Structure and function of the UVDE repair protein

Paspaleva, K.

Crystal structure of the DNA repair enzyme ultraviolet damage endonuclease

Chapter 2

Keti Paspaleva[1], Ellen Thomassen[2], Navraj S. Pannu[2], Shigenori Iwai[3], Geri F. Moolenaar[1], Nora Goosen[1] and Jan Pieter Abrahams[2]

[1]Laboratory of Molecular Genetics, Leiden Institute of Chemistry, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands

[2]Department of Biophysical Structural Chemistry, Leiden Institute of Chemistry, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands

[3]Division of Chemistry, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

SUMMARY

The ultraviolet damage endonuclease (UVDE) performs the initial step in an alternative excision repair pathway of UV-induced DNA damage, nicking immediately adjacent to the 5' phosphate of the damaged nucleotides. Unique for a single-protein DNA repair endonuclease, it can detect different types of damage. Here we show that *Thermus thermophilus* UVDE shares some essential structural features with Endo IV, an enzyme from the base excision repair pathway that exclusively nicks at abasic sites. A comparison between the structures indicates how DNA is bound by UVDE, how UVDE may recognize damage, and which of its residues are involved in catalysis. Furthermore, the comparison suggests an elegant explanation of UVDE's potential to recognize different types of damage. Incision assays including point mutants of UVDE confirmed the relevance of these conclusions.

INTRODUCTION

Guarding genetic integrity by repairing damaged DNA is one of the most fundamental processes of life. To meet this challenge, life has evolved a range of DNA repair pathways, including nucleotide excision repair (NER), base excision repair (BER), and the alternative ultraviolet damage endonuclease (UVDE) repair. NER recognizes many types of damage for which it requires a large, multiprotein complex (reviewed in Truglio *et al*., 2006). BER employs a collection of glycosylases, each recognizing and removing a different set of modified DNA bases (reviewed in Friedberg *et al*., 2005). The UVDE pathway distinguishes itself by recognizing and subsequently nicking DNA containing different types of damage with the single, multifunctional UVDE enzyme (Takao *et al*., 1996). Although functionally characterized, there are no structural data of the UVDE enzyme.

There are some similarities between the BER and UVDE pathways. In the first step of BER, a glycosylase removes the damaged base, leaving just the deoxyribose. In the next step, the DNA 5' adjacent to this abasic site is nicked by an apurinic/apyrimidinic endonuclease. Subsequently, other enzymes remove the nicked residue and resynthesize the damaged strand (Friedberg *et al*., 2005). In UVDE repair the first step is skipped, and the UVDE enzyme immediately recognizes and nicks the damaged DNA strand. Even though the UVDE enzyme is much more versatile in damage recognition than the BER endonuclease, it was proposed to have a similar topology (Aravind *et al*., 1999). The structure of the Endo IV endonuclease of the BER pathway and its complex with abasic, nicked DNA are known. (Hosfield *et al*., 1999).

Although the UVDE DNA repair pathway was originally thought to be specific for UV damage, enzymatic studies revealed that UVDE from *Schizosaccharomyces pombe* possesses a broader substrate specificity including pyrimidine dimers (CPD), 6-4 photoproducts (6-4PP), apurinic/apyrimidinic (AP) sites, uracil (U), dihydrouracil (DHU), and other non-UV-induced DNA adducts (Avery *et al*., 1999; Kanno *et al*., 1999). Biochemical and genetic analysis also suggest that UVDE may be involved in orchestrating mismatch repair in vivo (Kaur *et al*., 1999).

The broad substrate specificity of *S. pombe* UVDE was recently reviewed by Paul W. Doetsch (Doetsch *et al*., 2006), showing that this enzyme is also active on insertion-deletion loops, supporting the role of UVDE in mismatch repair. In conclusion, the broad substrate specificity of UVDE indicates that it recognizes a common distortion in the DNA helix rather than the chemical structure of the DNA damage.

The UVDE pathway was described for the first time in the fission yeast *S. pombe*. Homologues of UVDE are present in many fungal species but also in a number of bacteria, such as *Bacillus subtilis* and the thermophilic bacterium *Thermus thermophilus*. Overexpressed, purified full-length UVDE protein from *S. pombe* (68 kDa) is unstable (Kaur *et al*., 1998), and a number of biochemical studies have been done using a truncated protein lacking its N-terminal 228 amino acids. The bacterial homologues of UVDE lack this N-terminal region and also the

highly charged C-terminal domain of their eukaryotic counterparts. We determined the structure of UVDE from *T. thermophilus* to answer some of the main questions in the field: what is the catalytic mechanism of UVDE and how does it recognize so many different types of damage?

