

Viswanath Bandaru, Wendy  
Cooper, Susan S. Wallace and  
Sylvie Doublie\*

Department of Microbiology and Molecular  
Genetics, The Markey Center for Molecular  
Genetics, The University of Vermont, Stafford  
Hall, 95 Carrigan Drive, Burlington,  
VT 05405-0068, USA

Correspondence e-mail: sdoublie@uvm.edu

# Overproduction, crystallization and preliminary crystallographic analysis of a novel human DNA-repair enzyme that recognizes oxidative DNA damage

DNA glycosylases repair oxidative DNA damage caused by free radicals. Recently, NEIL1, a human homolog of *Escherichia coli* DNA glycosylase endonuclease VIII, has been identified and shown to exhibit broad substrate specificity for a variety of types of pyrimidine-base damage. An active C-terminal deletion construct of NEIL1 was overexpressed in *E. coli* and crystallized. The unliganded NEIL1 crystallizes in space group *R*3, with unit-cell parameters  $a = b = 132.2$ ,  $c = 51.1$  Å. Complete data sets were collected from native, selenomethionyl and iodinated NEIL1 to 2.1, 2.3 and 2.4 Å, respectively.

Received 4 March 2004

Accepted 1 April 2004

## 1. Introduction

Reactive oxygen species are generated endogenously; when produced in the vicinity of DNA, these free radicals cause a plethora of types of DNA damage, including base modifications, base loss and single-strand breaks (reviewed in Wallace, 2002). A battery of DNA glycosylases initiate the repair process by first removing the oxidatively damaged bases and then, in a concerted lyase reaction, cleaving the DNA backbone to generate a single-nucleotide gap. DNA polymerization followed by ligation completes the base-excision repair process.

In *Escherichia coli*, two different DNA glycosylases, endonuclease III (EcoNth) and endonuclease VIII (EcoNei), recognize and repair oxidized pyrimidine DNA bases (Wallace, 2002). Single *E. coli* mutants defective in either EcoNth or EcoNei do not show a mutator phenotype; however, *nth nei* double mutants exhibit elevated spontaneous mutation frequencies, suggesting that they serve as backup activities for each other (Blaisdell *et al.*, 1999; Jiang *et al.*, 1997; Saito *et al.*, 1997). hNTH1, a homolog of EcoNth, was the first human pyrimidine DNA glycosylase to be characterized (Aspinwall *et al.*, 1997; Hilbert *et al.*, 1997); we and others have recently identified three new human homologs (NEIL1, 2 and 3) of EcoNei (Bandaru *et al.*, 2002; Hazra, Izumi *et al.*, 2002; Hazra, Kow *et al.*, 2002; Morland *et al.*, 2002; Takao *et al.*, 2002; Wallace *et al.*, 2003). Multiple sequence alignments showed that the active-site residues, the helix–two-turns–helix and zinc-finger motifs are conserved between EcoNei and the human NEIL proteins, with the exception of NEIL1, which lacks the zinc-finger motif (Bandaru *et al.*, 2002). Unlike EcoNei, the NEIL1 amino-

acid sequence does not contain any of the four cysteine residues that coordinate zinc in the *E. coli* enzyme. The substrate specificity of NEIL1, however, resembles that of EcoNei in that NEIL1 recognizes all oxidized pyrimidine bases (Bandaru *et al.*, 2002; Hazra, Izumi *et al.*, 2002). As a first step towards elucidating the nature of substrate recognition by the human NEIL1 enzyme, we have overexpressed, purified and crystallized this glycosylase.

## 2. Material and methods

### 2.1. Protein expression and purification

A pET vector carrying full-length NEIL1 was transformed into Rosetta (DE3)pLysS cells (Novagen) and grown at 289 K for ~16 h. Recombinant C-terminally His-tagged NEIL1 protein (398 amino acids, with an estimated molecular weight of 44.7 kDa) was purified as described previously (Bandaru *et al.*, 2002). Crystallization trials for full-length C-terminally His-tagged NEIL1 failed to yield any crystals. This inability to grow crystals was corroborated by the fact that the protein was polydisperse regardless of the temperature or buffer conditions used, based on dynamic light-scattering (DLS) experiments performed on a DynaPro MS-X instrument (Protein Solutions). Hence, two different approaches were employed to engineer a protein construct that would be more amenable to crystallization.

