity likely to be paramount when interpreting the biological consequences of variants.

Graphical abstract



Introduction

DNA is subject to damage from a multitude of endogenous and exogenous sources, and repair or bypass of these sites of damage utilize an array of context-dependent systems—most of which are damage or structure-specific (1). Additionally, numerous essential cellular processes produce DNA structures, elements or intermediates that require enzymes from repair pathways for processing (2–5). The primary role for the base excision repair (BER) pathway is in genomic maintenance, serving as the front line of defense by recognition and removal of damaged purines and pyrimidines before they are encountered by the replication machinery. Failures or inefficiencies within BER can result in mutagenesis via mismatch nucleotide incorporation during replication or the ill-timed accumulation of intermediates producing catastrophic events such as strand breaks during replication or transcription (6–8). Sites of base loss or abasic (AP) sites are one of the most common types of DNA damage and intermediates. Processes

Received: July 31, 2023. **Revised:** October 17, 2023. **Editorial Decision:** October 18, 2023. **Accepted:** October 30, 2023

such as class-switch recombination and ten eleven translocase (TET) enzyme-mediated demethylation are examples that produce AP site intermediates and rely on enzymes of the base excision repair pathway (BER) for completion (3).

The recognition and removal of oxidatively damaged DNA lesions by glycosylases of the BER pathway utilizes two structural families, which act on oxidized base lesions in a wide range of DNA contexts: duplex, single-stranded and bubble structures arising during replication and transcription (9,10). In humans, the Nth superfamily contains OGG1 (8-oxoguanine glycosylase), NTHL1, and MUTYH, which share a common α-helical fold and an active site helix–hairpin–helix (HhH) motif followed by a Gly/Pro-rich loop and an invariant aspartic acid residue (11). The Fpg/Nei family comprises the three NEIL (Nei-Like) glycosylases. NEIL1, 2 and 3 each contain unique structural features within the conserved two-domain architecture fold: an N-terminal domain made of a two layered β-sandwich and a mostly helical C-terminal domain containing the helix-two-turn helix and zinc (or zincless) finger motifs for binding DNA. The mammalian NEIL2 glycosylase acts on oxidized base lesions in several DNA contexts, favoring single-stranded and bubble DNA, and displays a preference for oxidized cytosine products and AP sites (12,13). NEIL2 is of significant interest due to its implications in activities outside of canonical BER functions of genomic integrity (4,7). The enzyme has been shown to be uniquely conformationally dynamic compared to NEIL1 and NEIL3 (13). Detailed structural and mechanistic understanding of this unique behavior is needed to provide insight into how this glycosylase achieves its broad function, which in turn can lead to a novel avenue for pathway manipulation.

We have determined the 2.08 Å resolution structure of mammalian NEIL2 in complex with an abasic site analog-containing DNA duplex. A comparison to the unliganded structure revealed a large interdomain conformational change upon binding the DNA substrate, accompanied by local conformational changes in the C-terminal domain zinc finger and in the N-terminal domain void-filling loop necessary to position the enzyme on the DNA. The protein-DNA complex structure also reveals the position of a 10-residue insert unique to NEIL2, relative to NEIL1 and NEIL3, adjacent to the orphaned base on the complementary strand. Detailed activity studies for a range of oxidized lesions, in both ssDNA and ds-DNA contexts, paint a picture of NEIL2 as significantly favoring abasic sites over oxidized base lesions. The structural analysis provides insight into how NEIL2 is primed to be uniquely versatile, with features that also make it an intriguing target for manipulation of its BER-associated activities.

Materials and methods

Reagents

All reagents utilized were of highest quality available and purchased via Fisher Scientific or Sigma Aldrich unless indicated in the methods. Oligonucleotides for crystallization were synthesized at IDT.

Biological resources

Constructs used with protein expression were synthesized and codon optimized for expression in *Escherichia coli* at Genscript, received in the pUC57 shuttle vector and subcloned to pET30 vector (Novagen) for expression. Plasmid propagation used DH5α while protein expression used BL21DE3 from New England Biolabs.

Protein expression and purification

The gene construct for *Monodelphis domestica* NEIL2 (MdoNEIL2ΔD1) where the large disordered region (D1; Phe68 to Leu125) was replaced with the sequence Gly-Ser-Gly-Ser-Gly was synthesized by Genscript and subcloned into the pET30 (Novagen) vector using NdeI 5′ and XhoI 3′ restriction sites (13). The vector was transformed into *E. coli* BL21(DE3) for expression in terrific broth (TB) supplemented with 25 μM $ZnSO_4$ utilizing a 6-h induction at 20°C and 500 μM IPTG after reaching OD_{600} of 0.8 at 37°C, using the same protocol published for the WT MdoNEIL2 and HsaNEIL2 (13). Expression for selenomethionine incorporation substituted the TB media with minimal medium as described previously (14). Cells were pelleted and resuspended at 4 ml per gm of cell pellet in buffer containing 50 mM sodium phosphate pH 8, 300 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, 0.01% NP-40 and 25 mM imidazole, lysed by ultrasonification, and clarified by ultra-centrifugation. The supernatant was subjected to batch binding and elution from Ni-NTA agarose with elution at 400 mM imidazole. Eluant was adjusted to 50 mM HEPES pH 7.5, 100 mM NaCl, 10% glycerol and 1 mM DTT for purification over an SP FF cation exchange column (Cytiva) with a 20-column volume gradient to 1 M NaCl. Eluted protein was concentrated to 10 mg/ml and flash frozen for storage at −80°C. The MdoNEIL2ΔD1 Asn182Asp variant was prepared in a similar manner. An additional deletion construct, MdoNEIL2ΔD2, converts the sequence 155-Lys-Ala-Asn-Lys-Arg-Gly-Asp-Trp-Lys-163 to the linker Gly-Ser-Gly-Ser-Gly and was cloned, expressed and purified similarly to WT.

Sample preparation and crystallization

The DNA used for crystallization was synthesized by Integrated DNA Technologies (IDT), PAGE purified and annealed in 10 mM Tris pH 8 and 50 mM NaCl. The damage strand is a 14-mer oligonucleotide and the complementary strand a 13-mer oligonucleotide, resulting in a 5′-overhang for the damage strand (X signifies the location of the lesion or tetrahydrofuran (THF)).

5′-d(GCG TCC AXG TCT AC)-3′
3′-d(GC AGG TCC AGA TG)-5′

Duplex DNA at 135 μM was mixed with 90 μM MdoNEIL2 in the presence of 25 mM HEPES pH 7.5, 60 mM NaCl, 1 mM TCEP, 2% glycerol and incubated at room temperature for 30 minutes. Hanging drop vapor diffusion was employed using a reservoir solution containing 16–22% PEG 3350, 50 mM sodium acetate, 1 mM TCEP, 0.1M BisTris pH 6.5, and 3–7% DMSO. Complex and reservoir were mixed at equal volume and incubated at 24°C. Cryoprotection was achieved by soaking in reservoir reagent containing 20% PEG 3350 and 12% DMSO prior to flash cooling in liquid nitrogen. Heavy atom derivatives were produced by soaking of crystals in cryoprotection reagent at room temperature in the presence of 5 mM $KAuCl_2$, 100mM $BaCl_2$ or 100 mM hexamine cobalt.

Table 1. Crystallographic data collection and structure refinement statistics for MdoNEIL2ΔD1 in complex with THF-containing dsDNA

PDB ID	8TH9
Crystals (selenomethionine)	2
Wavelength (Å)	0.9794
Space group	I4(1)22
Cell a = *b*, *c* (Å)	121.430, 117.690
Resolution (Å)	38–2.08
(High resolution)[a]	(2.14–2.08)
Reflections non-anomalous	26 705
Reflections all	50 513
Completeness (%)	100 (100)
Wilson *B* (Å^2)	43.8
Multiplicity	21.6 (11.9)
I/sigI	12.4 (1.3)
R_{meas} %	13.6 (282)
R_{pim} %	2.9 (81.0)
$CC_{1/2}$	0.997 (0.522)
Model refinement	
R_{work}/R_{free} %	22.2/24.4
RMS bonds (Å)	0.009
RMS angles (°)	1.190
Rotamer outlier	8.3
Clash score	16.4
Ramachandran (%)	
Favored	98.0
Allowed	2.0
Outlier	0
Mean *B* (Å^2)	67.3
Correlation coefficient	0.843
Atoms protein	2946
Zn^{2+}	1
DNA	543
Solvent	59
Coordinate error (Å)	0.39
Phase error (°)	34.0

[a]Numbers in parentheses denote high resolution bin.

Structure determination

Primary data collection was performed at GM/CA Advanced Photon Source at Argonne National laboratory on the 23ID-B Eiger 16M and 23ID-D Pilatus 6M detectors. Additional data were collected at the National Synchrotron Light Source (NSLS-II) at Brookhaven National Laboratory on 17ID AMX on an Eiger 9M. Data were processed with multiple software platforms and evaluated for anomalous signal and isomorphism between crystal types. Best integration for the SeMet (Table 1) crystals and native (Supplementary Table S1) used Mosflm while the barium data set used Dials and the cobalt, low energy and gold data sets were processed with Mosflm (15), Dials (16) and XDS (17) depending on the crystal (Supplementary Table S1). Blend was used to merge data sets from 2–6 crystals within a crystal type (18) with final scaling and merging performed by Aimless. Initial phases were determined by combining multiple isomorphous methods using Sharp (19) in comparison with Se-Single Anomalous Dispersion (SAD) partial phases from Crank2 (20) within CCP4 (21) performing iterative rounds of search and refinement parameter optimization until a common origin could be found with both methods. Phased maps provided automated quality building for the C-terminal domain using Parrot (22) and Buccaneer (23) and revealed clear density for DNA duplex which was built manually into the phased map in Coot (24) using DNA from PDBID 3A46 (25) as a reference model. Placement of the N-terminal domain proved elusive using automated methods with the exception of the fraction of residues in proximity of the DNA. Anomalous difference Fourier maps from both Se-SAD and MIR consistently showed 8 peaks about a two-fold rotational axis where Met36 and Met39 were expected (Figure 1). Anomalous peaks for sulfur atoms and phosphate atoms verified the positions of cysteines and DNA backbone (Supplementary Figure S1) while elongation of the cyteine anomalous peaks of the N-terminal domain also suggested conformational heterogeneity. Iterative rounds of rigid body placement and rebuild were performed in high and low symmetry space groups until two conformations of the N-terminal domain were identified that resulted in similar R_{free} values. Iterative independent builds for each conformation were performed aided by real space refinement routines under morphing parameters within Phenix (26) restrained by the multiple isomorphous replacement (MIR) phases and geometry improvement by subsequent morphing using data scaled F_{calc}. Final rounds of refinement (Table 1) were performed in Phenix after setting the two N-terminal domain conformations to 0.5 occupancy within the single model with MIR phase restraints and domain-based TLS parameters. Refine-