EXPERIMENTAL PROCEDURES

Cloning, expression and purification of the UVDE homolog from *T. thermophilus*

The full-length *T. thermophilus* UVDE-coding region was initially amplified from the chromosomal DNA by PRC with primers 5'-GCTTCTCATATGATCCGCCTGGGCTACCCC -3' and 5'-TCGTCTCTGCAGTCAAGGGGTTGCTAGGCCCTGCTC-3' for the 5' and the 3' ends of the fragment, respectively. Primers were designed using the genome sequence of *T. thermophilus* (Henne *et al*., 2004). The UVDE-coding fragment was cloned into the *Nde*I and *Pst*I restriction sites of pETUVDEΔ228, a plasmid previously made in our lab to overproduce a truncated *S. pombe* UVDE protein (unpublished data). Cloning into this pET plasmid allowed for a T7 promoter-driven expression of the UVDE protein with a fusion of ten histidines. The ten His residues are attached to the N-terminal part of the protein by a short, nine amino acid linker containing a factor Xa cleavable site. The *T. thermophilus* UVDE expression vector (pUD24) was then transformed into *E. coli* BL21 (Studier *et al*., 1990). The UVDE protein was purified from cells of a 2 l culture, harvested 2 h after induction by IPTG, and lysed by sonication in 6 ml lysis buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 mM β-mercaptoethanol, 10 % glycerol, 1% Triton X-100). The lysate was separated into soluble and insoluble fractions by centrifugation at 37, 000 rpm for 30 min. The supernatant was loaded on a HiTrap-chelating column, which was equilibrated with buffer A (20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 10 % glycerol, 10 mM β-mercaptoethanol) containing 20 mM imidazole and the protein was eluted with a 20 – 250 mM gradient of imidazole in buffer A. Pooled fractions of the *T. thermophilus* UVDE were loaded on a Resource Q column equilibrated with 20 mM Tris (pH 7.5), 10 % glycerol. The protein was eluted with a 0 – 1 M NaCl gradient in 20 mM Tris-HCl (pH 7.5). Finally, the UVDE-containing fractions were loaded on a Nap5 gel-filtration column (Amersham), and equilibrated in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl. The protein fractions, which showed high purity were used in *in vitro* assays and crystallization trials.

For crystallization purposes, a selenomethionine (SeMet)-substituted UVDE protein was expressed in a methionine auxotrophic derivative of BL21 (Sohi *et al*., 2000). The SeMet-labelled UVDE was purified by the same procedure as the non-labelled protein and showed identical enzymatic activity as the non-labelled protein. Mutants E176A and Y105A containing alanine substitutions were constructed by site-directed mutagenesis using PCR and purified like the wild-type UVDE protein. All point mutations in the *T. thermophilus* UVDE resulted in proteins showing the same elution/purification profiles as the wild-type enzymes on ion-exchange and size-exclusion columns, suggesting they were properly folded.

Cloning, expression, and purification of the UVDE homolog from *S. pombe*

In this study, we used an N-terminal truncation of the full-length *S. pombe* UVDE protein. The first 223 amino acids were removed as described before (Kaur *et al*., 1998). All UVDE mutants were constructed by PCR and verified by sequencing for the absence of additional PCR-induced mutations. The wild-type Δ228-UVDE and mutant proteins were further purified on a HiTrap-chelating column, and hydroxyapatite and P11 phosphate cellulose columns. Detailed description of Δ228-UVDE protein purification will be given elsewhere (unpublished data). All *S. pombe* UVDE mutants showed the same elution/purification profiles as the wild-type enzymes.

DNA substrates

The DNA substrates used in this study are 30 bp substrates containing either a CPD or a 6-4PP adduct in the sequence 5'-CTCGTCAGCATC*TT*CATCATACAGTCAGT-3', the *TT* representing the position of the UV lesion. The oligonucleotides containing CPD or 6-4PP lesions were synthesized as described (Iwai, 2006).

Incision assay

The DNA substrates were labelled at the 5' side of the top strand using polynucleotide kinase as described (Verhoeven *et al*., 2002). The DNA substrates (0.2 nM) were incubated with 5 nM UVDE in 20 µl reaction mix (20 mM HEPES (pH 6.5), 100 mM NaCl, 1 mM $MnCl_2$). After 15 min incubation at 30°C for *S. pombe* UVDE or 55°C for the *T. thermophilus* protein, the reactions were terminated by adding 3 µl EDTA/SDS (0.33 M EDTA, 3.3% SDS) and 2.4 µl glycogen (4 µg/µl) followed by ethanol precipitation. The incision products were visualized on a 15 % denaturing polyacrylamide gel.

Crystallization

The purified protein in 20 mM Tris (pH 7.5), 150 mM NaCl, 10 % glycerol was dialyzed against 1× PBS (137 mM NaCl, 10 mM phosphate, 2.7 mM KCl (pH 7.4)) and concentrated to 3-5 mg/ml by centrifugation using an Ultrafree filter device (Millipore). The protein was stored at 4°C till further use. After 1 week, plate-like crystals appeared in the Eppendorf tube and grew to a size of 0.1 × 0.1 × 0.01 mm. Crystals were transferred to 25 % (v/v) glycerol in crystallization buffer prior to data collection.

Crystallographic data collection and processing

A MAD data set was collected on beamline BM14 at the European Synchrotron Radiation Facility (ESRF) at a wavelength of 0.97800 Å (peak), 0.97850 Å (inflection point), and 0.91840 Å (high-energy remote) using the anomalous signal of the selenium atoms. The crystal was flash-frozen and kept at 100K during data collection and 360 images were collected with a rotation angle of 1°. Reflections were integrated with MOSFLM (Leslie, 1999) and merged with SCALA (Evans, 1993) from the CCP4 suite (CCP4, 1994). For data statistics, see Table 1.

Structure solution and refinement

A strong nonorigin peak in the Patterson map indicated the presence of noncrystallographic translational symmetry. Fa values were calculated by AFRO (http://www.bfsc.leidenuniv.nl/software/crank/) and passed into CRUNCH2 (De Graaff *et al*., 2001), which found six selenium sites. These sites confirmed the translational symmetry seen in the Patterson map. The positions, occupancies, and temperature factors of the sites were refined and phased using BP3 (Pannu *et al*., 2003). Solvent flattening including the noncrystallographic operator was performed in dm (Cowtan, 1994). For autobuilding and iterative refinement of the model, ARP/wARP (Perrakis *et al*., 1999) and REFMAC (Murshudov *et al*., 1999), with the maximum likelihood function incorporating Hendrickson-Lattman coefficients (Pannu *et al*., 1998) from BP3, were used. ARP/wARP was able to build a model containing 495 residues. An anomalous difference Fourier map found six additional anomalous scatterers related by the translational noncrystallographic symmetry. Manual rebuilding of the model and addition of the unbuilt residues were done with Coot (Emsley and Cowtan, 2004). Refinement to 1.55 Å resolution was done using REFMAC including noncrystallographic symmetry restraints for residues 1 - 277, and water molecules were added using ARP/wARP and Coot. Illustrations were prepared using PyMOL (Figures 1, 2, 5B, 5D, and 6; DeLano, 2002), and an Accelrys DS visualizer (Figure 5C; http://www.accelrys.com/).

