HHS Public Access

Author manuscript

Structure. Author manuscript; available in PMC 2022 January 07.

Published in final edited form as:

Structure. 2021 January 07; 29(1): 29–42.e4. doi:10.1016/j.str.2020.08.001.

Unique Structural Features of Mammalian NEIL2 DNA Glycosylase Prime Its Activity for Diverse DNA Substrates and Environments

Brian E. Eckenroth[1,2,*], **Vy Bao Cao**[1], **April M. Averill**[1], **Julie A. Dragon**[1], **Sylvie Doublié**[1,*]

[1]Department of Microbiology and Molecular Genetics, University of Vermont, Stafford Hall, 95 Carrigan Drive, Burlington, VT 05405, USA

[2]Lead Contact

SUMMARY

Oxidative damage on DNA arising from both endogenous and exogenous sources can result in base modifications that promote errors in replication as well as generating sites of base loss (abasic sites) that present unique challenges to maintaining genomic integrity. These lesions are excised by DNA glycosylases in the first step of the base excision repair pathway. Here we present the first crystal structure of a NEIL2 glycosylase, an enzyme active on cytosine oxidation products and abasic sites. The structure reveals an unusual "open" conformation not seen in NEIL1 or NEIL3 orthologs. NEIL2 is predicted to adopt a "closed" conformation when bound to its substrate. Combined crystallographic and solution-scattering studies show the enzyme to be conformationally dynamic in a manner distinct among the NEIL glycosylases and provide insight into the unique substrate preference of this enzyme. In addition, we characterized three cancer variants of human NEIL2, namely S140N, G230W, and G303R.

In Brief

The X-ray crystal structure, SAXS analysis, and characterization of cancer-based variants of the mammalian NEIL2 DNA glycosylase from the base excision repair pathway are presented by Eckenroth et al. This enzyme contributes to maintaining genomic integrity and appears conformationally dynamic in its search for oxidatively damaged DNA bases.

Graphical Abstract

*Correspondence: beckenro@uvm.edu (B.E.E.), sdoublie@uvm.edu (S.D.).

AUTHOR CONTRIBUTIONS

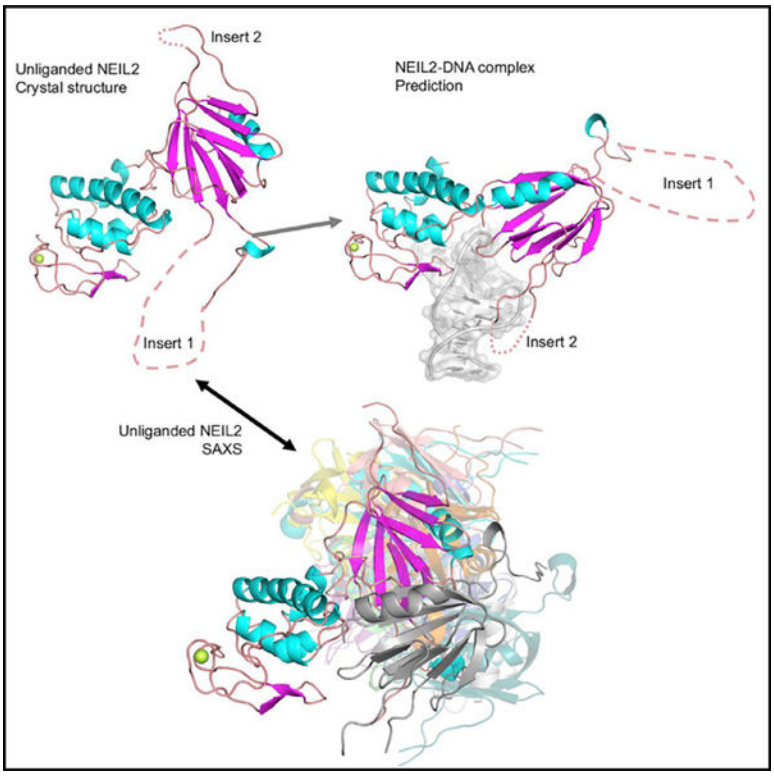
B.E.E. and S.D. wrote the manuscript. S.D. designed the project. B.E.E. designed and performed experiments, including protein expression and purification, crystallizations, processing of crystal and solution diffraction data, solution and refinement of structures, and activity assays. A.M.A. expressed and purified proteins. V.C. expressed and purified proteins and helped with the activity assays. J.A.D. performed analysis of cancer databases for variant selection.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.str.2020.08.001.

DECLARATION OF INTERESTS

The authors declare no competing interests.



INTRODUCTION

Sites of oxidative damage on DNA arise from numerous sources, including environmental agents, endogenously produced oxygen species from normal cellular respiration, and other vital processes, like demethylation, that produce intermediates, including abasic sites. Improperly managed or repaired sites can result in polymerase stalling during replication and ultimately mutagenesis and cancer. Mammals employ a battery of enzymes aimed at repair of DNA damage for the faithful replication of the genome and/or transcript interpretation. One such enzyme, Nei-like 2 (NEIL2), belongs to the Fpg/Nei family of DNA glycosylases (Figure 1) of the base excision repair (BER) pathway. This bifunctional enzyme is capable of excising an oxidized base (glycosylase activity) as well as cleaving the DNA backbone (lyase activity), with the ability to function in an apurinic endonuclease (APE)-independent manner (Das et al., 2006). NEIL2 displays activity on both single- (ssDNA) and double-stranded DNA (dsDNA) with a preference for bubble DNA structures, typically selecting oxidation products of cytosine (Hazra et al., 2002). Proteins of the Fpg/Nei family display remarkable structural conservation of the two-domain architecture while being highly divergent at the sequence level, particularly within the predominantly β-strand-rich N-terminal domain. The C-terminal domain contains the conserved helix-two-turn-helix (H2TH) and zinc or zincless finger (Doublié et al., 2004) DNA-binding motifs and is reasonably conserved within the family.

Accumulating evidence suggests that NEIL2 activity likely fulfills a context-dependent function rather than simple global DNA damage surveillance (Mullins et al., 2019). Studies

using NEIL2 knockdown mouse models have produced mixed results with regard to the accumulation of oxidative DNA damage (Chakraborty et al., 2015; Rolseth et al., 2017), with biochemical and cell-based data indicating a preference for repair of damage in transcriptionally active regions of the DNA (Banerjee et al., 2011). This is in line with the correlation to transcription-associated repair via the observed interaction of NEIL2 with Cockayne Syndrome Protein B (Aamann et al., 2014). Although a direct linkage to a cancer phenotype is yet to be established for NEIL2, studies have identified elevated risk associations in carriers of BRCA2 mutations—identifying NEIL2 as a risk modifier. A SNP in the NEIL2 gene evaluated from patients in the Consortium of Investigators of Modifiers of BRCA1/2 showed significant association between NEIL2 and BRCA2 mutation carriers and breast cancer risk (Osorio et al., 2014), while further analysis suggested that the elevated production of NEIL2 resulting from this SNP correlated with elevated oxidative DNA damage for the BRCA2 mutation carriers (Benitez-Buelga et al., 2017). Recent developments also point to alternative functions of the NEIL glycosylases beyond DNA repair via their contributions to the demethylation pathway of 5-methyl cytosine (Schomacher et al., 2016; Shayevitch et al., 2018).

While some hereditary-based cancers follow a predictable paradigm based on mutations to a specific gene or set of genes, the majority of cancers are considerably more complex and likely arise from an ensemble of mutations in an array of gene products modulated by the basal genomic variability within individual patients. Identifying all contributing factors and establishing their contextual roles are essential to understanding the development and progression of cancers as well as to selection of the most effective therapeutic protocol. Herein we report the first X-ray crystal structure of any Neil2 enzyme, in concert with solution-based studies. The protein adopts a unique open orientation upon crystallization in the absence of DNA substrate, predicting a necessary large conformational change to assemble a catalytically competent complex. The results provide insight into the substrate diversity of NEIL2 and shed light on a unique protein appendage that would allow interactions with multiple protein partners without impeding enzyme function. *In vitro* activity assays of cancer-associated variants of NEIL2 suggest that structural perturbations affect the overall enzymatic activity with significant downstream implications for BER.

RESULTS

Crystallization and Structure Determination

Human NEIL2 (HsaNEIL2) contains 332 amino acid residues with sequence analysis predicting large disorder content between residues 55 (immediately after β strand 2) and 125. Not surprisingly, full-length HsaNEIL2 failed to crystallize. In addition, HsaNEIL2 was problematic for solution studies due to multimodal polydispersity in dynamic light-scattering experiments; aggregation and radiation damage were also observed in small-angle X-ray scattering (SAXS) experiments. The N-terminal domain of the Fpg/Nei structural family contains eight β strands, and secondary structure predictions using HsaNEIL2 and several orthologs were inconclusive with regard to the boundaries for proper truncation of the disordered region due to predicted β-strand segments within that region for some orthologs. Two vertebrate orthologs with predicted lower disorder content in the N-terminal

domain were cloned, expressed, purified, and crystallized: The ortholog from the frog *Xenopus tropicalis* produced crystalline needle clusters that failed to optimize, whereas the mammalian ortholog from the gray short-tailed opossum *Monodelphis domestica* (Mdo) yielded diffraction-quality single crystals.

Because of the relatively low sequence conservation within the Fpg/Nei structural family and the symmetrical nature of the N-terminal fold, resulting in a 4-fold ambiguity in secondary structure matching, structure determination using molecular replacement was unsuccessful and necessitated experimental phasing methods. Numerous trials for single-wavelength anomalous diffraction (SAD) and multiwavelength anomalous diffraction (MAD) utilizing selenomethionine-substituted protein, as well as single-wavelength isomorphous replacement with anomalous scattering (SIRAS) methods, were explored and produced partial phase solutions that stalled during building and refine-ment. The combination of Se-MAD and multiple-wavelength isomorphous replacement (MIR) with anomalous scattering (MI-RAS), as well as phasing using anisotropically processed data, verified a strong noncrystallographic symmetry (NCS) axis within Laue group $\bar{3}$ parallel to that of a crystallographic axis for the higher ordered Laue group $3\overline{m}$ (Figure S1). The orienta- tion of the axis mimicked the $P3_212/P3_112$ space groups with a combined ambiguity of the screw axis. The SAD/MIR phases resolved the final space group to $P3_2$ with two molecules in the asymmetric unit. The solution included native data, Se-SAD at peak wavelength, along with iodide and gold derivatives. A native crystal collected at the peak wavelength for Zn was utilized for verification of the metal ion in the zinc finger motif, which displays anomalous signal at all wavelengths used for phasing.

Mutation of leucines 7, 210, and 284 to Met with subsequent data collection of the selenomethionine derivative at peak wavelength was used to verify their locations. They were chosen strategically, as L7 is in the N-terminal helix, L210 lies in the interdomain linker, and L284 is in the C-terminal domain away from the noncrystallographic 2-fold. Final data and model analysis revealed the hurdles that had plagued the phase determination using the Se-MAD data: of the four endogenous Met residues in the wild-type protein, M39 displays broad and diffuse anomalous signal indicating two conformations, M123 is located in a disordered region, and M301 lies directly across the NCS 2-fold rotation axis. A list of the anomalous and isomorphous difference peaks used for heavy atom placement and structure solution is provided in Table S1.

Overall Structure of NEIL2

The structure of the MdoNEIL2 was determined to 2.54 Å with R_{free} 27.5% and crystallized with two molecules in the asymmetric unit related by a noncrystallographic 2-fold axis. The final model for the two chains contains 251 and 255 amino acids of the 336 amino acid sequence with the primary omitted section being the large internal NEIL2 insert (insert 1: residues 66–131). Some residual density was observed for this region, but its position along the noncrystallographic 2-fold negated its inclusion due to ambiguity (Figure S1). A pair of weakly anomalous scattering peaks could be observed for both the wild-type and the L7/210/284M selenomethionine data, indicating partial ordering of M123 in the insert region. The two chains in the asymmetric unit show a 0.25 Å root-mean-square deviation

(RMSD) for the C-alpha trace with only minor differences between the two chains in the N-terminal domain loops and the respective refined B factors (Figure S1). For simplicity, the remainder of the description of the structure will be limited to a single chain (chain A).

