Crystal structure of *Escherichia coli* UvrB C-terminal domain, and a model for UvrB-UvrC interaction

Maninder Sohi[a,1], Alexander Alexandrovich[a,1], Geri Moolenaar[b], Rob Visse[b], Nora Goosen[b], Xavier Vernede[c], Juan C. Fontecilla-Camps[c], John Champness[a], Mark R. Sanderson[a,*]

[a] *The Randall Institute, King's College, 26–29 Drury Lane, London WC2B 5RL, UK*
[b] *Molecular Genetics, Leiden Institute of Chemistry Gorlaeus Laboratories, Leiden Univerisity, Postbus 9502, 2300 RA Leiden, The Netherlands*
[c] *CEA, CNRS, Institut de Biologie Structurale Jean-Pierre Ebel, Laboratoire de Cristallographie et Cristallogenese, F-38027 Grenoble 1, France*

Received 25 October 1999; received in revised form 1 December 1999

Edited by Hans Eklund

Abstract A crystal structure of the C-terminal domain of *Escherichia coli* UvrB (UvrB′) has been solved to 3.0 Å resolution. The domain adopts a helix-loop-helix fold which is stabilised by the packing of hydrophobic side-chains between helices. From the UvrB′ fold, a model for a domain of UvrC (UvrC′) that has high sequence homology with UvrB′ has been made. In the crystal, a dimerisation of UvrB′ domains is seen involving specific hydrophobic and salt bridge interactions between residues in and close to the loop region of the domain. It is proposed that a homologous mode of interaction may occur between UvrB and UvrC. This interaction is likely to be flexible, potentially spanning >50 Å.

Key words: Nucleotide excision repair; X-ray crystallography; UvrB protein; UvrB-C interaction

1. Introduction

The *Escherichia coli* UvrABC endonuclease system catalyses repair of DNA damage by recognising and removing many types of DNA lesions. The enzymatic mechanism consists of a $UvrA_2B$ complex recognising, then binding UvrB to a DNA lesion with the release of $UvrA_2$. The UvrB-DNA assembly, the so-called pre-incision complex, then recruits UvrC, and incisions to the DNA strand containing the lesion are made. The UvrBC complex hydrolyses the fourth or fifth phosphodiester to the 3′ side of the lesion followed by cutting of the eighth phosphodiester bond to the 5′ side of the lesion [1–3]. It has been shown that the presence of the C-terminal domain of UvrB is essential for the 3′ incision but not for the 5′ incision [4]. The C-terminal domain of UvrB is probably flexible with respect to the rest of the protein. The flexibility most likely stems from a linker region which is susceptible to proteolytic cleavage, this linker region is poorly conserved among UvrBs [3]. In *E. coli*, this region is cleaved by OmpT at position 607 or 609 leading to the formation of a truncated form of the protein called UvrB* [4]. This form of the protein is perfectly capable of forming a pre-incision complex on damaged DNA. The addition of UvrC to a UvrB* pre-incision complex with a double-stranded damaged DNA substrate does not lead to either 5′ or 3′ incision. However, on a 3′ pre-nicked DNA substrate, incubation of the UvrB* pre-incision complex with UvrC results in proficient 5′ incision, indicating that the presence of the UvrB C-terminal domain is only essential for the 3′ incision [4].

In solution, UvrB and UvrC form a complex [5], which is stabilised by the C-terminal domain of UvrB [5]. A damage-independent endonuclease activity is displayed by UvrBC and analysis of mutant UvrB and UvrC proteins revealed that this activity also requires the C-terminal domain of UvrB [5]. Residues 634–668 of UvrB have a high degree of sequence homology with residues 205–239 of UvrC. Biochemical data and mutational analyses have shown that these homologous regions are important for the interaction between UvrB and UvrC [4,6]. The introduction of a F652L mutation in UvrB resulted in a loss of 3′ incision, as seen for UvrB* [4]. Introduction of the identical mutation in the homologous region of UvrC had the same effect, corroborating the suggestion that the domains interact [4]. The C-terminal domain of UvrB was found to have a high score for a coiled-coil structure [4]. Nuclear magnetic resonance (NMR) analysis later demonstrated that this domain exhibits a two-helix structure [7].