RESULTS AND DISCUSSION

Structure determination

The structure of UVDE was determined to a resolution of 1.55 Å by three-wavelength multiple anomalous dispersion (MAD) using the anomalous signal from selenium atoms. The protein crystallized in space group P1, with unit cell dimensions of 47.34 × 48.70 × 68.76 Å and angles of $\alpha = 106.1°$, $\beta = 94.4°$, and $\gamma = 114.2°$. The crystals contained two molecules in the asymmetric unit, had a solvent content of 38.2 %, and a Wilson temperature factor of 17.3 $Å^2$. The resulting map showed good electron density for the residues –2 to 277 (monomer A) and 1 to 277 (monomer B). The His tag and most of the linker residues and the C-terminal residues 278 – 280 were disordered. There are only small differences between monomers A and B, which are related by a noncrystallographic translational symmetry vector of ~ 0 ½ ½. The two molecules in the asymmetric unit are very similar: the root-mean-square deviation (rmsd) between the atomic positions of corresponding main-chain atoms of the two monomers is 0.23 Å. The final model has good stereochemistry and R factors (Table 1) and it contains 4433 protein atoms, 6 anomalously scattering metal ions, 2 phosphate ions, and 356 water molecules.

Table 1. Crystallographic data and refinement statistics.

	Peak set	Inflection point	High remote
A. *Data collection*			
Beam-line	ESRF BM14	ESRF BM14	ESRF BM14
Wavelength (Å)	0.97800	0.97850	0.91840
Detector	MAR225 CCD	MAR225 CCD	MAR225 CCD
Resolution range (Å)	64.6-1.6 (1.69-1.60)[a]	23.6-1.55 (1.63-1.55)	
		24.0-1.50 (1.58-1.50)	
Multiplicity	3.9 (3.7)	3.9 (3.8)	3.9 (3.8)
Completeness (%)	67.9 (15.8)[b]	73.8 (23.6)	78.9 (31.6)
Rsym[c] (%)	4.0 (33.1)	3.9 (31.5)	4.4 (34.9)
B. *Phasing*			
Number of Se-sites		6	
FOM Overall		0.433	
FOM 1.73 – 1.55 Å		0.178	
C. *Refinement*			
Resolution range (Å)		20.0-1.55 (1.63-1.55)	
No. of reflections used in refinement	56303 (1053)		
No. of reflections used for R-free	2860 (66)		
R-factor[d]		0.18 (0.25)	
R-free		0.21 (0.35[g])	
No. of protein / water atoms	4433 / 356		
Average B-value	protein / solvent (Å^2)	18.6 / 28.4	
	phosphate / metal ions	40.0 / 18.5, 68.7, 74.4	
Ramachandran statistics[e] (%)	91.1 / 8.0 / 0.4 / 0.4		
R.m.s. deviations[f] (bonds, Å / angles, °)	0.015 / 1.5		

[a] Values in parentheses are for the highest resolution bin, where applicable.
[b] Data in the higher resolution shell are less complete because of data collection on a square detector.
[c] $R_{sym} = \Sigma_h \Sigma_i |I_{hi} - \langle I_h \rangle| / \Sigma_h \Sigma_i | I_{hi}|$, where I_{hi} is the intensity of the i^{th} measurement of the same reflection and $\langle I_h \rangle$ is the mean observed intensity for that reflection.
[d] $R = \Sigma||Fobs(hkl)| - |Fcalc(hkl)|| / \Sigma|Fobs(hkl)|$.
[e] According to the program PROCHECK (Laskowski, 1993). The percentages are indicated of residues in the most favored, additionally allowed, generously allowed and disallowed regions of the Ramachandran plot, respectively.
[f] Estimates provided by the program REFMAC (Murshudov *et al*., 1999).
[g] Bin free R value set count is 66 reflections.

Structure overview and the active site

The crystal structure shows UVDE to be a single-domain TIM barrel lacking the α8 helix of the prototypical TIM-barrel fold (Figure 1). Apart from the eight central β strands forming the barrel, two additional β strands (β1* and β2*) are present in the N-terminal part of the protein, before the α1 helix. Unexplained density at the tip of the side chain of Lys229 suggested this residue

may be modified, but the biochemical characterization and potential role of this modification are beyond the scope of this paper.

