5006–5016 *Nucleic Acids Research, 2005, Vol. 33, No. 15*
doi:10.1093/nar/gki796

Structure of the uncomplexed DNA repair enzyme endonuclease VIII indicates significant interdomain flexibility

Gali Golan, Dmitry O. Zharkov[1,2,3], Hadar Feinberg, Andrea S. Fernandes[3], Elena I. Zaika[3], Jadwiga H. Kycia[4], Arthur P. Grollman[3] and Gil Shoham*

Department of Inorganic Chemistry and the Laboratory for Structural Chemistry and Biology, The Hebrew University of Jerusalem, Jerusalem 91904, Israel, [1]Institute of Chemical Biology and Fundamental Medicine, Siberian Division of the Russian Academy of Sciences, Novosibirsk 630090, Russia, [2]Novosibirsk State University, Novosibirsk 630090, Russia, [3]Laboratory of Chemical Biology, Department of Pharmacological Sciences, Stony Brook University, Stony Brook, NY 11794-8651, USA and [4]Department of Biology, Brookhaven National Laboratories, Upton, NY 11973, USA

Received May 24, 2005; Revised and Accepted August 10, 2005

Protein Data Bank ID codes 1Q39, 1Q3C and 1Q3B

ABSTRACT

***Escherichia coli* endonuclease VIII (Nei) excises oxidized pyrimidines from DNA. It shares significant sequence homology and similar mechanism with Fpg, a bacterial 8-oxoguanine glycosylase. The structure of a covalent Nei–DNA complex has been recently determined, revealing critical amino acid residues which are important for DNA binding and catalysis. Several Fpg structures have also been reported; however, analysis of structural dynamics of Fpg/Nei family proteins has been hindered by the lack of structures of uncomplexed and DNA-bound enzymes from the same source. We report a 2.8 Å resolution structure of free wild-type Nei and two structures of its inactive mutants, Nei-E2A (2.3 Å) and Nei-R252A (2.05 Å). All three structures are virtually identical, demonstrating that the mutations did not affect the overall conformation of the protein in its free state. The structures show a significant conformational change compared with the Nei structure in its complex with DNA, reflecting a ~50° rotation of the two main domains of the enzyme. Such interdomain flexibility has not been reported previously for any DNA glycosylase and may present the first evidence for a global DNA-induced conformational change in this class of enzymes. Several local but functionally relevant structural changes are also evident in other parts of the enzyme.**

INTRODUCTION

Endonuclease VIII (Nei) is a bacterial DNA base excision repair enzyme with both *N*-glycosylase and AP lyase activities. Nei efficiently excises a number of redox-modified pyrimidines, including thymine glycol (Tg), dihydrothymine and dihydrouracil (DHU) (1–3). This substrate specificity overlaps with that of the bacterial endonuclease III (Nth). Initially, Nei was considered to serve as a backup activity for Nth; however, the kinetics of the two enzymes for different substrates vary significantly (2–4). Nei was also shown to excise (though less efficiently) 8-oxoguanine (8-oxoG) and 2,6-diamino-4-hydroxy-5*N*-methyl-formamidopyrimidine (5,6), possibly functioning as a part of the enzymatic system for repair of redox-modified purines (7).

Nei shares significant sequence similarity with bacterial Fpg protein (8), a DNA glycosylase/AP lyase that excises 8-oxoG and formamidopyrimidines from damaged DNA. Both sequence homology and structural similarity put these two proteins into their own Fpg/Nei family of DNA glycosylases. Fpg and Nei initiate cleavage of the *N*-glycosydic bond by a nucleophilic attack at C1′ of the damaged nucleotide, with the absolutely conserved N-terminal proline acting as the nucleophile (9,10). The base excision step is followed by two sequential β-elimination steps (11,12). Three mammalian Nei homologs (NEIL proteins) (13–15) are recent additions to the Fpg/Nei family. Members of this family share the overall fold and several critical motifs required for DNA binding or catalysis, including an N-terminal PE helix, a helix–two turn–helix motif (H2TH), and a zinc finger-type DNA binding domain, although NEIL proteins may lack some of these features (16).

*To whom correspondence should be addressed. Tel: +972 2 6585611; Fax: +972 2 6585319; Email: gil2@vms.huji.ac.il

We have previously reported the three-dimensional structure of *E.coli* Nei covalently cross-linked to 13mer DNA [PDB ID 1K3W, 1K3X (17)]. This structure demonstrated that Nei consists of two domains, N- and C-terminal, connected by a flexible hinge and perfectly positioned to bind DNA between them. Structural motifs and amino acids from both domains are involved in DNA binding and catalysis. Based on this structure and biochemical data, we proposed a detailed catalytic mechanism for Nei, suggesting specific roles of critical amino acids. The involvement of some of these residues in binding, recognition and catalysis have been tested by site-specific mutagenesis [(17), Dmitry O. Zharkov and Arthur P. Grollman, in preparation]. Besides the Nei–DNA covalent complex, reported structures of enzymes of the Fpg/Nei family include uncomplexed Fpg from *Thermus thermophilus* [Tth-Fpg (18)], covalent complexes of Fpg from *E.coli* [Eco-Fpg (19)] and *Bacillus stearothermophilus* [Bst-Fpg (20)], non-covalent complexes of Bst-Fpg and Fpg from *Lactococcus lactis* (Lla-Fpg) with DNA containing various lesions (20–23), and uncomplexed human NEIL1 (24). Notably, the known structures of free and DNA-bound Fpg/Nei proteins come from different species, complicating the analysis of initial non-specific enzyme–DNA interactions, induced conformational changes during DNA binding and potential relevance of interdomain flexibility for binding, recognition and catalysis.

In the present paper, we report the three-dimensional structures of free wild-type (WT) *E.coli* Nei and two catalytically deficient mutants, E2A and R252A. These structures demonstrate that DNA binding by Nei is accompanied with a substantial and unique re-orientation of the two protein domains, as well as several important local conformational changes. To our knowledge, such gross changes have not been observed in other DNA glycosylases. *E.coli* Nei represents the first example of an Fpg/Nei family enzyme structurally determined in both the uncomplexed and complexed forms for the same species, permitting the analysis of the enzyme structural dynamics free of interpretation uncertainty associated with variations in protein sequence and sources of protein.

MATERIALS AND METHODS

Enzymes and oligonucleotides

WT Nei was purified from an overexpressing strain of *E.coli* as described previously (10). Nei mutants E2A and R252A were produced using the QuikChange site-directed mutagenesis kit (Stratagene) and purified in the same way. The sequences of oligonucleotides were identical to those used in the crystallization of the Nei–DNA covalent complex (17).

Crystallization and X-ray diffraction data collection

Crystals of free WT Nei and the two Nei mutants were obtained under two significantly different crystallization conditions, as described in detail elsewhere, together with the respective data collection parameters and data processing statistics (25). Briefly, crystals of WT Nei were obtained by mixing the purified protein (~3 mg/ml) with reservoir solution containing 0.16 M calcium acetate, 0.08 M sodium cacodylate (pH 6.5) and 14.4% (w/v) PEG 8000. Diffraction data for these crystals were collected with a B4-CCD detector (Brandeis University) at the NSLS X25 beamline (Brookhaven National Laboratory, NY) to 2.6 Å resolution. Crystals of uncomplexed Nei-E2A were obtained by mixing the purified protein (~3 mg/ml) with a DHU-containing oligonucleotide duplex in reservoir solution containing 25% PEG 400, 0.1 M Tris–HCl (pH 8.0) and 0.2 M $MgCl_2$. Crystals of free Nei-R252A were obtained by mixing the purified protein (~3 mg/ml) with a Tg-containing duplex in reservoir solution containing 35% PEG 400, 0.1 M Tris–HCl (pH 8.4) and 0.2 M $MgCl_2$. Diffraction data for both Nei-E2A and Nei-R252A were collected on a MAR-CCD (165 mm) detector at the 5-ID-B beamline of the Advanced Photon Source synchrotron facility (Argonne National Laboratory, USA). All the data were reduced and scaled using DENZO and SCALEPACK (26).

