Structure-based insights into the repair of UV-damaged DNA

Meulenbroek, E.M.

6

Unravelling UVDE's uncanny ability to recognize and incise different types of damaged DNA

UV damage endonuclease (UVDE) is a DNA repair enzyme that can recognize and incise a diverse set of DNA lesions including UV-induced DNA damage and DNA containing abasic sites. In this chapter, we present the structure of UVDE from *Sulfolobus acidocaldarius* (*Sac*UVDE) and a pre-catalytic structure of *Sac*UVDE in complex with DNA containing a 6-4 photoproduct. These structures show that UVDE has an intriguing "dual flip" mechanism: the two purines opposite to the damaged pyrimidine bases are flipped into a dipurine-specific pocket, whilst the damaged bases are also flipped into a pocket. In contrast to UVDEs from other species, *Sac*UVDE shows a marked preference for DNA substrates containing 6-4 photoproducts compared to cyclobutane pyrimidine dimers: biochemical assays and mutagenic studies show that the flexibility of the damage binding pocket and positive charges on both sides of the substrate binding groove are main contributors to the broad substrate specificity for UVDE in general.

E.M. Meulenbroek, C. Peron Cane, I. Jala, S. Iwai, G.F. Moolenaar, N. Goosen, N.S. Pannu, *UV damage endonuclease employs a novel dual-dinucleotide flipping mechanism to recognize and incise different types of damaged DNA* (to be submitted)

6.1 Introduction

The UV damage endonuclease (UVDE) repair pathway, present in some prokaryotes and lower eukaryotes, was first identified in *S. pombe* where Nucleotide Excision Repair deletion mutants were seen to be only moderately UV-sensitive (McCready *et al*., 1993). Its central enzyme, UVDE, was found to be a DNA endonuclease that not only recognizes and incises DNA 5′ to cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4PPs) (Bowman *et al*., 1994), but also non-UV-induced DNA damage such as abasic sites, nicks and gaps (Avery *et al*., 1999). The activity on abasic sites, nicks and gaps was seen to depend on the presence of neighboring pyrimidines, suggesting that UVDE's active site is most optimal for binding distorted pyrimidines (Paspaleva *et al*., 2009). Later, UVDE homologues were discovered in *Neurospora crassa* and from *Bacillus subtilis* that incise at least CPD, 6-4PP and abasic site very efficiently (Kanno *et al*., 1999; unpublished results) and UVDE from *Thermus thermophilus* that efficiently cleaves CPDs and 6-4PPs while incising abasic sites only with moderate efficiency (Paspaleva, Thomassen *et al*., 2007).

Insights into the mechanism of UVDE came from the crystal structure of UVDE from *Thermus thermophilus* (Paspaleva, Thomassen *et al*., 2007). *Tth*UVDE has a TIM-barrel fold with a large groove with positive charges on the edges where DNA was proposed to bind based on the structural similarity to Endonuclease IV (Hosfield *et al*., 1999). The active site containing three metal ions is located on the bottom of this groove and two conserved residues (Gln and Tyr; together called the 'probing finger') were proposed to aid in flipping out the damaged bases from the DNA helix. Thus, the structure suggested plausible mechanisms for DNA binding and three-metal-ion catalysis, but could not explain UVDE's wide substrate specificity.

To explain the broad specificity, we have determined the structures of UVDE *Sulfolobus acidocaldarius* (*Sac*UVDE) on its own to 1.5 Å resolution and in a pre-catalytic complex with DNA containing a 6-4 photoproduct determined to 2.7 Å. The structures show that UVDE recognizes damaged pyrimidine dimers by flipping the two bases opposite to the damage into a dipurine-specific pocket and by flipping the damage itself into a flexible damage-binding pocket. Positive charges on both sides of the DNA binding-groove probably aid in binding different damaged DNA substrates.

6.2 Materials and methods

6.2.1 Cloning

Genomic DNA was isolated from *S. acidocaldarius* by resuspension of cells in TEN-buffer (20 mM Tris pH 8, 1 mM EDTA, 100 mM NaCl), followed by lysis in TENST-buffer (20 mM Tris pH 8, 1 mM EDTA, 100 mM NaCl, 1.6 % sarcosyl, 0.12 % Triton) and phenol/ chloroform extraction. The gene for UVDE was amplified using the primers 5′ ATTAATAACATATGAGAGTAGGTTACGTATCCAC 3′ and 5′ TAGGATCCATTAATCCAGTTTGTTTAACTCCTTTAAC 3′. Subsequently, it was cloned into the pETUVDEΔ228 vector (Paspaleva *et al*., 2009) using *Nde*I and *Bam*HI, re-

Table 6.1: 30-mer DNA substrates used in this study.

No damage	5′ CTCGTCAGCATCTTCATCATACAGTCAGTG 3′ 3′ GAGCAGTCGTAGAAGTAGTATGTCAGTCAC 5′
CPD and 6-4PP	5′ CTCGTCAGCATC**TT**CATCATACAGTCAGTG 3′ 3′ GAGCAGTCGTAGAAGTAGTATGTCAGTCAC 5′
Abasic site	5′ CTCGTCAGCATC**X**TCATCATACAGTCAGTG 3′ 3′ GAGCAGTCGTAGAAGTAGTATGTCAGTCAC 5′

The positions of the CPD (**TT**), (6-4)PP (**TT**) and AP site (**X**) are indicated in bold.

sulting in the gene for SacUVDE with a N-terminal 10x His-tag and factor Xa cleavage site. Mutants of *Sac*UVDE were created by PCR and cloned into the same vector with *Nde*I and *Bam*HI. All constructs were verified by sequencing.

6.2.2 Expression and purification

The plasmid with the gene for *Sac*UVDE was transformed to *E.coli* BL21(DE3)-codon+ and overexpressed for 2 hours at 37°C after induction by 0.5 mM IPTG. After harvesting, the pellet was resuspended in Ni buffer A (20 mM Tris pH 7.5, 500 mM NaCl, 50 mM imidazole, 8 mM β-mercaptoethanol and 10 % glycerol) and the cells were lysed by sonication. The lysate was spun down at 37.000 rpm (100.000 g) for 30 min and the soluble fraction was loaded on a His-trap column (GE healthcare) equilibrated with Ni buffer A. The column was washed with 20 column volumes Ni buffer A and the protein was then eluted with a 60 column volumes gradient to Ni buffer B (20 mM Tris pH 7.5, 500 mM NaCl, 500 mM imidazole, 8 mM β-mercaptoethanol and 10 % glycerol). Fractions containing *Sac*UVDE were dialyzed to 20 mM Tris pH 8 and were then loaded on a HiTrap Q column (GE healthcare) equilibrated with Q buffer A (20 mM Tris pH 8 and 10 % glycerol). The column was washed with 10 column volumes Q buffer A and eluted with a 60 column volumes gradient to Q buffer B (20 mM Tris pH 8, 1 M NaCl and 10 % glycerol).

