Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Domain closure and action of uracil DNA glycosylase (UDG): structures of new crystal forms containing the *Escherichia coli* enzyme and a comparative study of the known structures involving UDG

K. Saikrishnan,[a] M. Bidya Sagar,[a] R. Ravishankar,[a] S. Roy,[b] K. Purnapatre,[b] P. Handa,[b] U. Varshney[b] and M. Vijayan[a]*

[a]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India, and [b]Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560012, India

Correspondence e-mail: mv@mbu.iisc.ernet.in

Received 9 January 2002
Accepted 28 May 2002

PDB References: UDG, 1lqj, r1lqjsf; UDG–Ugi, form I, 1lqg, r1lqgsf; UDG–Ugi, form II, 1lqm, r1lqmsf.

The structures of a new crystal form of free *Escherichia coli* uracil DNA glycosylase (UDG), containing four molecules in the asymmetric unit, and two forms of its complex with the proteinaceous inhibitor Ugi, containing two and four crystallographically independent complexes, have been determined. A comparison of these structures and the already known crystal structures containing UDG shows that the enzyme can be considered to be made up of two independently moving structural entities or domains. A detailed study of free and DNA-bound human enzyme strengthens this conclusion. The domains close upon binding to uracil-containing DNA, whereas they do not appear to do so upon binding to Ugi. The comparative study also shows that the mobility of the molecule involves the rigid-body movement of the domains superposed on flexibility within domains.

1. Introduction

Uracil DNA glycosylase (UDG) is an important DNA-repair enzyme which recognizes and removes uracil occurring in DNA owing to either incorporation of dUMP or spontaneous deamination of cytosine and initiates the base-excision repair pathway (David & Williams, 1998; McCullough *et al.*, 1999). UDG was the first DNA glycosylase to be discovered and purified from *E. coli* (Lindahl, 1974; Lindahl *et al.*, 1977) and subsequently cloned (Duncan & Chambers, 1984) and sequenced (Varshney *et al.*, 1988). Although *E. coli* UDG (*Ec*UDG) has served as a prototype for much of the biochemical studies, enzymes from herpes simplex virus type-1 and human were the first to be X-ray analysed (Savva *et al.*, 1995; Savva & Pearl, 1995; Mol, Arvai, Sanderson *et al.*, 1995; Mol, Arvai, Slupphaug *et al.*, 1995; Slupphaug *et al.*, 1996; Parikh *et al.*, 1998). The X-ray determination of the *E. coli* enzyme in complex with the proteinaceous inhibitor Ugi encoded by *Bacillus subtilis* phage (PBS-1 or PBS-2) was first reported by our group (Ravishankar *et al.*, 1998). Subsequently, X-ray studies on the enzyme and several complexes were reported from two other laboratories (Putnam *et al.*, 1999; Xiao *et al.*, 1999; Drohat *et al.*, 1999; Werner *et al.*, 2000). The three-dimensional structures and the sequences show UDG to be a highly conserved enzyme (Pearl, 2000). The structural and other studies have also led to the proposal of a common pinch–push–pull mechanism for the detection and flipping out of uracil in the substrate for repair (Parikh *et al.*, 1998). According to this proposal, the initial lesion detection occurs *via* a pinch mechanism involving three Ser-Pro-rich

loops. Subsequently, the damaged nucleotide is flipped out through the DNA major groove into the enzyme active site by a push–pull mechanism. The Leu intercalation loop, on productive binding, moves about 4 Å, allowing the leucine side chain to penetrate the DNA minor groove, causing the complete flipping of uridine. This pushing is accompanied by a pull by the binding pocket that specifically recognizes the uracil nucleotide by complementary active-site residues. Recently, an additional role for this leucine has been suggested in stabilization of the productive enzyme–substrate complex (Handa *et al.*, 2001). The closing of the active site which occurs upon substrate binding distorts uracil and deoxyribose in the flipped-out nucleotide. The resulting strain is relieved by glycosylic bond cleavage (Parikh, Walcher *et al.*, 2000).

Part of the work reported here is concerned with the structures of a new crystal form of *Ec*UDG and two forms of its complex with Ugi. One of the two crystal forms of the complex is new. The other is the same as reported previously, but diffracts to higher resolution with more residues defined. The crystal structure of the free enzyme contains four molecules in the asymmetric unit, while there are two and four crystallographically independent molecules in the two forms of the complex. The new crystal form of the free enzyme has the highest solvent content observed in crystals containing UDG. Consequently, the resolution of the data is comparatively poor, but the molecules in it are less constrained by packing interactions when involved in domain or segmental motions. With the structure determinations reported here, structural data on 20 crystallographically independent *Ec*UDG molecules, free as well as complexed with different ligands in different environments, are available. They provide a rich resource for analysing the mobility of the molecule and its functional significance. Such an analysis, presented here in addition to the new structures, has yielded interesting results. The conclusions are further strengthened by a detailed study of the known structures of the free, Ugi-bound and DNA-bound human enzyme.