Structure determination and refinement

Molecular replacement was used to solve the phase problem of the higher resolution data (Nei-R252A). Only a combined search with separated protein domains of the Nei–DNA complex (1K3X; residues 1–120 and 125–262) as two consecutive reference models was successful. A clear solution was obtained with AMoRe (27), a supported program of the CCP4 package (28). The connecting polypeptide (residues 120–135) was built with the LSQ utility of the program 'O' (29). The structure was further refined with the Crystallography and NMR System (CNS) software (30) employing the maximum likelihood amplitude target procedure. The model was subjected to simulated annealing and iterative cycles of positional and temperature factor refinement, followed by manual fitting and rebuilding with 'O'. Water molecules were assigned to peaks in the (Fo–Fc) difference electron density maps at a contour level >2.8σ, which were also at a suitable distance and orientation to form a hydrogen bond with a potential partner. Glycerol molecules and divalent ions originating from the crystallization and cryo-protectant solutions were added manually and refined using CNS. The progress of the refinement was monitored by following the overall R_{free} value (31), calculated for 10% randomly selected reflections. After the completion of the structure determination of Nei-R252A and its full refinement, this structure was used as a molecular replacement reference model to get initial structures for Nei-E2A and WT Nei, which were further built and refined in a similar way. Examination and validation of local bond lengths and angles, dihedral angles, peptide bond geometry, canonical secondary structures and hydrogen-bonding interactions was done with WHATCHECK (32) and PROCHECK (33).

Graphics and presentations

Matrices for structural superpositions were calculated by least-squares distance minimization algorithm implemented in 'O' using Cα atoms as the guide coordinates. MOLSCRIPT (34), BOBSCRIPT (35), Raster3D (36) and SPOCK (37) were used to create the figures.

RESULTS

Structure determination

Solution and refinement. Attempts to solve the phase problem in any of the three structures by standard molecular replacement methods were unsuccessful with all Fpg/Nei family structures used as references. The phase problem was solved only when we examined the data of Nei-R252A crystals and used a combined molecular replacement search using two separate domains of Nei (from its DNA-bound structure) as reference models. The best combined solution for the Nei-R252A structure resulted in an initial R factor of 28.04% and R_{free} of 33.04%. At later stages of the refinement of Nei-R252A, it became clear that the domain orientation in free Nei is significantly different from that in the Nei–DNA complex, explaining the poor initial results of the first molecular replacement attempts (with the full enzyme). The initial electron density map of the Nei-R252A structure at 2.05 Å resolution accounted for most of the residues of the two domains. More importantly, the map unambiguously revealed the polypeptide chain connecting the two domains, allowing complete amino acid tracing throughout the protein. The refinement of the Nei-R252A structure converged to a final R of 22.8% and R_{free} of 26.0% (Table 1).

The refined Nei-R252A model was used as the reference model to obtain the initial structures of Nei-E2A and WT Nei by molecular replacement (Table 1). In all three structures, most protein residues are experimentally observed, including a polypeptide loop consisting residues Gly-213–Ala-223 which was missing from the structure of the Nei–DNA complex (17).

Table 1. Representative parameters from the crystallographic data processing and refinement of the structures of WT Nei, Nei-E2A and Nei-R252A

	WT-Nei	Nei-E2A	Nei-R252A
Data collection			
Resolution (Å)	30.0–2.60	30.0–2.30	30.0–2.05
(last shell)	(2.64–2.60)	(2.34–2.30)	(2.09–2.05)
Completeness	85.2% (88.8%)	91.7% (88.2%)	99.9% (100.0%)
(last shell)			
R_{merge} (last shell)	9.8% (51.8%)	6.2% (48.7%)	6.0% (42.1%)
I/σ	4.8	6.6	9.4
Structure refinement			
Resolution (Å)	30.0–2.80*	30.0–2.30	30.0–2.05
Completeness	81.7	85.6	99.7
R factor			
Work (%)	21.6	23.9	22.8
Free (%)	29.9	29.9	26.0
No. of atoms			
Protein	2016	1879	1952
Solvent	15	77	133
Zinc ions	1	1	1
Divalent ions	$2Ca^{2+}$	$2Mg^{2+}$	$2Mg^{2+}$
Glycerol molecules	–	3	1
R.m.s.d. bonds (Å)	0.007	0.007	0.007
R.m.s.d. angles (°)	1.3	1.3	1.5
Ramachandran plot			
Most favored (%)	83.3	85.8	89.3
Allowed (%)	15.4	10.8	9.8
Generous (%)	0.5	1.4	0.4
Disallowed (%)	0.9	1.9	0.4

*The resolution of the final refinement of WT-Nei was lowered in order to get a more reliable structure. R_{merge} for the last shell being used for the refinement of this structure (2.84–2.80 Å) is 33.8%.

Several small regions of the polypeptide chain (Pro-1–Gly-3, Pro-83–Thr-85, Leu-249–Arg-252) proved difficult to locate in all three structures due to diffuse electron density and probably a large degree of conformational disorder in the crystal.

Structure validation. Considering the sequential phase determination procedure and the unexpected domain orientation obtained, special efforts were made to validate the overall fold of the free Nei structures. All symmetry operations of the space group (I222) were applied on each of the two separate domains in order to examine other possible cross-domain relationships. Such tests indicated that the current fold is the only way in which the two domains may be positioned relative to each other in the present experimental unit cell. The clarity and interpretation of the experimental electron density was checked through a series of 'omit' maps, where a particular section of the structure was removed, a new electron density was calculated and a new model was built. Such re-building procedure was applied to a number of critical polypeptide sections of the protein with a special attention to the interdomain hinge polypeptide and its neighborhood (residues 115–140). In all cases, the re-built model was found to be identical to the original structure. These procedures clearly ruled out interpretation errors throughout the structure. Additional support for the validity of the reported structures comes from the striking similarity between all three Nei forms, despite significant differences in the way they were crystallized, solved and built.

The overall structure of free Nei

The current study provided three independent crystal structures of free *E.coli* Nei protein: WT Nei and two of its mutants, in which catalytically important residues Glu-2 or Arg-252 were changed to alanines (Nei-E2A and Nei-R252A, respectively). Despite the latter two were set up to co-crystallize with DNA, in all three cases only the protein was found in the crystals, which were isomorphous with identical space groups and crystallographic unit cell dimensions (25). All three Nei structures appeared to be very similar, as demonstrated by the very close overlap of their polypeptide chains (Figure 1a). Thus, for a detailed structural analysis, we have selected the Nei-R252A mutant, which yielded data of higher resolution and completeness, and better internal agreement (25).

The overall size of free Nei is $\sim 68 \times 31 \times 28$ Å, as measured by the maximal distances along the three dimensions of the molecule, a significant deviation from $\sim 57 \times 37 \times 38$ Å reported for Nei complexed with DNA (17). This difference indicates that a global conformational change has occurred in the protein on the way from its free state to the catalytic complex with DNA, transforming the overall shape of the enzyme from an elongated 'open' form to a more compact 'closed' form. A schematic structure of free Nei is presented in Figure 1b. As in the case of DNA-bound Nei (17), two distinct domains are evident, N-terminal (residues 1–120) and C-terminal (residues 135–262), connected with a linker (residues 120–135). Despite the global rearrangement, all secondary structure elements defined in the structure of DNA-bound Nei are retained in the free protein, including the conserved PE helix, the H2TH motif and the zinc finger, all found to participate in important interactions with the bound DNA.

