Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Structure of uracil-DNA glycosylase from *Mycobacterium tuberculosis*: insights into interactions with ligands

Prem Singh Kaushal,[a] Ramappa K. Talawar,[b] Umesh Varshney[b] and M. Vijayan[a]*

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

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

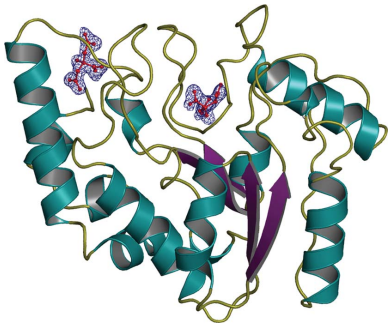
Received 16 April 2010
Accepted 15 June 2010

PDB Reference: uracil-DNA glycosylase, 3a7n.

Uracil *N*-glycosylase (Ung) is the most thoroughly studied of the group of uracil DNA-glycosylase (UDG) enzymes that catalyse the first step in the uracil excision-repair pathway. The overall structure of the enzyme from *Mycobacterium tuberculosis* is essentially the same as that of the enzyme from other sources. However, differences exist in the N- and C-terminal stretches and some catalytic loops. Comparison with appropriate structures indicate that the two-domain enzyme closes slightly when binding to DNA, while it opens slightly when binding to the proteinaceous inhibitor Ugi. The structural changes in the catalytic loops on complexation reflect the special features of their structure in the mycobacterial protein. A comparative analysis of available sequences of the enzyme from different sources indicates high conservation of amino-acid residues in the catalytic loops. The uracil-binding pocket in the structure is occupied by a citrate ion. The interactions of the citrate ion with the protein mimic those of uracil, in addition to providing insights into other possible interactions that inhibitors could be involved in.

1. Introduction

The occurrence of uracil in DNA is promutagenic and arises from spontaneous deamination of cytosine, resulting in G·U base pairing (Lindahl & Nyberg, 1974). A deficiency in uracil excision from a G·U base pair prior to replication can lead to GC–AT mutations. Uracil may also arise in DNA by misincorporation of dUMP (in place of dTMP) by DNA polymerase(s), resulting in A·U base paring (Tye & Lehman, 1977). In an A·U base pair uracil is not directly mutagenic, but it can cause cytotoxicity and hamper recognition of regulatory sequences by proteins (Mosbaugh & Bennett, 1994). This event is kept to a minimum by dUTPase present in the cell, which maintains a low level of dUTP. In order to maintain genomic integrity, the cells possess uracil-DNA glycosylases (UDGs), which are the first enzymes of the base-excision repair (BER) pathway for the excision of uracil (Lindahl, 1974). UDGs excise uracil by cleaving the N-glycosidic bond. The resulting AP site is cleaved by AP endonucleases; the damaged sugar residue is then removed by the presence of exonuclease activities and the single-nucleotide gap thus generated is filled in by DNA synthesis (Kubota *et al.*, 1996; Nicholl *et al.*, 1997; Parikh *et al.*, 1997). Five different families of UDGs have been reported to date. Of these, family 1 UDGs are referred to as UNG when the protein is eukaryotic and as Ung when it is prokaryotic, or collectively as UNGs. UNGs are a highly efficient, ubiquitous and conserved class of UDGs. *Escherichia coli* Ung (EcUng) is the prototype of the family 1 UDGs and is encoded by the *ung* gene (Lindahl *et al.*, 1977; Varshney *et al.*, 1988). UNGs are inhibited by a proteinaceous inhibitor known as uracil-DNA glycosylase inhibitor (Ugi) which is coded by *Bacillus subtilis* phages PBS-1 and PBS-2 (Cone *et al.*, 1980; Warner *et al.*, 1980; Wang & Mosbaugh, 1988). Ugi protects the phage genome, which contains uracil, from host UNG attack. UNGs are also inhibited by uracil and some of its derivatives (Krokan & Wittwer, 1981; Blaisdell & Warner, 1983; Focher *et al.*, 1993; Jiang *et al.*, 2005; Krosky *et al.*, 2006; Chung *et al.*, 2009).