For crystallization, the purification protocol was adapted: size exclusion (Superdex 200, GE healthcare) was performed in GF buffer (20 mM HEPES 7.2, 200 mM NaCl, 5 mM DTT) after the Ni purification instead of ion exchange. The protein from either protocol was found to be more than 95 % pure as judged from SDS PAGE. All purification steps were performed at 4 °C.

6.2.3 DNA substrates

The DNA substrates used in this study can be seen in Table 6.1. The oligos containing CPD or 6-4PP were synthesized as described in Iwai, 2006. The top strands of the DNA substrates were 5′ radioactively labelled using polynucleotide kinase as reported previously (Verhoeven *et al.*, 2002).

6.2.4 Incision assays

Labelled DNA substrates (1 nM) were incubated for 15 min at 55 °C with the indicated amount of UVDE (in the range 0.05 to 50 nM) in 20 mM HEPES pH 6.5, 100 mM NaCl and 1 mM $MnCl_2$ in a reaction mix of 20 μl. The reaction was then stopped by adding 3 μl stop mix (0.33 M EDTA, 3.3 % SDS), after which 2.4 μl 4 mg/ml glycogen was added and the DNA was precipitated by ethanol. Samples were loaded on a 15 % denaturing polyacrylamide gel and visualized by autoradiography. For kinetics incision assays, a mix was prepared of buffer, cofactor, protein (25 nM) and DNA and put at 55 °C. At the indicated time points, samples were taken out and the reaction was stopped in these samples.

6.2.5 Bandshift assays

Labelled DNA substrates (0.1 nM) were incubated for 10 min on ice with 660 nM UVDE in 20 mM Tris pH 6.5, 100 mM NaCl and 1 mM $MnCl_2$ in a reaction mix of 10 μl. Samples were loaded on a 6 % native gel, which was run at 4 °C in 1x TBE. The gel was dried and the result was visualized by autoradiography.

6.2.6 Filter-binding assays

Labelled 50-mer DNA substrates (4 nM) were incubated for 7 min with 50 nM UVDE at 55 °C in a reaction buffer containing 20 mM Tris pH 6.5, 100 mM NaCl and 1 mM $MnCl_2$ in a reaction mix of 20 μl. After incubation, 0.1 ml reaction buffer was added and the mixture was poured over a nitrocellulose filter. The filters were washed with 0.2 ml reaction buffer. The amount of DNA retained on the filter (by binding to UVDE) was determined using a scintillation counter. Each sample was corrected for the amount of DNA retained on a filter in the absence of protein. Binding is expressed as the percentage of the DNA retained on the filter divided by the input DNA.

6.2.7 Crystallization

*Sac*UVDE was concentrated to 3-5 mg/ml with a 3 kDa MWCO centrifugal filter unit (Millipore). Crystallization trials were performed using the sitting-drop vapour diffusion method and the JCSG+ and PACT (Qiagen) screens. *Sac*UVDE crystals were obtained in 20 % PEG3350 with 0.2 M NH_4Cl or 0.2 M NaI. The conditions were optimized by a systematic screen around these conditions and the largest crystals were grown in 14 to 28 % PEG3350 with 0.15 to 0.3 M NH_4Cl.

The damaged strand of the oligo containing 6-4PP for crystallization was synthesized as previously described (Iwai, 2006) while the undamaged strand was purchased from Eurogentec, Belgium. The sequence of the oligo was: 5' GCGTCC**TT**GACGACG 3' with the site of the damage printed in bold. The two strands were hybridized by heating to 80 °C for two minutes in 20 mM Tris pH 7 and then allowed to slowly cool down to room temperature. For co-crystallization, protein (at 0.11 mM) and DNA (at 0.21 mM) were incubated on ice for 15 minutes,

after which sitting-drop vapour diffusion experiments were set up in the NucPro screen (Jena Biosciences). Damaged DNA:protein complex crystals appeared after several days in 30 % PEG2000-MME, 100 mM acetate buffer pH 4.6, and 200 mM $(NH_4)_2SO_4$.

6.2.8 Data collection

Crystals were caught with SPINE sample loops and put in cryoprotectant solution (precipitant solution with 10-15 % glycerol) and flash-frozen. Data were collected at the European Synchrotron Radiation Facility (Grenoble, France). 180 images were collected with an oscillation angle of 1.0° with transmission of 13 % and exposure time of 0.5 s per frame at 0.9393 Å at 100 K on beamline ID14-4 for the apoprotein crystals. For the DNA:protein complex crystals, 150 images were collected with an oscillation angle of 1.0° and an exposure time of 25 s per frame at 0.934 Å at 100 K on beamline ID14-1. The images were processed with *iMosflm* (Leslie, 2011). Scaling and merging were done with *SCALA* (Evans, 2006) from the *CCP4* suite (Winn *et al*., 2011). For the apoprotein structure, two datasets (from two different crystals) were merged to yield the final dataset used for determining the structure. Data collection statistics are shown in Table 6.2.

6.2.9 Structure solution and refinement

The phase problem for the apoprotein was solved by molecular replacement using the structure of *Tth*UVDE (PDB entry 2j6v) as a search model. The model was automatically rebuilt using *ArpWarp* (Perrakis *et al*., 1999) and refined with *Refmac* (Murshudov *et al*., 2011). Manual fitting was performed using *Coot* (Emsley *et al*., 2010). For the DNA:protein complex, the phase problem was solved by molecular replacement using the structure of *Sac*UVDE as a search model. Clear difference density was visible for the DNA (Figure 6.1(e)) and the DNA was built in manually in *Coot*. The model was refined with *Refmac* and further manual fitting was also performed using *Coot*. The final R-factor and R_{free} for uncomplexed *Sac*UVDE were 0.177 and 0.214 respectively and 0.201 and 0.271 for *Sac*UVDE in complex with DNA. Refinement statistics are shown in Table 6.2.

Superpositions were done with the ssm function in *Coot*. Root-mean-square deviation calculations were done using *Theseus* (Theobald & Wuttke, 2006). Structure-based sequence alignment was performed using the program *VAST* (Thompson *et al*., 2009). Atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank (accession code 3tc3 for the apoprotein and 4gle for the cocrystal structure). All figures were made with *CCP4MG* (Potterton *et al*., 2004).