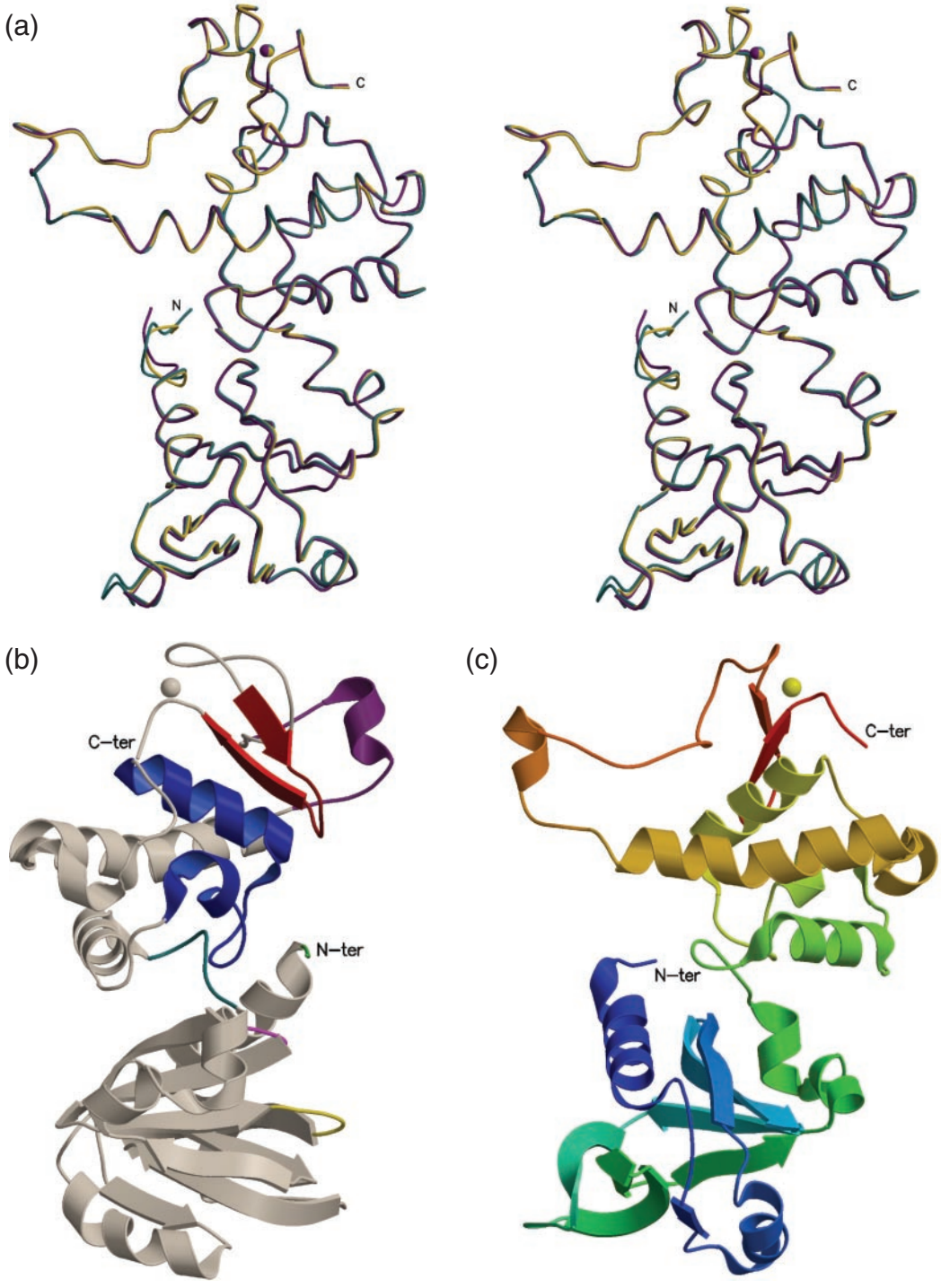

Figure 1. Overall structure of the DNA-free Nei. (**a**) Stereoview comparison of the three Nei structures described in the present paper (main-chain atoms only), demonstrating the high structural similarity between WT Nei (cyan), Nei-E2A (magenta) and Nei-R252A (yellow). (**b**) Schematic ribbon diagram, showing the overall structure of the molecule and its secondary structure elements. This scheme shows the DNA-binding elements of the enzyme, as determined from the Nei–DNA structure (17). These elements include the N-terminal loop (green), Lys-52 (pink), the DNA-inserted tripeptide including Gln69, Leu70 and Tyr71 (yellow), the H2TH motif (blue), and the zinc finger motif (red). The hinge polypeptide is shown in cyan and the 'missing loop' is shown in purple. The zinc atom is represented by a gray sphere (top). (**c**) A similar presentation of free Nei from a view perpendicular to (b). Here the polypeptide chain is 'Rainbow' color-coded, so that it starts (N-terminus) in dark blue and ends in red (C-terminus).

A wide groove formed between the domains (Figure 2) is where the bound DNA is found in the covalent Nei–DNA complex and is probably also used for initial non-specific DNA binding. The DNA-binding face of Nei, and especially the groove, is highly positively charged, explaining the high general affinity of Nei for polyanionic DNA, whereas several patches of negative charge are distributed over the opposite side of the protein.

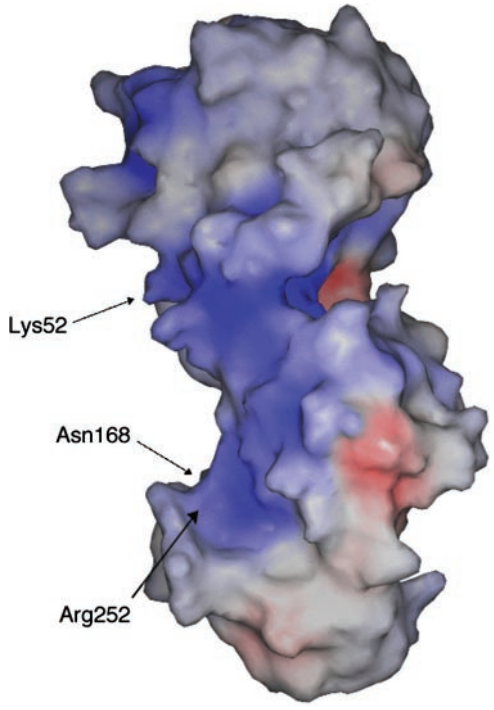

Figure 2. Solvent-accessible surface of free Nei, colored according to electrostatic potential (positive in blue, negative in red and neutral in gray), demonstrating the highly positive DNA-binding cleft of the enzyme (left center). Locations of three important residues involved in DNA binding are indicated by arrows.

Interdomain movement

Domain rotation. Overall least-squares fit of the current free Nei structure with the structure of its covalent DNA complex based on all Cα atoms results in a relatively high r.m.s.d. value of 2.52 Å. Even the visual inspection of the superposition shows very poor overlap between the two structures. In contrast, the individual domains show a relatively good overlap (Figure 3). The superposition of the N-terminal domains (data not shown) results in an r.m.s.d. value of 1.06 Å (Cα atoms of residues 1–120), and that of the C-terminal domains (Figure 3) shows an r.m.s.d. value of 0.98 Å (Cα atoms of residues 135–262). These values clearly indicate that the structures of the individual domains barely differ in the free and DNA-bound enzyme, and that it is the relative position and orientation of the two domains that changes significantly between the two Nei states.

Figure 3 illustrates the very close structural similarity between the C-terminal domains of the free and the DNA-bound Nei structures and the large orientation difference between the N-terminal domains of the two proteins after a C-terminal least-squares fit. A pronounced rotation of the domains (~50°) relative to each other is evident, probably induced by DNA binding. A similar rotation angle has been obtained from the superposition of the N-terminal domains (data not shown). The observed rotation is a combination of a closing movement of the two domains relative to a vertex on the long axis of the molecule, and twisting around the long axis. The vertex point of both bending and twisting motions is located around residues 123–124, at the beginning of the interdomain linker.