2. Materials and methods

2.1. Crystallization

Free UDG was overexpressed and purified from pTrcUng (Handa *et al.*, 2001). The UDG–Ugi complex was purified from a clone (pTrcUDG:Ugi1) containing the UDG and the Ugi genes in tandem downstream of the P_{trc} promoter with TAG as the termination codon for the UDG ORF. The plasmids were overexpressed in *E. coli* TG1 carrying *supE*, which results in low-level suppression of the UAG codon. The UDG–Ugi complex whose structure was reported previously was expressed from a similar construct which contained TAA as the termination codon at the end of the UDG ORF (Roy *et al.*, 1998). Crystallization was carried out using the hanging-drop method. Crystals of free UDG were grown by equilibrating 5 mg ml^{-1} protein in 20 m*M* Tris–HCl buffer pH 7.4 containing 100 m*M* NaCl against 15%(*w*/*v*) PEG 8000 in the same buffer at 293 K. Crystals grew in 15 d to dimensions of 0.16 × 0.04 × 0.04 mm. Two crystal forms of the complex were obtained from two different crystallization conditions. Crystals of form I were grown from 10 mg ml^{-1} protein in 20 m*M* Tris–HCl buffer pH 7.6, with 10%(*w*/*v*) PEG 1250 as the precipitant at 293 K. Crystals of maximum dimensions 0.12 × 0.10 × 0.08 mm took 32 weeks to grow. A solution of 10 mg ml^{-1} protein in 20 m*M* imidazole–maleate buffer pH 7.6 with 10%(*w*/*v*) PEG 4000 as the precipitant yielded crystals of form II. The setup was initially incubated at 281 K for eight weeks before being transferred to 293 K. After a period of 16 weeks, crystals of dimensions 0.14 × 0.10 × 0.12 mm were obtained. In all cases crystals appeared only very infrequently.

Table 1
Data-collection statistics.

Values in parentheses refer to the final resolution shell.

	Free	UDG–Ugi (form I)	UDG–Ugi (form II)
Space group	$P3_2$	$P2_12_12_1$	$P2_12_12$
Unit-cell parameters			
a (Å)	125.21	51.16	98.76
b (Å)	125.21	89.90	158.87
c (Å)	90.05	141.78	91.22
Z	12	8	16
Solvent content (%)	67.8	45.0	50.1
Data resolution (Å)	3.35	2.90	3.20
Final resolution shell	3.47–3.35	3.00–2.90	3.31–3.20
Total No. of reflections	38941	51769	94765
No. of unique reflections	20394	14803	21106
Completeness of data (%)	90.2 (94.8)	98.0 (98.2)	87.6 (79.3)
R_{merge} (%)	15.4 (36.9)	11.0 (37.8)	16.7 (39.8)
Average $I/\sigma(I)$	6.9 (1.4)	7.1 (1.5)	9.2 (1.4)

2.2. Data collection

Intensity data from all three crystals were collected at 293 K on a MAR Research imaging plate with X-rays generated by a Rigaku rotating-anode generator operating at 40 kV and 56 mA, using Cu $K\alpha$ radiation. The data set in each case was collected from a single crystal, even if it meant that completeness was not as good as it could be, and was processed using the *DENZO*/*SCALEPACK* suite of programs (Otwinowski, 1993). Data-collection statistics for both the free and complex crystal forms are given in Table 1. The space group and unit-cell parameters of form I of the complex are the same as those of the structure reported earlier. However, the crystals now diffracted better.

2.3. Structure solution and refinement

The structure solution of the crystals of free UDG turned out to be extremely difficult. The data quality was poor and processing indicated several possible hexagonal and trigonal space groups. Therefore, a large number of molecular-replacement calculations using *AMoRe* (Navaza, 1994) had to be carried out with different space groups and search models (PDB codes 1uug, 2uug, 1eui) before arriving at a solution which eventually turned out to be correct. The solution used the UDG molecule in the UDG–Ugi complex reported by

Putnam *et al.* (1999) (PDB code 1uug) as the search model and gave a correlation coefficient and *R* factor of 0.395 and 0.403, respectively, for data in the 15–4 Å resolution shell. It led to acceptable packing of 12 molecules in the unit cell with space group $P3_2$. Structure refinement was carried out using *X-PLOR* (Brünger, 1992). The electron-density map calculated after rigid-body refinement was used to rebuild the model using *FRODO* (Jones, 1978). Non-crystallographic symmetry restraints with an energy constant of 1256 kJ mol^{-1} were applied in the subsequent refinement. The NCS restraint was progressively reduced in steps of 209 kJ mol^{-1}, each step involving 15 steps of minimization. Only overall and group *B* factors were refined throughout. Water molecules were defined based on peaks with height greater than 3σ in $F_o - F_c$ maps and those with height 1σ in $2F_o - F_c$ maps. Omit-type maps (Vijayan, 1980; Bhat & Cohen, 1984) were used to confirm the location of water molecules and to remove the effects of any model bias. The structure of form I was refined using the coordinates of the *Ec*UDG–Ugi complex (PDB code 1eui) solved at 3.2 Å (Ravishankar *et al.*, 1998), as the lattice settings for both crystals were the same. The form II crystal structure was determined by molecular replacement using *AMoRe*. The *Ec*UDG–Ugi complex reported by Putnam *et al.* (1999) (PDB code 1uug) was used as the search model. The correct solution was obtained with a correlation coefficient of 73.9% and an *R* factor of 28.8%. The protocol used for refining the structures of the complexes was similar to that used in the case of the structure of free UDG. The refinement statistics for all three structures are given in Table 2.

