4880–4887 *Nucleic Acids Research, 1998, Vol. 26, No. 21*

X-ray analysis of a complex of *Escherichia coli* uracil DNA glycosylase (*Ec*UDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG

R. Ravishankar, M. Bidya Sagar, S. Roy[1], K. Purnapatre[1], P. Handa[1], U. Varshney[1] and M. Vijayan*

Molecular Biophysics Unit and [1]Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560 012, India

Received July 13, 1998; Revised and Accepted September 21, 1998

PDB accession no. 1EUI

ABSTRACT

Uracil–DNA glycosylase (UDG), a key highly conserved DNA repair enzyme involved in uracil excision repair, was discovered in *Escherichia coli*. The *Bacillus subtilis* bacteriophage, PBS-1 and PBS-2, which contain dUMP residues in their DNA, express a UDG inhibitor protein, Ugi which binds to UDG very tightly to form a physiologically irreversible complex. The X-ray analysis of the *E.coli* UDG (*Ec*UDG)–Ugi complex at 3.2 Å resolution, leads to the first structure elucidation of a bacterial UDG molecule. This structure is similar to the enzymes from human and viral sources. A comparison of the available structures involving UDG permits the delineation of the constant and the variable regions of the molecule. Structural comparison and mutational analysis also indicate that the mode of action of the enzyme from these sources are the same. The crystal structure shows a remarkable spatial conservation of the active site residues involved in DNA binding in spite of significant differences in the structure of the enzyme–inhibitor complex, in comparison with those from the mammalian and viral sources. *Ec*UDG could serve as a prototype for UDGs from pathogenic prokaryotes, and provide a framework for possible drug development against such pathogens with emphasis on features of the molecule that differ from those in the human enzyme.

INTRODUCTION

DNA glycosylases excise damaged or unconventional bases in DNA and initiate the base excision repair (BER) pathway to maintain genomic integrity. Of these, uracil–DNA glycosylase (UDG) excises uracil from DNA by cleavage of the N-glycosidic bond between uracil and deoxyribose sugar. Uracil residues in DNA arise as a result of deamination of cytosine or incorporation of dUMP by DNA polymerase. However, in cells, the latter event is kept to a minimum by the presence of dUTPase (*dut*) which helps maintain a low level of dUTP. UDG (Ung) was the first DNA glycosylase to be discovered (1) and purified from *Escherichia coli* (2). It has since been found in most organisms including some of the viruses that infect eukaryotic cells. The *E.coli* UDG (*Ec*UDG) gene (*ung*) was also the first uracil DNA glycosylase to be cloned (3) and sequenced (4). Subsequent cloning and sequencing of UDG genes from *Saccharomyces cerevisiae* (5), human (6) and other sources established that UDG is a highly conserved protein. Interestingly, the organisms also contain a second group of UDGs consisting of a diverse group of proteins such as the cyclin-like UDG (7,8), dsUDG (9) and glyceraldehyde 3 phosphate dehydrogenase (10). None of the UDGs studied so far require metal ions or other cofactors for its activity (11–14). While for much of the biochemical studies *Ec*UDG has served as a prototype (11,14–17), its 3D structure is still unavailable. However, the crystal structure solutions of the N-terminally deleted active UDGs from human and herpes simplex virus type-1 (HSV-1) have shown that their 3D structures are highly conserved. These crystal structures have revealed that the active site grooves of the conserved UDGs are also highly conserved and that they bind to uridine in its extrahelical conformation (18–22). Binding of extrahelical bases was earlier reported for two bacterial DNA methyltransferases and several other DNA repair enzymes (23–25). Recently, the structure of an engineered mutant of human UDG complexed with double-stranded DNA suggested that the uracil base flipping is achieved by the enzyme mediated ‘push’ and ‘pull’ mechanism (22,26) involving a highly conserved motif HPSPLS. However, a recent study of the interactions of HSV type-2 UDG mutants with single- or double-stranded DNA oligomers, which either contained or lacked dU, using surface plasmon resonance, suggested UDG binding to the spontaneously flipped out bases (27). In any case, the glycosidic bond between the uracil and the sugar is cleaved by the attack of a hydroxyl nucleophile onto the deoxyribose C1′ atom. The hydroxyl nucleophile is most likely generated by the activation of a water molecule by the aspartate residue of yet another conserved motif GQDPYH (12,18,28).

UDGs are inhibited by free uracil and some of its derivatives (11–14). Another category of UDG inhibitors is represented by *Bacillus subtilis* phage PBS-1 and PBS-2 encoded inhibitor

*To whom correspondence should be addressed. Tel: +91 80 334 6765; Fax: +91 80 334 1683; Email: mv@mbu.iisc.ernet.in

protein Ugi (29) and phage T5 induced proteins (30). The mechanism of UDG inhibition by Ugi has been well studied. Ugi is a heat stable, acidic, low molecular weight protein of 9.5 kDa which interacts with UDG in a 1:1 molar stoichiometry to form a complex that does not dissociate under physiological conditions (13). Because of its low molecular weight, the UDG–Ugi complex is well suited to study the mechanism of protein–protein interaction by various biophysical methods. The co-crystal structures of Ugi with human and HSV-1 UDGs have been solved recently (20,21).