This paper reports a crystallographic analysis of the C-terminal domain of UvrB which confirms the NMR structure of the domain as a two-helix motif capable of self-association. A model for the homologous domain of UvrC satisfies the sequence conservation between the two domains, and suggests a model for the interaction between UvrB and UvrC.

2. Materials and methods

2.1. Crystallisation and data collection

The cloning, over-expression and purification of a MH_6L-tagged 55 residue fragment of UvrB has been described previously for the preparation of the NMR samples [7]. Protein incorporating Se-Met was obtained by transforming the auxotrophic strain *E. coli* PP3398, which was constructed as a *metA* derivative of BL21:DE3 (Molecular Genetics Lab), with overproduction plasmid and grown on selenomethionine media [8,9]. Crystals in several crystal forms were grown using the hanging drop vapour diffusion method. Crystals in space group $P6_2$ (cell dimensions $a=b=84.81$ Å, $c=53.19$ Å) were initially grown at 5°C and 18–21°C from drops containing 4 μl of 2 mg/ml protein solution in 20 mM Tris–HCl, 150 mM NaCl, 0.1% NaN_3 pH 7.0 and 4 μl of reservoir solution containing 35–45% saturated ammonium sulphate, 100 mM Tris–HCl, 0.1% NaN_3 pH 8.8–9.4, and equilibrated against the reservoir solution. The largest crystals used

*Corresponding author. Fax: (44)-171-497 9078.
E-mail: mark.sanderson@kcl.ac.uk

[1] These authors contributed equally to the work.

Abbreviations: NMR, nuclear magnetic resonance; MAD, multiple anomalous dispersion

PII: S0014-5793(99)01690-7

for data collection and phasing were grown at 5°C from drops containing 9 μl protein solution at 2 mg/ml in 20 mM Tris–HCl, 150 mM NaCl, 0.1% NaN_3 pH 7.0 mixed with 3 μl of reservoir solution containing 35% saturated ammonium sulphate, 100 mM Tris–HCl, 0.1% NaN_3 pH 8.8. Another hexagonal crystal form with cell dimensions ($a = b = 38.64$ Å, $c = 145.05$ Å) was grown between 18 and 21°C by equilibration of the 3 μl of the afore-mentioned protein solution mixed with 3 μl of reservoir solution containing 2 M ammonium sulphate, 5% isopropanol, unbuffered at pH 6.47 and equilibrated against this solution. Reservoir volumes of 500 μl were used throughout all the crystallisations.

Several strategies for crystal growing and cryo-cooling were adopted, crystals were grown both at room temperature and in the cold room at 5°C. Cryo-cooling directly into liquid nitrogen and also cryo-cooling into liquid propane cooled to liquid nitrogen temperatures [10] using appropriate cryo-protecting solutions were tried. The latter method using the largest cold room grown crystals was found to be most successful provided that extreme care was taken not to allow moisture from the cold room air to be deposited onto the crystal. For multiple wavelength data collection on the ESRF MARCCD on beamline BM14, 32 crystals were mounted and exposed until a crystal was obtained with suitable mosaicity, diffracting properties, whose intensities could be integrated. Measurement of the crystal fluorescence spectrum identified suitable wavelengths about the Se-edge for the multiple anomalous dispersion (MAD) data collection.

2.2. Structure solution and refinement

The MAD data sets were processed using HKL suite [11]. The selenium sites were located using SOLVE [12]. Treating the MAD data as MIR data [13], electron density maps were generated and solvent-flattened using programs from the CCP4 suite [14]. The protein structure was modelled using XFIT from the Xtalview suite [15]. Initially, the structure was refined using XPLOR [16,17], and subsequently refined using SHELXL97 [18] applying differently weighted NCS restraints between regions of the homodimer and a tight restraint to minimise the deviation of isotropic temperature factors between bonded atoms. No water atoms could be clearly identified.