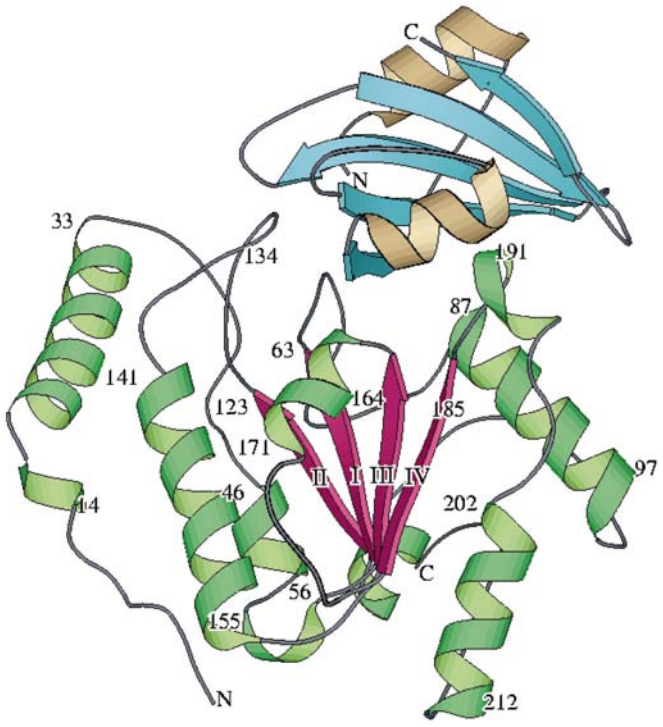

Figure 1
A ribbon representation of the *Ec*UDG–Ugi complex, illustrating the secondary-structure elements. The α-helices and the β-strands of the enzyme molecule are coloured green and pink, respectively, while those in the inhibitor molecule are coloured brown and cyan, respectively. The four strands of UDG are labelled. The figure was generated using *MOLSCRIPT* (Kraulis, 1991).

Table 2
Refinement parameters.

	Free	UDG–Ugi (form I)	UDG–Ugi (form II)
Resolution limits used in refinement (Å)	20.0–3.35	20.0–2.90	15.0–3.20
Final *R* factor (%)	24.1	18.3	18.8
R_{free} (%)	32.6	27.7	26.0
R.m.s. deviations from ideal			
Bond lengths (Å)	0.014	0.010	0.011
Bond angles (°)	1.9	1.7	1.6
Dihedral angles (°)	27.1	27.2	26.4
Improper angles (°)	0.91	0.83	0.95
Estimated coordinate error from Luzzati plot (Å)	0.40	0.25	0.30
No. of protein atoms	7157	4611	9690
No. of water molecules	175	156	58

Table 3
Structures containing *Ec*UDG used in the analysis.

PDB code	Resolution (Å)	Bound ligand	No. of molecules in asymmetric unit	Solvent content (%)	Reference
1uug	2.40	Ugi	2	52.7	Putnam *et al.* (1999)
2uug	2.60	Ugi	2	36.7	Putnam *et al.* (1999)
1eug	1.60	—	1	39.9	Xiao *et al.* (1999)
2eug	1.50	Uracil	1	35.5	Xiao *et al.* (1999)
3eug	1.43	Glycerol	1	36.5	Xiao *et al.* (1999)
4eug	1.40	—	1	39.9	Drohat *et al.* (1999)
5eug	1.60	Uracil	1	36.5	Drohat *et al.* (1999)
1lqg	2.90	Ugi	2	45.0	This study
1lqm	3.20	Ugi	4	50.1	This study
1lqj	3.35	—	4	67.8	This study
1flz	2.30	ssDNA	1	46.8	Werner *et al.* (2000)

2.4. Structural comparisons and helical parameters

ALIGN (Cohen, 1997) was used to carry out structural superpositions. Difference distance matrices (Nishikawa *et al.*, 1972; Kundrot & Richards, 1987; Madhusudan & Vijayan, 1991) were used for structural comparisons. If the distance between the C^{α} atom of the *i*th residue and that of the *j*th residue in one structure is R_{ij} and the corresponding distance in another structure is R'_{ij}, then the element d_{ij} in the difference distance matrix between the two is $R_{ij} - R'_{ij}$.

Helical parameters were calculated using a slightly modified version of *HELANAL* (Kumar & Bansal, 1996).

3. Results and discussion

3.1. Overall structure

*Ec*UDG, like the enzyme from other sources, is a classical α–β–α protein which has a small parallel four-stranded β-sheet (58–62, 119–123, 160–164 and 181–185), with II–I–III–IV strand connectivity, at the centre of the molecule (Fig. 1). The sheet is flanked by several helices. Four of the helices are ten or more residues long (18–33, 87–99, 141–155 and 202–212), while each of the remaining half a dozen helices contain six residues or less. The lengths of the loops vary from two to 26

Table 4
Root-mean-square deviations (Å) among *Ec*UDG molecules in the 11 different crystal forms.

C^{α} atoms of all residues were used in the superposition using *ALIGN* (Cohen, 1997).

			UDG–Ugi complex								Free						UDG–uracil		UDG–glycerol	UDG–DNA
		Form I	Form II				1uug		2uug		Free				1eug	4eug	2eug	5eug	3eug	1flz
		B	*A*	*B*	*C*	*D*	*A*	*B*	*A*	*B*	*A*	*B*	*C*	*D*	*A*	*A*	*A*	*A*	*A*	*A*
Form I	*A*	0.24	0.37	0.37	0.38	0.37	0.43	0.40	0.50	0.46	0.59	0.55	0.54	0.57	0.63	0.65	0.65	0.72	0.73	1.48
	B		0.38	0.37	0.38	0.37	0.40	0.39	0.49	0.46	0.59	0.57	0.49	0.59	0.64	0.71	0.70	0.77	0.74	1.47
Form II	*A*			0.22	0.20	0.20	0.29	0.32	0.43	0.45	0.47	0.47	0.45	0.49	0.62	0.62	0.61	0.66	0.70	1.41
	B				0.19	0.21	0.29	0.32	0.43	0.43	0.47	0.47	0.44	0.48	0.59	0.61	0.59	0.64	0.68	1.40
	C					0.19	0.28	0.32	0.42	0.44	0.47	0.47	0.45	0.47	0.60	0.61	0.60	0.65	0.69	1.41
	D						0.28	0.33	0.42	0.44	0.46	0.47	0.43	0.49	0.60	0.61	0.60	0.65	0.70	1.41
1uug	*A*							0.21	0.43	0.42	0.54	0.53	0.49	0.56	0.63	0.64	0.62	0.71	0.72	1.41
	B								0.44	0.42	0.56	0.53	0.49	0.56	0.64	0.62	0.64	0.73	0.73	1.39
2uug	*A*									0.38	0.54	0.54	0.53	0.57	0.54	0.58	0.55	0.63	0.60	1.39
	B										0.53	0.52	0.53	0.56	0.53	0.56	0.55	0.59	0.59	1.40
Free	*A*											0.35	0.33	0.35	0.54	0.56	0.56	0.60	0.61	1.27
	B												0.34	0.36	0.54	0.56	0.54	0.60	0.59	1.30
	C													0.33	0.53	0.55	0.53	0.60	0.59	1.25
	D														0.57	0.60	0.58	0.61	0.63	1.33
1eug	*A*															0.25	0.20	0.21	0.18	1.13
4eug	*A*																0.19	0.18	0.30	1.14
2eug	*A*																	0.18	0.30	1.09
5eug	*A*																		0.27	1.07
3eug	*A*																			1.19