6.3 Results

6.3.1 Overall structure of *Sac*UVDE with and without damaged DNA

To gain insight into UVDE's damage recognition and incision mechanism, a co-crystal structure of UVDE with damaged DNA is needed. Crystallization of *Tth*UVDE and *Bsu*UVDE with different damage-containing oligos (abasic site, CPD, different lengths and ends of oligos) was unsuccessful: only crystals of either protein or DNA alone were obtained. This is probably due to the relatively low affinity of UVDE for damaged DNA and the high propensity of *Tth*UVDE to crystallize on its own. Luckily, we found a UVDE homologue that more strongly interacts with DNA: UVDE from *Sulfolobus acidocaldarius*, an aerobic thermoacidophilic crenarchaeon. This homologue is more specific for 6-4PPs and also binds slightly stronger to this damage than *Tth*UVDE (see section 6.3.3). Crystallization of this protein with and without DNA were successful: a crystal of *Sac*UVDE diffracted to 1.5 Å and a crystal of *Sac*UVDE in complex with DNA containing 6-4 photoproduct, diffracted to 2.7 Å (statistics are shown in Table 6.2). The quality of the map of both structures is shown in Figure 6.1(a) and (c). Clear difference density for DNA was seen in the complex after molecular replacement with the apoprotein, into which the damaged DNA could be built manually. An OMIT map for the DNA can be seen in Figure 6.1(e). For uncomplexed *Sac*UVDE, two very similar chains are present in the asymmetric unit (root mean square deviation of the Cα atoms is 0.064 Å), while the DNA:protein complex crystals contain only one chain in the asymmetric unit.

*Sac*UVDE has a TIM-barrel fold (Figure 6.1(b)) and the *Sac*UVDE backbone structure is very similar to *Tth*UVDE: the root mean square deviation of the Cα atoms is only 0.651 Å. The sequence identity between these two proteins is 38 %. Figure 6.2 shows a structure-based sequence alignment. Density is only present for one metal in the *Sac*UVDE structure; this ion is located near (0.42 Å) the position of metal ion Mn1 in *Tth*UVDE (the most stably bound metal ion). During refinement, the occupancy for this metal ion was set to unity, but the B-factor of this ion refined to a relatively high value (50.78), suggesting either high mobility or low occupancy. The metal coordinating residues in *Sac*UVDE are at similar positions as those of all three metal ions in *Tth*UVDE. Addition of manganese is needed for activity in incision assays and therefore it is very likely that active *Sac*UVDE also uses a three-metal ion catalysis. Another notable difference between the structures of *Sac*UVDE and *Tth*UVDE, is that *Sac*UVDE has one extra α-helix (α8; Figure 6.1(b)) at its C-terminus. Thus, *Sac*UVDE does have a complete $\alpha 8\beta 8$ TIM-barrel, like the majority of UVDEs, in contrast to *Tth*UVDE.

The overall fold of *Sac*UVDE in the complex structure with DNA containing a 6-4 photoproduct is very similar to that of the apoprotein; the root mean square deviation of the Cα atoms is only 0.11 Å and thus hardly any changes occur upon DNA binding (Figure 6.1(d)). In the DNA:protein complex structure, the DNA is bound in the previously predicted DNA-binding groove. The DNA makes a bend of around 90 ° (Figure 6.1(f)), like in the related structure of Endonuclease IV with DNA containing an abasic site (Hosfield *et al.*, 1999). Numerous interactions of side-

Table 6.2: Data collection and refinement statistics.

Data collection		
	*Sac*UVDE	*Sac*UVDE with 6-4PP
Space group	P1	$C222_1$
Cell dimensions		
a, b, c (Å)	42.08 x 53.59 x 77.39	57.20 x 112.51 x 153.85
α, β, γ (°)	102.09, 93.02, 111.76	90.00, 90.00, 90.00
Resolution (Å)	46.05-1.50 (1.58-1.50)[a]	52.83-2.70 (2.85-2.70)[a]
Wilson plot B-factor	17.5	64.7
R_{merge}	0.099 (0.520)	0.121 (0.832)
I/σI	9.3 (1.9)	11.1 (1.9)
Completeness (%)	97.3 (94.0)	99.9 (99.2)
Redundancy	3.4 (2.4)	5.9 (5.2)
Total n° observations	317749	82168
N° unique reflections	94666	14041
Refinement		
Resolution (Å)	46.05-1.50	52.89-2.70
N° reflections	89933	13316
N° molecules in ASU	2	1
R_{work}/ R_{free}	0.177/ 0.214	0.191/ 0.262
N° atoms		
Protein	4706	2353
Metal ions	2	0
Water	579	4
DNA	0	609
Other ions	0	10
B-factors		
Protein	20.37	55.28
Metal ions	50.78	NA
Water	31.93	34.14
DNA	NA	67.83
Other ions	NA	60.77
R.m.s. deviations		
Bond lengths (Å)	0.026	0.011
Bond angles (°)	2.21	1.57
N° TLS bodies	2	NA
Ramachandran favored[b]	96.87 %	92.8 %
Ramachandran outliers[b]	0.00 %	0.69 %
Rotamer outliers[b]	1.48 %	2.95 %

[a] Values in parentheses are for the highest resolution shell.
[b] As determined by *Molprobity* (Davis *et al*, 2007).

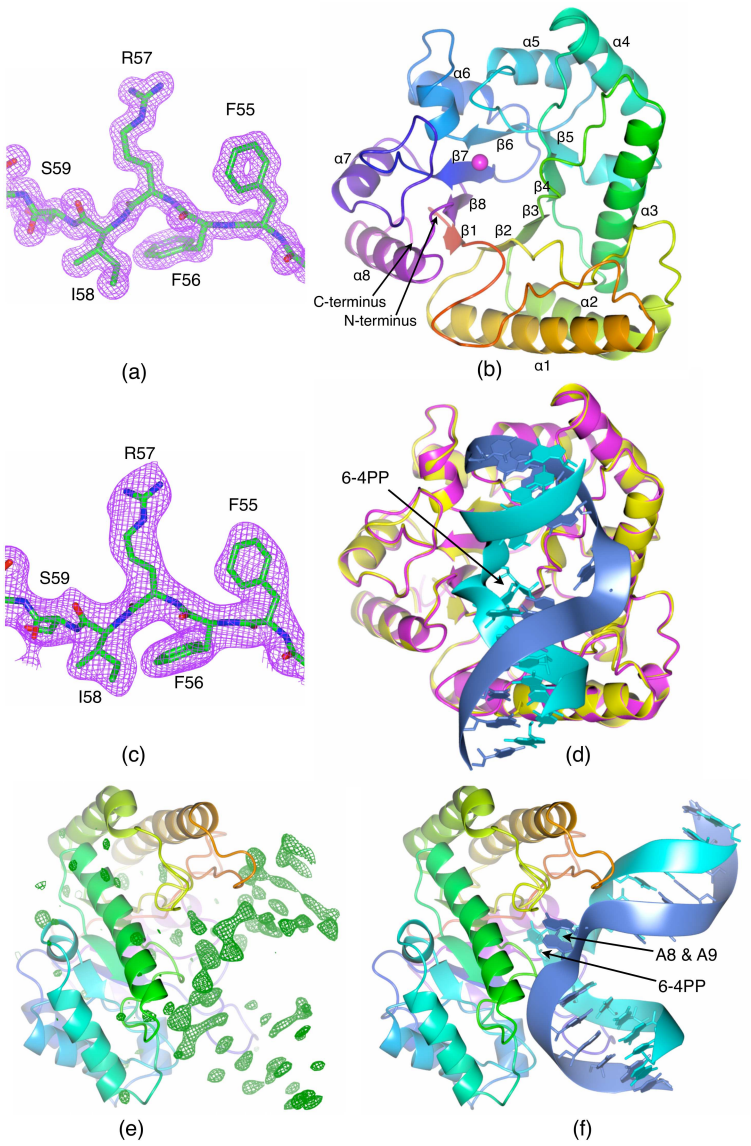

Figure 6.1: Overall structure of SacUVDE with and without DNA.
(a) Representative part of the electron density map of SacUVDE.
(b) Overall fold of SacUVDE, which has a TIM-barrel fold. The metal ion is shown in magenta.
(c) Representative part of the electron density map of SacUVDE in complex with DNA.
(d) Superposition of SacUVDE with (magenta) and without (yellow) DNA (blue and cyan), showing that the two structures are very similar.
(e) Omit map of the SacUVDE cocrystal with DNA showing clear positive difference density for the DNA (contoured at 3 σ).
(f) Overall fold of the SacUVDE-DNA complex, showing the 90 $^\circ$ bend in the DNA.