2.3. Homology modelling

The sequence of the C-terminal domain of UvrB whose structure is described in this paper and its alignment with the sequence of UvrC are shown in Fig. 1. The homologous domains will be named UvrB′ and UvrC′. The region of UvrC with sequence homologous to the structure was made into a homology model using InsightII (Molecular Simulation, San Diego, CA, USA; Version 97.2).

3. Results

3.1. Crystal structure

The crystal form possesses $P6_2$ symmetry, and a homodimer of the UvrB C-terminal domain is observed in the asymmetric unit cell (Fig. 2A). A part of the electron density map corresponding to the loop region is shown in Fig. 2B. The crystal-

```
UvrB
    631       641         651         661         671
SPK ALQQKIHELE GLMMQHAQNL EFEEAAQIRD QLHQLRELFI AAS
      : . .:  . :   .::: ::::::.::: :.. .:..
LSG KDDQVLTQLI SRMETASQNL EFEEAARIRD QIQAVRRVTE KQF
    202         212         222         232         242
UvrC
```

Fig. 1. Alignment of the C-terminal domain of UvrB with the sequence of UvrC.

lographic statistics are given in Table 1. Defined regions of electron density correspond to residues 628–673 of the UvrB sequence. All residues were modelled to the map generated from the solvent-flattened map phases, the refinement of the structure did not significantly contribute to better phased electron density maps. Some atoms on exposed side-chains were not found and left out of the structure (K630, Q634, K635, I636, E638, L642, Q645, Q648, I658, Q661, I670). The relatively high crystallographic R-factor of 28.6% for the refined structure can partly be explained in that one quarter of the protein is disordered, which includes MH_6L tail and nine residues up to S628 in the N-terminus of this domain, and a very high solvent content $V_m = 3.71$.

3.2. UvrB C-terminal domain

The two-helix fold (Fig. 3A) of the protein confirms an earlier NMR secondary structure analysis [7]. The main chain torsional angles are represented by a Ramachandran plot (Fig. 3B). Residues which deviate from torsional ideality are found in the tight loop region (N649-L650-E651-F652-E653). The association of the two helices forms a hydrophobic core which stretches the full length of the monomer (Fig. 3C). Salt bridges exist between R659 and D660; E640 and R666 on the surface of the domain.

3.3. Dimer structure of UvrB′ and model for UvrB′ and UvrC′ interaction

In this crystal structure, the loop regions of the two monomers associate end-to-end (Fig. 4A). A core of hydrophobic residues is packed together, reducing the solvent accessible surface of a monomer by ~11% (~333 $Å^2$). The residue contacts, related by non-crystallographic symmetry, important to dimerisation include ^{Se}M643 and L650 (A647, A655, A656) and F652, and a salt bridge between E653 and R659. Alternative monomer-monomer interactions arise out of crystal packing, however, the specificity of residue contacts and the 3-fold lower thermal displacements of residues in the loop