residues. A groove made up of residues Gln63, Asn123, Ser166, His187, Ser189, Leu191 and Ser192 is involved in DNA binding, while the uracil-binding pocket involves Gln63, Asp64, Tyr66, Phe77, Asn123 and His187. The binding region of Ugi on UDG is the same as that which accommodates DNA (Ravishankar *et al.*, 1998; Putnam *et al.*, 1999). The four crystallographically independent molecules in the crystals of free UDG and the six in the two forms of its complex with Ugi have essentially the same structure. Also, the structures of the free molecule and the complex derived from the present study are very similar to those obtained previously. The free structure consists of residues 1–228, 1–226, 1–229 and 3–228 in subunits *A*, *B*, *C* and *D*, respectively. Form I of the UDG–Ugi complex consists of residues 3–228 and 2–225 in enzyme molecules *A* and *B*, respectively, and 12–84 and 15–84 in the corresponding inhibitor molecules. In the form II UDG–Ugi complex, the enzyme molecules *B* and *C* contain residues 4–227, whereas *A* and *D* contain residues 4–226 and 2–226, respectively. Residues 3–84 in molecules *A* and *C*, 2–84 in molecule *B* and 1–84 in molecule *D* are defined in the inhibitor. In the earlier work from this laboratory, five N-terminal residues and three C-terminal residues of UDG could not be defined. In fact, the N-terminal residue has not been defined in any of the structures reported so far. The structures reported here, however, involve definition of the entire polypeptide chain.

The Ugi molecule consists of a five-stranded antiparallel β-sheet and two helices, one near the N-terminus and the other in the polypeptide stretch connecting strands I and II in the sheet. The N-terminal helix was disordered in the complex studied earlier in this laboratory, unlike in form II reported here. Indeed, form II is the only structure in which the entire polypeptide chain is visualized. The structure of Ugi has been defined, except for the two N-terminal residues, in other crystals of the UDG–Ugi complex (Putnam *et al.*, 1999). Also available are the structures of ten crystallographically independent molecules of free Ugi in two crystal forms (Putnam *et al.*, 1999). The structures of the bound Ugi in different crystals of its complex with UDG, including those reported here, have essentially the same structure. The free Ugi molecules have also a similar structure. However, as pointed out by Putnam *et al.* (1999), nine out of the ten free Ugi molecules have conformational differences at residues Gln19, Glu20 and Ser21 from the bound ones. These conformational differences enable Glu20 in the bound molecule to establish an important interaction with UDG (Putnam *et al.*, 1999).

3.2. DNA binding and domain movement

12 crystal structures containing the *Ec*UDG molecule are available. As form I of the UDG–Ugi complex reported here is essentially the same as that reported previously (Ravishankar *et al.*, 1998), albeit with a better resolution, only the former is considered for further analysis. The 11 structures contain a total of 20 crystallographically independent enzyme molecules (Table 3). The r.m.s. deviations in C^{α} positions on their pairwise superposition are listed in Table 4. The deviations are markedly high in the molecule bound to DNA, indicating considerable conformational changes caused by interactions with DNA.

UDG is often described as a single-domain enzyme (Putnam *et al.*, 1999; Parikh, Putnam *et al.*, 2000). The closing of the active-site cleft subsequent to binding of uracil-containing DNA has been described as arising from the clamping motion of active-site loops resulting from concerted conformational changes in them (Slupphaug *et al.*, 1996;

Werner *et al.*, 2000). The DNA complexes of the *E. coli* and human enzymes are available (Werner *et al.*, 2000; Slupphaug *et al.*, 1996; Parikh *et al.*, 1998; Parikh, Walcher *et al.*, 2000). Their comparison with the corresponding DNA-free enzyme appeared to indicate that the movement upon complexation encompassed the whole molecule. To further investigate this possibility, the difference distance matrices relating DNA-free and DNA-bound molecules were constructed for the *E. coli* and human enzymes. Only residues 6–220 (88–301 in human enzymes), which are present in all structures, were used in the calculations. The respective average values of R_{ij} values in the relevant structures were used in the calculations. The two difference distance matrices are shown in Fig. 2. The plots clearly delineate four regions of the molecule, within each of which the internal movements are minimal. They by and large appear to move as rigid bodies. Furthermore, the four rigid regions can be grouped into two pairs, with the two regions in each pair maintaining their spatial relationship while the pairs undergo relative movement. Thus, with respect to DNA binding, the molecule can be divided into two relatively rigid regions or domains. Careful examination of the difference distance matrices and the models led to the following delineation of different regions in the molecule.