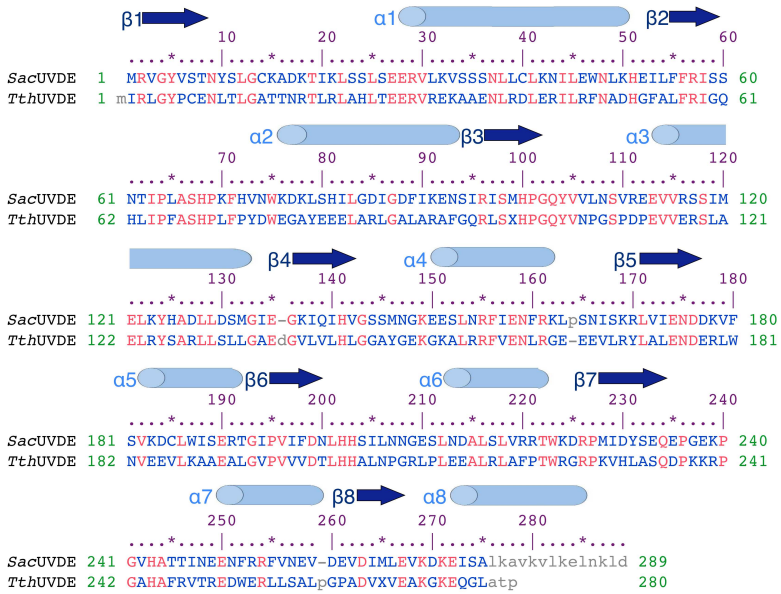

Figure 6.2: Structure-based sequence alignment of SacUVDE and TthUVDE. Secondary structure elements are indicated with cylinders and arrows for α-helices and β-strands respectively. Identical residues are in red, variable residues in blue, and unaligned residues in grey.

chains or backbone amides of the protein mostly with the DNA phosphates (and some with the base or the deoxyribose ring) hold the DNA in the DNA binding groove (Figure 6.3). The residues responsible for these interactions are partially or fully conserved in UVDE from different organisms, hence our structure is of general relevance for the understanding of the mode of action of the entire UVDE family. In the structure of *Sac*UVDE in complex with damaged DNA, no metal ions can be seen and thus the DNA is not incised and the complex is in a pre-incision state. The absence of metals in the structure is probably an effect of the crystallization condition. The position of the metal ions can be estimated by superposition with the crystal structure of *Tth*UVDE that contains all three metals. In the DNA:protein complex structure, the scissile P-O bond is still several Å away from the correct position for cleavage, but the positive charge of the metal ions is likely required to draw the scissile bond inwards to the correct position, allowing incision to take place.

6.3.2 Recognition of pyrimidine-dimer lesions by UVDE

Strikingly, the two bases opposite to the damage (nucleotides A8 and A9) are flipped from the DNA helix into a "undamaged-bases-binding" pocket of UVDE (Figure 6.4(a)). In this pocket, the conserved Tyr104 residue of the probing fin-

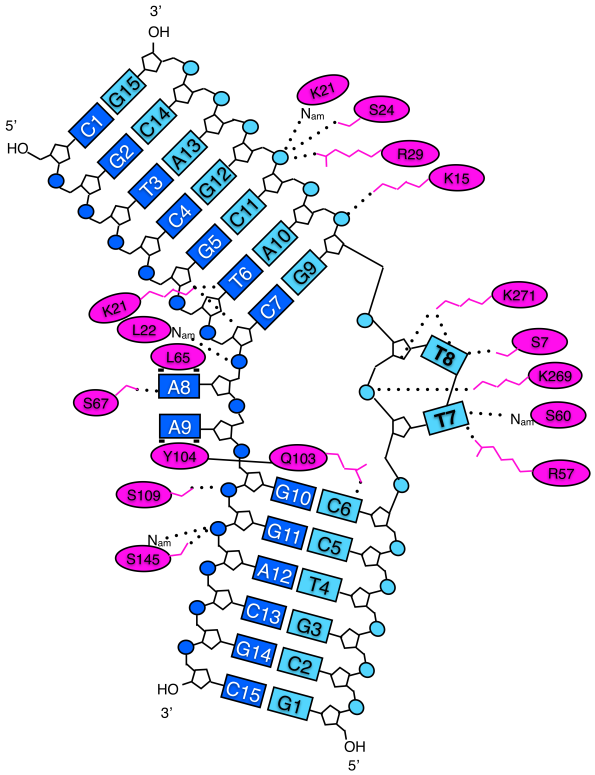

Figure 6.3: Schematic representation of UVDE-DNA interactions with the undamaged DNA strand in blue, the damaged strand in cyan and the protein in magenta. In three dimensions, the damaged bases T7 and T8 and the undamaged bases A8 and A9 are actually below the plane of the figure and the residues Q103 and Y104 are inserting in the helix.

ger forms stacking interactions with the base of A9. Leu65 aids in creating a hydrophobic environment on the other side of this pocket near A8. In other UVDEs, usually a leucine or phenylalanine is found at this position and in a superposition of *Tth*UVDE with the DNA:protein complex structure, this phenylalanine indeed forms stacking interaction with the base of A8. A hydrogen bond between the conserved Ser67 and the N1 of the A8 base is present that would be lost if a pyrimidine was present in this position due to its smaller size. The size of the pocket together with these stacking and hydrogen bond interactions make that the pocket is customized to fit two purines. Such a specific pocket for two bases of the undamaged DNA strand seems to be a novel property for a DNA repair enzyme. The existence of this pocket is a ingenious, useful feature for an enzyme that needs to recognize UV-damaged DNA, since dipyrimidines always have two purines opposite the lesion. It also gives an explanation for the strong preference of UVDE for incising abasic sites flanked by a pyrimidine over those flanked by a purine (Pas-