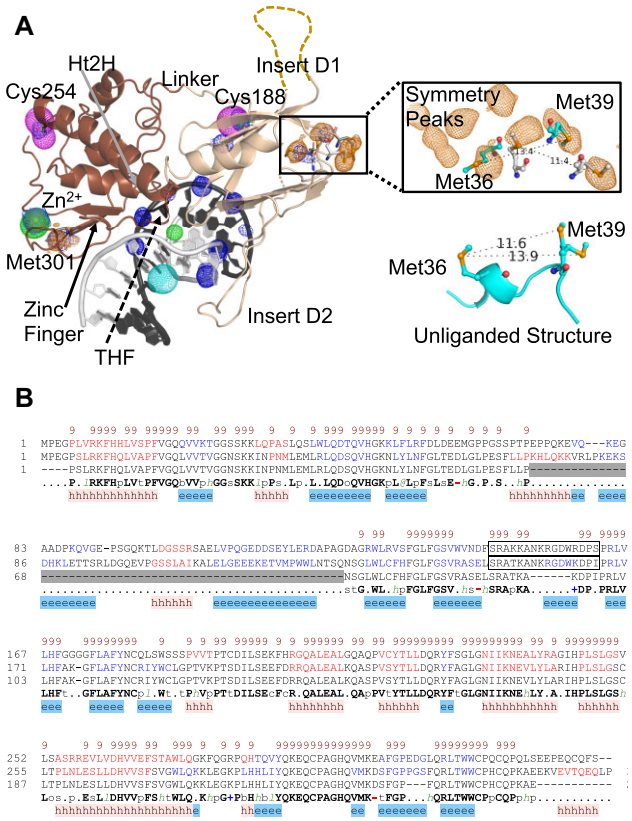


Figure 1. Crystal structure of NEIL2 from *Monodelphis domestica* bound to DNA. (**A**) MdoNEIL2 is colored by domain with tan for the N-terminal and brown for the C-terminal domains. Anomalous difference Fourier maps contoured at 3 σ for selenomethionine (orange), gold (maroon), barium (cyan), hexamine cobalt (green) and sulfur/phosphate (blue) used for structure determination. In the boxes are methionines 36 and 39 at the crystallographic two-fold axis, showing peak duplication indicating the dual conformation. The distance measurement of the two rotamers for the equivalent residues in the previously solved unliganded structure PDBID 6VJI are shown for (13). (**B**) Structure based sequence alignment using Promals3D (63) between HsaNEIL2 and MdoNEIL2. Greyed region is insert D1 and boxed region is insert D2. Rows for the alignment from top to bottom are conservation, HsaNEIL2, MdoNEIL2, PDBID:6vji, consensus sequence, consensus secondary structure.

ment was stabilized using mask parameter refinement in combination with the Waasmaier and Kirfel scattering function (27). Multiple lower Laue groups were explored, including solutions at space group P1 with no improvement in refinement or residual maps. The final best model resulting from refinement against the selenomethionine data using space group *I*4(1)22 and containing a single complex in the asymmetric unit with dual conformation of the N-terminal domain has been deposited along with structure factor amplitudes under PDB ID 8TH9.

Oligonucleotide preparation

For all substrates tested in this manuscript, the lesion substrate of interest was placed in the X position of a 30-mer oligonucleotide sequence (#1). To analyze removal activity within the context of duplex DNA, oligonucleotide (#1) was annealed with oligonucleotide (#2), where position Y was occupied by the canonical base that would be paired with the base the lesion is derived from (*i.e.* Gh was placed across from C, Tg was placed opposite A). In the case of the AP site, C was positioned opposite for comparison with Gh.

(#1)5′ -- d (TGT TCA TCA TGG GTC**X**TC GGT ATA TCC CAT) – 3′
(#2)3′ -- d (ACA AGT AGT ACC CAG**Y**AG ACC TAT AGG GTA) – 5′

Oligonucleotides containing 8-oxo-7,8-dihydroguanine (OG), thymine glycol (Tg) and 5-hydroxy-methyl uracil (5-hmU) were synthesized from phosphoramidite monomers purchased from Glen Research. Oligonucleotides containing uracil glycol (Ug) and 8-oxoinosine (OI) were synthesized using an Ug phosphoramidite that was synthesized using the Ug or OI phosphoramidite monomers that were synthesized as described previously (28–30). The oligonucleotides were synthesized using automated solid-phase DNA synthesis at the DNA/peptide synthesis core facility at the University of Utah Medical School. Oligonucleotides containing Dihydrothymine (DHT), 5-hydroxycytosine (5-OHC), 5-hydroxyUracil (5-OHU) were purchased from Midland Reagents. Uracil containing sequences were obtained from Integrated DNA Technologies. Oligonucleotides containing guanidinohydantoin (Gh) and both isomers of spiroiminodihydantoin (*R*-Sp and *S*-Sp) were prepared by oxidation of the OG containing oligonucleotide using Na_2IrCl_6 either in water at room temperature, or in 10 mM sodium phosphate buffer (pH 7.5–8.5) with 100 mM NaCl and heated to 65°C as previously reported (31). All Oligonucleotides were purified with a Dionex ion-exchange column using Beckman Gold Nouveau HPLC or a Shimadzu LC 10AT HPLC as described previously (28,32,33). Purified oligonucleotides were lyophilized, desalted with a Sep-Pak C18 desalting cartridge (Waters) and concentrations were determined based on the Abs_{260nm}. All oligonucleotide sequences were confirmed by MALDI-MS (Supplementary Table S2), except the Gh and Sp oligonucleotides which were confirmed by Q-exactive HF Orbitrap (Supplementary Figure S2).

For kinetic experiments, purified substrate containing oligonucleotide (#1) was radioactively end-labeled with γ-^{32}P-ATP purchased from PerkinElmer using T4 polynucleotide kinase from New England Biolabs (NEB) following the manufacturer's protocol. The labeled oligonucleotide was mixed with unlabeled oligonucleotide to an approximate concentration of 5% of the total to provide for accurate DNA substrate concentration determination. In the case of duplex substrates, the mixture of labeled and unlabeled oligonucleotides was mixed with unlabeled oligonucleotide (#2) in 20% excess. Both single strand and duplex substrates were annealed heating at 90°C for 5 min and then slowly cooled to 4°C in 20 mM Tris–HCl, pH7.6, 10 mM EDTA, 150 mM NaCl.

The abasic site (AP) substrates were prepared using an oligonucleotide containing uracil at the desired position (purchased from IDT). The uracil-containing DNA was then radiolabeled and mixed with unlabeled uracil containing strand to provide a sample that contained approximately 5% ^{32}P-label. The single stranded U-oligonucleotide (10 μM) was mixed with 10 units of uracil DNA glycosylase (UDG), purchased from New England Biolabs (NEB). One unit of UDG catalyzes and releases 60 pmol of uracil per minute from double-stranded DNA as defined by the manufacture. The reaction was allowed to proceed in the manufacturer's supplied buffer solution (10 mM Tris–HCl, 1 mM DTT, 1 mM EDTA at pH 8) and incubated for 30 minutes. Subsequently, the DNA was ethanol precipitated, quantified and annealed in buffer solution as described, with the exception of heating to 55°C to avoid background cleavage of the AP-site.

Kinetic and statistical analysis

Active site titrations were performed to determine the active fraction (%) for each enzyme using methods similar to previous reports (34,35). These experiments were performed under multiple turnover conditions (MTO) where the $[DNA]>[E]_{total}$ and using 20 nM AP-site containing single stranded DNA. The amplitude of the burst phase derived from production curves was divided by the total protein concentration (based on absorption at 280 nm) for each enzyme preparation to give an active enzyme fraction (Supplementary Table S3). Single turnover (STO) kinetic experiments where [E]>[DNA], were performed using 200 nM active enzyme, and 20 nM DNA by monitoring the extent of strand scission induced at the lesion site as a function of time after quenching reactions with heat and either denaturing dye containing 80% formamide, 0.025% xylene cyanol, 0.025% bromophenol blue in TBE buffer (for AP site DNA), or 1M NaOH (all bases). All reactions were quenched for 4 minutes at 90°C except for the AP site containing substrates, which were quenched at 55°C to avoid background strand scission. All reactions were performed in buffer containing 150 mM NaCl, 20 mM Tris–HCl, pH 7.6, 10 mM EDTA, and 0.1 mg/ml BSA. Time points of 20 s, 40 s, 1, 3, 5, 10, 20, 40 and 60 min were collected. All time courses were loaded onto 15% denaturing polyacrylamide gels in 1× TBE to separate the substrate and the product bands. Gels were then exposed to storage phosphor screens overnight, scanned with an Amersham Typhoon, quantified with ImageQuant TL v8.2, and fit to equations 1 or 2 using GraFit 5.0 for STO and MTO reactions respectively. Reaction time course were analyzed using three different enzyme aliquots for each enzyme form, and at least three trials were performed with each aliquot. The plots shown in the text represent the average (with standard deviation at each point) resulting from at least three trials from separate enzyme aliquots with three replicates for each aliquot. The production curves and relevant resulting kinetic parameters obtained from fits were fit and plotted using GraphPad Prism v9.3.