Recently, we described the design of a transcriptionally coupled system for co-expression of the two proteins to form UDG–Ugi complex in the cellular milieu (31). The *Ec*UDG–Ugi complex so obtained was purified and used to obtain the crystals of the complex. These crystals diffract to 3.2 Å. In this report we describe the long awaited 3D structure of *Ec*UDG in its complex with Ugi. Furthermore, the present analysis, the first of a bacterial UDG, permits a detailed comparison of the enzyme from mammalian, bacterial and viral sources.

MATERIALS AND METHODS

UDG and Ugi were co-expressed using a bicistronic construct, and the *in vivo* formed complex of *Ec*UDG with Ugi was purified to apparent homogeneity as described by Roy *et al.* (31). The crystals of the complex were grown from a hanging drop of 5 mg/ml protein in 20 mM Tris–HCl buffer, pH 7.4, containing 100 mM NaCl, 8% (w/v) PEG 3350 equilibrated against 30% (w/v) PEG 3350 in water. The crystals grew to their maximum size of 0.2 × 0.15 × 0.15 mm in 3–4 weeks. The asymmetric unit could contain two complexes for a Matthews co-efficient of 2.28 $Å^3$/Da assuming the molecular weight of the complex to be 36 kDa.

A crystal of the dimensions mentioned above was mounted in a glass capillary. X-ray diffraction data were collected at room temperature using a 30 cm MAR Research imaging plate system mounted on a Rigaku RU200 X-ray generator operating at 40 kV and 58 mA. Data collection statistics are given in Table 1. Though the crystals diffract to ~2.8 Å resolution, useful data could be processed only to 3.2 Å. The data set was collected from a single crystal (space group $P2_12_12_1$, a = 51.37; b = 89.77; c = 142.13 Å) and were processed using the DENZO/SCALEPACK suite of programs (32).

The structure of the *Ec*UDG–Ugi complex was determined by molecular replacement using AMoRe (33). The search model used was derived from the HSV-1 UDG–Ugi structure downloaded from the Brookhaven Protein Data Bank (PDB code: pdb1udi.ent) (34), which shares 49% sequence identity with *E.coli* UDG. Though the 'correct' solution had a clearly demarcated low R-factor and high correlation coefficient, an examination of the packing of the molecules (two complexes in the asymmetric unit) showed severe clashes involving the first 14 residues of the inhibitor. Subsequently, a search for two UDG molecules and one Ugi molecule did not show any short contacts. Data in the 15.0–4.0 Å resolution range and an integration radius of 20.0 Å, were used in the calculations of the rotation and translation functions. The solution had a correlation coefficient of 0.465 and an R-factor of 42.2%. In comparison, the next suggested solution had a correlation coefficient and R-factor of 0.392 and 44.4%, respectively. An examination of both $2F_o–F_c$ and $F_o–F_c$ maps calculated after an initial round of refinement, showed density for a second Ugi molecule but no density for the first 14 residues in both the independent inhibitor molecules. The second molecule was then built into the difference map and further refinement proceeded.

Table 1. Summary of data collection and refinement statistics

Data collection	
Total number of reflections measured	78855
Number of unique reflections I/ (σI) > 0.0	11074
Number of unique reflections I/ (σI) > 2.0	10140
Redundancy	7.1 (6.8)
Completeness (resolution range 10.0–3.2 Å) (%)	96.8 (89.8)
R_{merge} (%)[a]	13.1 (47.1)
Refinement statistics	
Final R[b] (%)	20.5 (22.7)
Final free R[b] (%)	27.4 (29.9)
Number of protein atoms	4573
Number of solvent atoms	116
r.m.s. deviation in bond angles (°)	1.8
r.m.s. deviation in bond lengths (Å)	0.01

[a]$R_{merge} = \Sigma |<I> – I_h| / \Sigma <I> \times 100$; I_h is the intensity of a measurement and <I> is the average of the measurements for a reflection h.
[b]$R = \Sigma |F_o – kF_c| / \Sigma F_o \times 100$.
The values provided in parenthesis in the 'Data collection' section are for the (3.3–3.2 Å) resolution shell. The values in parenthesis in the 'Refinement statistics' section, are for the crystal structure with the solvent molecules removed.

The initial free and the conventional R-factors were 45.7 and 44.8%, respectively, to a resolution 3.2 Å. Both $2F_o–F_c$ and $F_o–F_c$ maps calculated at this stage were easy to interpret. An atomic model of *Ec*UDG–Ugi complex consistent with the amino acid sequence deduced from the gene sequence was built using the interactive graphics program FRODO (35). Refinement proceeded routinely using XPLOR (36). Engh and Huber parameters (37) were used throughout the refinement. Non-crystallographic symmetry restraints were imposed in all the refinement cycles with an effective energy constant of 300 kcal/mol. A total of 10% of the reflections not included in the refinement were used for the computation of the free R-factor. At each stage, the refinement proceeded until the free R-factor converged. Rounds of refinement were interspersed with manual map building. Only the overall and group B-factors were refined throughout. When the free R-factor dropped to 29%, probable water molecules making hydrogen bonds with the protein atoms were located in difference Fourier maps contoured at 3σ level. A total of 116 water molecules were added to the model in stages. Extensive use was made of omit type maps (38,39) to confirm the location of water molecules and to remove the effects of any model bias. All the insertions and deletions in the sequence compared with the model from the herpes virus source could easily be built using these maps. The final model consists of residues from 6–226 (C-terminus) in enzyme molecule 1, 6–225 in enzyme molecule 2, 15–84 in both the molecules of the inhibitor and 116 water oxygens. However, density was absent beyond Cβ in residues 9, 11, 42, 177, 195 in enzyme molecule 1 and in residues 6, 10, 171, 177, 179, 195 and 211 in enzyme molecule