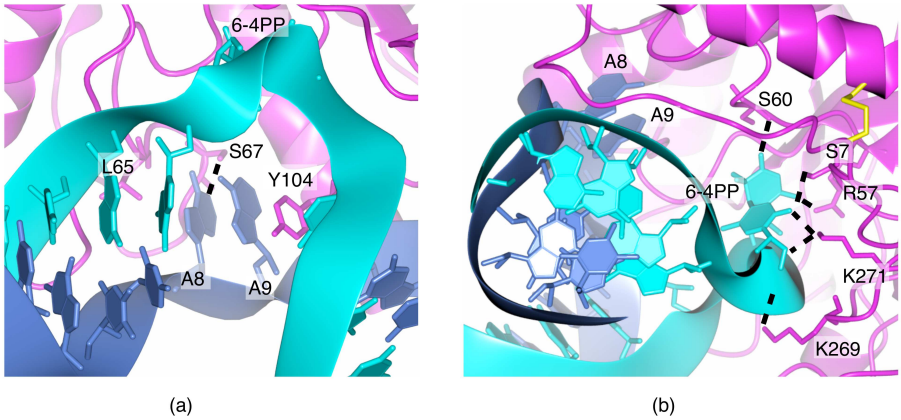

Figure 6.4: Detailed view of the pockets for the undamaged bases (a) and the damaged bases (b), showing their environment in the pockets. Hydrogen bonds are indicated with dashed lines.

paleva *et al.*, 2009), since the abasic sites flanked by a pyrimidine used in (Paspaleva *et al.*, 2009) had two purines opposite the damage. The importance of the pocket for the enzyme's function has been previously shown, since the mutation of the Tyr104 residue to alanine in *Tth*UVDE abolishes its incision activity (Paspaleva, Thomassen *et al.*, 2007). Moreover, flipping of the bases opposite to the damage into a pocket concurs with the previously reported fluorescence studies (Paspaleva *et al.*, 2009). In that study, it was noticed that the bases opposite to the damage are flipped from the helix into a partially to not solvent-exposed area, which agrees with bases flipped into a pocket.

Not only are the undamaged bases flipped, but the damage itself is also flipped into a protein pocket. A hydrogen bond by the probing finger residue Gln103 to the base just 5′ to the damage probably helps in stabilizing this flipped conformation (Figure 6.3). The damage pocket is lined with several residues capable of making hydrogen bonds to the damage (shown in Figure 6.3 and Figure 6.4(b)) and is close to the active site responsible for the enzyme's catalytic function. The pocket is tight around the 6-4PP, but the shape does allow a CPD to fit and be stabilized, while purines would be excluded from the pocket due to their larger size. To enter the pocket, the DNA backbone must deform substantially. The loss of base-pairing and amount of deformation that an undamaged DNA substrate would have to undergo is probably too energetically unfavorable to allow an undamaged DNA substrate to enter the pocket. This explains the broad specificity of the enzyme for many damaged types of DNA, while not cutting undamaged DNA.

6.3.3 *Sac*UVDE has a preference for 6-4 photoproduct over CPD damaged DNA

Although the structure of the apoprotein shows a high structural similarity to *Tth*UVDE, *in vitro* assays surprisingly showed that *Sac*UVDE has a very high in-

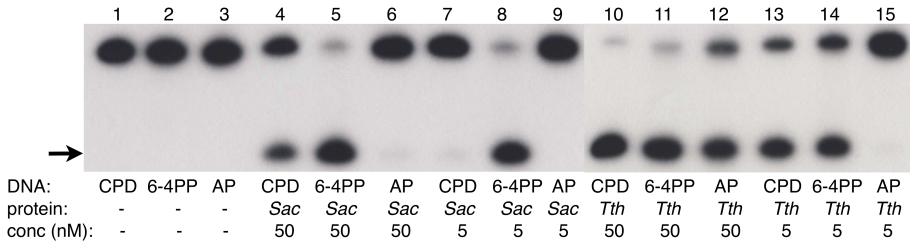

Figure 6.5: Activity of SacUVDE versus TthUVDE.
Incision assay with SacUVDE or TthUVDE (indicated below the lanes) showing that SacUVDE has a strong preference for incising 6-4PP compared to CPD in contrast to TthUVDE. The incision product is indicated with an arrow. The assay was carried out at 55 °C for both proteins.

Table 6.3: DNA binding by SacUVDE wildtype.

DNA lesion	**% binding**
CPD	8.6 ± 4
6-4PP	42.9 ± 2
Abasic site	8.1 ± 1
No damage	8.3 ± 7

cision activity for the 6-4PP, even higher than that of *Tth*UVDE, but it is low for CPD and virtually absent for abasic sites (Figure 6.5 lanes 4 to 9). This is in contrast to *Tth*UVDE that has a high incision activity for both CPD and 6-4PP and a lower incision activity for abasic sites (Figure 6.5 lanes 10 to 15). We set out to compare the structure and function of these two homologues to explain this difference and, by extending this explanation, to get insight into the generally broad substrate specificity of UVDE.

We tested whether the different substrate specificity is caused by experimental conditions, such as DNA length, pH, metal cofactor and temperature (results not shown). Much higher incision on 6-4PP than on CPD has been confirmed on 50-mer DNA substrates at both 55 °C (the standard temperature for our thermophilic incision assays) and 80 °C. The optimum activity is at pH 5.5 to 6.5 (range of 3.5 to 8.5 attempted), which is not surprising since the cellular pH is 6.5 in *S. acidocaldarius* (Moll & Schaefer, 1988). The optimal cofactor is manganese or cobalt; absence of metal ions, calcium, zinc and nickel gives little to no activity, and the addition of magnesium lowers the activity slightly. At all pHs and with all cofactors tested, the activity is always much higher for the 6-4PP than for the CPD, confirming that our result is a genuine characteristic of the protein and is not an artifact caused by experimental conditions.

Bandshift assays were performed on 6-4PP and CPD substrates to determine if the difference in activity is caused by binding or by catalysis. As can be seen in

Figure 6.8 lanes 3 and 4, binding to 6-4PP seems stronger than to the CPD. Both for CPD and 6-4PP the amount of free DNA can be seen to decrease, so the protein binds to both, but only on 6-4PP it gives a complex that is stable and specific enough to see a tight band on the gel. To confirm the results of the bandshift (stronger binding to 6-4PP than CPD), filter-binding assays were performed. As can be seen in Table 6.3, the filter-binding assays show that *Sac*UVDE indeed binds stronger to 6-4PP than to CPD. We therefore conclude that the difference in activity is caused by a difference in binding to the damage.

6.3.4 Structure-based mutagenesis explains *Sac*UVDE's substrate specificity

To explain *Sac*UVDE's preference for binding and incising 6-4PP, we took a closer look at the structural differences between *Sac*UVDE and *Tth*UVDE. In contrast to *Tth*UVDE, *Sac*UVDE has an extra helix (α8) at its C-terminus. There is a shift of around 2 Å in the neighboring α-helix of residues 248 to 260 (α7) and a shift of up to 3.3 Å in the loop of residues 235 to 242 (see Figure 6.6(a)). This latter loop is positively charged in *Tth*UVDE and is close to the DNA groove in the UVDE-DNA structure (Figure 6.9(b)), hence a shift of this loop might cause differences in DNA binding. To test the function of the α8-helix, we made a C-terminal truncation (*Sac*UVDE 1-275). Hardly any overexpression of soluble protein was detected for this construct, indicating that the α8-helix is important for stability in *Sac*UVDE. Moreover, other UVDEs like *Sp*UVDE and *Bs*UVDE have this additional helix as well, yet they can recognize both CPD and 6-4PP. It is therefore unlikely that this helix is the main reason for the difference.