E. coli UDG: domain 1, 6–76 and 112–150; domain 2, 77–105 and 155–220; link region, 106–111 and 151–154.

Human UDG (values in parentheses indicate the corresponding residues in *Ec*UDG): domain 1, 88–159 (6–78) and 201–235 (120–154); domain 2, 160–191 (79–110) and 240–301 (159–220); link region, 192–200 (111–119) and 236–239 (155–158).

The regions thus delineated in *Ec*UDG are illustrated in Fig. 3.

A rigid-body movement can be described in terms of a screw motion involving a rotation about and a translation along an axis. The domain closure subsequent to DNA binding can be described to a good approximation as the rigid-body movement of one domain with respect to the other. In order to quantify the movement, domain 1 of each free *Ec*UDG molecule of known structure and that in the *Ec*UDG–ssDNA complex were superposed using *ALIGN* (Cohen, 1997). A rotation in the range 8–10° with a negligibly small translation along an axis passing through the interface between the two

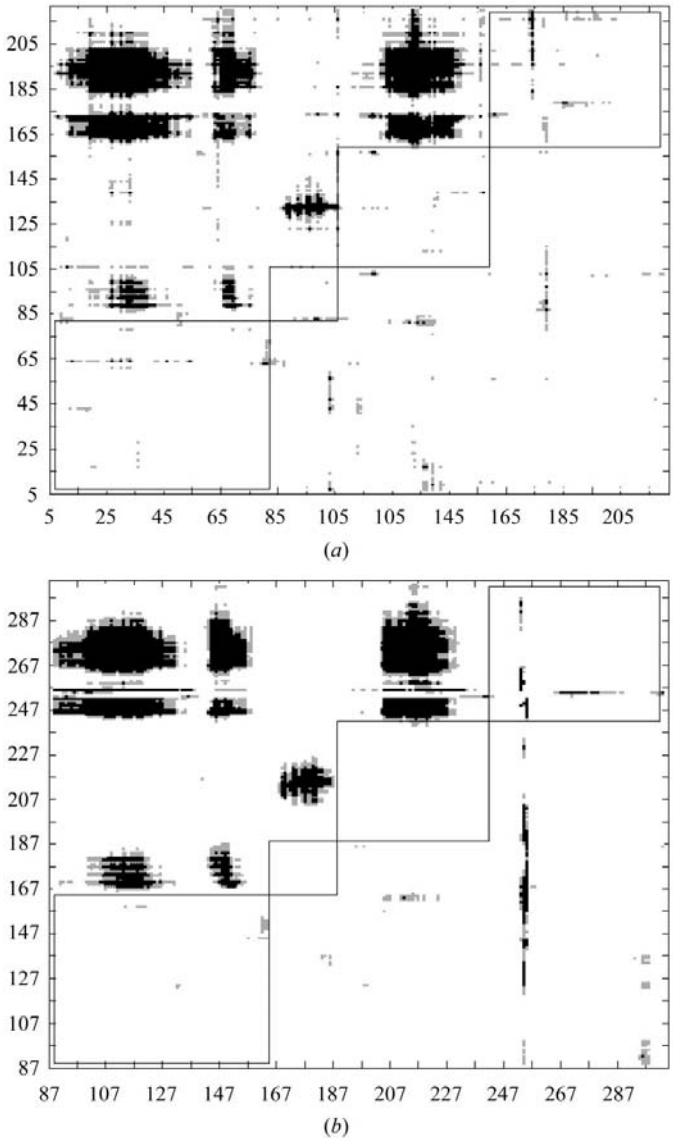

Figure 2
Difference distance matrix plots obtained by comparing the enzyme molecules in DNA-free structure and that in the DNA-bound structure. (*a*) *Ec*UDG and (*b*) human UDG. The half above the diagonal contains positive difference distances, while that below the diagonal contains the negative difference distances. Lightly shaded regions correspond to a difference of 1–1.5 Å. Regions with differences greater than 1.5 Å are indicated darkly shaded.

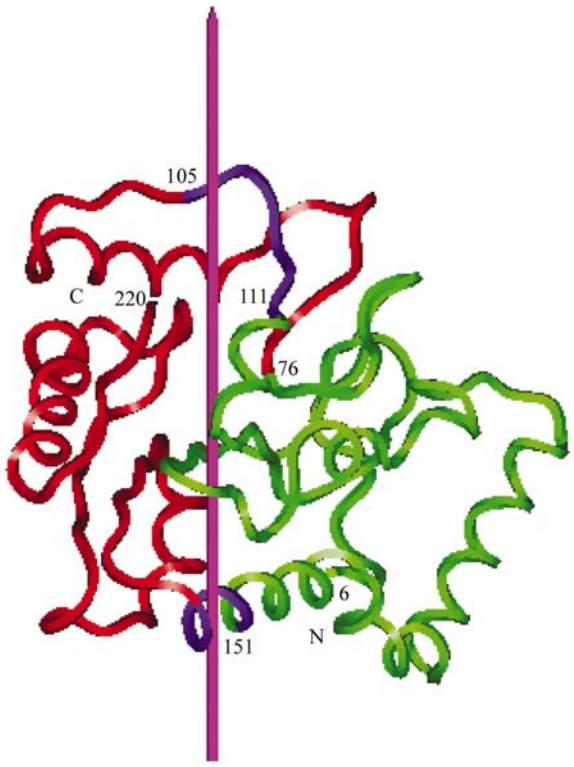

Figure 3
Delineation of domains in *Ec*UDG. Domain 1 is in green and domain 2 is in red. The linker region is in purple. The axis about which the domains move during closure is also shown.

