BIOLOGICAL CRYSTALLOGRAPHY
ISSN 1399-0047

Structure determination of uracil-DNA *N*-glycosylase from *Deinococcus radiodurans* in complex with DNA

Hege Lynum Pedersen,[a] Kenneth A. Johnson,[a] Colin E. McVey,[b] Ingar Leiros[a] and Elin Moe[a,b]*

[a]The Norwegian Structural Biology Center (NorStruct), Department of Chemistry, UiT – The Arctic University of Norway, 9037 Tromsø, Norway, and [b]Instituto de Tecnologia Química e Biológica (ITQB), Universidade Nova de Lisboa, 2780-157 Oeiras, Portugal. *Correspondence e-mail: elin.moe@uit.no

Received 17 November 2014
Accepted 27 July 2015

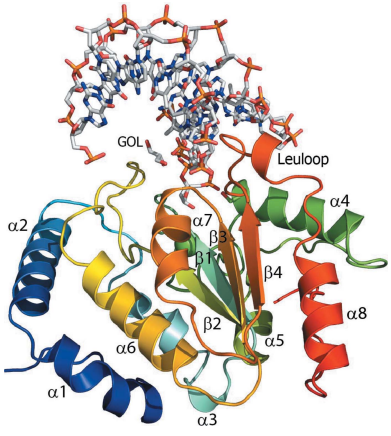
Edited by G. J. Kleywegt, EMBL-EBI, Hinxton, England

Keywords: protein–DNA co-crystal structure; uracil-DNA *N*-glycosylase; *Deinococcus radiodurans*; base-excision repair.

PDB reference: uracil-DNA *N*-glycosylase in complex with DNA, 4uqm

Supporting information: this article has supporting information at journals.iucr.org/d

Uracil-DNA *N*-glycosylase (UNG) is a DNA-repair enzyme in the base-excision repair (BER) pathway which removes uracil from DNA. Here, the crystal structure of UNG from the extremophilic bacterium *Deinococcus radiodurans* (*Dr*UNG) in complex with DNA is reported at a resolution of 1.35 Å. Prior to the crystallization experiments, the affinity between *Dr*UNG and different DNA oligonucleotides was tested by electrophoretic mobility shift assays (EMSAs). As a result of this analysis, two 16 nt double-stranded DNAs were chosen for the co-crystallization experiments, one of which (16 nt AU) resulted in well diffracting crystals. The DNA in the co-crystal structure contained an abasic site (substrate product) flipped into the active site of the enzyme, with no uracil in the active-site pocket. Despite the high resolution, it was not possible to fit all of the terminal nucleotides of the DNA complex into electron density owing to disorder caused by a lack of stabilizing interactions. However, the DNA which was in contact with the enzyme, close to the active site, was well ordered and allowed detailed analysis of the enzyme–DNA interaction. The complex revealed that the interaction between *Dr*UNG and DNA is similar to that in the previously determined crystal structure of human UNG (hUNG) in complex with DNA [Slupphaug *et al.* (1996). *Nature (London)*, **384**, 87–92]. Substitutions in a (here defined) variable part of the leucine loop result in a shorter loop (eight residues instead of nine) in *Dr*UNG compared with hUNG; regardless of this, it seems to fulfil its role and generate a stabilizing force with the minor groove upon flipping out of the damaged base into the active site. The structure also provides a rationale for the previously observed high catalytic efficiency of *Dr*UNG caused by high substrate affinity by demonstrating an increased number of long-range electrostatic interactions between the enzyme and the DNA. Interestingly, specific interactions between residues in the N-terminus of a symmetry-related molecule and the complementary DNA strand facing away from the active site were also observed which seem to stabilize the enzyme–DNA complex. However, the significance of this observation remains to be investigated. The results provide new insights into the current knowledge about DNA damage recognition and repair by uracil-DNA glycosylases.



1. Introduction

Uracil-DNA *N*-glycosylase (UNG) is part of the base-excision repair (BER) pathway which is highly conserved from bacteria to man, and primarily repairs endogenous DNA damage such as deamination, alkylation, oxidation and single-strand breaks in genomic DNA (reviewed in Krokan & Bjørås, 2013). UNG removes uracil in DNA, which may occur by both the deamination of cytosine (Lindahl & Nyberg, 1974) and misincorporation during replication (Tye *et al.*, 1977). The

crystal structure of the catalytic domain of UNG is known from human (Mol, Arvai, Slupphaug *et al.*, 1995), herpes simplex virus 1 (HSV-1; Savva *et al.*, 1995), *Escherichia coli* (Saikrishnan *et al.*, 2002), Atlantic cod (Leiros *et al.*, 2003), *Deinococcus radiodurans* (Leiros *et al.*, 2005), *Vibrio cholerae* (Raeder *et al.*, 2010), *Mycobacterium tuberculosis* (Kaushal *et al.*, 2010), *Bacillus subtilis* (Baños-Sanz *et al.*, 2013), *Staphylococcus aureus* (Wang *et al.*, 2014), *Leishmania naffi* (PDB entry 3cxm; Structural Genomics of Pathogenic Protozoa Consortium, unpublished work) and *Coxiella burnetii* (Franklin *et al.*, 2015). All 11 enzymes have conserved structures and consist of a classic single-domain α/β-fold with a central four-stranded parallel and twisted β-sheet surrounded by 8–11 α-helices. The N- and C-termini are on opposite sides of the central β-sheet and the active site is located within a positively charged groove at the C-terminal end of the β-sheet.

A common characteristic of UNGs is their inhibition by Ugi (Zharkov *et al.*, 2010), and a large number of crystal structures have been determined of UNG in complex with Ugi (Mol, Arvai, Sanderson *et al.*, 1995; Savva & Pearl, 1995; Putnam *et al.*, 1999; Ravishankar *et al.*, 1998; Saikrishnan *et al.*, 2002; Géoui *et al.*, 2007; Kaushal *et al.*, 2008; Baños-Sanz *et al.*, 2013; Wang *et al.*, 2014; Assefa *et al.*, 2014). The complex structures have revealed that the UNG–Ugi interaction closely resembles the interaction of UNG with DNA and has provided an alternative strategy for understanding the nature of UNG–DNA interaction. However, in order to obtain a full understanding of DNA-damage identification it is crucial to study the crystal structures of UNG in complex with DNA.

To date, crystal structures of UNG–DNA complexes have only been determined for human UNG (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998; Parikh *et al.*, 2000). These crystal structures revealed that the uracil is flipped out of the DNA base stack and into the active site of UNG prior to catalysis (Slupphaug *et al.*, 1996). Furthermore, they suggested that the detection of uracil involves phosphate-backbone compression, minor-groove binding, abasic gap plugging and protein–DNA attraction in a 'pinch–push–plug–pull' mechanism (Parikh *et al.*, 1998, 2000; Stivers, 2004). Three serine-rich and/or proline-rich loops are involved in this process: the 4-Pro loop (165-PPPPS-169), the Gly-Ser loop (246-GS-247) and the minor-groove intercalation loop (leucine loop; 268-HPSPLSVYR-276) (hUNG numbering). The serines in these loops (Ser169, Ser247, Ser270 and Ser273) form hydrogen bonds to the phosphates 5′ and 3′ to uracil, and participate in orienting the enzyme correctly for DNA scanning. A fourth loop, the water-activating loop (143-GQDPYH-148) is also important in positioning an Asp residue for catalysis. The initial backbone compression is suggested to be coupled to a minor-groove reading head formed by the leucine-loop residues Tyr275 and Arg276 in hUNG, which make water-bridged hydrogen bonds to structurally conserved purine N3 sites and widen the minor groove. When the uracil is detected, it is flipped out of the DNA by further backbone compression, penetration of Leu272 into DNA and pulling of the uracil into the specificity pocket (Parikh *et al.*, 1998). It has further been suggested that Leu272 has an additional role in plugging the cavity in DNA after uracil flipping, thereby increasing the lifetime of the extrahelical base (Jiang *et al.*, 2001).

D. radiodurans is a pigmented pink/orange bacterium which was first identified in 1956 in canned meat sterilized by ionizing radiation (Anderson *et al.*, 1956). *D. radiodurans* exhibits an outstanding resistance to ionizing radiation and desiccation and tolerates radiation doses of up to 5000 Gy without loss of viability. Most other organisms cannot survive doses above 50 Gy (Mattimore & Battista, 1996). Such a massive radiation dose is estimated to induce several hundred double-strand breaks (DSBs), thousands of single-strand gaps and about 1000 sites of DNA-base damage per chromosome. The annotated sequence of the *D. radiodurans* genome was published in 1999 (White *et al.*, 1999) and allowed a detailed analysis of the genomic composition of this organism. The resistance mechanism of *D. radiodurans* is not yet fully known, but initial investigations have suggested that it is complex and most likely determined by a combination of factors such as genome packing, cell structure and a highly efficient DNA-repair machinery (White *et al.*, 1999; Makarova *et al.*, 2001; Liu *et al.*, 2003; Levin-Zaidman *et al.*, 2003).