In 1995, X-ray crystal structures of the UNG–Ugi complex and of uncomplexed UNG from human (HsUNG; Mol, Arvai, Sanderson *et al.*, 1995; Mol, Arvai, Slupphaug *et al.*, 1995) and herpes simplex virus (HSVUNG; Savva & Pearl, 1995; Savva *et al.*, 1995) were reported. The structure of EcUng in complex with Ugi was first reported from this laboratory (Ravishankar *et al.*, 1998). Subsequently, several other crystal structures of EcUng and its complexes have been reported (Xiao *et al.*, 1999; Putnam *et al.*, 1999; Saikrishnan *et al.*, 2002). More recently, the structures of uncomplexed Ungs from *Gadus morhua* (GmUNG; Leiros *et al.*, 2003), *Deinococcus radiodurans* (DrUng; Leiros *et al.*, 2005) and *Vibrio cholerae* (VcUng; Raeder *et al.*, 2010) and the structure of a complex of Ugi with UNG from Epstein–Barr virus (EBVUNG; Geoui *et al.*, 2007) have also become available. The structures of complexes of HsUNG and EcUng with DNA have been reported (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998, 2000; Werner *et al.*, 2000; Bianchet *et al.*, 2003). The structures of complexes of UNG with uracil and its derivatives have also been reported (Savva *et al.*, 1995; Slupphaug *et al.*, 1996; Xiao *et al.*, 1999; Parikh *et al.*, 1998; Werner *et al.*, 2000; Bianchet *et al.*, 2003; Krosky *et al.*, 2006; Chung *et al.*, 2009).

The genome of *Mycobacterium tuberculosis*, the causative agent of tuberculosis in humans, is highly G+C-rich. *M. tuberculosis* also infects the host macrophage, which generates both reactive nitrogen intermediates (RNI) and reactive oxygen species (ROS), which are known to deaminate cytosines in DNA. Thus, the genome of *M. tuberculosis* could be more susceptible to the generation of G·U pairs (Wink *et al.*, 1991) and make Ung a crucial enzyme for survival of the mycobacterium. In fact, studies with *M. smegmatis* and *M. tuberculosis* have also suggested that Ung is a crucial enzyme for survival of mycobacteria in mouse or macrophage models (Venkatesh *et al.*, 2003; Sassetti & Rubin, 2003). Therefore, as part of our structural genomics program on mycobacterial proteins (Datta *et al.*, 2000; Saikrishnan *et al.*, 2003, 2005; Roy *et al.*, 2004; Krishna *et al.*, 2006; Selvaraj *et al.*, 2007; Das *et al.*, 2006; Kaushal *et al.*, 2008; Chetnani *et al.*, 2009), we undertook structural studies of *M. tuberculosis* Ung (MtUng). The recent structural analysis of the MtUng–Ugi complex reported from this laboratory (Kaushal *et al.*, 2008) led to a rationale for the decreased stability of the MtUng–Ugi complex in comparison to the EcUng–Ugi complex. It also provided insights into the molecular mechanism of action of the enzyme and the relatively invariant and variable features of the molecule. Here, we report the crystal structure of MtUng at 1.95 Å resolution with one citrate ion bound to the uracil-recognition pocket in the active site and another bound on the surface of the protein molecule. A comparison of the present structure with the Ugi and DNA complexes provides insights into the structural variations resulting from complexation. The structure also yields further information on possible enzyme–inhibitor interactions. An analysis of available UNG sequences in terms of structure helps in identifying the central conserved region of the molecule and its importance.

2. Materials and methods

2.1. Expression and purification of *M. tuberculosis* Ung

The MtUng–Ugi construct that was previously used to produce the MtUng–Ugi complex (Singh *et al.*, 2006) was subjected to *ugi* removal by digesting it with *Eco*RI (MBI Fermentas) to release the appropriate 500 bp fragment. The vector was religated using T4-DNA ligase to obtain MtUng. The above construct was transformed into *E. coli* BL21 (DE3) cells and MtUng was purified in essentially the same way as the MtUng–Ugi complex (Singh *et al.*, 2006). Fractions containing a near-homogenous preparation of MtUng were pooled, dialyzed against 10 m*M* Tris–HCl pH 7.5 and estimated by Bradford's method using bovine serum albumin as standard (Sedmak & Grossberg, 1977). The final preparation was concentrated to 7 mg ml^{-1} and used for crystallization.

2.2. Crystallization and data collection

MtUng was screened for crystallization using the Crystal Screen, Crystal Screen 2 and Index kits from Hampton Research. 3 µl 7 mg ml^{-1} protein solution in 10 m*M* Tris–HCl pH 7.4, 200 m*M* NaCl and 3 µl precipitant solution were used in the microbatch method of crystallization at 298 K. After 7 d, broom-shaped crystals appeared when 1.8 *M* triammonium citrate pH 7.0 was used as the precipitant (Index kit condition No. 21). This condition was further optimized by the addition of organic solvents. Diffraction-quality single crystals were obtained when 10% 2-propanol was added to the precipitant. These crystals diffracted to a resolution of 1.95 Å. No additional cryoprotectant was used prior to data collection. Diffraction data were collected at low temperature (100 K) using a MAR Research image-plate system (diameter 345 mm) with Osmic mirrors and a Bruker AXS MICROSTAR ULTRA II rotating-anode X-ray generator. Intensity data were processed and scaled using *MOSFLM* and *SCALA* from the *CCP*4 program package (Collaborative Computational Project, Number 4, 1994). Intensities were converted to structure-factor amplitudes using *TRUNCATE* in *CCP*4.