Another reason for the different phenotypes could be that the the damage-binding pockets differ (see Figure 6.6(b)). However, the damage-binding pockets of both enzymes are very similar. The most prominent side-chain difference in the pocket is the Tyr10 in *Sac*UVDE (versus leucine in *Tth*UVDE and valine, methionine or threonine in most other UVDEs). A tyrosine at this position might favor binding of 6-4PP over CPD, because it might have some specific interactions with 6-4PP (the oxygen atom in the deoxyribose ring of base 8 will come into hydrogen bonding distance of the hydroxyl group of Tyr10 when metal ions pull down the lesions deeper into pocket). To test this, we mutated Tyr10 to alanine and tested its activity. As can be seen in Figure 6.7, the activity of the *Sac*UVDE Y10A mutant is very similar to wildtype, with high incision at the 6-4PP and hardly any incision at the CPD and abasic site (compare lanes 4-6 to lanes 11-13). The binding to either substrate is also comparable to wildtype (Figure 6.8). Hence we conclude that the different phenotype is most probably not due to side-chain differences in the damage-binding pocket.

Another remarkable difference between the structures *Sac*UVDE and *Tth*UVDE is a backbone shift (up to 2.8 Å) of residues 10-18 in *Sac*UVDE (Figure 6.6(c)), which is located near the damage-binding pocket. This shift is most probably caused by the disulfide bridge between C14 and C40 in *Sac*UVDE that is not present in *Tth*UVDE nor most other UVDEs. To study the importance of the disulfide bridge, we performed activity assays in the presence of reducing agents (5 - 250 mM β-

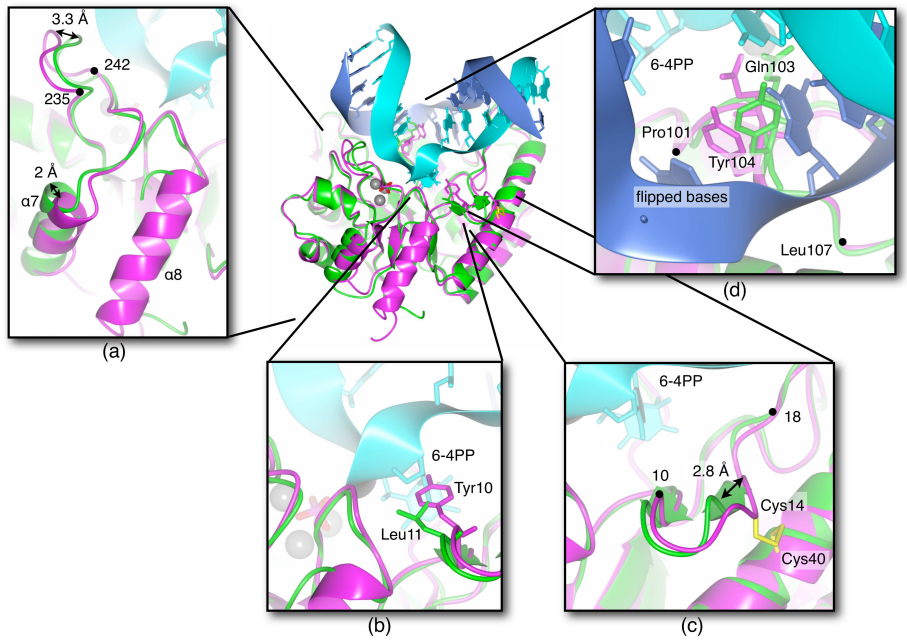

Figure 6.6: Most notable differences between structures of SacUVDE and TthUVDE. Superposition is shown with SacUVDE depicted in magenta, TthUVDE in green, the metal ions in grey, undamaged DNA in blue and damaged DNA in cyan. The following features can be seen. (a) The presence of α8-helix in SacUVDE causes a shift of the α7-helix and the 235-242 loop. (b) The substrate-binding pocket with residue Tyr10 in SacUVDE and Leu11 in TthUVDE. (c) The disulfide bridge C14-C40 causes a shift of residues 10-18 near the damage binding pocket. (d)The probing finger loop of UVDE starting at Pro101 and ending at Leu107. The probing finger itself (Gln103 and Tyr104) is also indicated. The numbering of SacUVDE is used.

mercaptoethanol or DTT; results not shown). No difference in activity could be observed even at the highest reducing agent concentration. Since there is 5 mM DTT present in the crystallization buffer and still the disulfide bridge was seen in the crystal structure, it is likely that the disulfide bridge is not easily broken. During the assay in presence of reducing agents this might also have been the case.

To disrupt the disulfide bridge with certainty and hence be able to verify its importance, we mutated C14 to alanine and tested its activity. Figure 6.7 shows that *Sac*UVDE C14A (lanes 14-16) has a strongly reduced activity compared to *Sac*UVDE wt (lanes 4-6). This can be seen more clearly in an assay with reduced concentrations of protein, which shows a much lower incision activity of *Sac*UVDE C14A (lanes 20-26) compared to wildtype (lanes 4-10). At all temperatures tested (35 °C, 45 °C and 55 °C), the relative difference between mutant and wildtype stays the same (results not shown). The solubility of the mutant during purification is comparable to wildtype. Based on these two observations, we conclude that the re-

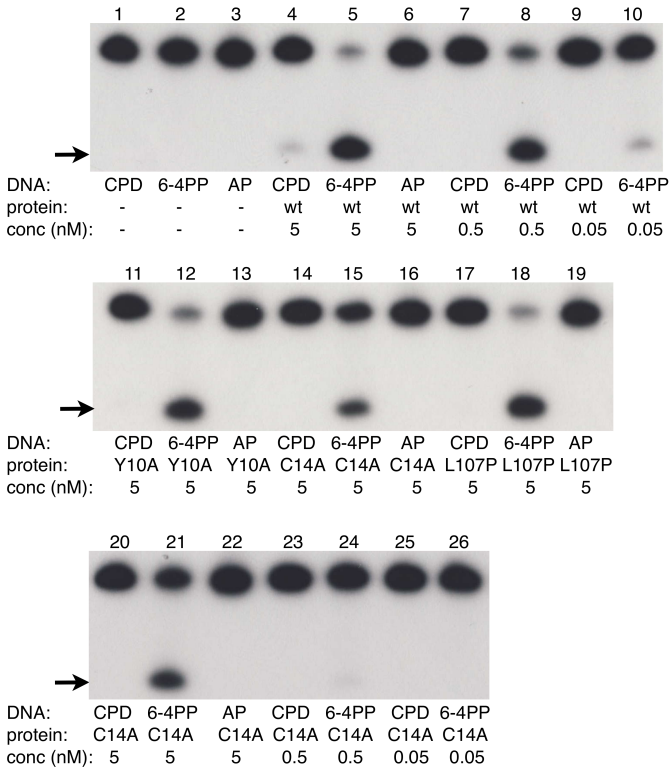

Figure 6.7: Activity of mutants of SacUVDE.
Lanes 1-10: Incision activity of SacUVDE wt at concentrations of 0.05 to 5 nM showing a strong preference for incision on 6-4PP. The incision product is indicated with an arrow.
Lanes 11-19: Incision activity of SacUVDE mutants showing that mutants Y10A and L107P have a similar activity to SacUVDE wt and that SacUVDE C14A has a strongly reduced activity.
Lanes 20-26: Incision activity of SacUVDE C14A in more detail confirming the strongly reduced activity of this mutant.

