Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Unique features of the structure and interactions of mycobacterial uracil-DNA glycosylase: structure of a complex of the *Mycobacterium tuberculosis* enzyme in comparison with those from other sources

Prem Singh Kaushal,[a] Ramappa K. Talawar,[b] P. D. V. Krishna,[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 17 October 2007
Accepted 22 February 2008

PDB Reference: uracil-DNA glycosylase–uracil-DNA glycosylase inhibitor complex, 2zhx, r2zhxsf.

Uracil-DNA glycosylase (UNG), a repair enzyme involved in the excision of uracil from DNA, from mycobacteria differs from UNGs from other sources, particularly in the sequence in the catalytically important loops. The structure of the enzyme from *Mycobacterium tuberculosis* (*Mt*Ung) in complex with a proteinaceous inhibitor (Ugi) has been determined by X-ray analysis of a crystal containing seven crystallographically independent copies of the complex. This structure provides the first geometric characterization of a mycobacterial UNG. A comparison of the structure with those of other UNG proteins of known structure shows that a central core region of the molecule is relatively invariant in structure and sequence, while the N- and C-terminal tails exhibit high variability. The tails are probably important in folding and stability. The mycobacterial enzyme exhibits differences in UNG–Ugi interactions compared with those involving UNG from other sources. The *Mt*Ung–DNA complex modelled on the basis of the known structure of the complex involving the human enzyme indicates a domain closure in the enzyme when binding to DNA. The binding involves a larger burial of surface area than is observed in binding by human UNG. The DNA-binding site of *Mt*Ung is characterized by the presence of a higher proportion of arginyl residues than is found in the binding site of any other UNG of known structure. In addition to the electrostatic effects produced by the arginyl residues, the hydrogen bonds in which they are involved compensate for the loss of some interactions arising from changes in amino-acid residues, particularly in the catalytic loops. The results arising from the present investigation represent unique features of the structure and interaction of mycobacterial Ungs.

1. Introduction

Uracil-DNA glycosylases (UDGs) constitute a highly conserved superfamily of enzymes that are present in a wide variety of organisms. Uracil, a promutagenic base, mostly occurs in DNA as a result of the spontaneous deamination of cytosine (Lindahl & Nyberg, 1974). Occasionally, it can also occur owing to the misincorporation of dUMP in place of dTMP by DNA polymerase (Tye & Lehman, 1977). UDG is involved in the excision of uracil from DNA as a first step in the base-excision repair pathway (Lindahl, 1974). This is followed by further repair catalyzed by AP endonuclease, deoxyribophosphodiesterase, DNA polymerase and DNA ligase (Kubota *et al.*, 1996; Nicholl *et al.*, 1997; Parikh *et al.*, 1997). Five different families of UDGs have been reported to

doi:10.1107/S090744490800512X

date. Of these, the family 1 UDGs (referred to as UNG when the protein is eukaryotic and Ung when the protein is prokaryotic) have been thoroughly characterized.

UNG is inhibited by free uracil and some its derivatives (Krokan & Wittwer, 1981; Blaisdell & Warner, 1983; Focher *et al.*, 1993; Jiang *et al.*, 2005; for convenience, the acronym UNG is used to collectively refer to DNA glycosylases from different sources). It is also inhibited by uracil-DNA glycosylase inhibitor (Ugi), which forms a tight 1:1 complex with UNG (Bennett *et al.*, 1993). Ugi is an early gene product of *Bacillus subtilis* phages PBS-1 and PBS-2 (Cone *et al.*, 1980; Warner *et al.*, 1980; Wang & Mosbaugh, 1988). Interestingly, the DNA of this virus naturally contains uracil and Ugi is designed to protect it from attack by UNG. Ugi has been used extensively in the study of UNG. In particular, the crystal structures of the Ugi complexes of UNG from four different organisms, namely humans (*Hs*UNG; Mol, Arvai, Sanderson *et al.*, 1995), herpes simplex virus (HSVUng; Savva *et al.*, 1995), *Escherichia coli* (*Ec*Ung; Ravishankar *et al.*, 1998; Putnam *et al.*, 1999; Saikrishnan *et al.*, 2002) and Epstein–Barr virus (EBVUng; Geoui *et al.*, 2007), have been determined. The structures of uncomplexed UNG from three of these organisms (Mol, Arvai, Slupphaug *et al.*, 1995; Savva & Pearl, 1995; Xiao *et al.*, 1999; Putnam *et al.*, 1999; Saikrishnan *et al.*, 2002), *Gadus morhua* (*Gm*UNG; Leiros *et al.*, 2003) and *Deinococcus radiodurans* (*Dr*Ung; Leiros *et al.*, 2005) are also available. Complexes of *Hs*UNG and *Ec*Ung with DNA fragments have also been reported (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998, 2000; Werner *et al.*, 2000; Bianchet *et al.*, 2003). These structural studies have led to detailed characterization of the interactions of UNG with Ugi and DNA. These studies, together with extensive biochemical and genetic investigations, have provided a reasonably comprehensive picture of the molecular mechanism of action of the enzyme.