2.3. Structure determination and refinement

The crystal structure of MtUng was determined by molecular replacement using the coordinates of the Ung molecule in its complex with the inhibitor Ugi (PDB code 2zhx; Kaushal *et al.*, 2008) as the search model. *Phaser* (Storoni *et al.*, 2004) gave the best solution for one molecule in the asymmetric unit, with a log-likelihood gain of 1009.1 and a *Z* score of 25.1. The Matthews coefficient (2.5 Å^3 Da^{-1}; Matthews, 1968) also confirmed the presence of one molecule in the asymmetric unit, with a solvent content of 51.2%. The model was initially refined using *CNS* v.1.2 (Brünger *et al.*, 1998). Model building using *Coot* (Emsley & Cowtan, 2004) was alternated with iteration of rigid-body refinement, positional refinement and simulated annealing followed by individual temperature-factor refinement. Water O atoms were identified on the basis of peaks in both $2F_o - F_c$ and $2F_o - F_c$ maps. During the initial stage of refinement, two patches of unexplained electron density appeared (Supplementary Fig. S1[1]). Careful examination of the density indicated that it might correspond to two citrate ions that had been picked up from the mother liquor. The two citrate ions were included in the refinement along with the protein molecule. A simulated-annealing OMIT map was calculated by omitting the citrate ions. Electron density in the OMIT map confirmed the presence of these ions (Fig. 1). The final cycles of refinement were carried out using *REFMAC* (Murshudov *et al.*, 1997) in *CCP*4 employing the TLS option (Winn *et al.*, 2003), using six groups indicated by the application of *TLSMD* (Painter & Merritt, 2006). The structure refined to acceptable values of *R* and R_{free} with good geometry. In the final refined model of MtUng, two N-terminal residues which could not be located previously in the MtUng–Ugi complex and six of the 11 residues of the histidine tag were defined. The final model of MtUng consisted of 233 amino-acid residues in a single polypeptide chain comprising residues −5 to 227 of the amino-acid sequence, two citrate ions and 208 water molecules. The side

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: BE5148).

Table 1
X-ray crystal data, refinement and model statistics.

Values in parentheses are for the highest resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	
a	44.48
b	63.67
c	86.40
Unit-cell volume (Å^3)	24668.7
V_M (Å^3 Da^{-1})	2.52
Solvent content (%)	51.2
No. of molecules in asymmetric unit	1
Resolution range (Å)	26.24–1.95 (2.06–1.95)
No. of observed reflections	88441 (11929)
No. of unique reflections	18238 (2583)
Completeness (%)	98.2 (97.7)
Multiplicity	4.8 (4.6)
Average $I/\sigma(I)$	13.7 (3.2)
R_{merge}† (%)	10.0 (49.3)
Refinement and model statistics	
R factor (%)	18.1
R_{free}‡ (%)	19.9
R.m.s. deviation from ideal	
Bond lengths (Å)	0.007
Bond angles (°)	1.1
B values (Å^2)	
From Wilson plot	30.0
From refinement	18.9
Residues in Ramachandran plot§ (%)	
Core regions	93.7
Allowed regions	5.8
Generously allowed regions	0.5
Disallowed regions	0.0

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th observation of reflection *hkl* and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations *i* of reflection *hkl*. ‡ R_{free} was calculated from a randomly selected 5% (885) of unique reflections that were omitted from structure refinement. § Calculated for nonglycine and nonproline residues using *PROCHECK* (Laskowski *et al.*, 1993).

chains and residues in loop regions are better defined in this structure than in that of the MtUng–Ugi complex (Supplementary Fig. S2) presumably on account of the better resolution of the data in the present study. Crystal data and data-collection, refinement and model statistics are given in Table 1.

2.4. Analysis of the structure

The refined model was evaluated using *PROCHECK* (Laskowski *et al.*, 1993). Structures were superposed using *ALIGN* (Cohen, 1997). *HBPLUS* was used to identify hydrogen bonds (McDonald & Thornton, 1994). Figures were generated using *PyMOL* (DeLano, 2002).

2.5. Analysis of sequences

Sequence analysis was carried out using the online server *ConSurf* v.3.0 (Glaser *et al.*, 2003; Landau *et al.*, 2005; http://consurf.tau.ac.il/index.html). Homologous sequences of UNG were retrieved from the SWISS-PROT database (Bairoch & Apweiler, 1999) using *PSI-BLAST* (Altschul *et al.*, 1997) and were aligned using *MUSCLE* (Edgar, 2004). The phylogenic tree was constructed using the neighbour-joining algorithm (Saitou & Nei, 1987). Position-specific conservation scores for residues in the sequences were calculated using the empirical Bayesian method (Mayrose *et al.*, 2004) and then mapped onto the three-dimensional structure of MtUng.