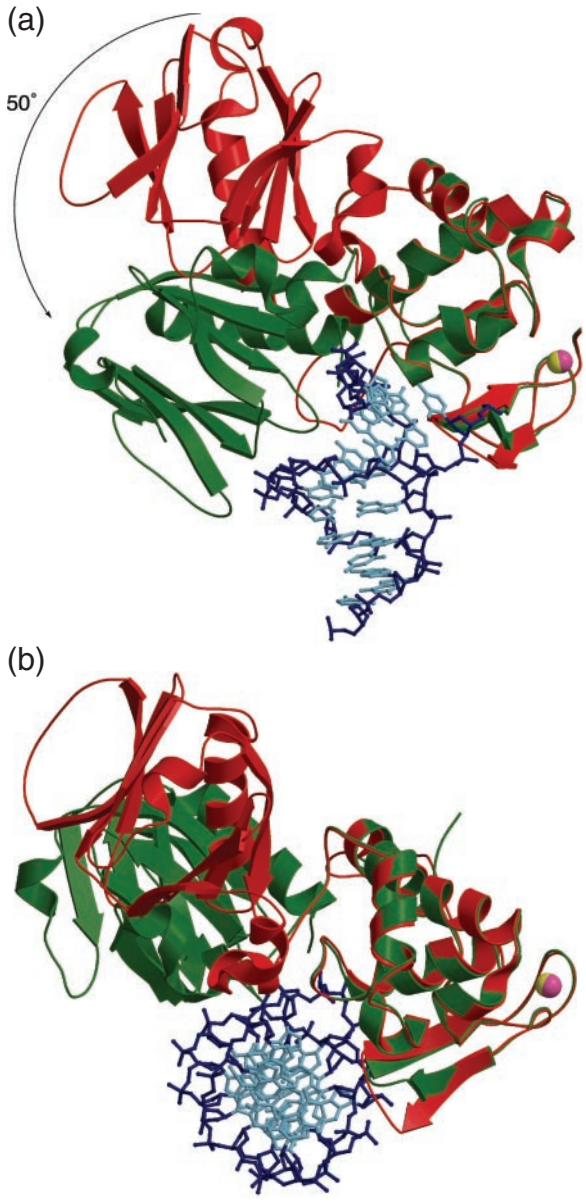

Figure 3. Superposition of free Nei on the Nei–DNA complex, as resulted from a least-squares fit of the C-terminal domains of the free Nei structure (red) with the structure of the Nei–DNA covalent complex (1K3W). In the complex, the protein is shown in a ribbon diagram (green) and the DNA is shown in a ball-and-stick model (backbone, blue; bases, cyan). The zinc atom (bottom right) is shown as a magenta/yellow ball in the free/complex structures, respectively. (**a**) A view perpendicular to the long axis of the Nei molecule in the Nei–DNA complex, demonstrating the 50° rotation of the two domains of the free enzyme relative to each other. (**b**) The same superposition model viewed along the axis of the bound DNA [approximately orthogonal to (a)], demonstrating the induced fit of the free protein onto the DNA in the complex.

Interdomain hinge. The location of the flexible hinge connecting the two domains in the free Nei structure is shown in Figure 4a. This interdomain polypeptide makes a wide loop, the opening of which is facing the side opposite to the DNA-binding groove. Figure 4a also shows the relative rotation of the two domains from the open conformation in the free enzyme to the closed conformation in the DNA-bound enzyme. When the angular opening of the flexible loop

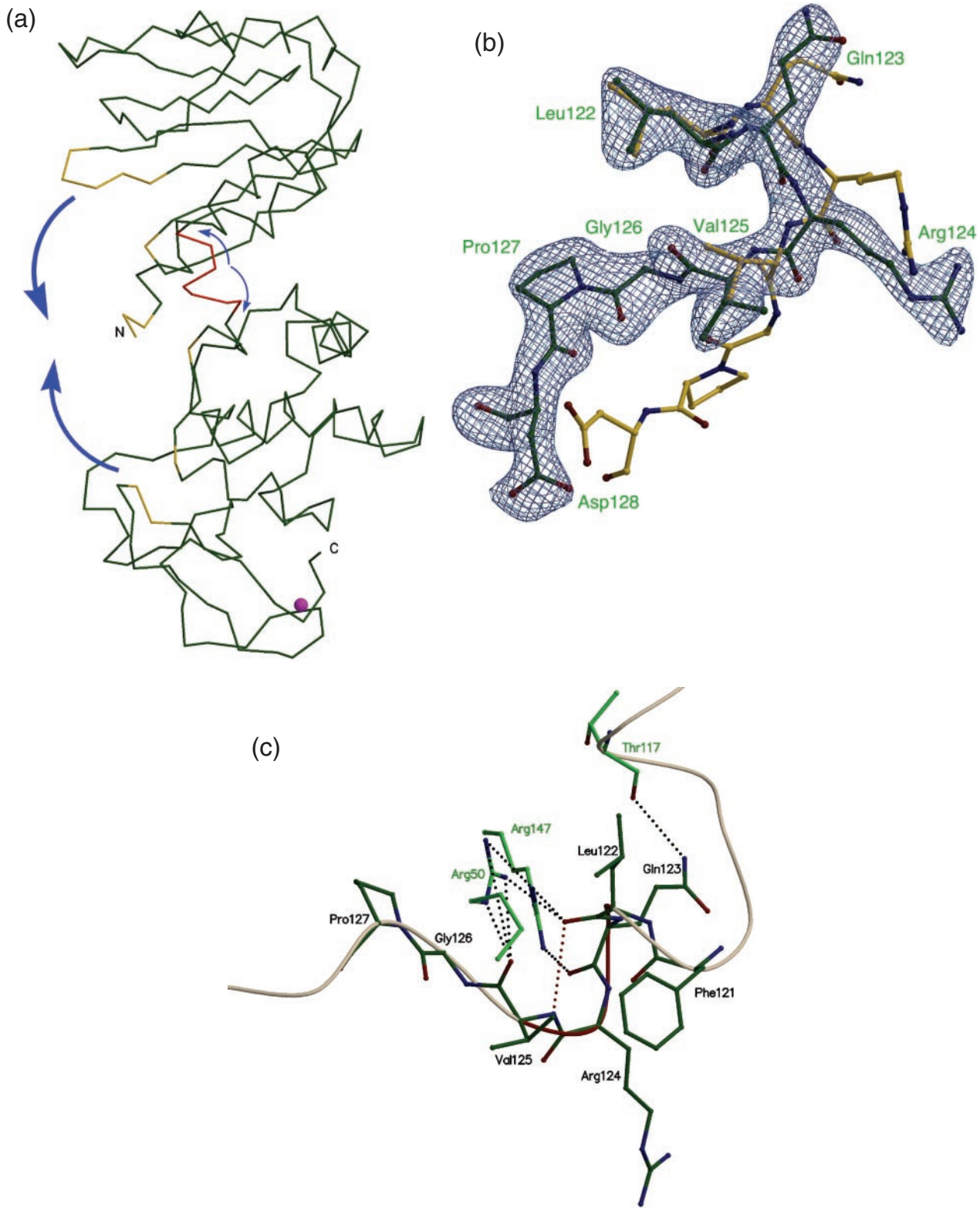

Figure 4. Conformational changes in the flexible hinge connecting the two Nei domains. (**a**) Cα trace of the free Nei structure. The areas involved in interactions with DNA are shown in yellow. The hinge region (residues 121–127) is shown in red (center). The movements of the protein elements from the open to the closed state are indicated by arrows (large arrows, protein domains; small arrows, hinge). (**b**) A close view on the hinge (residues 122–128) in the free Nei structure (carbons in green), with the corresponding section of the experimental electron density map ('omit' map at contour level of 4.0σ, blue), demonstrating the unequivocal model building of this part of the structure. Superimposed on it, is the same polypeptide chain in the Nei–DNA complex (carbons in yellow), showing the major conformational differences in this area between free and complexed Nei. (**c**) The main hydrogen bonds (dotted lines) stabilizing the open state of the free protein. Carbon atoms of the hinge residues are shown in dark green, carbon atoms of the interacting residues, in light green. The main chain of the hinge is colored beige, except for the closed turn (residues 122–125), which is colored purple.

increases, the angular opening between the two domains on the DNA-binding face is decreased, leading to a more compact structure that fits better onto a DNA duplex.