The general fold of NEIL2 is composed of an N-terminal β-sandwich domain and a C-terminal domain containing the H2TH and zinc finger DNA-binding motifs typical of the Fpg/Nei superfamily of glycosylases (Figure 2) (Prakash et al., 2012). Unique to NEIL2 is an extended interdomain linker along with a large insert within the N-terminal domain (aa 66–131 in MdoNei2). The positions of the naturally occurring (36, 39, and 301) and engineered (7, 210, and 284) methionines were identified using anomalous scattering from the selenomethionine-substituted protein crystals (Doublié, 2007). The heavy atom soaks provided identification of cysteines 183 and 254 (Figures 2A and 2B), while the zinc atom in the predicted CHCC-type zinc finger was verified by X-ray fluorescence at the peak wavelength for zinc.

The core portions of the N- and C-terminal domains were compared with other proteins in the Fpg/Nei family, with the N-terminal domain showing a 1.75 Å RMSD (22% sequence identity over 92 residues, 377 atoms) and 0.78 Å RMSD (29% sequence identity over 100 residues, 483 atoms) for the N- and C-terminal domains of mouse NEIL3, respectively. Modest molecular replacement scores were obtained when a trimmed C-terminal domain of NEIL3 was used as a search model, but it did not provide enough phasing power for structure completion, particularly with the challenge of the rotational NCS, as the aligned portion represented less than 30% of the molecular mass. In addition, the N-terminal domain shares very little sequence conservation with other glycosylases of the same family outside of the N-terminal helix active-site P2-E3-G4 motif and conserved lysine (K50 in MdoNeil2 and HsaNEIL2), further hindering structure solution via molecular replacement. The lack of sequence conservation in the N-terminal domain for this family of proteins is significantly challenging to molecular replacement when considering that the topology of the β-sandwich fold (Figure 2) generates 4-fold ambiguity in a model search routine.

Since the large insert in NEIL2 contains significant disorder preventing its inclusion in the final model and resulting elevated B factors, a protein construct was engineered with the region omitted. Omission of this region had little impact on enzymatic activity (below); however, the construct proved more difficult to crystallize, with the diffraction limit reaching only to 4.3 Å, with pronounced anisotropic diffraction. The crystals were of the same space group with minimal change in unit cell parameters. Isomorphous difference methods with an R_{merge} of 31% for F_{obs} between full-length and truncated datasets (Table 1, Figure S2), a 34.5% R_{free} with the isomorphous replacement model upon rigid-body refinement, and the molecular replacement solution all show the truncated protein to be in the same conformation as the full-length protein, suggesting that the insert aided in crystal growth but does not influence the overall structure of the protein within the crystal.

The NEIL2 Ortholog Is Active on Both Single- and Double-Stranded DNA

The activity of NEIL2 was assessed initially using cyanoborohydride trapping on SDS-PAGE (data not shown) followed by glycosylase activity assays using radiolabeled DNA substrates. The glycosylase activity quantification is shown in Figures 3 and S3. The

opossum ortholog shows 57% sequence identity and 70% homology with the human enzyme, with the variability in sequence of the unique insert accounting for 10% of the divergence. Not surprisingly, the human and opossum ortholog showed a similar substrate profile using the initial trapping assay screen with greater activity toward ssDNA substrates as summarized quantitatively in Figure 3. NEIL2 shows modest activity toward lesions within duplex DNA. However, unlike what was observed for NEIL3, NEIL2 shows robust activity for the apurinic/apyrimidinic (AP) site within duplex DNA. In both single and double-stranded contexts, glycosylase activity reaches completion with enzyme-to-substrate ratios in the 4:1 to 16:1 range for both the full-length protein and the construct with the disordered region removed. The abasic site was the preferred substrate in both contexts. For the oxidized base lesions, DHU showed the greatest activity within ssDNA, reaching completion at the highest enzyme ratio, while OHU showed the greatest activity in dsDNA but reached only ~30% completion under the conditions employed. The poor Tg activity along with AP site preference is in agreement with what was previously reported (Katafuchi et al., 2004). Deletion of the large insert (MdoNEIL2cut variant) did not adversely affect the enzymatic activity (Figure 3). Interestingly, the deletion variant appears to be more active than wild type on Tg in a dsDNA context, but this activity represents just a fraction of the activity on its best substrate, abasic site-containing DNA. We observed in our assays that the primary product of the reaction for ssDNA appears to be the δ-elimination product, with the β-elimination product being favored for duplex DNA substrates (Figures 3 and S3).

Comparison with Other Fpg/Nei Glycosylases

The NEIL2 structure revealed, unexpectedly, that the orientation of the C-terminal domain relative to the N-terminal domain differs significantly from that of other proteins in the family (Figure 2C). While NEIL1 adopts a “closed” conformation whether it is bound to DNA or not, NEIL2 exhibits an “open” conformation, which places the N-terminal catalytic residue on the opposite face of the C-terminal domain DNA-binding motifs, requiring a near 80 rotation (Figure 4) of the domains relative to each other to achieve catalytic competency. The only other time this has been observed for proteins in this family is for the more distantly related endonuclease VIII glycosylase from *E. coli* (EcoNei) (Figures 4A and 4B) (Golan et al., 2005; Zharkov et al., 2002). It was suggested by the authors that other glycosylases may also display interdomain flexibility. Comparison of Figures 4A–4C for the unliganded conformations of EcoNei and MdoNEIL2 shows a drastic difference in relative orientation and the type of movement required to achieve a typical catalytically competent interdomain orientation. Whereas EcoNei domain movement can occur through a simple hinge motion to place the N-terminal active-site helix (blue) into position, MdoNEIL2 would need to utilize a twisting rotation, as the N-terminal helix in the crystal structure resides on the opposite face of the structure in the equivalent figure orientation.

The NEIL glycosylases differ from their bacterial counterparts because they contain long disordered regions (Liu et al., 2013a). Whereas NEIL1 and NEIL3 harbor these flexible regions at their C terminus, sequence alignments revealed that NEIL2 is unique in that the disordered region is located within the glycosylase fold, more precisely in the N-terminal domain. Two interesting unique NEIL2 features are revealed within the N-terminal domain pertaining to insertion elements: a largely disordered region (aa 65–130) resides between β

strands 3 and 4. In the open conformation the insert projects toward the typical DNA binding interface with the C-terminal domain (Figure 1C). In the predicted closed catalytically competent NEIL2, this region projects away from the DNA binding interface (Figure 4E). This insert is likely involved in protein-protein interactions with proteins of the BER pathway and other DNA repair proteins (Das et al., 2006). A second, smaller insert of 8–10 residues (insert 2 residues 155–165), also unique to NEIL2, is located between β strands 5 and 6 and contains three lysines and one arginine in the opossum ortholog, and three lysines and two arginines in the human sequence, which could influence activity on the unique substrate range for NEIL2. The N-terminal domain rotation expected for assembly of a catalytically competent complex would put this smaller insert in proximity to the DNA substrate-binding cleft (Figure 5).

Despite an overall low sequence conservation, NEIL2 retains some key features within the N-terminal domain compared with other glycosylases of the Fpg/Nei superfamily: an N-terminal active site (Figure 5), a conserved lysine critical to activity located in a short loop between β strands 2 and 3 (Lys50 in MdoNeil2/HsaNEIL2 versus Lys54 in HsaNEIL1), and a loop between β strands 4 and 5 containing a hydrophobic residue (typically Leu or Met) that stabilizes the everted damaged base into the active site of the enzyme (Imamura et al., 2009). Because of the large insert in NEIL2, prediction of the location for the β4-β5 loop was uncertain. The structure of MdoNEIL2 determined here reveals that the β4-β5 loop comprises residues 140–143, with Leu141 as a candidate to serve the function of stabilizing the damaged base everted into the active site.

The position of a third loop, located between β strands 7 and 8, which in NEIL1 harbors two residues stabilizing the orphaned base on the nondamaged strand (typically an Arg and a Tyr/Phe), was equally uncertain. Similar to what was observed for the ssDNA-dependent NEIL3, the β7-β8 loop containing two of the void-filling residues in NEIL1 is significantly shorter for NEIL2 (Liu et al., 2013b); NEIL2 therefore lacks two of the void-filling residues, implying that lesion search and recognition will differ from what was described for NEIL1 (Imamura et al., 2009, 2012; Zhu et al., 2016).

CONFORMATIONAL DYNAMICS IN SOLUTION

The observation of the unique interdomain conformation of NEIL2 prompted further investigation into the solution behavior of the protein utilizing SAXS (Table 3, Figure 6). Although crystal structures exist for *E. coli* endonuclease VIII (EcoNei) showing distinct conformations of the enzyme between unliganded and bound to DNA (Golan et al., 2005; Zharkov et al., 2002), this behavior has not been investigated in solution. We conducted SAXS experiments for the unliganded protein to see if the structure observed in the crystal represents the solution state of the enzyme (Figure 6B). Data were collected under static conditions as well as utilizing in-line size-exclusion chromatography (SEC), with the subsequent scattering curves fit to the theoretical curves generated from the published crystal structures, using fast X-ray scattering (FoXS) and utilizing PCNA as a protein control (Figure 6A). Both methods provide strong evidence that the conformation observed for the unliganded EcoNei represents the state observed in solution, suggesting that this enzyme undergoes a specific conformational event upon binding DNA. Evaluation of

particle shape in the form of Rg and Dmax shows PCNA to fall within values predicted from the crystal structures as well as being in agreement with previous SAXS studies (Prakash et al., 2017). While no SAXS studies have been published to date for EcoNei, values for Rg and Dmax were consistent with the crystal structure.

Both full-length and insert-deleted NEIL2 (MdoNEIL2cut) were evaluated, with the latter only amenable to the in-line SEC data collection due to instability within batch methods (Figure 6C). SEC, dynamic light scattering, and multiangle light scattering showed both constructs to be homogeneous monomeric populations in solution. The full-length NEIL2 showed Rg as well as pairwise distance distribution functions consistent with being larger than EcoNei. However, the Dmax, particularly under batch conditions, was considerably larger (Figure 6D). Like the MdoNEIL2cut construct, the full-length protein showed improvement by application of in-line SEC methods. The distance distribution profile, like the primary scattering curve for MdoNEIL2cut, was quite similar to that observed for EcoNei, indicating some agreement in particle shape. As the large disordered insert cannot directly be fit to the scattering curve of the full-length protein, structure fitting was performed using only the MdoNEIL2cut data. Like that of EcoNei, the best fit to data could be achieved to the crystal structure and not to a model produced by allowing the N- and C-terminal domains to rotate to a catalytically competent conformation (Figure 6E). However, the quality of the fit in the low Q and Porod regions deviates significantly from the experimental curve compared with that of the PCNA control or EcoNei. This finding indicates that neither the MdoNEIL2 apo crystal structure conformation nor the expected catalytic conformation, nor an average between the two theoretical curves (representing equilibrium of the two conformations), represents the true behavior of MdoNEIL2 in solution.

The data for both EcoNei and MdoNEIL2 were further evaluated with MultiFoXS (Schneidman-Duhovny et al., 2016), an ensembling method not restricted to the input conformation of the protein. Both proteins were allowed flexibility in the linker (residues 125–134 for EcoNei and 190–214 for MdoNEIL2) and evaluated for up to five states with the final selection of 10 models for each state (Figure S4). The method suggested two or three states to provide the best fit to the data. For EcoNei, the input unliganded structure was selected for 3 of the 10 models for conformation 1, and in none of the models produced for any of the five states was the catalytically competent conformation produced. Conversely, neither the crystal structure conformation nor the conformation predicted for catalytic competency was sampled during the simulation. Additional analysis of the particle envelopes using Dammif (Franke and Svergun, 2009) displayed convergence to a discrete population for EcoNei but high variation for MdoNEIL2 (Figure S4). Overall the data suggest that, while EcoNei likely exists in discrete populations, MdoNEIL2 is free to sample more conformational space in the absence of DNA.

ANALYSIS OF HsaNEIL2 CANCER VARIANTS

Several human NEIL2 variants (S140N, G230W, and G303R) were selected from cancer databases by the UVM Bioinformatics Shared Resource for evaluation of biochemical function guided by the crystal structure of MdoNEIL2. G230W and G233 in MdoNEIL2

showed a significant defect in activity (Figure 3B), predominantly with ssDNA. This glycine is the N-terminal cap of a helix in the H2TH motif of the C-terminal domain. This residue is absolutely conserved (Figure 5) within the Fpg/Nei family. It provides the N-terminal cap of the helix coordinating the catalytic residue Glu3, and the adjacent residue, Asn231, is also strictly conserved as it forms a hydrogen bond with a nonbridging oxygen of the phosphate backbone 5′ to the lesion. Introducing a large aromatic residue like tryptophan at the location of G230 is predicted to induce local changes that would negatively affect DNA binding and activity. A second variant, G303R in the zinc finger motif, also showed a significant decrease in glycosylase activity, albeit less severe compared with G230W, whereas lyase activity was largely retained for both ssDNA and dsDNA. Mutation to arginine in this location with such proximity to the site of lesion has the potential to influence either interaction with the DNA or the chemistry. Interestingly, the assay product observed for ssDNA for these two variants was β elimination, like that seen for dsDNA substrate, and not δ elimination (Figure S3). A third variant, S140N was also evaluated. It resides at the N-terminal end of β strand 5, which resides between the canonical void-filling loop and the small insert in NEIL2. S140N showed no change in overall glycosylase activity relative to wild type but appeared to follow the trend of the other variants, with a modest shift in the equilibrium between β elimination and δ elimination (Figure S3E). It is possible that a more significant effect may be seen for this variant with a different substrate or sequence context.