duced activity is not caused by reduced stability during the time of the assay, but by a genuine defect in the protein's function. We tested DNA binding of this mutant by a bandshift assay (Figure 6.8), which also shows a reduced affinity, suggesting that mutant C14A has a defect in binding to damaged DNA. This result was confirmed by filter-binding assays (results not shown). It is possible that the rigidity caused by the disulfide bridge prevents *Sac*UVDE to incise efficiently both CPD and 6-4PP, but that removing this disulfide bridge leads to a pocket that is too flexible to cut anything efficiently.

An important factor for DNA recognition and binding is the probing finger of the residues Gln103 and Tyr104 in *Sac*UVDE (Gln104 and Tyr105 in *Tth*UVDE; Figure 6.6(d)), which helps in flipping out the damaged bases and the opposite bases from the double DNA helix. In *Tth*UVDE, the loop towards this finger is more

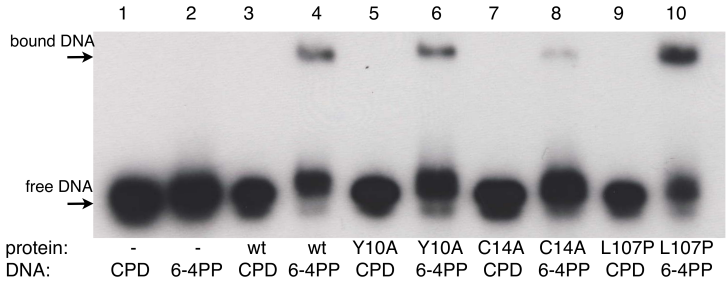

Figure 6.8: Bandshift assay of SacUVDE wt and mutants Y10A, C14A and L107P, showing all proteins have a strong preference of binding 6-4PP to CPD and reduced binding of mutant C14A. Protein and DNA substrates used are indicated below the lane.

rigid, since it has a proline at the beginning and the end of it (Pro102 and Pro108), while *Sac*UVDE only has one proline at the corresponding positions (Pro101 and Leu107). A more rigid finger loop might be more suited to flip out less distorted DNA substrates and hence a protein with such a rigid finger might have a broader substrate range. To test this hypothesis, we mutated Leu107 in *Sac*UVDE to a proline. The resulting protein was poorly soluble (most of the protein was in the pellet after lysis and centrifugation), but the soluble part could still be purified to near homogeneity. As can be seen in Figure 6.7 the mutant protein (lanes 17-19) has an incision activity and substrate specificity comparable to wildtype (lanes 4-6): it only incises 6-4PP efficiently. Also binding studies show the same pattern (see Figure 6.8), with a large preference for 6-4PP over CPD. The binding efficiency of L107P looks slightly higher than wildtype, but this is probably due to inaccurate concentration determination of the mutant due to its low concentration caused by problems in its solubility. Taken together the above results, we deem it unlikely that the difference in rigidity of the probing finger causes the different phenotypes of *Sac*UVDE and other UVDEs.

A charge surface plot of both *Sac*UVDE and *Tth*UVDE (Figure 6.9) shows a difference in the positive charges on the end of the DNA-binding groove that insert into the DNA helix, which have previously been hypothesized to be involved in DNA binding (Paspaleva *et al.*, 2007). For *Tth*UVDE there is a stretch of three positively charged residues (Lys238, Lys239 and Arg240) and another lysine close by (Lys179) on one side of the groove, while *Sac*UVDE only has two (Lys178, Lys239). On the other side of the groove, *Tth*UVDE also has more positive charges: four (Arg22, His25, Arg30, Lys34), while *Sac*UVDE has only two (Lys21, Arg29). Having a prominent positive charge on both sides of the groove likely helps in fixing and bending the DNA and might explain the different phenotype of *Sac*UVDE and *Tth*UVDE. To verify the importance of this positive charge at both sides of the DNA-binding groove, we mutated residues Gly237 and Glu238 to lysine in *Sac*UVDE. To test its activity in detail, we performed a kinetics incision assay (shown in Figure 6.10). In this assay it can be seen that the difference in incision activity at CPD and

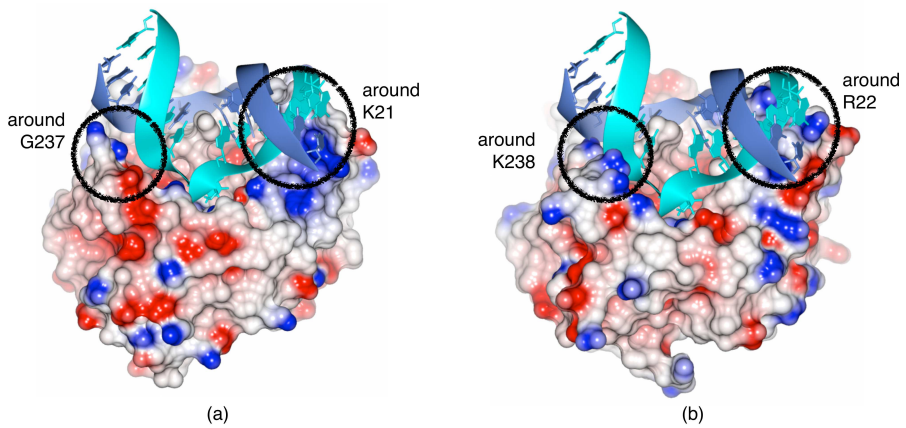

Figure 6.9: Charge surface representations of SacUVDE and TthUVDE with positive charge in blue and negative charge in red. DNA is depicted in cyan (damaged strand) and blue (undamaged strand).
(a) SacUVDE; regions corresponding to that with more positive charge in TthUVDE are shown in circles with one residue number for orientation.
(b) TthUDVE; regions which have more positive charge shown in circles.