A close view of the interdomain hinge peptide in free Nei is shown in Figure 4b, together with the corresponding section of the electron density map. Although likely to be flexible in solution (see Discussion), this linker seems to be fixed and well defined in the present crystal structure. An overlay of the hinge region of the Nei–DNA complex on that of free Nei shows the local conformational changes that are involved in loop opening in the closed protein form. The specific interactions stabilizing the open free form of the protein which are lost in the closed form in the Nei–DNA complex are shown in Figure 4c. These are mainly the strong hydrogen bonds between the hinge residues and three neighboring residues Arg-50, Thr-117 and Arg-147. Additional stabilization comes from the interaction between Leu-122 and Val-125 within the hinge peptide, fixing a more compact conformation of the turn taken by residues 122–125. Most of these stabilizing interactions are lost upon transition to the closed form. A pronounced twist of the protein main chain takes place around residues Arg-124 and Val-125, moving apart Leu-122 and Val-125, as well as the entire polypeptide chain away from Arg-50, Thr-117 and Arg-147. As a result, the angular opening of the hinge loop widens, and the protein transforms into its closed form.

Interdomain interactions. Given the overall closing movement of the Nei domains upon DNA binding, it is not surprising that many other interactions are formed between them outside of the hinge region. Most of these involve elements from the DNA-binding motifs. For example, a number of strong new hydrogen bonds are formed in the Nei–DNA complex between Glu-2, Glu-5 and the neighboring residues of the H2TH motif at the DNA-binding face of the protein. Some of these new interdomain interactions are mediated by water molecules, and a number of new hydrogen bonds are mediated by the bound DNA. An example is the network of hydrogen bonds that connect residues in the zinc finger motif (e.g. Arg-252) and the H2TH motif (e.g. Asn-168) of the C-terminal domain, through the DNA and water molecules, to Lys-52 from the N-terminal domain. Together, the interactions that form upon DNA binding outnumber the interactions stabilizing the open form of the free enzyme, energetically justifying the transition into the closed form in the complex. The closed form of Nei is absolutely required for catalysis, where protein elements from both domains are concomitantly needed.

Local changes within the two domains

PE helix. In addition to the global interdomain movement, there are several local but clear and functionally relevant conformational changes within each domain of free Nei as compared with the Nei–DNA complex. One of these local differences is observed in the first four N-terminal amino acids (Pro-1–Pro-4). In the present structures of free Nei, the experimental electron density around this region is weak and diffuse, indicating crystallographic disorder or multiple conformations. At lower contour levels of the electron density map, it was nevertheless possible to build most of the N-terminal tetrapeptide at a partial occupancy. This partially occupied conformation is likely to be only one of several possible conformations of the N-terminus, which is probably rather flexible in the uncomplexed enzyme. In contrast, the N-terminal tetrapeptide is very well defined in the structure of the Nei–DNA covalent complex, conformationally fixed through a number of interactions with DNA, including the covalent bond to Pro-1 (17).

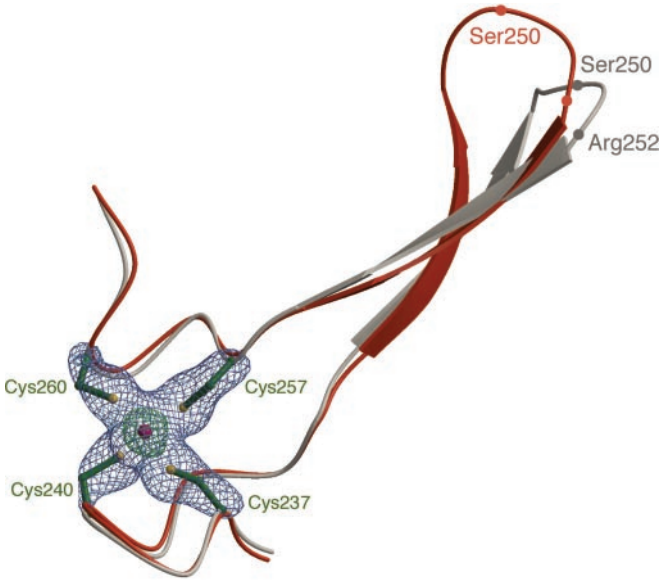

Figure 5. The zinc finger motif of Nei. An electron density map is shown around the bound zinc atom ['omit' map at contour levels of 5σ (blue), and 17σ (green)]. The four Cys residues (with typical Zn–S bond lengths of 2.28–2.38 Å) are shown in a ball-and-stick representation and the two anti-parallel β-strands are shown in a ribbon diagram (beige). Superimposed on this region is the corresponding region of the Nei–DNA complex (red), demonstrating a movement of the tip of the zinc finger between the free and the DNA-bound forms of the enzyme.

Zinc finger. Another local conformational change observed in the present structure of free Nei involves the zinc finger motif (residues 245–256). In the covalent Nei–DNA complex, Arg-252, located on the turn connecting two antiparallel β-strands of the zinc finger, was shown to participate in a number of critical interactions with DNA (17). Similarly, in the N-terminal peptide, the experimental electron density around this region in free Nei is weak and diffuse, indicating considerable flexibility and multiple conformations. Again, it was possible to build the major conformation of this turn at a lower contour level. In this conformation, the turn at the tip of the zinc finger in free Nei structure was shifted by ∼4 Å from its location in the Nei–DNA complex (Figure 5).

The 'missing loop'. In the structure of the covalent Nei–DNA complex, the polypeptide loop consisting of residues 213–223 was crystallographically disordered and could not be built experimentally (and was hence termed the 'missing loop') due to very weak and unclear electron density. This loop had to be therefore reconstructed by molecular modeling to obtain a better picture of the base recognition site (17). Similarly, this loop was disordered in all crystal structures of Fpg bound to AP site (or AP-site analogs) DNA (19–21), but relatively well-ordered in several Fpg/DNA structures with the modified base still present (22,23). In the present structure

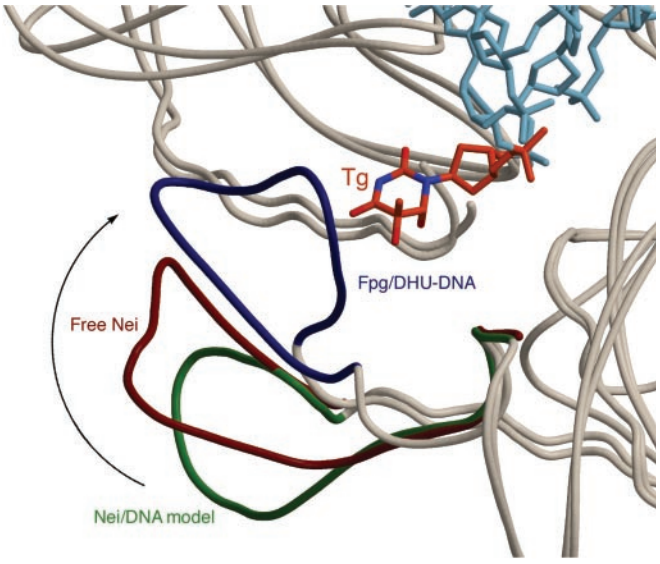

Figure 6. The 'missing' polypeptide loop in Nei and Fpg. A superposition of the polypeptide chain in free Nei (red), and the corresponding region of the DNA-complexed Nei (green) with modeled thymine glycol (Tg, orange; other parts of the DNA are shown in cyan) and the loop region as described previously (17). Also superimposed is the same region in the crystal structure of *Bacillus stearothermophilus* Fpg complexed with DHU-containing DNA (PDB 1R2Z) (22) (blue). Similar parts of the polypeptide chain are shown in khaki. Note the major conformational change of this flexible loop (residues 212–223) in the different structures and upon DNA binding (left), closing onto the modeled Tg binding pocket.

of free Nei, the experimental electron density observed around this region was clear and complete, and the conformation of the peptide could be unequivocally determined. Figure 6 shows a superimposition of this loop in free Nei, the model built for the Nei–DNA complex, and one of the Fpg/DNA structures in which this loop was experimentally observed (1R2Y, Bst-Fpg bound to DHU-containing DNA). Although no definite measurements are possible due to the theoretical nature of the modeled conformation, this comparison illustrates the trend in Nei dynamics opposite to what was seen for the PE helix and the zinc finger, namely, DNA-induced disordering of a previously ordered structural element.