Pathogenic mycobacteria such as *Mycobacterium tuberculosis* are at a high risk of cytosine deamination not only because of the high G+C content of their genomes but also because of their exposure to reactive oxygen species (ROS) and the reactive nitrogen intermediate (RNI) produced by the host macrophage. Therefore, the role of Ungs is particularly important in these organisms. The Ung from *M. tuberculosis* (*Mt*Ung) has recently been cloned, expressed, biochemically characterized and crystallized (Acharya *et al.*, 2003; Singh *et al.*, 2006). Transposon-site hybridization (TraSH) experiments have suggested that Ung is essential for the survival of *M. tuberculosis* in the mouse model (Sassetti & Rubin, 2003). Very recently, a Ung belonging to a different family (UdgB) has also been identified in the organism and has been characterized (Srinath *et al.*, 2007). As part of structural studies on mycobacterial proteins (Datta *et al.*, 2000, 2003; Roy *et al.*, 2004, 2008; Saikrishnan *et al.*, 2003; Saikrishnan, Kalapala *et al.*, 2005; Saikrishnan, Manjunath *et al.*, 2005; Das *et al.*, 2006; Krishna *et al.*, 2006, 2007; Selvaraj *et al.*, 2007), we report here the crystal structure of *Mt*Ung in complex with Ugi. The structure of the enzyme exhibits unique features which differ from those of UNGs of other organisms. It provides a rationale for the decreased stability of the *Mt*Ung–Ugi complex in comparison to the *Ec*Ung–Ugi complex. It indicates a higher stability for the *M. tuberculosis* enzyme and a greater contribution of the electrostatic component to its interaction with DNA. A comparative study also involving the structures of Ung from other organisms provides further insights into the molecular mechanism of action of the enzyme and the relatively invariant and variable features of the molecule.

Table 1
Crystal data and data-collection, refinement and model statistics.

Values in parentheses are for the highest resolution shell.

Space group	*C*2
Unit-cell parameters	
a (Å)	201.14
b (Å)	64.27
c (Å)	203.68
β (°)	109.7
$V_{\rm M}$ (Å^3 Da^{-1})	2.6
Solvent content (%)	52.8
No. of molecules in ASU	7
Resolution range (Å)	30.0–3.1 (3.21–3.1)
No. of observed reflections	119763
No. of unique reflections	43788 (4095)
Completeness (%)	97.2 (91.8)
Multiplicity	2.7
Average $I/\sigma(I)$	8.2 (2.4)
$R_{\rm merge}$† (%)	14.9 (42.1)
Refinement and model statistics	
R factor (%)	23.4
$R_{\rm free}$‡ (%)	27.6
Total No. of protein atoms	16321
No. of solvent atoms	519
R.m.s. deviation from ideal	
Bond lengths (Å)	0.013
Bond angles (°)	1.4
Residues in Ramachandran plot§ (%)	
Core regions	82.8
Allowed regions	14.4
Generously allowed regions	2.3
Disallowed regions	0.5

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th observation of reflection *hkl* and $\langle I(hkl)\rangle$ is the weighted average intensity for all *i* observations of reflection *hkl*. ‡ 5% of reflections were used to calculate $R_{\rm free}$. § Calculated for nonglycine and nonproline residues using *PROCHECK*.

2. Materials and methods

2.1. Structure determination and refinement

*Mt*Ung was crystallized and diffraction data were collected as described previously by Singh *et al.* (2006). The Matthews coefficient (Matthews, 1968) varies from 3.04 to 2.28 Å^3 Da^{-1} for six to eight copies of the *Mt*Ung–Ugi complex in the asymmetric unit. The corresponding solvent content varies from 59.5 to 46.1%. *Mt*Ung shares the highest sequence identity (45%) with *Hs*UNG of the UNG–Ugi complexes of known three-dimensional structure. Molecular replacement with *AMoRe* (Navaza, 1994) using the *Hs*UNG–Ung complex as the search model did not yield a satisfactory solution. The removal of a few N- and C-terminal residues from the search model also did not help. Eventually, the structure was solved using *Phaser* (Storoni *et al.*, 2004) with the Ugi complexes of *Hs*UNG (PDB code 1ugh), *Ec*Ung (PDB code 1uug) and HSVUng (PDB code 1udi) as search models. Parts of the N-and C-terminal tails were removed from these models in the successful attempt. The best solution was obtained for seven molecules in the asymmetric unit. The model was initially refined using *CNS* v.1.1 (Brünger *et al.*, 1998). Initially, each of the seven copies was treated as a rigid body. Subsequently, the Ung and Ugi molecules were also treated as separate rigid bodies. The rigid-body refinement was followed by molecular-dynamics simulation accompanied by torsion-angle refinement. The definition of side chains was specifically addressed in the subsequent examination of maps and positional refinement. During most of the refinement process, NCS restraints (1256 kJ mol^{-1}) were used. Sevenfold-averaged maps were also used extensively in the initial stages. *Coot* (Emsley & Cowtan, 2004) was used for manual model building. Water O atoms were first identified on the basis of peaks at the 1σ level and the 3σ level in $2F_o - F_c$ and $F_o - F_c$ maps, respectively. The threshold values were subsequently reduced to 0.8σ and 2.5σ, respectively. The final cycles of refinement were carried out using the program *REFMAC* (Murshudov *et al.*, 1997) from *CCP*4 (Collaborative Computational Project, Number 4, 1994) without NCS restraints. In this step, TLS refinement (Winn *et al.*, 2003) was carried out, treating the 14 molecules in the asymmetric unit as separate groups. Crystal data and data-collection, refinement and model statistics are given in Table 1.

2.2. Modelling of the *Mt*Ung–dsDNA complex

Previous studies in this laboratory have shown that UNG is made up of two domains and a hinge region (Saikrishnan *et al.*, 2002). It was further demonstrated that the domains close by about 11° on binding DNA. To effect this domain closure in *Mt*Ung, domain I of *Mt*Ung was superposed onto that of *Hs*UNG in its DNA complex (Parikh *et al.*, 1998). Subsequently, domain II of *Mt*Ung was moved to superpose it onto that in the complex. This involved a rotation of 9.9° and a translation of 0.39 Å; the two Ung molecules were then in good superposition. The DNA molecules in the *Hs*UNG complex now interacted satisfactorily with the domain-closed *Mt*Ung molecules, thus providing a model for the *Mt*Ung–DNA complex. This model was soaked in a 4 Å shell of water using *INSIGHT*II after H atoms had been generated. Energy minimization and simulated annealing were carried out using *CNS* v.1.1. A dielectric constant of unity was used. A restraint of 21 kJ mol^{-1} was applied to all C$^{\alpha}$ atoms throughout refinement. In the first step, the model was subjected to conjugate-gradient energy minimization with the introduction of a small repulsive van der Waals term and with the electrostatic term switched off. In the next step, the electrostatic term was switched on and the structure was minimized for 100 cycles. This was followed by simulated annealing. The model was heated to 3000 K and the simulation was performed in 25 K steps, with each step containing 50 cycles of 5 fs each. Subsequently, one further step of conjugate minimization was carried out. The refinement was terminated when the gradient of the total energy was less than 0.21 kJ mol^{-1} Å^{-1}.

