The EMBO Journal (2005) 24, 885–894 | © 2005 European Molecular Biology Organization | All Rights Reserved 0261-4189/05
www.embojournal.org

THE EMBO JOURNAL

Structural insights into the first incision reaction during nucleotide excision repair

James J Truglio[1], Benjamin Rhau[1], Deborah L Croteau[2], Liqun Wang[1], Milan Skorvaga[2,3], Erkan Karakas[1], Matthew J DellaVecchia[2], Hong Wang[2], Bennett Van Houten[2] and Caroline Kisker[1,*]

[1]Department of Pharmacological Sciences, State University of New York at Stony Brook, Stony Brook, NY, USA, [2]Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA and [3]Department of Molecular Genetics, Cancer Research Institute, Slovak Academy of Sciences, Bratislava, Slovakia

Nucleotide excision repair is a highly conserved DNA repair mechanism present in all kingdoms of life. The incision reaction is a critical step for damage removal and is accomplished by the UvrC protein in eubacteria. No structural information is so far available for the 3′ incision reaction. Here we report the crystal structure of the N-terminal catalytic domain of UvrC at 1.5 Å resolution, which catalyzes the 3′ incision reaction and shares homology with the catalytic domain of the GIY-YIG family of intron-encoded homing endonucleases. The structure reveals a patch of highly conserved residues surrounding a catalytic magnesium-water cluster, suggesting that the metal binding site is an essential feature of UvrC and all GIY-YIG endonuclease domains. Structural and biochemical data strongly suggest that the N-terminal endonuclease domain of UvrC utilizes a novel one-metal mechanism to cleave the phosphodiester bond.

The EMBO Journal (2005) **24,** 885–894. doi:10.1038/sj.emboj.7600568; Published online 3 February 2005
Subject Categories: structural biology; genomic stability & dynamics
Keywords: crystallography; DNA damage; DNA repair; nucleotide excision repair; UvrC

Introduction

Nucleotide excision repair (NER) stands apart from other DNA repair mechanisms available to the cell in its ability to recognize a broad range of structurally unrelated DNA damages (Van Houten, 1990; Friedberg *et al*, 1995; Lloyd and Van Houten, 1995; Sancar, 1996; Goosen and Moolenaar, 2001) including carcinogenic cyclobutane pyrimidine dimers induced by UV radiation, benzo[a]pyrene-guanine adducts caused by smoking and burning of fossil fuels, and guanine-cisplatinum adducts formed during cancer chemotherapy (Sancar, 1994). The strategy employed by NER is the same in all three kingdoms of life. NER in prokaryotes was one of the first repair mechanisms discovered (Boyce and Howard-Flanders, 1964; Setlow and Carrier, 1964) and is mediated by the UvrA, UvrB and UvrC proteins. These three proteins recognize and cleave damaged DNA in an ATP-dependent multistep reaction. UvrA is involved in damage recognition and either forms a heterotrimeric ($UvrA_2UvrB$) (reviewed in Theis *et al*, 2000) or heterotetrameric ($UvrA_2UvrB_2$) (Verhoeven *et al*, 2002) complex with UvrB. This complex is believed to scan the DNA helix for conformational perturbations induced by DNA lesions (Theis *et al*, 2000). After the damage has been identified, UvrA dissociates from the protein–DNA complex, leaving UvrB bound to the DNA (Orren and Sancar, 1990) forming a stable preincision complex (Theis *et al*, 2000; Skorvaga *et al*, 2002). UvrC binds to this complex and mediates the incision three or four nucleotides 3′ to the damaged site, followed by a second incision seven nucleotides 5′ to the damaged site (Sancar and Rupp, 1983; Lin and Sancar, 1992a, b; Verhoeven *et al*, 2000). UvrD (helicase II) and DNA polymerase I (polI) are required for turnover of the UvrABC proteins (Caron *et al*, 1985; Husain *et al*, 1985). UvrD removes both UvrC and the oligonucleotide containing the lesion, while UvrB remains bound to the gapped DNA until it is displaced by DNA polI (Orren *et al*, 1992). The reaction is completed by DNA ligase, which closes the nicked DNA. This multistep process of DNA recognition and repair ensures a high degree of discrimination between the damaged and nondamaged strand.

Site-directed mutagenesis and sequence alignments have shown that UvrC catalyzes both the 3′ and 5′ incisions and each of these incisions is performed by a distinct catalytic site that can be inactivated independently (Lin and Sancar, 1992b; Verhoeven *et al*, 2000). The domain responsible for 3′ incision is located in the N-terminal half of the molecule and consists of approximately the first hundred residues. This domain shares limited homology with a small module found in members of the GIY-YIG endonuclease family (Aravind *et al*, 1999). Also included in the N-terminal half is a region that interacts with the C-terminal domain of UvrB (Aravind *et al*, 1999). The 5′ catalytic domain, which is distantly related to *Escherichia coli* endonuclease V, is located in the C-terminal half of the protein along with two helix-hairpin-helix motifs employed in DNA binding (Aravind *et al*, 1999). After recruitment to the UvrB:DNA preincision complex, UvrC first catalyzes cleavage of the DNA on the 3′ side of the lesion (Verhoeven *et al*, 2000). This incision requires the interaction between the C-terminal domain of UvrB and the homologous UvrB binding domain of UvrC (Moolenaar *et al*, 1995, 1998a), which is not required for 5′ incision (Moolenaar *et al*, 1995).

In order to obtain a better understanding of the 3′ incision event, we have solved the crystal structure of the N-terminal

*Corresponding author. Department of Pharmacological Sciences, State University of New York at Stony Brook, Stony Brook, NY 11794-5115, USA. Tel.: +1 631 632 1465; Fax: +1 631 632 1555;
E-mail: kisker@pharm.sunysb.edu

Received: 3 December 2004; accepted: 7 January 2005; published online: 3 February 2005

endonuclease domain of UvrC from two different thermophilic organisms, *Bacillus caldotenax* and *Thermotoga maritima*, at 2.0 and 1.5 Å resolution, respectively. This domain shares structural and sequence similarity to the catalytic domain found in I-*Tev*I, a GIY-YIG homing endonuclease (Van Roey *et al*, 2002). I-*Tev*I is one of at least 60 known GIY-YIG endonuclease family members that are present in bacteriophage T4, bacteria, archaea, algal chloroplasts and mitochondria, and fungal mitochondria. Members of this family are characterized by a 70–100 residues long module containing a conserved GIY-(X_{9-11})-YIG motif.

We have identified a patch of highly conserved residues on the surface of the N-terminal domain of UvrC to which a single divalent cation is bound. The residues that form the metal binding pocket are conserved throughout all GIY-YIG endonucleases, suggesting the site to be a common feature of all family members. We mutated seven amino acids within the conserved patch and analyzed whether full-length UvrC was still able to incise damaged DNA, in the complete UvrABC reaction. Combined with the structural data, the results indicate that the conserved patch is the active site and the bound divalent cation is the catalytic metal. Based on our data, we propose that UvrC uses a novel one-metal mechanism to catalyze cleavage of the fourth or fifth phosphodiester bond 3′ to the DNA lesion.

Results

Crystal structure of the N-terminal endonuclease domain of UvrC

The N-terminal catalytic domain of UvrC was initially cloned from *B. caldotenax* ($UvrC^{N\text{-}Bca}$; residues 1–98) and its structure was solved by multiwavelength anomalous diffraction (MAD) (Table I). The protein crystallized in space group C2 and contained four molecules in the asymmetric unit arranged as a tetramer with C_4 symmetry. Each of the four subunits has nearly identical conformations, with an average root mean square (r.m.s.) deviation of 0.62 Å for the Cα atoms of residues 8–94. The structure was refined at 2.0 Å resolution to an *R*-factor of 0.203 and R_{free} of 0.252 (Table I). Residues 95–98 of subunit A, 1–8 and 95–98 of subunit B, 95–98 of subunit C, as well as 1 and 96–98 of subunit D are disordered.

The corresponding domain was also cloned from *T. maritima* ($UvrC^{N\text{-}Tma}$; residues 1–97) and crystallized in space group $P4_32_12$ containing one molecule in the asymmetric unit. Crystals were soaked with manganese or magnesium chloride. The structure of $UvrC^{N\text{-}Tma}$ bound to manganese was solved by molecular replacement using the $UvrC^{N\text{-}Bca}$ monomer as a search model (see Materials and methods). This structure was refined at 1.5 Å resolution to an *R*-factor of 0.167 and R_{free} of 0.185 (Table I), and consists of residues 1–89. The structure of magnesium-bound $UvrC^{N\text{-}Tma}$ was solved using difference Fourier methods (Table I). For the remainder of the discussion, residue numbering will correspond to UvrC from *T. maritima*, unless specified otherwise.