DISCUSSION

The current structure of free Nei provides a first example of an Fpg/Nei family enzyme structurally characterized both as a free protein and a complex with DNA. Hence, useful comparisons can be made between the free Nei structure and the previously reported structure of Nei–DNA covalent complex (17), and between free Nei and other free Fpg/Nei proteins.

In the present work, we have used both WT Nei and two Nei site-directed mutants, E2A and R252A, in which the amino acids forming critical interactions with DNA in the Nei–DNA complex have been eliminated. The catalytic activity of these mutants on Tg-and DHU-containing substrates is greatly reduced (17). Remarkably, all three proteins produced nearly identical structures, differing only in minor details of conformations of a few side chains. This high structural similarity indicates that Glu-2 and Arg-252 are not involved in significant structural stabilization of the enzyme, at least in its free state, and that their mutations abolish Nei catalytic activity by disrupting its active site contacts with DNA.

The most obvious change between free and DNA-bound Nei is the gross closing motion of the protein domains, a 'hinge motion' according to the domain movement classification (38). Although hinge domain motions are known for several DNA-binding proteins, including DNA polymerases and transcription factors, we could not find any other DNA glycosylase for which such a pronounced hinge motion has been reported. Human uracil-DNA glycosylase (UNG) undergoes a hinge movement upon binding damaged DNA, however, this enzyme does not contain two well-separated domains, and the observed motion is much smaller.

Given this degree of conformational movement in Nei, two important questions can be considered. First, is the open state the only conformation adopted by free Nei in solution, or is it one of several possible conformations trapped in the crystal? Second, what possible roles can have the differences in the flexibility of the hinge and several other Nei motifs observed between the free and DNA-bound enzyme?

The only other bacterial base excision repair enzyme of the Fpg/Nei family which has been structurally characterized in its free state is Tth-Fpg (18). Interestingly, free Tth-Fpg has the same closed conformation as DNA-bound Nei (17), Eco-Fpg (19), Bst-Fpg (20,22) and Lla-Fpg (21,23). Despite the differences in protein and DNA sequences, the structures of Fpg in these complexes are very similar, each of them showing little difference from the structure of the DNA-free Tth-Fpg. For example, the overall r.m.s.d. between DNA-bound Eco-Fpg and free Tth-Fpg is 1.45 Å (1.60 Å for the N-terminal domains and 0.89 Å for the C-terminal domains). Hence, the available Fpg structures reveal no major interdomain conformational changes, and definitely nothing that resembles the transformation from the open to the closed form observed for Nei upon DNA binding. Uncomplexed human NEIL1 also appears to adopt the closed form (24). Based on a series of structures of thermophilic Bst-Fpg bound to DNA and on its comparison with Tth-Fpg, it has been suggested that Fpg in general does not undergo notable conformational changes upon DNA binding (20).

A comparison between the structures of free Nei and Tth-Fpg confirms that indeed they have been determined in different overall conformations (Figure 7). As in the case with free versus complexed Nei, a superposition of the two structures has a significantly better domain-by-domain overlap. The r.m.s.d. value calculated for all Cα atoms of free Nei and Tth-Fpg is 2.45 Å, while it is 1.83 Å for their N-terminal domains and 1.42 Å for the C-terminal domains. Therefore, the individual domains of the proteins are reasonably similar, and it is the interdomain movement that accounts for the larger difference between the full-length proteins. A closer examination of the corresponding domains indicates that the main differences between the N-terminal domains are located in the N-terminal PE helix (residues 1–8), demonstrating that this catalytically important element is rather flexible in both proteins. The largest difference between the two C-terminal domains is located around the 'missing loop' (residues 212–226 in Nei), which is longer in Tth-Fpg, contains additional Gly and Pro residues, and assumes a very different conformation, forming different base-binding pockets that underlie the different substrate specificities of the two enzymes.

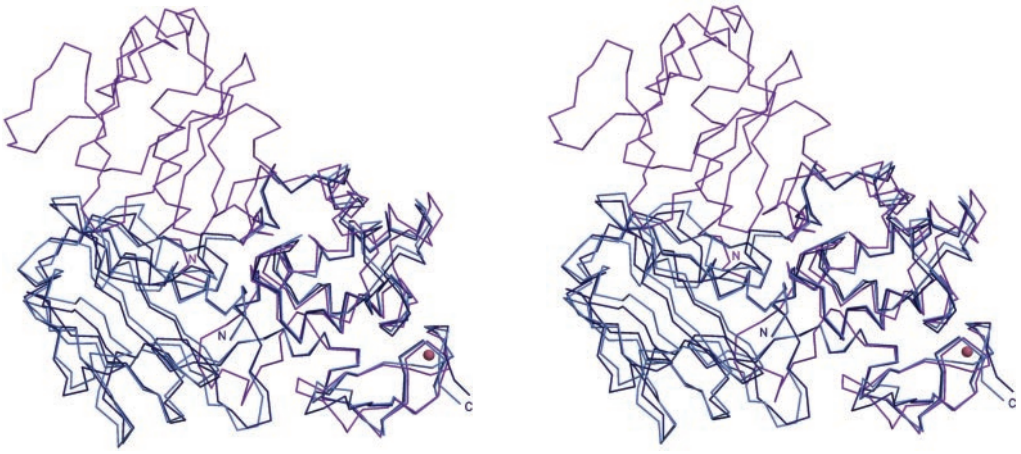

Figure 7. Stereoview superposition of free Tth-Fpg (blue) and free Nei (purple). Two crystallographically independent Tth-Fpg molecules (PDB 1EE8) are shown in dark and light blue, respectively. All three structures (shown here as Cα diagrams) have been superimposed based on their C-terminal domains. Note the relatively small interdomain flexibility of the two Tth-Fpg molecules, in comparison with the significant interdomain rotation in the free Nei structure.

The obvious possible reasons for the difference between free Nei and Tth-Fpg could be the expected higher rigidity of thermophilic Tth-Fpg compared with mesophilic Nei, or differences in sources, sequences and purification procedures of the enzymes. We suggest, however, that all or most of the Fpg/Nei enzymes may in fact possess high interdomain flexibility. NMR evidence indicates high mobility in parts of Eco-Fpg, including most of the hinge region, both with and without DNA (39,40). Very few backbone assignments could be made in the first and the last α-helices (αA and αF) of free Eco-Fpg, indicating that these elements are quite mobile in solution. In free Tth-Fpg and in all DNA-bound Fpg and Nei structures, these two helices partially pack against each other, but this packing is completely disrupted in the open form of free Nei. Thus, the high mobility of αA and αF in free Eco-Fpg suggests that this enzyme partially exists in an open form resembling the DNA-free Nei structure reported here. Stopped-flow measurements with intrinsic tryptophan fluorescence detection also reveal several large movements in Eco-Fpg during substrate processing, some of which have been attributed to the DNA binding step (41,42). Thus, the open and closed interdomain conformations observed, respectively, for Nei and Tth-Fpg, could represent two extremes within this dynamic range of conformations of the free protein. The complexed forms of Fpg/Nei enzymes are probably more rigid due to a large number of interactions with DNA.