2.3. Structure analysis and comparisons

The refined model was evaluated using *PROCHECK* (Laskowski *et al.*, 1993). Structure-based sequence alignment was performed using *STRAP* (Gille & Frommel, 2001). Structures were superposed using *ALIGN* (Cohen, 1997). *HBPLUS* was used to identify hydrogen bonds (McDonald & Thornton, 1994). The buried surface area was taken to be the difference between the sum of the accessible surface areas of the components and that of the complex. The accessible surface area was calculated employing *NACCESS* (Hubbard & Thronton, 1993) using a probe size of 1.4 Å^2. The electrostatic surface potential was calculated using *ABPS* (Baker *et al.*, 2001). Figures were generated using *PyMOL* (DeLano, 2002), *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Bacon, 1997).

3. Results and discussion

3.1. Overall features

Surprisingly, the structure of the *Mt*Ung–Ugi complex contains seven chemically identical copies of the complex in the asymmetric unit. No rational (twofold, threefold *etc.*) noncrystallographic symmetry could be discerned between pairs of crystallographically independent complexes. This is unusual, but corresponds to the experimental intensity data as evidenced by the ready refinement of the structure to acceptable values of R and $R_{\rm free}$ with reasonable geometry.

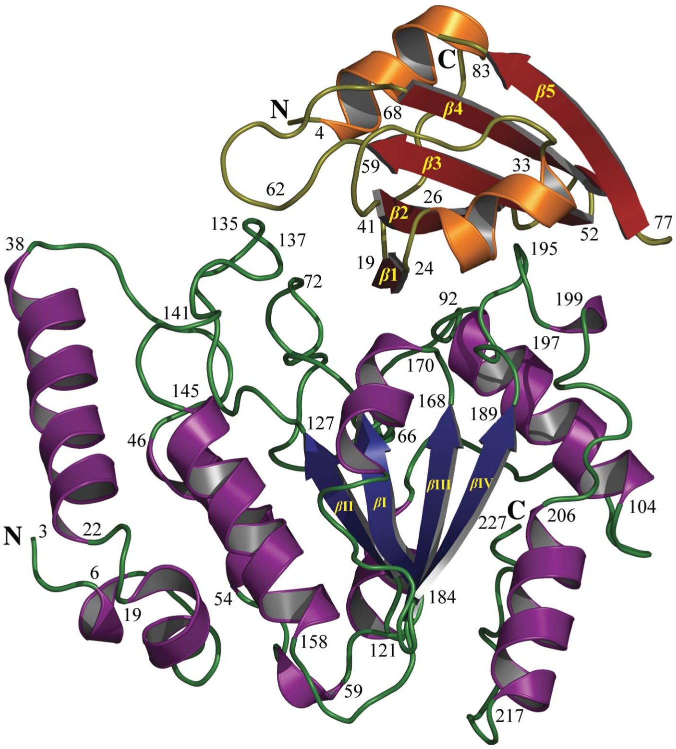


Figure 1
Structure of the complex between *Mt*Ung (α-helices in deep purple, β-strands in blue and loops in green) and Ugi (α-helices in orange, β-strands in red and loops in deep olive). The numbering of the β-strands is indicated.

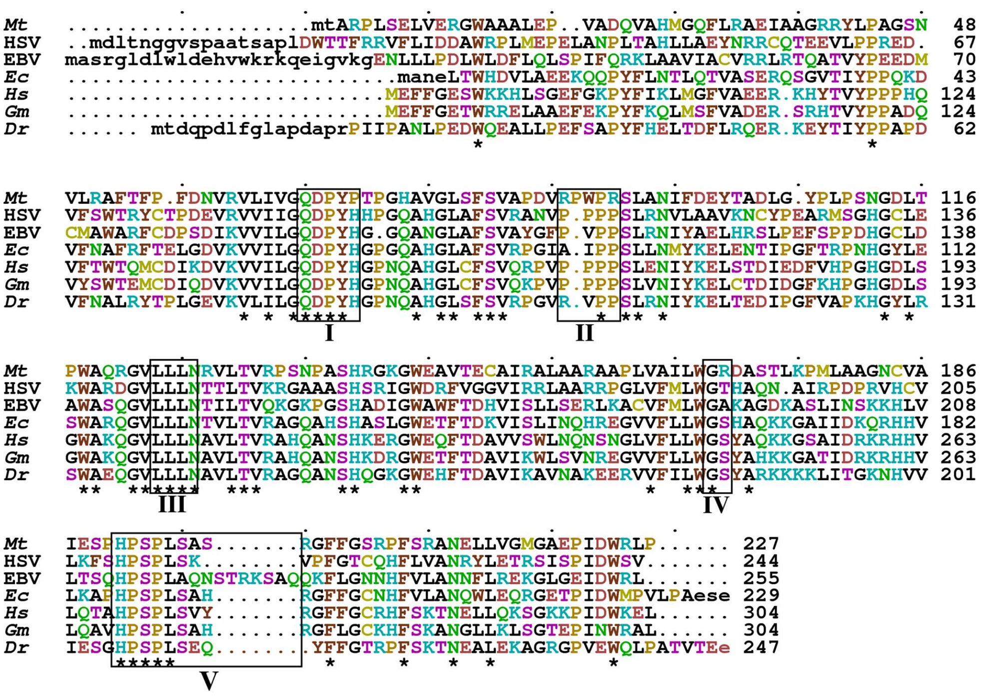
Although no NCS restraints were applied in the final cycle of refinement, the crystallographically independent Ung and Ugi molecules have very nearly the same structure. The r.m.s. deviations between pairs of Ung molecules range between 0.2 and 0.4 Å. The corresponding range in the case of Ugi molecules is 0.4–0.6. Thus, the Ugi molecule exhibits a slightly larger variability in structure. The fact that all the molecules exhibit the same structure in spite of the different crystallographic environments demonstrates the robustness of the structure. The existence of seven nearly identical copies of the same structure also compensates for the somewhat limited resolution of the crystal structure.