We had originally identified human NEIL1 in the genomic database with PSI-BLAST using *Arabidopsis thaliana* formamido-pyrimidine DNA glycosylase (AthFpg) as the seed sequence. Subsequently, we cloned and expressed human NEIL1 and showed it to be a homolog of *E. coli* endonuclease VIII

(Bandaru *et al.*, 2002). Moreover, it was known that the C-terminal 109 amino acids of AthFpg are not required for either substrate binding or DNA-glycosylase activity (Ohtsubo *et al.*, 1998). We therefore surmised that based on the sequence alignment of AthFpg and NEIL1, a C-terminal deletion construct of NEIL1 could be designed that might retain glycosylase activity. Multiple sequence alignments of AthFpg and NEIL1 along with other members of the Fpg/Nei family predicted that deletion of the C-terminal 56 amino acids in NEIL1 should not affect the glycosylase activity or DNA-binding capability. Accordingly, a deletion construct of NEIL1 missing the C-terminal 56 amino acids (NEIL1 $\Delta$ 56) was constructed by PCR and subcloned into a pET30a vector (Novagen) between *Nde*I and *Xho*I sites.

In the second approach, a Predictor of Naturally Disordered Regions (PONDR; Li *et al.*, 1999) was used to detect disordered region(s) in NEIL1 that may hinder crystallization. PONDR analysis of AthFpg and NEIL1 showed that both proteins have a disordered C-terminal region (Fig. 1). Interestingly, the disordered region predicted in AthFpg by the PONDR analysis corresponds to the C-terminal 109 residues shown to be dispensable for glycosylase activity (Ohtsubo *et al.*, 1998). Based on this observation, we designed a

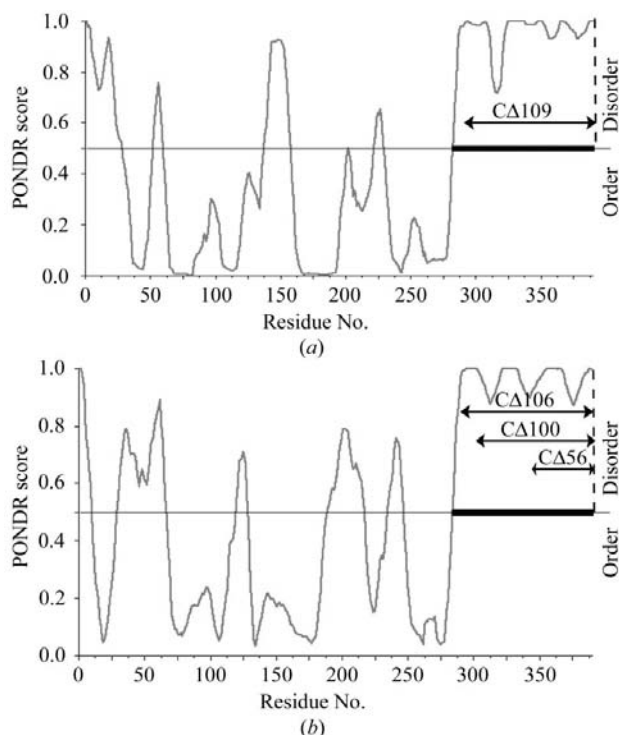
construct missing the entire disordered region (C-terminal 106 amino acids) in NEIL1 (Fig. 1). Little or no soluble protein expression was observed with this construct. A series of shorter C-terminal deletion constructs were cloned and checked for expression. These studies showed that deletions of >100 amino acids did not yield any protein expression. Therefore, a NEIL1 construct missing the C-terminal 100 amino acids (NEIL1 $\Delta$ 100) was constructed for crystallization studies (Fig. 1).

The truncated NEIL1 $\Delta$ 56 and NEIL1 $\Delta$ 100 protein constructs carrying a C-terminal hexa-His tag, which comprises eight amino acids (LEHHHHHH), were expressed and purified as described above. As shown in Fig. 2, both protein constructs were purified to homogeneity. Both freshly prepared and flash-frozen truncated NEIL1 proteins retained DNA glycosylase/lyase activity on a thymine glycol-containing double-stranded substrate, although NEIL1 $\Delta$ 100 was less stable and lost activity over time.

## 2.2. Crystallization and X-ray diffraction experiments

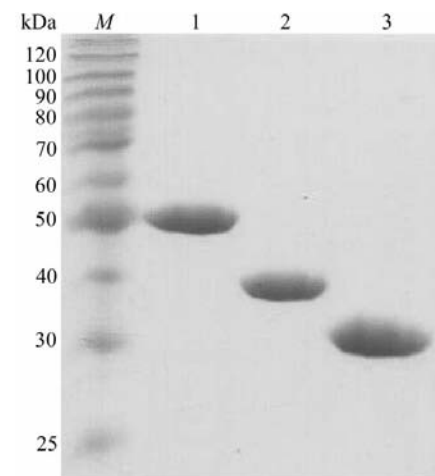
Preliminary crystallization conditions for both NEIL1 $\Delta$ 56 and NEIL1 $\Delta$ 100 were obtained using a sparse-matrix screen (Jancarik & Kim, 1991; Crystal Screen I,