Table 5
Domain closure.

(*a*) Domain closure in *E. coli* UDG upon binding to ssDNA (1flz) with respect to free UDG. See text for details.

Free UDG	Rotation (°)	Translation (Å)
1eug	8.3	−0.09
2eug	8.2	−0.09
3eug	8.9	−0.09
4eug	8.5	−0.02
5eug	8.0	−0.01
Free *A*	9.5	−0.06
Free *B*	9.8	0.01
Free *C*	9.8	−0.01
Free *D*	9.8	−0.02

(*b*) Domain closure in human UDG upon binding to dsDNA with respect to free UDG (1akz). See text for details.

UDG–dsDNA	Rotation (°)	Translation (Å)
1ssp	11.3	0.21
2ssp	11.7	0.20
1emj	11.4	0.19
1emh	11.1	0.19
4skn	9.5	0.25

domains was then necessary to bring domain 2 in the two molecules into superposition (Table 5; Fig. 4). The angle of rotation is slightly higher in the four molecules in the structure reported here than in the others, presumably on account of the somewhat higher flexibility afforded by the higher solvent content in the former.

Extensive information on domain closure is in fact provided by the complexes involving the human enzyme. Five structures of the human UDG bound to double-stranded DNA are available. The rotation and translation during domain closure in each of these structures with respect to the free molecule are listed in Table 5. All but one structure (4skn) have comparable values. In the case of 4skn the rotation angle is 2° lower than that in other structures. The enzyme molecule in this structure has mutations D145N and L272R. Upon binding to the substrate, Leu272 penetrates the DNA minor groove, which is facilitated by the domain closure. It is possible that the larger side chain of arginine at this position prevents complete closure.

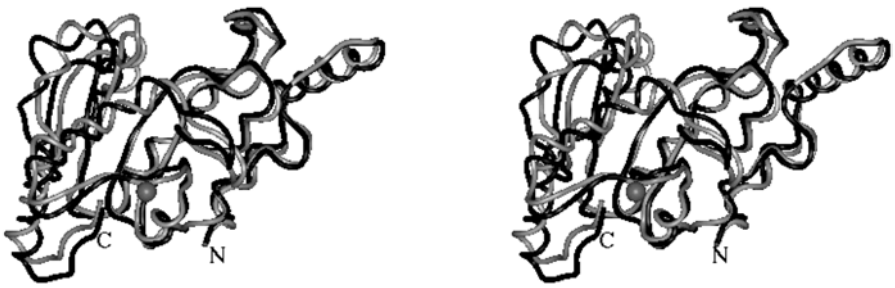

Figure 4
Stereoview illustrating domain closure in *Ec*UDG. Domain 1 of the free enzyme (dark) and that of the enzyme complexed with single-stranded DNA (light) are superposed. The axis about which the molecule has to rotate to bring domain 2 into superposition is perpendicular to the figure and is indicated by a ball.

Although the transformation from the open to the closed form is substantially a rotation, it does not appear to lend itself to description as a simple hinge motion. The interface between the two domains is extensive. It passes through the middle of the central β-sheet. Main-chain hydrogen bonds connect the two central strands (I and III) in the closed form. The residues at one end of the strands move apart in the open form so that the NH of one strand is no longer at a hydrogen-bonding distance from the CO of the other. A water molecule bridges the two groups. This water bridge has been observed in all structures with resolution better than 3 Å, including those of the human and HSV-1 enzymes (Putnam *et al.*, 1999).

In the only available HSV1 UDG–DNA complex, the enzyme is in the open conformation (Savva *et al.*, 1995). The molecule is complexed with 5′-p-dTdTdT-OH-3′. The 3′ thymine makes a favourable contact with the mouth of the uracil pocket but does not enter it (Savva *et al.*, 1995). The presence of free uracil within the active-site pocket of UDG does not in itself result in domain closure (Savva *et al.*, 1995; Xiao *et al.*, 1999). It is not even a necessary condition as an abasic DNA substrate can induce the closed conformation (Parikh *et al.*, 1998). The single-stranded trinucleotide induces domain closure in the *E. coli* enzyme to nearly the same extent as the larger double-stranded DNA fragment does in the human enzyme. Thus, it would appear surprising that a single-stranded trinucleotide does not induce domain closure in the viral enzyme. A major difference between the complexes of *E. coli* and herpes enzymes is in the location of the trinucleotide in the binding region. In the *Ec*UDG complex with 5′-OH-dUAAp-3′, the 5′dU, which has two nucleotides towards the 3′ end, interacts with the binding pocket. In the complex of the viral enzyme, however, the location in the vicinity of the binding pocket is occupied by the 3′-terminal nucleotide. It has been shown previously that a minimum of two nucleotides are necessary at the 3′ side of the uracil at the binding site for enzyme activity (Varshney & van der Sande, 1991).

In earlier discussions, the emphasis was on concerted conformational changes of residues and loops at the active site upon binding of UDG with DNA (Slupphaug *et al.*, 1996; Werner *et al.*, 2000), although the possibility of two halves of the enzyme molecule closing in had been suggested (Putnam *et al.*, 1999). The analyses discussed above, however, conclusively demonstrate and quantitatively describe the domain motion in UDG associated with DNA binding.

In addition to the inter-domain movement, DNA binding also results in movements within the domains. The movement of the 166–171 helix in *Ec*UDG and the corresponding 247–254 helix in the human enzyme in domain 2 is particularly pronounced. This helix is on the protein surface with its N-terminus forming part of the DNA-binding region. The interactions of the N-terminus with

Table 6
Helix motion.

(*a*) Motion of the 166–171 helix in UDG–DNA complex (1flz) with respect to free *Ec*UDGs. See text for details.