Like all other known structures of the family 1 UDGs, *Mt*Ung has a classical $\alpha/\beta/\alpha$ fold with a four-stranded parallel β-sheet in the middle (Fig. 1). This sheet is made up of strands 62–65 (β1), 123–127 (β2), 163–168 (β3) and 184–189 (β4). The structure contains a number of helices. Of these, four are long and contain more then ten residues. These long helices are made up of residues 20–35, 92–104, 145–158 and 206–216. 50% of the residues are in loops, most of which are involved in interconnecting regular secondary-structural elements. As in other similar structures, each Ugi molecule is made up of a five-stranded antiparallel sheet, two helices and connecting loops. Five short stretches in Ung have been identified as being important for catalytic activity (Parikh *et al.*, 1998). These are the water-activating loop (residues 66–72; labelled I in Fig. 2*a*), the proline-rich motif (89–93; II), the uracil-recognition loop (124–128; III), the Gly-Ser loop (169–170; IV) and the leucine loop (191–199; V).

3.2. Molecular structure of Ung: constant and variable regions and stability

The alignment of the amino-acid sequences of *Mt*Ung with those of other UNGs of known three-dimensional structure is illustrated in Fig. 2(*a*). There is considerable variation in the length of the N-terminal tail among these proteins. In particular, human UNG (*Hs*UNG) and *Gadus marhua* UNG (*Gm*UNG) have tails that are used for routing, which are cleaved to yield catalytic domains starting at residue 82. The tails are shortest in *Mt*Ung and *Ec*Ung. The variation in the length of the C-terminal tail is less pronounced. However, the lengths of the defined regions in different crystal structures are comparable and vary between 223 and 230 residues. *Mt*Ung has a sequence identity that ranges from 41 to 47% with the other six UNGs. The sequence identity is highest with *D. radiodurans* Ung. The C$^\alpha$ atoms of the two structures superpose with an r.m.s. deviation of 0.8 Å for 199 pairs. The r.m.s. deviation in C$^\alpha$ position with the other structures varies between 1.05 and 1.44 Å for 203–207 pairs.

A superposition of the C$^\alpha$ traces of all seven UNG structures is shown in Fig. 3. Detailed analysis of the deviations between pairs of structures clearly shows that the three-dimensional geometry of the contiguous stretch from residues 58 to 195 is reasonably well preserved in all the structures, although short variable regions exist within it. Indeed, this stretch starts four residues prior to the first β-strand (β1) and ends six residues after the last β-strand (β4), thus emphasizing the centrality of the β-sheets to the structure. In the 138 residues in this central stretch, there are only 58 in which the C$^\alpha$ positions deviate by 1 Å or more in more than one of the possible 25 pairs among the seven UNGs of known structure. Including residues which are not



(*a*)

M. gilvum	gi|145224794|	RPLPRS
M. vanbaalenii	gi|120403144|	RPLPRS
M. sp. MCS	gi|108798901|	RPLPRS
M. smegmatis	gi|118469778|	RPLPRS
M. tuberculosis	gi|15610113|	RPWPRS
M. leprae	gi|15827886|	RPLPRS
M. avium	gi|118465585|	SPLPRS
M. ulcerans	gi|118617565|	RPLPRS

(*b*)

Figure 2
(*a*) Structure-based sequence alignment of *Mt*Ung, *Hs*UNG (PBD code 1ugh), *Ec*Ung (PDB code 1uug), HSVUng (PDB code 1udi), EVBUng (PDB code 2j8x), *Gm*UNG (PDB code 1okb) and *Dr*Ung (PDB code 2boo). The five catalytic motifs are shown in boxes. Stars indicate conserved residues and dots indicate every tenth residue in the sequence alignment. (*b*) Alignment of sequences of the proline loop within *Mycobacterium* sp.

defined in one or the other structure in each pair, the corresponding number is 82 in the 88 remaining residues in the N-terminal and C-terminal tails. Thus, these tails as a whole are clearly the variable regions of the molecule (Fig. 3). They are also substantially devoid of catalytically important residues. There are a total of 51 residues which are identical in the seven sequences of UNGs of known structure. Of these, 44 belong to the 138 residues comprising the central region in *Mt*Ung. Only seven belong to the 89 residues in the N- and C-terminal stretches (Fig. 2*a*). The relatively invariant nature of the central region can also be discerned from the r.m.s.d. in C^{α} positions when the central region of *Mt*Ung is superposed on those of the other six structures. The r.m.s.d. now varies between 0.61 and 0.77 Å, whereas most of the values were greater then 1 Å when the whole molecules were superposed.