Hampton Research) by the hanging-drop vapor-diffusion method, mixing 1  $\mu$ l protein solution (5 mg ml<sup>-1</sup> in 20 mM HEPES pH 7.6, 150 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 10% glycerol) with 1  $\mu$ l well solution and equilibrating against 0.6–1 ml reservoir solution. The incubation temperatures for the crystal trays were determined based on DLS experiments performed on a DynaPro MS-X instrument (Protein Solutions). Temperatures at which the protein was monodisperse were deemed to be suitable for crystallization. Despite the fact that it was missing most of the C-terminal disordered region, unliganded NEIL1 $\Delta$ 100 protein precipitated in most of the conditions, even at low protein concentration. In contrast, NEIL1 $\Delta$ 56, which still retains some of the region predicted to be disordered (Fig. 1), crystallized in Crystal Screen I condition 17 (30% polyethylene glycol 4000, 0.1 M Tris-HCl pH 8.5, 0.2 M lithium sulfate) at 285 K. After microseeding, unli-



**Figure 1**

PONDR analysis of (a) AthFpg and (b) full-length NEIL1. Regions with a PONDR score of >0.5 are predicted to be disordered. Disordered regions of 40 amino acids or longer are identified by thick lines.



**Figure 2**

Coomassie-stained SDS-PAGE gel of recombinant full-length and truncated NEIL1 proteins. Lane M, molecular-weight standards in kDa; lane 1, full-length NEIL1; lane 2, C-terminal 56-residue deletion construct; lane 3, C-terminal 100-residue deletion construct.



**Figure 3**

Crystals of wild-type NEIL1 $\Delta$ 56. Typical dimensions are 0.2  $\times$  0.15  $\times$  0.05 mm.

**Table 1**  
Data-collection statistics.

Values in parentheses refer to the highest resolution shell.

	Native	SeMet	SeMet-I	I
Resolution (Å)	20–2.10 (2.17–2.10)	20–2.30 (2.40–2.30)	20–2.40 (2.55–2.40)	20–2.30 (2.40–2.30)
Unique reflections	18495	14770	12854	14357
Redundancy	11.9 (11.9)	5.2 (4.7)	6.0 (3.6)	3.0 (2.0)
$R_{\text{merge}}^{\dagger}$	0.085 (0.383)	0.095 (0.437)	0.089 (0.292)	0.078 (0.241)
Completeness (%)	95.1 (100)	99.9 (100)	98.4 (91)	97.0 (85.5)
Overall $I/\sigma(I)$	39.6 (5.6)	13.7 (2.8)	19.5 (2.8)	15.8 (2.5)
$R_{\text{iso}}^{\ddagger}$	—	0.093	0.147	0.158
No. sites	—	4 Se	4 Se, 5 I	8 I

$\dagger R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$ , where  $\langle I \rangle$  is the average intensity from multiple observations of symmetry-related reflections.  $\ddagger R_{\text{iso}} = \sum |F_{PH} - F_P| / \sum |F_P|$ , where  $F_P$  is the observed structure-factor amplitude for the native data set and  $F_{PH}$  is the observed structure-factor amplitude for the heavy-atom derivative.

ganded NEIL1CΔ56 yielded diffraction-quality crystals at 285 K in 2–7 d (Fig. 3)

A selenomethionyl variant of NEIL1CΔ56 was prepared by inhibiting methionine biosynthesis (Doublie, 1997) and was purified as described above. A litre of induced culture yielded ~0.5 mg pure SeMet-NEIL1CΔ56 protein, enough to grow crystals for diffraction studies. Crystallization conditions for SeMet-NEIL1CΔ56 were similar to those used for the native protein.

Crystals were harvested from their crystallization drops and flash-cooled in liquid nitrogen. Data from NEIL1CΔ56 crystals were collected on a MAR345 image-plate detector using Cu  $K\alpha$  radiation generated by a Rigaku HR-300 X-ray generator. Assuming one monomer per asymmetric unit, a Matthews coefficient (Matthews, 1968) of 2.25 Å<sup>3</sup> Da<sup>-1</sup> was calculated, which corresponds to ~45% solvent content. Molecular-replacement attempts were made using other members of the Nei family as search models, but they uniformly failed to produce a clear solution. In addition to producing the selenomethionyl protein, we chose to attempt phasing with halides (Dauter & Dauter, 2001; Dauter *et al.*, 2000), which we performed while waiting for synchrotron beamtime to collect multi-wavelength data from the selenomethionyl crystals. The native NEIL1CΔ56 crystals were soaked in sodium iodide solutions (0.125–0.25 M NaI for 2–15 min; Dauter & Dauter, 2001; Dauter *et al.*, 2000). SeMet crystals were also soaked in iodide solutions, which provided a double derivative (Rould, 1997). All derivative data were collected at 100 K using Cu  $K\alpha$  radiation on a MAR 345 image-plate detector and care was taken to collect accurate and redundant data. A summary of the data-collection statistics is

shown in Table 1. Data were integrated and scaled using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