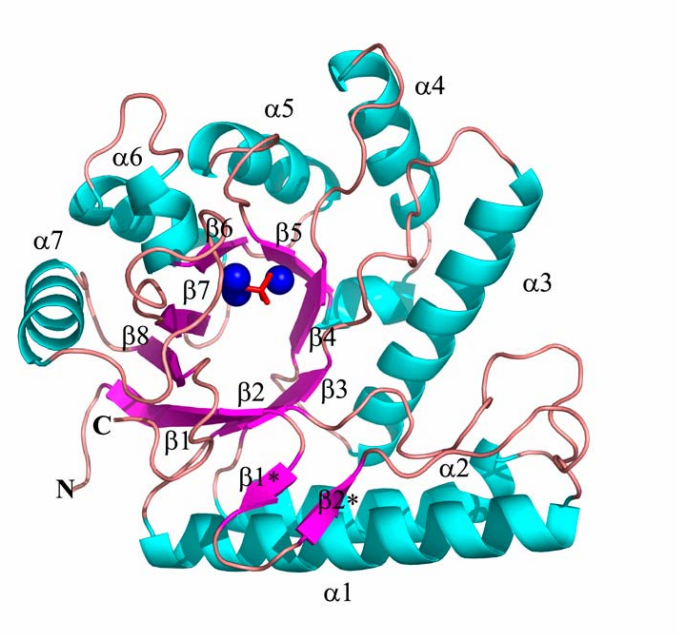

Figure 1. UVDE secondary structure.

View of the UVDE overall fold and topology. The α helices and β strands are labelled according to the canonical TIM-barrel fold and colored light blue and purple respectively. The metal ions are colored blue with the coordinating phosphate in red.

The refined structure of UVDE has three anomalously scattering metal ions, located closely to the C terminus. This, together with the close proximity of the protein's N and C termini, classifies UVDE as a member of the TIM-barrel family of divalent metal-dependent enzymes. The metal ions are well ordered, with a mean temperature factor of 18.5, 68.7, and 74.4 $Å^2$. Because the occupancies of all metal ions were set to one, the high mean temperature factors for two of them probably represent low occupancy rather than high mobility. One of the metal ions is octahedrally coordinated by the side chains of residues Glu175, Glu269, His231, and Asp200 and two oxygen atoms from a phosphate ion (Figure 2A). Another metal ion shows distorted bipyramidal coordination by His101, His143, and Glu175 and two oxygen atoms from the phosphate ion (Figure 2B), while the third metal has an irregular four-fold coordination by one oxygen atom from the phosphate, His244, His203, and one water molecule (Figure 2B).

As a result of the low occupancy for two of the three metal ions, a fluorescence scan performed at the beamline was unable to determine the nature of the metal ions. Because we did not include any metal ions in the crystallization buffer, UVDE must have picked them up

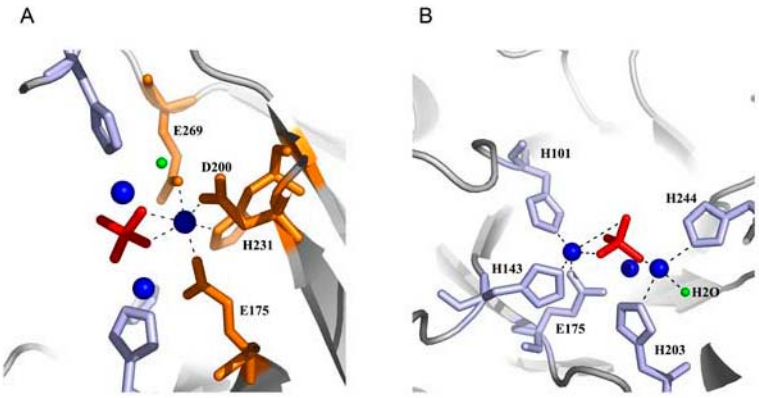

Figure 2. Metal coordination.

A. The octahedrally coordinated metal ion is colored blue. The four coordinating residues, H231, D200, E269 and E175 are shown in ball-and-stick representation and colored orange. The phosphate coordinated by two oxygen atoms is colored red.

B. The distorted bipyrimidal coordination of the second metal ion by His101, His143, Glu175 and two oxygen atoms from the phosphate ion, while the third metal ion has an irregular four fold coordination by one oxygen atom from the phosphate, His244, His203 and one water molecule. The residues involved in the coordination are colored in light blue (H101, H143, H244 and H203). The phosphate is colored red, the water molecule is in green and the metal ions are colored blue.

from its heterologous expression host *Escherichia coli* or during the purification procedure, which included the use of a nickel column. For incision of DNA containing CPD and 6-4PP UV lesions, however, the additional presence of 1 mM Mn^{2+} was required (see below). Inclusion of 10 mM Mg^{2+} instead of Mn^{2+} could only marginally activate the enzyme: CPD-damaged DNA was not incised, whereas the 6-4PP-containing substrate was incised with extremely low efficiency. Apparently Mn^{2+} is a required cofactor, and possibly one or more metal sites picked up nickel during the purification procedure, which has to be exchanged with manganese for enzyme activity. However, it remains to be determined how many ordered Mn^{2+} ions are required for full activity of UVDE. These observations agree with the structurally related enzymes Endo IV (Hosfield *et al*., 1999) and xylose isomerase (Carrell *et al*., 1989), which need a cluster of three or two divalent ions for catalytic activity, respectively.

Except for H244, all the metal-coordinating residues are fully conserved in all known UVDE homologs (Figure 3), and we tested the importance of some of these residues in mutational studies. Residue Glu175 was mutated into an alanine (E175A) and the incision efficiency of this point mutant was compared to that of the wild-type protein using 5'-labelled substrates containing either a CPD or a 6-4PP. Incubation of these substrates with wild-type UVDE resulted in incision efficiencies of 95 % and 90 %, respectively (Figures 4A and 4B, lane 2, both panels). Incubation of the same substrates with the E175A mutant did not result in any detectable incision on the

CPD substrate and a very low incision on the 6-4PP-containing substrate (Figures 4A and 4B, lane 7, both panels). Having a severe kink of ~ 44°, the DNA duplex is more distorted in the 6-4PP than in the CPD substrate (Kim and Choi, 1995). This significant distortion might facilitate the recognition of the 6-4PP damage, which could result in the observed residual activity of the catalytically impaired UVDE E175A mutant.