An unusually high number of DNA glycosylases have been identified in the genome of *D. radiodurans* (Makarova *et al.*, 2001) and five of them have been characterized as uracil-DNA glycosylases: uracil-DNA *N*-glycosylase (*Dr*UNG/DR_0689), mismatch-specific uracil-DNA glycosylase (*Dr*MUG/DR_0751), thermophilic uracil-DNA glycosylase (*Tm*UDG/DR_1751) and two hypothetical UDGs (DR_0022 and DR_1663). We have cloned the genes encoding all of the uracil-DNA glycosylases from *D. radiodurans* and have obtained crystals and determined the crystal structures of *Dr*UNG (Leiros *et al.*, 2005) and *Dr*MUG (Moe *et al.*, 2006). The results of structure–function analysis of these proteins revealed that they possess catalytic modifications compared with non-extremophilic proteins, which may optimize the DNA-repair efficiency and repertoire of this organism. Early studies of uracil-DNA glycosylase (UDG) activity in *D. radiodurans* have also shown that the addition of Ugi to crude extracts resulted in an ∼95% reduction of the overall UDG activity, indicating that *Dr*UNG is the main uracil-DNA glycosylase in this organism (Sandigursky *et al.*, 2004).

The structures of hUNG–DNA co-crystals have all been determined with ∼10 bp dsDNA oligonucleotides and to a minimum resolution of 1.9 Å (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998, 2000). Here, we present the first high-resolution (1.35 Å) co-crystal structure of a bacterial UNG in complex with a 16 nt dsDNA oligonucleotide with an abasic site (a processed AU base pair). The structure of *D. radiodurans* UNG (*Dr*UNG) in complex with DNA shows that despite several amino-acid substitutions in catalytically important loops, the protein–DNA interaction and most likely the DNA-repair mechanism of *Dr*UNG are similar to those of the previously studied hUNG (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998). A comparative analysis of the *Dr*UNG–DNA structure with hUNG–DNA structures (PDB entries 4skn, 1ssp and 2ssp) provides support for a hypothesis that the DNA-backbone compression and minor-groove stabilization by the

leucine loop is an important part of the damage-identification mechanism of UNGs. It provides further support to the theory that the previously observed high catalytic efficiency (Leiros *et al.*, 2005) is caused by additional positively charged residues around the DNA-binding cleft which result in the high substrate affinity of *Dr*UNG compared with hUNG.

2. Materials and methods

2.1. Cloning, expression and purification of *Dr*UNG

The gene encoding the *D. radiodurans* uracil-DNA *N*-glycosylase *Dr*UNG (DR_0968) has previously been cloned into a pDEST14 Gateway expression vector (Invitrogen) with nucleotides encoding an N-terminal six-histidine tag (His tag) using Gateway technology (Leiros *et al.*, 2005). Prior to this work, *Dr*UNG was recloned into the same vector with the addition of nucleotides encoding a tobacco etch virus (TEV) protease cleavage site between the His tag and the *Dr*UNG gene. The primers used for amplification of the gene were as follows (Sigma–Aldrich): FPdrUNGHISTEV, 5′-CATCACC-ATCACCATCAC*GAAAACCTGTATTTCCAGGGAGCA*-**ACCGACCAACCCGACCTG**-3′, RPdrUNG, 5′-GGGGAC-CACTTTGTACAAGAAAGCTGGGTCCTATTCCTCCGT-CACCGTGGC-3′, and FPdrHISTAG, 5′-GGGGACAAGTT-TGTACAAAAAAGCAGGCTTCGAAGATAGAACCATG-CATCACCATCACCATCAC-3′. The gene was first amplified using the FPdrUNGHISTEV primer in order to engineer the N-terminal hexahistidine tag (underlined) and the TEV cleavage site (italicized) and the RPdrUNG primer. The final gene product was further amplified using the FDRHISTAG primer (containing the *att*B1 site and nucleotides encoding the His tag) and the RPdrUNG primer and used in a BP reaction along with the pDONR201 vector and in an LR reaction along with the pDEST14 vector. The sequences of the clones were confirmed by DNA sequencing using the BigDye v.3.1 sequencing protocol and a 3130xl Genetic Analyzer (Applied Biosystems).

The expression and purification was performed as described previously (Leiros *et al.*, 2005) with some modifications. The cells were resuspended in cell-lysis buffer (50 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl) with one tablet of cOmplete Mini EDTA-free protease-inhibitor cocktail (Roche Applied Science) before disruption by sonication using a Sonics VC Ultrasonic processor (Sonics & Materials). After isolation of the recombinant protein from the cell extract by centrifugation, the protein solution was filtered using a 0.45 µ*M* syringe filter (Millipore). The filtrate was loaded onto a 1 ml HisTrap HP column equilibrated with 50 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl. Fractions containing *Dr*UNG with a His tag and a TEV cleavage site were mixed with 1:10(*w*:*w*) TEV protease (18 mg ml^{-1}) followed by cleavage of the His tag and dialysis in 50 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl at 4°C overnight. The TEV protease and His tag were removed from the *Dr*UNG protein after cleavage by HisTrap purification using the protocol described previously (Leiros *et al.*, 2005). Fractions of the flowthrough containing *Dr*UNG without the His tag were pooled and loaded onto a 1 ml HP Q Sepharose column equilibrated with 50 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl and eluted with a gradient from 0 to 100% buffer *B* (50 m*M* Tris–HCl pH 7.5, 1 *M* NaCl). Fractions containing purified *Dr*UNG in 50 m*M* Tris–HCl pH 7.5, 300 m*M* NaCl were concentrated to 5 mg ml^{-1} and stored at 4°C.

Oligo name	Nucleotides and nucleotide numbering																			Chain
	5′-	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16			B
		3′-	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	-5′	C
16 nt	5′-	C	C	T	G	T	C	C	A	**Ab**	G	T	C	T	C	C	G	-3′		
ApG*		3′-	G	A	C	A	G	G	T	**G**	C	A	G	A	G	G	C	G	-5′	
1 6nt		5′-	C	T	G	T	C	C	A	**Ab**	G	T	C	T	C	C	G	-3′		
ApG		3′-	G	A	C	A	G	G	T	**G**	C	A	G	A	G	G	C	-5′		
16 nt	5′-	C	C	T	G	T	C	**C**	**A**	**U**	**G**	**T**	**C**	**T**	**C**	**C**	G	-3′		
AU*		3′-	G	A	C	A	**G**	**G**	**T**	**A**	**C**	**A**	**G**	**A**	**G**	**G**	C	G	-5′	
16 nt		5′-	C	T	G	T	C	C	A	**U**	G	T	C	T	C	C	G	-3′		
AU		3′-	G	A	C	A	G	G	T	**A**	C	A	G	A	G	G	C	-5′		
12 nt			5′-	T	G	T	C	C	A	**Ab**	G	T	C	G	T	-3′				
ApG*				3′-	C	A	G	G	T	**G**	C	A	G	C	C	A	-5′			
12 nt			5′-	C	G	T	C	C	A	**U**	G	T	C	T	C	-3′				
AU*				3′-	C	A	G	G	T	**A**	C	A	G	A	G	G	-5′			
10 nt			5′-	T	G	T	C	C	A	**Ab**	G	T	C	-3′						
ApG*				3′-	C	A	G	G	T	**G**	C	A	G	A	-5′					
10 nt					5′-	G	A	T	G	**ΨU**	T	T	A	G	C	-3′				
ΨU					3′-	C	T	A	C	**G**	A	A	T	C	G	-5′				
9 nt				5′-	G	T	C	C	A	**Ab**	G	T	C	-3′						
ApG				3′-	C	A	G	G	T	**G**	C	A	G	-5′						
9 nt					5′-	C	G	G	G	**U**	G	G	C	G	-3′					
UG						3′-	C	C	C	**G**	C	C	G	C	G	-5′				

Figure 1
Oligonucleotides tested for complex formation with *Dr*UNG. The oligonucleotide (oligo) shaded in dark grey (16 nt AU*) is the oligo which co-crystallized with *Dr*UNG, and the bases in white letters were included in the final model of the structure. All of the oligonucleotides in bold have an identical core sequence. The nucleotides that are shaded in light grey are within the boundaries of interacting with specific amino acids in *Dr*UNG according to a *NUCPLOT* analysis (Luscombe *et al.*, 1997) of the *Dr*UNG–DNA crystal structure. An asterisk indicates oligonucleotides with a 1 nt overhang.

Table 1
Crystallographic data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

PDB code	4uqm
Beamline	ID29, ESRF
Space group	$P2_12_12$
Unit-cell parameters (Å)	a = 84.3, b = 98.72, c = 43.94
Data-collection statistics	
Wavelength (Å)	1.0000
Resolution range (Å)	30–1.35 (1.38–1.35)
No. of unique reflections	78013 (5484)
Completeness (%)	95.7 (91.7)
R_{merge}† (%)	4.5 (59.8)
Wilson B factor (Å^2)	25.1
Average multiplicity	4.4 (4.4)
Mean $\langle I/\sigma(I)\rangle$	13.0 (2.15)
Refinement and model statistics‡	
Resolution range (Å)	27–1.35
R_{work}§ (%)	18.3
R_{free}§ (%)	20.5
R.m.s.d. from ideal bond lengths¶ (Å)	0.014
R.m.s.d. from ideal bond angles¶ (°)	1.717
Clashscore	3.36
Validation by *MolProbity*	
Ramachandran favoured (%)	97.83
Ramachandran outliers (%)	0
Rotamer outliers (%)	1
Overall rank†† (percentile)	96th

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an individual measurement of reflection hkl and $\langle I(hkl)\rangle$ is the mean intensity of the same reflection. ‡ Five TLS groups for protein and four for DNA; isotropic B factors. § $R_{work} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$; R_{free} is the same but for 5% of the total reflections that were never used in refinement. ¶ Engh & Huber (1991). †† For 3057 structures with resolution 1.35 ± 0.25 Å, where the 100th percentile contains the structures with the best validation scores using the *MolProbity* criteria.