While under STO conditions, both the glycosylase reaction, and the lyase reactions were fit to Equation 1 where k_{obs} is equated to k_g, the rate of all steps involved in glycosidic bond cleavage for NEIL2 with base lesion substrates, and k_L represents the rate of the AP lyase strand scission for NEIL2 reactions with the AP-containing substrates.

$$[P]_t = A\left[1 - \exp(-k_{obs}t)\right] \quad (1)$$

The MTO kinetics were fit to Equation 2 where the two rates represent k_{obs} (rate constant for the burst) and k_{ss} (rate of the linear phase), which can be simplified to the turn-over rate k_{to} by dividing k_{ss} by the amplitude of the burst phase A. Under the MTO conditions, A = [active enzyme].

$$[P]_t = A\left[1 - \exp(-k_{obs}t)\right] + k_{ss}t \quad (2)$$

Results

Overall structure of the protein–DNA complex

We attempted to crystallize both human NEIL2 (HsaNEIL2) and *Monodelphis domestica* NEIL2 (MdoNEIL2) with DNA. NEIL2 contains two internal flexible inserts, a large insert (D1) comprising Phe68 to Leu125 and a smaller insert (D2) comprising residues 155–163 (13). When crystals did not form, we deleted insert D1 for MdoNEIL2 based on the structure of the unliganded form and a structure-based sequence alignment (13). Sequence comparison of HsaNEIL2 with MdoNEIL2 shows 57% identity and 69% similarity for the full-length protein, and 65% identity with 73% similarity when omitting insert D1. Insert D1 was structurally predicted to be away from the DNA-binding interface and previous studies have implicated insert D1 in protein-protein interactions with other BER factors (36,37). Importantly the resulting protein construct folds properly and is active. Crystals diffracting to 2.0 Å were obtained in the presence of DNA containing tetrahydrofuran (THF), a non-hydrolyzable abasic site analog. The crystal structure was solved by a combination of multiple isomorphous replacement, single wavelength anomalous dispersion, and molecular replacement data, as molecular replacement alone using the unliganded enzyme as model failed to provide a solution. Two conformations of the N-terminal domain relative to the DNA were observed in the crystal. While a rotational difference is observed between the two orientations of the N-terminal domain, they converge to a pivot point at the DNA duplex interface such that all key residues and features are retained at a similar point of action. For the remaining description of the protein-DNA complex we will therefore limit the key points to a single conformation for simplicity.

The overall structure of NEIL2 bound to a DNA duplex shows the general characteristics of the Fpg/Nei superfamily where the DNA is bound in the cleft between the N- and C-terminal domains (Figure 2) with the exception that NEIL2 possesses an extended looped arm (insert D2) encircling the complementary DNA strand at the orphaned base position (Figure 2). Mapping of the interactions between NEIL2 and the DNA reveals an asymmetric binding mode favoring 3′-directional binding to the duplex on both the damage and complementary strands (Supplementary Figure S3). The tightest interactions appear to be within two nucleotides of the lesion in the 3′ direction on each strand but contacts extend for most of the length of each strand.

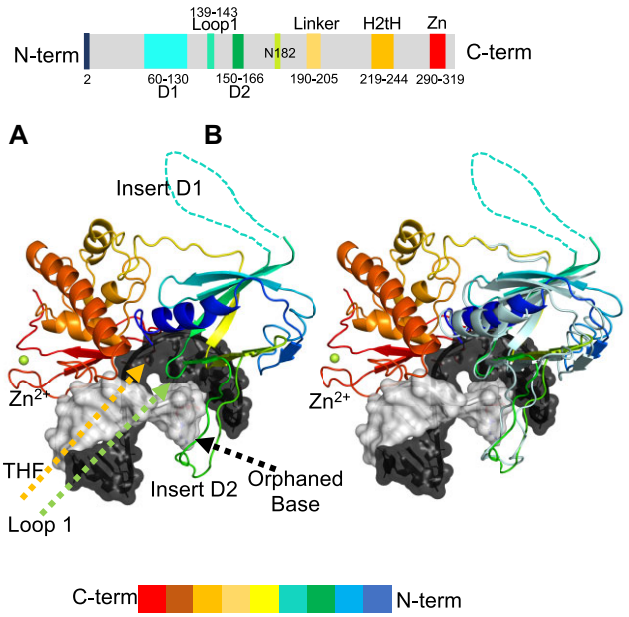


Figure 2. Overall structure of NEIL2 in complex with AP-site containing DNA. Crystal structure of *Monodelphis domestica* NEIL2 (MdoNEIL2) with dsDNA containing an AP-site analog. (**A**) NEIL2 is colored blue to red from the N- to C-terminus, the damaged-containing DNA strand in black and complementary strand in grey. The NEIL2 large insert D1 removed for crystallization is indicated by a dashed line (cyan) and insert D2 is indicated (green). (**B**) The pale blue model is a second conformation of the N-terminal domain in the crystal.

NEIL2 undergoes conformational changes upon binding DNA

As anticipated from the solution structure studies of the unliganded protein, a large interdomain conformational event was required for binding the DNA duplex (Figure 3A) to bring key structural elements into orientation necessary for enzymatic activity (13). The DNA interaction elements also displayed local conformational events necessary to interact with the DNA. The calculated RMSD for the N-terminal domain (residues 2:194) between unliganded and DNA-bound was 0.51 Å with the primary deviations being around residues 47–51, which encompass the conserved catalytic lysine Lys50, the void-filling loop 139:144, and residues 152:164 (insert D2). Additionally, we observe an ordering of the N-terminal residues for catalysis as residues Pro2, Glu3 and Gly4 were disordered in the unliganded structure (Figure 3B). In the N-terminal domain, the loop between β-strands 4 and 5 positions Leu141 and Phe142 into the DNA duplex opposite the abasic site lesion (Figure 4A) while a very short loop connecting β-strands 7 and 8 orients Asn182 to hydrogen bond with the +1 and +2 positions relative to the lesion. The conformation also positions β-strand 8 for backbone to phosphate interactions with the damage-containing strand. Beta strand 8 also contains Arg184 which flanks the phosphate backbone of the +2 position. On the opposite side of the DNA duplex, the complementary strand, in particular the orphaned base opposite the lesion, rests on the surface generated between the two strands of the extended arm of insert 2 (Figure 4B). The orphaned cytosine base is observed flipped out of the DNA duplex rather than being intrahelical. Insertion of the Leu141-Phe142 void-filling loop within the DNA duplex and the extrahelical conformation of the orphaned base combine to

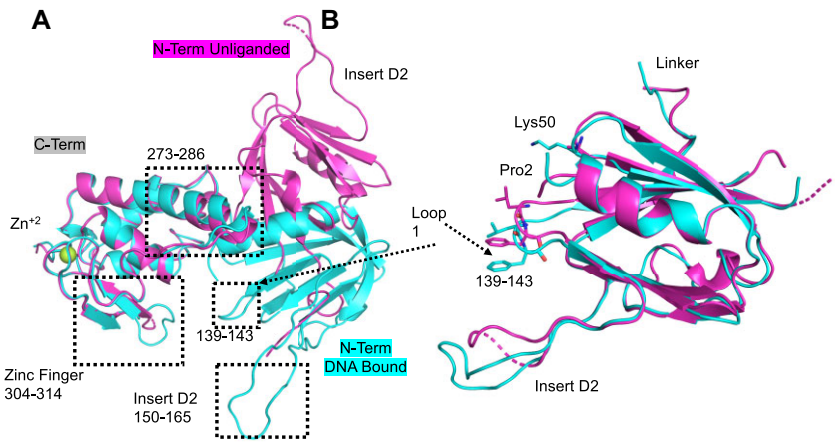


Figure 3. (**A**) Global and local conformational changes are observed in NEIL2 upon binding damaged DNA. Shown are the unliganded MdoNEIL2 (13) (magenta) and DNA-bound (cyan) after least squares superposition of the C-terminal domains with DNA omitted for clarity. Boxed areas indicate local conformational changes between unliganded and DNA-bound. (**B**) Local conformations after least squares superposition of the N-terminal domains.

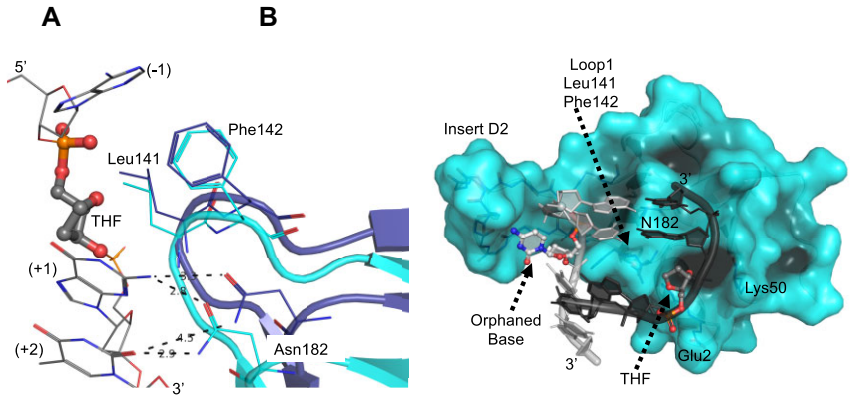


Figure 4. N-terminal domain interactions between NEIL2 and DNA. (**A**) NEIL2 has a single loop containing Leu141-Phe142 for filling the void generated by the everted lesion. It also contains a short loop containing Asn182 in coordinating position for bases immediately 3′ to the damage site. Both N-terminal domain conformations converge at the AP-site, as do active site residues. Shown in cyan and dark blue are the two N-terminal domain conformations. The complementary strand would be in the foreground and has been omitted for clarity. (**B**) The unique insert 2 in NEIL2 coordinates the flipped out orphaned base of the dsDNA complementary strand. The data indicate that this insert is flexible.

distort the neighboring base pair tilt at the −1 position (5′-) relative to the lesion.

The C-terminal domain of NEIL2 provides two regions of contact to the DNA. The H2tH and zinc finger motifs partake in highly conserved interactions to the non-bridging oxygens of the phosphate backbone flanking the site of damage (Figure 5A). Asn234 and Tyr289 contact the lesion phosphate via the H2tH motif while backbone contacts are observed via the zinc finger. Another conserved residue typically associated with lyase activity is an arginine from the zinc finger, Arg313 in the case of NEIL2. Residues 311 through 314 provide backbone mediated contacts to the damage strand backbone flanking the lesion, although the sidechain of Arg313 is rotated away from the lesion. This is likely an influence from the crystal packing as it is forming a salt bridge with Glu326 from a symmetry mate. The coil connector preceding the helix-two-turn-helix (Ht2H) residue, Tyr289, provides a clamp-like feature for grasping the −2 and −3 positions of the complementary strand (Figure 5B). The sequence Pro283-Leu284-His285 is identical between MdoNEIL2 and HsaNEIL2 and creates a cup around the −2 position with the main chain amide from Leu284 along with the amide from His286 coordinating the non-bridging phosphate oxygens. The calculated RMSD of the C-terminal domain (residues 195:325) between unliganded and DNA-bound NEIL2 was 0.45 Å for 729 atoms, with most of the deviation originating from residues 274:286 and 304:314.