An indication of the possible role of the variable N- and C-terminal stretches can be gleaned by comparing *Hs*UNG and *Gm*UNG. Both have similar sequences, with an identity of 76%. The C^{α} positions in the three-dimensional structures superpose with an r.m.s.d. of 0.38 Å for 218 pairs. However, humans function at ambient temperature, while *Gadus morhua* survives at cold temperatures. Cold adaptation involves a decrease in thermal stability accompanied by increased catalytic efficiency. Biochemical, mutational and molecular-dynamics investigations indicate that the N- and C-terminal regions behave differently during unfolding and probably have contacts that are decisive for stability (Moe *et al.*, 2004; Olufsen *et al.*, 2005, 2007).

Several stabilizing interactions of the main body of the molecule with the N- and C-terminal tails exist in UNGs. Such interactions also exist within the individual tails. Their numbers are comparable in UNGs from different sources. However, several interesting differences are observed between them (Fig. 4). For example, the carbonyl O atom of the conserved Gly144 in the body of the molecule is involved in a hydrogen bond to Gln24 NE2 in the N-terminal tail in *Mt*Ung. The corresponding donor in the N-terminal region in *Ec*Ung is Thr23 OG1 (24 in *Mt*Ung numbering). The hydrogen bond does not exist in *Hs*UNG. Indeed, in the human enzyme there is a hydrophobic interaction between Leu105 (24 in *Mt*Ung numbering) and Trp222 (144 in *Mt*Ung numbering) (Fig. 4*a*). Among the interactions between the C-terminal stretch and the central region, there are interactions in which the basic pattern is retained even when the residues differ between species. For example, Arg121 NH1 and NH2 hydrogen bond to Trp224 O and Arg225 O, respectively, in *Mt*Ung. Arg121 is replaced by Gln117 in *Ec*Ung. The side chain of this glutamine interacts with Asp219 N and Asp219 O. Exactly the same interaction also exists in *Hs*UNG. Arg61 NH1 forms a hydrogen bond to Glu222 O in addition to the hydrogen bond that NH2 of the same residue makes to Arg121 O within the central region in *Mt*Ung. Lys57 in *Ec*Ung corresponds to Arg61 in *Mt*Ung. The side chain of this lysine makes two hydrogen bonds to O and OG of Thr216. In human UNG Lys138 NE makes a hydrogen bond to Lys297 O (Fig. 4*b*). These two lysine residues have been suggested to have a repulsive interaction (Olufsen *et al.*, 2007).

Among the five short stretches that have been identified as being involved in catalysis, the uracil-recognition loop has identical sequences in all UNGs of known structure. The sequence of the leucine loop is also substantially conserved, except for the insertion of a short stretch in EBVUng. In the water-activating loop, the crucial highly conserved histidine which is believed to be involved in correctly orienting the substrate (Dinner *et al.*, 2001) as well as interacting with water molecule when binding to DNA (Parikh *et al.*, 1998) is replaced by proline in *Mt*Ung. Again, in what is described as the Gly-Ser loop, the second position is occupied by an arginine residue in *Mt*Ung, whereas the residue is serine in most of the other UNGs. In the proline loop, there is a single amino-acid insertion in *Mt*Ung. Incidentally, this is also true of Ungs from other mycobacteria (Fig. 2*b*). This insertion leads to a substantial protrusion of the loop in *Mt*Ung in comparison to other UNGs of known structure (Fig. 3). The composition of the loop is also substantially different in *Mt*Ung: the loop contains two arginines and one tryptophan in addition to two prolines. In contrast, all four of these residues are proline in *Hs*UNG and two are proline in *Ec*Ung (Fig. 2*a*).

In terms of overall amino-acid composition, *Mt*Ung is somewhat different from *Hs*UNG and the prototype *Ec*Ung. It contains 20 arginine residues, which account for 8.8% of the residues in the protein. In addition, there are two lysines, making *Mt*Ung highly basic with a pI of 9.2. *Hs*UNG is equally basic, but contains only eight arginines, while the number of lysines is 18. The content of basic amino acids is lower in *Ec*Ung, with nine arginines and nine lysines in the sequence. Arginine and lysine are highly basic; the basicity of arginine is a little higher than that of lysine. Furthermore, arginyl residues are in general known to confer more stability to the protein than lysyl residues (Mrabet *et al.*, 1992), presumably on

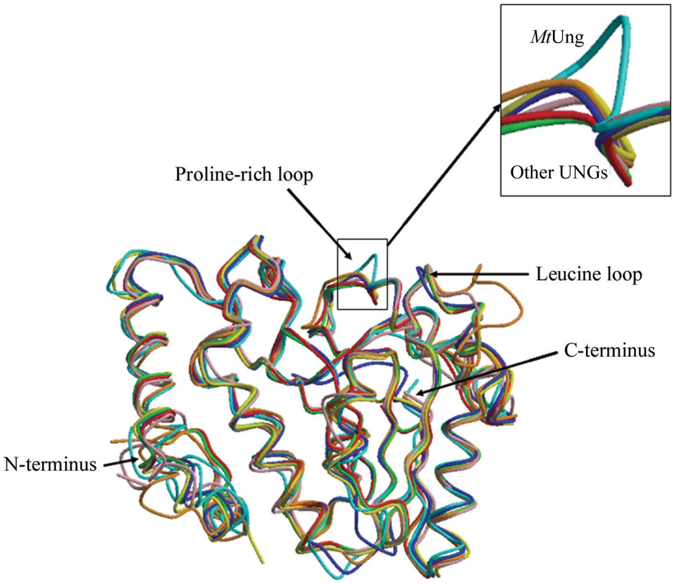


Figure 3
Superposition of UNGs of known structure. *Mt*Ung is in cyan. The proline-rich loop is highlighted in the inset.

account of the greater hydrogen-bonding potential of the guanidyl group in comparison to the amino group.