### 2.3. Phasing

*CNS* was used to locate and refine the four Se sites in the diffraction data collected using Cu  $K\alpha$  radiation (Brünger *et al.*, 1998). The resulting selenium phases were used in isomorphous difference Fourier calculations to identify the iodide sites. All sites were then refined with *SOLVE*; *RESOLVE* was used for density modification (Terwilliger, 2002). It should be noted here that combining the selenomethionyl and iodide derivatives provided enough phasing information to yield an interpretable electron-density map, confirming the usefulness of halide soaks (Dauter & Dauter, 2001; Dauter *et al.*, 2000) and selenium substitution in the phasing of protein structures, even in cases where the number of methionines is below average (one methionine in 86 amino acids compared with an average of one in 50; Lemke *et al.*, 2002). In addition, all of the data sets were collected using Cu  $K\alpha$  radiation, giving credence to the assertion that when selenomethionyl protein crystals are in hand, one should not wait for synchrotron time to attempt phasing, provided of course that the diffraction data are accurate (Lemke *et al.*, 2002). Model building and refinement of the structure are in progress and structural details will be described elsewhere.

We thank Dr Jeffrey Bond for stimulating discussions and Dr Mark A. Rould for critically reading the manuscript. This work was supported by an NIH award (PHS R37CA33657) to SSW. The crystallographic work was supported by an award to the

University of Vermont under the Howard Hughes Medical Institute Biomedical Research Support Program for Medical Schools.

### References

- Aspinwall, R., Rothwell, D. G., Roldán-Arjona, T., Anselmino, C., Ward, C. J., Cheadle, J. P., Sampson, J. R., Lindahl, T., Harris, P. C. & Hickson, I. D. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 109–114.
- Bandaru, V., Sunkara, S., Wallace, S. S. & Bond, J. P. (2002). *DNA Repair*, **1**, 517–529.
- Blaisdell, J. O., Hatahet, Z. & Wallace, S. S. (1999). *J. Bacteriol.* **181**, 6396–6402.
- Brünger, 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., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst. D* **54**, 905–921.
- Dauter, Z. & Dauter, M. (2001). *Structure*, **9**, R21–R26.
- Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst. D* **56**, 232–237.
- Doublie, S. (1997). *Methods Enzymol.* **276**, 523–530.
- Hazra, T. K., Izumi, T., Boldogh, I., Imhoff, B., Kow, Y. W., Jaruga, P., Dizdaroglu, M. & Mitra, S. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 3523–3528.
- Hazra, T. K., Kow, Y. W., Hatahet, Z., Imhoff, B., Boldogh, I., Mokkaipati, S. K., Mitra, S. & Izumi, T. (2002). *J. Biol. Chem.* **277**, 30417–30420.
- Hilbert, T. P., Chung, W., Boorstein, R. J., Cunningham, R. P. & Teebor, G. W. (1997). *J. Biol. Chem.* **272**, 6733–6740.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Jiang, D., Hatahet, Z., Blaisdell, J. O., Melamed, R. J. & Wallace, S. S. (1997). *J. Bacteriol.* **179**, 3773–3782.
- Lemke, C. T., Smith, G. D. & Howell, P. L. (2002). *Acta Cryst. D* **58**, 2096–2101.
- Li, X., Romero, P., Rani, M., Dunker, A. K. & Obradovic, Z. (1999). *Genome Inform. Ser. Workshop Genome Inform.* **10**, 30–40.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Morland, I., Rolseth, V., Luna, L., Rognes, T., Bjaras, M. & Seeberg, E. (2002). *Nucleic Acids Res.* **30**, 4926–4936.
- Ohtsubo, T., Matsuda, O., Iba, K., Terashima, I., Sekiguchi, M. & Nakabeppu, Y. (1998). *Mol. Gen. Genet.* **259**, 577–590.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Rould, M. A. (1997). *Methods Enzymol.* **276**, 461–472.
- Saito, Y., Uraki, F., Nakajima, S., Asaeda, A., Ono, K., Kubo, K. & Yamamoto, K. (1997). *J. Bacteriol.* **179**, 3783–3785.
- Takao, M., Kanno, S. I., Kobayashi, K., Zhang, Q. M., Yonei, S., van der Horst, G. T. J. & Yasui, A. (2002). *J. Biol. Chem.* **277**, 42205–42213.
- Terwilliger, T. C. (2002). *Acta Cryst. D* **58**, 1937–1940.
- Wallace, S. S. (2002). *Free Radic. Biol. Med.* **33**, 1–14.
- Wallace, S. S., Bandaru, V., Kathe, S. D. & Bond, J. P. (2003). *DNA Repair*, **2**, 441–453.