2.2. Electrophoretic mobility shift assay (EMSA)

The binding of *Dr*UNG to DNA was investigated using EMSA. 60 µ*M* *Dr*UNG was incubated in binding buffer (2.5 m*M* HEPES pH 7.5, 5 m*M* NaCl, 2.5 m*M* $MgCl_2$, 1% glycerol) for 10 min before being mixed with 50 µ*M* of each dsDNA substrate (Fig. 1) and incubated for 1 h. The DNA oligonucleotides were ordered from Sigma–Aldrich. Both incubations were performed at 4°C. The *Dr*UNG and DNA samples were run on 1% agarose gel in cold 0.5× Tris–borate pH 8.0 buffer to analyse the binding of *Dr*UNG to each dsDNA.

2.3. DNA purification prior to co-crystallization

The DNA oligonucleotides used for co-crystallization were dissolved and annealed in buffer consisting of 25 m*M* HEPES pH 7.5, 50 m*M* NaCl. A 1 ml Mono Q 5/50 GL column (GE Healthcare) equilibrated with 50 m*M* HEPES pH 8.0 and 50 m*M* NaCl was used for purification of the dsDNAs. The flow was maintained at 1 ml min^{-1} and the absorbance was monitored at 260 and 280 nm. The bound DNA was eluted with 50 m*M* HEPES pH 8.0, 1 *M* NaCl over a gradient from 0 to 100% buffer *B*.

2.4. Crystallization

Crystals of the *Dr*UNG–DNA complex were grown in 0.05 *M* sodium citrate pH 4.6, 20% PEG 3350. The sequence of the DNA which resulted in crystals was the 16 bp AU dsDNA (5′-CCTGTCCAUGTCTCCG-3′) with a 1 bp overhang on the 5′ ends and an adenine opposite to the uracil (Fig. 1). *Dr*UNG and DNA were mixed in a 1:1.1 ratio and incubated for 20 min on ice before the *Dr*UNG–DNA and reservoir solutions were mixed in a 1:1.5 ratio. Crystals grew to full size within 3 d at room temperature. The *Dr*UNG–DNA crystals were flash-cooled in liquid nitrogen in 100 m*M* sodium citrate pH 4.5, 12.5% glycerol, 125 m*M* NaCl, 22% PEG 3350.

2.5. Data collection and structure determination

An in-house data set was collected to 1.6 Å resolution (not shown) from a single crystal at cryogenic temperature (93 K) on a R-AXIS IV^{++} double image-plate detector system (Rigaku). The same crystal was used to collect a 1.35 Å resolution data set on beamline ID29 at the European Synchrotron Radiation Facility at cryogenic temperature (100 K) using a Pilatus 6M detector (Dectris) (Table 1). Both data sets were processed and scaled using the *XDS* program suite (Kabsch, 2010). The structure was solved by molecular replacement using the home data set, with the structure of the *D. radiodurans* UNG apoprotein as a search model (PDB entry 2boo; 230 amino acids, residues Pro17–Glu246; Leiros *et al.*, 2005). Calculation of the Matthews coefficient (Matthews, 1968) suggested that only one molecule of protein would be present in the asymmetric unit. For molecular replacement, *Phaser* (McCoy *et al.*, 2007) from the *CCP*4*i* graphical user interface (Potterton *et al.*, 2003) to the *CCP*4 suite of crystallographic programs (Winn *et al.*, 2011) was used to search for one protein molecule in the eight possible orthorhombic space groups. A correct solution identifying the two screw axes was found, and the data were re-indexed to the conventional setting. Refinement was carried out with *REFMAC*5 (Murshudov *et al.*, 2011) in the *CCP*4 program suite with 5% of the data chosen randomly for use as a set of test reflections that were not used in refinement. All inspection and building of the structure was performed in *Coot* (Emsley *et al.*, 2010). After two cycles of refinement at 2.7 Å resolution, R factors of R_{work} = 36.1% and R_{free} = 45.1% were obtained (see Table 1 for definitions of R_{work} and R_{free}). The double-stranded DNA bound to *Dr*UNG was clearly visible in the electron-density maps and difference maps, and was built using the DNA molecule found in the structure of the human hUNG–DNA complex (PDB entry 1emh; Parikh *et al.*, 2000) as a guide. Water molecules were added and validated using *Coot*. In addition, one molecule of glycerol and one chloride ion were placed in the structure. For further refinement, two translation–libration–screw (TLS) groups consisting of the protein chain and the two DNA chains were chosen. For the home 1.6 Å resolution data, R factors of R_{work} = 17.1% and R_{free} = 20.5% were attained. When the synchrotron data to 1.35 Å resolution were obtained, the structure was further refined using both TLS and anisotropic B-factor refinement strategies in *REFMAC*5. The free set of reflections was the same as in the 1.6 Å resolution home data but extended to 1.35 Å resolution for the new data.

After several iterations of refinement and rebuilding, a final model comprising 230 amino acids (Arg16–Val244), 32 DNA bases, 350 water molecules, one chloride anion and one glycerol molecule that exhibited excellent geometry as analyzed by the *MolProbity* server (Chen *et al.*, 2010) was obtained with final R factors of R_{work} = 18.3% and R_{free} = 20.5% (Table 1). The structure and the 1.35 Å resolution data have been deposited as PDB entry 4uqm.

2.6. Sequence analysis

A structural sequence alignment of the catalytic domains of structurally determined uracil-DNA *N*-glycosylases was prepared using *Expresso* (Armougom *et al.*, 2006) and generated by *ESPript* 3 (Gouet *et al.*, 1999).

3. Results and discussion

3.1. Identification of candidate DNA for co-crystallization

In order to identify the best candidate dsDNA for co-crystallization experiments, EMSA experiments (1% agarose gel) were performed with different oligonucleotides as described in Fig. 1. Different lengths (nine, ten, 12 and 16 nucleotides), overhangs and base damages (tetrahydrofuran abasic site, uracil and pseudouracil) were tested. The results showed that *Dr*UNG had the strongest affinity for the 16 nt ApG and AU oligonucleotides; thus, these were selected for the co-crystallization experiments (Fig. 2). Crystals were obtained with *Dr*UNG and a 16 nt AU oligonucleotide with 5′ overhang ends and were subjected to data collection for structure determination.

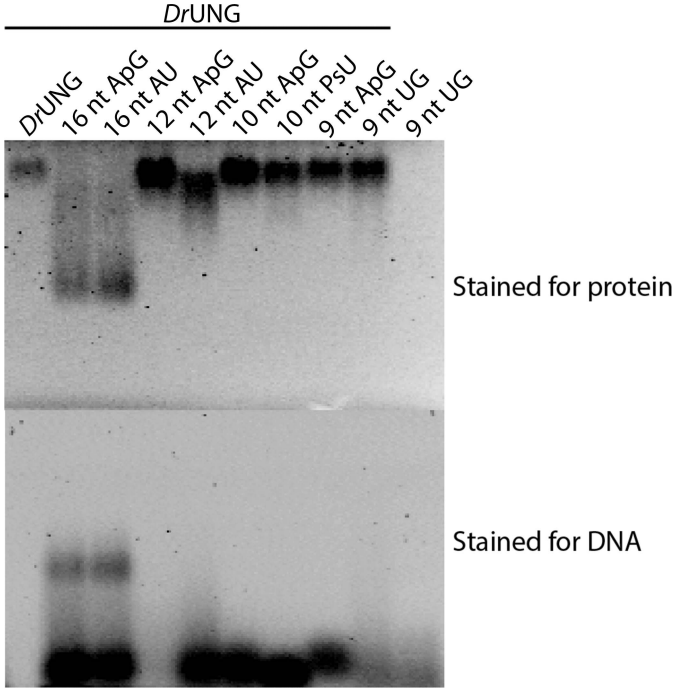


Figure 2
EMSA of *Dr*UNG in complex with different dsDNAs. The interactions of *Dr*UNG with dsDNAs ranging from 9 to 16 nt were explored using EMSA. The nucleotide length and sites for repair are indicated at the top of the figure. The EMSA gel was stained for protein visualization (SYPRO Ruby) and nucleotide visualization (SYBR Green) as indicated to the right of the figure. *Dr*UNG has an apparent preference for the 16 nt dsDNA with an apurinic or a uracil site. 16 nt dsDNA with a 1 nt overhang was selected for co-crystallization experiments.