3. Results and discussion

3.1. Overall features

The amino-acid sequences of UNGs of known X-ray structure are highly homologous, with sequence identity between pairs of molecules varying between 41 and 47% despite being obtained from very different species. The homology is also reflected in the three-dimensional structure. R.m.s. deviations in C^α positions between pairs of these structures are in the range 1–1.3 Å. Each molecule, including that for which the structure is reported here, is made up of a central β-sheet containing four strands (β1, 62–66; β2, 123–127; β3, 163–169; β4, 184–189; MtUng numbering) flanked by eight helices (−2 to 3, 12–17, 22–37, 92–105, 116–120, 145–158, 170–197 and 206–216) (Fig. 1). These secondary-structural elements are connected by loops, which account for about 50% of the amino-acid residues in the molecule. Most of the loops are concentrated in a region that is involved in substrate binding. As demonstrated earlier, the sequences and conformation of the central region of the molecule ranging from residue 58 to residue 195 are substantially conserved among UNGs of known three-dimensional structure (Kaushal *et al.*, 2008). All of the catalytic loops, namely loop I (66–72), loop II (89–93), loop III (124–128), loop IV (169–170) and loop V (191–199), belong to this region. The N- and C-terminal stretches exhibit large variations in sequence and structure. In the present structure, a citrate ion is located in the uracil-recognition pocket and another is located on the surface of the molecule between the N-terminal histidine tag and the 41–46 stretch in a loop (Fig. 1).

3.2. Comparison with Ugi and DNA complexes

The complexation of the proteins from different sources with the proteinaceous inhibitor Ugi has been extensively investigated (Cone *et al.*, 1980; Warner *et al.*, 1980; Wang & Mosbaugh, 1988; Mol, Arvai,

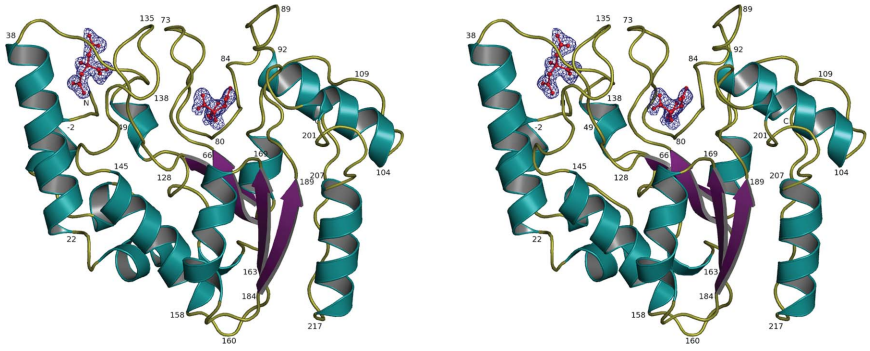


Figure 1
Stereoview of the Ung molecule. The bound citrate ions are shown in ball-and-stick representation. Electron density corresponding to the two bound citrate ions in an $F_o - F_c$ OMIT map is also shown. The contours are at the 3.5σ level.

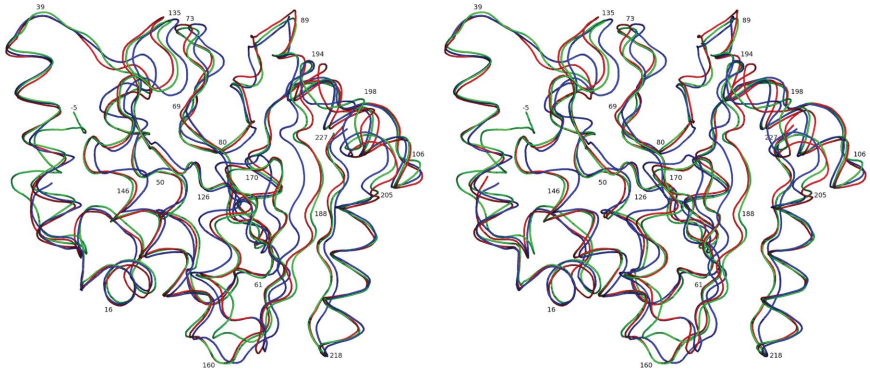


Figure 2
Stereoview of the superposition of the C^{α} traces of the structure of MtUng (green), MtUng in the MtUng–Ugi complex (red) and MtUng in the model of the MtUng–DNA complex (blue).