The *B. caldotenax* and *T. maritima* structures have similar conformations with an r.m.s. deviation of 1.55 Å for 88 Cα atoms. Both have an αββααβα(α) topology with the C-terminal α-helix (α5) not present in $UvrC^{N\text{-}Tma}$. The core of the molecule is a three-stranded β-sheet, which is flanked by the first three helices on one side and helix 4 on the other (Figure 1). Helices α1, α2 and α4 run approximately parallel to the β-sheet, while helix α3 is positioned almost perpendicular to these secondary structural elements, making contact only with the bottom edge of the sheet. In $UvrC^{N\text{-}Bca}$ helix α3

Table I Crystallographic statistics

Data set	Native (Bca)	SeMet peak (Bca)	SeMet inflection (Bca)	SeMet remote (Bca)	Native (Tma)	Mn^{2+} bound native (Tma)	Mg^{2+} bound native (Tma)
Resolution (Å)	2.0	2.0	2.0	2.0	1.8	1.5	1.8
Wavelength (Å)	1.0	0.9794	0.9798	0.9538	1.1	1.1	1.1
Unique reflections	29 029	56 919	56 982	56 947	16 312	27 459	16 305
$\langle I\rangle/\langle\sigma I\rangle$	37.2 (5.58)	31.1 (2.74)	31.6 (3.36)	31.6 (3.34)	47.5 (4.88)	40.5 (4.57)	35.8 (4.02)
Completeness (%)	100.0 (100.0)	98.2 (97.6)	98.2 (97.6)	98.2 (97.6)	99.9 (100.0)	98.4	99.7
R_{sym}	0.07 (0.43)	0.07 (0.70)	0.06 (0.53)	0.06 (0.53)	0.06 (0.60)	0.04 (0.31)	0.06 (0.52)
Phasing to 3.0 Å							
FOM (from SOLVE)			0.78 (0.68)				
Map correlation			0.45				
Mean phase difference (deg)			55.0 (60.7)				
R_{cryst} (R_{free})	0.203 (0.252)				0.185 (0.199)	0.167 (0.185)	0.175 (0.202)
r.m.s. deviation bond lengths (Å)	0.014				0.017	0.013	0.017
r.m.s. deviation bond angles (deg)	1.5				1.4	1.5	1.5
Mean *B*-factor	36.1				26.0	24.8	21.8
Ramachandran Statistics	94.3/4.8/0.9/0.0				94.9/3.8/1.3/0.0	94.9/5.1/0.0/0.0	96.2/2.6/1.3/0.0

Tma and Bca refer to *T. maritima* and *B. caldotenax* UvrC, respectively. $R_{sym} = \sum_{hkl}\sum_i|I_i - \langle I\rangle|/\sum_{hkl}\sum_i\langle I\rangle$, where I_i is the *i*th measurement and $\langle I\rangle$ is the weighted mean of all measurements of *I*. $\langle I\rangle/\langle\sigma I\rangle$ indicates the average of the intensity divided by its average standard deviation. Numbers in parentheses refer to the respective highest resolution data shell in each data set. $R_{cryst} = \sum||F_o|-|F_c||/\sum|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes. R_{free} is same as R_{cryst} for 5% of the data randomly omitted from the refinement. The map correlation coefficient describes the correlation between the electron density map calculated from the final model and the map corresponding to the experimental set of phases, averaged over all grid points. The mean phase difference is the mean differences between the initial phases calculated from SOLVE and phases calculated from the final wild-type model. Ramachandran statistics indicates the fraction of residues in the most favored, additionally allowed, generously allowed and disallowed regions of the Ramachandran diagram, as defined by the program PROCHECK (Laskowski *et al*, 1993).

also interacts with the C-terminus, connecting the two helical halves of the protein across the β-sheet.

Comparison to the catalytic domain of I-TevI

A search using DALI (Holm and Sander, 1995), a network service for comparing three-dimensional protein structures (http://www.ebi.ac.uk/dali), identified the catalytic domain of I-*Tev*I (Van Roey *et al*, 2002) (PDB codes 1LN0 and 1MK0) as the only structure with a fold similar to the N-terminal domain of UvrC (*Z*-score of 7.8). I-*Tev*I is a member of the GIY-YIG family of homing endonucleases, which in turn belong to the larger GIY-YIG superfamily that includes UvrC. The only entity in common among all superfamily members is the domain presented here. It is a small, 70–100 residues module, containing a conserved GIY-(X_{9-11})-YIG motif (Van Roey *et al*, 2002) (Gly 17, Val 18, Tyr 19–Tyr 29, Ile 30, Gly 31), four invariant residues (Gly 31, Arg 39, Glu 76, Asn 88) and two highly conserved residues (Tyr 19, Tyr 29) (Figure 1). These amino acids are all located in proximity to each other and form, in part, a highly conserved surface.

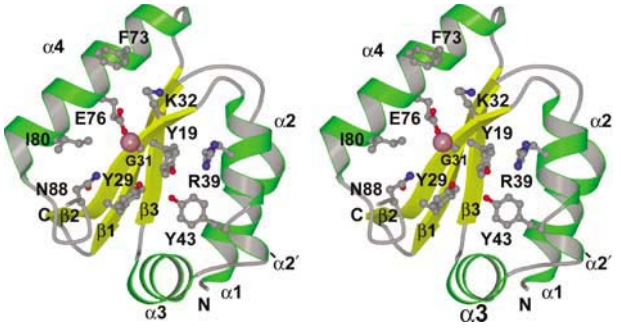

Figure 1 : Stereo view of the N-terminal endonuclease domain of $UvrC^{N\text{-}Tma}$. The central β-sheet (β1–β3) is shown in yellow and the surrounding helices (α1–α4) in green. Conserved residues are shown in ball-and-stick representation and the Mg^{2+} ion as a magenta sphere. The N- and C-termini of the domain are indicated. This figure and Figures 2A, 4 and 8 were generated with the programs MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt and Murphy, 1994).

The catalytic domains of UvrC and I-*Tev*I have likely diverged long ago as reflected by their low sequence identity of only 15% (Figure 2B). $UvrC^{N\text{-}Tma}$ superimposes onto the catalytic domain of I-*Tev*I with an r.m.s. deviation of 2.2 Å for 60 out of 89 possible Cα atoms. However, there are notable differences in secondary and tertiary structure (Figure 2A). $UvrC^{N\text{-}Tma/Bca}$ contains an additional helix, α1, as compared to I-*Tev*I. The function of α1 in UvrC is undoubtedly structural, as the residues that form this helix are not conserved among UvrCs, and sequence alignments show that like I-*Tev*I, certain UvrC proteins lack this N-terminal helix (Figure 3B). Secondly, the fragment in $UvrC^{N\text{-}Tma/Bca}$ spanning α2 and β3, which includes α3, is not structurally conserved compared to I-*Tev*I. Regardless of this dissimilarity, the position of Ile 54 from $UvrC^{N\text{-}Tma}$ (Leu 56 in $UvrC^{Bca}$) is conserved and superimposes well with Leu 45 in I-*Tev*I (Figure 2A and B). This residue is structurally important and stabilizes the hydrophobic core of the domain. Lastly, $UvrC^{N\text{-}Bca}$ contains an additional helix, α5, at its C-terminus that is replaced by a loop region in both I-*Tev*I and $UvrC^{N\text{-}Tma}$ (Figure 2A and B).