3.2. The overall structure

The *Dr*UNG–DNA co-crystal structure was determined by molecular replacement to a resolution of 1.35 Å. The final model of *Dr*UNG consisted of 230 amino-acid residues in a single polypeptide chain comprising residues Arg16–Val244 of the amino-acid sequence and DNA (nucleotides 7–15 in chain *B* paired with nucleotides 12–3 in chain *C*; Fig. 1), a glycerol molecule and a chloride anion. The overall structure of the protein–DNA complex can be seen in Fig. 3. The DNA base stack is generally well defined in electron density (Supplementary Fig. S1), but some of the terminal nucleotides are disordered owing to a lack of stabilizing interactions with the protein and thus were difficult to fit into the electron density despite the high resolution of the structure (Supplementary Figs. S2*a*, S2*b* and S2*c*). However, the nucleotides which are in close proximity to the enzyme and the active site are very well defined and enabled us to perform a thorough analysis of the enzyme–DNA interactions of the complex. The glycerol molecule is found in the interface between the enzyme and the DNA and is stabilized through hydrogen bonds to Gln82 and a phosphate group of G10 on the damage-containing DNA

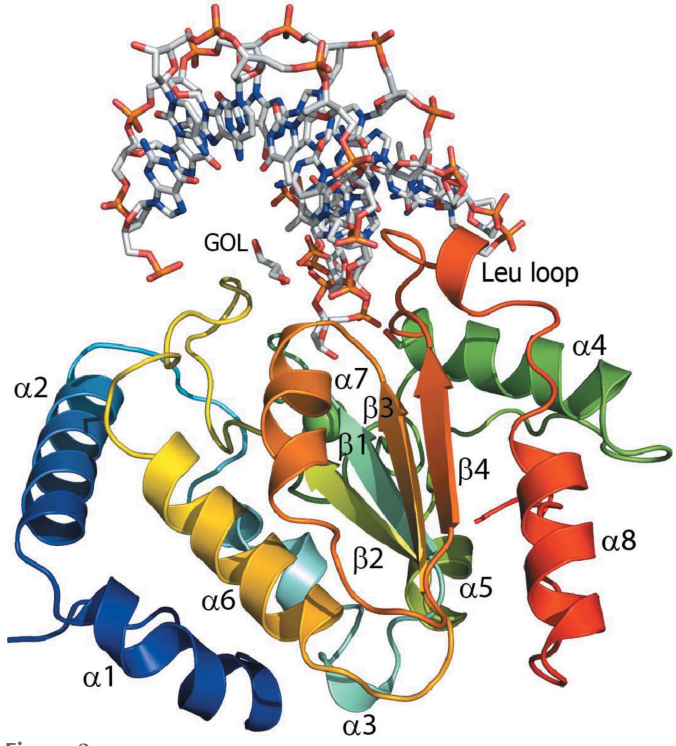


Figure 3
Overall structure of *Dr*UNG–DNA. The protein backbone is shown in rainbow colours ranging from blue (N-terminus) to red (C-terminus). The complexed DNA chains and the glycerol molecule are shown in sticks coloured by element and with the backbone in grey. The chloride anion in green can be seen behind the β-sheet.

strand. The chloride anion is coordinated by Phe96, Tyr85 and two nearby water molecules. Both ligands were part of the cryosolution that was used for data collection.

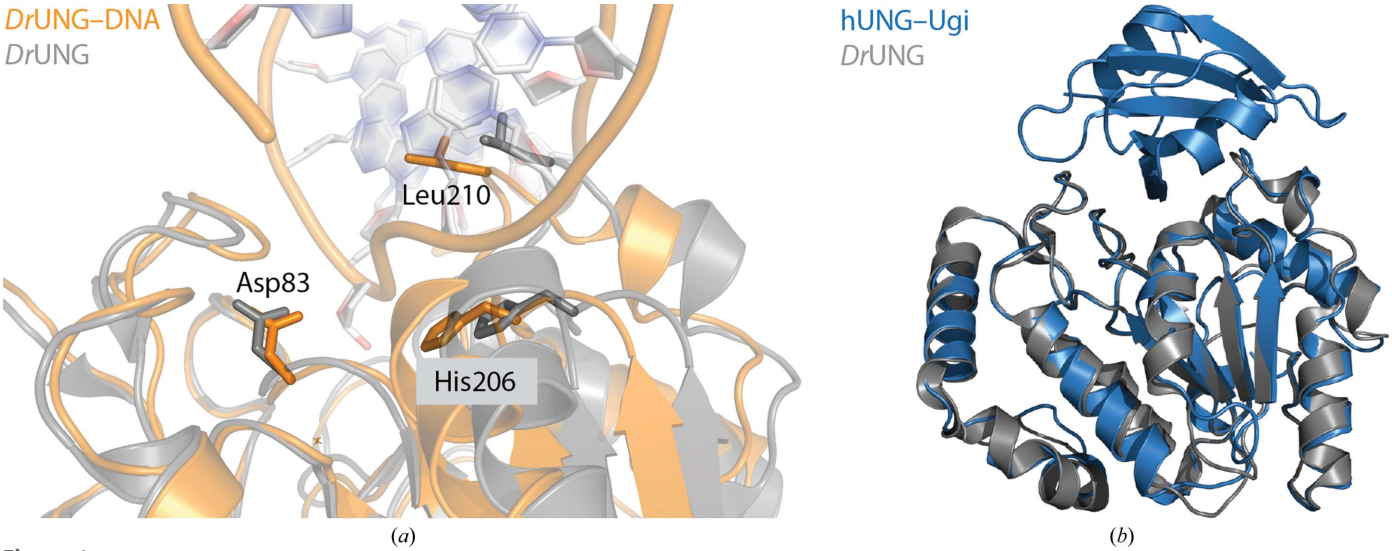
Overall, the structure of *Dr*UNG in complex with DNA is similar to the previously determined crystal structures of the catalytic domains of mutant and wild-type hUNG in complex with DNA (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998, 2000). The DNA binding introduces similar conformational changes in *Dr*UNG as observed for hUNG in the hUNG–DNA structure (Fig. 4*a*). The main chains holding the catalytic His206 and Asp83 (*Dr*UNG numbering) move 2.7 and 0.6 Å, respectively, and correspond well to the observed movement of the equivalent residues in hUNG (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998, 2000). The most severe alteration is observed in the leucine loop, where Leu210 moves 3.1 Å in order to fill the gap of the flipped-out base. The observed conformational change of *Dr*UNG in complex with DNA provides support for the suggestion that the DNA interaction leads to domain closure in UNG, which has been demonstrated by a comparative analysis of apo, DNA-bound and Ugi-bound structures of UNG (Saikrishnan *et al.*, 2002; Kaushal *et al.*, 2008). A superpositioning of the hUNG–Ugi complex (Mol, Arvai, Sanderson *et al.*, 1995) onto the apo structure of *Dr*UNG (Leiros *et al.*, 2005) (Fig. 4*b*) also demonstrates a high potential for specific interactions between *Dr*UNG and Ugi, and provides support for the previously observed inhibition of UNG activity in *D. radiodurans* extracts, suggesting that *Dr*UNG is the major UDG in this organism (Sandigursky *et al.*, 2004).

Compared with the apo *Dr*UNG structure, the *Dr*UNG in this study contains a single amino-acid substitution (Gly150Ala) caused by the erroneous incorporation of a cytosine (GCG instead of GGG) during the amplification of DNA for re-cloning of *Dr*UNG with the TEV cleavage site. Analysis of the structure suggests that the substitution will not affect the protein–DNA interaction. This residue is furthermore on the surface of the protein and does not affect the overall structure of the protein.

3.3. The absence of uracil in the complex structure

The uracil has been cleaved off the DNA and, unlike most of the previously determined crystal structures of hUNG in complex with DNA (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998, 2000), the uracil was not found in the active site of the enzyme. However, in one of the structures of hUNG in complex with DNA (PDB entry 2ssp) the uracil is also absent (Parikh *et al.*, 1998). This structure has an L272A substitution and was pre-incubated with DNA with an AU base pair as in the case of *Dr*UNG. Interestingly, in both the *Dr*UNG–DNA and 2ssp structures the deoxyribose is found in a β-anomer conformation rather than the α-anomer conformation as observed in the wild-type hUNG–DNA structures (PDB entries 1ssp and 4skn) (Fig. 5). The authors suggested that the observed β-conformation in PDB entry 2ssp had been generated by initial cleavage of the *N*-glycosidic bond by the L272A mutant followed by product release, which allowed the uracil to diffuse and the abasic deoxyribose to isomerize in solution. The L272A mutant then rebound the AP DNA upon pre-incubation (at room temperature) prior to co-crystallization.

They also explain that the α-anomer observed in the wild-type UDG–DNA structure is the result of a *trans* attack of a water molecule on the uracil–sugar C1′ atom, as seen in the Ab-site product-bound structure (PDB entry 1ssp). However, by looking at the structure it is clear that the deoxyribose group in wild-type UNG could not be in the β-conformation owing to a steric clash between the hydroxyl group and the uracil (Fig. 5). In both *Dr*UNG–DNA and 2ssp there is no



(*a*) (*b*)

Figure 4
Superposition of *Dr*UNG–DNA (protein shown in orange) with uncomplexed *Dr*UNG (protein shown in grey; PDB entry 2boo; Leiros *et al.*, 2005) and hUNG–Ugi (protein shown in blue; PDB entry 1ugh; Mol, Arvai, Sanderson *et al.*, 1995). (*a*) Illustration of overall conformational changes in the catalytic domain when *Dr*UNG is bound to DNA. The most distinct move is observed for Leu210, which moves 3.1 Å upon DNA binding, thereby filling the gap left by the flipped-out uracil base. The protein main chain holding the catalytically important residues His206 and Asp83 moves by 2.7 and 0.6 Å, respectively. (*b*) Superposition of hUNG–Ugi (blue) and uncomplexed *Dr*UNG (grey) shows that Ugi aligns very well with *Dr*UNG.

Table 2
Interactions between residues of *Dr*UNG (this study) and hUNG (PDB entry 1ssp) with DNA.