Sanderson *et al.*, 1995; Mol, Arvai, Slupphaug *et al.*, 1995; Ravishankar *et al.*, 1998; Putnam *et al.*, 1999; Saikrishnan *et al.*, 2002; Acharya *et al.*, 2002, 2003; Geoui *et al.*, 2007; Kaushal *et al.*, 2008). The interactions of UNG with DNA have also been structurally explored, although much less extensively (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998, 2000; Werner *et al.*, 2000; Bianchet *et al.*, 2003). Although Ung was originally believed to be a single-domain protein, we subsequently showed that it contains two domains (domain I, 12–81 and 124–156; domain II, 83–114 and 163–224) connected by two linker regions (115–123 and 159–162). We also showed that these two domains close by about 10° on the DNA molecule in the known structures of complexes involving EcUng and HsUNG (Saikrishnan *et al.*, 2002). This information was made use of in constructing a model of MtUng with DNA (Kaushal *et al.*, 2008). In the refined model of the MtUng–DNA complex the domain closure was 6°. Interestingly, the domains open up by 4.5° in the crystal structure of the MtUng–Ugi complex (Kaushal *et al.*, 2008).

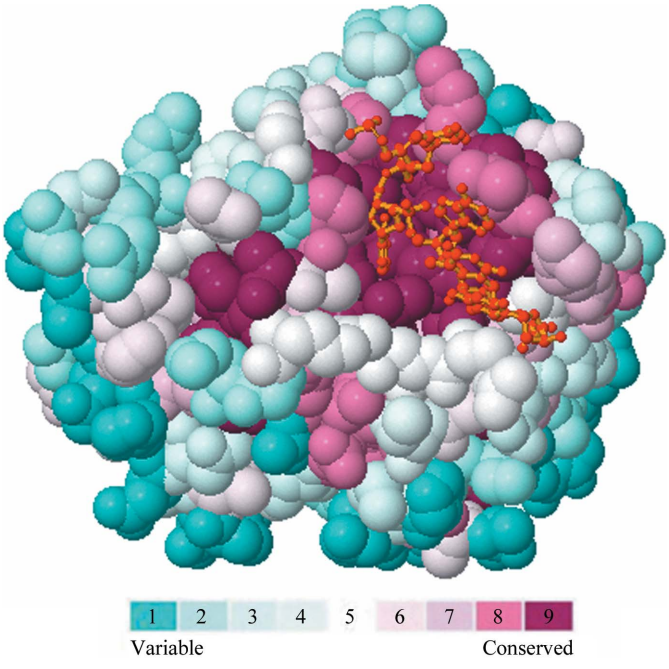


Figure 3
The evolutionary conservation score of amino acids in 293 sequences homologous to MtUng mapped onto the structure of MtUng. The bound DNA molecule in the model constructed previously (Kaushal *et al.*, 2008) is also shown.

In addition to the gross effects involving domain movement, it is also interesting to explore the effect of Ugi and DNA binding on specific regions of the protein, particularly the catalytic loops, in MtUng (Fig. 2). The water-activated loop (loop I) is largely unaffected by Ugi binding. The same is true for the uracil-recognition loop (loop III) and the Gly-Ser loop (loop IV) in the case of Ugi binding. The movement of the former and that of loop II (the proline-rich loop) in the case of DNA binding can be explained almost exclusively in terms of domain closure. In loop IV (the Gly-Ser loop), Ser in HsUNG and EcUng is replaced by an arginine in MtUng. This arginine interacts with DNA, which explains the movement of loop IV on DNA binding. Loop II (the proline-rich loop) in MtUng is different in length and composition compared with those of Ung proteins from other sources. It is one residue longer in MtUng and protrudes towards Ugi in the complex between the two proteins, thus engendering additional interactions between the two. Therefore, not surprisingly, complexation with Ugi results in movement of this loop. The interactions of this loop with DNA are much less intimate. The leucine loop (loop V) moves on complexation in both cases. This is hardly surprising as Leu195 makes strong hydrophobic interactions in both cases. These interactions are stronger in the Ugi complex, in which the leucyl side chain is embedded in the hydrophobic pocket of Ugi. The movements in the loop are therefore larger in the complex with Ugi.

A secondary-structural element which moves as a whole on DNA binding is the 170–176 helix. Interestingly, movement of this helix was noticed when the crystal structures of EcUng and HsUNG were compared with those of their complexes with DNA (Saikrishnan *et al.*, 2002), although the nature of the movement is not exactly the same in the three cases. It is satisfying, and indeed reassuring, that movements of the type experimentally observed in EcUng and HsUng also appeared in a model.