Both $UvrC^{N\text{-}Tma/Bca}$ and I-*Tev*I contain a strictly conserved glutamate, arginine and asparagine, and a pair of highly

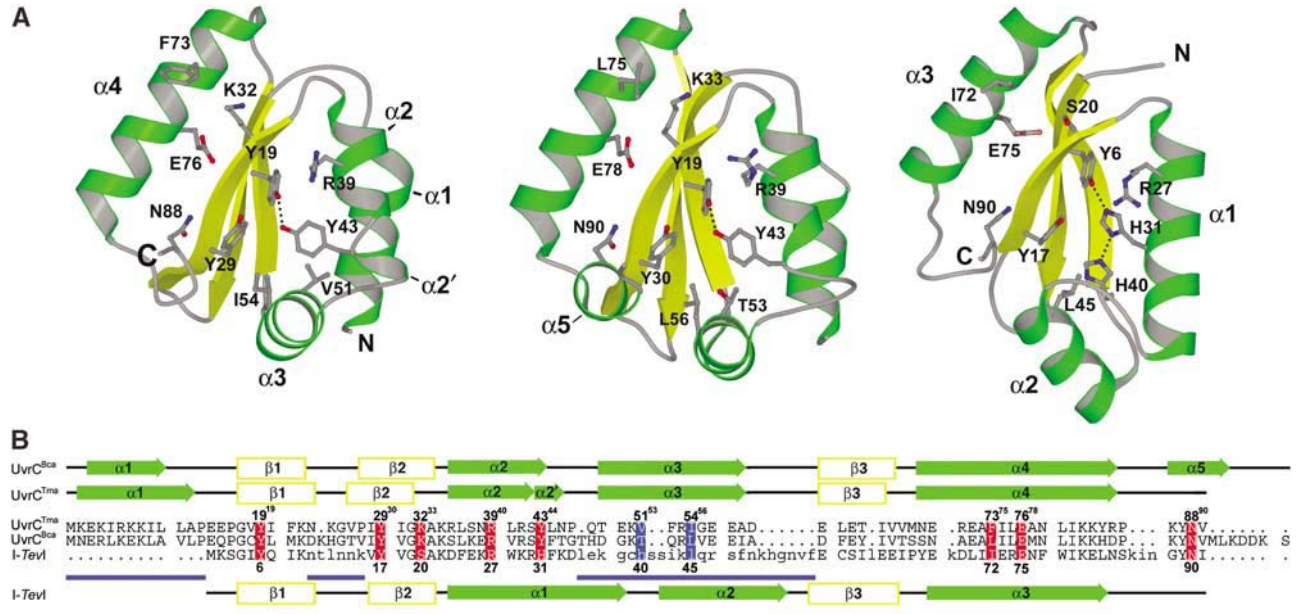

Figure 2 Structural comparison of the two 3′ endonuclease domains from UvrC and the analogous domain in I-*Tev*I. (**A**) Following their superposition, the three proteins were separated and displayed side by side: $UvrC^{N\text{-}Tma}$ (left), $UvrC^{N\text{-}Bca}$ (middle) and I-*Tev*I (right). Selected residues are shown in ball-and-stick representation; hydrogen bonds are indicated by dotted lines. (**B**) Structure-based sequence alignment of UvrC from *T. maritima* and *B. caldotenax*, and I-*Tev*I. Secondary structure elements are indicated above and below the sequence alignment corresponding to UvrC and I-*Tev*I, respectively. The blue lines below the secondary structure elements indicate large regions of structural dissimilarity. Uppercase letters indicate residues that align structurally, while lowercase letters indicate residues that are not structurally aligned. Selected, structurally aligned residues are highlighted in red. Residues in a similar position, but not structurally aligned are highlighted in blue. Numbers above the sequence alignment correspond to residue numbering in $UvrC^{N\text{-}Tma}$ and $UvrC^{N\text{-}Bca}$ with numbering from $UvrC^{N\text{-}Bca}$ in superscript. Numbers below the alignment relate to residue numbering in I-*Tev*I.

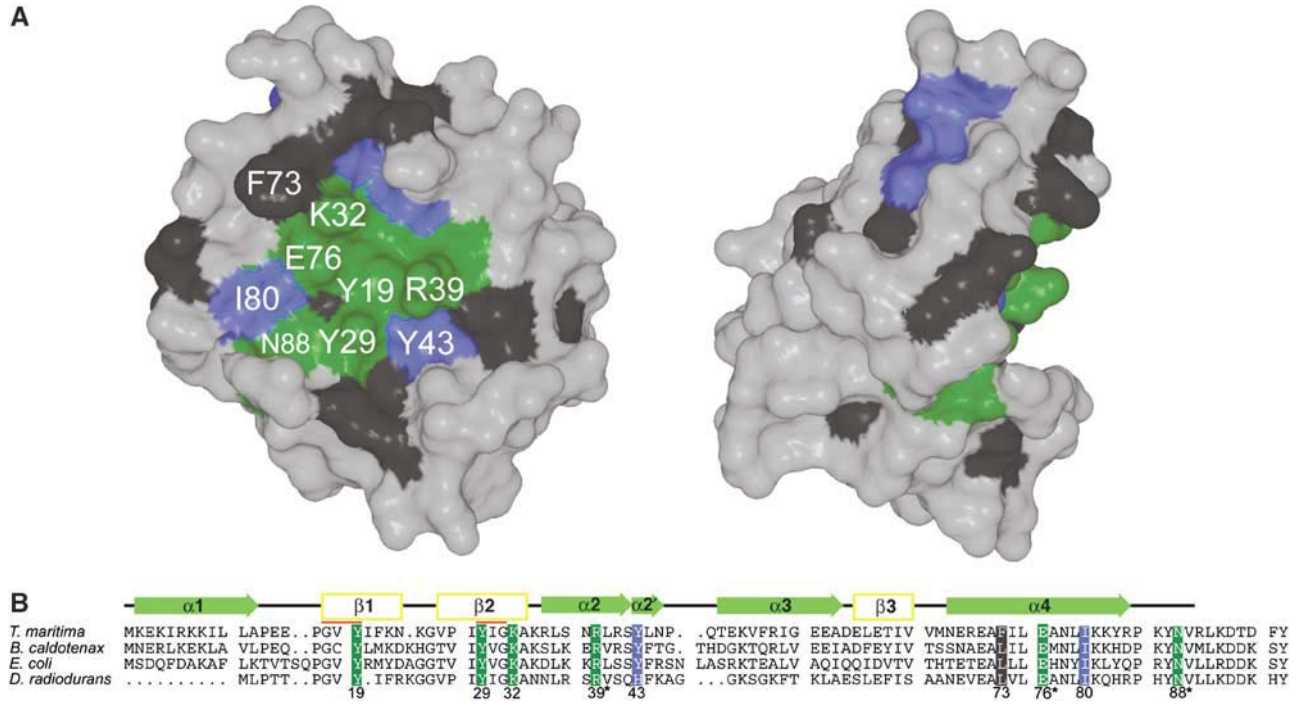

Figure 3 Sequence conservation of the 3′ endonuclease domain of UvrC. (**A**) Two different views of the N-terminal domain in surface representation. The right view is rotated 90° relative to the left view around the vertical axis. Only the most conserved residues with solvent-accessible side chains are labeled with the addition of Phe 73. Color-coding is according to conservation (green: strictly conserved; blue: highly conserved; black: moderately conserved). (**B**) Sequence alignment of the N-terminal GIY-YIG domain from four selected UvrCs. The sequences are from *T. maritima* (gi:8134799), *B. caldotenax*, *E. coli* (gi:38704033) and *Deinococcus radiodurans* (gi:6116772). Secondary structure elements are indicated above the sequence and refer to the structure of $UvrC^{N\text{-}Tma}$. They are color-coded according to Figure 1. Conserved residues are color-coded as in (A). The conserved GIY-(X_{9-11})-YIG sequence for which the domain was named is indicated by two red lines above the sequences. Part A and Figure 6 were generated with the programs SPOCK (Christopher and Baldwin, 1998) and RASTER3D.

conserved tyrosine residues (Tyr 19 and Tyr 29 in $UvrC^{Tma}$; Tyr 6 and Tyr 17 in I-*Tev*I) (Figure 2A). The third tyrosine, Tyr 43 in $UvrC^{N\text{-}Tma}$, corresponds to His 31 in I-*Tev*I (Figure 2A and B). Sequence alignments reveal that this residue is either a tyrosine or a histidine in all UvrCs (Figure 3B) and all other GIY-YIG family members. Both Tyr 43 of $UvrC^{N\text{-}Tma}$ and His 31 of I-*Tev*I form a hydrogen bond to Tyr 19 and Tyr 6, respectively (Figure 2A). His 31 of I-*Tev*I in turn forms a hydrogen bond to His 40, which occupies the position of Val 51 in UvrC. In place of a hydrogen bond, Val 51 forms van der Waals interactions with Tyr 43.

The conserved surface

Mapping the sequence conservation of UvrC proteins from different organisms onto the surface reveals a conserved patch of amino acids (Figure 3A and B). Six strictly conserved residues are located in the center on one side of the surface: Tyr 19, Tyr 29, Lys 32, Arg 39, Glu 76 and Asn 88. These strictly conserved amino acids are surrounded by a number of highly conserved residues: Tyr 43, Phe 73 and Ile 80. The conserved residues form a shallow, concave surface with dimensions of 16 Å × 15 Å and 5 Å deep, which could easily accommodate double-stranded DNA. This is in agreement with biochemical data, which have shown that the 3′ endonuclease site only recognizes double-stranded DNA as a substrate and DNA containing an unpaired region of more than eight nucleotides overlapping the 3′ incision site is not incised (Zou and Van Houten, 1999). The cleft is sufficient for nuclease activity; however, the isolated domain does not bind to DNA (data not shown) and requires either the UvrB interacting domain and/or the C-terminal helix-hairpin-helix DNA binding domain of UvrC for catalysis to occur.