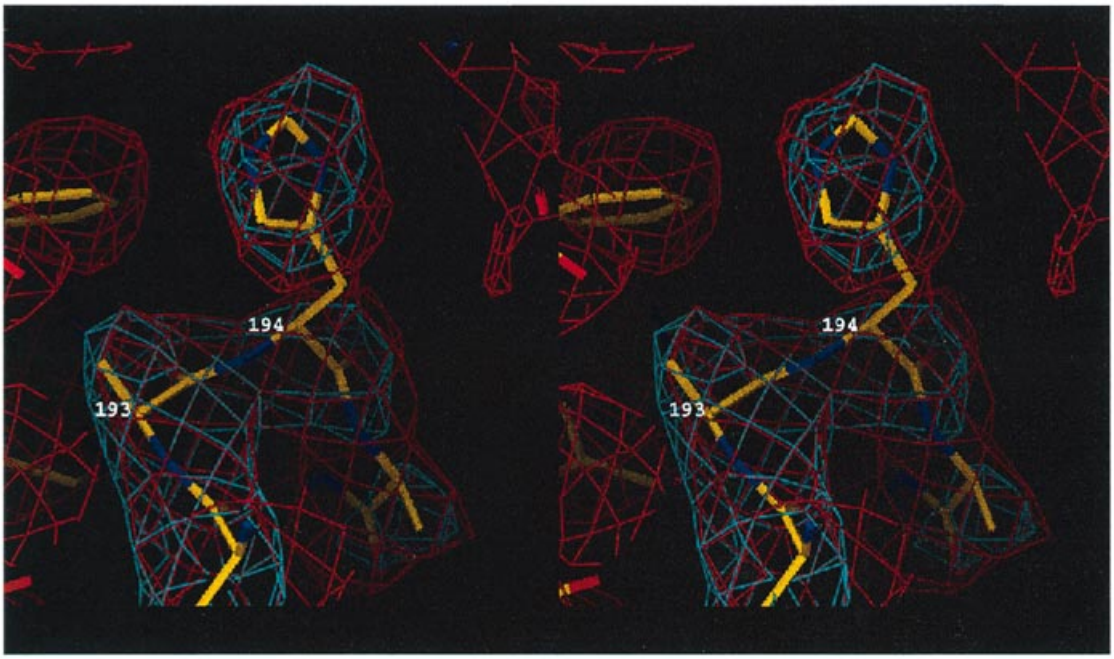

Figure 1. Stereo view of the electron density, F_o–F_c in cyan and $2F_o$–F_c in red, corresponding to the insertions (with respect to the HSV-1 enzyme) Ala 193 and His 194. These residues were removed from the model, from which the phases for the maps were calculated. The contour levels are at 2.5 and 0.8 σ respectively. Figure prepared using O (47).

2. The refinement converged to an R-value of 20.6 and a free R-value of 27.4% for all reflections with positive intensity in the resolution range 10–3.2 Å. A Luzzati plot (40) calculated at the end of the refinement showed an average coordinate error of 0.3 Å. The r.m.s. deviations of the bond lengths and angles from standard values are 0.01 Å and 1.8°, respectively. The final refinement statistics are shown in Table 1. The quality of the final $2F_o$–F_c and F_o–F_c map is illustrated in Figure 1. The coordinates and the structure factors have been deposited in the PDB with accession codes 1EUI and R1EUISF, respectively.

Escherichia coli ung gene was mutated at codons 64 and 187 using a modification of the method of Kunkel (41). Relevant segments of *ung* gene were subcloned into pTrc99C (Pharmacia) based expression system and verified by sequence analysis. Log phase cultures of freshly obtained transformants were induced by 0.25–0.5 mM IPTG, and the activities of the pure proteins were determined from range finding experiments (17).

RESULTS AND DISCUSSION

Molecular structure

As expected, partly on account of the NCS restraints, the two crystallographically independent complexes in the structure have remarkably similar conformation. When superposed, the r.m.s. deviation in Cα positions between the two molecules is as low as 0.36 Å, while the corresponding figure for the inhibitor molecules is still lower at 0.08 Å. The mutual disposition of the enzyme and the inhibitor is also the same in the two complexes. For example, when the orientation matrix and the translation vector which relate the two enzyme molecules are used to superpose the inhibitor molecules, the r.m.s. deviation in Cα positions between the two inhibitor molecules still remains low at 0.35 Å. Thus, only one of the complexes is used in the discussion that follows.

As in the case of the enzymes from other sources, *Ec*UDG has a classical α-β-α structure with a small four-stranded β-sheet sandwiched between the helices at the core of the molecule (Fig. 2). The number of helices varies depending upon the criteria used for identifying them. The current widely used criteria provided by Kabsch and Sanders (42), leads to the identification of 11 separate helical stretches of varying lengths (Fig. 3). Of these, four contain >10 residues each. Two such long helices lie on either side of the sheet. The remaining helices contain five residues or less. The N-terminal segment up to residue 33 is reasonably straight, and contains an extended stretch, a short three-residue helix and the long 18–33 helix. The polypeptide chain then takes a U-turn and follows a zig-zag course consisting of a long loop, a short helix and a short loop before entering the β structure as the first five-residue long strand. The strand is followed by a 13-residue helix, another long loop and a short helix before entering the sheet as strand 2. Strands 2 and 3 are connected by a long loop, a long helix and a short loop. The polypeptide stretch that connects strands 3 and 4 contain two short helical stretches in addition to loops. The C-terminal stretch that follows strand 4 contains a short loop, a short helix, a long loop, a long helix and a coil, in that order.