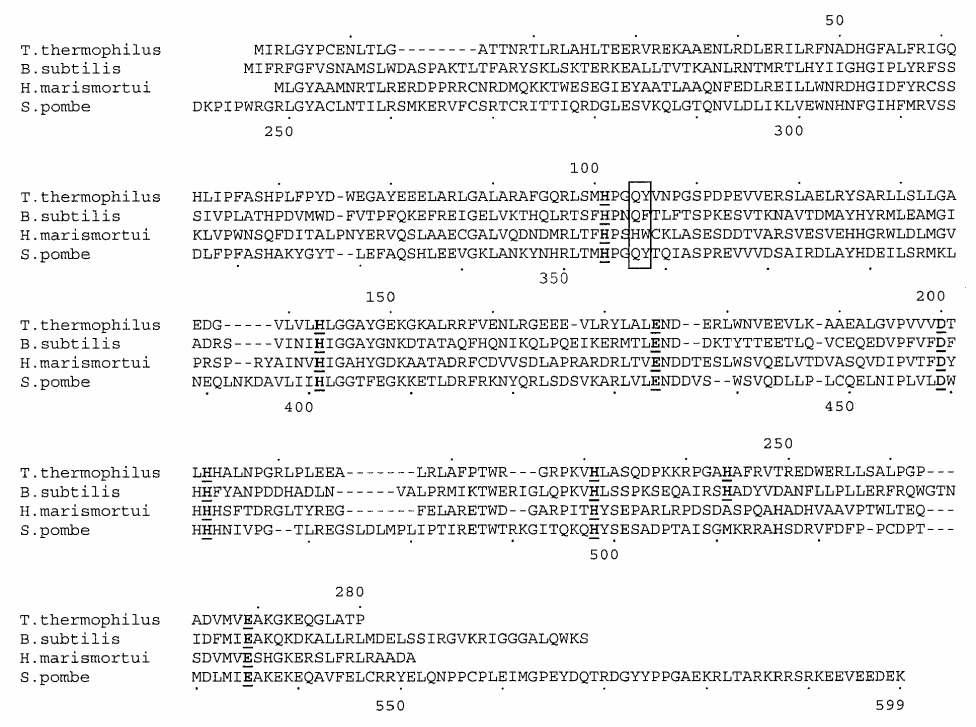

Figure 3. Alignment of UVDE homologues.

The amino acid sequence of UVDE from *T. thermophilus* (Henne *et al*., 2004) is aligned with a homologue from another eubacterium, *B. subtilis* (Kunst *et al*., 1997), a homologue from an archaebacterium *H. marismortui* (Baligal *et al*., 2004) and a homologue from a eukaryote, *S. pombe* (Takao *et al*., 1996). Note that the *S. pombe* protein has an additional N-terminal extension of 240 amino acids. The metal-coordinating residues (H101, H143, E175, D200, H203, H231, H244 and E269) are in bold and underlined. The Gln104 and Tyr105 residues proposed to intercalate the DNA are boxed.

Mutating homologous metal-coordinating residues in *S. pombe* UVDE (Glu434, Asp459, and His498; see Figure 3) also resulted in proteins with impaired incision. No detectable incision was observed with mutants E434A and D459A, either on the CPD or on the 6-4PP, showing the crucial role of these residues for the *S. pombe* catalytic site. Mutant H498A showed a reduced activity on the two substrates, indicating that in the absence of this residue the active site metals can still bind, albeit with lower affinity. Taken together, these data suggest that the UVDE active

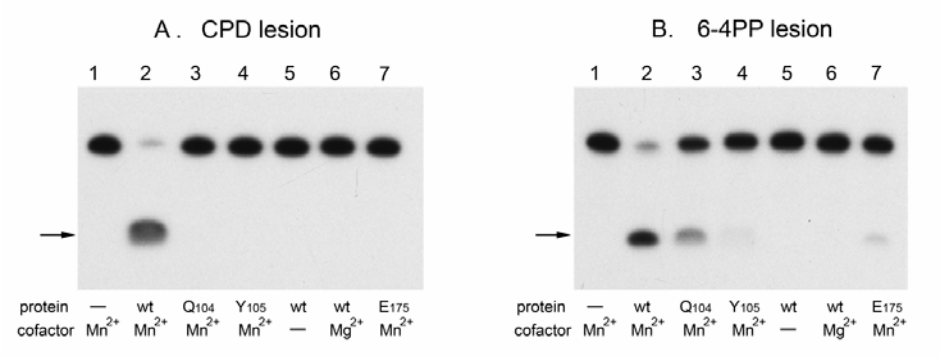

Figure 4. DNA incision by UVDE. Terminally labeled 30 bp DNA substrates with a CPD.

A or 6-4PP lesion B were incubated with (mutant) UVDE protein in the presence or absence of the indicated metal ions. The incision product is indicated with an arrow. The lanes marked Q104, Y105 and E175 contain the corresponding alanine substitutions at these positions.

site metal ions are likely to be directly involved in phosphodiester cleavage, as observed in other DNA repair enzymes such as Endo IV. Indeed, the three metal sites that were observed superimpose on the three zinc sites of the Endo IV structure. The positively charged metal ions can act as a Lewis acid to stabilize a water-derived hydroxide attacking the phosphodiester backbone of DNA and counteracting the developing negative charge on the DNA during the cleavage reaction.

Comparison with endo IV and potential interactions with damaged DNA

Superposition of UVDE with the DNA-repair enzyme Endo IV (Hosfield *et al*., 1999) from the BER pathway shows that the two enzymes share major structural features, including the TIM-barrel fold (Figure 5A) and a wide groove (29 Å) that houses the active site at the bottom (Figure 5B). An extensive positive charge on both sides of the solvent-accessible groove of UVDE (Figure 5C) can be seen, which is suited to bind a DNA duplex.