Domains	*Dr*UNG		Bond	Chain *B*	Chain *C*	hUNG		Bond	Chain *B*	Chain *C*
Water-activating loop	Gln82	N	Water-mediated H bond	Ab		Gln144	OE1 and NE2	Water-mediated H bond	Ab5P and A6P	
	Asp83	O	Water-mediated H bond	Ab		Asp145				
	His86	NE2	H bond	A8P		His148				
	Gln90		Water-mediated H bond	Ab9P		Gln152				
4-Pro loop	Ser107	N and OG	H bonds	Ab9P		Ser169				
GS loop	Ser185	N	H bond, nonbonded contact <3.35 Å	T11P		Ser247	N	H bond	A6P	
Leucine loop	His206	NE2	H bond	T11P		His268	N	H bond	A6P	
	Ser208	OG	Nonbonded contact <3.35 Å	G10P		Ser270		Nonbonded contact <3.35 Å	Ab5P	
	Pro209	O	Nonbonded contact <3.35 Å		T10	Pro271	O	Water-mediated H bond		A28
	Leu210	O	Water-mediated H bond	T11		Leu272	O	Water-mediated H bond	A6 and T7	A27
	Ser211	OG	Nonbonded contact <3.35 Å		G10	Ser273	OG	H bond	A6P	
	Glu212	OE2	Nonbonded contact <3.35 Å		T10					
	Gln213					Tyr275	OH	Water-mediated H bond		A27
	Tyr214	OH	Water-mediated H bond	T11		Arg276	NE	Water-mediated H bond	A6 and T7	
Surface		Long-range electrostatic interactions					Long-range electrostatic interactions			
	His86	+		Ab9		His148	+			
	Arg103	+			G12	Pro165	−			
	Arg109	+			G12	Glu171	−			
	Lys113	+			G11	Lys175	+			
	Arg188	+		C12		Gln250	−			
	Lys189	+			G3	Lys251	+			
	His206	+		Ab9		His268	+			

uracil; thus, it is possible for the hydroxyl group to remain in the β-conformation, which is further stabilized by ionic interactions (∼3 Å) with two water molecules in the active-site pocket.

The explanations for the absence of uracil in the *Dr*UNG–DNA structure could be related to that suggested for PDB entry 2ssp (Parikh *et al.*, 1998). The L272A mutant was shown to have a significant reduction in catalytic efficiency and substrate specificity compared with the wild-type enzyme, which could allow the slow association–catalysis–dissociation–re-association scenario during the 30 min pre-incubation of the L272A mutant with DNA at room temperature prior to crystallization. This could also be the case for the *Dr*UNG–DNA complex. Even though the enzyme has been shown to possess very high catalytic efficiency at 37°C (Leiros *et al.*, 2005), one would expect this efficiency to be considerably lower at 4°C, which was the temperature during the pre-incubation of this complex, and thus allow the same process to take place for the *Dr*UNG–DNA complex.

3.4. The interface and residue–DNA interactions

The interactions between *Dr*UNG and DNA were analysed by *NUCPLOT* (Luscombe *et al.*, 1997) and manual inspection of the structure, and revealed that the protein–DNA interaction was stabilized by a number of hydrogen bonds between some of the conserved amino acids in the water-activating, 4-Pro, GS and leucine loops and the DNA (Supplementary Figs. S3 and S4 and Table 2), as previously observed for hUNG–DNA (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998).

Firstly, it can be observed that four conserved serines (Ser107, Ser185, Ser208 and Ser211 in *Dr*UNG) form strong interactions with the sugar phosphates surrounding the abasic site (Ab9) of the DNA in the *Dr*UNG–DNA structure. Ser107 forms N- and OG-mediated hydrogen bonds to the Ab9 phosphate (Ab9P), and Ser185 makes a hydrogen bond and a direct nonbonded contact of <3.35 Å, respectively, to the T11 phosphate (T11P), while Ser211 makes a direct nonbonded contact of <3.35 Å to G10, and finally Ser208 forms two direct nonbonded contacts of <3.35 Å with the G10 phosphate (G10P) (Fig. 6*a*). In human UNG the equivalent serines (Ser169, Ser247, Ser270 and Ser273) have been suggested to compress the DNA backbone through these phosphate interactions, thereby facilitating the flipping of the damaged base into the active site of the enzyme (Slupphaug *et al.*, 1996), and in a comparison with the *NUCPLOT* representation of a hUNG–DNA structure (PDB entry 1ssp; DNA sequence 5′-CTGTUATCTT-3′ with the complementary strand

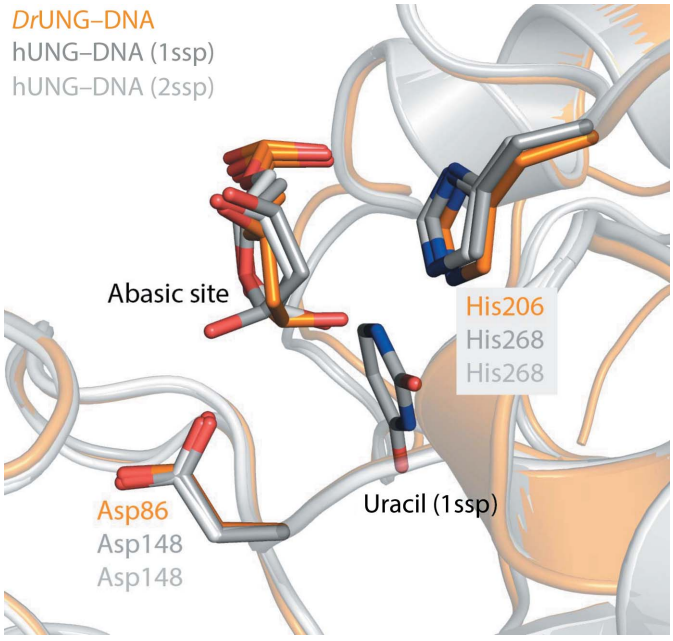


Figure 5
Illustration of the α- and β-conformations of the flipped-out abasic site in *Dr*UNG–DNA (orange), the hUNG–DNA structure with PDB code 1ssp (dark grey) and the hUNG–DNA structure with PDB code 2ssp (light grey). The deoxyribose is found in the β-conformation in *Dr*UNG and 2ssp, while it is in the α-conformation in 1ssp, which also has uracil in the active site. The absence of uracil in the active site in *Dr*UNG and 2ssp allows the β-conformation, which otherwise would be prevented owing to a steric clash with the hydroxyl group of the abasic site.

5′-AAGATAACAG-3′) it can be observed that the same interactions are conserved for Ser247, Ser270 and Ser273 on the 5′ side of the AP site, while no interaction is indicated for Ser169 (Table 2 and Supplementary Figs. S2 and S3). However, in Parikh *et al.* (1998) it is described that Ser169 is engaged in interactions with uracil and this explains why it does not appear. The observed interaction between the four serines in *Dr*UNG and the DNA backbone suggest that they are most likely to serve the same function as in hUNG; thus, this function is a conserved feature of UNGs.

Close to the water-activating loop, Gln90 makes one water-mediated hydrogen bond to the Ab9 phosphate. Furthermore, His86, which is close to the catalytic Asp83, makes a NE2-mediated hydrogen bond to the A8 phosphate on the 5′ side of the abasic site (A8P) (Fig. 6*a*). In hUNG the equivalent residue to His86 is His148, which is described as being engaged in hydrogen bonds to catalytic waters (Parikh *et al.*, 1998). Since there is no uracil in the *Dr*UNG–DNA structure this is not observed here; however, because the positions of these residues are structurally conserved they are most likely to serve the same function as in hUNG: stabilizing the uracil upon insertion of the base in the specificity pocket followed by participation in the cleavage of the *N*-glycosidic bond to generate the abasic site.

In the 4-Pro loop, the two first Pro residues are substituted by arginine (Arg103) and valine (Val104) in *Dr*UNG followed by two Pro residues. In this loop, Arg103 makes long-range electrostatic interactions with the sugar moiety of G12 on the

