The EMBO Journal Vol. 21 No. 22 pp. 6257–6266, 2002

Solution structure and DNA-binding properties of the C-terminal domain of UvrC from *E.coli*

S.Singh, G.E.Folkers, A.M.J.J.Bonvin, R.Boelens, R.Wechselberger, A.Niztayev and R.Kaptein[1]

Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[1]Corresponding author
e-mail: kaptein@nmr.chem.uu.nl

The C-terminal domain of the UvrC protein (UvrC CTD) is essential for 5′ incision in the prokaryotic nucleotide excision repair process. We have determined the three-dimensional structure of the UvrC CTD using heteronuclear NMR techniques. The structure shows two helix–hairpin–helix (HhH) motifs connected by a small connector helix. The UvrC CTD is shown to mediate structure-specific DNA binding. The domain binds to a single-stranded–double-stranded junction DNA, with a strong specificity towards looped duplex DNA that contains at least six unpaired bases per loop ('bubble DNA'). Using chemical shift perturbation experiments, the DNA-binding surface is mapped to the first hairpin region encompassing the conserved glycine–valine–glycine residues followed by lysine–arginine–arginine, a positively charged surface patch and the second hairpin region consisting of glycine–isoleucine–serine. A model for the protein–DNA complex is proposed that accounts for this specificity.

Keywords: bubble DNA/DNA repair/HhH motif/ nucleotide excision repair/UvrC C-terminal domain

Introduction

Cells are equipped with a number of DNA repair defence strategies to maintain the integrity of the genetic material. One of the important repair systems, nucleotide excision repair (NER), is highly conserved in various organisms. It removes a wide variety of damaged nucleotides from DNA by incision on both sides of the lesion. In prokaryotes, NER is mediated by the joint action of the UvrA, UvrB and UvrC proteins (Sancar, 1996). First, a $UvrA_2B$ complex is formed, which recognizes the DNA lesion site. Next, UvrB, which by itself does not have any affinity for damaged DNA, displaces $UvrA_2$, accompanied by ATP hydrolysis, thereby forming a stable UvrB–DNA pre-incision complex. Upon binding to the UvrB–DNA complex, UvrC triggers the dual incisions. The 3′ incision, which precedes the 5′ incision, takes place at the fourth or fifth phosphodiester bond from the lesion. The 5′ incision then follows at the eighth phosphodiester bond from the lesion (Sancar, 1996). Various experiments show that UvrC protein contains the catalytic sites for both 3′ and 5′ incisions (Lin and Sancar, 1992; Verhoeven *et al*., 2000).

Figure 1A shows the domain organization of UvrC. The N-terminus, or Uri domain, shows significant similarity to the intron-encoded endonucleases. This domain consists of conserved tyrosine and glutamate residues that participate in the catalysis of the 3′ incision (Aravind *et al*., 1999). The second domain, called UvrBC, is responsible for binding UvrB. The endo V domain contains the conserved catalytic aspartate residues and catalyses 5′ incision. Finally, the C-terminal region of the UvrC contains the helix–hairpin–helix (HhH) motif shown to be essential for 5′ incision (Moolenaar *et al*., 1998). This domain shows 33% sequence homology to the C-terminal domain (CTD) of ERCC1, the protein that, together with XPF, is responsible for 5′ incision in humans. The HhH motif has been predicted to exist in several DNA repair proteins (Figure 1B) and has been shown to mediate a non-sequence-specific interaction with DNA (Doherty *et al*., 1996; Aravind *et al*., 1999; Shao and Grishin, 2000). Sequence alignments of several HhH motifs show conservation of the GhG pattern in the hairpin region (where h is a hydrophobic residue) and that positively charged residues are often concentrated near the N-terminus of the second α-helix (Figure 1B).

The importance of UvrC CTD in 5′ incision, and the more general role of HhH domains in DNA binding, suggest an important function for this domain in the NER system. We show here that the HhH domain of UvrC is crucial for binding to DNA that contains a region of at least six unpaired bases ('bubble DNA'). We have determined the solution structure of this domain using NMR spectroscopy, which showed that it is composed of two consecutive HhH motifs that are linked by a connector helix. The conserved lysine and glycine residues in the hairpins were found to be important for DNA binding as determined by DNA titration experiments. On the basis of these experiments, we propose a model for DNA recognition by UvrC CTD.

Results and discussion

UvrC CTD is a DNA-binding domain

The loss of 5′ incision by UvrC protein and its inability to bind to single-stranded DNA (ssDNA) in the absence of the UvrC CTD (Moolenaar *et al*., 1998) suggests that the UvrC CTD is directly involved in DNA binding. To test this hypothesis, we performed electrophoretic mobility shift assays (EMSAs) using ssDNA, double-stranded (ds) DNA and various forms of DNA with ss–ds junctions. (Figure 2A). No binding was detected on dsDNA or ssDNA, while binding to bubble type DNA was observed when at least six unpaired bases are present. The binding strength increased with the number of unpaired bases. Weaker binding was observed on hairpin substrates, although the difference from the bubble substrate is not

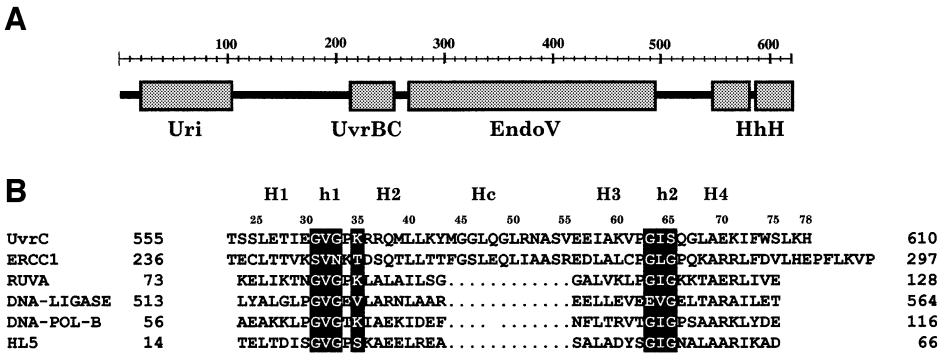

Fig. 1. (**A**) Domain organization of the UvrC protein. (**B**) Sequence alignment of HhH domains of UvrC, ERCC1, RuvA, DNA ligase, DNA polymerase β and HL5. H1, H2, H3 and H4 indicate the presence of an α-helix, Hc represents the presence of a connector helix, and h1 and h2 are the hairpin loops. Conserved regions are indicated in boxes.