DISCUSSION

Mammalian DNA is a dynamic and crowded space with numerous processes occurring both sequentially and simultaneously. Coordination of DNA replication, sensing of damage sites for repair, and changes in methylation patterns for transcription of activated genes require the orchestration of recruitment for multifunctional protein factors for use in specific functions. Increasing evidence has placed NEIL2 at the junction of several of these processes through identification of binding partners, their influence on activity, and the range of substrates suitable for activity. Its fundamental activity of recognition and removal of oxidized bases and abasic sites within unique structures of DNA is temporally and spatially dependent on the specific function. Currently identified binding partners of NEIL2 include BER and small-patch-repair DNA polymerase β and XRCC1 (Campalans et al., 2005; Das et al., 2006), transcription-associated repair CSB (Aamann et al., 2014), and APE1 and TDG (Schomacher et al., 2016) and represent different steps in DNA repair and maintenance. Here we present the first crystal structure of a NEIL2 glycosylase. The protein adopts an unexpected open conformation not observed in NEIL1 and NEIL3. We demonstrate that NEIL2 is conformationally dynamic and suggest that the flexibility of NEIL2 allows tethering to other proteins without impeding its interactions with the DNA. This feature would allow utilization of NEIL2 in alternate environments. Placement of the protein-interaction module within the N-terminal domain of NEIL2 indicates that the repair environment or dynamic interplay between BER partners differs from that of NEIL1, which contains a long unstructured extension that serves a similar function (Hegde et al., 2013; Sharma et al., 2018), but at its C terminus.

The crystal structure also revealed that NEIL2 lacks two of the three void-filling or intercalating residues (Arg and Tyr) observed in other glycosylases of the same family (Leu/Met, Arg, Phe/Tyr) (e.g., Met81 on the β4-β5 loop or Arg118 and Phe120 on the β7-β8 loop in human NEIL1) (Doublié et al., 2004; Liu et al., 2013b; Zhu et al., 2016).The Leu/Met hydrophobic residue stabilizes the flipped out lesion everted from the double helix. The Arg and Phe/Tyr residues invade the DNA double helix on the side of the orphaned base. The aromatic residue in the triad has been referred to as the wedge residue, because of its role in lesion search. An *E. coli* Fpg variant where Phe111 was mutated to alanine displayed a reduced glycosylase activity on lesion-containing dsDNA compared with the wild-type enzyme. This decrease in activity could be attributed to a reduced ability to search for lesions (Dunn et al., 2011). The Arg and Phe/Tyr residues are absent in NEIL2 because the loop they reside on is much shorter than that in NEIL1. We described the same situation in the NEIL3 structure (Liu et al., 2013b). Because NEIL2 and NEIL3 prefer ssDNA over dsDNA it is not all that surprising that they lack the loop that harbors residues whose function is to buttress the orphaned base on the opposite strand. But this leaves open the question as to how NEIL2 and NEIL3 search for lesions and which residue, if any, serves as the “wedge” residue.

Of the currently determined structures of enzymes within the Fpg/Nei family, NEIL2 represents only the second example of a glycosylase undergoing large-scale conformational changes. Numerous studies have investigated the conformational events in glycosylases as observed by changes in tryptophan fluorescence, including for NEIL1, which experiences small-scale dynamics (Kladova et al., 2019) and EcoNei (Kuznetsov et al., 2012), which shows large-scale dynamics. These studies can measure only solvent environment change about the Trp residues and cannot speak of distance in the manner that fluorescence resonance energy transfer can, but both provide timescale comparisons. Interestingly, the rate constant of the initial measurable step for both enzymes is extremely fast and on equivalent scales, at 180 3 10^6 s^{-1} for NEIL1 and 12 3 10^6 s^{-1} for EcoNei, comparing abasic site substrates. These results follow the structural distinction that the enzyme with the smallest conformational change (NEIL1) has the faster rate for this step by about 10-fold. However, this must be taken into context by considering that the overall rate of the reaction (chemistry) is only 0.1–0.5 s^{-1}. In fact, the additional observed microscopic rate constants were quite similar for the two proteins. These results suggest that the initial binding and scanning of the DNA occur 7 orders of magnitude faster than the actual rate of chemistry, regardless of the conformational distance traveled. It is important in the case of NEIL2, which we predict must experience a large domain rearrangement to assemble the catalytically competent complex in the manner observed for EcoNei, that the conformational step not constitute a functional or kinetic barrier.

Mouse knockdown models of NEIL1 and NEIL2 resulting in metabolic stress and inflammation but modest or no elevation in mutational load (Rolseth et al., 2017) suggest a substrate repertoire that overlaps with other enzymes or that proteins like NEIL2 serve functions outside of basal genome maintenance. The strong AP-site activity on both ssDNA and dsDNA suggests that the abasic site is a true substrate for NEIL2, the favored base modification substrate is as of yet unidentified, or sequence-context dependence for a particular lesion may exist. Indeed, it was recently shown that both EcoNei and Fpg will

excise N4,5-dimethylcytosine while leaving 5-methylcytosine and 4-methylcytosine uncut (Alexeeva et al., 2018). In addition, both NEIL1 and NEIL2 were shown to function in TDG-mediated turnover of 5hmC (Schomacher et al., 2016) at sites of TET-dependent DNA demethylation, while 5hmC accumulated at sites of DNA damage (Kafer et al., 2016) and thus exhibited the ability to substitute for APE1 in the demethylation pathway. Interestingly, knockdown of NEILs resulted in elevated 5fC and 5caC levels, whereas APE1 knockdown did not. The proposed association of NEIL2 with transcription-associated repair (Chakraborty et al., 2015), juxtaposed with the difficulty in producing cancer phenotypes in mice (Rolseth et al., 2017), suggests that the true function of NEIL2 may be yet to be established. Transient sources of bubble or transient ssDNA occur during transcription and recombination events. While errors in replicated DNA strands may be tracked experimentally, sites left unrepaired during translation resulting in aberrant protein synthesis are more difficult to measure but contribute to inflammation.

Recent studies implicate a potential role for NEIL2 in cytosine deaminase APOBEC3-mediated mutagenesis, likely via its activity toward abasic sites (Shen et al., 2020), where data suggest NEIL2 may outcompete APE1 for the site in particular contexts. This presents an interesting contribution of NEIL2 variants to the APOBEC3 mutation signature by aberrant NEIL2 activity toward the abasic sites produced during the deamination process as well as reduced activity toward oxidized cytosine lesions. The variants studied here displayed the most significant decrease in activity for the cytosine-derived substrates 5-OHU in both ssDNA and dsDNA contexts, and with DHU in the single-stranded substrate. The variant G230W also showed a pronounced decrease in activity for the abasic site in both contexts. Recent data also suggest that both NEIL1 and NEIL2 show preference for a D loop mimic over the single-stranded exposed lesion within bubble structure substrates (Makasheva et al., 2020) while displaying differential activity depending on the location of the lesion within the bubble. Moreover, removal of hydantoin lesions by NEIL1 was shown to be context dependent (Zhao et al., 2010). The structural and biochemical investigations presented here provide insight into the dynamic behavior of NEIL2. Our work also illuminates how variants of the enzyme can contribute to disease either through mutagenesis-based genomic instability or by influencing gene expression through action on unique DNA structures encountered during transcription.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

RESOURCE AVAILABILITY

Lead Contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Sylvie Doublié (sdoublie@uvm.edu). Questions regarding the technical aspects of the studies within the manuscript can be directed to Brian E. Eckenroth (beckenro@uvm.edu).

Materials Availability—Plasmids generated will be deposited into Addgene. In the meantime researchers can contact directly Sylvie Doublié for published materials.

Data and Code Availability—The model and data associated with the crystal structure of MdoNEIL2 has been deposited in the Protein Data Bank (rcsb.org) under PDB ID code 6VJI. The deposit includes the data used for model refinement as well as the data sets used for phase determination (native, selenomethionine at peak wavelength, zinc peak wavelength, gold derivative at peak wavelength and iodide derivative). All raw diffraction data (images) have been deposited at proteindiffraction.org.

The SEC-SAXS data for EcoNei (accession code SASDJA4), MdoNEIL2 cut (accession code SASDJC4) and their associated model fits have been deposited into the Small Angle Scattering Biological Data Bank (sasbdb.org) as has the full length MdoNEIL2 (accession code SASDJB4) SEC data.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All protein used for biochemical and structural studies were expressed recombinantly in *E. coli* using commercially available strains BL21(DE3) from New ENgland Biolabs for the MdoNEIL2 codon optimized by Genscript and E. coli BL21-Rosetta2(DE3) pLysS Novagen for non-codon optimized.

METHOD DETAILS

Protein Expression and Purification—Both full-length and truncated gray short tailed opossum *Monodelphis domestica* Neil2 (MdoNEIL2) sequences were synthesized and codon-optimized for expression in *E. coli* by GenScript and subcloned into a pET30 vector (Novagen) using NdeI and XhoI restriction sites, then transformed into BL21(DE3) plated onto LB-agar under kanamycin selection. Positive colonies were used for bulk expression using Terrific Broth (TB) supplemented with 4% (v/v) glycerol and 10 mM zinc sulfate. Culture was grown to an OD_{600} of 1.0, induced with 500 μM IPTG, and expressed for 6 hours at room temperature. Cells were harvested then lysed in buffer containing 50 mM sodium phosphate pH 8, 300 mM NaCl, 10% (v/v) glycerol, 10 mM imidazole, 2 mM β-mercaptoethanol, 0.01% (v/v) NP-40 and 1 mM PMSF and purified by nickel affinity resin (Thermo Fisher Scientific) with elution in buffer containing 400 mM imidazole. The eluant was exchanged to 25 mM Hepes pH 7.5, 100 mM NaCl, 10% (v/v) glycerol and 1 mM DTT prior to ion exchange chromatography using a Fast Flow SP column (GE Healthcare). Elution from the SPFF column was performed with a 100 mM to 1M gradient over 20 column volumes. Concentrated aliquots were stored in 50 mM HEPES pH 7.5, 300 mM NaCl, 1 mM DTT and 10% glycerol and flash frozen in liquid nitrogen for storage at −80°C. Expression for selenomethionine incorporation substituted the TB media with minimal media as described in (Doublié, 2007). For phasing verification three sites were mutated from Leu to Met (7, 210 and 284) to increase the number of heavy atom sites described below. Upon structure solution of the full-length protein, a construct lacking the internal disordered region was engineered, replacing residues [F68-N127] of insert 1 with a short flexible linker of sequence GSGSG. The loop deletion construct for MdoNEIL2 was expressed and purified as described above. Purification of human NEIL2 (HsaNEIL2) followed the same protocol as MdoNEIL2, with the exception of the expression step: The pET30-HsaNEIL2 construct is not codon optimized for *E. coli* expression and thus required Rosetta2 DE3 pLysS cells (Novagen) for expression.

Glycosylase Activity Assays—Initial activity for NEIL2 was screened using a cyanoborohydride trapping assay. Reactions containing 17 μM enzyme and 43 μM DNA substrate were incubated at 25°C for 30 minutes in 10 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM DTT, 1 mM $MgCl_2$ in the presence of 100 mM sodium cyanoborohydride. Reactions were quenched by the addition of 100 mM Tris-HCl pH 8.0 and run on a denaturing 12% SDS-PAGE gel. Activity was determined using ^{32}P labeled DNA oligos and performed using 35-mer oligonucleotides with a damage-containing strand sequence of 5′-TGTCAATAGCAAG(X)GGAGAAGTCAATCGTGAGTCT-3′, where X was Tg, 5-OHU, 5,6-DHU were purchased from Midland Certified Reagent Co. (Midland, TX), purified by urea PAGE and labeled using T4 polynucleotide kinase in the presence of $\gamma-^{32}P$ as previously described (Prakash et al., 2014). The assays were performed using 25 nM DNA and 25–800 nM enzyme for 30 minutes at 37°C in 10 mM Tris-HCl pH 8.0, 75 mM NaCl, 100 ug/mL BSA and 1 mM DTT. The reactions were quenched with solution containing 98% formamide, 5 mM EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue then heated at 95°C for 5 min before separation on 12% urea-PAGE. Maximum activity under the assay conditions for the most active enzymes could be estimated using fits to standard single exponential curves. However, assays with lower activity fit poorly. All assay titration curves are provided in Figure S3 and are plotted as the fraction of DNA substrate cleaved versus protein concentration, where 1.0 represents 100% cleavage. Final comparison of maximal activity under the assay conditions was performed using the mean activity at the highest enzyme concentration and summarized in Figure 3.