Comparison of NEIL2 to NEIL1 and NEIL3 reveals differences and an unexpected similarity

Upon encountering lesion-containing DNA, glycosylases of the Fpg/Nei family position the lesion in the cleft between the N- and C-terminal domains such that the damage-containing strand backbone is coordinated by conserved residues from the H2tH and zinc finger motifs of the C-terminal domain, and the key active site residues Pro2, Glu3 and a conserved lysine in the loop between β-strands 2 and 3 in the N-terminal domain, as shown in Figure 6. NEIL1 utilizes a two-loop invasion mechanism of the DNA duplex. The lesion is everted from the DNA duplex to be accessed by active site residues. The loop connecting β-strands 4 and 5 contains a hydrophobic residue (usually a leucine or a methionine; M81 in human NEIL1), which fills the void created by the ever-

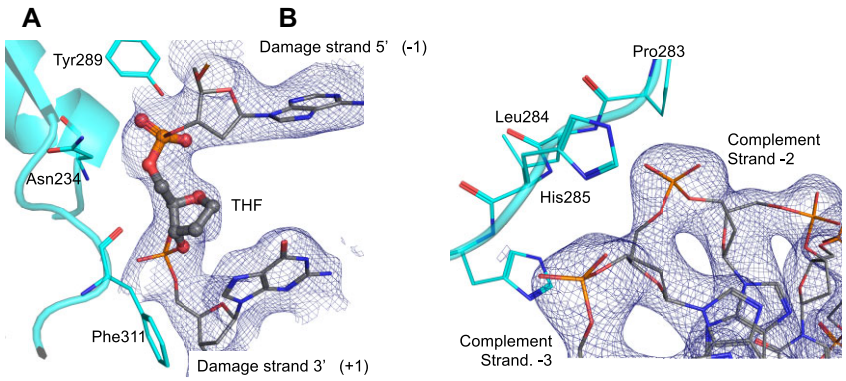


Figure 5. Interaction between the C-terminal domain and the damaged strand (**A**) or complementary strand (**B**). The C-terminal domain and zinc-finger arm bind the backbone of the damaged strand. All residues are conserved between MdoNEIL2 and the human enzyme, except Phe311, which is a conservative change to a bulky hydrophobic leucine residue. The phosphate backbone of the complementary strand (−2 and −3 positions) is bound by His285 and His286 sidechains along with the amide nitrogen of Leu284 and His286. Outside of the orphaned base contact with the base of insert 2 shown in Figure 4B, this is the only significant contact with the complementary strand. Shown is the 2.08 Å resolution MIR (multiple isomorphous replacement) phased map after density modification contoured at 1 σ.

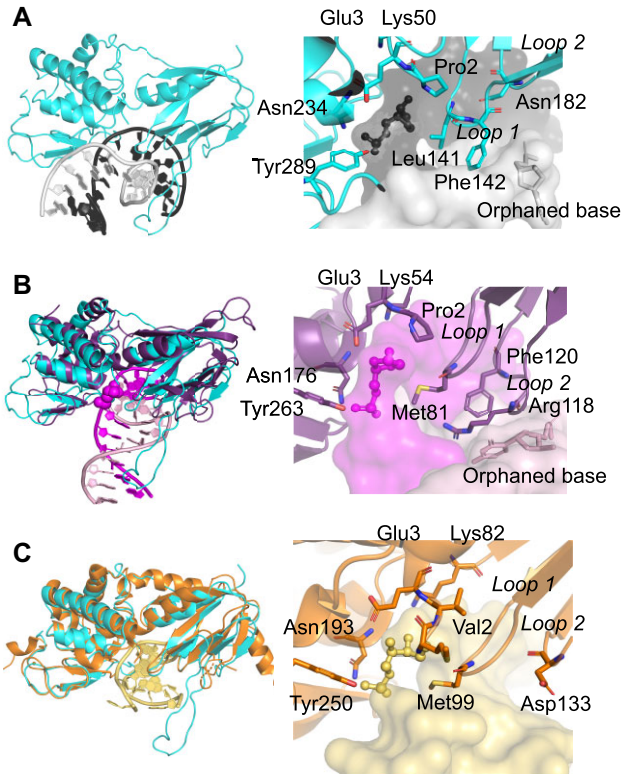


Figure 6. Comparison of NEIL1, 2 and 3 in complex with AP-site containing DNA. NEIL1 and NEIL2 orient dsDNA differently but maintain a similar lesion position in the active site while NEIL3 is bound to a ssDNA hairpin with partial dsDNA in the active site. (**A**) Shown is NEIL2 (cyan), the AP-site containing strand (black) and complementary strand (grey). (**B**) Shown is NEIL1 (purple) (PDBID 5ITT) (64), its AP-site containing strand (maroon) and complementary strand (pink). AP-sites for both are highlighted (black or magenta spheres). Zoom panels show NEIL1 orthologs have two void-filling loops, including one with a residue (R118) to stabilize the intrahelical orphaned base whereas NEIL2 has the shorter Asn182 containing loop. (**C**) Comparison to NEIL3 (orange) in complex with AP-site trapped ssDNA hairpin (PDBID 7Z5A (42)). Asp133 is equivalent to Asn182 in NEIL2 and was proposed to contribute to ssDNA requirement of NEIL3 via charge repulsion. NEIL3 also has a shortened form of this loop similar to NEIL2.

sion of the damaged base. A second loop, between β-strands 7 and 8, provides two residues (usually an aromatic residue and an arginine; Phe120 and Arg118). The aromatic residue is inserted close to the orphaned base and has been termed 'wedge residue' for its role in sensing DNA lesions (38,39). The arginine residue contacts and stabilizes the intrahelical orphaned base. Inspection of the NEIL2 complex shows a striking, and unexpected, similarity to NEIL1. NEIL2 contains only one void-filling loop because the second loop is truncated in both NEIL2 and NEIL3 glycosylases (40,41). In NEIL2, two hydrophobic void-filling residues, Leu141 and Phe142, are located on one loop, yet cover a similar footprint in the DNA duplex void as the residues that reside on two separate loops in NEIL1. Notably absent in NEIL2 is a residue similar to Arg118 in NEIL1 to stabilize the intrahelical orphaned base. In NEIL2, the orphaned base is extrahelical and is instead contacted by insert D2 primarily via backbone contacts. Residue Asn182 in the truncated second loop is within hydrogen bonding with the + 1 and + 2 bases of the damaged strand (Figures 4A and 6).

A comparison to the structure of NEIL3 covalently trapped to an AP-site shows that the truncated second void-filling loop harbors an aspartate (Asp133) in the position analogous to Asn182 in NEIL2 (Figure 6C) (42). The Asp133 position in human NEIL3 is one of two acidic residue substitutions that were proposed earlier to contribute to its strict ssDNA preference through electrostatic repulsion with the DNA phosphate backbone (40). While the NEIL3 complex structure provides global insight into the general orientation of the DNA interacting with the enzyme, it is important to note that the nature of the cross-linked complex is likely artifactual, as the chemical trapping reaction to the AP-site involved the N-terminal Met1 (42). Fpg/Nei glycosylases require post-translational processing by methionine aminopeptidase (MAP) to remove the N-terminal methionine, making the N-terminus, and active site nucleophile, residue number 2 (Pro2 in NEIL1 and NEIL2; Val2 in NEIL3). The MAP processing resulting from endogenous *E. coli* expression is sufficient to cleave Met1 in most DNA glycosylases, but NEIL3 necessitates MAP supplementation in order to obtain an active enzyme (43,44). This finding is congruent with the very long time course required

for the chemical crosslinking reaction to reach completion when the N-terminal methionine is still present (42).

Lesion specificity of WT NEIL2 on ssDNA and dsDNA

The WT HsaNEIL2 and MdoNEIL2 enzymes were tested for activity against a panel of lesions previously shown to be substrates for the NEIL1 glycosylase (28). The glycosylase activity of the two mammalian NEIL2 enzymes was assessed under the single turnover conditions (STO, [E] > [DNA]) with OG, Gh, *R*-Sp, *S*-Sp, Tg, DHT, Ug, 5-OHU, 5-hmU, 8OI and 5-OHC (Figure 7) localized within a 30-nucleotide single strand or duplex across from pre-replicative base pairing context by monitoring strand scission induced by base quenching at the enzyme-produced AP site (Supplementary Figure S4). In the same oligonucleotide sequence, we also analyzed the AP lyase activity of the NEIL2 constructs, by monitoring enzyme-induced strand scission at an AP site substrate generated by removal of uracil from the U-containing oligonucleotide by uracil–DNA glycosylase (Supplementary Figure S4). Of the damaged base substrates analyzed, only Gh, *R*- and *S*-Sp, 5-OHU and Ug were removed to an appreciable extent, with more base removal activity with ssDNA over duplex substrates. For the most part, both enzymes removed lesions similarly, with the exception of the excision of the Sp isomers only by HsaNEIL2. Notably, previous work has observed weak removal activity of DHU and Tg by NEIL2 at high enzyme concentration (13). However, most strikingly, even in these qualitative assays, the AP lyase activity was significantly greater than the glycosylase activity for all the lesions tested.

Based on these initial results, a more thorough kinetic analysis of the activity of HsaNEIL2 and MdoNEIL2 towards the AP site, Gh, Ug and 5-OHU substrates in single strand, and AP:C, Gh:C, Ug:G and 5-OHU:G in duplex was performed (Figure 7, Table 2). We first performed active site titration assays to evaluate the stability of the NEIL2 enzymes in assay buffer and conditions, and to compare the relative stability of the two enzymes. Stability was assessed by pre-incubating the enzyme in the reaction buffer with AP-site containing single stranded DNA at time-points ranging from 0 to 60 min, followed by gauging of the active fraction via the burst amplitudes in MTO reaction at each time point. We observed that HsaNEIL2 exhibited a significant loss of activity under these conditions within 15 min, while this was not the case with MdoNEIL2 (Supplementary Table S4). To mitigate potential complications due to enzyme activity loss, analysis of the STO production curves focused on the first 10 min of the enzymatic reaction (Figure 7). Overall, the trends observed mirrored those from the qualitative assays with NEIL2 enzymes exhibiting a strong preference towards the AP site substrate in both ssDNA and dsDNA, followed by base lesion substrates containing Gh, 5-OHU, then Ug. In single-strand DNA, both enzymes behaved almost identically, showing similar amounts of product formed at 10 min (Figure 7A) and at similar rates (Table 2). In duplex DNA, the human enzyme showed reduced amount of product formed and lower rates of product formation for the AP:C and Gh:C, while 5-OHU:G and Ug:G showed minimal product formation. In contrast, MdoNEIL2 was observed to be largely insensitive to lesion context, forming similar amounts of product, albeit at slower rates for 5-OHU:G and Ug:G. Both HsaNEIL2 and MdoNEIL2 enzymes exhibited nearly identical slow turnover rates under MTO conditions with the preferred substrate ssDNA-AP (0.001 and 0.006 min^{-1}, respectively) suggestive of slow release of the cleaved AP site product (Supplementary Table S3).