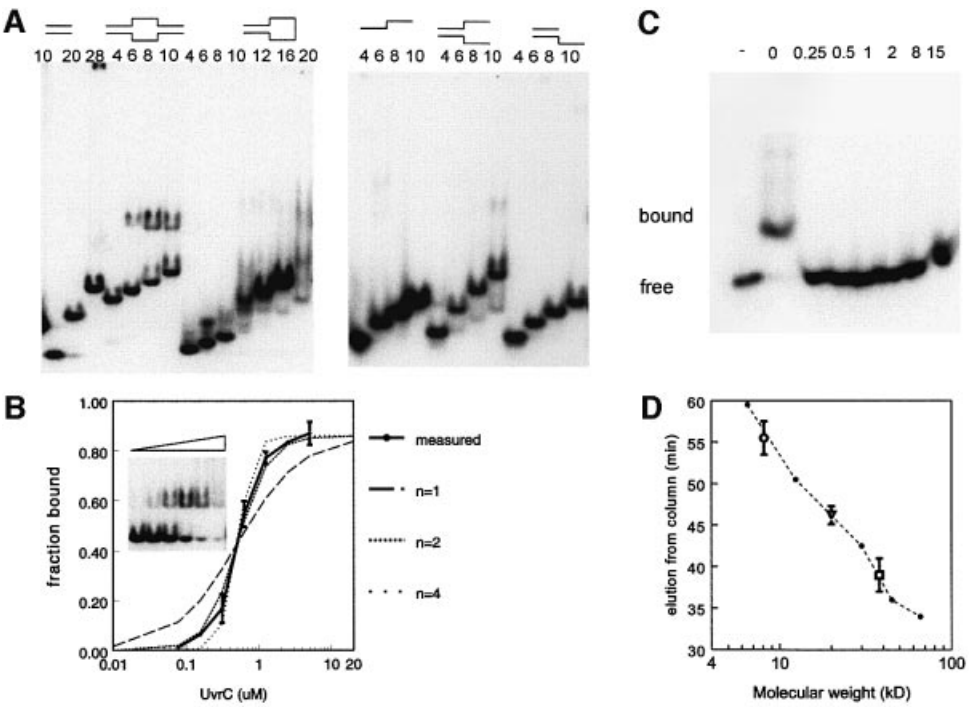

Fig. 2. EMSA assay of UvrC CTD binding to different DNA substrates. (**A**) DNA binding of 1 μM UvrC CTD with dsDNA (10, 20 and 28 bp), bubble DNA, hairpin DNA, ssDNA, a fork with unpaired bases at the 3′ and 5′ end, and a fork with unpaired bases at the 3′ end. The type of DNA and the size of the unpaired bases (T stretch) are indicated schematically at the top of the gel. Free and bound represent the unbound DNA and the DNA in complex with UvrC CTD, respectively. (**B**) Quantification of three representative binding experiments using the bubble with an 8 bp spacer as a probe. The average fraction bound and the SD was determined for the indicated UvrC CTD concentrations (circles with error bars). Using these data, we calculated the theoretical binding curves assuming respectively one (n = 1, dashed line), two (n = 2, small dashed line) or four (n = 4, dotted line) UvrC CTD molecules per DNA substrate, as described in Materials and methods. The inset shows one of these experiments with 0.08, 0.16, 0.31, 0.625, 1.25 and 5.0 μM UvrC CTD. (**C**) Off-rate determination. Binding competition by the addition of a 200-fold molar excess of unlabelled homologous DNA at the indicated time points (in min) of a pre-formed complex of UvrC (3 μM) with a bubble with eight unpaired bases. –, no protein added to the binding mixture; 0, no competitor added to the protein–DNA complex. (**D**) Graphical representation from a gel filtration experiment showing the time after elution from the column (in min) against the molecular weight, calibrated using BSA (66 kDa), ovalbumin (46 kDa), carbonic anhydrase (30 kDa), cytochrome *c* (12 kDa) and aprotinin (6 kDa) (these proteins are indicated by black dots). The average molecular weight and SD for UvrC (open circle), bubble DNA with 10 unpaired bases (open triangle) and protein–DNA complex (open square) was determined by measuring the elution time of two injections for both the references and the indicated samples.

very large. However, a stable complex was formed only when 16–20 unpaired bases were present. A 'fork' with unpaired DNA on the 3′ end of the dsDNA was a poor substrate for UvrC CTD. Weak binding was detected with a fork containing 8 or 10 unpaired bases on both the 3′ and 5′ ends. Interestingly, the protein–DNA complexes formed with various substrates with an increasing number of unpaired bases appear in almost equal positions in the gel, despite their differences in size. This suggests a similar distortion or bending of the various DNA substrates upon binding. The specificity of the complex was confirmed by binding competition using an excess of unlabelled DNA

Table I. Quantification of UvrC CTD binding to various DNA substrates

DNA	*n*	Apparent K_d (μM)	Hill coefficient
Bubble	7	0.7 ± 0.2	2.3 ± 0.4
Hairpin	8	1.2 ± 0.4	1.5 ± 0.3
Fork	5	3.0 ± 0.6	1.5 ± 0.3
ssDNA	7	19.5 ± 17	1.2 ± 0.6

Binding of UvrC CTD to a bubble with eight unpaired T bases, a hairpin with 16 unpaired T bases, a fork with eight unpaired T bases with an overhang on both the 5′ and 3′ end, and ssDNA (5′-GGGCGGCGGGTTTTTTTT).
The apparent K_d and Hill coefficient ± SD is calculated from the indicated number of experiments (*n*) as described in Materials and methods.

(data not shown). The N-terminal His tag does not contribute to the observed binding since its removal did not change the binding affinity (data not shown).

The observed binding of UvrC CTD contrasts with results obtained using full-length UvrC on comparable substrates. No binding was observed for UvrC on DNA fragments that served as good substrates for the UvrBC complex (Zou *et al*., 1997; Moolenaar *et al*., 1998, 2000). Although experimental differences such as binding conditions (or DNA sequences) could not be excluded, we propose that the HhH domain is exposed differently due to steric hindrance or conformational changes when present in a UvrB–UvrC complex, in UvrC or in isolation. Indeed, the activities of other domains of UvrB or UvrC also change upon complex formation. For example, UvrB is unable to bind to lesion-containing dsDNA, whereas it can bind to a damaged DNA substrate with a 5′ overhang. The UvrB–UvrC complex, however, can bind to both substrates (Moolenaar *et al*., 1998, 2000). Furthermore, formation of a complex is essential for DNA repair since neither UvrB nor UvrC alone can perform incisions (Moolenaar *et al*., 1995). Together, these data indicate that the two individual proteins function differently when present in isolation or in the UvrB–UvrC complex. This might explain the absence of DNA binding of UvrC, despite the presence of a domain that is shown to be required for DNA binding (Moolenaar *et al*., 1998).

The observed binding preferences of the UvrC CTD to various (non-damaged) DNA substrates are in good agreement with the reported binding preference of the UvrB–UvrC complex. This complex binds to a DNA substrate containing an ss–ds junction (denoted Y substrate) with an apparent K_d of 5–10 nM (Zou *et al*., 1997). Studies by Moolenaar *et al*. (1998) extended these observations, showing that the UvrB–UvrC complex binds a probe containing a 5′ ssDNA overhang but not a 3′ ssDNA. The prerequisite for at least six unpaired bases for UvrC CTD to bind to a bubble substrate is in good agreement with the minimal size of the bubble required for incision by the UvrB–UvrC complex. However, under these conditions, the incision was strictly dependent on DNA containing a modified base (Zou *et al*., 1999). The weak binding to ssDNA is in accordance with the observation by Moolenaar *et al*. (1998) showing CTD-dependent binding to an ssDNA cellulose column (estimated to be in the micromolar range by Moolenaar *et al*., 1998). The inability to detect binding with dsDNA is compatible with the reported absence of binding affinity of the monomeric form of UvrC (Tang *et al*., 2001).

Quantitative analysis of the binding of UvrC CTD to DNA

Data on binding of UvrC CTD to various DNA substrates were fitted to determine the K_d. The apparent binding affinity for the bubble substrate was calculated to be 0.7 ± 0.2 μM. While the hairpin was a slightly worse substrate, the fork binds with a 4-fold lower affinity (Table I). We were unable to determine the K_d for ssDNA accurately, especially for dsDNA due to the low affinity of the protein for these substrates (Table I; data not shown). These binding data could be fitted to cooperative binding models with Hill coefficients >1, indicating the formation of at least two UvrC CTD molecules per DNA substrate. The Hill coefficient of 2.3 ± 0.4 for bubble DNA points towards an even higher order complex. For comparison, we show the simulated cooperative binding curves (Figure 2B) for this protein–DNA complex when composed of one, two or four UvrC CTD molecules per DNA substrate. Due to binding of UvrC CTD to additional low affinity binding sites (as illustrated by smearing at higher protein concentrations, inset of Figure 2B), we are unable to conclude from these DNA-binding studies whether two or more UvrC CTD proteins are binding to a bubble substrate.