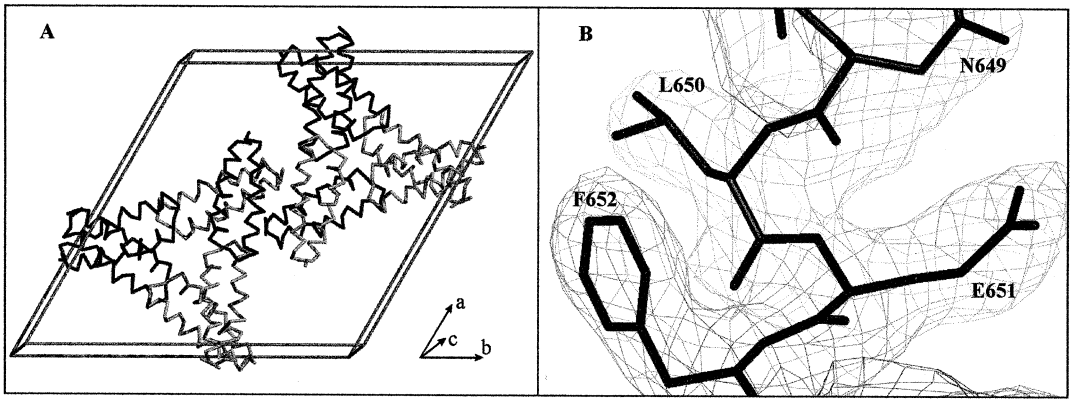

Fig. 2. Crystal packing and electron density map of the C-terminal domain of UvrB. A: Unit cell crystal packing. B: Fragment of the electron density map for the loop region.

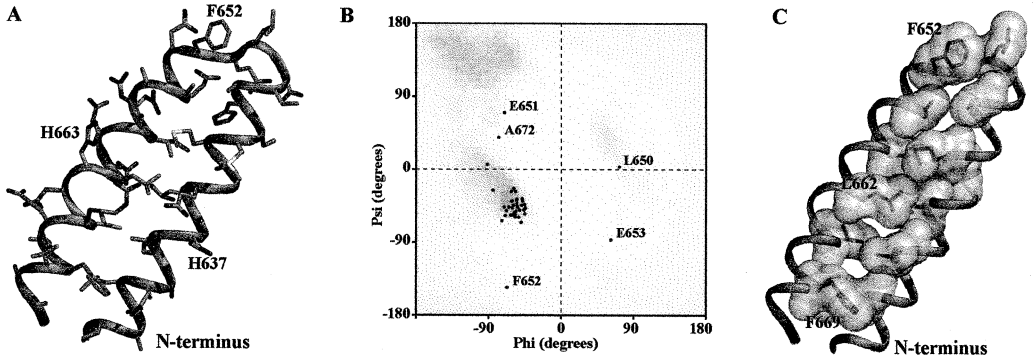

Fig. 3. Crystal structure of the C-terminal domain of UvrB. A: Three-dimensional structure of the C-terminal domain of UvrB. B: Ramachandran plot. C: Hydrophobic core of the C-terminal domain of UvrB.

region favour the model that the loop contacts between monomers are significant.

Structure alignment of a homologous domain of UvrC (UvrC′) with the helix-loop-helix motif of UvrB′ show that hydrophobicity of the core and loop region would be completely preserved. If the UvrB′ dimer interactions as seen in the crystal are taken as a model for dimerisation of the homologous domains in UvrB and UvrC (Fig. 4B), then, in addition to the hydrophobic residues in the loop, E653 and R659 of UvrB could make salt bridges with R230 and E224 of UvrC.

4. Discussion

From the structure of the UvrB′ domain, a model for its homologue UvrC′ has been made. The extent of the helices expected in UvrC′ may be slightly shorter than those observed in UvrB′. Considering the high conservation of specific residues in the loop region, a convincing model for UvrB and UvrC association is proposed from the UvrB′ dimer interaction via loop regions as seen in the crystal. Favouring this type of interaction, as opposed to the formation of a four-helical bundle, is the 3-fold higher ratio of positively charged side-chains found in the helical regions of UvrB′ and UvrC′ model. Further evidence comes from NMR experiments and measurement of the UvrB′ Stokes radius [19], which also indicate that the presence of ^{Se}Met residues is not significant. Mutational analyses supporting this type of interaction include the amino acid substitutions F652L in UvrB and F223L in UvrC which have both been shown to disturb the interaction between UvrB and UvrC [6]. However, a strain with an R659A mutation in UvrB is still UV-resistant, indicating a functional interaction between UvrB(659A) and UvrC [20]. R659 is one of the six residues involved in UvrB′ dimerisation and, according to the model proposed, for UvrB′-UvrC′ association. The R659A mutation would be expected to reduce the dimerisation affinity but apparently does not altogether abolish it.