However, there are important differences, which reflect the functional divergence between the two enzymes: Endo IV only nicks at abasic sites, whereas UVDE also recognizes other types of damage. Moreover, incision of the *T. thermophilus* UVDE on the abasic site lesion is only 20 % (Figure 6), which is lower compared to the reported efficiency of *S. pombe* UVDE (Kanno *et al*., 1999). The difference in substrate specificity of the *T. thermophilus* homolog in comparison with UVDE proteins from other species will be addressed elsewhere (unpublished data).

In the structure of Endo IV in complex with a DNA duplex nicked at the abasic site, the DNA helix has a ~ 90° kink and both the abasic phosphoribose and the base opposing this damage are flipped out of the double helix (Hosfield *et al*., 1999). The nicked, abasic DNA of the Endo IV-DNA complex could be fitted analogously into UVDE: the deep groove of UVDE can comfortably harbour the kinked duplex. The active sites of UVDE and Endo IV are both located at the bottom of the groove.

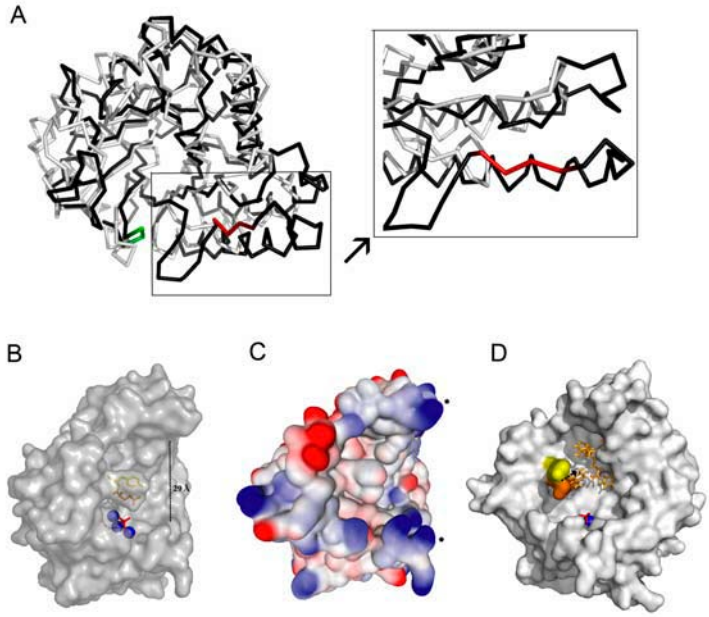

Figure 5. DNA binding site.

A. Cα superpositions of UVDE and Endo IV showing that both enzymes share major structural features. UVDE is shown in black and Endo IV in gray. Residues 18 - 21 in the NRTL-strand of UVDE causing a clash between the DNA en UVDE are colored red. The residues Lys273 and Glu274, which need to make a small rearrangement for fitting of the flipped-out base, are colored green. The superposition was done using the program Theseus (Theobald and Wuttke, 2006)

B. Surface representation of UVDE showing the (semi)conserved residues Tyr105 and Gln104 in ball-and stick representation and colored in yellow and orange respectively. The metal ions are colored blue and the coordinating phosphate is in red.

C. Electron surface potential of UVDE. Positive charges are marked in blue, and negative charges are in red. The positively charged rims of the groove (marked with *) suggest the DNA binding site.

D. Surface representation of UVDE showing the cavity allowing Q104 and Y105 movement. The surface of UVDE is rotated 90° clockwise compared to B and C. The surfaces of residues Q104 and Y105 are colored yellow and orange, respectively. The possible movement of Q104 and Y105 is indicated with a black arrow. The residues of Endo IV (Phe32, Asn35, Gln36, Arg37 and Tyr72) that show that this cavity is not present in Endo IV are colored orange and are shown in ball-and-stick representation.

At the bottom of the proposed DNA binding groove of UVDE is a loop with the conserved sequence GQY, in which Gln104 and Tyr105 point straight into the solvent (Figure 5B). Endo IV has the same loop, albeit with different, Endo IV-specific residues. In Endo IV, the side-chain residues Tyr72 and, to a lesser extent, Leu73, project into the kink of the DNA duplex (Hosfield *et al*., 1999). We propose that the two loops are functionally equivalent and that Gln104 and

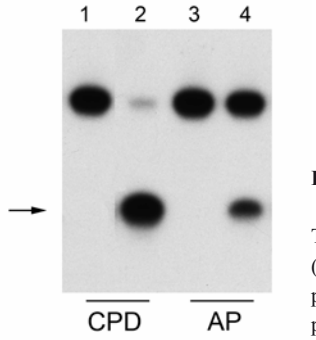

Figure 6. Incision of an abasic site DNA lesion by UVDE.

Terminally labelled 30 bp DNA substrates with a CPD or an Abasic site (AP) lesion (as indicated) were incubated with *T. thermophilus* UVDE protein (lanes 2 and 4) in the presence of 1 mM Mn^{2+}. The incision product is indicated with an arrow.

Tyr105 of the UVDE GQY loop stabilize the kink in the DNA duplex at the position of damage in a similar fashion. For Gln104 and Tyr105 to take up a similar position as Tyr72 and Leu73, the loop needs to shift by about 3 Å upon DNA binding, while the side chains may need to adopt a different conformation. There is a cavity in the UVDE structure that allows such a shift and the presence of the conserved Gly103 suggests potential flexibility of this loop (Figure 5D).