3.3. Conservation of catalytic and other central residues

The sequences of 293 UNGs from different sources, with a sequence identity of 30–93%, were aligned employing the procedure outlined in §2. Of these, 276 are bacterial and 13 are viral. Two are from yeast and two are mammalian. Thus, bacterial Ungs dominate the set. Alignment of sequences from a representative set of sources is shown in Supplementary Figure S3. There are a total of 17 residues that are fully conserved in all 293 sequences. Most of them belong to the catalytic loops or are in their immediate neighbourhood. Of the remaining few, two are glycines which could be conformationally important. Another, Phe81, stacks on uracil in the substrate-binding

pocket. Another invariant residue, Trp118, is involved in several hydrophobic interactions, especially in the C-terminal region.

The pattern outlined above is also discernable when the position-specific conservation scores calculated using the empirical Bayesian method (Mayrose *et al.*, 2004) are mapped onto the three-dimensional structure of MtUng (Fig. 3). The score indicates 22 residues as the most conserved. The 17 fully conserved residues form a subset of these 22 residues. As can be clearly seen from Fig. 3, much of the conserved region is involved in interaction with DNA.

3.4. Interactions at the binding pocket

The catalytic mechanism of the family 1 UDGs has been thoroughly explained using a variety of approaches (Lindahl *et al.*, 1977; Varshney *et al.*, 1988; Parikh *et al.*, 1998; Slupphaug *et al.*, 1996). It involves the flipping-out of the uracil base, which is then locked in a pocket in UNG before excision. The Ung–uracil interactions have been characterized by X-ray studies of complexes of uracil (Savva *et al.*, 1995; Slupphaug *et al.*, 1996; Xiao *et al.*, 1999; Bianchet *et al.*, 2003), uridine (Krusong *et al.*, 2006), ligands containing uracil (Krosky *et al.*, 2006; Chung *et al.*, 2009) and DNA itself (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998, 2000; Werner *et al.*, 2000). The residues constituting this pocket are highly conserved across species. In the present structure the uracil-binding pocket is occupied by a citrate ion from the precipitant solution.

The interactions of the citrate ion at the binding pocket are illustrated in Fig. 4(*a*). Those of the uracil molecule in its complex with HsUng are shown in Fig. 4(*b*) for comparison. The interactions of the ligand with the enzyme are essentially the same in all the relevant crystal structures. In addition to the hydrogen bonds that are specifically indicated in Fig. 4(*b*), uracil also makes a stacking interaction with the side chain of Phe158 (HsUNG numbering). One of the planar carboxylate groups in the citrate ion is nearly coplanar with the six-membered ring of uracil and hence stacking is also seen in the citrate complex. The binding pocket of previously reported structures contain an invariant water molecule hydrogen bonded to Pro146 O and the side chains of Asp145 and His148 (HsUNG numbering; Parikh *et al.*, 1998). This water molecule is believed to be involved in catalysis. A water molecule occurs at the same location in the present structure, although the interaction involving histidine is abolished as this amino acid is substituted by a proline in MtUng.