The increased flexibility and local structural changes have also been observed in several functionally important parts of Nei, including the N-terminal tetrapeptide and the zinc finger. A lack of rigidity in the former motif, which bears the catalytic nucleophile Pro-1 and the critical carboxyl of Glu-2, may be important for the initial stages of DNA binding and catalysis, where Pro-1 should be positioned in a suitable orientation for a nucleophilic attack on the target sugar. This high flexibility of the N-terminal α-helix in Nei agrees well with the high mobility of the same helix revealed by NMR experiments in free Eco-Fpg (39,40). The zinc finger also has to attain optimal geometry of interaction with the incoming DNA. We have observed movement of the tip of the zinc finger between the open and closed forms of Nei (Figure 5); a similar motion of ∼3 Å has been reported for the non-covalent Lla-Fpg–DNA complex (21) and in several Bst-Fpg complexes (20), in comparison with the structure of free Tth-Fpg. Although in that case the shift could be attributed to the difference between the Fpg sequences, our observation of the same conformational change in Nei suggests that it could be a general feature associated with DNA binding by the enzymes of the Fpg/Nei family. NMR experiments conducted on Eco-Fpg indicate that the very tip of the zinc finger is mobile in the free enzyme, and that this mobility may increase even more when the enzyme binds to DNA (40).

The opposite tendency is evident in the 'missing loop' structure, which is more ordered in free Nei than in the Nei–DNA covalent complex. As mentioned above, a similarly positioned loop was observed in the structure of free Tth-Fpg (18) and NEIL1 (24), as well as in Bst-Fpg and Lla-Fpg complexed with DNA ligands containing a damaged base (8-oxoG or DHU for Bst-Fpg and Fapy-G for Lla-Fpg) (22,23). The loop, however, was not observed in structures of Nei–DNA or Fpg–DNA complexes, covalent or not, containing no DNA base in the active site of the enzyme (17,19–21). As suggested by NMR data, the loop is well-ordered in free Eco-Fpg but becomes partially disordered in the enzyme bound to AP site-containing DNA (40). The sequence of the loop is mostly unconserved in the Fpg/Nei family except for a relatively high content of glycines, which could lead to the high flexibility of the loop observed in the DNA-bound enzymes. There is also a relatively high content of hydrophobic side chains, which may partly account for the higher rigidity of this region in the absence of DNA. In the case of free Nei, the rigidity of this loop is additionally increased by several local interactions with divalent cations (Ca^{2+} in WT Nei; Mg^{2+} in Nei-E2A and Nei-R252A), which have not been found in the DNA-bound form and may have to be displaced during DNA binding.

The increased disorder in the 'missing loop' after base excision may be rationalized as follows. The loop has been experimentally shown to form a part of the base recognition pocket in Bst-Fpg (22) and Lla-Fpg (23); the same role was suggested for Eco-Fpg and Nei based on molecular modeling (17,19,43,44). Although the order of product release in

Fpg/Nei enzymes has not been specifically addressed, in several other DNA glycosylases the base is invariably released first (45,46). Disorder of the loop in all complexes with abasic DNA could thus reflect the flexibility required for the excised base to leave the active site without releasing DNA. The overlay of the loop region (Figure 6) shows that the loop is located on the opposite side of the base recognition pocket, forming a 'gate' which can either lock the everted base in the pocket (as in Bst-Fpg/DHU), or unlock and enable its release from the pocket after its excision from the DNA (as in abasic cross-linked Nei/DNA). This loop is observed in three distinct conformations representing 'snapshots' along its movement between the 'locked' and the 'releasing' states. In the free Nei structure, this region adopts a conformation half-way between the locked and unlocked states, but is also rather open and distant from the C-terminal DNA-binding motifs (H2TH and zinc finger). It is therefore likely that in the absence of a tightly bound base, the 'missing loop' is rather free and mostly found away from the DNA interface.

In summary, the reported structures of free Nei suggest a high degree of interdomain flexibility in the protein and the mobility of several of its critical parts. These features may reflect more general principles of protein–DNA binding. For example, residues at the DNA-binding face of *lac* repressor are flexible in the free protein and in the protein bound to non-specific DNA but become structured upon binding specific DNA sequence (47). A larger number of structures of DNA glycosylases, analyzed in their free and DNA-bound forms, will be required to examine the general applicability of such principles for recognition and repair of damaged DNA.

ACKNOWLEDGEMENTS

We thank the staff at the National Synchrotron Light Source facility (X25 and X26C beamlines) of the Brookhaven National Laboratory and the staff at the Advanced Photon Source synchrotron facility (5-ID-B beamline) of the Argonne National Laboratory for their help in the X-ray synchrotron data collection and analysis. This research was supported in part by grant (CA17395) (to A.P.G) from the National Institutes of Health. D.O.Z. acknowledges support from the Presidium of the Russian Academy of Sciences (program 10.5), Russian Foundation for Basic Research (04-04-48254, 05-04-48619), Russian Science Support Foundation, US Civil Research and Development Foundation, and the Wellcome Trust UK (070244/Z/03/Z). G.G. was supported by fellowships from the Wolf Foundation (Israel), the Chorafas Foundation (Switzerland) and the Clore Foundation (Israel). Funding to pay the Open Access publication charges for this article was provided by HU internal funds.

Conflict of interest statement. None declared.