In UVDE, the close proximity of Gln104 and Tyr105 to the metal coordination site suggests that these residues might similarly probe the DNA for damage and present the scissile phosphodiester bond at the 5' side of the lesion to the active site. Tyr105 appears to be semiconserved: in UVDE from other species, phenylalanine or tryptophan can also be found (Figure 3), and these residues could also fulfil the proposed intercalating role.

To confirm the significance of Gln104 and Tyr105 for the UVDE function we created point mutants, changing the residues into alanine (Q104A and Y105A). Indeed, the incision assay with Q104A did not reveal any detectable activity on the CPD (Figure 4A). Some residual activity on the 6-4PP substrate (Figure 4B) is observed, which is again an indication that the 6-4PP might be an easier target for damage recognition and processing. Mutant Y105A did not show any incision on the CPD substrate and an extremely low incision on the 6-4PP (Figures 4A and 4B), underlining the crucial role of Tyr105 for enzyme activity. If DNA binds in the same way in UVDE as in Endo IV, there would be a major clash between the DNA and the NRTL strand of UVDE (comprising residues Asn18 – Leu21; Figures 7A and 7B). There is no equivalent of this strand in Endo IV (Figure 5A, zoom-in). An intriguing possibility is that this strand moves toward the GQY loop, to take up a position similar to the Gln36 – Trp39 loop in Endo IV. Such a movement would reposition Asn18 and Arg19 into the kink of the DNA, where they could make similar interactions with the DNA as Gln36 and Arg37 in Endo IV. These residues in Endo IV are vital for stabilizing the kink and flipping out the base opposing the damage (Hosfield *et al*., 1999).

The conformational similarities between UVDE and Endo IV suggest that also in UVDE, both the damaged base(s) and the base opposing this damage flip out of the double helix. However, is there space for the damaged base to flip out in UVDE? In Endo IV this is not an issue, as it only recognizes DNA with an abasic site. Space for one or more flipped-out bases in

UVDE can only be created by a small rearrangement of two residues at its C terminus: Lys273 and Glu274 (Figure 5A). The presence of Gly272 and Gly276 just before and after these residues suggests they may indeed have the freedom to move. This movement would create a pocket in which the damaged base could stack against Tyr6, while Lys273, Glu274, and possibly also Asn10 could form polar interactions with the damaged base.

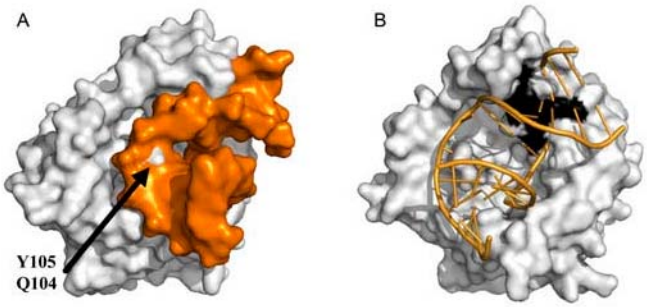

Figure 7. DNA binding.

A. Model of UVDE bound to DNA. For modeling the DNA fragment of the Endo IV co-crystal was used (Hosfield *et al*., 1999). The protein surface is colored in light gray and the DNA is presented in orange surface representation.

B. Same as A, but rotated 90° clockwise. The DNA is presented as a cartoon showing the clash between the DNA and the NRTL-strand of UVDE (amino acids N18-L21 are shown in black surface representation).

CONCLUSION

Our results shed light on the remarkable ability of UVDE to recognize different types of DNA damage. It is much more versatile than its closest homolog Endo IV, which only nicks adjacent to abasic sites. The two enzymes share some characteristic features: a deep DNA binding groove, at the bottom of which are the catalytic site and residues capable of intercalating in the double helix at the site of damage. Furthermore, we observed three metal ions in the same location as the cluster of three Zn^{2+} ions in Endo IV's active site. The intrinsic metals of UVDE, however, seem not to be able to perform an enzymatic activity, but the addition of manganese appears to be required. Superposition of UVDE with Endo IV predicts that the rearrangements in UVDE upon DNA binding are likely to be more substantial than those observed in Endo IV. This structural flexibility might be part of the explanation of the broader substrate specificity of UVDE.

REFERENCES

Aravind L, Walker DR and Koonin EV, Conserved domains in DNA repair proteins and evolution of repair systems, *Nucleic Acids Res*. **27** (1999), pp. 1223-1242

Avery AM, Kaur B, Taylor JS, Mello JA, Essigmann JM and Doetsch PW, Substrate specificity of ultraviolet DNA endonuclease (UVDE/Uve1p) from *Schizosaccharomyces pombe*, *Nucleic Acids Res*. **27** (1999), pp. 2256-2264

Baliga NS, Bonneau R, Facciotti MT, Pan M, Glusman G, Deutsch EW, Shannon P, Chiu Y, Weng RS and Gan RR, Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea, *Genome Res*. **14** (2004), pp. 2221-2234

Carrell HL, Glusker JP, Burger V, Manfre F, Tritsch D and Biellmann JF, X-ray analysis of D-xylose isomerase at 1.9 Å: native enzyme in complex with substrate and with a mechanism-designed inactivator, *Proc*. *Natl*. *Acad*. *Sci*. *USA*. **86** (1989), pp. 4440-4444

CCP4 (Collaborative Computational Project, Number 4), The CCP4 suite: programs for protein crystallography, *Acta Crystallogr*. *D Biol*. *Crystallogr*. **50** (1994), pp. 760-763

Cowtan K, dm: An automated procedure for phase improvement by density modification, *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31** (1994), pp. 34-38