The divalent cation

Phosphodiester bonds, although thermodynamically labile, require large activation energies for cleavage at physiological pH (Galburt and Stoddard, 2002). This is mainly due to the negative charge of the phosphoryl group at this pH, which repels potential attacking nucleophiles. Three chemical entities are required to catalyze efficiently the cleavage of the phosphodiester bond: a general base to position and activate the nucleophile (usually a water molecule) for inline attack of the 5′ phosphate, a general acid to protonate the 3′ leaving group and a Lewis acid to stabilize the pentacovalent phosphoanion transition state. A number of different strategies to satisfy these requirements have evolved. A common feature of most nucleases is the use of one, two or even three divalent cations in the active site to lower the free energy of the transition state (Galburt and Stoddard, 2002). Metal ions can decrease the pK_a of coordinating water molecules, resulting in a bound hydroxide, which could take the role of either a nucleophile or a general base. Alternatively, a metal-coordinated water molecule can be acidic and has the potential to serve as the general acid necessary to protonate the 3′ OH leaving group. However, the most substantial reduction in free energy derives from the metal's ability to stabilize the negative charge of the phosphoanion transition state (Galburt and Stoddard, 2002).

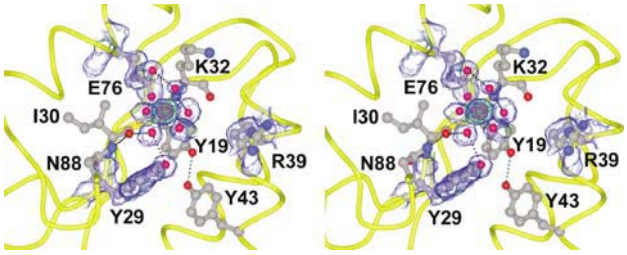

Figure 4 Stereo view of the active site of the 3′ endonuclease domain. The metal ion is shown as a magenta sphere and the five surrounding water molecules as red spheres. Hydrogen bonds are shown as dotted lines. A simulated annealing omit map omitting the magnesium-water cluster, Glu 76, Arg 39 and Tyr 29 is shown at 1σ (blue, transparent) and an anomalous map is shown at 7σ (green cage). Residues in close proximity to the metal ion are shown in ball-and-stick representation.

No divalent cation was observed in the *B. caldotenax* structure after soaking and cocrystallization attempts. This is presumably due to the crystallization conditions, which contained high salt concentrations at low pH (5.4). In addition, there were protein–protein interactions in proximity to the only strictly conserved negatively charged amino acid, Glu 76, where metal binding was predicted to occur. Fortunately, crystals of $UvrC^{N\text{-}Tma}$ grew in PEG 8000 at high pH (8.5), a condition more suitable for soaking experiments. Additionally, Glu 76 was completely solvent accessible with no nearby protein–protein interactions. These crystals were briefly soaked with either $MnCl_2$ or $MgCl_2$ and the resulting structures provided similar results: a single divalent cation coordinated by Glu 76 and five well-ordered water molecules in an octahedral arrangement (Figure 4). These structures provide the first view of the metal and its exact coordination geometry for any member of the GIY-YIG superfamily and thus structural insight into the catalytic mechanism. The bound manganese was clearly identified in anomalous density maps, while omit maps revealed the octahedral arrangement of the chelating water molecules (Figure 4). The refined density for the water molecules in both structures is unambiguous with *B*-factors ranging from 17.1 to 23.6 $Å^2$ for the manganese structure and 21.1 to 28.1 $Å^2$ for the magnesium structure. The *B*-factors for the manganese and magnesium ion are 18.8 and 25.6 $Å^2$, respectively. The manganese structure superimposes onto the magnesium structure with an r.m.s. deviation of 0.1 Å, and the two metal-water clusters superimpose with an r.m.s. deviation of 0.1 Å as well.

Although Glu 76 is the only protein residue bound directly to the metal, the water molecules coordinating to the metal form additional contacts to the protein (Figure 4). One of the waters forms hydrogen bonds to both the hydroxyl group of Tyr 29 and the main-chain carbonyl of Ile 30. A second water forms a hydrogen bond to the carboxylate of Glu 76, while a third water forms a hydrogen bond to the backbone amide of Lys 32. Residues forming the binding pocket for the metal-water cluster are highly conserved, suggesting the metal binding site to be a conserved feature of all GIY-YIG family members.

Mutational analysis of the conserved surface

Based on our structural data, we generated mutations of the highly conserved surface residues Y29A, Y29F, K32A, R39A, F73A, F73E, E76A, I80A, I80E and N88A in full-length

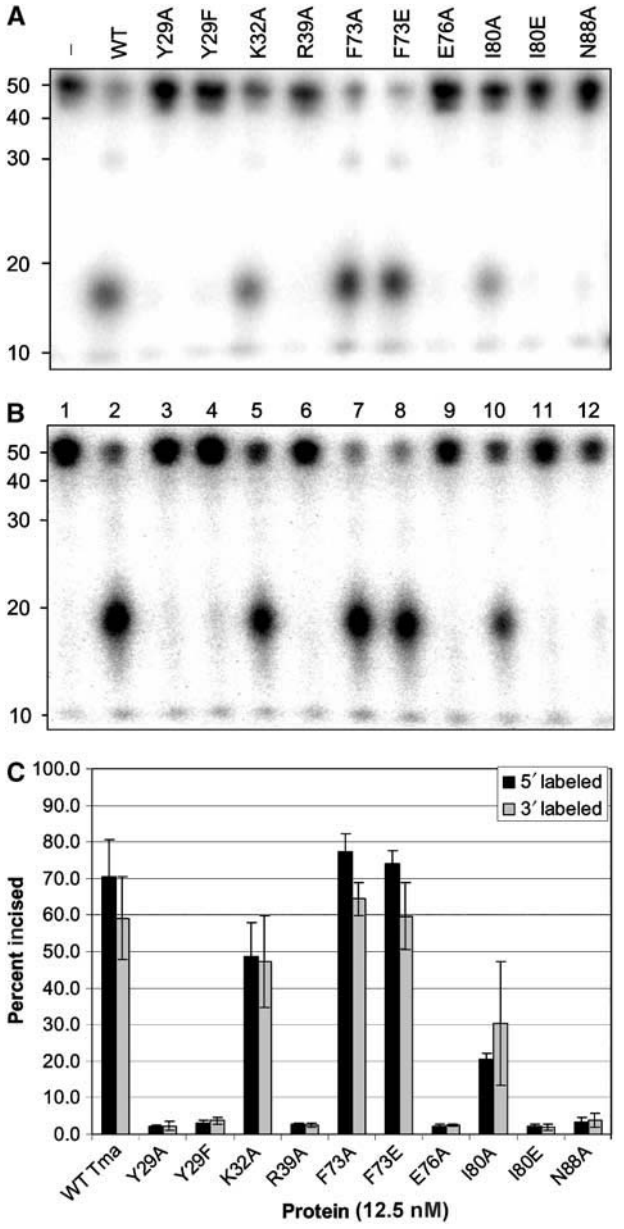

Figure 5 Incision activity of *T. maritima* UvrC mutants. The 5′ (**A**) or 3′ (**B**) end-labeled 50-mer double-stranded DNA substrates containing a centrally located fluorescein (FldT) were incubated with 20 nM $UvrA^{Bca}$, 100 nM $UvrB^{Bca}$ and 12.5 nM of the indicated $UvrC^{Tma}$ protein for 30 min at 55°C in reaction buffer. The reactions were terminated with stop buffer, and the incision products were analyzed on a 10% denaturing polyacrylamide gel. (**C**) Comparison of the incision activity using the 3′-labeled substrate (gray bars) and 5′-labeled substrate (black bars) and the indicated UvrC proteins. Data are reported as the mean ± the standard deviation of the mean of at least four incision assays per UvrC protein.

$UvrC^{Tma}$ and tested these mutant proteins in an incision assay using a defined substrate and the two other NER proteins $UvrA^{Bca}$ and $UvrB^{Bca}$ (Figure 5). In addition, we verified that all of the mutants were able to bind and crosslink to double-stranded DNA. Electrophoretic mobility shift assays were performed using wild-type and mutant UvrC proteins (in the absence of UvrA or UvrB) with a 40 bp duplex containing a centrally located (+)-*trans*-BPDE adduct. Crosslinking was performed using wild-type and

mutant UvrC proteins (in the absence of UvrA or UvrB) with 50 bp duplex containing a site-specific arylazido-modified photoaffinity reagent (as described in DellaVecchia *et al*, 2004) (data not shown).