The observed smearing indicates that during electrophoresis, the complex dissociates, suggesting a high off-rate. Dissociation rate experiments confirm the presence of a high off-rate for all the substrates tested (bubble 8, hairpin 16, fork 8, ssDNA or dsDNA). At the first time point (15 s) after the addition of unlabelled DNA, all protein–DNA complexes were competed away (Figure 2C; data not shown). Given the detection limits of the experiment, this indicates a $k_{off} > 0.5$/s.

The binding experiments show that UvrC CTD binds with an apparent K_d of ~1 μM to specific DNA substrates containing ss–ds junctions. Although systematic analyses of DNA-binding affinities of isolated HhH domains are not available, the observed K_d is in the same range as reported for other HhH domain-containing proteins (Doherty *et al*., 1996; Shao and Grishin, 2000). This binding affinity is at least an order of magnitude lower than for the UvrB–UvrC complex on similar substrates (Zou *et al*., 1997; Moolenaar *et al*., 1998). The lower affinity could be caused by DNA substrates that differ in both length and sequence. However, it is more likely that other part(s) of the proteins present in the UvrB–UvrC complex also fulfil an essential role in complex formation. Indeed, isolated UvrB can bind to some DNA substrates without lesions (Moolenaar *et al*., 2000), and atomic force microscopy revealed that extensive UvrB-dependent DNA wrapping is observed (Verhoeven *et al*., 2001), underscoring the idea that other domains in the UvrB–UvrC complex also contribute to DNA binding.

UvrC CTD is a monomer in solution and binds as a dimer to bubble DNA

From binding studies, we were unable to determine accurately the protein–DNA ratio, although a 2:1 ratio seems most likely. This leads to the question of whether

Table II. Structural statistics for the UvrC CTD[a]

R.m.s.d. (Å) with respect to the mean	
Heavy backbone atoms (residues 28–75)	1.00 ± 0.2
Heavy backbone atoms (residues 36–75)	0.46 ± 0.1
All heavy atoms (residues 28–75)	1.45 ± 0.2
All heavy atoms (residues 36–75)	1.00 ± 0.1
No. of experimental restraints	
Intra-residue NOEs	625
Inter-residue sequential NOEs ($\|i-j\|=1$)	305
Inter-residue medium-range NOEs ($1<\|i-j\|<5$)	213
Inter-residue long-range NOEs ($\|i-j\|>4$)	183
Total NOEs	1326
Dihedral angle restraints	68
Restraint violations[b]	
NOE distances with violations >0.3 Å	2.2 ± 1.7
Dihedrals with violations >3°	1.3 ± 1.2
R.m.s.d. for experimental restraints[c]	
All distance restraints (1326) (Å)	0.032 ± 0.003
Torsion angles (68) (°)	0.7 ± 0.2
CNS energies from SA[d]	
F_{vdw} (kcal/mol)	–512 ± 10
F_{elec} (kcal/mol)[e]	–2135 ± 74
R.m.s.d. (Å) from idealized covalent geometry	
Bonds (°)	0.0062 ± 0.00
Angles (°)	0.61 ± 0.03
Impropers(°)	0.56 ± 0.05
Ramachandran analysis (residues 23–78[f], 36–74[g])	
Residues in the favoured region (%)	69.1[f], 79.7[g]
Residues in additional allowed regions (%)	27.8[f], 19.5[g]
Residues in generously allowed regions (%)	3.1[f], 0.7[g]
Residues in disallowed regions (%)	0.0[f], 0.0[g]

[a]Based on the 22 structures, obtained by simulated annealing in CNS followed by refinement in explicit water using NOE distance restraints, dihedral angle restraints, bonds, angles, impropers, dihedral angle, van der Waals and electrostatic energy terms.
[b]No distances were violated by >0.5 Å, and no dihedral angle restraints were violated by >5°.
[c]The number of each class of experimental restraints is given in parentheses.
[d]Force constants were described in Materials and methods.
[e]The Lennard-Jones 6–12 and coulomb energy terms were calculated within CNS using the OPLS non-bonded parameters (as described in Materials and methods).
[f]Ramachandran analysis of the UvrC CTD (residues 23–78).
[g]Ramachandran analysis of the structured region of the UvrC CTD (residues 36–74).

UvrC CTD is a dimer in solution or dimerizes upon DNA binding. From a blue native gel electrophoresis experiment (Schägger *et al.*, 1994), we noticed that UvrC CTD is monomeric in solution (data not shown). Gel filtration experiments further confirm this since UvrC CTD elutes from the column corresponding to a mol. wt of 8.1 ± 0.8 kDa (Figure 2D). The determined molecular weights for the bubble DNA and protein–DNA complex from the elution profile were 20 ± 2 and 38 ± 4 kDa, respectively. This indicates that the complex most probably contains two UvrC CTD molecules per DNA molecule.

Three-dimensional structure

The structure of UvrC CTD ($UvrC^{554–610}$) was solved at 27°C (20 mM $NaPO_4$ pH 6.8, 300 mM NaCl) using two- and multidimensional heteronuclear NMR spectroscopy, making use of uniformly ^{15}N- and $^{15}N/^{13}C$-labelled protein. NMR shows that the UvrC CTD is monomer in solution (by estimating the line width of the signals in the $^{1}H–^{15}N$ HSQC spectrum), which is in agreement with the biochemical data presented in the previous section. Nearly complete ^{1}H, ^{15}N and ^{13}C assignments were obtained using standard resonance assignment procedures (see Materials and methods). The global fold was established using unambiguously assigned long-range nuclear Overhauser effects (NOEs). Several long-range NOEs between the helices were obtained from a 3D-NOESY(^{13}C,^{1}H)-HSQC spectrum and a 3D-NOESY(^{15}N,^{1}H)-HSQC spectrum. The unassigned NOEs with multiple possible assignments were used in ARIA (Linge and Nilges, 1999) as ambiguous restraints. The final ensemble of structures was calculated by simulated annealing on the basis of 1326 unambiguous and 60 ambiguous distance restraints, with the inclusion of 68 dihedral angle constraints obtained from TALOS on the basis of backbone chemical shift values. A summary of the experimental restraints and structural statistics is provided in Table II. A superposition of the final ensemble of 22 simulated annealing structures is shown in Figure 3A, and a stereo view of the representative (closest to average) structure in Figure 3B. It should be noted that inclusion of the automatically assigned restraints from ARIA increases the coordinate precision from 0.57 ± 0.1 to 0.46 ± 0.1 Å for the backbone atoms (Figure 4) and from 1.17 ± 0.1 to 1.0 ± 0.1 Å for all heavy atoms (residues 36–75).