The first 14 N-terminal residues in the inhibitor molecules are not seen in the electron density map and are presumed to be disordered. The rest of the molecule is essentially an antiparallel β-sheet made up of five contiguous strands with a helix in the

Figure 2. The molecular structure of the *Ec*UDG–Ugi complex. The constant regions of the enzyme and the inhibitor molecules are in blue and green, respectively, while the variable regions in both the molecules are in yellow (see text for details). Figure prepared using MOLSCRIPT (48) and Raster 3D (49).

polypeptide stretch that connects strands 1 and 2 (Fig. 2). The helix lies on one side of the sheet towards the end of the molecule.

Comparison with the human and HSV-1 enzymes and their complexes

The structures of human and HSV-1 UDG and their inhibitor complexes including those involving Ugi have been reported (18–21). Also available is an NMR structure of Ugi (43). In terms of the sequence, the major difference among the *E.coli*, human and HSV-1 enzymes lies at the N-terminus. The first residue in *Ec*UDG corresponds to residue 26 in the HSV-1 enzyme and residue 83 in the human enzyme. The first 71 residues have been removed from the recombinant human enzyme used for structure analysis. A further nine residues are not seen in the electron density map. In the case of the HSV-1 enzyme, the first 16 residues are not seen in the maps. This number is five in the electron density maps of *Ec*UDG. Thus, the first structurally defined residue in *E.coli* is the sixth one from the N-terminus. It corresponds to the residue 31 in HSV-1 UDG and the residue 88 in the human enzyme. Furthermore, there are deletions at residues 35, 170, 193 and 194 (*Ec*UDG numbering scheme) and an insertion between residues 199 and 200 in the HSV-1 enzyme when compared with the *E.coli* enzyme. The sequence of the human enzyme has only one deletion at residue 34 when compared with *Ec*UDG. The sequence identity between the human and the HSV-1 enzymes is 39%, whereas it is 49% between *E.coli* and HSV-1 enzymes and 56% between *E.coli* and human enzymes. Thus, the *E.coli* enzyme is more similar to the human enzyme than to the HSV-1 enzyme. It also turns out that the human and the HSV-1 enzymes are closer to the *E.coli* enzyme than they are to each other.

The 3D structure of the *Ec*UDG in its complex with Ugi is similar to that of the HSV-1 and human enzymes and their complexes. The coordinates of the free enzymes and the complexes of the HSV-1 enzyme with uracil (18), a trinucleotide (18) and Ugi (21) are available. There are altogether seven crystallographically independent UDG molecules distributed among the crystal structures for which the atomic coordinates are available in the PDB. The r.m.s. deviations in the Cα position among them are listed in Table 2. The values involving the HSV-1 enzyme and its complexes clearly show that the geometry of the enzyme molecule remains relatively unaffected by interaction with inhibitors. Those involving enzymes from the three sources partially conform to the conclusion drawn from sequence information. The r.m.s. deviation in Cα position of *Ec*UDG from the human enzyme is less than that from the HSV-1 enzyme. However, that between the human and the HSV-1 enzyme is slightly lower than that between the *E.coli* and HSV-1 enzymes, though it is considerably higher than that between *E.coli* and human enzymes.

The availability of several structures of the same molecule from different species and under different environmental conditions

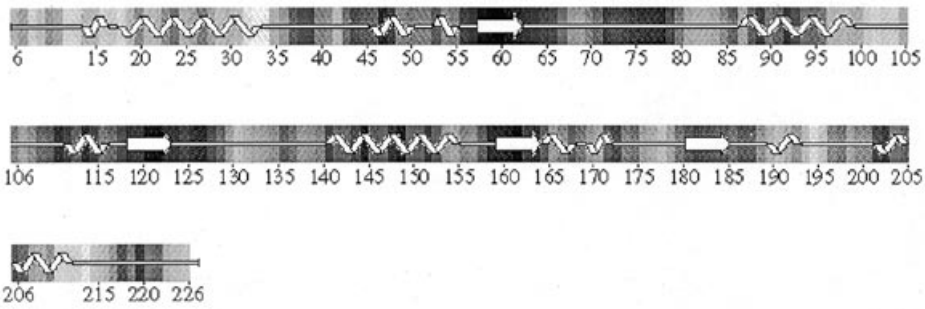

Figure 3. Secondary structural elements in the enzyme molecule delineated using PROCHECK (50).