Crystallization and Data Collection of MdoNEIL2—Frozen aliquots of full-length MdoNEIL2 were thawed and exchanged into buffer containing 50 mM HEPES pH 7.5, 100 mM NaCl and 1 mM TCEP. Crystallization optimization was performed by hanging drop vapor diffusion after high-throughput screening identified suitable starting conditions (PEG-Ion HT, condition C11; Hampton Research). For the initial matrix screening, the protein was retained in the storage buffer containing 10% (v/v) glycerol. Crystallization hits were amorphous quasi-crystals with subsequent experimentation showing removal of the glycerol to be essential for producing properly formed crystals. Protein concentration was optimal between 1.5 and 3 mg/ml. Optimized conditions contained 40 mM HEPES pH 7.5, 1 mM TCEP, 50 mM sodium succinate, 1% propylene glycol and 12–16% PEG 3350. Cryoprotection was achieved by increasing the PEG concentration to 20%, with the inclusion of 30% glucose. Crystals diffracted to 2.7 Å for the native full-length MdoNEIL2, 2.5 Å for the selenomethionyl version, and 3.1 Å for the selenomethionyl L7/210/284M triple mutant. Heavy atom soaks used native crystals and were performed in the cryoprotection stabilizing reagent for 2 hours at room temperature in the presence of 200 mM NaI, 5 mM $KAu(CN)_2$, or 5 mM K_2PtCl_4. Crystals were screened for diffraction quality on a home source (Bruker D8 Quest). High-resolution data were acquired at the Advanced Photon Source GM/CA Sector 23 on either a Mar300 CCD or Pilatus 6M detector. Data for each heavy atom were collected at the peak wavelength (Table 1), with the WT selenomethionine also collected at a high-energy remote wavelength. The truncated form of MdoNEIL2 crystallized in similar conditions to that of full-length, with the distinction that crystals were consistently smaller and diffracted poorly.

Structure Determination of MdoNEIL2—Diffraction data were processed with XDS (Kabsch, 2010) and scaled using STARANISO (Tickle et al., 2018), with the exception of the data sets for the selenomethionyl L7/210/284M variant, Platinum derivative, and the truncated form of MdoNEIL2, which were processed with Proteum3 (Bruker AXS). Initial selenomethionine sites for the full-length construct along with anomalous scattering from the zinc atom were identified using ShelxD (Sheldrick, 2008) as incorporated into AutoSHARP (Vonrhein et al., 2007). The full-length protein sequence contains 336 residues and 4 methionines. Sites were refined using ShelxE and phasing performed using AutoSHARP followed by improved density modification using Solve/Resolve within Phenix (Adams et al., 2010). Cross comparison between datasets was performed using Scaleit after data merging using CAD. An initial solution was achieved using space group $P3_212$, as suggested by Laue group interpretation and self-rotation function analysis. Model building and refinement stalled with R-factors greater than 30% and elevated B-factors. Detailed analysis of anomalous reflection correlations identified the lower symmetry space group $P3_2$ as a more probable solution. A final phased solution using AutoSHARP was achieved using the WT selenomethionine, iodide and gold derivatives using STARANISO ellipsoidal truncated data. The L7/210/284M selenomethionyl variant, WT platinum, and a native crystal collected at the peak wavelength for zinc were used for validation (Table 1). Model building was performed in Coot (Emsley et al., 2010) and final refinement performed with Phenix using the spherically truncated data processed by either Proteum3 for the single highest resolution crystal of the WT selenomethionine or aP_scale/Aimless for two crystals (Table 2). The final stages of refinement were challenged by the disordered regions of the N-terminal domains with final model B-factors at the high end of the distribution when compared to others structures of similar resolution. This is likely due to 25% of the molecular mass within the crystal unable to be included in the final model. While the majority of the refinement stages were performed within Phenix using the spherically truncated data from Proteum3, issues during deposition validation prompted a systematic comparison of data processing and final refinement programs, part of which is shown in Table 2. The first four columns in the table show the final iteration of refinement for the same input molecule by either Phenix or Refmac utilizing data processed with Proteum3 or aP_scale/Aimless. Upon deposition validation, significant variant, most notably in RSRZ, were observed. The final model refined has been deposited with PDB ID 6VJI and includes the associated spherically truncated structure factors. The deposition also includes the native crystal data, selenomethionine peak wavelength, iodide derivative and gold derivative ellipsoidally truncated used for phasing along with data collected at the zinc peak wavelength. Raw diffraction data will also be made available.

The truncated form of MdoNEIL2 was evaluated via both isomorphous differences (Phenix) and molecular replacement (Phaser (McCoy et al., 2007)), applying search routines with the N and C terminal domains separately. Due to the low resolution, only rigid body refinement was evaluated using Phenix (Table 2).

SAXS Data Collection and Interpretation—Samples for SAXS analysis performed in batch/static mode were prepared by additional purification over a Superdex S200 gel filtration column (GE Life Sciences Inc.) in 25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM

DTT and 2–4% glycerol with the equivalent buffer utilized for background subtraction. A concentration series was screened for each protein to evaluate concentration-dependent oligomerization. Data were collected at the SIBYLS beamline (Advanced Light Source, Lawrence Berkeley National Laboratory) through the mail-in data collection program at 11–12 KeV at 0.3–0.5 sec/frame for a 10–15 second run on a Pilatus 6M detector (Table 3) with background correction performed at the beamline. Samples for in-line SEC-SAXS were run in 25 mM BisTris pH 8, 150 mM NaCl, 2% glycerol and 1 mM TCEP using a Shodex PROTEIN KW-802.5 at 12 KeV and 3 sec/frame. Background regions were selected from the region prior to the first sample elution peaks and corrections utilizing a minimum of 10 points were performed using Chromixs (Panjkovich and Svergun, 2018). Sample peak points of full width above half max were were selected for background correction. All SAXS primary data analysis was performed using ATSAS 2.8.4 (Franke et al., 2017) with Guinier approximation using data with qRg <1.3 performed with PRIMUS (Konarev et al., 2003). Scattering ranges and radiation wavelengths for data collection are provided in Table 3. Fitting of scattering curves to theoretical scattering based on deposited crystal structures was performed using FoXS (Schneidman-Duhovny et al., 2016). The EcoNei protein used for SAXS experiments was expressed and purified as described in (Bandaru et al., 2006; Jiang et al., 1997) via chitin agarose affinity and Fast Flow SP column (GE Healthcare) while the expression and purification protocol of human PCNA (used as control) was previously described in (Prakash et al., 2017). The pET11a-PCNA plasmid was a lind gift from Dr. Todd Washington (University of Iowa, USA), expressed using Rosetta2 DE3 pLysS, and purified using His-tag affinity and Superdex 200 gel filtration. The data utilized for SAXS analysis will be made available via the Small Angle Scattering Biological Data Bank (SASBDB) (www.sasbdb.org).

***Selection of HsaNEIL2 cancer variants* G303R:** The rs528951528 single nucleotide polymorphism (SNP) of NEIL2, a G to C transversion mutation is predicted to be deleterious (SIFT) or probably damaging (Polyphen)(Adzhubei et al., 2010; Gao et al., 2013). This mutation leads to the substitution of glycine to arginine at residue 303. G230W: A NEIL2 mutation, where G is mutated to T at position 8: 11640908 (hg19/GRCh37) is known from one lung adenocarcinoma (TCGA-55–8089-01). It is predicted to be deleterious (SIFT) or probably damaging (Polyphen). This mutation leads to the substitution of glycine to tryptophan at residue 230 (G230W). Another mutation at the same position, G230R, is associated with esophageal adenocarcinoma (TCGA-R6-A8WC-01). S140N: A NEIL2 G to A transition mutation at position 11637387 is predicted to be deleterious (SIFT) or probably damaging (Polyphen). This mutation leads to the substitution of serine to asparagine at residue 140. It is associated with bladder urothelial carcinoma (TCGA-XF-A9SZ-01).

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of activity data was performed using ImageJ and data analysis performed using PRISM. Data are reported as the mean of duplicates with error representing standard deviation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by the National Cancer Institute via NIH grant P01 CA098993 awarded to S.D. with B.E.E. supported by NIH grant R50 CA233185. Single-crystal X-ray diffraction experiments were conducted at GM/CA@APS, which has been funded in whole or in part by federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract DE-AC02-06CH11357. The SAXS data collections were conducted at the Advanced Light Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by the DOE Office of Biological and Environmental Research. Additional support comes from the National Institutes of Health project MINOS (R01GM105404) and a High-End Instrumentation grant S10OD018483. SAXS data were collected at SIBYLS, which is funded by DOE/BER IDAT grant, contract DEAC02-05CH11231 and NIH MINOS RO1. Identification and selection of human variants along with DNA sequencing services were performed using the Bioinformatics Shared Resource and Vermont Integrative Genomics Resource at the University of Vermont. Support for the X-ray facility by the Trunk Foundation is gratefully acknowledged. The authors thank Clemens Vonrhein of Global Phasing Limited for assistance with STARANISO data processing and SHARP/AutoSharp.

REFERENCES

Aamann MD, Hvitby C, Popuri V, Muftuoglu M, Lemminger L, Skeby CK, Keijzers G, Ahn B, Bjørå s M, Bohr VA, et al. (2014). Cockayne Syndrome group B protein stimulates NEIL2 DNA glycosylase activity. Mech. Ageing Dev. 135, 1–14. [PubMed: 24406253]

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221. [PubMed: 20124702]

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, and Sunyaev SR (2010). A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249. [PubMed: 20354512]

Alexeeva M, Guragain P, Tesfahun AN, Tomkuviene M, Arshad A, Gerasimaite R, Ruksenaite A, Urbanaviciute G, Bjørå s M, Laerdahl JK, et al. (2018). Excision of the doubly methylated base N(4),5-dimethylcytosine from DNA by Escherichia coli Nei and Fpg proteins. Philos. Trans. R. Soc. Lond. B Biol. Sci. 373, 20170337.

Bandaru V, Blaisdell JO, and Wallace SS (2006). Oxidative DNA glycosylases: recipes from cloning to characterization. Methods Enzymol. 408, 15–33. [PubMed: 16793360]

Banerjee D, Mandal SM, Das A, Hegde ML, Das S, Bhakat KK, Boldogh I, Sarkar PS, Mitra S, and Hazra TK (2011). Preferential repair of oxidized base damage in the transcribed genes of mammalian cells. J. Biol. Chem. 286, 6006–6016. [PubMed: 21169365]

Benitez-Buelga C, Baquero JM, Vaclova T, Fernandez V, Martin P, Inglada-Perez L, Urioste M, Osorio A, and Benitez J (2017). Genetic variation in the NEIL2 DNA glycosylase gene is associated with oxidative DNA damage in BRCA2 mutation carriers. Oncotarget 8, 114626–114636. [PubMed: 29383107]

Campalans A, Marsin S, Nakabeppu Y, O'Connor TR, Boiteux S, and Radicella JP (2005). XRCC1 interactions with multiple DNA glycosylases: a model for its recruitment to base excision repair. DNA Repair (Amst) 4, 826–835. [PubMed: 15927541]

Chakraborty A, Wakamiya M, Venkova-Canova T, Pandita RK, Aguilera-Aguirre L, Sarker AH, Singh DK, Hosoki K, Wood TG, 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. [PubMed: 26245904]

Das A, Wiederhold L, Leppard JB, Kedar P, Prasad R, Wang H, Boldogh I, Karimi-Busheri F, Weinfeld M, Tomkinson AE, 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. [PubMed: 16982218]

Doublié S (2007). Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. Methods Mol. Biol. 363, 91–108. [PubMed: 17272838]

Doublié S, Bandaru V, Bond JP, and Wallace SS (2004). 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, 10284–10289. [PubMed: 15232006]

Dunn AR, Kad NM, Nelson SR, Warshaw DM, and Wallace SS (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. [PubMed: 21666255]

Emsley P, Lohkamp B, Scott WG, and Cowtan K (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501. [PubMed: 20383002]

Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT, Kikhney AG, Hajizadeh NR, Franklin JM, Jeffries CM, et al. (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J. Appl. Crystallogr. 50, 1212–1225. [PubMed: 28808438]

Franke D, and Svergun DI (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Crystallogr. 42, 342–346. [PubMed: 27630371]

Fromme JC, and Verdine GL (2002). Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. Nat. Struct. Biol. 9, 544–552. [PubMed: 12055620]

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. 6, pl1.