Mutation of the sole metal ligand, Glu 76, to alanine renders UvrC unable to mediate either the 3′ or 5′ incision (Figure 5). UvrC from *Tma*, like *E. coli* UvrC, cannot achieve 5′ incision without prior 3′ incision, and thus inactivation of the 3′ nuclease active site inhibits the 5′ nuclease activity. Other inactive mutants were N88A, Y29A, Y29F, R39A and I80E. Mutant I80A showed an approximately 50% reduction in activity. Mutation of Lys 32 to alanine resulted in a protein that was 25–30% less active than wild-type UvrC and mutation of Phe 73 to either alanine or glutamate resulted in an enzyme with wild-type UvrC activity (Figure 5). The Y19F and Y43F mutants did not overexpress and could not be studied. In addition to the full-length mutants, we generated mutations in the isolated N-terminal domain (Y19F, Y29F, Y43F and N88A) and determined the structures of these mutants to inspect whether structural changes within the mutants lead to the inactivation of the enzyme (see Supplementary Table I).

Discussion

In this study, we have crystallized and solved the structures of the N-terminal endonuclease domain of UvrC from two different thermophilic bacteria, *T. maritima* and *B. caldotenax*. Multiple sequence alignments have revealed that this domain is the only domain shared among all GIY-YIG family members (Kowalski *et al*, 1999), suggesting it to contain all the information necessary for catalysis. We have identified a patch of highly conserved residues on the surface of the N-terminal domain of UvrC containing a single divalent cation, which is coordinated by Glu 76 and five water molecules in an octahedral arrangement. We have mutated residues within the conserved region in full-length UvrC and analyzed these mutants for incision activity in the complete UvrABC reaction. These results identify residues important in catalysis and clearly indicate that the conserved region is the active site and the bound divalent cation is the catalytic metal.

The active site

A single divalent cation is bound to the conserved surface of the N-terminal domain of UvrC and coordinated exclusively by Glu 76, which is invariant in all known GIY-YIG family members (Figure 4 and 2B). A similar interaction between the equivalent glutamate in I-*Tev*I (Glu 75) and a manganese ion has been reported (Van Roey *et al*, 2002). Mutation of this glutamate to alanine renders UvrC inactive (Figure 5). The analogous mutation in I-*Tev*I gives similar results (Derbyshire *et al*, 1997; Kowalski *et al*, 1999). Likewise, mutation of Ile 80, which forms part of the metal binding pocket, significantly reduces UvrC's activity in the case of I80A and essentially inactivates the protein when mutated to glutamate (Figure 5). In addition, the invariant residue Gly 31 is positioned just behind the bound metal and a side chain at this position would lead to steric interference (Figure 1). Mutation of the analogous glycine (Gly 19) to alanine in I-*Tev*I yields a protein with no detectable activity (Kowalski *et al*, 1999). Furthermore, all residues that hydrogen bond to the metal-coordinated water molecules are conserved in UvrC. A similar octahedral arrangement is seen in the crystal structure of *Serratia* endonuclease (Miller *et al*, 1999), which also coordinates a single, catalytic divalent cation using a lone protein residue, in this case an asparagine. These observations strongly suggest that the bound divalent cation in UvrC is catalytically important and fulfills the role of the Lewis acid in catalysis as suggested for most nucleases (Galburt and Stoddard, 2002).

Since there is no obvious candidate to act as the general acid in the active site, we hypothesize that one of the water molecules coordinated to the metal performs this function. A similar one-metal mechanism is observed in the unrelated His-Cys box homing endonuclease I-*Ppo*I, which, like UvrC, coordinates the divalent cation with only one residue, Asn 119 (Galburt, 1999). The active site of I-*Ppo*I is similar to that of *Serratia* endonuclease mentioned above. In addition, there is no obvious general base in the N-terminal domain of UvrC, which is required to activate a nucleophilic water. In I-*Ppo*I, a conserved histidine in the active site (His 98) fulfills this role by positioning and deprotonating the nucleophilic water (Galburt, 1999). In UvrC, one of the highly conserved tyrosines (Tyr 19, Tyr 29 or Tyr 43) may fulfill this function as there is no histidine present. Since the two full-length UvrC mutants Y19F and Y43F did not overexpress, we were only able to analyze the activity of the Y29F mutant showing that it has no detectable activity (Figure 5). Furthermore, the side-chain hydroxyl of Tyr 29 protrudes from the domain's surface and is in close proximity to the divalent cation (4.3 Å), which places it in a region of high positive charge (Figure 6). This arrangement could be required to lower its pK_a so that it is able to serve as a general base. The hydroxyl of Tyr 29 also forms a hydrogen bond to one of the metal-coordinated water molecules, which in turn hydrogen bonds to the backbone carbonyl of Ile 30 (Figure 4). Assuming that this water molecule is a hydroxide, Tyr 29 could accept a proton from a nucleophilic water while transferring its proton to the metal-bound hydroxide. This suggests that Tyr 29 acts as a proton shuttle with the final acceptor being the metal-bound hydroxide (Figure 7). To analyze the importance of this residue, we solved the crystal structure of the isolated N-terminal domain Y29F mutant bound to manganese (see

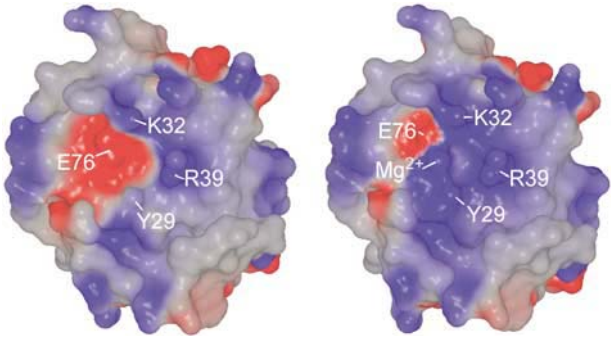

Figure 6 Electrostatic surface representation of the 3′ endonuclease domain calculated in the absence (left) and presence (right) of the bound metal ion. Surface charge was calculated using SPOCK at an ionic strength of 100 mM and is contoured at $\pm 10\,k_BT$. Blue, positively charged; red, negatively charged. The calculations clearly show that the active site is mostly positively charged and Tyr 29 is located in a positively charged environment, which may lower its pK_a.

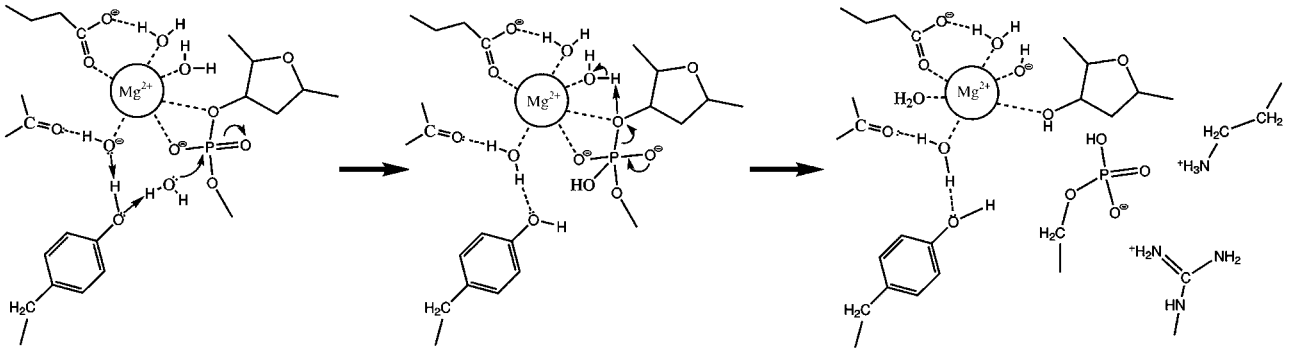

Figure 7 Proposed reaction mechanism for 3′ phosphodiester bond cleavage by UvrC. The metal ion fulfills the role of the Lewis acid and one of the water molecules coordinated to the metal acts as a general acid. Tyr 29 acts as the general base and hydrogen bonds to a metal-coordinated hydroxide. Due to this coordination scheme, Tyr 29 can accept a proton from a nucleophilic water molecule while simultaneously transferring its proton to the metal-bound hydroxide. Arg 39 and Lys 32 are responsible for stabilizing the negative charge of the free 5′-phosphate after DNA cleavage.