The UvrC CTD consists of two consecutive HhH motifs, called HhH-I and HhH-II, linked by a connector helix (Figure 3). In the first motif, HhH-I, the first presumed helix, H1, is composed of just a single helical turn, while in general this is a well-defined helix in most of the HhH domains determined so far. In principle, this could be due to inadequate domain selection. However, (^{15}N,^{1}H)-HSQC and 3D-NOESY(^{15}N,^{1}H)-HSQC experiments with an extended protein ($UvrC^{481–610}$) exclude this possibility, since neither significant differences in chemical shift values nor extra NOEs that would be indicative of formation of an α-helix were seen for the H1 residues. The ^{15}N- and ^{13}C-edited 3D-NOESY-HSQC spectra display only a few sequential and no long-range NOEs for residues 22–27. The $^{1}H–^{15}N$ NOE data (Figure 4C) reveal high backbone mobility in this region of the protein. The hairpin loop h1 (residues 28–34) shows mainly sequential NOEs but only a few long-range NOEs, and this results in the disorder observed for this region. The hairpin includes the conserved GhG pattern (where h is valine in this case). After this h1 loop, the second helix, H2, follows (residues 35–43). The N-terminal part of H2 containing the positively charged residues Lys35 and Arg36 shows very few sequential or long-range NOEs. This could be due to the mobility of their side chains. The connector helix, Hc, consists of residues 47–52. The connector helix is oriented in a perpendicular direction with respect to adjacent

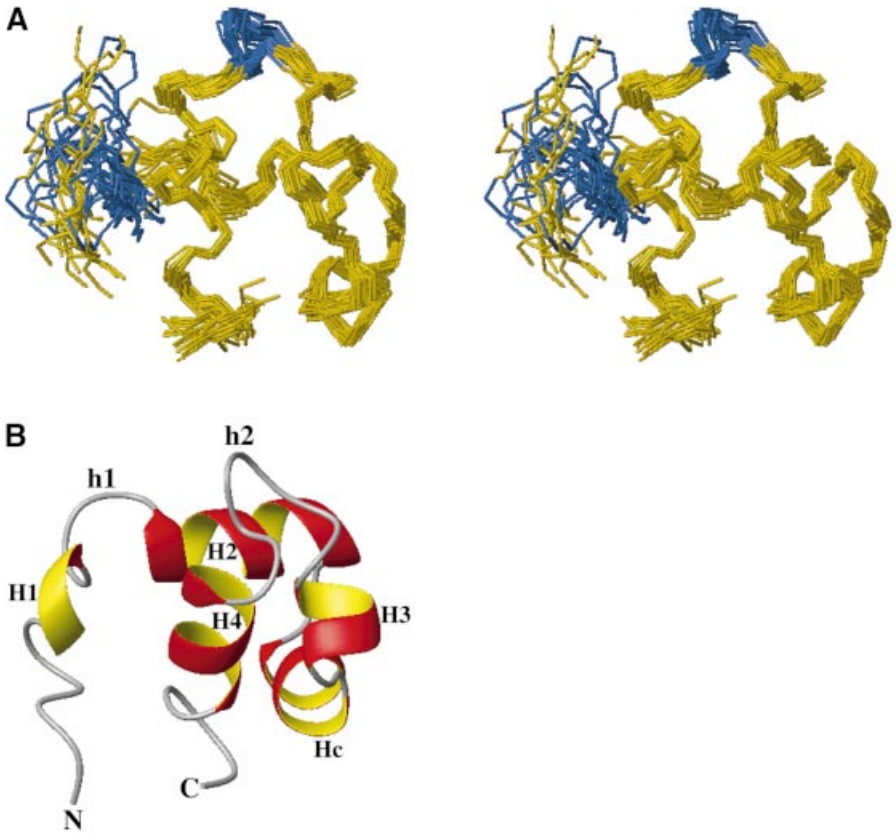

Fig. 3. The NMR solution structure of the UvrC CTD. (**A**) Backbone stereo view (residues 28–78) of the NMR ensemble (22 structures); the hairpins are coloured in blue. (**B**) Ribbon view of a representative UvrC CTD structure (closest to average) for residues 23–78. h1 and h2 are the hairpins of HhH motifs. The structures were displayed using the molecular graphics program MOLMOL (Koradi *et al*., 1996).

helices. The last two helices, H3 and H4, consisting of residues 55–58 and 67–77, form the HhH-II motif, with a hairpin loop, h2, consisting of residues 59–66. The conserved pattern GhG is broken in h2 by a serine in the place of the second glycine (GIS). This type of replacement of a conserved sequence is often seen in the case of HhH domains.

Structural similarity to other HhH proteins

By searching the Protein Structure DataBase (PDB) with the coordinates of the UvrC CTD using the program DALI (Holm *et al*., 1993), structural similarity was found to HhH domains of two proteins, namely RuvA and DNA ligase (Thayer *et al*., 1995; Pelletier *et al*., 1996; Roe *et al*., 1998; Ariyoshi *et al*., 2000; Lee *et al*., 2000). Backbone r.m.s.ds of 1.63 and 1.47 Å were obtained between the representative UvrC CTD structure and RuvA and DNA ligase, respectively (Figure 5). The most prominent difference between UvrC CTD and these two proteins is that UvrC has just a single turn in the first helix of HhH-I, while the others have a well-defined helix. In fact, comparison of the presently known HhH domain structures reveals that there is considerable variation in the detailed structures of various HhH domains: e.g. the absence, or distortion of the first α-helix, the absence or presence of the connector helix, variation in the size of the connector helix, etc. (Velankar *et al*., 1999; Wada *et al*., 2000). The localized sequence and structural differences in HhH motifs might result in different substrate specificities.

Mapping of the binding site by chemical shift perturbation

The DNA interaction surface of the UvrC CTD protein was determined using the NMR chemical shift perturbation method (Dekker *et al*., 1993), in which the 2D-^{1}H–^{15}N HSQC spectra of UvrC CTD were recorded with successive additions of DNA. The method detects residues that are interacting directly with DNA or that are indirectly affected by binding (Figure 6). Large chemical shift changes (>0.2 p.p.m. in ^{1}H and 3.0 p.p.m. in ^{15}N) were observed upon DNA binding mainly in the hairpin regions, indicating their involvement in the interaction with DNA. The changes increased on continued addition of DNA. Except for the signals corresponding to Glu30 and the side chain of Arg36, the rest of the signals that were involved in binding disappeared when half equivalent of DNA was added. As it became difficult to determine the end point of the titration due to broad or disappearing signals, determination of the K_d and the stoichiometry of the complex became impossible. Since, the maximum observed chemical shift change is 100 Hz (which is the last point before disappearance of the amide proton of Gly31), we find a lower limit of ~600/s for the dissociation rate constant. The titration results suggest a complex binding mechanism that includes both high- and low-affinity binding sites. The protein seems to be in intermediate to fast exchange on the NMR time scale in its complex with DNA, as judged by ^{1}H–^{15}N HSQC cross-peaks that showed changes in their position and broadened resonances. In most cases, these

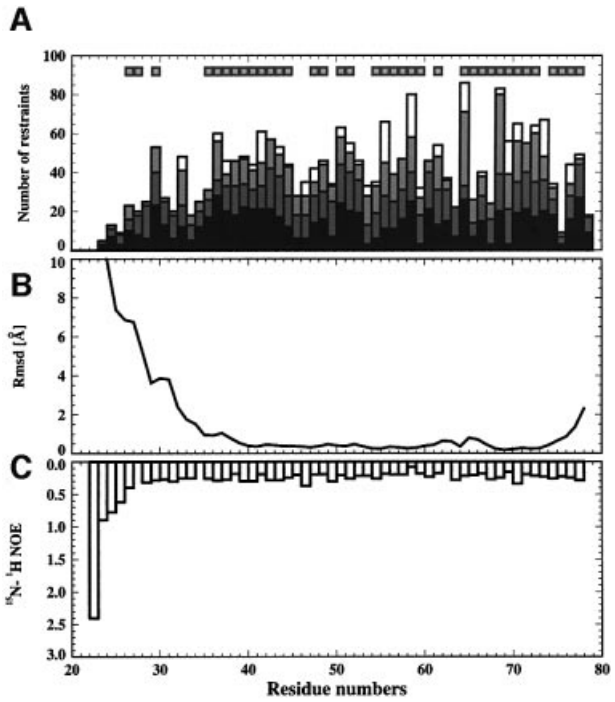

Fig. 4. (**A**) NMR restraints statistics. (**B**) Backbone r.m.s.d. (**C**) Backbone $\{^1H\}$–^{15}N NOEs as a function of residue number. The NOE restraints are intra-residue NOEs, inter-residue short-range sequential NOEs ($|i - j| = 1$), inter-residue medium-range NOEs ($1 < |i - j| < 5$) and long-range NOEs ($|i - j| > 4$) from bottom to top. The boxes at the top indicate the inclusion of TALOS-derived ϕ and χ angle restraints for that residue. The residue r.m.s.ds were calculated over the ensemble of 22 structures after superposition of residues 36–74.