3.3. Ung–Ugi interactions

The UNG–Ugi complex is mainly stabilized by two types of interactions (Putnam *et al.*, 1999). The first is a hydrophobic interaction resulting from the protrusion of the side chain of Leu195 of UNG into the hydrophobic cavity in Ugi, involving Ile33, Val43, Met56, Leu58 and Val71. The second set of interactions involves hydrogen bonds between the DNA-binding groove of UNG and the β1 edge of Ugi. The surface areas buried on complexation are 2250, 2100 and 2200 Å^2 in the *Mt*Ung, *Ec*Ung and *Hs*UNG complexes, respectively. The nonpolar components of the buried area are 1450, 1370 and 1350 Å^2, respectively. Thus, the surface area buried on complexation is comparable in the three cases. Consequently, the difference in the stability of the complexes must be caused by differences in hydrogen bonding.

It has been shown that the *Mt*Ung–Ugi complex dissociates in 5–6 *M* urea, while the *Ec*Ung–Ugi complex is stable even at 8 *M* urea (Purnapatre & Varshney, 1998; Acharya *et al.*, 2003). The Ung–Ugi interactions in the *E. coli* and human enzymes are identical. Two of these interactions are lost on account of the substitution of His by Pro and Gln by His in positions 71 (67 in *Ec*Ung) and 75 in *Mt*Ung (71 in *Ec*Ung), respectively (Fig. 5*a*). On the other hand, the substitution of serine by arginine at position 170 results in an additional hydrogen bond in the *Mt*Ung complex (Fig. 5*b*). The insertion in the proline loop referred to earlier and its protrusion towards Ugi results in a bifurcated hydrogen bond between Gln19 NE2 of Ugi and Pro89 O and Trp90 O of *Mt*Ung (Fig. 5*a*). However, this protrusion abolishes a water bridge that is present in the *Ec*Ung complex between Gln19 NE2 of Ugi and the carbonyl O atom of Ala at position 88 in the *Ec*Ung numbering scheme. Another additional interaction in the *Mt*Ung–Ugi complex is an NH–π hydrogen bond involving Arg92 NH2 (Pro in other known Ung structures) and the aromatic ring of Tyr47 in Ugi (Fig. 5*c*). Among these differences in Ung–Ugi interactions, the loss of the hydrogen bond between Ser21 OG in Ugi and His67 NE2 in *Ec*Ung is perhaps the most significant. A complex of a mutant of Ugi in which Ser21 is replaced by Pro dissociates in 2 *M* urea (Acharya *et al.*, 2002). Thus, the substitution of His by Pro in *Mt*Ung and the consequent loss of the hydrogen bond is expected to substantially reduce the stability of the

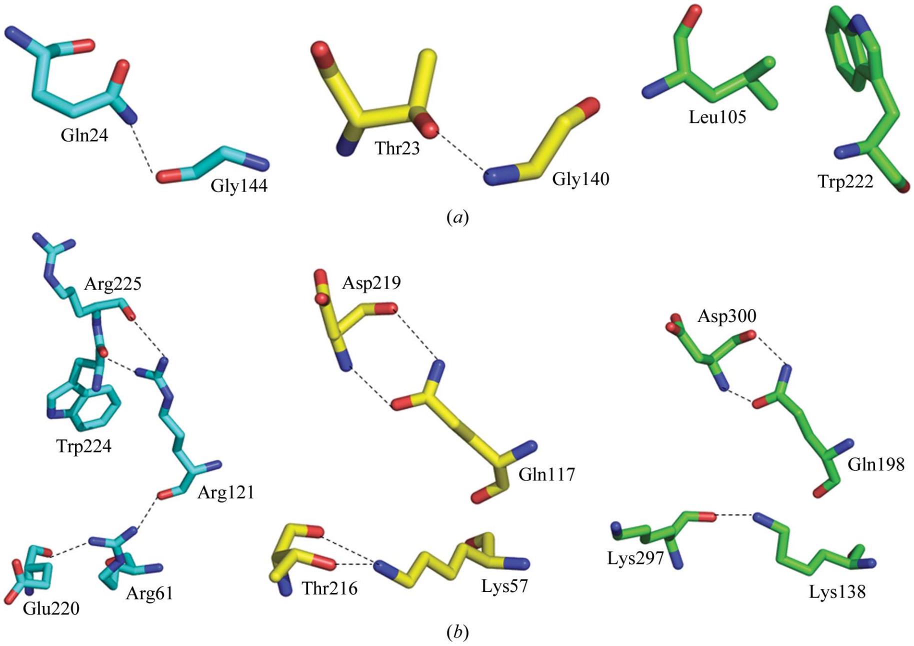


Figure 4
Comparable interactions in *Mt*Ung (cyan), *Ec*Ung (yellow) and *Hs*UNG (green) between the core region and (*a*) the N-terminal tail and (*b*) the C-terminal tail.