Activity of NEIL2 variants with ssDNA and dsDNA containing an abasic site

As the AP site was the preferred lesion for NEIL2, we analyzed the AP lyase activity under MTO and STO conditions using MdoNEIL2 constructs informed by the DNA-bound structure: ΔD1, ΔD2 and Asn182Asp. The MdoNEIL2-ΔD1 construct lacks the larger disordered insert region, while MdoNEIL2-ΔD2 has the segment observed to interact with the complementary strand deleted. We also examined MdoNEIL2 Asn182Asp, based on the observation that Asn182 interacts with DNA, and that there is an Asp at this position in NEIL3, which was predicted to disfavor binding to ds DNA via electrostatic repulsion (40). We hypothesized that the Asn182Asp variant would exhibit a decreased propensity to bind ds DNA.

To begin, we assessed the % activity of each enzyme under MTO conditions with the preferred substrate, ssDNA-AP (Table S3). Compared to WT MdoNEIL2, MdoNEIL2ΔD1 showed a markedly higher active fraction, while MdoNEIL2ΔD2 showed a significantly lower active fraction. Using the active fractions obtained from the MTO experiments, we then performed STO reactions with the AP site in ss and ds DNA contexts (Figure 8). While all the enzymes behaved almost identically in ssDNA, showing efficient AP lyase activity, the activity of several variants was severely reduced when the AP site was localized within duplex DNA. With MdoNEIL2ΔD2, a significantly reduced amount of product was observed (Figure 8B), as well as a 12-fold reduction in reaction rate (Table 3). Similarly, with the MdoNEIL2 Asn182Asp construct an even more significant reduction in the reaction rate and overall product formed was observed. The WT MdoNEIL2 and the MdoNEIL2ΔD1 showed no change in activity between dsDNA and ssDNA. The observation that MdoNEIL2ΔD1 had the highest % active fraction, greater than WT MdoNEIL2 (Supplementary Table S3), suggests that the deletion of this larger insert may stabilize an active conformation *in vitro* allowing for improved overall catalytic performance. The reduction in AP lyase activity in duplex DNA by deleting unique insert D2, and replacing Asn182 with Asp, supports the hypothesis that D2 and neighboring residues play a critical role in allowing NEIL2 to function on duplex DNA.

Discussion

We present here the first crystal structure of a mammalian NEIL2 enzyme complexed with DNA. This structure solution required a combination of MIR, SAD, and molecular replacement methods. As expected from the unliganded structure, a large conformational change occurs when the enzyme binds DNA in order to form a catalytically competent complex. The new structure also revealed smaller local conformational changes to accommodate binding to duplex DNA in a binding mode favoring the 3′ backbone coordination for each strand (Supplementary Figure S3). We tested a wide array of oxidized lesions: OG, Gh, Sp, Tg, DHT, Ug, 5-OHU, 5-hmU, 8OI and 5-OHC, along with AP sites. We also examined protein variants informed by the crystal structure and

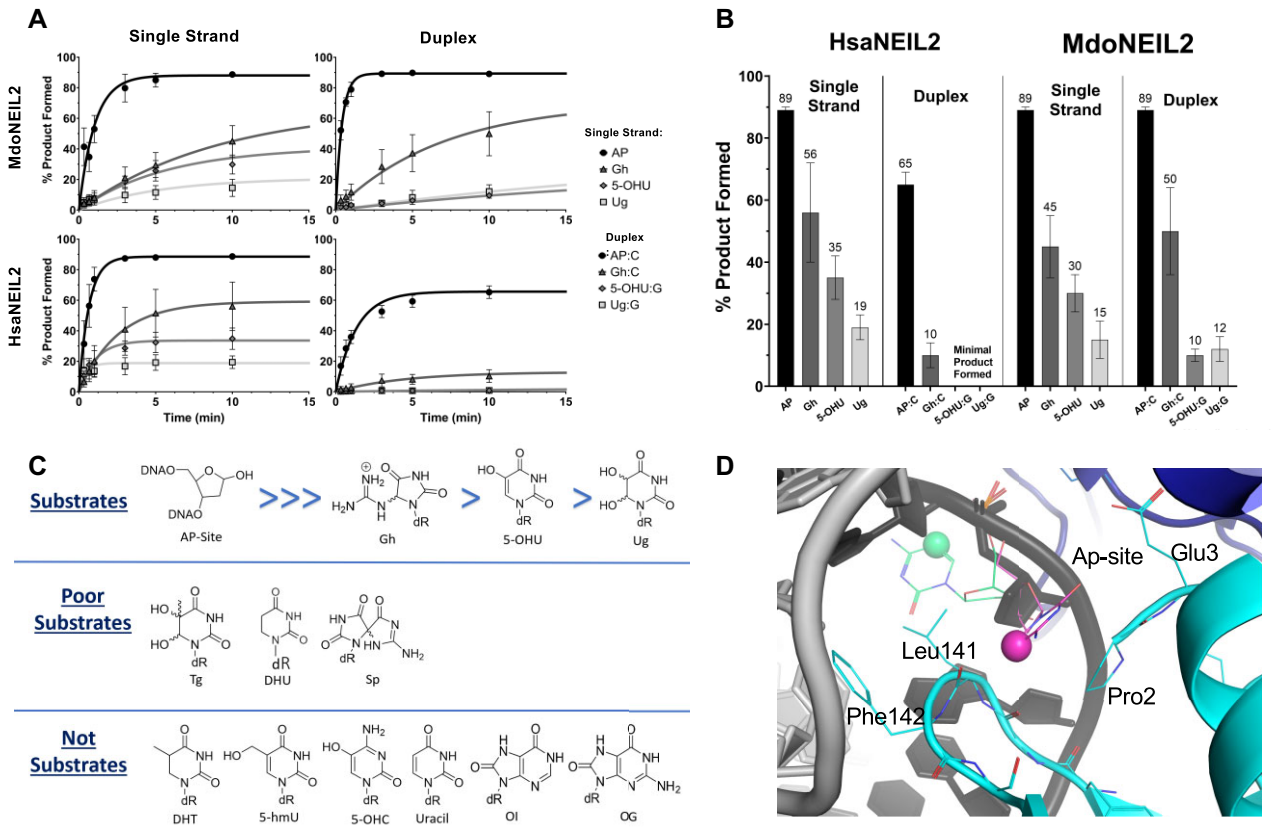


Figure 7. Base excision versus AP lyase activity in ssDNA and dsDNA of WT full-length HsaNEIL2 and WT full-length MdoNEIL2 under STO conditions. (**A**) STO production curves for WT HsaNEIL2 (top) and WT MdoNEIL2 (bottom) processing AP, Gh, 5-OHU and Ug substrates in single strand (left) and duplex (right) DNA. The plots show that amount product formed over the course of 10 min, and the data was fit to single exponential (see Materials and methods). Duplex lesion substrates were placed across from pre-replicative base pairs (dG for 5-OHU and Ug, and dC for Gh), and AP was placed across from dC; data were fitted to a single-exponential equation (see methods), and represent averages of at least three trials from separate enzyme aliquots with three sub-replicates for each aliquot. (**B**) Relative processing of substrates was gauged by plotting the % product formed at 10 min by WT HsaNEIL2 (left) and WT MdoNEIL2 (right) with AP, Gh, 5-OHU and Ug substrates in single strand and duplex DNA. Bar graphs were created using GraphPad prism 9. (**C**) Represents structure activity relationships (SAR) for NEIL2 defining substrates versus non-substrates. (**D**) Shown is the active site of MdoNEIL2 bound to the AP- site (dark grey) with a cytosine modeled in two orientations. The sphere indicates the C5 position. Also shown are the complementary strand (light grey), N-terminal domain (cyan), C-terminal domain (blue). The magenta cytosine is modeled into the catalytic pocket via the compressed minor groove whereas the green cytosine is modeled flipped out towards the major groove.

Table 2. AP lyase and glycosylase activity of NEIL2 constructs under STO conditions

	WT full-length HsaNEIL2			
	Single-strand DNA		Duplex DNA	
	k_g *or* k_l (min^{-1})[a]	Product@10 min (%)[b]	k_g *or* k_l (min^{-1})[a]	Product @ 10 min (%)[b]
AP	1.6 ± 0.6*	89 ± 1	0.8 ± 0.2*	65 ± 4
Gh	0.4 ± 0.1	56 ± 16	0.3 ± 0.1	10 ± 4
5-OHU	1.0 ± 0.3	35 ± 7	NA	NA
Ug	>2	19 ± 4	NA	NA
	WT full-length MdoNEIL2			
	Single strand		Duplex	
	k_g *or* k_l (min^{-1})[a]	Product@10 min (%)[b]	k_g *or* k_l (min^{-1})[a]	Product@10 min (%)[b]
AP	1.2 ± 0.3*	89 ± 2	2.4 ± 0.4*	88 ± 2
Gh	0.11 ± 0.04	45 ± 10	0.2 ± 0.1	50 ± 14
5-OHU	0.2 ± 0.1	30 ± 6.2	0.04 ± 0.01	10 ± 2
Ug	0.2 ± 0.2	15 ± 6	0.04 ± 0.02	12 ± 4

[a]Rate constants derived from fitting of data from single-turnover glycosylase or AP lyase assays with HsaNEIL2 and MdoNEIL2 on AP site and Gh, 5-OHU, and Ug in single strand and duplex DNA. Gh and AP were placed across form C in duplex DNA, while 5-OHU and Ug were placed across from G. Production curves were fit with equation 1 (see Materials and methods), where k_g *or* k_L is the rate representing the rate constant for the AP lyase reaction on the AP substrate (emphasized by '*') or glycosylase reaction with the Gh, 5-OHU and Ug substrates. Values listed are the average and standard deviation of time courses collected from 3 separate enzyme aliquots, with three trials for each aliquot.
[b]Indicates the amount of product formed at this time point.