changes allowed unambiguous assignment of the cross-peaks of the ^1H–^{15}N HSQC spectrum by following the cross-peak position during the DNA titration. The ^1H and ^{15}N chemical shift changes for backbone amides are shown in Figure 6A and define the DNA-binding surface for this DNA substrate with the UvrC CTD protein. The interaction site includes residues Thr28, Ile29, Glu30, Gly31, Val32 and Gly33 from the HhH-I hairpin. Positively charged residues such as Lys35, Arg36 and Arg37 along with Met39 of the second helix of HhH-I also show changes in their chemical shift values. In addition, the peaks of Gly63, Ile64 and Ser65 from the second hairpin loop HhH-II shift upon DNA addition. The maximum amide chemical shift perturbation is observed for the residues Gly31 and Gly33 of the first hairpin. The downfield shift of the backbone nitrogens as well as that of the amide protons of Gly31 and Gly33 of the HhH-I hairpin indicate the formation of hydrogen bonds between amide protons and the phosphate oxygens of DNA, as also found in the case of RuvA (Ariyoshi *et al*., 2000). The amide chemical shift changes of the basic residues Lys35 and Arg36 may indicate an interaction with the phosphate backbone, although these changes could also be due to the interaction of the side chains with the unpaired bases as there is a substantial downfield shift of the side chain amide proton of Arg36, indicating hydrogen bond formation.

From the NMR chemical shift mapping method, we see that UvrC CTD shows a similar pattern of interaction with

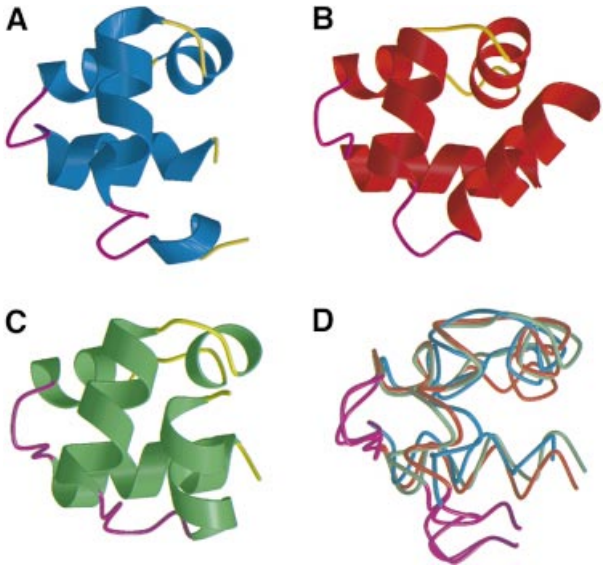

Fig. 5. Comparison of the overall topologies of the HhH domain of (**A**) UvrC in blue, (**B**) RuvA in red, (**C**) DNA ligase in green and (**D**) an overlay of all three. The hairpin loops are coloured purple and the helix H1 was removed for the sake of clarity. The backbone r.m.s.d. of UvrC CTD with RuvA is 1.63 Å and with DNA ligase is 1.47 Å. The structures were displayed using the software Molscript and Raster 3D (Esnouf, 1997).

DNA to that observed for RuvA (Ariyoshi *et al*., 2000) and DNA ligase (Lee *et al*., 2000). RuvA is a tetrameric Holliday junction-recognizing protein, containing two copies of the HhH motif in each monomer that interact with the minor groove of the DNA. The crystal structure of the RuvA–Holliday junction complex shows that DNA interaction with the protein is mainly through hydrogen bond formation between the amides of the residues in the hairpin and the phosphate backbone of the Holliday junction DNA. Polar interactions between the side chains of the basic residues close to the hairpin loop of the HhH motif and the phosphate oxygens have also been found. Despite the absence of the well-defined first helix, the UvrC CTD can interact with the DNA through the hairpin loops and the neighbouring positively charged residues of the HhH-I.

Model for DNA binding

We have shown above that the UvrC CTD binds weakly to either ssDNA or dsDNA. However, when binding experiments were performed with looped duplex DNA, binding to DNA was found with a minimum loop size of six bases. Strong binding was not observed with hairpin or fork DNA substrates, suggesting that a ds–ss junction is not sufficient for binding. The cooperative nature (Figure 2A) of the protein binding to bubble DNA requires that at least two molecules of UvrC CTD are involved in DNA binding. To accommodate two UvrC CTD molecules at the two ds–ss junctions of the bubble DNA, the junctions need to be at a certain minimum distance. Thus, to account for the loop size requirement and the cooperative binding, we suggest a model where two properly interacting molecules of UvrC CTD bind to the two ds–ss junctions of the bubble DNA.

The NMR shift perturbation results of UvrC CTD with DNA have shown that both hairpins h1 and h2 are involved in binding. As compared with the structurally

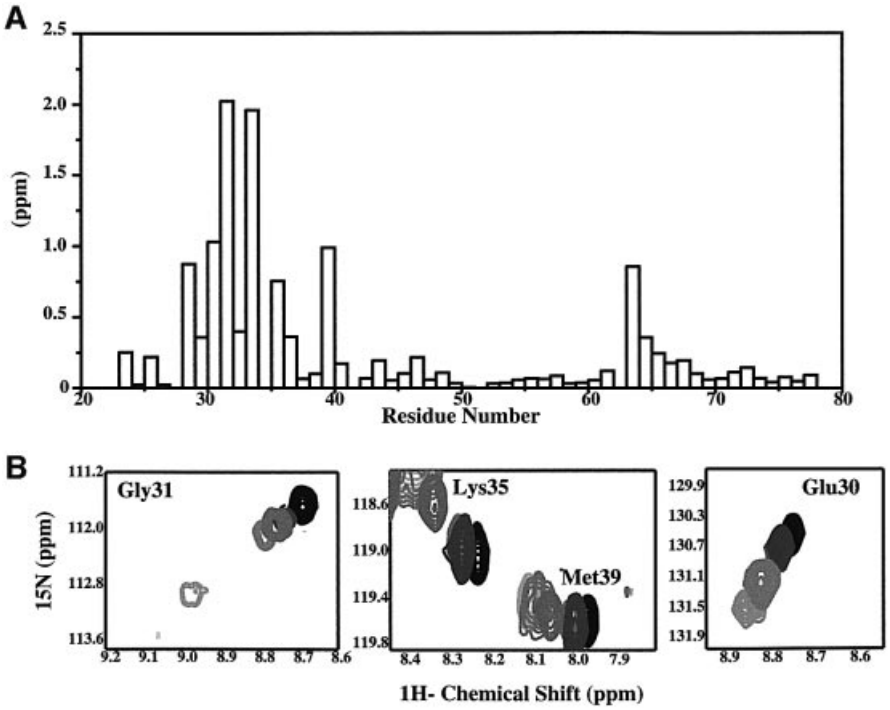

Fig. 6. (**A**) Plot of chemical shift value changes (the r.m.s.d. of chemical shift changes in backbone amide nitrogen and amide proton) upon titration of UvrC CTD and the bubble DNA versus residue number. The absence of a bar in the plot indicates the presence of a proline residue or an unmeasured shift due to overlap. (**B**) Part of the HSQC spectrum displaying changes in the chemical shift values of Gly31, Lys35, Met39 and Glu30 residues, respectively. The peak position corresponds to 0, 0.032, 0.068, 0.13, 0.5 and 1.0 equivalent of DNA to UvrC CTD (increasing from light grey to black).