Supplementary Table I), confirming that it is identical to the wild-type protein with an r.m.s. deviation of 0.07 Å for all 89 Cα atoms. The analogous mutation in I-*Tev*I (Y17A) retains 1% activity compared to the wild-type I-*Tev*I although it is structurally compromised (Kowalski *et al*, 1999). This does not appear to be the case for the Y29F UvrC mutant since overexpression of the full-length mutant protein is comparable to wild type and the structures of the Y29F and wild type isolated domains are essentially identical. Perhaps the instability of I-*Tev*I is due to the alanine mutation as opposed to phenylalanine in UvrC. However, we observe no change in expression level or solubility of full-length UvrC Y29A, suggesting that the GIY-YIG domain of UvrC could be more stable in general.

We do recognize that there are two other conserved tyrosines in the active site, Tyr 19 and Tyr 43, and we solved the structures of the N-terminal domain mutants Y19F and Y43F, which, in contrast to the full-length enzyme, overexpressed well. Similar to Y29F, both Y19F and Y43F bound the divalent cation and were nearly identical to the wild-type N-terminal domain with an r.m.s. deviation of 0.24 and 0.26 Å for all 89 Cα atoms, respectively. These two tyrosines were initially thought to be structurally important since they form hydrogen bonds to each other (Figure 4), but no significant structural perturbations can be observed upon mutation to phenylalanine. A mutation similar to Y19F was made in I-*Tev*I (Y6A), which inactivated the protein and rendered it insoluble and/or unstable (Kowalski *et al*, 1999). The role of Tyr 19 and Tyr 43 in catalysis awaits further analysis. However, a close look at the position of the three tyrosines favors Tyr 29 as the general base, since Tyr 19 and Tyr 43 do not extend from the surface and are more distant from the divalent cation, resulting in an environment that is less positively charged and therefore less likely to lower the pK_a (Figure 6).

I-*Ppo*I contains a strictly conserved arginine, Arg 61, as do all GIY-YIG family members. The function of the arginine in I-*Ppo*I is to stabilize the negative charge of the free 5′-phosphate after DNA cleavage (Galburt, 1999). A similar role can be imagined for Arg 39 in UvrC (Figure 7). The side chain of Arg 39 is completely solvent exposed and highly flexible as observed by incomplete side-chain density. The critical role of this conserved arginine for the activity of the N-terminal domain of UvrC and I-*Tev*I has been previously demonstrated by site-directed mutagenesis (Derbyshire *et al*, 1997; Verhoeven *et al*, 2000). Here we confirm that a mutation of this conserved Arg 39 to alanine renders UvrC catalytically inactive (Figure 5). A second strictly conserved residue in UvrC, which may also be involved in stabilizing the product, is Lys 32. Lysine 32 is completely solvent accessible and is positioned in close proximity to the bound metal and Arg 39 (Figure 4). However, this residue is not strictly conserved within other GIY-YIG family members and is replaced by a serine in I-*Tev*I (Figure 2A). Mutation of Lys 32 to alanine reduces its activity by 25–30% compared to wild-type UvrC (Figure 5), suggesting that it may play a role in stabilizing the negative charge of the free 5′-phosphate, but is clearly less important than Arg 39.

The role of Asn 88

The last of the four invariant residues is Asn 88. We originally predicted this residue to be structurally important since it forms two hydrogen bonds to Ile 30: one to the backbone amide and one to the backbone carbonyl (Figure 4). The backbone carbonyl of Ile 30, in turn, hydrogen bonds to the same manganese-coordinated water molecule as Tyr 29 (Figure 4). We solved the structure of the N88A mutant since it is in close proximity to the bound metal (Figure 4) and also renders the protein inactive (Figure 5). In I-*Tev*I, the analogous mutation (N90A) leads to an enzyme with only 3% activity compared to the wild type protein (Kowalski *et al*, 1999). Like the three tyrosine mutant structures, N88A can still bind the divalent cation and the structure is nearly identical to wild-type UvrC with an r.m.s. deviation of 0.12 Å over all 89 Cα atoms. Therefore, N88A does not perturb the active site architecture as a water molecule substitutes for the side chain of Asn 88 in the structure. Due to its position at the C-terminus of the 3′ endonuclease domain, it can be speculated that this residue is important to position the catalytic domain correctly with respect to the other domains of UvrC. This is a critical requirement since it has been shown that the interaction between UvrC and the UvrB–DNA complex is located in a region C-terminal to the 3′ endonuclease domain (Moolenaar *et al*, 1997). Upon interaction of UvrC with the UvrB–DNA complex, the N-terminal

domain would then be located in close proximity to the DNA, thus allowing the incision reaction.

Protein–DNA interactions

The GIY-YIG endonuclease domain, when isolated from either UvrC (data not shown) or I-*Tev*I (Van Roey *et al*, 2002), is not able to bind or incise DNA. Apparently, a separate DNA binding and/or protein interacting domain is required to position the GIY-YIG domain properly on the DNA. For UvrC, this would be the UvrB binding domain and/or the C-terminal helix-hairpin-helix DNA binding motif (Moolenaar *et al*, 1997, 1998b; Sohi *et al*, 2000). Van Roey *et al* (2002) proposed that the GIY-YIG domain might interact with DNA in a similar way as I-*Ppo*I (Galburt, 1999) based on similarities in the active sites of the two proteins. The amino acids in I-*PpoI* and UvrC that coordinate the divalent metal ion are both located in an α-helix (Asn 119 in α2 of I-*PpoI* and Glu 76 in α4 of UvrC$^{N\text{-}Tma}$). I-*Ppo*I inserts this helix into the minor groove (Flick *et al*, 1998) of the DNA (Figure 8), which leads to several protein–DNA interactions. We superimposed helix α4 of UvrC with the corresponding helix in I-*Ppo*I while maintaining the position of Glu 76 (UvrC) with respect to Asn 119 (I-*Ppo*I). This results in the overlap of Tyr 29 (UvrC) onto His 98 (I-*Ppo*I) (Figure 8), which was initially encouraging since we hypothesize that Tyr 29 is the general base similar to His 98. Similar results were observed when I-*Tev*I was superimposed with I-*Ppo*I (Van Roey *et al*, 2002). However, in addition to Asn 119, another important residue is located in helix α2 of I-*Ppo*I, Leu 116, which points into the minor groove and forms van der Waals contacts with an adenine. Mutation of Leu 116 has been reported to reduce the catalytic efficiency of I-*Ppo*I (Galburt, 1999). After superposition with UvrC, Leu 116 aligned structurally with Phe 73 (Figure 8), a type conserved hydrophobic residue (Figure 3A and B). We mutated Phe 73 to examine its importance in comparison to I-*Ppo*I. A mutation to either alanine or glutamate resulted in an enzyme that was as active as the wild-type protein (Figure 5), indicating that Phe 73 is most likely not involved in protein–DNA interactions or in a different way compared to I-*PpoI*. In addition to the mutational data, the superposition of UvrC and I-*Ppo*I revealed that the catalytic metals of both enzymes are 3.6 Å apart (Figure 8). These results strongly suggest that the interaction of the GIY-YIG domain with DNA is different from that observed for I-*Ppo*I despite the similarities in the positions of active site residues.

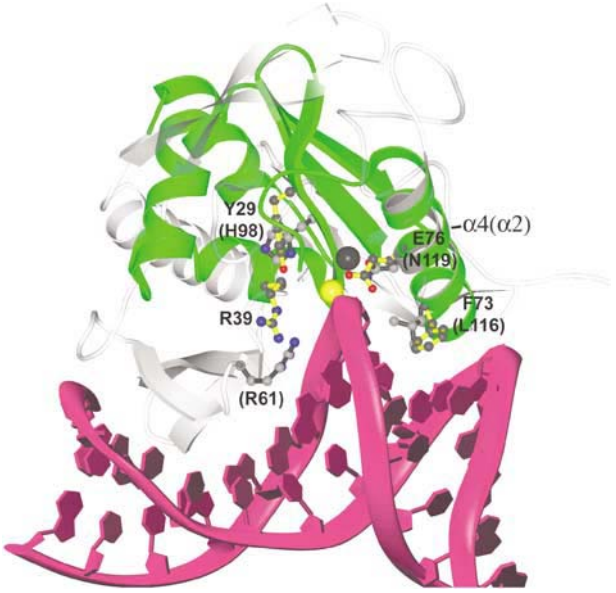

Figure 8 Superposition of helix α4 of the 3′ endonuclease domain of UvrC onto helix α2 of the I-*Ppo*I–DNA complex. I-*Ppo*I and UvrC$^{N\text{-}Tma}$ are illustrated as gray and green ribbon diagrams, respectively. Selected side chains of I-*Ppo*I and UvrC$^{N\text{-}Tma}$ are drawn in ball-and-stick with dark gray bonds for I-*Ppo*I and yellow bonds for UvrC$^{N\text{-}Tma}$. The magnesium ion for each protein is illustrated as a large sphere and color-coded to match the protein's bond color. The DNA is colored magenta and was generated using the program MOLMOL (Koradi *et al*, 1996). Residues are labeled and residue numbers from I-*Ppo*I are shown in parentheses.