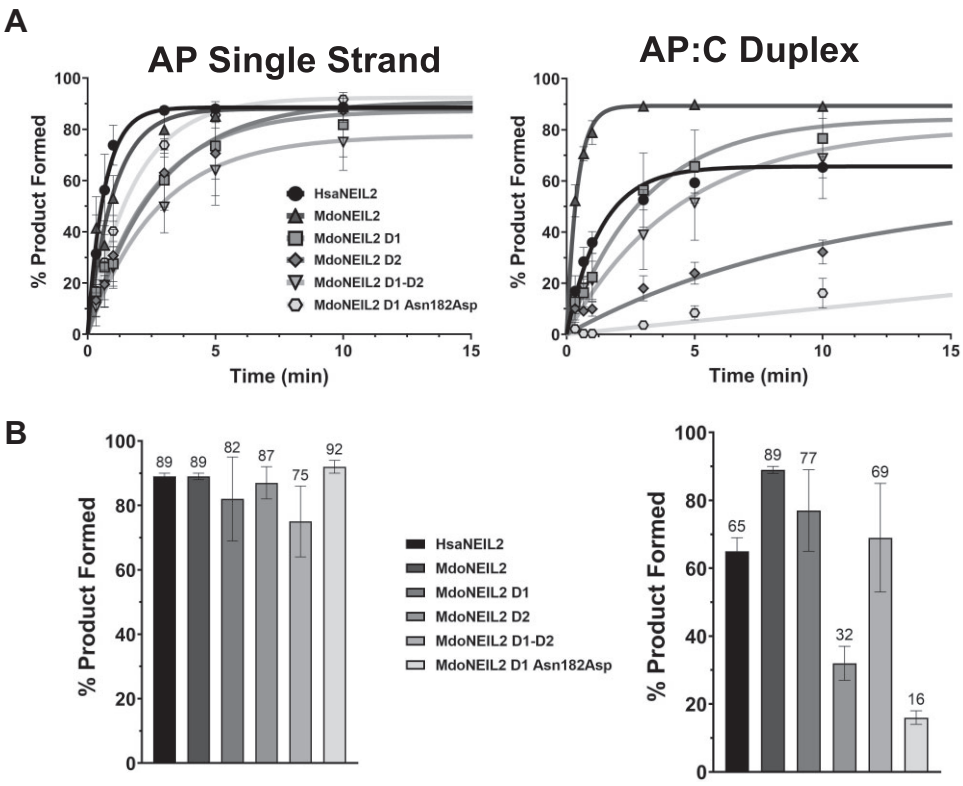


Figure 8. Impact of insert 2 deletion on NEIL2 dsDNA activity under STO conditions. (**A**) AP lyase activity with WT full-length HsaNEIL2 (circles), WT full-length MdoNEIL2 (upward triangle), MdoNEIL2ΔD1 (square), MdoNEIL2ΔD2 (rhombus), MdoNEIL2ΔD1 Asn182Asp (hexagon) in single strand (left) and duplex across from dC (right). Curves represent the amount of product formed over the course of 10 minutes at various time points under STO conditions; data were fitted to a single exponential equation (see methods), and represent averages of trials of three separate enzyme aliquots, with three sub-replicates for each aliquot. Note, reactions were analyzed within the first 10 minutes to prevent interference due to loss of enzyme activity during the reaction. (**B**) The total amount of product produced at 10 min for AP lyase activity in single-strand (left) and duplex DNA across from dC (right). Bar graphs represent the average and standard deviation.

Table 3. AP lyase activity of NEIL2 constructs lacking flexible inserts under STO conditions

	Single-strand DNA		Duplex DNA	
	k_l (min^{-1})[a]	Product@10 min (%)[b]	k_l (min^{-1})[a]	Product@10 min (%)[b]
WT Full-length HsaNEIL2	1.6 ± 0.6	89 ± 1	1.2 ± 0.3	65 ± 4
WT Full-length MdoNEIL2	0.7 ± 0.2	89 ± 1	2.4 ± 0.4	89 ± 1
MdoNEIL2ΔD1	0.4 ± 0.2	82 ± 13	0.3 ± 0.1	77 ± 12
MdoNEIL2ΔD2	0.4 ± 0.2	88 ± 5	0.1 ± 0.04	32 ± 5
MdoNEIL2ΔD1 Asn182Asp	0.6 ± 0.3	92 ± 2	0.02 ± 0.01	16 ± 6

[a]Rates of NEIL2 enzyme AP lyase activity at AP-sites in single-strand and duplex DNA across from dC. The rate of the lyase reaction (k_l) was determined using an hour-long time course and fitting to Equation 1 (see methods). Values listed are the average and standard deviation of time courses collected from three separate enzyme aliquots, with three trials for each aliquot.
[b]The values reflect the amount of product form at the 10 min time point.

tested their activity on AP site-containing DNA, the preferred substrate.

Of the lesions tested, only the AP site, Gh, *R*-Sp, *S*-Sp, 5-OHU, and Ug were identified as substrates for NEIL2. Previous studies have shown that NEIL2 has a weak activity towards removal of Sp, with more robust removal of Gh from both ssDNA and duplex, consistent overall with our findings (45). A previous study also suggested that under some conditions, NEIL2 exhibits a limited ability to remove DHU and Tg in ssDNA (13). By comparing the different structural properties of each lesion identified as a substrate, to those identified to be poor substrates, or not substrates at all, we can glean structure activity relationships (SAR) for preferred NEIL2 base-lesion substrates (Figure 7). All of the lesion substrates that have been identified have a T- or U-like Watson-Crick face. This is consistent with the NEIL2 base excision activity on 5-OHU, but not 5-OHC. This feature alone does not explain NEIL2 specificity, however, since U, T, DHT and 5-hmU are not substrates. The addition of a hydroxyl group at the C5 position of U in 5-OHU provided for NEIL2 base removal activity over U, while substitution of OH with a methyl, as present in T, did not. Similarly, we found Ug was a better substrate than Tg, and that DHU was a poor substrate, while DHT was not processed by NEIL2. These findings suggest that a polar substituent at C5 is preferred, while a hydrophobic methyl group at C5 is disfavored. This may be due to a

combination of contacts favoring polar substituents and minimizing steric clashes.

In addition, structural features of the base will impact its inherent lability, as indicated by N1-H acidity. Lesion structure also impacts the preferred tautomeric form and N3 proton affinity, features that may also influence steps of base excision catalysis (28). Notably, calculations indicate a slightly higher N1-H acidity of 'U' derived over 'T' lesions (28,46). Of the base lesions examined, Gh is the best substrate for NEIL2, consistent with it being positively charged and therefore poised to be a good leaving group; these features are also reflected in gas-phase calculations as high N1-H acidity (28). Notably, while Gh was a good substrate in both DNA contexts, Sp was not, being only removed from ssDNA by HsaNEIL2. In addition to the lack of Sp harboring a positive charge, steric clashes may arise due to its locked chiral center between the two rings. However, somewhat surprising, no clear difference was observed between the two isomers. These results are in contrast to NEIL1 that removes both Gh and Sp, and distinguishes between the two Sp diastereomers. OG was not a substrate for NEIL2, like the other NEILs (43,47). However, unlike NEIL1, NEIL2 did not show improved activity for OI, a derivative of OG lacking the 2-amino group, further underscoring differences between the two glycosylases. Indeed, NEIL1 has a much more robust glycosylase activity on a broad array of substrates and prefers base lesions in duplex contexts compared to NEIL2. These factors are consistent with hypotheses that NEIL2 activity is much more context-dependent, in DNA structures formed during transcription rather than NEIL1 that is more base and duplex specific due to a primary role in global genome repair.

A challenge with the interpretation of lesion preference for NEIL2 is that the C5 position of pyrimidines is located in the major groove and glycosylases predominantly interrogate DNA via the minor groove using a void-filling loop (containing Leu141 and Phe142 for NEIL2; Figure 6). During this process, the major groove surrounding the lesion is distorted and compressed as the lesion is flipped out of the duplex. For NEIL1, the orphaned base remains intrahelical, stabilized by Arg118 while for NEIL2 the orphaned base is extrahelical and coordinated by backbone contacts within insert D2 (Figure 4). Based on the orientation one can anticipate that the orphaned base extrusion trajectory is via the major groove, *i.e.* pushed out by the loop. If so, one would expect the lesion to follow a similar trajectory which does not put the base in the active site. However, the 3′ orientation bias diagramed in Supplementary Figure S3 would permit the lesion to flip towards the catalytic residues of the N-terminus. The substitution preference at the C5 position can be interpreted as follows: there is an energy barrier to opening the damaged base pair to flip out the lesion that is lowered by polar substitution at the C5 position. An alternative is that interactions or clashes generated by the C5 substitution as it is extruded out of the duplex (Figure 7D.) dictate reactivity. The DNA is bent, promoting extrusion of the lesion as is typical for this family of glycosylases. Modeling of a base at the abasic site in the structure was thus limited to the extrusion trajectory. The modeling of the lesion site in two orientations supports both an in and out trajectory for the base, with the out being major groove and towards the zinc finger and no clashes with the modelled cytosine. This proposed trajectory, however, rotates the glycosidic bond beyond the reach of the catalytic residue Pro2. The inward trajectory puts the base and glycosidic bond nearest Pro2, with the only apparent clash with the C5 position being the backbone carbonyl located between Leu141 and Phe142 of Loop 1. The 3′-favored polar orientation of the duplex for NEIL2 (Supplementary Figure S3) may support major groove trajectory of the orphaned base but minor groove trajectory of the lesion. For the lesion to extrude into the active site, the prime interaction candidates for the C5 position are the backbone carbonyls of the void-filling loop between residues 139–143.