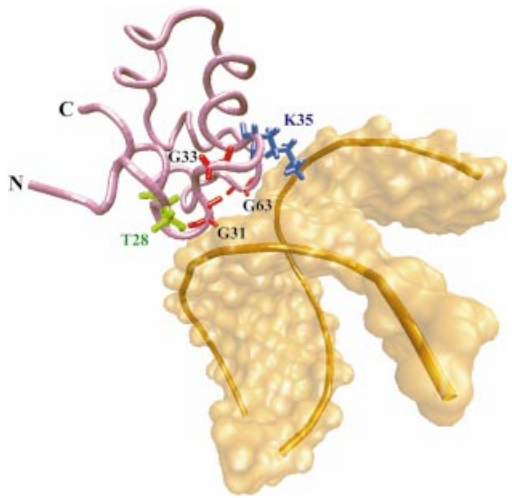

Fig. 7. Model representing the interaction of the UvrC CTD with the ds–ss junction. Glycines are coloured red, lysine in blue and threonine in green. This model was obtained by superimposing the two hairpins of UvrC CTD on to the corresponding loops of RNA polymerase II (domain containing the active site of the Rpb1 subunit, PDB accession No. 16IH, see Materials and methods). The structure was generated using the software VMD (Humphrey *et al*., 1996).

homologous HhH domain of RuvA, similar regions are interacting with DNA: the hairpin loops and the positively charged residues from the second helix of HhH-1. In RuvA, these make contacts with the DNA backbone in the minor groove at a Holliday junction (ds–ss) (Roe *et al*., 1998). In the deposited structure of the RuvA complex (Rafferty *et al*., 1996; PDB 1CUK), however, the hairpins contact DNA within the duplex region and the DNA coordinates for the junction are missing. Since our biochemical experiments have shown that UvrC CTD binds to bubble DNA, we have modelled the protein–DNA complex using the ds–ss junction coordinates of the DNA of the RNA polymerase II complex (Cramer *et al*., 2001; PDB 1I6H). Figure 7 shows an interaction surface of the protein in the UvrC CTD–DNA complex where the two hairpins of the HhH domain interact with the two strands of the DNA. In our NMR titration experiments with the UvrC CTD, the conserved glycine residues in the first hairpin and the positively charged residues (Lys35) at the N-terminus of the second helix of the HhH-I show the largest chemical shift perturbation. The chemical shift changes for the backbone amides of Gly31 and Gly33 are attributed to the interaction between these amides and the phosphate oxygens, similarly to that seen for RuvA. The side chain of Lys35 might interact either with the phosphate oxygens as in the case of RuvA, or with the unpaired bases in the loop. This type of interaction with the phosphate backbone is likely to be non-sequence specific, but may require a distortion in the backbone. Studies of the pre-incision complexes of UvrB protein with substrates such as *cis*-Pt–DNA (Visse *et al*., 1994), BPDE–DNA (Zou and Van Houten, 1999) and AAF–DNA (Pierce *et al*., 1989) have shown that the formation of the UvrB–DNA pre-incision complex is associated with significant changes in the DNA structure. These reports suggest that those changes may include unpairing of base pairs, unwinding of the helix, destacking of the base pairs,

etc. The bending of the DNA in the pre-incision complex has also been revealed by electron microscopy experiments (Shi *et al*., 1992) and by linear dichroism (Takahashi *et al*., 1992) of the DNA. These studies, along with the preference of the UvrC CTD for bubble DNA might suggest that bubble DNA resembles the intermediate state of DNA corresponding to the pre-incision complex. In summary, we have established that UvrC CTD is crucial for DNA binding, specifically to bubble DNA with a loop size of at least six bases. The structure of the domain has been solved and the DNA interaction surface characterized. Based on this, a model for the protein–DNA complex is presented that accounts for the interaction with DNA. More detailed structural studies of the UvrC CTD–DNA complex are in progress.

Materials and methods

Plasmid

UvrC CTD (residues 554–610) and an extended UvrC construct (residues 471–610) were PCR amplified using *Escherichia coli* genomic DNA as a template, with primers (sequence available upon request) that were extended with an *Nde*I and *Bam*HI site at the 5′ and 3′ end, respectively. Purified PCR products, after digestion with *Nde*I and *Bam*HI, were cloned into the corresponding sites of pET15B (Novagen). Constructs were confirmed by sequencing.

Purification of UvrC$^{554–610}$

Escherichia coli BL21 (DE3) containing the plasmid were grown in minimum medium at 30°C, induced at A_{600} = 0.6 with 0.5 mM isopropyl-β-D-thiogalactopyranoside and grown for a further 5–6 h before harvesting. For the production of ^{15}N-labelled protein $^{15}NH_4Cl$ was used, and for the $^{13}C/^{15}N$-labelled sample, 0.2% (U-6)[^{13}C]glucose and $^{15}NH_4Cl$ was used. All purification steps were carried out at 4°C. The cells were resuspended in lysis buffer [50 mM $NaPO_4$ pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.2% Triton X-100 (w/v), 1 mM β-mercaptoethanol, 0.2 mM phenylmethylsulfonyl fluoride (PMSF) and protease inhibitors (Sigma)] and stored at –80°C.

Frozen cells were lysed by two freeze–thaw cycles, followed by sonication using a Soniprep 150 model (MSE). The sonicate was cleared of cell debris by centrifugation at 15 000 *g* for 1 h. The supernatant solution containing UvrC CTD was loaded on an Ni-NTA column (Qiagen) equilibrated with buffer containing 50 mM sodium phosphate, 300 mM NaCl and 20 mM imidazole at pH 8.0, and eluted with 350 mM imidazole in the same buffer. The protein-containing fraction was dialysed against buffer containing 300 mM NaCl, as the solubility of UvrC CTD depends on the concentration of NaCl. After 2-fold dilution, it was loaded on a cation-exchange column and eluted with a 150 mM–1 M NaCl linear gradient. The fraction containing the protein was loaded on a high-load Superdex75 prep grade gel filtration column (APB) with a buffer containing 300 mM NaCl, 50 mM $NaPO_4$ pH 6.8. The volume of the sample was reduced by ultrafiltration (amicon) with a 1 K cut-off filter.

DNA fragments used for binding studies

dsDNA: 1, 5′-GGGCGGCGGG and 5′-CCCGCCGCCC; 2, 5′-GGGCGGGGGGCGGCGGG and 5′-CCCGCCGCCCCCCGCCGCCC; and 3, 5′-GGGCGGCGGG$(T)_8$GGCGGGGCGG and 5′-CCGCCCCG$CC(A)_8$CCCGCCGCCC. Bubble DNA: 5′-GGGCGGCGGG(T_x)GGCGGGGCGG and 5′-CCGCCCCGCC(T_x)CCCGCCGCCC, where x = 4, 6, 8 or 10. Hairpin: 5′-GGGCGGCGGG(T_x)CCCGCCGCCC, where x = 4, 6, 8, 10, 12, 16 or 20. Fork: 5′-GGGCGGCGGG(T_x) with lower strand 5′-(T_x)CCCGCCGCCC, where x = 4, 6, 8 or 10. Half fork 5′-GGGCGGCGGG(T_x) with lower strand 5′-CCCGCCGCCC where x = 4, 6, 8 or 10. ssDNA: 5′-GGGCGCGGG(T_x), where x = 4, 6, 8 or 10.