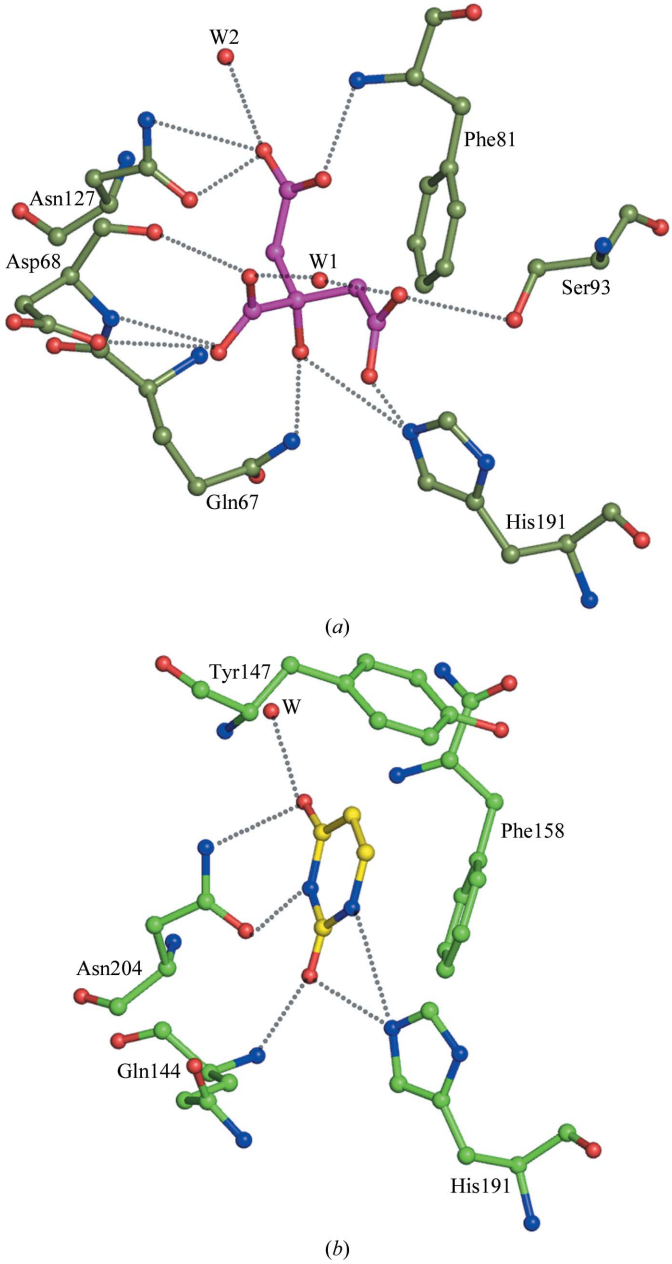


Figure 4
Interactions of (*a*) MtUng with citrate and (*b*) HsUNG with uracil.

4. Conclusions

The structure of MtUng, although very similar to that of the enzyme from other sources, exhibits differences in the terminal regions and in some catalytic loops. The differences in the loops are reflected in their movements during complexation with Ugi or DNA. Interestingly, the domains close slightly on DNA binding, while they open slightly on Ugi binding. A detailed analysis of 293 homologous sequences brings out the highly conserved nature of the DNA-binding site. The uracil-binding pocket of DNA in the present structure is occupied by a citrate ion. The citrate ion makes all the interactions that uracil does with the enzyme. The former is also involved in other interactions with the binding pocket. Information on those interactions could be useful in inhibitor design.

Data were collected at the X-ray Facility for Structural Biology at the Molecular Biophysics Unit supported by the Department of Science and Technology, Government of India. Computations were performed partially at the Bioinformatics Centre and the Graphics Facility, both of which are supported by the Department of Biotechnology (DBT). The work forms part of a DBT-sponsored research programme. During the period of the work, MV was supported successively by a DBT Distinguished Research Professorship and a DAE Homi Bhabha Professorship.

References

Acharya, N., Kumar, P. & Varshney, U. (2003). *Microbiology*, **149**, 1647–1658.
Acharya, N., Roy, S. & Varshney, U. (2002). *J. Mol. Biol.* **321**, 579–590.
Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Bairoch, A. & Apweiler, R. (1999). *Nucleic Acids Res.* **27**, 49–54.
Bianchet, M. A., Seiple, L. A., Jiang, Y. L., Ichikawa, Y., Amzel, L. M. & Stivers, J. T. (2003). *Biochemistry*, **42**, 12455–12460.
Blaisdell, P. & Warner, H. (1983). *J. Biol. Chem.* **258**, 1603–1609.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read,