Golan G, Zharkov DO, Feinberg H, Fernandes AS, Zaika EI, Kycia JH, Grollman AP, and Shoham G (2005). Structure of the uncomplexed DNA repair enzyme endonuclease VIII indicates significant interdomain flexibility. Nucleic Acids Res. 33, 5006–5016. [PubMed: 16145054]

Hazra TK, Kow YW, Hatahet Z, Imhoff B, Boldogh I, Mokkapati SK, 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. [PubMed: 12097317]

Hegde ML, Tsutakawa SE, Hegde PM, Holthauzen LM, Li J, Oezguen N, Hilser VJ, Tainer JA, and Mitra S (2013). The disordered C-terminal domain of human DNA glycosylase NEIL1 contributes to its stability via intramolecular interactions. J. Mol. Biol. 425, 2359–2371. [PubMed: 23542007]

Hoffmann S, Smedegaard S, Nakamura K, Mortuza GB, Raschle M, Ibanez de Opakua A, Oka Y, Feng Y, Blanco FJ, Mann M, et al. (2016). TRAIP is a PCNA-binding ubiquitin ligase that protects genome stability after replication stress. J. Cell Biol. 212, 63–75. [PubMed: 26711499]

Imamura K, Averill A, Wallace SS, and Doublié S (2012). Structural characterization of viral ortholog of human DNA glycosylase NEIL1 bound to thymine glycol or 5-hydroxyuracil-containing DNA. J. Biol. Chem. 287, 4288–4298. [PubMed: 22170059]

Imamura K, Wallace SS, 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. [PubMed: 19625256]

Jiang D, Hatahet Z, Blaisdell JO, Melamede RJ, and Wallace SS (1997). Escherichia coli endonuclease VIII: cloning, sequencing, and overexpression of the nei structural gene and characterization of nei and nei nth mutants. J. Bacteriol. 179, 3773–3782. [PubMed: 9171429]

Kabsch W (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132. [PubMed: 20124692]

Kafer GR, Li X, Horii T, Suetake I, Tajima S, Hatada I, and Carlton PM (2016). 5-Hydroxymethylcytosine marks sites of DNA damage and promotes genome stability. Cell Rep. 14, 1283–1292. [PubMed: 26854228]

Katafuchi A, Nakano T, Masaoka A, Terato H, Iwai S, Hanaoka F, and Ide H (2004). Differential specificity of human and Escherichia coli endonuclease III and VIII homologues for oxidative base lesions. J. Biol. Chem. 279, 14464–14471. [PubMed: 14734554]

Kladova OA, Grin IR, Fedorova OS, Kuznetsov NA, and Zharkov DO (2019). Conformational dynamics of damage processing by human DNA glycosylase NEIL1. J. Mol. Biol. 431, 1098–1112. [PubMed: 30716333]

Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, and Svergun DI (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. 36, 1277–1282.

Kontopidis G, Wu SY, Zheleva DI, Taylor P, McInnes C, Lane DP, Fischer PM, and Walkinshaw MD (2005). Structural and biochemical studies of human proliferating cell nuclear antigen complexes provide a rationale for cyclin association and inhibitor design. Proc. Natl. Acad. Sci. U S A 102, 1871–1876. [PubMed: 15681588]

Kuznetsov NA, Koval VV, Zharkov DO, and Fedorova OS (2012). Conformational dynamics of the interaction of Escherichia coli endonuclease VIII with DNA substrates. DNA Repair (Amst) 11, 884–891. [PubMed: 23000248]

Liu M, Bandaru V, Bond JP, Jaruga P, Zhao X, Christov PP, Burrows CJ, Rizzo CJ, Dizdaroglu M, and Wallace SS (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. [PubMed: 20185759]

Liu M, Doublié S, and Wallace SS (2013a). Neil3, the final frontier for the DNA glycosylases that recognize oxidative damage. Mutat. Res. 743744, 4–11. [PubMed: 23274422]

Liu M, Imamura K, Averill AM, Wallace SS, and Doublié S (2013b). Structural characterization of a mouse ortholog of human NEIL3 with a marked preference for single-stranded DNA. Structure 21, 247–256. [PubMed: 23313161]

Makasheva KA, Endutkin AV, and Zharkov DO (2020). Requirements for DNA bubble structure for efficient cleavage by helix-two-turn-helix DNA glycosylases. Mutagenesis 35, 119–128. [PubMed: 31784740]

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ (2007). Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674. [PubMed: 19461840]

Mullins EA, Rodriguez AA, Bradley NP, and Eichman BF (2019). Emerging roles of DNA glycosylases and the base excision repair pathway. Trends Biochem. Sci. 44, 765–781. [PubMed: 31078398]

Osorio A, Milne RL, 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.

Panjkovich A, and Svergun DI (2018). CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. Bioinformatics 34, 1944–1946. [PubMed: 29300836]

Pei J, Kim BH, and Grishin NV (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. 36, 2295–2300. [PubMed: 18287115]

Prakash A, Carroll BL, Sweasy JB, Wallace SS, and Doublié S (2014). Genome and cancer single nucleotide polymorphisms of the human NEIL1 DNA glycosylase: activity, structure, and the effect of editing. DNA Repair (Amst) 14, 17–26. [PubMed: 24382305]

Prakash A, Doublié S, and Wallace SS (2012). The Fpg/Nei family of DNA glycosylases: substrates, structures, and search for damage. Prog. Mol. Biol. Transl. Sci. 110, 71–91. [PubMed: 22749143]

Prakash A, Eckenroth BE, Averill AM, Imamura K, Wallace SS, and Doublié S (2013). Structural investigation of a viral ortholog of human NEIL2/3 DNA glycosylases. DNA Repair (Amst) 12, 1062–1071. [PubMed: 24120312]

Prakash A, Moharana K, Wallace SS, and Doublié S (2017). Destabilization of the PCNA trimer mediated by its interaction with the NEIL1 DNA glycosylase. Nucleic Acids Res. 45, 2897–2909. [PubMed: 27994037]

Rolseth V, Luna L, Olsen AK, Suganthan R, Scheffler K, Neurauter CG, Esbensen Y, Kusnierczyk A, Hildrestrand GA, Graupner A, et al. (2017). No cancer predisposition or increased spontaneous mutation frequencies in NEIL DNA glycosylases-deficient mice. Sci. Rep. 7, 4384. [PubMed: 28663564]

Schneidman-Duhovny D, Hammel M, Tainer JA, and Sali A (2016). FoXS, FoXSDock and MultiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Res. 44, W424–W429. [PubMed: 27151198]

Schomacher L, Han D, Musheev MU, 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. [PubMed: 26751644]

Sharma N, Chakravarthy S, Longley MJ, Copeland WC, and Prakash A (2018). The C-terminal tail of the NEIL1 DNA glycosylase interacts with the human mitochondrial single-stranded DNA binding protein. DNA Repair (Amst) 65, 11–19. [PubMed: 29522991]

Shayevitch R, Askayo D, Keydar I, and Ast G (2018). The importance of DNA methylation of exons on alternative splicing. RNA 24, 1351–1362. [PubMed: 30002084]

Sheldrick GM (2008). A short history of SHELX. Acta Crystallogr. A 64, 112–122. [PubMed: 18156677]

Shen B, Chapman JH, Custance MF, Tricola GM, Jones CE, and Furano AV (2020). Perturbation of base excision repair sensitizes breast cancer cells to APOBEC3 deaminase-mediated mutations. Elife 9, e51605.

Strzalka W, Oyama T, Tori K, and Morikawa K (2009). Crystal structures of the Arabidopsis thaliana proliferating cell nuclear antigen 1 and 2 proteins complexed with the human p21 C-terminal segment. Protein Sci. 18, 1072–1080. [PubMed: 19388052]

Tickle IJ, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C, and Bricogne G (2018). STARANISO (Global Phasing Ltd).

Vonrhein C, Blanc E, Roversi P, and Bricogne G (2007). Automated structure solution with autoSHARP. Methods Mol. Biol. 364, 215–230. [PubMed: 17172768]

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. 67, 235–242. [PubMed: 21460441]

Zhao X, Krishnamurthy N, Burrows CJ, and David SS (2010). Mutation versus repair: NEIL1 removal of hydantoin lesions in single-stranded, bulge, bubble, and duplex DNA contexts. Biochemistry 49, 1658–1666. [PubMed: 20099873]

Zharkov DO, Golan G, Gilboa R, Fernandes AS, Gerchman SE, Kycia JH, Rieger RA, Grollman AP, and Shoham G (2002). Structural analysis of an Escherichia coli endonuclease VIII covalent reaction intermediate. EMBO J. 21, 789–800. [PubMed: 11847126]

Zhu C, Lu L, Zhang J, Yue Z, Song J, Zong S, Liu M, Stovicek O, Gao YQ, 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. [PubMed: 27354518]

Highlights

- This work presents the first crystal structure of a NEIL2 glycosylase
- The structure revealed an unexpected open conformation
- SAXS data illustrate that this DNA repair protein is conformationally dynamic
- Two cancer-associated variants exhibit a decrease in glycosylase activity

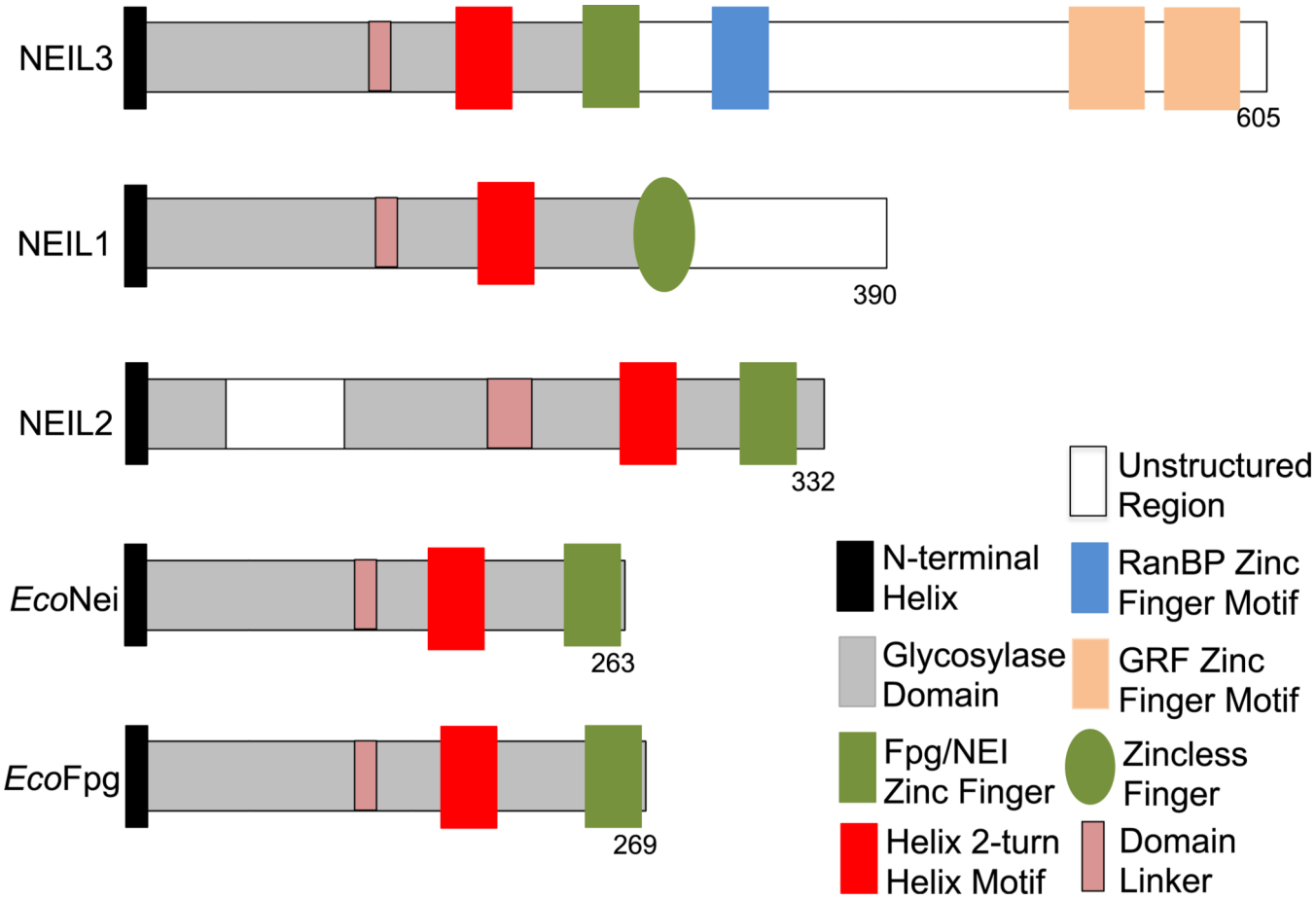


Figure 1. Diagram of Fpg/Nei Glycosylase Domain Composition Highlighting Their Structural Features

NEIL enzymes are substantially larger than their bacterial counterparts. The eukaryotic glycosylases harbor disordered regions. NEIL2 is unique in that the flexible region is internal. Figure adapted from Liu et al. (2010).