permits the delineation of the relatively constant and variable regions in it. A simple computational approach developed earlier in this laboratory (44) was used for this purpose. If the distance between the Cα atoms of the ith residue and that of the jth is R_{ij}, then the difference in R_{ij} between the kth and lth molecule may be specified as $d_{ijkl} = R_{ijk} - R_{ijl}$. The r.m.s. deviation for each residue is then defined as r.m.s. (i) = $(<|d_{ijkl}|^2>)^{1/2}$, where the average is taken over j, k and l. The residue with the highest r.m.s. (i) is taken as the most variable residue. The calculation is repeated after eliminating the most variable residue to identify the next most variable residue. The calculation is repeated until the r.m.s. (i) of the most variable residue of the set is less than a specified value d. Residues 62–67, 77–79, 118–123 and 187–192 are known to be invariant. These stretches contain most of the residues involved in DNA binding and catalysis, and are expected to have the same mutual disposition in all UDG structures. One of them, 118–123, constitutes a strand in the β-sheet as well. On account of its central position, the β-sheet itself is expected to be relatively rigid. Therefore, d was chosen as the minimum value of r.m.s. (i) (1.8 Å) that would leave the four invariant stretches and the central β-sheet in the constant region. The average of the r.m.s. deviations among the enzymes from the three sources is also close to 1.8 Å. The results of the calculation indicated that residues 35–47, 53–54, 56–68, 70–80, 83, 85–97, 99, 108–136, 139–147, 150–165, 167–168, 175–177 and 179–194 constitute the relatively constant region while the remaining residues belong to the variable region. These are illustrated in Figure 2. The residues constituting the constant region are spatially contiguous and contain the central β-sheet and its immediate neighbourhood, particularly residues facing the substrate/inhibitor. The variable regions are, as expected, at the periphery of the molecule. The longest stretches of the polypeptide chain in these regions are the N-terminal residues up to 34 and the C-terminal residues, 195–226. The N-terminal residues and residues 48–52 and 148–149 constitute a variable region at one edge of the molecule. The C-terminal residues and the residues 100–107 form a variable region at the opposite edge of the molecule. These account for 76 out of the 93 residues in variable regions. The remaining variable residues are distributed in a non-contiguous manner.

Coordinates of only two crystal structures containing Ugi are currently available. Of these, the present structure contains two independent molecules and the complex involving the HSV-1 enzyme contains one independent molecule. Of the two helices in Ugi found in the complex with HSV-1 UDG, one is disordered in the present structure. An NMR study on the inhibitor and its complex (43) shows that the orientation of the two helices with respect to the sheet is highly variable. Presumably, in the crystal structure, the complex with *Ec*UDG permits variable orientation of one of the helices, while the other helix has a fixed orientation. Calculation of the type outlined earlier was performed on residue 15 to the C-terminus in the two Ugi molecules in the present structure and the molecule in the complex with HSV-1 enzyme, to delineate the constant and the variable regions in it. In this calculation, the threshold value d was chosen as 0.60 Å, a value close to the average r.m.s. deviation in the Cα positions among the three molecules. In addition to residues 1–14 (disordered in the *Ec*UDG complex), the other residues in the variable regions are 15–17, 26, 34, 37–38, 42, 47–53, 62, 75–76 and 82–84. All these residues belong to loops or chain termini.

Active site, mechanism of action and inhibitor binding

The binding site of the enzyme consists of a long positively charged groove which binds DNA and an adjacent pocket which binds uracil (Fig. 4a). Residues Gln 63, Asp 64, His 67, Phe 77, Ser 88, Asn 123, Ser/Thr 166, His 187, Ser 189, Leu 191 and Ser 192 are involved in the groove in the enzymes from the three sources. The uracil binding pocket is made up of Gln 63, Asp 64, Tyr 66, Phe 77 and Asn 123. Earlier structural studies and mutational analysis of the human and HSV-1 enzymes indicate equivalents of Asp 64 and His 187 to be catalytic residues (18,22).

When the HSV-1-UDG uracil complex was superposed on the *Ec*UDG molecule, the uracil molecule fits in exactly without short contacts in the binding pocket of the latter. The interactions involving uracil are also preserved. The same result was obtained when the HSV-1 UDG-trinucleotide complex was superposed on *Ec*UDG. The relation between the mechanism of action of *Ec*UDG and that of the human and HSV-1 enzymes was further explored through mutational studies involving Asp 64 and His 187. It was found that the mutants D64N and D64H as well as H187L and H187R were <1% active compared with the wild-type enzyme. These mutations, however, do not affect the binding of *Ec*UDG to Ugi, suggesting that the mutations do not affect the 3D structure of the enzyme and that the reduction of activity is caused by interference with the catalytic mechanism. A water molecule adjacent to the catalytic residue Asp 64, has been implicated in catalysis on the basis of the work on the human and HSV-1 enzyme. This water molecule is present in the *Ec*UDG–Ugi complex as well. To sum up, the above observations conclusively demonstrate that the nature of the binding site and mechanism of action of *Ec*UDG are essentially the same as in the mammalian and the viral enzymes.

Table 2. Root mean square deviations (Å) among UDG molecules in different crystal structures

	*Ec*UDG–Ugi Molecule 2	Human (free)	HSV (free)	HSV–Ugi	HSV–uracil	HSV–trinucl.
*Ec*UDG–Ugi						
Molecule 1	0.36	1.24	1.95	1.86	2.00	1.86
Molecule 2		1.21	1.95	1.86	2.00	1.85
Human (free)			1.83	1.77	1.86	1.78
HSV (free)				0.41	0.21	0.41
HSV–Ugi					0.48	0.40
HSV–uracil						0.49

The low value for the *Ec*UDG molecules could be partly due to NCS restraints.