REFERENCES

1. Melamede,R.J., Hatahet,Z., Kow,Y.W., Ide,H. and Wallace,S.S. (1994) Isolation and characterization of endonuclease VIII from *Escherichia coli*. *Biochemistry*, **33**, 1255–1264.
2. Purmal,A.A., Lampman,G.W., Bond,J.P., Hatahet,Z. and Wallace,S.S. (1998) Enzymatic processing of uracil glycol, a major oxidative product of DNA cytosine. *J. Biol. Chem.*, **273**, 10026–10035.
3. Jiang,D., Hatahet,Z., Melamede,R.J., Kow,Y.W. and Wallace,S.S. (1997) Characterization of *Escherichia coli* endonuclease VIII. *J. Biol. Chem.*, **272**, 32230–32239.
4. Dizdaroglu,M., Burgess,S.M., Jaruga,P., Hazra,T.K., Rodriguez,H. and Lloyd,R.S. (2001) Substrate specificity and excision kinetics of *Escherichia coli* endonuclease VIII (Nei) for modified bases in DNA damaged by free radicals. *Biochemistry*, **40**, 12150–12156.
5. Asagoshi,K., Yamada,T., Okada,Y., Terato,H., Ohyama,Y., Seki,S. and Ide,H. (2000) Recognition of formamidopyrimidine by *Escherichia coli* and mammalian thymine glycol glycosylases. Distinctive paired base effects and biological and mechanistic implications. *J. Biol. Chem.*, **275**, 24781–24786.
6. Hazra,T.K., Izumi,T., Venkataraman,R., Kow,Y.W., Dizdaroglu,M. and Mitra,S. (2000) Characterization of a novel 8-oxoguanine-DNA glycosylase activity in *Escherichia coli* and identification of the enzyme as endonuclease VIII. *J. Biol. Chem.*, **275**, 27762–27767.
7. Blaisdell,J.O., Hatahet,Z. and Wallace,S.S. (1999) A novel role for *Escherichia coli* endonuclease VIII in prevention of spontaneous G→T transversions. *J. Bacteriol.*, **181**, 6396–6402.
8. Jiang,D., Hatahet,Z., Blaisdell,J.O., Melamede,R.J. and Wallace,S.S. (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.
9. Zharkov,D.O., Rieger,R.A., Iden,C.R. and Grollman,A.P. (1997) NH_2-terminal proline acts as a nucleophile in the glycosylase/AP-lyase reaction catalyzed by *Escherichia coli* formamidopyrimidine-DNA glycosylase (Fpg) protein. *J. Biol. Chem.*, **272**, 5335–5341.
10. Rieger,R.A., McTigue,M.M., Kycia,J.H., Gerchman,S.E., Grollman,A.P. and Iden,C.R. (2000) Characterization of a cross-linked DNA-endonuclease VIII repair complex by electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.*, **11**, 505–515.
11. Bailly,V., Verly,W.G., O'Connor,T. and Laval,J. (1989) Mechanism of DNA strand nicking at apurinic/apyrimidinic sites by *Escherichia coli* [formamidopyrimidine]DNA glycosylase. *Biochem. J.*, **262**, 581–589.
12. O'Connor,T.R. and Laval,J. (1989) Physical association of the 2,6-diamino-4-hydroxy-5N-formamidopyrimidine-DNA glycosylase of *Escherichia coli* and an activity nicking DNA at apurinic/apyrimidinic sites. *Proc. Natl Acad. Sci. USA*, **86**, 5222–5226.
13. Bandaru,V., Sunkara,S., Wallace,S.S. and Bond,J.P. (2002) A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to *Escherichia coli* endonuclease VIII. *DNA Repair*, **1**, 517–529.
14. Hazra,T.K., Izumi,T., Boldogh,I., Imhoff,B., Kow,Y.W., Jaruga,P., Dizdaroglu,M. and Mitra,S. (2002) Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc. Natl Acad. Sci. USA*, **99**, 3523–3528.
15. Morland,I., Rolseth,V., Luna,L., Rognes,T., Bjoras,M. and Seeberg,E. (2002) Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic Acids Res.*, **30**, 4926–4936.
16. Zharkov,D.O., Shoham,G. and Grollman,A.P. (2003) Structural characterization of the Fpg family of DNA glycosylases. *DNA Repair*, **2**, 839–862.
17. Zharkov,D.O., Golan,G., Gilboa,R., Fernandes,A.S., Gerchman,S.E., Kycia,J.H., Rieger,R.A., Grollman,A.P. and Shoham,G. (2002) Structural analysis of an *Escherichia coli* endonuclease VIII covalent reaction intermediate. *EMBO J.*, **21**, 789–800.
18. Sugahara,M., Mikawa,T., Kumasaka,T., Yamamoto,M., Kato,R., Fukuyama,K., Inoue,Y. and Kuramitsu,S. (2000) Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, *Thermus thermophilus* HB8. *EMBO J.*, **19**, 3857–3869.
19. Gilboa,R., Zharkov,D.O., Golan,G., Fernandes,A.S., Gerchman,S.E., Matz,E., Kycia,J.H., Grollman,A.P. and Shoham,G. (2002) Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *J. Biol. Chem.*, **277**, 19811–19816.
20. Fromme,J.C. and Verdine,G.L. (2002) Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nat. Struct. Biol.*, **9**, 544–552.
21. Serre,L., Pereira de Jesus,K., Boiteux,S., Zelwer,C. and Castaing,B. (2002) Crystal structure of the *Lactococcus lactis* formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA. *EMBO J.*, **21**, 2854–2865.

22. Fromme,J.C. and Verdine,G.L. (2003) DNA lesion recognition by the bacterial repair enzyme MutM. *J. Biol. Chem.*, **278**, 51543–51548.
23. Coste,F., Ober,M., Carell,T., Boiteux,S., Zelwer,C. and Castaing,B. (2004) Structural basis for the recognition of the FapydG lesion (2,6-diamino-4-hydroxy-5-formamidopyrimidine) by formamidopyrimidine-DNA glycosylase. *J. Biol. Chem.*, **279**, 44074–44083.
24. Doublie,S., Bandaru,V., Bond,J.P. and Wallace,S.S. (2004) The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. *Proc. Natl Acad. Sci. USA*, **101**, 10284–10289.
25. Golan,G., Zharkov,D.O., Fernandes,A.S., Zaika,E., Kycia,J.H., Wawrzak,Z., Grollman,A.P. and Shoham,G. (2004) Crystallization and preliminary crystallographic analysis of endonuclease VIII in its uncomplexed form. *Acta Crystallogr.*, **D60**, 1476–1480.
26. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
27. Navaza,J. and Saludjian,P. (1997) AMoRe: An automated molecular replacement program package. *Methods Enzymol.*, **276**, 581–594.
28. Number 4 Collaborative Computional Project. (1994) The *CCP*4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
29. Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.
30. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) *Crystallography & NMR system*: a new software suite for macromolecular structure determination. *Acta Crystallogr.*, **D54**, 905–921.
31. Brunger,A.T. (1992) Free R-value—a novel statistical quantity for assessing the accuracy of crystal-structures. *Nature*, **355**, 472–475.
32. Hooft,R.W.W., Vriend,G., Sander,C. and Abola,E.E. (1996) Errors in protein structures. *Nature*, **381**, 272.
33. Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) *PROCHECK*: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
34. Kraulis,P.J. (1991) *MOLSCRIPT*—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
35. Esnouf,R.M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.*, **15**, 132–134.
36. Merrit,E.A. and Bacon,D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.
37. Christopher,J.A. (1998) *SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual)*. The Center for Macromolecular Design, Texas A&M University.
38. Echols,N., Milburn,D. and Gerstein,M. (2003) MolMovDB: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res.*, **31**, 478–482.
39. Buchko,G.W., Wallace,S.S. and Kennedy,M.A. (2002) Base excision repair: NMR backbone assignments of *Escherichia coli* formamidopyrimidine-DNA glycosylase. *J. Biomol. NMR*, **22**, 301–302.
40. Buchko,G.W., McAteer,K., Wallace,S.S. and Kennedy,M.A. (2005) Solution-state NMR investigation of DNA binding interactions in *Escherichia coli* formamidopyrimidine-DNA glycosylase (Fpg): a dynamic description of the DNA/protein interface. *DNA Repair*, **4**, 327–339.
41. Fedorova,O.S., Nevinsky,G.A., Koval,V.V., Ishchenko,A.A., Vasilenko,N.L. and Douglas,K.T. (2002) Stopped-flow kinetic studies of the interaction between *Escherichia coli* Fpg protein and DNA substrates. *Biochemistry*, **41**, 1520–1528.
42. Koval,V.V., Kuznetsov,N.A., Zharkov,D.O., Ishchenko,A.A., Douglas,K.T., Nevinsky,G.A. and Fedorova,O.S. (2004) Pre-steady-state kinetics shows differences in processing of various DNA lesions by *Escherichia coli* formamidopyrimidine-DNA glycosylase. *Nucleic Acids Res.*, **32**, 926–935.
43. Zaika,E.I., Perlow,R.A., Matz,E., Broyde,S., Gilboa,R., Grollman,A.P. and Zharkov,D.O. (2004) Substrate discrimination by formamidopyrimidine-DNA glycosylase: a mutational analysis. *J. Biol. Chem.*, **279**, 4849–4861.
44. Perlow-Poehnelt,R.A., Zharkov,D.O., Grollman,A.P. and Broyde,S. (2004) Substrate discrimination by formamidopyrimidine-DNA glycosylase: distinguishing interactions within the active site. *Biochemistry*, **43**, 16092–16105.
45. McCann,J.A.B. and Berti,P.J. (2003) Adenine release is fast in MutY-catalyzed hydrolysis of G:A and 8-oxo-G:A DNA mismatches. *J. Biol. Chem.*, **278**, 29587–29592.
46. Kuznetsov,N.A., Koval,V.V., Zharkov,D.O., Nevinsky,G.A., Douglas,K.T. and Fedorova,O.S. (2005) Kinetics of substrate recognition and cleavage by human 8-oxoguanine-DNA glycosylase. *Nucleic Acids Res.*, **33**, 3919–3931.
47. Kalodimos,C.G., Biris,N., Bonvin,A.M.J.J., Levandoski,M.M., Guennuegues,M., Boelens,R. and Kaptein,R. (2004) Structure and flexibility adaptation in nonspecific and specific protein–DNA complexes. *Science*, **305**, 386–389.