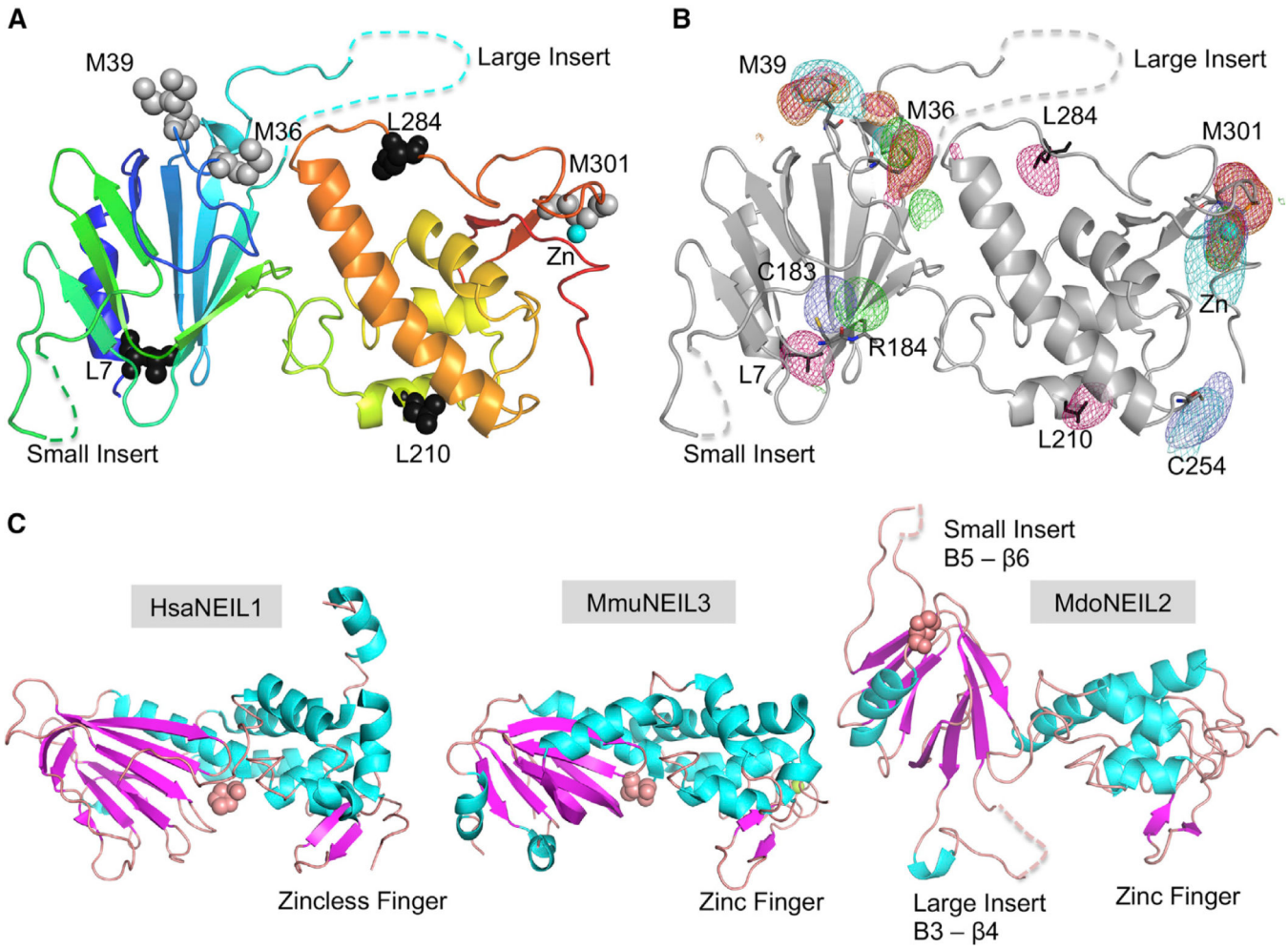


Figure 2. Structure of MdoNEIL2 and Comparison to Other Glycosylases

(A) Cartoon representation of the NEIL2 structure from N terminus (blue) to C terminus (red) with wild-type methionines highlighted with gray spheres and methionines engineered for additional phase verification shown in black.

(B) Cartoon representation of NEIL2 (gray) and overlaid anomalous difference Fourier maps contoured at 3σ for the wild-type SeMet (orange), L-M SeMet (purple), sodium iodide (green), $KAu(CN)_2$ (blue), and K_2PtCl_4 (cyan).

(C) Comparison of MdoNEIL2 to human NEIL1 (HsaNEIL1; PDB: 1TDH) (Doublié et al., 2004) and mouse NEIL3 (MmuNEIL3; PDB: 3W0F) (Liu et al., 2013b) displayed with the C-terminal domain in equivalent orientations and demonstrating the unique interdomain orientation of NEIL2. The N-terminal active-site residue is shown in pink, in space-fill mode. The N-terminal domain large insert (insert 1) and small insert (insert 2) unique to NEIL2 are disordered in the structure and shown as dotted lines.

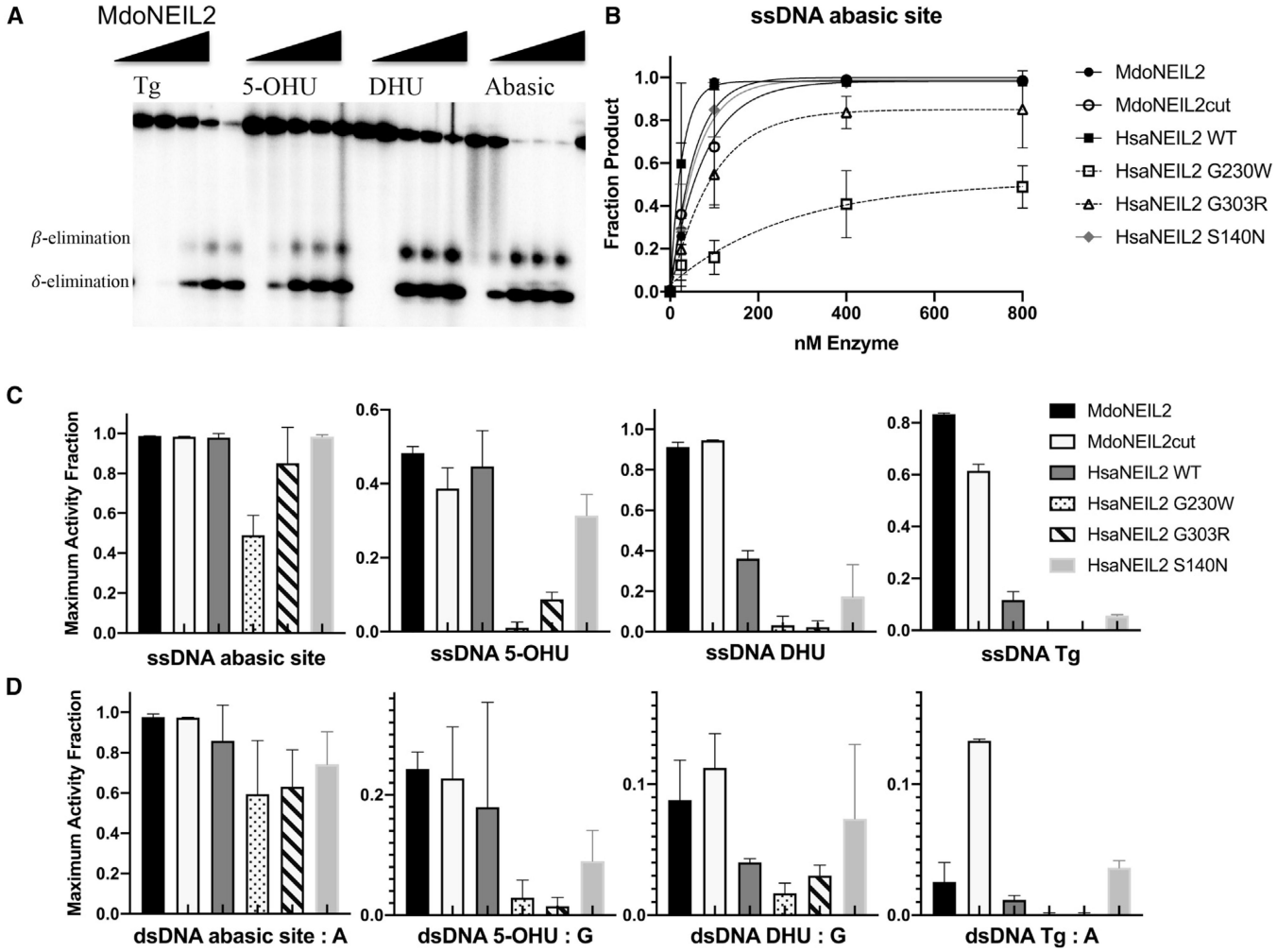


Figure 3. Activity of MdoNEIL2 toward DNA Lesions

(A) Glycosylase activity of MdoNEIL2 for ^{32}P-labeled single-stranded DNA (ssDNA) substrates run on urea-PAGE. Assays were performed using 25 nM DNA substrate and enzyme concentrations of 0, 25, 100, 400, and 800 nM.

(B) Shown is the concentration-based activity titration for single-stranded abasic site substrate with data fit to a single exponential.

(C) Shown is the summary of the maximum activity under the assay conditions expressed as fraction of strand cleavage for all single-stranded substrates andenzyme variants of MdoNEIL2 and HsaNEIL2.

(D) Shown is the summary of the maximum activity for all double-stranded (dsDNA) substrates. Titration curves for all assays are provided in Figure S3.