As suggested from our previous study and supported here with a wider array of lesion substrates, NEIL2 is not a particularly robust glycosylase but is a very good AP lyase. This poses the question as to what is the true substrate for NEIL2. A number of studies have implicated NEIL1 and NEIL2 in processing AP sites generated from pathways outside of canonical BER of oxidized lesions. AP sites are produced in abundance from BER and are primarily processed by APE1 (48). However TET-mediated cytosine demethylation (49) during gene expression regulation and APOBEC3 cytosine deamination (50) in the immune response converge on TDG/UNG dependence for removal of the resulting dU intermediate, thereby generating AP sites (4,51) as well. These processes are among the more intriguing potential functions for NEIL2 as knockout mouse models do not exhibit global increases in mutagenesis load or increased cancer risk (52) but more localized damage increases to transcribed regions of the genome and elevated inflammatory phenotypes. Additionally, both of these processes are transcription-coupled, which is a hallmark proposal for NEIL2 core activity (53,54). These inflammatory phenotypes for NEIL2 deficient models provide a partial mechanistic link to *Helicobacter pylori*-associated gastric cancer, where regulation of NEIL2 expression is perturbed (55). The prototypical gastric cancer signature contains microsatellite instability and defects through the mismatch repair pathway (56). Indeed, inefficient BER glycosylases can promote mismatch accumulation through two pathways: unrepaired lesions that mis-pair during DNA replication and elevated BER activity increasing the dependence on a lower fidelity polymerase like DNA pol β (57).

The consideration that NEIL2 processes AP sites and oxidized base lesions in DNA in a structure- and even sequence-specific manner would suggest a mutational or damage signature for NEIL2 knockout or variants that has yet to be identified. Interestingly, both the APOBEC3 deamination and TET-mediated demethylation processes have sequence and structure specific contexts often containing loops and bubbles that would fit into the NEIL2 repertoire (58–61). Recent studies have attempted to investigate positional dependence of lesions within a defined bubble for both NEIL1 and NEIL2 (62). For NEIL1, an increase in activity was observed when 5-OHU is at the ends of the bubble nearing the duplex, while for DHU a decrease when at the 2 position and an increase in activity for the 10 position were observed in the 12 base-bubble. NEIL2 showed fluctuations but no statistically significant positional dependencies across the bubble. However, in light of the stability limitations shown here for the human enzyme and the marked preference for the AP sites over oxidized base lesions, revisiting these experiments under different conditions or using MdoNEIL2 substitution may reveal greater details. The distinction in bubble position preference for NEIL1 compared to NEIL2 could be in part due to the truncation of loop 2 in NEIL2. As discussed above, this loop contains Arg118 for coordination of the orphaned

base in NEIL1. The short form of this loop in NEIL2 instead contains Asn182, seen coordinating the bases just 3′ to the AP site on the damage strand rather than the complementary strand. Asn182 is equivalent to Asp133 in NEIL3, an enzyme that strongly favors ssDNA. Mutation of Asn182 to Asp in NEIL2 showed a marked reduction in activity with dsDNA (Figure 8), supporting the original hypothesis that the negative charge from Asp could impart an electrostatic repulsion effect on duplex DNA. Unlike NEIL1, the orphaned base opposite the lesion site is not retained inside the helix due to the absence of loop 2 and a residue equivalent to Arg118 (Figure 6B) needed to stabilize the unpaired base within the duplex. Instead, it is flipped out of the DNA duplex and sitting in a pocket created at the base of insert 2 in NEIL2. Insert 2 is 12 residues in length with 10 residues being identical between MdoNEIL2 and HsaNEIL2 and contains 6 positively charged Arg/Lys residues. We anticipated that the conservation of these residues imparted some specific interactions with the DNA but instead observed that the side chains were largely flexible and not interacting with the dsDNA. Instead, backbone contacts provide the surface for the coordination of the orphaned base. When truncating insert 2 to just a Gly-Ser-Gly-Ser-Gly linker (MdoNEIL2ΔD2 construct), a significant decrease in activity towards dsDNA was observed (Figure 8) whereas truncation of insert 1 (MdoNEIL2ΔD1 construct) showed no such effect, supporting the structural interpretation. It is likely that residues in insert 2 impact activity or aid in binding to certain bubble structures or yet-to-be determined sequence context.

The dynamic nature of NEIL2 is likely to impart some functional advantage pertaining to activity on a particular lesion, sequence context, or DNA structure. These structural dynamics also need to be taken into consideration when interpreting the impact of variants during the unliganded, DNA-bound, or enzymatic transition states of NEIL2. Prediction of the impact on NEIL2 activity of certain variants may not be defined by their position at the unliganded or DNA bound points but rather in their roles in the conformational transition between the two states and may likely be substrate dependent. A similar consideration should be taken when considering structure-based predictions for inhibitor screening of a glycosylase like NEIL2, as the ideal target mode may lie at unique points along the conformational path.

Data availability

Processed data including phase information and structure for the selenomethionine-containing model have been deposited into the Protein Databank (rcsb.org) under PDB ID 8TH9 along with the processed data for the native, gold, barium, cobalt and low energy sulfur/phosphate native.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

The authors would like to thank Drs Susan Tsutakawa and James Holton of the SIBYLS group for helpful discussions. We also thank Drs John Patrick Rogers and Amelia Manlove for synthesis of Ug and 8OI-containing oligonucleotides, respectively. X-ray diffraction data were collected at GM/CA@APS which has been funded by the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006, P30GM138396). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. The Eiger 16M detector at GM/CA-XSD was funded by NIH grant S10 OD012289.

Funding

National Institutes of Health/National Cancer Institute program [P01-CA098993 to S.S.W., S.D.]; National Institutes of Health/National Institute for General Medical Sciences [R01-GM143557 to S.S.D.]; B.E.E. was supported by National Institutes of Health/National Cancer Institute [R50 CA233185]. Funding for open access charge: Institution issued purchase order, grant funds.

Conflict of interest statement

None declared.

References

1. Chatterjee,N. and Walker,G.C. (2017) Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.*, **58**, 235–263.
2. Shen,L., Song,C.X., He,C. and Zhang,Y. (2014) Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu. Rev. Biochem.*, **83**, 585–614.
3. Drohat,A.C. and Coey,C.T. (2016) Role of base excision "repair" enzymes in erasing epigenetic marks from DNA. *Chem. Rev.*, **116**, 12711–12729.
4. Sarker,A.H., Cooper,P.K. and Hazra,T.K. (2021) DNA glycosylase NEIL2 functions in multiple cellular processes. *Prog. Biophys. Mol. Biol.*, **164**, 72–80.
5. Jacobs,A.L. and Schar,P. (2012) DNA glycosylases: in DNA repair and beyond. *Chromosoma*, **121**, 1–20.
6. Osorio,A., Milne,R.L., Kuchenbaecker,K., Vaclova,T., Pita,G., Alonso,R., Peterlongo,P., Blanco,I., de la Hoya,M., Duran,M., *et al.* (2014) DNA glycosylases involved in base excision repair may be associated with cancer risk in BRCA1 and BRCA2 mutation carriers. *PLoS Genet.*, **10**, e1004256.
7. Mullins,E.A., Rodriguez,A.A., Bradley,N.P. and Eichman,B.F. (2019) Emerging roles of DNA glycosylases and the base excision repair pathway. *Trends Biochem. Sci.*, **44**, 765–781.
8. SenGupta,T., Palikaras,K., Esbensen,Y.Q., Konstantinidis,G., Galindo,F.J.N., Achanta,K., Kassahun,H., Stavgiannoudaki,I., Bohr,V.A., Akbari,M., *et al.* (2021) Base excision repair causes age-dependent accumulation of single-stranded DNA breaks that contribute to Parkinson disease pathology. *Cell Rep.*, **36**, 109668.
9. Dizdaroglu,M. (2003) Substrate specificities and excision kinetics of DNA glycosylases involved in base-excision repair of oxidative DNA damage. *Mutat. Res.*, **531**, 109–126.
10. Brooks,S.C., Adhikary,S., Rubinson,E.H. and Eichman,B.F. (2013) Recent advances in the structural mechanisms of DNA glycosylases. *Biochim. Biophys. Acta*, **1834**, 247–271.
11. Faucher,F., Doublie,S. and Jia,Z. (2012) 8-oxoguanine DNA glycosylases: one lesion, three subfamilies. *Int. J. Mol. Sci.*, **13**, 6711–6729.
12. Hazra,T.K., Kow,Y.W., Hatahet,Z., Imhoff,B., Boldogh,I., Mokkapati,S.K., Mitra,S. and Izumi,T. (2002) Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *J. Biol. Chem.*, **277**, 30417–30420.
13. Eckenroth,B.E., Cao,V.B., Averill,A.M., Dragon,J.A. and Doublie,S. (2021) Unique structural features of mammalian NEIL2 DNA