R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chetnani, B., Das, S., Kumar, P., Surolia, A. & Vijayan, M. (2009). *Acta Cryst.* D**65**, 312–325.
Chung, S., Parker, J. B., Bianchet, M., Amzel, L. M. & Stivers, J. T. (2009). *Nature Chem. Biol.* **5**, 407–413.
Cohen, G. E. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cone, R., Bonura, T. & Friedberg, E. C. (1980). *J. Biol. Chem.* **255**, 10354–10358.
Das, S., Kumar, P., Bhor, V., Surolia, A. & Vijayan, M. (2006). *Acta Cryst.* D**62**, 628–638.
Datta, S., Prabu, M. M., Vaze, M. B., Ganesh, N., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2000). *Nucleic Acids Res.* **28**, 4964–4973.
DeLano, W. L. (2002). *PyMOL Molecular Viewer.* http://www.pymol.org.
Edgar, R. C. (2004). *Nucleic Acids Res.* **32**, 1792–1797.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Focher, F., Verri, A., Spadari, S., Manservigi, R., Gambino, J. & Wright, G. E. (1993). *Biochem. J.* **292**, 883–889.
Geoui, T., Buisson, M., Tarbouriech, N. & Burmeister, W. P. (2007). *J. Mol. Biol.* **366**, 117–131.
Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). *Bioinformatics*, **19**, 163–164.
Jiang, Y. L., Krosky, D. J., Seiple, L. & Stivers, J. T. (2005). *J. Am. Chem. Soc.* **127**, 17412–17420.
Kaushal, P. S., Talawar, R. K., Krishna, P. D. V., Varshney, U. & Vijayan, M. (2008). *Acta Cryst.* D**64**, 551–560.
Krishna, R., Manjunath, G. P., Kumar, P., Surolia, A., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2006). *Nucleic Acids Res.* **34**, 2186–2195.
Krokan, H. & Wittwer, C. U. (1981). *Nucleic Acids Res.* **9**, 2599–2613.
Krosky, D. J., Bianchet, M. A., Seiple, L., Chung, S., Amzel, L. M. & Stivers, J. T. (2006). *Nucleic Acids Res.* **34**, 5872–5879.
Krusong, K., Carpenter, E. P., Bellamy, S. R., Savva, R. & Baldwin, G. S. (2006). *J. Biol. Chem.* **281**, 4983–4992.
Kubota, Y., Nash, R. A., Klungland, A., Schar, P., Barnes, D. E. & Lindahl, T. (1996). *EMBO J.* **15**, 6662–6670.
Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. & Ben-Tal, N. (2005). *Nucleic Acids Res.* **33**, W299–W302.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leiros, I., Moe, E., Lanes, O., Smalås, A. O. & Willassen, N. P. (2003). *Acta Cryst.* D**59**, 1357–1365.
Leiros, I., Moe, E., Smalås, A. O. & McSweeney, S. (2005). *Acta Cryst.* D**61**, 1049–1056.
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.
Lindahl, T. & Nyberg, B. (1974). *Biochemistry*, **13**, 3405–3410.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mayrose, I., Graur, D., Ben-Tal, N. & Pupko, T. (2004). *Mol. Biol. Evol.* **21**, 1781–1791.
McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.
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., Alseth, I., Krokan, H. E. & Tainer, J. A. (1995). *Cell*, **80**, 869–878.
Mosbaugh, D. W. & Bennett, S. E. (1994). *Prog. Nucleic Acid Res. Mol. Biol.* **48**, 315–370.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Nicholl, I. D., Nealon, K. & Kenny, M. K. (1997). *Biochemistry*, **36**, 7557–7566.
Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* D**62**, 439–450.
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., Mol, C. D. & Tainer, J. A. (1997). *Structure*, **5**, 1543–1550.
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.
Putnam, C. D., Shroyer, M. J., 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.
Raeder, I. L. U., Moe, E., Willassen, N. P., Smalås, A. O. & Leiros, I. (2010). *Acta Cryst.* F**66**, 130–136.
Roy, S., Gupta, S., Das, S., Sekar, K., Chatterji, D. & Vijayan, M. (2004). *J. Mol. Biol.* **339**, 1103–1113.
Saikrishnan, K., Bidya Sagar, M., Ravishankar, R., Roy, S., Purnapatre, K., Handa, P., Varshney, U. & Vijayan, M. (2002). *Acta Cryst.* D**58**, 1269–1276.
Saikrishnan, K., Jeyakanthan, J., Venkatesh, J., Acharya, N., Sekar, K., Varshney, U. & Vijayan, M. (2003). *J. Mol. Biol.* **331**, 385–393.
Saikrishnan, K., Kalapala, S. K., Varshney, U. & Vijayan, M. (2005). *J. Mol. Biol.* **345**, 29–38.
Saitou, N. & Nei, M. (1987). *Mol. Biol. Evol.* **4**, 406–425.
Sassetti, C. M. & Rubin, E. J. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 12989–12994.
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.
Sedmak, J. J. & Grossberg, S. E. (1977). *Anal. Biochem.* **79**, 544–552.
Selvaraj, M., Roy, S., Singh, N. S., Sangeetha, R., Varshney, U. & Vijayan, M. (2007). *J. Mol. Biol.* **372**, 186–193.
Singh, P., Talawar, R. K., Krishna, P. D. V., Varshney, U. & Vijayan, M. (2006). *Acta Cryst.* F**62**, 1231–1234.
Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E. & Tainer, J. A. (1996). *Nature (London)*, **384**, 87–92.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Tye, B. K. & Lehman, I. R. (1977). *J. Mol. Biol.* **117**, 293–306.
Varshney, U., Hutcheon, T. & van de Sande, J. H. (1988). *J. Biol. Chem.* **263**, 7776–7784.
Venkatesh, J., Kumar, P., Krishna, P. S., Manjunath, R. & Varshney, U. (2003). *J. Biol. Chem.* **278**, 24350–24358.
Wang, Z. & Mosbaugh, D. W. (1988). *J. Bacteriol.* **170**, 1082–1091.
Warner, H. R., Johnson, L. K. & Snustad, D. P. (1980). *J. Virol.* **33**, 535–538.
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.
Wink, D. A., Kasprzak, K. S., Maragos, C. M., Elespuru, R. K., Misra, M., Dunams, T. M., Cebula, T. A., Koch, W. H., Andrews, A. W., Allen, J. S. & Keefer, L. K. (1991). *Science*, **254**, 1001–1003.
Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
Xiao, G., Tordova, M., Jagadeesh, J., Drohat, A. C., Stivers, J. T. & Gilliland, G. L. (1999). *Proteins*, **35**, 13–24.