The above conclusion is supported by a comparison of the mode of Ugi binding to the enzyme from the three sources. The helix, comprising residues 3–13 of the inhibitor, which is disordered in the present structure, does not seem to play any part in the binding. There are no enzyme–inhibitor contacts with this helix in the HSV-1 UDG–Ugi complex. The enzyme–inhibitor interactions are nearly the same in all the three cases. All such interactions (distance < 3.6 Å) involving *Ec*UDG are given in Table 3. All of these are observed in the complex involving the HSV-1 enzyme as well. When the two complexes were superposed using the orientation matrix and the translation vector appropriate for the superposition of the two enzyme molecules, the inhibitors superposed with a low r.m.s. value of 0.95 Å for Cα atoms. Detailed calculations involving the inhibitor complex of the human enzyme could not be carried out in the absence of the availability of the coordinates. However, the reported structure indicates that the enzyme–inhibitor interactions in it are similar to those in the complexes involving *E.coli* and HSV-1 enzymes.

A solvent accessibility calculation carried out using the Lee and Richards algorithm (45) with a probe radius of 1.4 Å shows that the *Ec*UDG–Ugi complex formation buries as much as 21% (1064 $Å^2$) of the total accessible surface area of the Ugi molecule. In order to probe the role in UDG binding, if any, of the first 14 disordered residues in the present structure, the buried surface area calculation was carried out for the HSV-1 enzyme using the same algorithm and probe radius with and without the 14 residues in the HSV-1 UDG–Ugi complex. The results show that the buried surface area remains unaffected by the removal of these residues. The UDG–Ugi interface is highly polar and displays both electrostatic charge and shape complementarity. The coefficients of shape complementarity (46) are 0.61 and 0.59, respectively, for two molecules in the asymmetric unit compared with 0.64 observed in HSV-1 UDG–Ugi complex.

Despite the close similarity of the binding sites of the enzymes from the three sources, differences exist in details. It is particularly interesting to examine these differences between the *E.coli* and the human enzymes. The Cα atoms of the residues in the binding site of *Ec*UDG superpose with an r.m.s. deviation of 0.61 Å on those of the human enzyme and with an r.m.s. deviation of 0.32 Å on those in the HSV-1 enzyme (Fig. 4b). Thus, despite the greater sequence and overall structural similarity, the human enzyme exhibits larger deviations at the active site compared with the HSV-1 enzyme. The same result is obtained when all the atoms in the binding site are considered (1.14 and 0.62 Å in the

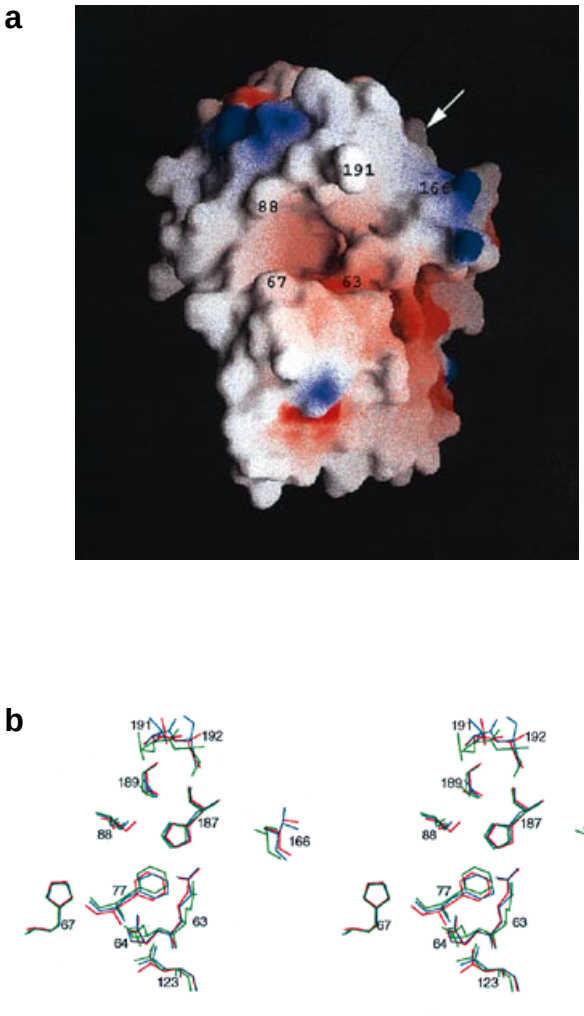

Figure 4. The binding of UDG. (**a**) The face of the molecule containing the binding site. The groove that binds DNA is indicated by an arrow. Figure prepared using GRASP (51). (**b**) Stereo view of the superposition of the catalytic site residues in the *E.coli* (red), human (green) and HSV-1 (blue) enzymes, prepared using INSIGHTII (52).

human and the HSV-1 enzymes, respectively). The largest overall deviations (1.2 Å in Cα) between the two enzymes occurs in the case of Ser 166 (*Ec*UDG numbering). Interestingly, the conformation of the catalytic residue Asp 64 is different in the two enzymes. The side chain carboxylate is staggered between the main chain amino and carbonyl groups ($\chi1 = 71°$ in molecule 1 and 65° in molecule 2) in the *E.coli* enzyme while it is *trans* to the amino group ($\chi1 = -158°$) in the human enzyme. The conformation of this aspartyl residue in the HSV-1 enzyme is similar to that in the *E.coli* enzyme. The same is the case with the adjacent Gln 63. Another residue which exhibits different conformations in the *E.coli* and human enzymes is Leu 191 which is believed to play an important role in the interaction of UDG with DNA (19,22). $\chi1$ in this residue is close to 180° and $\chi2$ close to 60° in both the molecules in *Ec*UDG and the HSV-1 enzyme, whereas these torsion angles are close to –60 and 180° in the human enzyme. It is also interesting to note that residues 194 and 195 present in the *E.coli* and the human enzymes are also involved in DNA binding (22). The first is a histidine in *Ec*UDG and a tyrosine in the human enzyme, while the second is an arginine in both the cases. Residue 194 is deleted and 195 is a lysine in the HSV-1 enzyme. The loop containing these residues exhibits structural differences among the enzymes and hence may influence enzyme activity. In addition, differences between the *E.coli* and the human enzymes in the variable regions of the molecule could well be important in interactions with accessory proteins *in vivo*.