6-4PP has become smaller in the G237K + E238K mutant. This can be explained by assuming a combination of the two following effects. First, the overall activity of the protein goes down, as can be seen with the 6-4PP, due to addition of non-natural charge for this homologue that might put the DNA slightly in the wrong position. Secondly, the CPD goes up relatively to the 6-4PP, potentially due to the easier bending of the DNA. Together, these two effects can cause the result we observed: a smaller difference in the activity between the two substrates. Thus, the positive charge on both sides of the groove can be indeed an important factor in *Sac*UVDE substrate preference and explains at least in part its difference with UVDEs from other organisms.

6.4 Discussion

The pre-incision complex structure has provided confirmation of previously proposed elements of the mechanism of UVDE, such as the DNA binding groove and the 90° bend in the DNA. It also presents a novel mechanism for recognizing damaged dipyrimidines by a dual flipping mechanism. UVDE flips the undamaged bases opposite to the damage (always two purines, since the damage consists of a dipyrimidine) in a dipurine-specific pocket. Not only the undamaged bases are flipped, but also the damaged ones are flipped in a protein pocket. This combination represents an elegant mechanism for recognizing and verifying the presence of the lesion. First, the flipping probably can only occur in damaged DNA, since the presence of the lesion causes a distortion in the DNA (such as the loss of hydro-

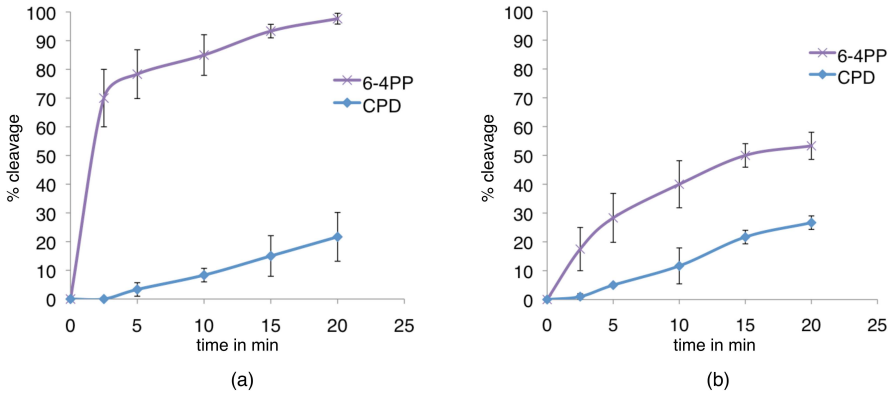

Figure 6.10: Activity of SacUVDE wt and SacUVDE G237K E238K.
(a) Kinetics incision assay with SacUVDE wildtype, showing a strong difference between incision on CPD and 6-4PP.
(b) Kinetics incision assay with SacUVDE G237K E238K, showing a smaller difference between incision on CPD and 6-4PP.

gen bonding). It is therefore highly likely that flipping is easier in damaged DNA. Second, dipurines are stabilized in the undamaged-DNA-binding pocket, providing a mechanism for preferentially repairing distorted dipyrimidines. Thirdly, the damaged bases are flipped into a pocket that can accommodate damaged dipyrimidines very well, but will exclude larger lesions. Smaller lesions such as abasic sites can be incised by UVDE, albeit with a lower efficiency, perhaps because the lesion is too small to be held stably in the damage pocket. And last, the DNA backbone is rather distorted at the point where the damaged bases are flipped, which likely helps in discriminating between damaged and undamaged DNA: for undamaged DNA such a distortion would be energetically quite unfavorable.

To gain more insight into the broad specificity of UVDE, we have compared the phenotype and structure of the more general *Tth*UVDE to the more specific *Sac*UVDE. The latter only efficiently incises 6-4PPs, the UV-lesion that causes severe distortion in the DNA backbone (a 44° kink; Kim & Choi, 1995). Surprisingly, *Sac*UVDE's structure was found to be very similar to the structure of *Tth*UVDE, which also incises CPDs well and abasic sites moderately well. We demonstrated that the different phenotype is not caused by a more specific damage-binding pocket nor by less rigidity of the probing finger. However, the disulfide bridge in *Sac*UVDE might have a role in this. Removing this disulfide bridge in *Sac*UVDE has a detrimental effect on the protein's ability to bind and incise damaged DNA. This could be explained as follows. Rigidity around the binding pocket potentially caused by the disulfide bridge in *Sac*UVDE might prevent the protein from binding to both CPD and 6-4PP. This non-flexible pocket might be shaped to only fit well, selectively, 6-4PP. Removing the disulfide bridge might then result in a protein with a binding pocket too flexible to bind anything stably, as we observed in

our biochemical assays. The disulfide bridge might have arisen in evolution to yield a more thermoresistant protein for this thermophilic organism, which has a 10 to 15 °C higher optimum growth temperature than *Thermus thermophilus*). Perhaps as trade-off for the extra stability, it had to give up the broader substrate specificity typical for UVDE. Potentially it could do this, because the organism already has a photolyase for repairing CPDs. This scenario is feasible, since a homologous protein to the potential ORF for *S. acidocaldarius* photolyase has been shown to reverse CPDs *in vitro* (Fujihashi *et al.*, 2007).

Less positive charge on both sides of the DNA binding groove might also make binding a broad range of DNA substrates more difficult for *Sac*UVDE, since this charge likely assists in bending the DNA; a 90° bend in the DNA is needed for optimal binding. The 6-4PP is already distorted more than the CPD in solution, explaining why the protein can bind 6-4PP without the charge on both sides of the groove. Indeed, addition of this charge in *Sac*UVDE decreases the difference in incision efficiency between CPD and 6-4PP. This could be taken as a confirmation of this hypothesis. In conclusion, we think that the difference in substrate specificity between *Tth*UVDE and *Sac*UVDE is caused by a combination of effects, such as the rigidity caused by the disulfide bridge and more difficulty in bending DNA caused by having less positive charges on both sides of the groove.

Comparison of the very specific *Sac*UVDE to the more broad *Tth*UVDE gives insight into the relatively broad substrate specificity of UVDE in general. It seems that flexibility around the binding pocket and capacity of bending the DNA are important factors for UVDE's capacity to recognize and incise different DNA lesions. *Sac*UVDE might have given up these features to obtain a more stable protein for living at extraordinarily high temperatures. Though *Sac*UVDE perhaps plays a lesser role in DNA repair in *S. acidocaldarius*, it may be useful for biochemical assays to distinguish between CPD and 6-4PP, since it is, to the authors' knowledge, the most specific enzyme for incision only next to 6-4PPs.

Taking together the above results, we can make the following model for UVDE activity. First, UVDE recognizes a distortion in the DNA, such as a kink and/ or loss of hydrogen bonding, and binds to it. The residues of the probing finger (Gln103 and Tyr104) flip the damaged bases as well as the opposite bases out of the helix into their respective pockets. Only if the opposite bases are two purines (so they fit well in their pocket) and if the damaged bases fit in the damage pocket, does the enzyme bind stably to the damage. The positive charge of the metal ions then draws in the scissile phosphodiester bond and incision by a hydroxyl ion takes place using three metal-ion mediated catalysis.