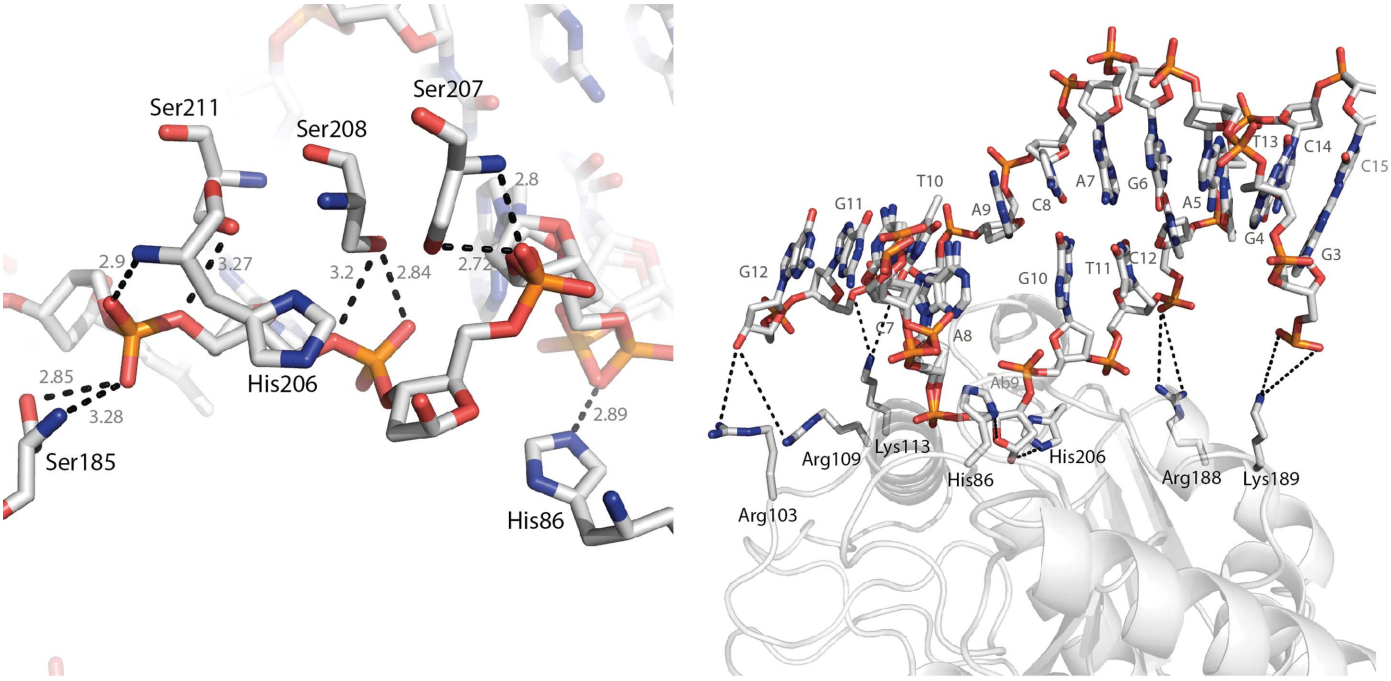


Figure 6
Short-range and long-range interactions between *Dr*UNG and DNA. (*a*) Short-range interactions (indicated for each interaction) are observed with Ser207, Ser185, Ser208, Ser211, His206 and His86 in *Dr*UNG. (*b*) Long-range interactions (<7 Å) are observed with His86, Arg103, Arg109, Lys113, Arg188, Lys189 and His206.

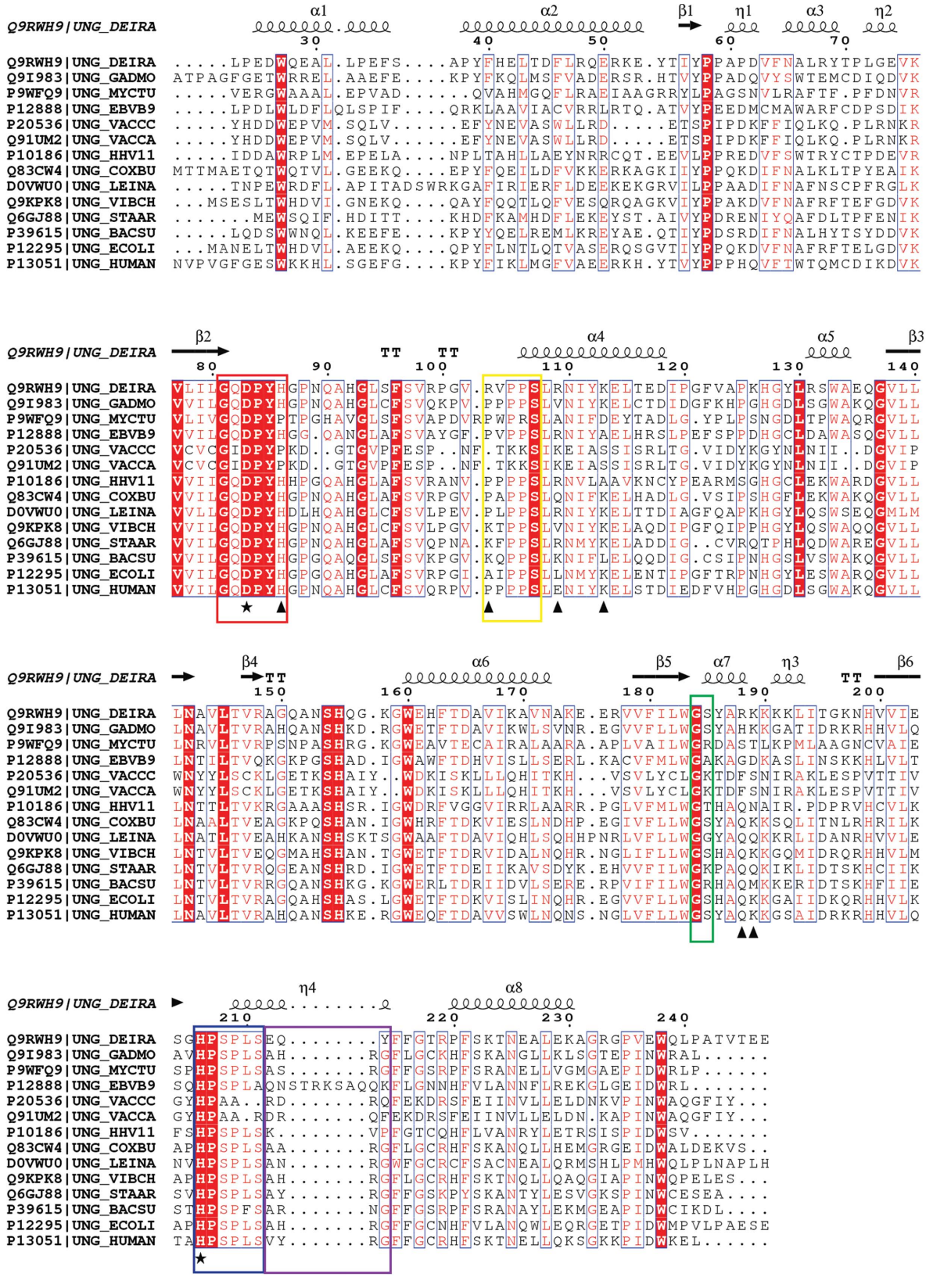


Figure 7
Structural sequence alignment of the catalytic domains of all structurally determined uracil-DNA glycosylases (generated by *ESPript* 3; Gouet *et al.*, 1999). The secondary structure of *Dr*UNG is shown at the top of the columns, and the *Dr*UNG sequence starts with residue 23. All of the conserved residues are shown in white on a red background, and residues with similar properties are shown in red on a white background. The catalytic Asp83 and His206 residues are indicated by black stars, and suggested positively charged long-range electrostatic interaction residues in *Dr*UNG are indicated by back triangles (His86, Arg103, Arg109, Lys113, Arg188 and Lys189). The conserved part of the leucine loop is shown in a dark blue box and the variable part in a purple box. The water-activating loop, the 4-Pro loop and the GS loop are indicated in red, yellow and green boxes, respectively.

complementary strand of the uracil-holding strand (Fig. 6*b*), in addition to the hydrogen bond to the phosphate on the 5′ side of the abasic site made by Ser185. A structural sequence alignment of the catalytic domain of structurally determined UNGs (Fig. 7) reveals that it is not only *Dr*UNG which has a substituted Pro residue in this position. Many UNGs have substituted both the first two Pro residues observed in hUNG, and in many cases, as in *Dr*UNG, the first Pro residue is substituted by a positively charged residue. *B. subtilis* (Baños-Sanz *et al.*, 2013), *V. cholerae* (Raeder *et al.*, 2010) and *S. aureus* (Wang *et al.*, 2014) UNGs possess a lysine, while *M. tuberculosis* (Kaushal *et al.*, 2010) UNG possesses an arginine in this position. There are no protein–DNA co-crystal structures of these enzymes; however, taking into consideration that they are in the same position as Arg103 and that the UNG structures are highly structurally conserved, it is likely that they interact with DNA in a similar way as in *Dr*UNG and thus help to stabilize the suggested DNA-damage interrogation complex prior to damage detection and removal.

The leucine loop has been suggested to be important for bringing His206 (*Dr*UNG numbering) into contact with uracil in DNA and to be involved in both catalysis and stabilizing the DNA upon flipping the uracil into the active site (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998). To date, this loop and its conserved residues have been described to cover a range of nine residues from the catalytic His268 to Arg276 in hUNG (268-HPSPLSVYR-276). However, a sequence comparison of the catalytic domains of all of the structurally determined UNGs to date (Fig. 7) shows that there is indeed a difference in the length of this loop: the nine-residue loop is well conserved among bacterial and mammalian UNGs, while the viral UNGs and *Dr*UNG have both extended and shortened loops. In particular, the Epstein–Barr virus UNG has a long loop, which despite its length has been suggested to serve the same function as in other UNGs (Géoui *et al.*, 2007). Based on our sequence comparison it seems that the leucine-loop region which is more conserved is HPSPLS (206-HPLPLS-211 in *Dr*UNG), while the following residues (212-EQY-214) in *Dr*UNG comprise a variable part of the loop. Thus, we will refer to these two parts of the loop as the conserved and the variable parts of the leucine loop.

In the conserved part of the leucine loop the catalytic His206 forms a hydrogen bond to the T11 phosphate on the 3′ side of the abasic site (Fig. 6*a*), as in hUNG where His268 makes a hydrogen bond to a phosphate on the same side of the abasic site. In hUNG, both Pro271 (Pro209 in *Dr*UNG) and Leu272 (Leu210 in *Dr*UNG) also make water-mediated hydrogen bonds to the DNA phosphate backbone, both to the complementary strand (both residues) and the damage-containing strand (Leu272). In *Dr*UNG there seems to be no direct interaction between these residues and DNA; however, Leu210 is intercalated between the two strands in the same position as Leu272 in hUNG and seems to be able to fulfil its task to stabilize the DNA upon flipping of the damaged base into the active site.