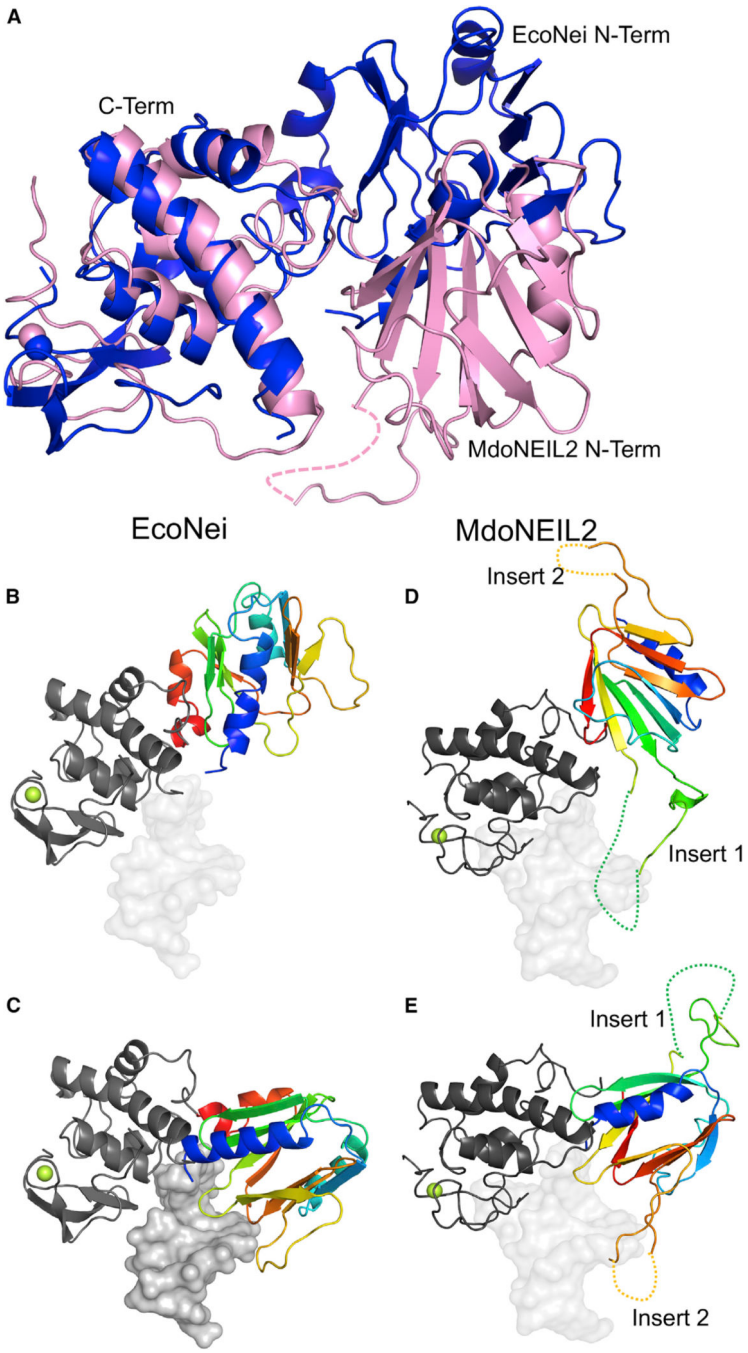


Figure 4. Comparison of the Open and Closed Forms of *E. coli* Endonuclease VIII (EcoNei) with MdoNEIL2

(A) Comparison of the unliganded forms of EcoNei (blue) and MdoNEIL2 (pink) oriented with the C-terminal domain on the left and N-terminal domain on the right. The zinc ions from the zinc finger are shown as spheres; the structures were superimposed on the C-terminal domain.

(B and C) (B) The unliganded EcoNei (PDB: 1Q3B) (Golan et al., 2005) is shown in comparison to (C) DNA-bound EcoNei (PDB: 2EA0) (Golan et al., 2007). Both forms of EcoNei were overlaid based on their C-terminal domains and the DNA duplex for the DNA-

bound form shown in both (B) and (C). A significant rotation (reported to be ~50°) of the N-terminal domain for the unliganded structure would be required for catalytic competency (Golan et al., 2005).

(D and E) (D) The MdoNEIL2 structure is overlaid with the C-terminal expected conformation of (E), a predicted catalytically competent complex. A n angle of ~80° for rotation of the NEIL2 N-terminal domain between the two conformations was determined using Superpose within CCP4 (Winn et al., 2011). For (B)–(E), the C-terminal domain is oriented on the left and in black while the N-terminal domain is oriented to the right and shown in rainbow coloring indicating the N-terminal end containing the catalytic residues P2-E3 as blue and the C-terminal end of the domain as red. The structures in (B), (D), and (E) do not contain DNA, so the duplexes are shown in transparent mode for reference, as the DNA model is from the structure in (C) after C-terminal domain superpositions. The disordered large insert (insert 1) and small insert (insert 2) of the MdoNEIL2 N-terminal domain are shown as dotted lines.

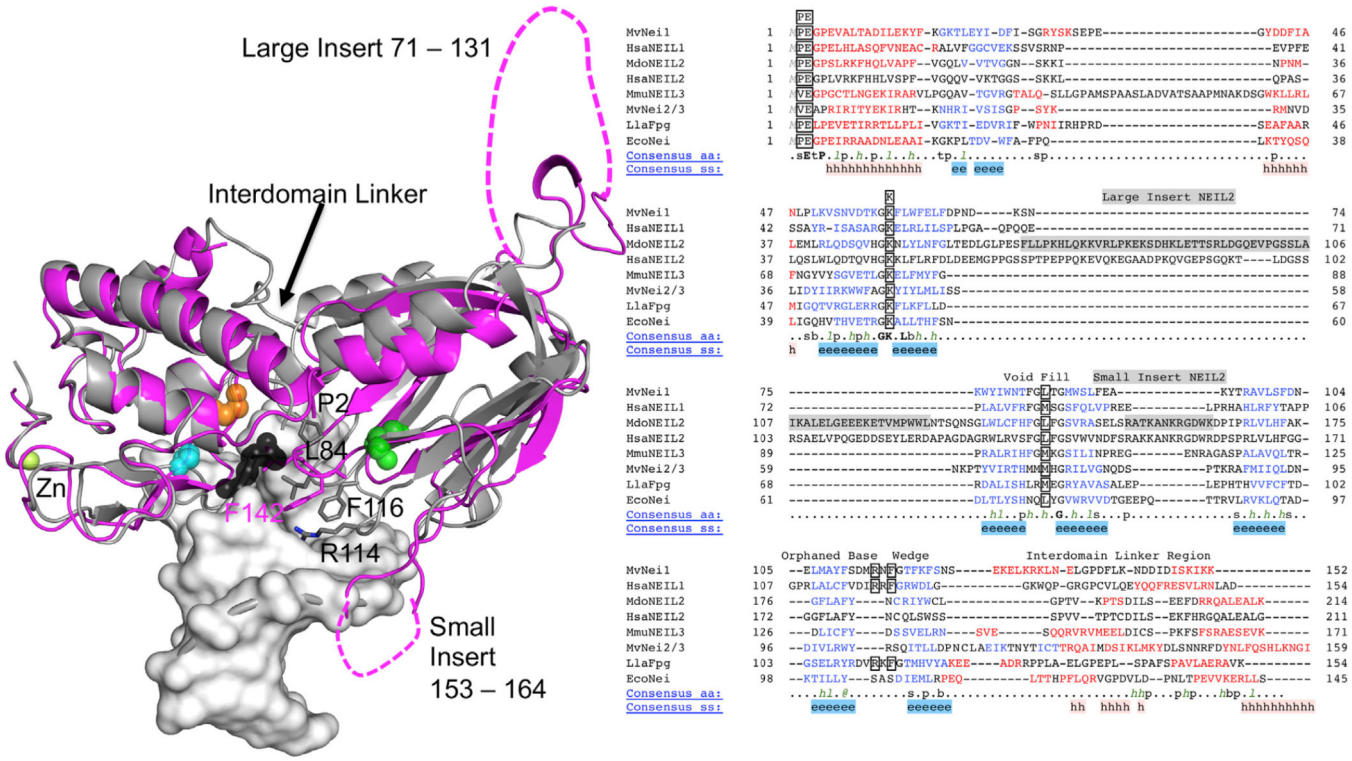


Figure 5. Comparison Using Least-Squares Superposition of the N-terminal Domain of Proteins in the Fpg/Nei Superfamily

The crystal structure of the mimivirus NEIL1 ortholog (MvNei1) (Imamura et al., 2009) is shown in complex with the abasic-site-containing duplex, colored gray, with key loops and the respective residues involved in catalytic function shown: P2-E3 and void-filling residues (L84, R114, and F116). The black spheres represent the abasic-site lesion. The MdoNEIL2 structure is shown in magenta upon independent superposition of the N-terminal and C-terminal domains onto the corresponding domains in MvNei1. Residues corresponding to HsaNEIL2 cancer variants S140 (green), G230 (orange), and G303 (cyan) are highlighted. The structure-based sequence alignment for the Fpg/Nei family was produced using PROMALS3D (Pei et al., 2008). Whereas the loop between β strands 4 and 5 containing the void-filling residue is consistent in length within the Fpg/Nei family, the loop between β strands 7 and 8 is quite variable and is significantly shorter in the NEIL2/3 orthologs. The structures used in the alignment are MvNei1 (PDB: 3A46) (Imamura et al., 2009), HsaNEIL1 (1TDH) (Doublié et al., 2004), MmuNEIL3 (3W0F) (Liu et al., 2013b), MvNei2/3 (4MB7) (Prakash et al., 2013), LlaFpg (1L1T) (Fromme and Verdine, 2002), and EcoNei (2EA0) (deposited but unpublished).

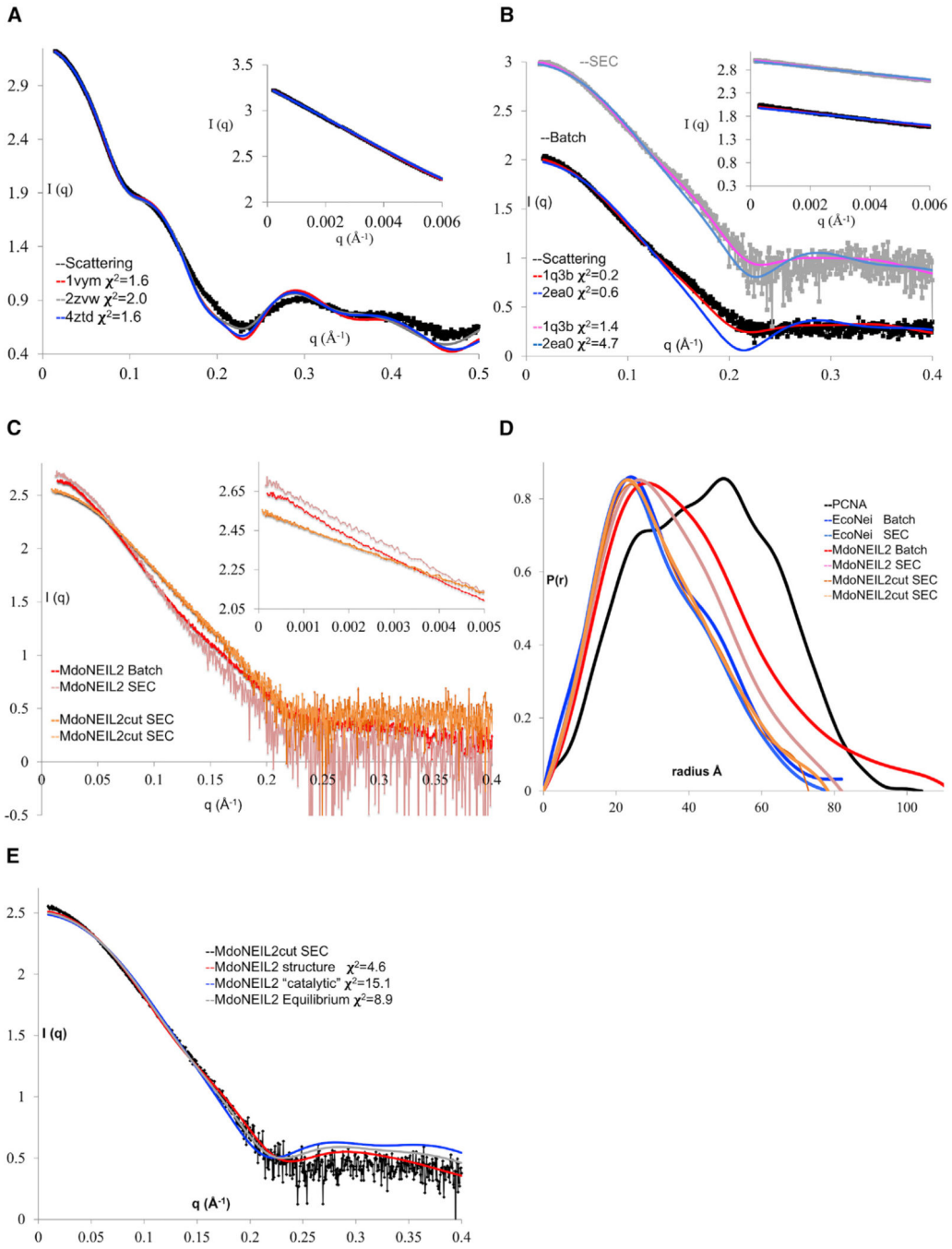


Figure 6. SAXS Analysis for PCNA, EcoNei, MdoNEIL2, and MdoNEIL2cut

(A) Shown is the scattering curve for PCNA (experimental control) and fits to three crystal structures (PDB: 4ZTD [Hoffmann et al., 2016], 2ZVW [Strzalka et al., 2009], 1VYM [Kontopidis et al., 2005]), after omission of any ligands, using FoXS (Schneidman-Duhovny et al., 2016).

(B) Shown is the SAXS profile for EcoNei under batch conditions (black) or upon elution from in-line SEC (gray). The fits shown represent the apo conformation (red and pink) and DNA bound (dark and light blue).