Table 3. Ugi–*Ec*UDG interactions

Inhibitor residue	Protein residue	Distance (Å) Molecule 1	Molecule 2
Hydrogen bonds			
Glu 19 O	Gln 71 NE2	3.2	3.2
Glu 20 OE1	Ser 88 N	3.0	3.0
OE1	Pro 87 N	3.3	3.3
OE2	Ser 88 OG	3.3	3.2
Ser 21 OG	His 67 NE2	3.1	3.0
Leu 23 N	Gln 63 OE1	2.7	2.7
O	Gln 63 NE2	2.9	2.8
Glu 28 OE2	Ser 166 N	2.7	2.9
Asp 61 OD1	His 134 ND1	3.3	3.0
Tyr 65 OH	Ala 133 O	3.1	2.7
Gln 73 NE2	Pro 190 O	2.8	3.1
Residues within 4 Å from the inhibitor			
Gln 63, Tyr 66, His 67, Gln 71, Ile 85, Pro 86, Pro 87, Ser 88, Leu 90, Gln 132, Ala 133, His 134, Gly 165, Ser 166, His 167, Lys 171, His 187, Ser 189, Pro 190, Leu 191			

CONCLUSIONS

Deamination of cytosine residues to uracil is a frequent occurrence in the cell. Such conversions are likely to be even more prominent in the organisms with G+C rich genomes. Deamination results in the appearance of uracil in DNA, which unless repaired, leads to G:C→A:T mutations. Intracellular pathogens such as mycobacteria are at increased risk of cytosine deamination not only because of their G+C rich genomes (up to 70%) but also because of the unfavorable habitat of the host macrophages where they multiply. Thus, in these prokaryotic organisms, UDG is likely to be a crucial enzyme in DNA repair.

Description of the crystallographically solved 3D structure of the full-length wild-type UDG from *E.coli* complexed with the UDG inhibitor protein from PBS2, is significant in that the *E.coli* enzyme has served as a prototype for the various biological studies on the uracil excision repair pathway. It will now be possible to design various site-directed mutants of *Ec*UDG to study, in detail, the interaction of the different accessory factors with this important DNA repair enzyme. The present analysis, the first of a prokaryotic UDG, further establishes the high level of structural homology among the animal, bacterial and viral UDGs. Their DNA binding properties and mechanism of action are also essentially the same. Yet structural differences exist among them. A detailed comparison of the enzyme from the three sources permits the delineation of the relatively constant and variable regions of the molecule. *Ec*UDG, reported here, could be a prototype for UDGs from pathogenic prokaryotic organisms and could serve as a model for development of drugs against them. The regions of the molecule where the *E.coli* enzyme differs from the human enzyme would be particularly important in this effort.

ACKNOWLEDGEMENTS

The work was funded by the Department of Science and Technology (DST) and the Department of Biotechnology (DBT). The diffraction data were collected at the National Area Detector facility supported by the DST and DBT. Facilities at the Supercomputer Education and Research Centre (SERC) and the Interactive Graphics Based Molecular Modelling Facility and the Distributed Information Centre (both supported by the DBT) were used in the work. M.B.S. and S.R. have been supported by DBT post doctoral Research Associateships, whereas K.P. and P.H. have been recipients of K.S.Krishnan and Council of Scientific and Industrial Research fellowships, respectively.

REFERENCES

1 Lindahl,T. (1974) *Proc. Natl Acad. Sci. USA*, **71**, 3649–3653.
2 Lindahl,T., Ljungquist,S., Siegert,W., Nybert,B. and Sperens,B. (1977) *J. Biol. Chem.*, **252**, 3286–3294.
3 Duncan,B.K. and Chambers,J.A. (1984) *Gene*, **28**, 211–219.
4 Varshney,U., Hutcheon,T. and van de Sande,J.H. (1988) *J. Biol. Chem.*, **263**, 7776–7784.
5 Percival,K.J., Klein,M.B. and Burgers,P.M.J. (1989) *J. Biol. Chem.*, **264**, 2593–2598.
6 Olsen,L.C., Aasland,R., Krokan,H.E. and Hellard,D.E. (1989) *EMBO J.*, **8**, 3121–3125.
7 Muller-Weeks,S.J. and Caradonna,S. (1996) *Exp. Cell Res.*, **226**, 346–355.
8 Muller,S.J. and Caradonna,S. (1991) *Biochim. Biophys. Acta*, **1088**, 197–207.
9 Gallinari,P. and Jiricny,J. (1996) *Nature*, **383**, 735–738.
10 Meyer-Siegler,K., Mauro,D.J., Seal,G., Wurzer,J., DeRiel,J.K. and Sirover,M.A. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 8460–8464.
11 Duncan,B.K. (1981) In Boyer,P. (ed.), *The Enzymes*. Academic Press, Orlando, pp. 565–586.
12 Krokan,H.E., Standal,R. and Slupphaug,G. (1997) *Biochem. J.*, **325**, 1–15.
13 Mosbaugh,D.W. and Bennett,S.E. (1994) *Prog. Nucleic Acid Res. Mol. Biol.*, **48**, 315–369.
14 Lindahl,T. (1982) *Annu. Rev. Biochem.*, **51**, 61–87.
15 Varshney,U. and van de Sande,J.H. (1991) *Biochemistry*, **30**, 4055–4061.
16 Kumar,N.V. and Varshney,U. (1994) *Nucleic Acids Res.*, **18**, 3737–3741.
17 Kumar,N.V. and Varshney,U. (1997) *Nucleic Acids Res.*, **25**, 2336–2343.
18 Savva,R., McAuley-Hecht,K., Brown,T. and Pearl,L. (1995) *Nature*, **373**, 487–493.