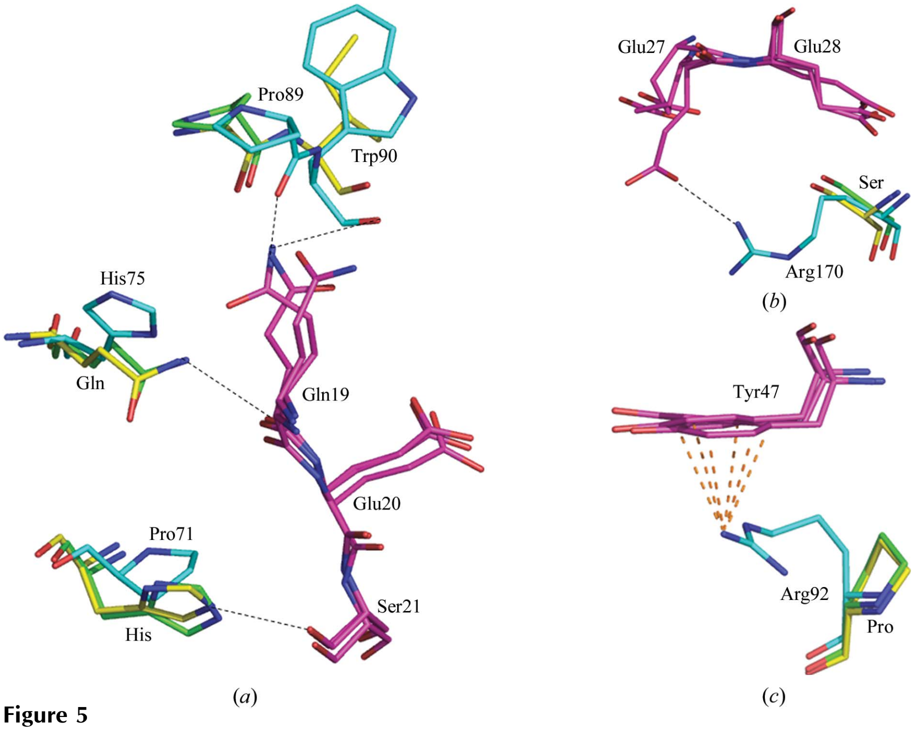
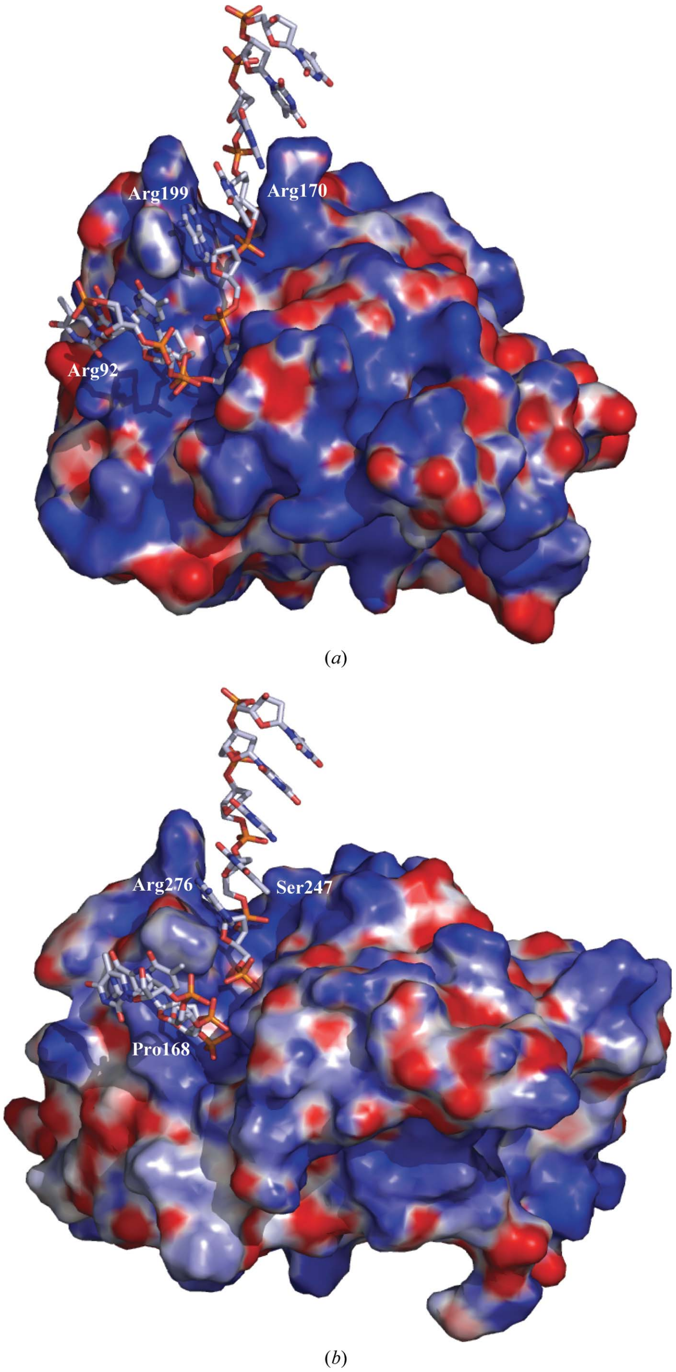


Figure 5
Interactions of Ugi (magenta) with *Mt*Ung (cyan), *Ec*Ung (yellow) and *Hs*UNG (green). *Mt*Ung residues are numbered. See text for details.

complex. The loss of a second interaction on account of the substitution of Gln by His at position 75 (*Mt*Ung numbering) must further reduce the stability of the complex. These losses are partially compensated by the additional interactions in the *Mt*Ung–Ugi complex outlined above. Even after this compensation, the *Mt*Ung–Ugi complex remains substantially less stable then the *Ec*Ung–Ugi complex.



(*a*)

(*b*)

Figure 6
Electrostatic surface potential of (*a*) *Mt*Ung and (*b*) *Hs*UNG along with one strand of bound DNA. Several critical residues mentioned in the text are indicated. Positive and negative charges are in blue and red, respectively.

3.4. Ung–DNA interactions

A model of the complex between *Mt*Ung and double-stranded DNA was constructed as described in §2 using the crystal structure of the *Hs*UNG–DNA complex (Parikh *et al.*, 1998). This involves the initial application of the domain closure deduced from comparison of the free and DNA-bound Ung molecules (Saikrishnan *et al.*, 2002). Incidentally, there is a distinct difference between the UNG–DNA and UNG–Ugi complexes. Complex formation with Ugi does not involve any domain closure, while that with DNA involves the closure of domain 1 comprising residues 12–81 and 124–156 by 9.9° with respect to domain 2 made up of 83–114 and 163–224 about a link region involving residues 115–123 and 159–162.