Conclusion

We have solved the crystal structure of the N-terminal GIY-YIG endonuclease domain from two thermophilic organisms, *B. caldotenax* and *T. maritima*. This domain is shared by a number of very different proteins all belonging to the GIY-YIG family of endonucleases. In this study, we have characterized the active site of the N-terminal GIY-YIG endonuclease domain of UvrC including the catalytic divalent cation and its coordination geometry. Our results indicate that the N-terminal endonuclease domain of UvrC utilizes a novel one-metal mechanism to cleave the phosphodiester bond. We propose a mechanism where the Lewis acid is the divalent cation, the general acid is a metal-coordinated water molecule and the general base is Tyr 29.

Materials and methods

Protein expression and purification

N-terminal domains of UvrC from *B. caldotenax* (UvrC$^{N\text{-}Bca}$; residues 1–98) and *T. maritima* (UvrC$^{N\text{-}Tma}$; residues 1–97) were cloned into the pTXB1 vector (New England Biolabs) under control of the T7 promoter. The proteins were purified using the T7 IMPACT^{tm} system (NEB). The proteins were expressed in BL21-CodonPlus$^{®}$(DE3)-RIL cells and purified by chitin-affinity chromatography and size-exclusion chromatography. The concentration of UvrC$^{N\text{-}Tma}$ and UvrC$^{N\text{-}Bca}$ was determined by its absorption at 280 nm using calculated extinction coefficients of 7680 and 6400 M^{-1} cm^{-1}, respectively.

Selenomethionine-substituted UvrC$^{N\text{-}Bca}$ was expressed in BL21-(DE3)-RIL cells by methionine biosynthesis inhibition. Cells were grown in M9 minimal medium with 0.2% glucose and 5% (v/v) glycerol as carbon sources. Purification was as described for the native protein with the exception of adding 5 mM DTT to the size-exclusion chromatography buffer.

Crystallization and data collection

Crystals of native and selenomethionine-derivatized UvrC$^{N\text{-}Bca}$ were grown by vapor diffusion, equilibrating equal volumes of protein solution (16.0 mg/ml) and precipitant solution containing 2 M $(NH_4)_2SO_4$, 5 mM DTT and 100 mM sodium acetate pH 5.4, against a reservoir solution containing 2 M $(NH_4)_2SO_4$, 5 mM DTT, 100 mM sodium acetate pH 5.4 and 250 mM NaCl. Crystals of UvrC$^{N\text{-}Tma}$ were grown by vapor diffusion, equilibrating equal volumes of protein solution (16.0 mg/ml) and precipitant solution containing 26% PEG 8000 and 100 mM Tris pH 8.5, against a reservoir solution containing 26% PEG 8000, 100 mM Tris pH 8.5 and 250 mM NaCl. Diffraction data of UvrC$^{N\text{-}Bca}$ and UvrC$^{N\text{-}Tma}$ crystals, cryocooled in liquid nitrogen, were collected at beam lines X26C and X12B, and X26C and X25, respectively, at the National Synchrotron Light Source at Brookhaven National Laboratory. Diffraction data were indexed, integrated and scaled using the HKL software (Otwinowski and Minor, 1997). Crystals of UvrC$^{N\text{-}Bca}$ belong to space group C2 with $a = 86.5$ Å, $b = 86.7$, $c = 67.9$ and $\beta = 120.2$ Å, and contain four molecules per asymmetric unit. Crystals of UvrC$^{N\text{-}Tma}$ belong to

space group $P4_32_12$ with $a = 55.0$ Å and $c = 109.0$ Å, and contain one molecule per asymmetric unit.

$UvrC^{N\text{-}Tma}$ crystals were transferred into 19% PEG 8000, 100 mM Tris pH 8.5, 25% glycerol and either 200 mM $MnCl_2$ or 300 mM $MgCl_2$. Crystals were soaked stepwise in solutions with increasing divalent cation and glycerol concentrations while simultaneously removing NaCl from the mother liquor. The total exposure time to the divalent cation was 1 h.

Structure solution and refinement

The structure of $UvrC^{N\text{-}Bca}$ was determined by MAD phasing using SOLVE (Terwilliger and Berendzen, 1999), which located 10 out of the 16 selenium sites. Phase refinement was performed with RESOLVE (Terwilliger, 2000), gradually extending the phases from 3.0 to 2.0 Å resulting in an ~50% complete model. The remainder of the model was built with O (Jones *et al*, 1991) and refinement was continued with REFMAC (Murshudov *et al*, 1997). Four-fold noncrystallographic symmetry (NCS) restraints were maintained initially and removed in the final refinement cycles. Water molecules were added with ARP (Perrakis *et al*, 1999). TLS refinement (one TLS group per monomer) was used in the final stages to account for overall anisotropic motion of the molecules. Additional calculations were performed using the CCP4 suite of programs (Bailey, 1994).

The structure of $UvrC^{N\text{-}Tma}$ bound to manganese was determined by molecular replacement using AMORE (Navaza, 1994). The search model was a modified monomer from the $UvrC^{N\text{-}Bca}$ structure where 18 dissimilar residues were changed to alanine. The resulting structure was initially refined at 1.5 Å by simulated annealing using CNS (Brunger *et al*, 1998). Side chains from $UvrC^{N\text{-}Bca}$ were replaced with those from $UvrC^{N\text{-}Tma}$ and refinement was continued with REFMAC and O.

$UvrC^{N\text{-}Tma}$ mutants

Single amino-acid residue substitution mutants of $UvrC^{N\text{-}Tma}$ were generated with the QuickChange Site-Directed Mutagenesis Kit (Stratagene) using pTXB1-*uvrC*$^{N\text{-}Tma}$ as template, and sense and antisense oligonucleotides specific for each mutant as PCR primers.

DNA substrates

All DNA substrates were synthesized by Sigma-Genosys (Woodlands, TX). The DNA sequence of the 50-mer double-stranded substrate containing a single internal fluorescein (FldT) adduct was F26 (5′-GACTACGTACTGTTACGGCTCCAT C[FldT]CTACCGCAAT CAGGCCAGATCTGC-3′) while the complementary strand was NDB (5′-GCAGATCTGGCCTGATTGCGGTAGAGATGGAGCCGTAACAGTAC GTAGTC-3′). The F26 strand was 5′ end-labeled using OptiKinase (USB Corporation) and [γ-^{32}P]ATP (3000 Ci/mmol; Amersham Biosciences) according to the manufacturer's instructions. The reaction was terminated by the addition of EDTA and the enzyme was heat denatured by incubation for 10 min at 65°C. Alternatively, the F26 oligo was 3′ labeled using terminal transferase (Roche) and [α-^{32}P]dideoxyATP (3000 Ci/mmol; Amersham Biosciences) according to the manufacturer's instructions. The reaction was terminated as described above. After either labeling reaction, free nucleotides were removed by gel filtration chromatography (Micro Biospin-6, Bio-Rad). The labeled oligonucleotide was annealed to the complementary oligonucleotide (NDB) using a 20% molar excess of NDB. The double-stranded character of the 50-mer duplex was confirmed by native polyacrylamide gel electrophoresis.

UvrABC incision assay

The 5′ or 3′ end-labeled duplex DNA (2 nM) was incised by the UvrABC enzymes (20 nM Bca UvrA, 100 nM Bca UvrB, 12.5 nM Tma UvrC) in 20 µl of UvrABC buffer (50 mM Tris–HCl pH 7.5, 50 mM KCl, 10 mM $MgCl_2$, 1 mM ATP and 5 mM DTT) at 55°C for 30 min. The reaction was terminated by addition of EDTA (20 mM). A 2 µl portion of the reaction was added to 5 µl of formamide and blue dextran, and then heated to 85°C for 10 min. The incision products were resolved on a 10% denaturing polyacrylamide gel and electrophoresis was performed at 325 V for 40 min in Tris–Borate–EDTA buffer (89 mM Tris, 89 mM boric acid and 2 mM EDTA). The gels were dried and exposed to a phosphorimager screen (Molecular Dynamics) overnight. The incision efficiency was calculated using the Molecular Dynamics software ImageQuant.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online.