In the crystal structure of UvrB from *Thermus thermophilus*, where its C-terminal domain is not observed [21], and the OmpT proteolysis of UvrB suggest that the UvrB′ domain exists at the end of a flexible region of protein. The significance of this flexibility is not altogether clear but may favour correct positioning of UvrC for 3′ incision, or conformational

Table 1
Crystallographic statistics

	Reciprocal space						
	Unit cell dimensions					$a = 84.81$ Å $c = 53.19$ Å	
	Wavelength	0.9789 Å		0.9788 Å		0.8856 Å	
	Resolution	20.0–3.0 Å		20.0–3.0 Å		20.0–3.0 Å	
Total *hkl*	Completeness	65307	100%	44863	100%	65208	100%
Unique *hkl*	$\|F(hkl)\| > 4\sigma$	4475	74.3%	4420	65.3%	4450	74.1%
	R_{merge}		9.6%		11.6%		10.2%
	Phasing	(treated as native)					
Phasing power	acentric/centric			0.50/0.35		1.17/0.77	
R_{Cullis}	acentric/centric			0.93/0.90		0.82/0.72	
	Mean FOM	0.511[a]	0.815[b]				
	Refinement						
R-factor	$F_o hkl > 4\sigma$ used					0.286	6381
Free *R*-factor	$F_o hkl > 4\sigma$ used					0.325	327
Number of parameters/restraints						3031/3887	
Bond RMSD	Angle RMSD					0.006 Å	0.020°

$R_{merge} = \Sigma_{hkl}\Sigma i|\langle I(hkl)\rangle - I(hkl)_i|/\Sigma_{hkl}\langle I(hkl)\rangle$ where I is intensity of a reflection. Friedel pairs are kept separate.
$R_{anom} = R_{merge}$ for Friedel pairs only.
Phasing power $= \langle \mathbf{F}_H\rangle - \langle E\rangle$ where $\mathbf{F}_H$ is the heavy atom structure factor and E is the lack of closure.
$R_{Cullis} = \langle E\rangle/\langle\Delta\rangle$ where E is the lack of closuree error and Δ is the isomorphous difference.
The crystallographic *R*-factor is defined $\Sigma||\mathbf{F}_o| - |\mathbf{F}_c||/\Sigma|\mathbf{F}_o|$ and indicates the accuracy of the model.
The free *R*-factor is a cross-validation residual using 5% of the native data, randomly chosen and excluded from the refinement.
[a]Before solvent flattening.
[b]After solvent flattening.

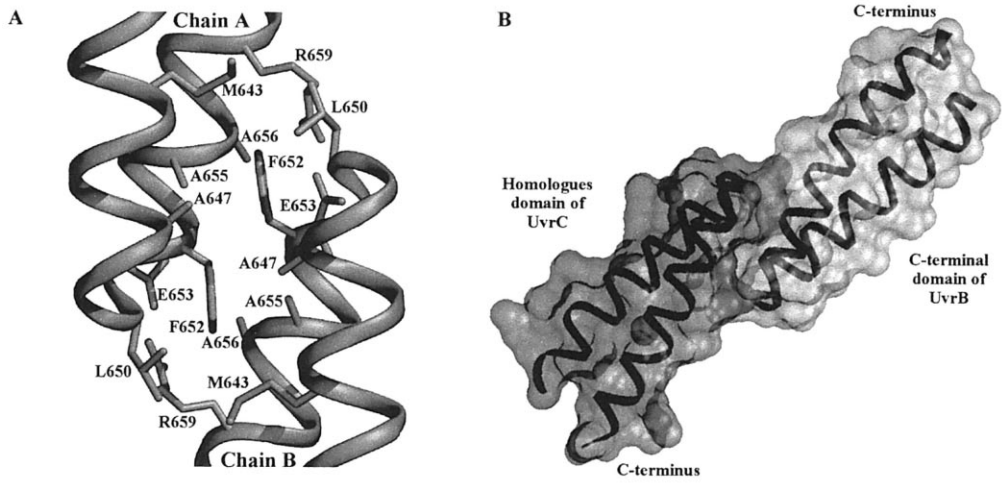

Fig. 4. Importance of the loop region in the oligomerisation mechanism. A: Dimerisation interface between C-terminal domains of UvrB. B: Model for dimerisation between the homologous domains of UvrB and UvrC.