			Helical parameters		
	Rotation (°)	Translation (Å)	Average twist (°)	Average *N*	Average unit height (Å)
UDG–DNA					
1flz			97.4	3.71	1.52
Free UDG					
1eug	8.5	0.02	97.9	3.68	1.44
2eug	8.3	−0.10	98.5	3.66	1.44
3eug	8.8	−0.08	97.9	3.68	1.43
4eug	8.5	−0.22	98.5	3.66	1.46
5eug	7.8	−0.15	98.0	3.67	1.47
Free *A*	9.6	−0.23	102.1	3.54	1.44
Free *B*	10.4	−0.09	99.7	3.63	1.50
Free *C*	10.2	0.02	99.1	3.64	1.49
Free *D*	10.7	−0.09	97.8	3.69	1.55

(*b*) Motion of the 247–254 helix in human UDG–DNA complexes with respect to free UDG (1akz). See text for details.

			Helical parameters		
	Rotation (°)	Translation (Å)	Average twist (°)	Average *N*	Average unit height (Å)
Free UDG					
1akz			97.6	3.69	1.48
UDG-DNA					
1ssp	10.4	−0.44	109.4	3.31	1.83
2ssp	11.9	−0.16	109.9	3.30	1.81
1emh	11.5	−0.57	109.1	3.32	1.84
1emj	10.8	−0.62	109.5	3.31	1.87
4skn	6.4	−0.49	107.2	3.37	1.79

the DNA substrate are presumably responsible for the movement of the helix on DNA binding. This region of the structure is well defined. In the case of the free structure, the overall *B* value (averaged over the four crystallographically independent molecules) of the helix is 27.4 Å^2, compared with 28.6 Å^2 for the whole structure. Also, the electron density in this region is well defined.

The parameters required to define the movement of the helix were obtained through two steps of superposition. In the first step, domain 2 of the DNA-bound structure was superposed onto domain 2 of the free structure. In the next step, the helix in the newly oriented structure was superposed onto the corresponding helix in the free molecule. The second step provided the rotation and translation parameters. The helix as a whole rotates about an axis nearly parallel to the helix axis by an angle ranging from 7.8 to 10.7°. Here again, the angle of rotation is slightly, though distinctly, higher in the superposition involving the free molecules in the structure presented here, perhaps on account of the higher mobility afforded by the higher solvent content. Although the helix moves as a whole in response to DNA binding, the helical parameters themselves remain unchanged (Table 6; Fig. 5*a*).

A rigid-body movement of the 247–254 helix similar to that found in *Ec*UDG is also observed in the human enzyme on DNA binding. Understandably, the movement is smaller in the DNA complex of the L272R mutant (4skn). The angle of rotation of the helix in the other complexes has values similar to that in *Ec*UDG. However, unlike in the case of *Ec*UDG, the helical parameters also change substantially in the human enzyme (Table 6; Fig. 5*b*). In the free molecule, the parameters correspond to those of an α-helix. Upon DNA binding, the helix winds more tightly, with parameters intermediate between those of an α-helix and a 3_{10} helix. The reason for the difference between the *E. coli* and human enzyme in this regard is not immediately obvious. It might be connected to the nature and length of the bound DNA. The known structures of the DNA-bound human enzyme involve 8–10 base-pair-long double-stranded DNA. The phosphate group of the

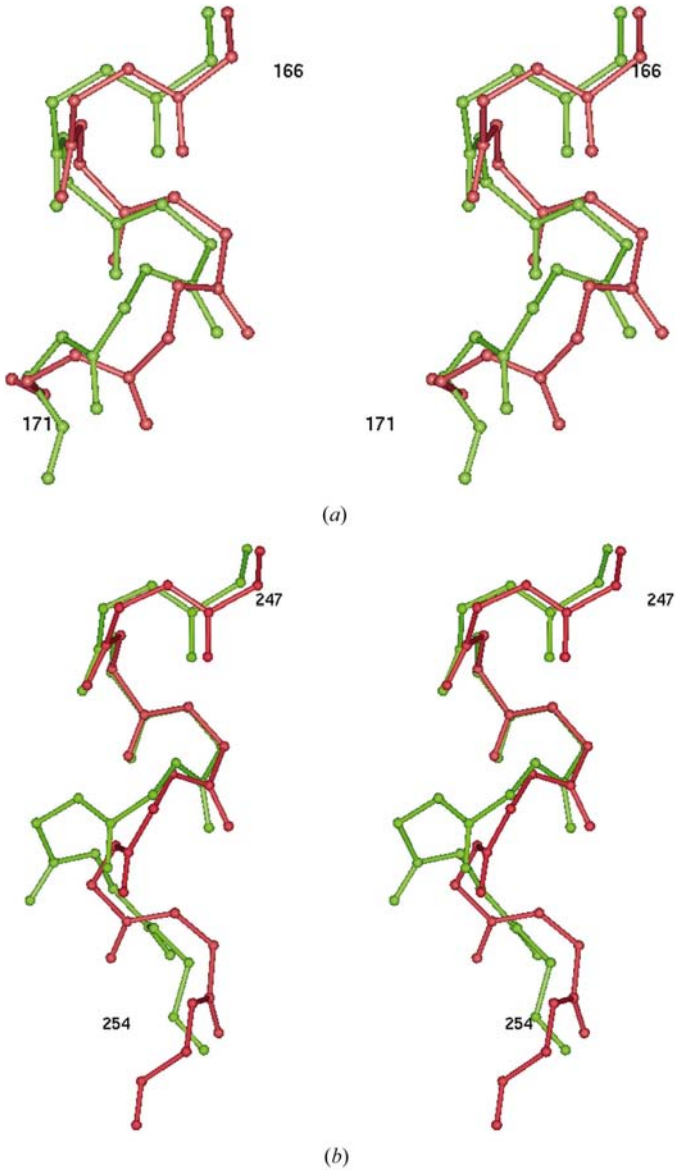

Figure 5
(*a*) Stereoview of the spatial disposition of the 166–171 helix in free (green) and DNA-bound (red) *Ec*UDG upon superposition of domain 2 from the respective structures. (*b*) Similar view in the corresponding 247–254 helix of human enzyme. See text for details.