glycosylase prime its activity for diverse DNA substrates and environments. *Structure*, **29**, 29–42.
14. Doublié,S. (2007) Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. *Methods Mol. Biol.*, **363**, 91–108.
15. Powell,H.R., Battye,T.G.G., Kontogiannis,L., Johnson,O. and Leslie,A.G.W. (2017) Integrating macromolecular X-ray diffraction data with the graphical user interface iMosflm. *Nat. Protoc.*, **12**, 1310–1325.
16. Clabbers,M.T.B., Gruene,T., Parkhurst,J.M., Abrahams,J.P. and Waterman,D.G. (2018) Electron diffraction data processing with DIALS. *Acta Crystallogr. D Struct. Biol.*, **74**, 506–518.
17. Diederichs,K. (2006) Some aspects of quantitative analysis and correction of radiation damage. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 96–101.
18. Foadi,J., Aller,P., Alguel,Y., Cameron,A., Axford,D., Owen,R.L., Armour,W., Waterman,D.G., Iwata,S. and Evans,G. (2013) Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **69**, 1617–1632.
19. Vonrhein,C., Blanc,E., Roversi,P. and Bricogne,G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.*, **364**, 215–230.
20. Ness,S.R., de Graaff,R.A., Abrahams,J.P. and Pannu,N.S. (2004) CRANK: new methods for automated macromolecular crystal structure solution. *Structure*, **12**, 1753–1761.
21. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G., McCoy,A., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.
22. Cowtan,K. (2010) Recent developments in classical density modification. *Acta. Crystallogr. D Biol. Crystallogr.*, **66**, 470–478.
23. Cowtan,K. (2008) Fitting molecular fragments into electron density. *Acta. Crystallogr. D Biol. Crystallogr.*, **64**, 83–89.
24. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta. Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
25. Imamura,K., Wallace,S.S. and Doublié,S. (2009) Structural characterization of a viral NEIL1 ortholog unliganded and bound to abasic site-containing DNA. *J. Biol. Chem.*, **284**, 26174–26183.
26. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta. Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
27. Waasmaier,D. and Kirfel,A. (1995) New analytical scattering-factor functions for free atoms and ions. *Acta Crystallogr., Sect. A*, **51**, 416–431.
28. Lotsof,E.R., Krajewski,A.E., Anderson-Steele,B., Rogers,J., Zhang,L., Yeo,J., Conlon,S.G., Manlove,A.H., Lee,J.K. and David,S.S. (2022) NEIL1 Recoding due to RNA editing impacts lesion-specific recognition and excision. *J. Am. Chem. Soc.*, **144**, 14578–14589.
29. Bodepudi,V., Shibutani,S. and Johnson,F. (1992) Synthesis of 2′-deoxy-7,8-dihydro-8-oxoguanosine and 2′-deoxy-7,8-dihydro-8-oxoadenosine and their incorporation into oligomeric DNA. *Chem. Res. Toxicol.*, **5**, 608–617.
30. Oka,N. and Greenberg,M.M. (2005) The effect of the 2-amino group of 7,8-dihydro-8-oxo-2′-deoxyguanosine on translesion synthesis and duplex stability. *Nucleic Acids Res.*, **33**, 1637–1643.
31. Yeo,J., Goodman,R.A., Schirle,N.T., David,S.S. and Beal,P.A. (2010) RNA editing changes the lesion specificity for the DNA repair enzyme NEIL1. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 20715–20719.
32. Zhao,X., Krishnamurthy,N., Burrows,C.J. and David,S.S. (2010) Mutation versus repair: NEIL1 removal of hydantoin lesions in single-stranded, bulge, bubble, and duplex DNA contexts. *Biochemistry*, **49**, 1658–1666.
33. Fleming,A.M., Orendt,A.M., He,Y., Zhu,J., Dukor,R.K. and Burrows,C.J. (2013) Reconciliation of chemical, enzymatic, spectroscopic and computational data to assign the absolute configuration of the DNA base lesion spiroiminodihydantoin. *J. Am. Chem. Soc.*, **135**, 18191–18204.
34. Porello,S.L., Leyes,A.E. and David,S.S. (1998) Single-turnover and pre-steady-state kinetics of the reaction of the adenine glycosylase MutY with mismatch-containing DNA substrates. *Biochemistry*, **37**, 14756–14764.
35. Krishnamurthy,N., Zhao,X., Burrows,C.J. and David,S.S. (2008) Superior removal of hydantoin lesions relative to other oxidized bases by the human DNA glycosylase hNEIL1. *Biochemistry*, **47**, 7137–7146.
36. Das,A., Wiederhold,L., Leppard,J.B., Kedar,P., Prasad,R., Wang,H., Boldogh,I., Karimi-Busheri,F., Weinfeld,M., Tomkinson,A.E., *et al.* (2006) NEIL2-initiated, APE-independent repair of oxidized bases in DNA: evidence for a repair complex in human cells. *DNA Repair (Amst.)*, **5**, 1439–1448.
37. Das,S., Chattopadhyay,R., Bhakat,K.K., Boldogh,I., Kohno,K., Prasad,R., Wilson,S.H. and Hazra,T.K. (2007) Stimulation of NEIL2-mediated oxidized base excision repair via YB-1 interaction during oxidative stress. *J. Biol. Chem.*, **282**, 28474–28484.
38. Dunn,A.R., Kad,N.M., Nelson,S.R., Warshaw,D.M. and Wallace,S.S. (2011) Single qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA. *Nucleic Acids Res.*, **39**, 7487–7498.
39. Nelson,S.R., Dunn,A.R., Kathe,S.D., Warshaw,D.M. and Wallace,S.S. (2014) Two glycosylase families diffusively scan DNA using a wedge residue to probe for and identify oxidatively damaged bases. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E2091–E2099.
40. Liu,M., Imamura,K., Averill,A.M., Wallace,S.S. and Doublié,S. (2013) Structural characterization of a mouse ortholog of human NEIL3 with a marked preference for single-stranded DNA. *Structure*, **21**, 247–256.
41. Prakash,A., Eckenroth,B.E., Averill,A.M., Imamura,K., Wallace,S.S. and Doublié,S. (2013) Structural investigation of a viral ortholog of human NEIL2/3 DNA glycosylases. *DNA Repair (Amst.)*, **12**, 1062–1071.
42. Huskova,A., Dinesh,D.C., Srb,P., Boura,E., Veverka,V. and Silhan,J. (2022) Model of abasic site DNA cross-link repair; from the architecture of NEIL3 DNA binding domains to the X-structure model. *Nucleic Acids Res.*, **50**, 10436–10448.
43. Liu,M., Bandaru,V., Bond,J.P., Jaruga,P., Zhao,X., Christov,P.P., Burrows,C.J., Rizzo,C.J., Dizdaroglu,M. and Wallace,S.S. (2010) The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 4925–4930.
44. Liu,M., Bandaru,V., Holmes,A., Averill,A.M., Cannan,W. and Wallace,S.S. (2012) Expression and purification of active mouse and human NEIL3 proteins. *Protein Expr Purif*, **84**, 130–139.
45. Hailer,M.K., Slade,P.G., Martin,B.D., Rosenquist,T.A. and Sugden,K.D. (2005) Recognition of the oxidized lesions spiroiminodihydantoin and guanidinohydantoin in DNA by the mammalian base excision repair glycosylases NEIL1 and NEIL2. *DNA Repair (Amst.)*, **4**, 41–50.
46. Krajewski,A.E. and Lee,J.K. (2021) Gas-phase experimental and computational studies of 5-halouracils: intrinsic properties and biological implications. *J. Org. Chem.*, **86**, 6361–6370.
47. Prakash,A., Doublié,S. and Wallace,S.S. (2012) The Fpg/Nei family of DNA glycosylases: substrates, structures, and search for damage. *Prog. Mol. Biol. Transl. Sci.*, **110**, 71–91.
48. Whitaker,A.M. and Freudenthal,B.D. (2018) APE1: a skilled nucleic acid surgeon. *DNA Repair (Amst.)*, **71**, 93–100.
49. Schomacher,L., Han,D., Musheev,M.U., Arab,K., Kienhofer,S., von Seggern,A. and Niehrs,C. (2016) Neil DNA glycosylases promote substrate turnover by tdg during DNA demethylation. *Nat. Struct. Mol. Biol.*, **23**, 116–124.
50. Shen,B., Chapman,J.H., Custance,M.F., Tricola,G.M., Jones,C.E. and Furano,A.V. (2020) Perturbation of base excision repair

sensitizes breast cancer cells to APOBEC3 deaminase-mediated mutations. *Elife*, **9**, e51605.
51. Lirussi,L. and Nilsen,H.L. (2023) DNA glycosylases define the outcome of endogenous base modifications. *Int. J. Mol. Sci.*, **24**, 10307.
52. Rolseth,V., Luna,L., Olsen,A.K., Suganthan,R., Scheffler,K., Neurauter,C.G., Esbensen,Y., Kusnierczyk,A., Hildrestrand,G.A., Graupner,A., *et al.* (2017) No cancer predisposition or increased spontaneous mutation frequencies in NEIL DNA glycosylases-deficient mice. *Sci. Rep.*, **7**, 4384.
53. Chakraborty,A., Tapryal,N., Islam,A., Mitra,S. and Hazra,T. (2021) Transcription coupled base excision repair in mammalian cells: so little is known and so much to uncover. *DNA Repair (Amst.)*, **107**, 103204.
54. Chakraborty,A., Wakamiya,M., Venkova-Canova,T., Pandita,R.K., Aguilera-Aguirre,L., Sarker,A.H., Singh,D.K., Hosoki,K., Wood,T.G., Sharma,G., *et al.* (2015) Neil2-null mice accumulate oxidized DNA bases in the transcriptionally active sequences of the genome and are susceptible to innate inflammation. *J. Biol. Chem.*, **290**, 24636–24648.
55. Sayed,I.M., Sahan,A.Z., Venkova,T., Chakraborty,A., Mukhopadhyay,D., Bimczok,D., Beswick,E.J., Reyes,V.E., Pinchuk,I., Sahoo,D., *et al.* (2020) Helicobacter pylori infection downregulates the DNA glycosylase NEIL2, resulting in increased genome damage and inflammation in gastric epithelial cells. *J. Biol. Chem.*, **295**, 11082–11098.
56. Yuza,K., Nagahashi,M., Watanabe,S., Takabe,K. and Wakai,T. (2017) Hypermutation and microsatellite instability in gastrointestinal cancers. *Oncotarget*, **8**, 112103–112115.
57. Sawyer,D.L. and Sweasy,J.B. (2022) DNA polymerase beta in the context of cancer. *Crit. Rev. Oncog.*, **27**, 17–33.
58. McDaniel,Y.Z., Wang,D., Love,R.P., Adolph,M.B., Mohammadzadeh,N., Chelico,L. and Mansky,L.M. (2020) Deamination hotspots among APOBEC3 family members are defined by both target site sequence context and ssDNA secondary structure. *Nucleic Acids Res.*, **48**, 1353–1371.
59. Ginno,P.A., Gaidatzis,D., Feldmann,A., Hoerner,L., Imanci,D., Burger,L., Zilbermann,F., Peters,A., Edenhofer,F., Smallwood,S.A., *et al.* (2020) A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nat. Commun.*, **11**, 2680.
60. Shi,K., Carpenter,M.A., Banerjee,S., Shaban,N.M., Kurahashi,K., Salamango,D.J., McCann,J.L., Starrett,G.J., Duffy,J.V., Demir,O., *et al.* (2017) Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.*, **24**, 131–139.
61. Ebrahimi,D., Alinejad-Rokny,H. and Davenport,M.P. (2014) Insights into the motif preference of APOBEC3 enzymes. *PLoS One*, **9**, e87679.
62. Makasheva,K.A., Endutkin,A.V. and Zharkov,D.O. (2020) Requirements for DNA bubble structure for efficient cleavage by helix-two-turn-helix DNA glycosylases. *Mutagenesis*, **35**, 119–128.
63. Pei,J., Tang,M. and Grishin,N.V. (2008) PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic Acids Res.*, **36**, W30–W34.
64. Zhu,C., Lu,L., Zhang,J., Yue,Z., Song,J., Zong,S., Liu,M., Stovicek,O., Gao,Y.Q. and Yi,C. (2016) Tautomerization-dependent recognition and excision of oxidation damage in base-excision DNA repair. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 7792–7797.

Received: July 31, 2023. **Revised:** October 17, 2023. **Editorial Decision:** October 18, 2023. **Accepted:** October 30, 2023