EMSAs

The oligonuceotides purchased from Life Technologies were end labelled (only the G strand e.g. 5′-GGGCGGCGGG) with [γ-^{32}P]ATP using T4 kinase and purified using a polyacrylamide gel. An equimolar amount of unlabelled ssDNA was added to the gel-purified ssDNA for the preparation of the various substrates, if required.

Binding reactions (20 µl) were performed at 4°C in buffer containing 10 mM Tris pH 7.3, 5% glycerol (w/v), 1 mM EDTA, 1 mM dithiothreitol (DTT) and 10 µg/ml bovine serum albumin (BSA). Unless otherwise indicated, a final concentration of 1 µM UvrC CTD was used. After the addition of 2000 c.p.m. of purified DNA (specific activity 10^9 c.p.m./µg), samples were incubated for 2 h on ice and loaded on a pre-run 6% (w/v) polyacrylamide gel (29:1) containing 0.5× TBE as a running buffer. The electrophoresis was carried out at 120 V for 3–4 h at 4°C. Gels were vacuum dried and exposed against a phosphoimager screen (Kodak) for at most 2 days. Quantification was performed using a Molecular Imager FX system using Quantity One software (Bio-Rad). For competion experiments, a 500-fold molar excess of unlabelled DNA was added to the reaction mixture prior to the addition of labelled bubble DNA.

Off-rate experiments were performed under identical conditions but in a larger volume (200 µl) and a UvrC CTD concentration that reaches ~80–90% binding saturation. After complex formation, a 200-fold molar excess of unlabelled DNA was added (t = 0) and, at the indicated time points, 20 µl of the sample was loaded on a constantly running gel.

Determination of the K_d and the Hill coefficient

Binding affinity was determined by the quantification of the bound and unbound DNA since [DNA] << $[protein]_{total}$, $[protein]_{unbound}$ = $[protein]_{total}$

All binding data were fitted using a non-linear regression method, against:

$$\text{Fraction bound} = 1/[1 + (K_d/[\text{protein}])^n] \quad (1)$$

where n represents the Hill coefficient, indicating the number of proteins in complex with DNA. Since complete binding of DNA was not reached for all probes at the highest protein concentration due to distortion of the complex during electrophoresis, we also fitted the binding data with:

$$\text{Fraction bound} = a/[1 + (K_d/[\text{protein}])] \quad (2)$$

where a represents the fraction of DNA that was bound at binding saturation. K_d values did not change significantly if this correction factor was omitted.

The presence of additional low affinity binding site(s) did not dramatically influence the calculated K_d for the higher affinity binding sites, although relatively large errors were obtained due to complications with the determination of binding saturation.

K_d and Hill coefficients were determined using equation (1). To confirm that the calculated Hill coefficient can also provide information about cooperativity for these non-ideally behaving DNA complexes, we compared the binding data of the most stable complex formed [with bubble (8T)]. An accurate fit requires knowledge about the K_d, fraction bound at binding saturation (a) and the Hill coefficient (n). These values could be determined combining equations (1) and (2):

$$\text{Fraction bound} = a/[1 + (K_d/[\text{protein}])^n] \quad (3)$$

Fitting of the binding data would not result in accurate determination of the variables due to the limited number of points and the steepness of the curve, together with the observed problems in determining the fraction bound at higher protein concentrations. Therefore, we took the average of three independent binding experiments and fitted these against equations (2) or (3) to determine the K_d and a. These values were then used in equation (3), and simulated curves were calculated for n = 1, n = 2 and n = 4, and plotted together with the average and SDs for the calculated fraction bound.

Molecular weight determination by gel filtration

For gel filtration, an analytical Superdex 75 HR 10/30 (APB) was equilibrated with NMR buffer at 4°C using a Pharmacia FPLC. A 100 µl aliquot of UvrC CTD (10 nmol), bubble DNA [5′-GGGCGG$CGGG(T)_{10}$GGCGGGGCGG and its complement, 5 nmol] or protein–DNA complex, incubated for 2 h on ice (5 and 10 nmol for DNA and UvrC CTD, respectively), were injected. For calibration, a mixture of BSA, ovalbumin, carbonic anhydrase, cytochrome *c* and aprotinin (Sigma) was used. Each sample was injected twice and mass was determined by fitting the elution time with the mass of the indicated proteins.

DNA titration

The oligonucleotides used for NMR measurements were purchased from RNA-TEC Belgium. The 22 bp oligonucleotides used for NMR has the sequence 5′-GGAAGATC$(T)_6$ACCGTACG; 5′-CCTTCTAG$(T)_6$TGGCATGC.

The DNA samples were annealed and dialysed in 5 mM sodium phosphate and 30 mM NaCl pH 6.8, so that the final concentration of the buffer becomes the same as that for protein. The 2D ^{1}H–^{15}N HSQC spectra of UvrC CTD with 0.024, 0.032, 0.048, 0.060, 0.068, 0.083, 0.12, 0.19, 0.22, 0.25, 0.45, 0.5, 0.60, 0.8, 1.0 and 2.0 equivalents of DNA were recorded at 20°C.

NMR measurements

NMR experiments were carried out at 27°C on a Bruker DRX 600 MHz instrument and Varian Inova 750 MHz. For the backbone resonance assignments, 3D-HNCO, CBCANH, CBCA(CO)NH, NOESY-(^{15}N,^{1}H)-HSQC, TOCSY-(^{15}N,^{1}H)-HSQC (as described in Cavanagh *et al*., 1996) spectra were recorded, and for the side chain resonance assignments, 3D-H(C)CH TOCSY and (H)CCH TOCSY spectra were recorded. NOE distance constraints were obtained from 2D-NOESY, 3D-NOESY(^{15}N,^{1}H)-HSQC and 3D-NOESY-(^{13}C,^{1}H)-HSQC spectra with mixing times of 70 and 150 ms. All NMR spectra were processed using the software package NMRPIPE (Delaglio *et al*., 1995) and analysed using NMRView (Johnson and Blevins, 1994).

Chemical shift-derived restraints

The chemical shifts of 97% of the nuclei were determined (BMRB accession No. 5217). The Cα, Cβ, C_O, Hα and N chemical shifts of 56 residues served as input for the TALOS program (Cornilescu *et al*., 1999). TALOS derives information on the Φ and Ψ backbone dihedral angles from a comparison of secondary chemical shift patterns of amino acid triplets against a database of secondary chemical shifts corresponding to known conformations. A conservative approach was chosen requiring that all 10 best matches agree for a prediction to be accepted. The TALOS predictions were converted into dihedral angle restraints as the average Φ and Ψ angles ± 2 SD or a minimum of ±10°.

Structure calculations

Distance restraints were obtained from 2D-NOESY and 3D-^{15}N- and ^{13}C-edited NOESY experiments (mixing time 150 ms). TALOS-derived dihedral angle constraints were used as described above. Structure calculations were performed (see methods in Bonvin *et al*., 2001) with CNS (Brünger *et al*., 1998) using the ARIA setup and simulated annealing protocols (Nilges and O'Donoghue, 1998; Linge and Nilges, 1999). The best 22 structures were selected based on the lowest total restraint energy.