19 Mol,C.D., Arvai,A.S., Slupphaug,G., Kavli,B., Alseth,I., Krokan,H.E. and Tainer,J.A. (1995) *Cell*, **80**, 869–878.
20 Mol,C.D., Arvai,A.S., Sanderson,R.J., Slupphaug,G., Kavli,B., Krokan,H.E., Mosbaugh,D.W. and Tainer,J.A. (1995) *Cell*, **82**, 701–708.
21 Savva,R. and Pearl,L.H. (1995) *Nat. Struct. Biol*., **2**, 752–755.
22 Slupphaug,G., Mol,C.D., Kavli,B., Arvai,A.S., Krokan,H.E. and Tainer,J.A. (1996) *Nature*, **384**, 87–92.
23 Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X. (1994) *Cell*, **76**, 357–369.
24 Reinisch,K.M., Chen,L., Verdine,G.L. and Lipscomb,W.N. (1995) *Cell*, **82**, 143–153.
25 Roberts,R.J. (1995) *Cell*, **82**, 9–12.
26 Kunkel,T.A. and Wilson,S.H. (1996) *Nature*, **384**, 25–26.
27 Panayotou,G., Brown,T., Barlow,T., Pearl,L.H. and Savva,R. (1998) *J. Biol. Chem*., **273**, 45–50.
28 Pearl,L.H. and Savva,R. (1995) *Trends Biochem. Sci*., **20**, 421–426.
29 Cone,R., Bonura,T. and Friedberg,E.C. (1980) *J. Biol. Chem*., **255**, 10354–10358.
30 Warner,H.R., Johnson,L.K. and Snustad,D.P. (1980) *J. Virol*., **33**, 535–538.
31 Roy,S., Purnapatre,K., Handa,P., Boyanapalli,M.M. and Varshney,U. (1998) *Protein Expr. Purif*., **13**, 155–162.
32 Otwinowski,Z. (1993) In Sawyer,L., Isaasc,N. and Bailey,S. (eds), *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*. SERC Daresbury Laboratory, Warrington, UK, pp. 56–62.
33 Navaza,J. (1994) *Acta Crystallogr*., **A50**, 157–163.
34 Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.F.,Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol*., **112**, 535–542.
35 Jones,T.A. (1978) *J. Appl. Crystallogr*., **11**, 268–272.
36 Brunger,A.T. (1992) *X-PLOR Manual, Version 3.1*. Yale University, New Haven, CT.
37 Engh,R.A. and Huber,R. (1991) *Acta Crystallogr*., **A47**, 392–400.
38 Bhat,T.N and Cohen,G.H., (1984) *J. Appl. Cryst*., **17**, 244–248.
39 Vijayan,M. (1980) In Diamond,R., Ramaseshan,S. and Venkatesan, K (eds), *Phase Evaluation and Some Aspects of the Fourier Refinement of Macromolecules*. Indian Academy of Sciences, Bangalore, pp 19.01–19.26.
40 Luzzati,V. (1952) *Acta Crystallogr*., **5**, 802–810.
41 Handa,P. and Varshney,U. (1998) *Ind. J. Biochem. Biophys*., **35**, 63–66.
42 Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
43 Lundqist,A.J., Berger,R.D., Bennett,S.E., Bolton,P.H and Mosbaugh,D.W. (1997) *J. Biol. Chem*., **272**, 21408–21412.
44 Sadasivan,C., Nagendra,H.G. and Vijayan,M. (1998) *Acta Crystallogr*., in press.
45 Lee,B. and Richards,F.M. (1971) *J. Mol. Biol*., **55**, 379–400.
46 Lawrence,M.C. and Colman,P.M. (1993) *J. Mol. Biol*., **234**, 946–950.
47 Jones,T.A., Zou,J.Y., Cowman,S.W. and Kjeldgaard,M. (1991) *Acta Crystallogr*., **A47**, 110–119.
48 Kraulis. P. (1991) *J. Appl. Cryst*., **24**, 946–950.
49 Merritt,E.A. and Murphy,M.E.P. (1994) *Acta Crystallogr*., **D50**, 869–873.
50 Laskowski,R.A., MacArthur,M.W., Moss,D.S and Thornton,J.M. (1993) *J. Appl. Cryst*., **26**, 283–291.
51 Nicholls,A., Bharadwaj,R. and Honig,B. (1993) *Biophys. J*., **64**, A116.
52 INSIGHTII (1993) *Insight II User Guide*. Biosym Technologies, San Diego, CA.