De Graaff RAG, Hilge M, van der Plas JL and Abrahams JP, Matrix methods for solving protein substructures of chlorine and sulfur from anomalous data, *Acta Crystallogr*. *D Biol*. *Crystallogr*. **57** (2001), pp. 1857-1862

DeLano WL, The PyMOL Molecular Graphics System (http://www.pymol.org) (2002)

Doetsch PW, Beljanski V and Song B, The ultraviolet damage endonuclease (UVDE) protein and alternative excision repair: a highly diverse system for damage recognition and processing. In: V. Beljanski and B. Song, Editors, *DNA Damage Recognition*, Taylor & Francis Press, New York (2006), pp. 211-223

Emsley P and Cowtan K, Coot: model-building tools for molecular graphics, *Acta Crystallogr*. *D Biol*. *Crystallogr*. **60** (2004), pp. 2126-2132

Evans PR, Data reduction. In: E. Dodson, M. Moore, A. Ralph and S. Bailey, Editors, *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, Daresbury Laboratory, Warrington, UK (1993), pp. 114-122

Friedberg EC, Walker GC, Siede W, Wood RD, Schultz AR and Ellenberger T, DNA Repair and Mutagenesis, ASM Press, Washington, DC (2005)

Henne A, Bruggemann H, Raasch C, Wiezer A, Hartsch T, Liesegang H, Johann A, Lienard T, Gohl O and Martinez-Arias R, *et al*., The genome sequence of the extreme thermophile *Thermus thermophilus*, *Nat*. *Biotechnol*. **22** (2004), pp. 547-553

Hosfield DJ, Guan Y, Haas BJ, Cunningham RP and Tainer JA, Structure of the DNA repair enzyme Endo IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis, *Cell* **98** (1999), pp. 397-408

Iwai S, Chemical synthesis of oligonucleotides containing damaged bases for biological studies, *Nucleosides Nucleotides Nucleic Acids* **25** (2006), pp. 561-582

Kanno S, Iwai S, Takao M and Yasui A, Repair of apurinic/apyrimidinic sites by UV damage endonuclease; a repair protein for UV and oxidative damage, *Nucleic Acids Res*. **27** (1999), pp. 3096-3103

Kaur B, Avery AM and Doetsch PW, Expression, purification, and characterization of ultraviolet DNA endonuclease from *Schizosaccharomyces pombe*, *Biochemistry* **37** (1998), pp. 11599-11604

Kaur B, Fraser JL, Freyer GA, Davey S and Doetsch PW, A Uve1p-mediated mismatch repair pathway in *Schizosaccharomyces pombe*, *Mol*. *Cell*. *Biol*. **19** (1999), pp. 4703-4710

Kim JK and Choi BS, The solution structure of DNA duplex-decamer containing the (6-4) photoproduct of thymidylyl (3'→5') thymidine by NMR and relaxation matrix refinement, *Eur. J. Biochem.* **228** (1995), pp. 849-854

Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A and Borchert S, *et al.*, The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*, *Nature* **390** (1997), pp. 249-256

Laskowski RA, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Cryst.* **26** (1993), pp. 283-291

Leslie AG, Integration of macromolecular diffraction data, *Acta Crystallogr. D Biol. Crystallogr.* **55** (1999), pp. 1696-1702

Murshudov GN, Vagin AA, Lebedev A, Wilson KS and Dodson EJ, Efficient anisotropic refinement of macromolecular structures using FFT, *Acta Crystallogr. D Biol. Crystallogr.* **55** (1999), pp. 247-255

Pannu NS, Murshudov GN, Dodson EJ and Read RJ, Incorporation of prior phase information strengthens maximum likelihood structure refinement, *Acta Crystallogr. D Biol. Crystallogr.* **54** (1998), pp. 1285-1294

Pannu NS, McCoy AJ and Read RJ, Application of the complex multivariate normal distribution to crystallographic methods with insights into multiple isomorphous replacement phasing, *Acta Crystallogr. D Biol. Crystallogr.* **59** (2003), pp. 1801-1808

Perrakis A, Morris R and Lamzin V.S, Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol.* **6** (1999), pp. 458-463

Sohi M, Alexandrovich A, Moolenaar GF, Visse R, Goosen N, Vernede X, Fontecilla-Camps J.C, Champness J and Sanderson M.R, Crystal structure of *Escherichia coli* UvrB C-terminal domain, and a model for UvrB-UvrC interaction, *FEBS Lett.* **465** (2000), pp. 161-164

Studier FW, Rosenberg AH, Dunn JJ and Dubendorf JW, Use of T7 RNA polymerase to direct expression of cloned genes, *Methods Enzymol.* **185** (1990), pp. 60-89

Takao M, Yonemasu R, Yamamoto K and Yasui A, Characterization of a UV endonuclease gene from the fission yeast *Schizosaccharomyces pombe* and its bacterial homolog, *Nucleic Acids Res.* **24** (1996), pp. 1267-1271.

Theobald DL and Wuttke DS, Theseus: maximum likelihood superpositioning and analysis of macromolecular structures, *Bioinformatics* **22** (2006), pp. 2171-2172

Truglio JJ, Croteau DL, Van Houten B and Kisker C, Prokaryotic nucleotide excision repair: the UvrABC system, *Chem. Rev.* **106** (2006), pp. 233-252

Verhoeven EE, van Kesteren M, Turner JJ, van der Marel GA, van Boom JH, Moolenaar GF and Goosen N, The C-terminal region of *Escherichia coli* UvrC contributes to the flexibility of the UvrABC nucleotide excision repair system, *Nucleic Acids Res.* **30** (2002), pp. 2492-2500