Acknowledgements

We thank Dr Annie Heroux of X26C, Dr Dieter Schneider of X12B and Dr Michael Becker of X25 for excellent technical assistance at the National Synchrotron Light Source (NSLS). This research was supported by DOE grant (DEFG02-01ER63073), NIH grant (GM070873), and the Pew Scholars Program in the Biomedical Sciences to CK. Beamline X26C at the NSLS in Brookhaven is supported in part by the State University of New York and its Research Foundation.

Accession numbers

Coordinates for native, Mn^{2+}-bound, Mg^{2+}-bound, Y19F, Y43F, Y29F and N88A *T. maritima* UvrC, and native *B. caldotenax* UvrC have been deposited in the Protein Data Bank with accession codes 1YCZ, 1YD0, 1YD1, 1YD2, 1YD3, 1YD4, 1YD5 and 1YD6, respectively.

References

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

Bailey S (1994) The CCP4 suite—programs for protein crystallography. *Acta Crystallogr D* **50:** 760–763

Boyce RP, Howard-Flanders P (1964) Release of ultra-violet light-induced thymine dimers from DNA in *E. coli* K-12. *Proc Natl Acad Sci USA* **51:** 293–300

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges N, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography and NMR systems (CNS): a new software system for macromolecular structure determination. *Acta Crystallogr D* **54:** 905–921

Caron PR, Kushner SR, Grossman L (1985) Involvement of helicase-II (UvrD gene product) and DNA Polymerase-I in excision mediated by the UvrABC protein complex. *Proc Natl Acad Sci USA* **82:** 4925–4929

Christopher JA, Baldwin TO (1998) SPOCK: real-time collaborative molecular modeling. *J Mol Graph* **16:** 285

DellaVecchia MJ, Croteau DL, Skorvaga M, Dezhurov OI, Lavrik B, Van Houten B (2004) Analyzing the handoff of DNA from UvrA to UvrB utilizing DNA–protein photoaffinity labeling. *J Biol Chem* **279:** 45245–45256

Derbyshire V, Kowalski JC, Dansereau JT, Hauer CR, Belfort M (1997) Two-domain structure of the td intron-encoded endonuclease I-*Tev*I correlates with the two-domain configuration of the homing site. *J Mol Biol* **265:** 494–506

Flick KE, Jurica MS, Monnat Jr RJ, Stoddard BL (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-*Ppo*I. *Nature* **394:** 96–101

Friedberg E, Walker J, Siede W (1995) *DNA Repair and Mutagenesis*. Washington, DC: American Society of Microbiology Press

Galburt EA (1999) A novel endonuclease mechanism directly visualized for I-*Ppo*I. *Nat Struct Biol* **6:** 1096–1099

Galburt EA, Stoddard BL (2002) Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry* **41:** 13851–13860

Goosen N, Moolenaar GF (2001) Role of ATP hydrolysis by UvrA and UvrB during nucleotide excision repair. *Res Microbiol* **152:** 401–409

Holm L, Sander C (1995) Dali: a network tool for protein structure comparison. *Trends Biochem Sci* **20:** 478–480

Husain I, Houten BV, Thomas DC, Abdel-Monem M, Sancar A (1985) Effect of DNA polymerase I and DNA helicase II on the turnover rate of UvrABC excision nuclease. *Proc Natl Acad Sci USA* **82:** 6774–6778

Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47:** 110–119

Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14:** 29–32

Kowalski JC, Belfort M, Stapleton MA, Holpert M, Dansereau JT, Pietrokovski S, Baxter SM, Derbyshire V (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-*Tev*I: coincidence of computational and molecular findings. *Nucleic Acids Res* **27:** 2115–2125

Kraulis PJ (1991) MOLSCRIPT—a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* **24:** 946–950

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26:** 283–291

Lin JJ, Sancar A (1992a) (A)BC excinuclease: the *Escherichia coli* nucleotide repair enzyme. *Mol Microbiol* **6:** 2219–2224

Lin JJ, Sancar A (1992b) Evidence for 5′ incision by UvrC through a catalytic site involving Asp399, Asp438, Asp466, and His538 residues. *J Biol Chem* **267:** 17688–17692

Lloyd RS, Van Houten B (1995) DNA damage recognition. In *DNA Repair Mechanisms: Impact on Human Diseases and Cancer*, Vos J-M (ed) pp 25–66. Austin, TX: RG Landes Company, Biomedical Publishers

Merritt EA, Murphy MEP (1994) Raster3D Version 2.0—a program for photorealistic molecular graphics. *Acta Crystallogr D* **50:** 869–873

Miller MD, Cai J, Krause KL (1999) The active site of *Serratia* endonuclease contains a conserved magnesium-water cluster. *J Mol Biol* **288:** 975–987

Moolenaar GF, Bazuine M, van Knippenberg IC, Visse R, Goosen N (1998a) Characterization of the *Escherichia coli* damage-independent UvrBC endonuclease activity. *J Biol Chem* **273:** 34896–34903

Moolenaar GF, Franken KLMC, Dijkstra DM, Thomas-Oates JE, Visse R, van de Putte P, Goosen N (1995) The C-terminal region of the UvrB protein of *Escherichia coli* contains an important determinant for UvrC binding to the preincision complex but not the catalytic site for 3′-incision. *J Biol Chem* **270:** 30508–30515

Moolenaar GF, Franken KLMC, van de Putte P, Goosen N (1997) Function of the homologous regions of the *Escherichia coli* DNA excision repair proteins UvrB and UvrC in stabilization of the UvrBC–DNA complex and in 3′-incision. *Mutat Res* **385:** 195–203

Moolenaar GF, Uiterkamp RS, Zwijnenburg DA, Goosen N (1998b) The C-terminal region of the *Escherichia coli* UvrC protein, which is homologous to the C-terminal region of the human ERCC1 protein, is involved in DNA binding and 5′ incision. *Nucleic Acids Res* **26:** 462–468

Murshudov G, Vagin A, Dodson E (1997) Refinement of macromolecular structures by the maximum likelihood method. *Acta Crystallogr D* **53:** 240–255

Navaza J (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr A* **50:** 157–163

Orren DK, Sancar A (1990) Formation and enzymatic properties of the UvrB–DNA complex. *J Biol Chem* **265:** 15796–15803

Orren DK, Selby CP, Hearst JE, Sancar A (1992) Post-incision steps of nucleotide excision repair in *Escherichia coli*. Disassembly of the UvrBC-DNA complex by helicase II and DNA polymerase I. *J Biol Chem* **267:** 780–788

Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology*, Carter CW, Sweet RM (eds) Vol. 276, pp 307–326. New York: Academic Press

Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6:** 458–463

Sancar A (1994) Mechanisms of DNA excision repair. *Science* **266:** 1954–1956

Sancar A (1996) DNA excision repair. *Annu Rev Biochem* **65:** 43–81

Sancar A, Rupp WD (1983) A novel repair enzyme: UvrABC excision nuclease of *Escherichia coli* cuts a DNA strand on both sides of the damaged region. *Cell* **33:** 249–260

Setlow RB, Carrier WL (1964) The disappearance of thymine dimers from DNA: an error-correcting mechanism. *Proc Natl Acad Sci USA* **51:** 226–231

Skorvaga M, Theis K, Mandavilli BS, Kisker C, Van Houten B (2002) The beta-hairpin motif of UvrB is essential for DNA binding, damage processing, and UvrC-mediated incisions. *J Biol Chem* **277:** 1553–1559

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

Terwilliger TC (2000) Maximum likelihood density modification. *Acta Crystallogr D* **56:** 965–972

Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D* **55:** 849–861

Theis K, Skorvaga M, Machius M, Nakagawa N, Van Houten B, Kisker C (2000) The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch. *Mutat Res* **460:** 277–300

Van Houten B (1990) Nucleotide excision repair in *Escherichia coli*. *Microbiol Rev* **54:** 18–51

Van Roey P, Meehan L, Kowalski JC, Belfort M, Derbyshire V (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-*Tev*I. *Nat Struct Biol* **9:** 806–811

Verhoeven EE, Wyman C, Moolenaar GF, Goosen N (2002) The presence of two UvrB subunits in the UvrAB complex ensures damage detection in both DNA strands. *EMBO J* **21:** 4196–4205

Verhoeven EEA, van Kesteren M, Moolenaar GF, Visse R, Goosen N (2000) Catalytic sites for 3′- and 5′ incision of *E. coli* excision repair are both located in UvrC. *J Biol Chem* **275:** 5120–5123

Zou Y, Van Houten B (1999) Strand opening by the $UvrA_2B$ complex allows dynamic recognition of DNA damage. *EMBO J* **18:** 4889–4901