regulation of the UvrBC-DNA incision complex. The most striking feature of the model proposed for UvrB′-C′ interaction is the large distance covered, potentially spanning >50 Å.

Acknowledgements: This work was supported by an EU Structural Biology Framework IV Programme grant. Mr. Matthew Bennett is thanked for his contribution to MAD phasing, refinement and model-building. Data on liquid N_2 frozen crystals were collected and cryo-freezing problems were investigated during time on NSLS beamlime X12B, Drs Malcolm Capel and Klaus Schroer are thanked for their help. Multiple wavelength data were collected on ESRF BM14 and Drs Vivian Stayanov and Gordon Leonard are thanked for their assistance during data collection. Mr. Tom Rutherford is thanked for excellent X-ray technical assistance. Further thanks to Dr Andrew Lane, who initially proposed the UvrB-UvrC dimer on the basis of NMR and ultracentrifugation experiments in solution on UvrB.

References

[1] Sancar, A. (1998) Ann. Rev. Biochem. 65, 43–65.
[2] van Houten, B. (1990) Microbiol. Rev. 54, 18–51.
[3] Goosen, N., Moolenaar, G.F., Visse, R. and van de Putte, P. (1998) in: Nucleic Acids and Molecular Biology (Eckstein, F. and Lilley, D., Eds).
[4] 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) J. Biol. Chem. 270, 30508–30515.
[5] Moolenaar, G.F., Bazuine, M., van Knippenberg, I.C., Visse, R. and Goosen, N.J. (1998) J. Biol. Chem. 273, 34896–34903.
[6] Moolenaar, G.F., Franken, K.L.M.C., van de Putte, P. and Goosen, N. (1998) Mut. Res. 385, 195–203.
[7] Alexandrovich, A., Sanderson, M.R., Goosen, N., Moolenaar, G.F. and Lane, A.N. (1999) FEBS Lett. 451, 181–185.
[8] Hendrickson, W.A., Horton, J.R. and LeMaster, D.M. (1990) EMBO J. 9, 1665–1672.
[9] Doublie, S. (1997) Methods Enzymol. 276, 523–530.
[10] Vernede, X. and Fontecilla-Camps, J.C. (1999) J. Appl. Cryst. 32, 505–509.
[11] Otwinowski, Z. and Minor, W. (1994) Methods Enzymol. 276, 307–326.
[12] Terwilliger, T.C. and Berendzen, J. (1999) Acta Cryst. D 55, 849–861.
[13] Ramakrishnanan, V. and Biou, V. (1997) Methods Enzymol. 276, 538–557.
[14] Collaborative Computational Project Number 4 (1994) Acta Crystallogr. D50, pp. 760–763.
[15] McCree, D.E. (1999) J. Struct. Biol. 125, 156–165.
[16] Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) Science 235, 458–460.
[17] Brünger, A.T (1992) Nature 355, 472–474.
[18] Sheldrick, G.M. and Schneider, T.R. (1997) Methods Enzymol. 277, 319–343.
[19] Alexandrovich, A., Czisch, M., Frenkiel, T.A., Sanderson, M.R., Moolenaar, G.F., Goosen, N. and Lane, A.N., unpublished work.
[20] Lin, J.J., Phillips, A.M., Hearst, J.E. and Sancar, A.J. (1992) Biol. Chem. 267, 17693–17700.
[21] Machius, M., Henry, L., Palnitkar, M. and Deisenhofer, J. (1999) Proc. Natl. Acad. Sci. USA 96, 11717–11722.