In hUNG the variable part of the leucine loop interacts with DNA through Tyr275 and Arg276, which form a water-mediated hydrogen bond to A27 on the complementary strand of the damaged DNA (Tyr275) and water-mediated hydrogen bonds to A6 and T7 on the 5′ side of the abasic site. These residues have previously been suggested to act as reading heads for the detection of uracil in DNA (Parikh *et al.*, 1998), and mutational studies of UNG from Atlantic cod (cUNG) showed that the substitution of His275 in cUNG by Tyr275 reduced the catalytic efficiency of the protein severely, supporting the reading-head role of residue 275 (Moe *et al.*, 2004). From Fig. 8, it can be observed that both Tyr275 and Arg276 in hUNG seem to go 'deeper' into the minor groove than Gln213 and Tyr214 in *Dr*UNG, which corresponds well to the observation that these residues form strong interactions with the DNA. However, despite the lack of charge, both Gln213 and Tyr214 seem to be able to widen the minor groove and 'push' Leu272 into the DNA in order to facilitate uracil flipping, which is indicated by the conformational change observed for both the residues in this region and the overall structure (Fig. 4).

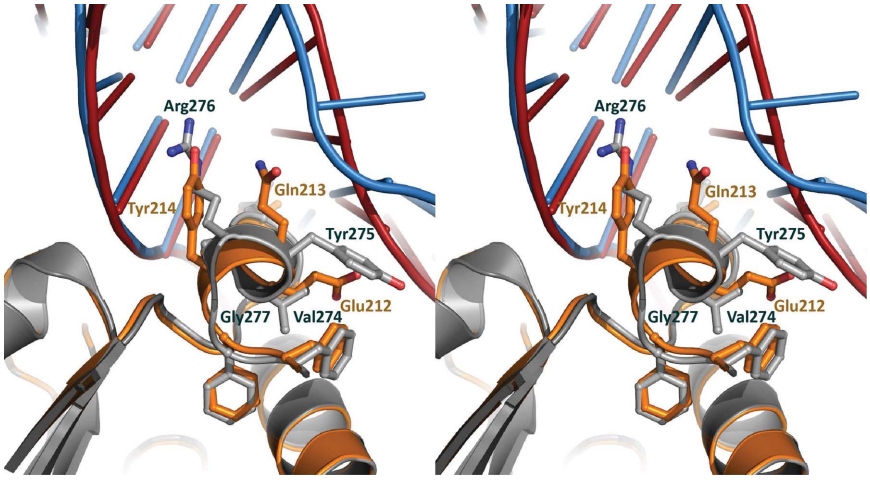


Figure 8
Stereoview comparing the conformations of amino-acid residues in the leucine-loop regions of *Dr*UNG–DNA (protein in gold, DNA in red) and hUNG–DNA (protein in grey, DNA in blue; PDB entry 1emh). The residues in the leucine loop of *Dr*UNG are substituted by 212-EQY-214, with one deletion and no sequence similarity compared with hUNG (274-VYRG-277).

3.5. Surface charge and DNA dynamics

The ability of UNGs to form strong interactions with DNA is determined to a high degree by charge and shape complementarity (Slupphaug *et al.*, 1996). We have previously observed that *Dr*UNG possesses a much higher catalytic efficiency for the removal of uracil in DNA compared with hUNG, and have assigned this property to a very high substrate affinity (K_m) of *Dr*UNG caused by an increased number of positive residues on the surface of the protein (Leiros *et al.*, 2005). An analysis of long-range electrostatic interactions between

positively charged residues on the surface of *Dr*UNG and DNA in this work strongly supports this hypothesis (Fig. 6*b* and Table 2). Compared with hUNG, we observe that the number of interactions is elevated and a total of seven positively charged residues in *Dr*UNG (His86, Arg103, Arg109, Lys113, Arg188, Lys189 and His206) make long-range ionic interactions with DNA, while the corresponding number in hUNG is five (His148, Lys175, Lys251, His268 and Arg276). In particular, the interaction with Arg109 and Arg188 may be of importance. In hUNG the equivalent residues are Glu171 and Gln250, and in a mutational study of cod and human UNG we have shown that a V171E and H250Q mutant of cod UNG and an E171V mutant of hUNG affected both the substrate affinity and the catalytic turnover of the enzyme (Moe *et al.*, 2004). Thus, the locations of these residues seem to be important for substrate interaction and catalytic efficiency of the UNG enzymes. This provides support for our hypothesis that long-range positive electrostatic interactions contribute to the high catalytic efficiency of *Dr*UNG.

Our analysis of the surface charge of *Dr*UNG compared with hUNG may also provide an explanation for why *Dr*UNG did not seem to form stable complexes with the shorter oligonucleotides in our EMSA analysis (Fig. 2). Based on the determined structure of *Dr*UNG–DNA in this work, we have analysed which parts of the oligonucleotides should be able to form stable contacts with the DNA (indicated in Fig. 1). This analysis reveals that the shorter oligonucleotides should also be able to form stable complexes with *Dr*UNG, especially those which have an identical sequence to the co-crystallized oligonucleotide. However, on a closer inspection of the long-range electrostatic contacts which are made between the enzyme and dsDNA it is observed that many of them are formed to the complementary strand of the damage-containing DNA. In Fig. 6(*b*), we show that Lys189 forms long-range electrostatic contacts (<7 Å) with the phosphate group of G3 on the 5′ side of the complementary strand of the DNA, which is beyond the boundary of the potential contact points for the 12 and 10 nt dsDNA oligonucleotides. Thus, we suggest that this interaction, in addition to a generally high positively charged electrostatic surface charge of *Dr*UNG (Leiros *et al.*, 2005), provides an explanation for this strong preference for longer oligonucleotides in our EMSA experiments.

In a recent paper, a combination of H/D-exchange mass (HDMX) spectroscopy and computational docking was used to analyse the interaction between UNG and long DNA fragments (30 bp; Roberts *et al.*, 2012). The results showed that the fragments increased the solvent protection of UNG at the active site, but also identified residues 210–220 and 251–264 (hUNG numbering) as potential DNA-interacting sites. Residues 210–220 are placed in the loop between β4 and α6 parallel to the active site and are equivalent to residues 184–194 in *Dr*UNG. The 251–264 region consists of residues on β5 and α7 and is equivalent to residues 189–202 in *Dr*UNG (Fig. 7). The authors suggest that the observed solvent protection in other regions close to the active site is caused by strand separation of the dsDNA during damage recognition and/or repair and substrate binding after base removal. It is not possible to demonstrate these dynamic movements in protein crystals owing to crystal packing. In our work, we have determined the crystal structure of UNG in complex with a longer oligonucleotide (16 bp) than previously used (10 bp; Slupphaug *et al.*, 1996; Parikh *et al.*, 1998, 2000). Even though we do not observe any dramatic dynamic movements close to the active site, the disorder (double conformation) that we observe for, for example, A8 and C7 on the 5′ side of the abasic site (Supplementary Fig. S1*b*) and the increasing disorder at the terminal ends for our longer DNA construct could provide some support for this hypothesis.

In the initial HDMX experiments the authors also observed some solvent protection close to the N-terminus of UNG (Roberts *et al.*, 2012). Curiously, in our structure we observe specific interactions between the DNA and amino acids in the N-terminus of a symmetry-related molecule (Fig. 9). Asn22, Arg48 and Arg68 generate hydrogen bonds to the phosphates of G6, A7 and C8 on the complementary strand to the damage-containing strand, respectively. These interactions seem to be further stabilized by two prolines (59 and 61) in the symmetry-related molecule which are positioned in the minor groove oppositely orientated to the abasic site. This interaction could be the result of crystal packing, but seems to be very specific, and taking the HDMX results into consideration it is tempting to speculate that it might be relevant for stabilization of the UNG–DNA complex before, during or after

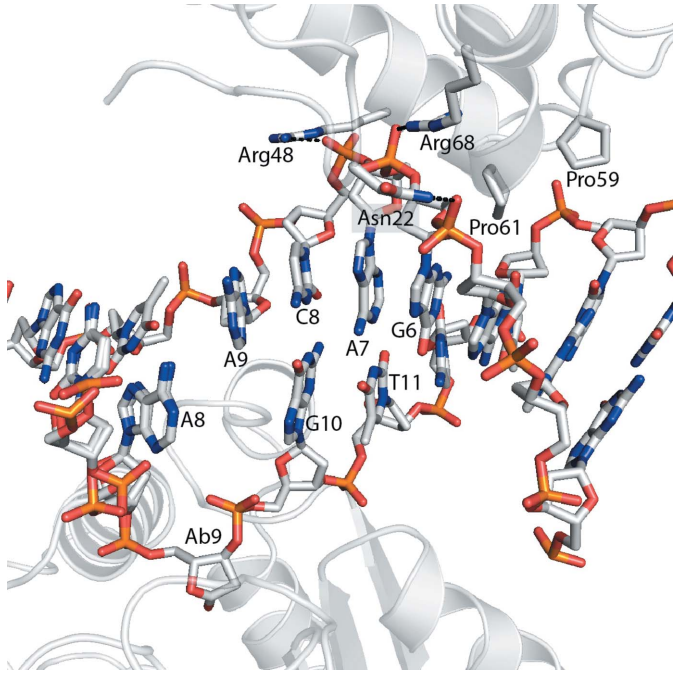


Figure 9
Specific interactions between DNA and a symmetry-related molecule. Interactions were observed between Asn22 and the G6 phosphate (2.73 Å), Arg68 and the A7 phosphate (3.6 Å), and Arg48 and the C8 phosphate (3.2 Å) on the DNA strand complementary to the damage-containing strand. The interactions are further stabilized by the presence of two prolines (59 and 61) from the symmetry-related molecule which are placed in the minor groove.

base removal; however, this clearly needs further investigation.