Model for DNA binding

In the model structure of the complex, the ss–ds junction coordinate of the nucleotide was taken from the PDB 1I6H of the RNA polymerase II–DNA complex (Cramer *et al*., 2001). In the RNA polymerase II–DNA complex, the two loops of the first domain (RPB1) were interacting with the nucleotide junction. When we superimposed the UvrC CTD hairpin loops with the loops of the RPB1 domain that interact with the junction, an r.m.s.d. of <1.6 Å was obtained. Also, the distance between the two loops of the RPB1 was comparable with that between the two hairpin loops of the UvrC CTD.

Coordinates

The PDB accession No. for the coordinates is 1KFT.

Acknowledgements

This work was financially supported by the Centre for Biomedical Genetics.

References

Aravind,L., Walker,D.R. and Koonin,E.V. (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.*, **27**, 1223–1242.

Ariyoshi,M., Nishino,T., Iwasaki,H., Shinagawa,H. and Morikawa,K. (2000) Crystal structure of the Holliday junction DNA in complex with a single RuvA tetramer. *Proc. Natl Acad. Sci. USA*, **97**, 8257–8262.

Bonvin,A.M.J.J., Houben,K., Guenneugues,M., Kaptein,R. and Boelens,R. (2001) Rapid protein fold determination using secondary chemical shifts and cross-hydrogen bond ^{15}N–^{13}C scalar couplings ($^{3hb}J_{NC'}$). *J. Biomol. NMR*, **21**, 221–233.

Brünger,A.T. *et al*. (1998) Crystallography a NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.

Cavanagh,J., Fairbrother, WJ., Palmer,A.G.,III and Skelton,N.J. (1996) *Protein NMR Spectroscopy*. Academic Press, San Diego, CA.

Cornilescu,G., Delaglio,F. and Bax,A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.

Cramer,P., Bushnell,D.A. and Kornberg,R.D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science*, **292**, 1863–1876.

Dekker,N., Cox,M., Boelens,R., Verrijzer,C.P., van der Vliet,P.C. and Kaptein,R. (1993) Solution structure of the POU-specific DNA-binding domain of Oct-1. *Nature*, **362**, 852–855.

Delaglio,F., Grzesiek,S., Vuister,G., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX PIPES. *J. Biomol. NMR*, **6**, 277–293.

Doherty,A.J., Serpell,L.C. and Ponting,C.P. (1996) The helix–hairpin–helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res.*, **24**, 2488–2497.

Esnouf,R.M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.*, **15**, 132–134

Holm,L. and Sander C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.

Johnson,B.A and Blevins,R.A. (1994) NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.

Koradi,R., Billeter,M. and Wúthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55.

Lee,Y.L., Chang,C., Song,H.K., Moon,J., Yang,J.K., Kwon,S.T. and Suh,S.W. (2000) Crystal structure of NAD(+)-dependent DNA ligase: modular architecture and functional implications. *EMBO J.*, **19**, 1119–1129.

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

Linge,J.P. and Nilges,M. (1999) Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. *J. Biomol. NMR*, **13**, 51–59.

Moolenaar,G.F., Franken,K.L.M.C., Dijkstra,D.M., Thomas-Oates,J.E., Visse,R., van de Putte,P. and 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,G.F., Uiterkamp,R.S., Zwijnenburg,D.A. and Goosen,N. (1998) 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.

Moolenaar,G.F., Monaco,V., van der Marel,G.A., van Boom,J.H., Visse,R. and Goosen,N. (2000) The effect of the DNA flanking the lesion on formation of the UvrB–DNA preincision complex. Mechanism for the UvrA-mediated loading of UvrB onto a DNA damaged site. *J. Biol. Chem.*, **275**, 8038–8043

Nilges,M. and O'Donoghue,S. (1998) Ambiguous NOEs and automated NOE assignment. *Prog. NMR Spectrosc.*, **32**, 107–119.

Pelletier,H. and Sawaya,M.R. (1996) Characterization of the metal ion binding helix–hairpin–helix motifs in human DNA polymerase β by X-ray structural analysis. *Biochemistry*, **35**, 12778–12787.

Pierce,J.R., Case,R. and Tang,M.S. (1989) Recognition and repair of 2-aminofluorene– and 2-(acetylamino)fluorene–DNA adducts by UVRABC nuclease. *Biochemistry*, **28**, 5821–5826.

Rafferty,J.B. *et al*. (1996) Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. *Science*, **274**, 415–421.

Roe,S.M., Barlow,T., Brown,T., Oram,M., Keeley,A., Tsaneva,I.R. and Pearl,L.H. (1998) Crystal structure of an octameric RuvA–Holliday junction complex. *Mol. Cell*, **2**, 361–372.

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

Schägger,H., Cramer,W.A. and von Jagow,G. (1994) Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein

complexes by two-dimensional native electrophoresis. *Anal. Biochem.*, **217**, 220–230

Shao,X. and Grishin,N.V. (2000) Common fold in helix–hairpin–helix proteins. *Nucleic Acids Res.*, **28**, 2643–2650.

Shi,Q., Thresher,R., Sancar,A. and Griffith,J. (1992) Electron microscopic study of (A)BC excinuclease. DNA is sharply bent in the UvrB–DNA complex. *J. Mol. Biol.*, **226**, 425–432.

Takahashi,M., Bertrand-Burggraf,E., Fuchs,R.P. and Norden,B. (1992) Structure of UvrABC excinuclease–UV-damaged DNA complexes studied by flow linear dichroism. DNA curved by UvrB and UvrC. *FEBS Lett.*, **314**, 10–12.

Tang,M.S., Nazimiec,M., Ye,X., Iyer,G.H., Eveleigh,J., Zheng,Y., Zhou,W. and Tang,Y.Y. (2001) Two forms of UvrC protein with different double-stranded DNA binding affinities. *J. Biol. Chem.*, **276**, 3904–3910.

Thayer,M.M., Ahern,H., Xing,D., Cunningham,R.P. and Tainer,J.A. (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *EMBO J.*, **14**, 4108–4120.

Velankar,S.S., Soultanas,P., Dillingham,M.S., Subramanya,H.S. and Wigley,D.B. (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell*, **97**, 75–84.

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

Verhoeven,E.E., Wyman,C., Moolenaar,G.F., Hoeijmakers,J.H. and Goosen,N. (2001) Architecture of nucleotide excision repair complexes: DNA is wrapped by UvrB before and after damage recognition. *EMBO J.*, **20**, 601–611

Visse,R., King,A., Moolenaar,G.F., Goosen,N. and van de Putte,P. (1994) Protein–DNA interactions and alterations in the DNA structure upon UvrB–DNA preincision complex formation during nucleotide excision repair in *Escherichia coli*. *Biochemistry*, **33**, 9881–9888.

Wada,T., Yamazaki,T. and Kyogoku,Y. (2000) The structure and the characteristic DNA binding property of the C-terminal domain of the RNA polymerase α subunit from *Thermus thermophilus*. *J. Biol. Chem.*, **275**, 16057–16063.

Zou,Y. and Van Houten,B. (1999) Strand opening by the UvrA(2)B complex allows dynamic recognition of DNA damage. *EMBO J.*, **18**, 4889–4901.

Zou,Y., Walker,R., Bassett,H., Geacintov,N.E. and Van Houten,B. (1997) Formation of DNA repair intermediates and incision by the ATP-dependent UvrB–UvrC endonuclease. *J. Biol. Chem.*, **272**, 4820–4827.

Received January 21, 2002; revised September 25, 2002; accepted September 30, 2002