(C) Scattering curves for batch and SEC-SAXS for full-length MdoNEIL2 (red and pink) and SEC-SAXS for MdoNEIL2cut (orange). For (A), (B), and (C), the inset shows the Guinier region.

(D) Distance distribution function for all samples.

(E) FoXS fitting for the MdoNEIL2cut.

Table 1.

Crystallographic Data Collection Statistics Used for Structure Determination of MdoNEIL2

	WT native	WT SeMet peak	WT SeMet remote	WT Zn edge	WT iodide	*WT Platinum peak*[b]	WT Gold peak	L7, 210, 284M SeMet peak
Number of crystals	2	2	2	1	1	*1*	1	1
Wavelength Å	1.033	0.979	0.870	1.28	1.77	*1.072*	1.039	0.979
Space group	P3(2)	P3(2)	P3(2)	P3(2)	P3(2)	*P3(2)*	P3(2)	P3(2)
Cell a = b, c	68.3, 149.2	67.9, 149.1	67.9, 149.3	68.2, 148.2	67.8, 147.3	*68.9, 151.6*	68.0, 148.5	67.7,149.1
Resolution Å (high resolution Å)	58.9–2.63 (2.79–2.63)[a]	40–2.54 2.68–2.54)	34.02–2.75 (2.94–2.75)	59–2.95 (3.05–2.95)	58.9–3.07 (3.36–3.07)	*39–3.89 (4.03–3.89)*	58.9–2.79 (2.98–2.79)	38–3.1 (3.21–3.1)
Resolution a* Å	2.91	2.86	2.96		3.32		3.22	
Resolution c* Å	2.63	2.54	2.75		3.07		2.79	
Reflections	22,928	25,470	20,072	16,233	14,323	*7,451*	18,570	13,806
Completeness %	100 (100)	100 (99.9)	99.8 (100)	99.6 (96.4)	99.1 (100)	*99.6 (99.0)*	96.2 (99.7)	99.6 (96.8)
Multiplicity	16.2 (9.7)	11.0 (9.0)	7.3 (7.2)	5.8 (5.6)	9.1 (7.5)	*3.6 (3.4)*	5.7 (5.9)	5.4 (5.4)
Wilson B	63.2	64.2						
I/Sigmal	14.1 (1.4)	18.7 (0.9)	17.6 (1.1)	7.8 (0.9)	13.1 (1.2)	*6.5 (1.4)*	9.3 (0.8)	8.5 (1.2)
Rmeas %	18.4 (575)	13.0 (376)	10.2 (197)	23.3 (369)	25.3 (245)	*15.4 (91.8)*	16.2 (288)	13.9 (143)
Rpim %	4.4 (155)	3.9 (121)	3.8 (72.8)	9.7 (155)	11.8 (126)	*8.2 (49.8)*	6.9 (117)	6.0 (60.8)
CC ½	99.9 (55.3)	99.8 (46.8)	99.7(60.1)	99.5 (33.0)	99.5(53.1)	*99.3 (29.2)*	99.6 (51.1)	99.7 (48.1)
D-Anom 1/2 sets								
Overall		0.53			0.1		0.15	
Inner shell		0.61			0.24		0.39	
Outer shell		0.01			0.02		0	
Anomalous cutoff Å		3.7			5.4		5.7	
Cross R_{factor} F_{obs} %	-	12.0	13.7	23.2	23.2	*76.0*	26.4	22.5
Phasing power								
Isomorphous		1.30			0.47		0.51	
Anomalous		0.77			0.29		0.48	
Heavy atom sites		10			6		6	
CCall/CCweak		45.2/27.0			34.7/15.3		50.0/30.6	
CFOM		72.2			50.0		80.6	

	WT native	WT SeMet peak	WT SeMet remote	WT Zn edge	WT iodide	*WT Platinum peak*[b]	WT Gold peak	L7, 210, 284M SeMet peak
Correlation		0.35			0.28		0.46	
Phasing R-cullis								
Isomorphous		0.42			0.84		0.92	
Anomalous		0.87			0.98		0.96	
Phasing FOM (E**2 corr)/contrast		0.312						
Density modif.		0.45						
Solvent flat.		3.1						
Initial model								
R/R_{free} %		21.6/39.3						

The data for phase determination were scaled anisotropically with ellipsoidal truncation using STARANISO; anisotropy resolution limits provided along a* and c* for the datasets are indicated. Datasets for validation were processed utilizing Proteum3. Final model refinement was performed using spherically truncated data shown in Table 2. WT, wild type.

[a] Numbers in parentheses denote high-resolution bin.

[b] The Platinum derivative is italicized to indicate that it was used for site verification and not phasing due to lack of isomorphism.

Table 2.

Crystallographic Data Processing Using Spherically Truncated Data and Model Statistics for Structure Refinement of MdoNEIL2 Using Multiple Programs

	Wild-type MdoNEIL2 SeMet peak				**MdoNEIL2cut (Δ67–133)**
PDB ID			6VJI		N/A
Data processing	Proteum3	Proteum3	AutoProc/Aimless	AutoProc/Aimless	Proteum3
Structure refinement	Phenix	Refmac	Phenix	Refmac	Phenix
No. of crystals	1	1	2	2	1
Wavelength A	0.979				1.033
Space group	P3(2)				P3(2)
Cell a = b, c	68.03, 149.3		67.89, 149.1		68.7, 146.5
Resolution A (high resolution A)	46–2.57 2.66–2.57		38–2.54 2.68–2.54		46–3.9 (3.99–3.9)
Reflections	24,594		25,470		7,271
Completeness %	100 (100)[a]		100 (99.9)		99.5 (97.7)
Multiplicity	6.0 (5.9)		11.0 (9.0)		5.1 (4.8)
I/SigmaI	8.4 (0.5)		18.7 (0.9)		3.8 (1.5)
Rmeas %	10.3 (292)		13.0 (376.1)		25.2 (96.3)
Rpim %	4.2 (120)		3.9 (121.3)		11.1 (91.6)
CC ½	0.998 (0.421)		0.998 (0.468)		0.976 (0.428)
Cross R_{factor} F_{obs} %					31.2
Model refinement					
Coordinate error Å	0.45	0.51	0.45	0.55	
R_{work} %	24.6	24.5	25.0	25.9	
R_{free} %	27.4	28.2	27.5	28.7	37.3[c]
RMS bonds Å	0.003	0.002	0.002	0.002	
RMS angles °	0.649	1.08	0.610	1.08	
Ramachandran					
Favored	97.0	94.3	97.6	94.3	
Disallowed	0.2	0.8	0	0.8	
Mean B $Å^2$	97.3	99.5	98.2	101.9	

	Wild-type MdoNEIL2 SeMet peak			MdoNEIL2cut (Δ67–133)
Correlation coeff.	0.700	0.788	0.70	0.788
Figure of merit	0.72	0.732	0.73	0.710
RSRZ outlier %[b]	47.3	38.4	15.5	36.8
Residues built (1–336)			Chain A	Chain B
			5–71	5–71
			130–156	130–154
			163–307	162–307
			311–326	311–325

Collection statistics are also shown for the construct where the large disordered region of the N terminus was removed. The deposited coordinates are from the column with the PDB ID.

[a]Numbers in parentheses denote high-resolution bin.

[b]The RSRZ was calculated via the OneDep Validation Server.

[c]Rigid-body refinement only.

Table 3.

SAXS Data Collection and Analysis Parameters

	PCNA 160323	EcoNei 160524	EcoNei SEC 180309	MdoNEIL2 160323	MdoNEIL2 SEC 180309	MdoNEIL2cut SEC 180309
Data collection parameters						
SASBDB code			SASDJA4		SASDJB4	SASDJC4
Wavelength (Å)	1.127	1.127	1.03	1.127	1.03	1.03
Q range ($Å^{-1}$)	0.013–0.512	0.017–0.323	0.016–0.344	0.019–0.320	0.015–0.323	0.016–0.371
Exposure time (s)	0.5 s/15	0.3 s/10	3 s	0.5 s/15	3 s	3 s
Concentration (mg/mL)	3.2–8.5	0.7–2.3	5	2.6–4.2	5	10
Structural parameters						
I(0) from Guinier	1823 ± 4.8	113.1 ± 1.7	264.9 ± 0.7	489.9 ± 3.6	185.6 ± 0.9	353.1 ± 0.9
R_g (Å) from Guinier	34.6 ± 0.1	24.2 ± 0.5	23.3 ± 0.9	30.9 ± 0.3	26.6 ± 0.6	24.4 ± 0.1
I(0) from P(r)	1817	114.4	268.0	493.8	185.2	354.8
R_g (Å) from P(r)	34.2	25.0	24.0	32.1	26.9	24.9
D_{max} (Å)	105	81.9	78.0	112.2	82.0	78.3
Porod volume (Å)	132,000	48,300	44,600	77,600	57,800	46,800
Molecular weight determination (kDa)						
$MW_{SAXSMoW}$	88.3	28.6	30.4	57.1	41.1	30.6
Expected theoretical	89.1	30.7	30.7	39.0	39.0	32.7

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
E. coli BL21(DE3)	New England Biolabs	C2527
E. coli BL21-Rosetta2(DE3) pLysS	Novagen	71403
Chemicals, Peptides, and Recombinant Proteins		
Uracil DNA Glycosylase	New England Biolabs	M0280
T4 polynucleotide kinase	New England Biolabs	M0201
NdeI	New England Biolabs	R0111
XhoI	New England Biolabs	R0146
Deposited Data		
MdoNEIL2 Crystal Structure	This paper	PDB ID code: 6VJI, rcsb.org
MdoNEIL2 Diffraction Data	This paper	PDB ID code: 6VJI, proteindiffraction.org
MdoNEIL2 SEC-SAXS	This paper	SASDB ID code: SASDJB4,
MdoNEIL2cut SEC-SAXS	This paper	SASDB ID code: SASDJC4
EcoNei SEC-SAXS	This paper	SASDB ID code: SASDJA4
HsaNEIL1 structure	Doublié et al., 2004	PDB ID code: 1TDH
MmuNEIL3 structure	Liu et al., 2013a, 2013b	PDB ID code: 3W0F
MvNei1 structure	Imamura et al., 2009	PDB ID code: 3A46
MvNei2/3 structure	Prakash et al., 2013	PDB ID code: 4MB7
LlaFPG structure	Fromme and Verdine, 2002	PDB ID code: 1L1T
EcoNei structure	Golan et al., 2007	PDB ID code: 2EA0
EcoNei Structure	Golan et al., 2005	PDB ID code: 1Q3B
PCNA structure	Kontopidis et al., 2005	PDB ID code: 1VYM
PCNA structure	Strzalka et al., 2009	PDB ID code: 2ZVW
PCNA structure	Hoffmann et al., 2016	PDB ID code: 4ZTD
Oligonucleotides		
5′-TGTCAATAGCAAG(X)GGAGAAGT CAATCGTGAGTCT-3′(X) = Thymine Glycol, 5′-hydroxyuracil, 5,6-Dihydrouracil, Uracil	Prakash et al., 2014, Midland Certified Reagent Company 2013	N/A
Recombinant DNA		
pET30-HsaNEIL2	This paper	N/A
MdoNEIL2, MdoNEIL2cut	This paper	N/A
pET28a	Novagen	69864
pET30a	Novagen	69909
pTYB2-EcoNei	Bandaru et al., 2006	N/A
pET11a-PCNA	M.T. Washington (University of Iowa, USA)	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
ImageJ	2.0.0-rc-59/1.51n	http://imagej.net
Proteum3	Bruker AXS	https://www.bruker.com/products/x-raydiffraction-and-elemental-analysis/single-crystalx-ray-diffraction/sc-xrd-software/overview/scxrd-software/proteum3.html
ShelX	Sheldrick, 2008	http://shelx.uni-goettingen.de/index.php
AutoSharp/Staraniso	Global Phasing Inc.	https://www.globalphasing.com/
Phenix 1.14	Adams et al., 2010	https://www.phenix-online.org/
CCP4	CCP4 7.0	http://www.ccp4.ac.uk/
Chromixs	ATSAS 2.8.4	https://www.embl-hamburg.de/biosaxs/software.html
PRIMUS	ATSAS 2.8.4	
FoXS/MultiFoXS	Schneidman-Duhovny et al., 2016	https://modbase.compbio.ucsf.edu/multifoxs/
Promals3D	Pei et al., 2008	http://prodata.swmed.edu/promals3d/promals3d.php
Polyphen	Adzhubei et al., 2010	http://genetics.bwh.harvard.edu/pph2/
Other		
TCGA genomics dataset	Gao et al., 2013	https://www.cbioportal.org/