4. Concluding remarks

In this work, we have presented the first high-resolution crystal structure (1.35 Å) of a bacterial uracil-DNA *N*-glycosylase (UNG) from the extremophile *D. radiodurans* (*Dr*UNG) in complex with DNA. In the determined structure, *Dr*UNG is found in a complex with its product (DNA with an abasic site) without the uracil, which has probably diffused out from the structure during the low-temperature pre-incubation of the protein–DNA mixture prior to crystallization. Compared with the apo *Dr*UNG structure (Leiros *et al.*, 2005), the overall structure of *Dr*UNG in complex with DNA appears in a closed conformation as observed for hUNG (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998, 2000; Fig. 4), and central amino acids for catalysis have moved as much as 2–3 Å from their original positions in the apo structure. A structural sequence alignment of crystallized catalytic domains of UNGs (Fig. 7) shows that the so-called leucine loop is of variable length across different kingdoms of life. Based on this alignment, we suggest splitting the definition of the so-called leucine loop into two: a conserved and a variable part. The conserved part consists of the catalytic histidine, two conserved prolines, the DNA-plugging leucine and the two conserved DNA-pinching serines (HPSPLS), while the residues following the second serine until the highly conserved phenylalanine are designated the variable part.

An analysis of the interaction between the DNA and the residues in the DNA-interacting loops of *Dr*UNG (Fig. 6, Table 2) shows that despite several amino-acid substitutions in important positions, the loop interactions are conserved to a high degree and are most likely to serve the same functions as observed for hUNG in complex with DNA (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998). We also suggest that even though the variable part of the leucine loop in *Dr*UNG is shorter than in many other bacterial UNGs, the loop is able to provide stable interactions with the DNA minor groove upon flipping of the uracil into the specificity pocket of the enzyme. We believe that the properties shown here by the substitution in the 4-Pro loop and the variable part of the leucine loop in *Dr*UNG support a model in which DNA compression and minor-groove stabilization are important parts of the DNA damage-detection mechanisms of uracil-DNA *N*-glycosylases.

The *Dr*UNG–DNA crystal structure also provides a rationale for the previously observed high catalytic efficiency of *Dr*UNG (Leiros *et al.*, 2005). An increased number of long-range electrostatic surface interactions between positively charged residues close to the DNA-binding site and the DNA explains the high substrate affinity resulting in high catalytic efficiency of *Dr*UNG.

Acknowledgements

We thank the FRIBIO and the National Functional Genomics Programme (FUGE) within the Research Council of Norway and the University of Tromsø for funding this project. Provision of beam time at the European Synchrotron Radiation Facility (ESRF) beamline ID29, Grenoble, France is gratefully acknowledged.

References

Anderson, A., Nordan, H., Cain, R., Parrish, G. & Duggan, D. (1956). *Food Technol.* **10**, 575–578.
Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V. & Notredame, C. (2006). *Nucleic Acids Res.* **34**, W604–W608.
Assefa, N. G., Niiranen, L., Johnson, K. A., Leiros, H.-K. S., Smalås, A. O., Willassen, N. P. & Moe, E. (2014). *Acta Cryst.* D**70**, 2093–2100.
Baños-Sanz, J. I., Mojardín, L., Sanz-Aparicio, J., Lázaro, J. M., Villar, L., Serrano-Heras, G., González, B. & Salas, M. (2013). *Nucleic Acids Res.* **41**, 6761–6773.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Franklin, M. C., Cheung, J., Rudolph, M. J., Burshteyn, F., Cassidy, M., Gary, E., Hillerich, B., Yao, Z.-K., Carlier, P. R., Totrov, M. & Love, J. D. (2015). *Proteins*, doi:10.1002/prot.24841.
Géoui, T., Buisson, M., Tarbouriech, N. & Burmeister, W. P. (2007). *J. Mol. Biol.* **366**, 117–131.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Jiang, Y. L., Kwon, K. & Stivers, J. T. (2001). *J. Biol. Chem.* **276**, 42347–42354.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kaushal, P. S., Talawar, R. K., Krishna, P. D. V., Varshney, U. & Vijayan, M. (2008). *Acta Cryst.* D**64**, 551–560.
Kaushal, P. S., Talawar, R. K., Varshney, U. & Vijayan, M. (2010). *Acta Cryst.* F**66**, 887–892.
Krokan, H. E. & Bjørås, M. (2013). *Cold Spring Harb. Perspect. Biol.* **5**, a012583.
Leiros, I., Moe, E., Lanes, O., Smalås, A. O. & Willassen, N. P. (2003). *Acta Cryst.* D**59**, 1357–1365.
Leiros, I., Moe, E., Smalås, A. O. & McSweeney, S. (2005). *Acta Cryst.* D**61**, 1049–1056.
Levin-Zaidman, S., Englander, J., Shimoni, E., Sharma, A. K., Minton, K. W. & Minsky, A. (2003). *Science*, **299**, 254–256.
Lindahl, T. & Nyberg, B. (1974). *Biochemistry*, **13**, 3405–3410.
Liu, Y., Zhou, J., Omelchenko, M. V., Beliaev, A. S., Venkateswaran, A., Stair, J., Wu, L., Thompson, D. K., Xu, D., Rogozin, I. B., Gaidamakova, E. K., Zhai, M., Makarova, K. S., Koonin, E. V. & Daly, M. J. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 4191–4196.
Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. (1997). *Nucleic Acids Res.* **25**, 4940–4945.
Makarova, K. S., Aravind, L., Wolf, Y. I., Tatusov, R. L., Minton, K. W., Koonin, E. V. & Daly, M. J. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 44–79.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 499–501.
Mattimore, V. & Battista, J. R. (1996). *J. Bacteriol.* **178**, 633–637.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Moe, E., Leiros, I., Riise, E. K., Olufsen, M., Lanes, O., Smalås, A. & Willassen, N. P. (2004). *J. Mol. Biol.* **343**, 1221–1230.
Moe, E., Leiros, I., Smalås, A. O. & McSweeney, S. (2006). *J. Biol. Chem.* **281**, 569–577.
Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., Mosbaugh, D. W. & Tainer, J. A. (1995). *Cell*, **82**, 701–708.

Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E. & Tainer, J. A. (1995). *Cell*, **80**, 869–878.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E. & Tainer, J. A. (1998). *EMBO J.* **17**, 5214–5226.
Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M. & Tainer, J. A. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 5083–5088.
Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). *Acta Cryst.* D**59**, 1131–1137.
Putnam, C. D., Shroyer, M. J., Lundquist, A. J., Mol, C. D., Arvai, A. S., Mosbaugh, D. W. & Tainer, J. A. (1999). *J. Mol. Biol.* **287**, 331–346.
Raeder, I. L. U., Moe, E., Willassen, N. P., Smalås, A. O. & Leiros, I. (2010). *Acta Cryst.* F**66**, 130–136.
Ravishankar, R., Bidya Sagar, M., Roy, S., Purnapatre, K., Handa, P., Varshney, U. & Vijayan, M. (1998). *Nucleic Acids Res.* **26**, 4880–4887.
Roberts, V. A., Pique, M. E., Hsu, S., Li, S., Slupphaug, G., Rambo, R. P., Jamison, J. W., Liu, T., Lee, J. H., Tainer, J. A., Ten Eyck, L. F. & Woods, V. L. Jr (2012). *Nucleic Acids Res.* **40**, 6070–6081.
Saikrishnan, K., Bidya Sagar, M., Ravishankar, R., Roy, S., Purnapatre, K., Handa, P., Varshney, U. & Vijayan, M. (2002). *Acta Cryst.* D**58**, 1269–1276.
Sandigursky, M., Sandigursky, S., Sonati, P., Daly, M. J. & Franklin, W. A. (2004). *DNA Repair (Amst.)*, **3**, 163–169.
Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995). *Nature (London)*, **373**, 487–493.
Savva, R. & Pearl, L. H. (1995). *Nature Struct. Mol. Biol.* **2**, 752–757.
Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E. & Tainer, J. A. (1996). *Nature (London)*, **384**, 87–92.
Stivers, J. T. (2004). *Prog. Nucleic Acid Res. Mol. Biol.* **77**, 37–65.
Tye, B. K., Nyman, P. O., Lehman, I. R., Hochhauser, S. & Weiss, B. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 154–157.
Wang, H.-C., Hsu, K.-C., Yang, J.-M., Wu, M.-L., Ko, T.-P., Lin, S.-R. & Wang, A. H.-J. (2014). *Nucleic Acids Res.* **42**, 1354–1364.
White, O. *et al.* (1999). *Science*, **286**, 1571–1577.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Zharkov, D. O., Mechetin, G. V. & Nevinsky, G. A. (2010). *Mutat. Res.* **685**, 11–20.