nucleotide two residues away from the uracil inserted into the binding pocket interacts with residues 246 and 247 in the helix (Parikh *et al.*, 1998). There is no electron density for the corresponding phosphate group in the single-stranded trinucleotide bound to *Ec*UDG and the group is presumably disordered (Werner *et al.*, 2000).

Thus, to summarize, the UDG molecule is made up of two domains. On DNA binding, the domains move in concert to close in on the nucleic acid molecule. The binding also results in movements within the domains.

The diffraction data were collected at the National Area-Detector Facility supported by Department of Science and Technology (DST) and the Department of Biotechnology (DBT). Facilities at the Supercomputer Education Research Centre (SERC) and the Interactive Graphics Based Molecular Modelling Facility and the Distributed Information Centre (both supported by DBT) were used in this work. MBS and SR were supported by DBT Postdoctoral Research Associateships and KP and PH were recipients of K. S. Krishnan and Council of Scientific and Industrial Research fellowships, respectively. The work was funded by DST and DBT.

References

Bhat, T. N. & Cohen, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.
Brünger, A. T. (1992). *X-PLOR: Version 3.1. A System for X-ray Crystallography and NMR.* Yale University, New Haven, CT, USA.
Cohen, G. E. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
David, S. S. & Williams, S. D. (1998). *Chem. Rev.* **98**, 1221–1261.
Drohat, A. C., Xiao, G., Tordova, M., Jagadeesh, J., Pankiewicz, K. W., Watanabe, K. A., Gilliland, G. L. & Stivers, J. T. (1999). *Biochemistry*, **38**, 11876–11886.
Duncan, B. K. & Chambers, J. A. (1984). *Gene*, **28**, 211–219.
Handa, P., Roy, S. & Varshney, U. (2001). *J. Biol. Chem.* **276**, 17324–17331.
Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
Kumar, S. & Bansal, M. (1996). *Biophys. J.* **71**, 1574–1586.
Kundrot, C. E. & Richards, F. M. (1987). *J. Mol. Biol.* **193**, 157–170.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Lindahl, T. (1974). *Proc. Natl Acad. Sci. USA*, **71**, 3649–3653.
Lindahl, T., Ljungquist, S., Siegert, W., Nyberg, B. & Sperens, B. (1977). *J. Biol. Chem.* **252**, 3286–3294.
McCullough, A. K., Dodson, M. L. & Lloyd, R. S. (1999). *Annu. Rev. Biochem.* **68**, 255–285.
Madhusudan & Vijayan, M. (1991). *Curr. Sci.* **60**, 165–170.
Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., Mosbaugh, D. W. & Tainer, J. A. (1995). *Cell*, **82**, 701–708.
Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Krokan, H. E. & Tainer, J. A. (1995). *Cell*, **80**, 869–878.
Navaza, L. (1994). *Acta Cryst.* A**50**, 157–163.
Nishikawa, K., Ooi, T., Isogai, Y. & Saito, N. (1972). *J. Phys. Soc. Jpn*, **32**, 1331–1337.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E. & Tainer, J. A. (1998). *EMBO J.* **17**, 5214–5226.
Parikh, S. S., Putnam, C. D. & Tainer, J. A. (2000). *Mutat. Res.* **460**, 183–199.
Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M. & Tainer, J. A. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 5083–5088.
Pearl, L. H. (2000). *Mutat. Res.* **460**, 165–181.
Putnam, C. D., Shroyer, M. J. N., Lundquist, A. J., Mol, C. D., Arvai, A. S., Mosbaugh, D. W. & Tainer, J. A. (1999). *J. Mol. Biol.* **287**, 331–346.
Ravishankar, R., Bidya Sagar, M., Roy, S., Purnapatre, K., Handa, P., Varshney, U. & Vijayan, M. (1998). *Nucleic Acids Res.* **26**, 4880–4887.
Roy, S., Purnapatre, K., Handa, P., Boyanapalli, M. & Varshney, U. (1998). *Protein Expr. Purif.* **13**, 155–162.
Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995). *Nature (London)*, **373**, 487–493.
Savva, R. & Pearl, L. H. (1995). *Nature Struct. Biol.* **2**, 752–757.
Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E. & Tainer, J. A. (1996). *Nature (London)*, **384**, 87–92.
Varshney, U., Hutcheon, T. & van de Sande, J. H. (1988). *J. Biol. Chem.* **263**, 7776–7784.
Varshney, U. & van de Sande, J. H. (1991). *Biochemistry*, **30**, 4055–4061.
Vijayan, M. (1980). *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 19.01–19.26. Bangalore: Indian Academy of Sciences.
Werner, R. M., Jiang, Y. L., Gordley, R. G., Jagadeesh, G. J., Ladner, J. E., Xiao, G., Tordova, M., Gilliland, G. L. & Stivers, J. T. (2000). *Biochemistry*, **39**, 12585–12594.
Xiao, G., Tordova, M., Jagadeesh, J., Drohat, A. C., Stivers, J. T. & Gilliland, G. L. (1999). *Proteins Struct. Funct. Genet.* **35**, 13–24.