The Ung molecule in the energy-minimized model of the *Mt*Ung–DNA complex is substantially similar to the UNG molecule in the *Hs*UNG–DNA complex. The C^{α} positions of the molecules in the two structures superpose with an r.m.s. deviation of 0.85 Å. The corresponding deviation in the positions of all atoms is 1.2 Å when the two DNA molecules are superposed. The mutual disposition of protein and DNA in the structures is also very similar (Fig. 6). The area buried on complexation in *Mt*Ung is 1531 Å^2, of which 865 Å^2 is nonpolar. The corresponding values in the case of the *Hs*UNG complex are somewhat lower at 1213 and 712 Å^2, respectively. These values in the energy-minimized model of the *Ec*Ung–DNA complex, constructed in the same way as the *Mt*Ung–DNA complex was, are close to those in the *Hs*UNG–DNA complex at 1300 and 722 Å^2, respectively.

The model of the *Mt*Ung–DNA complex is similar to the *Hs*UNG–DNA structure and the model of the *Ec*Ung–DNA complex except for the additional positive charge in the DNA-binding region of *Mt*Ung on account of the increased presence of arginine in this region (Figs. 6 and 7). Based on the criterion of a change in the surface area buried of 1 Å^2 or more, 23

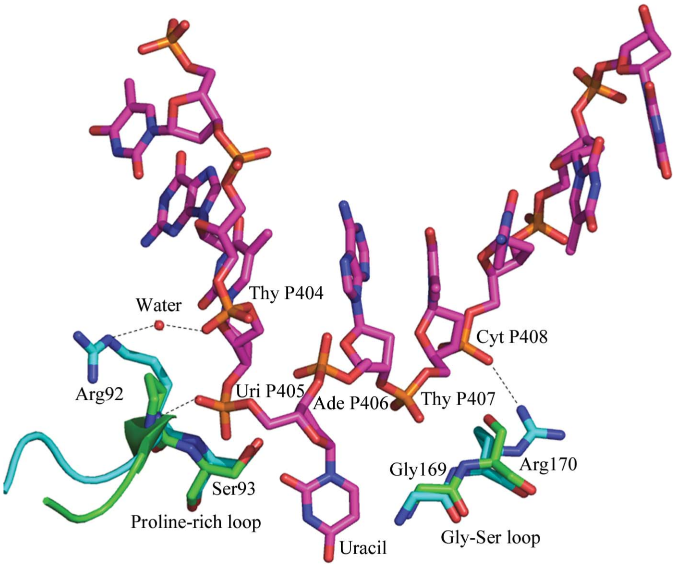


Figure 7
Unique interactions between DNA (magenta) and *Mt*Ung (cyan) in the modelled *Mt*Ung–DNA complex. Corresponding residues in *Hs*UNG (green) are shown for comparison.

residues are affected by DNA binding in *Mt*Ung. The corresponding number is 21 in *Hs*UNG. Of the 23 residues in *Mt*Ung, the change in surface area on complexation is 10 Å^2 or more for 18 residues. This number is 12 in the *Hs*UNG complex. This is in agreement with the greater burial of surface area on complexation in *Mt*Ung in comparison to that in *Hs*UNG. Almost all the residues in the leucine loop are involved in UNG–DNA interactions in both cases. The burial of surface area on complexation is the greatest for the leucine residue in this loop in both the structures. At the other extreme, only one of the four residues in the uracil-recognition loop is affected by DNA binding in *Mt*Ung; none is affected in *Hs*UNG. A few residues not belonging to the recognized catalytic loops also interact with DNA in both complexes. They include Ser135 (His212 in *Hs*UNG), Pro137 (Ala214 in *Hs*UNG) and Arg141 (Lys218 in *Hs*UNG), all of which belong to an extended loop following strand β2 and preceding helix 145–158.

Of the 23 and 21 residues affected by DNA binding in *Mt*Ung and *Hs*UNG, respectively, 18 occur at homologous positions in the two sequences. However, the residues are different at eight of these positions. The last residue in the water-activating loop is proline in *Mt*Ung (Pro71), while it is histidine in all other UNGs of known structure, including those from humans and *E. coli*. In the *Hs*UNG–DNA complex, a side-chain N atom of this residue and the side chain of the aspartyl residue and the carbonyl O atom of the propyl residue in the same loop are hydrogen bonded to water molecules (Parikh *et al.*, 1998). The hydrogen bonds between UNG and DNA are by and large similar in the two complexes. However, a couple of additional interactions in the *Mt*Ung complex, such as those illustrated in Fig. 7, resulting from the substitution of arginine for other residues presumably lead to additional stability in this complex. The substitution of Glu171 in *Hs*UNG by Ala95 in *Mt*Ung serves to remove a repulsive interaction with DNA (Moe *et al.*, 2004). The same is true in relation to *Ec*Ung, *Gm*UNG and *Dr*Ung, where the residue is Leu, Val and Arg, respectively.

Quantum-mechanical/molecular-mechanical studies have indicated that His67 in the water-activating loop of *Ec*Ung is important for positioning the reactants; it also makes an unfavourable energy contribution in achieving the transition-state intermediate (Dinner *et al.*, 2001). This residue is mutated to proline (Pro71) in *Mt*Ung. The next residue, a glycine in *Ec*Ung, is threonine in the mycobacterial enzymes. Studies involving a H67P mutant and a H67P/G68T double mutant of *Ec*Ung revealed that the substitution of histidine by proline results in an increased $K_{\rm m}$ (decreased substrate affinity) and $V_{\rm max}$, although the catalytic efficiency ($V_{\rm max}/K_{\rm m}$) remains unaltered (Acharya *et al.*, 2004).

The residues involved in DNA binding in *Hs*UNG are more similar to those in *Ec*Ung than to those in *Mt*Ung. In particular, the residue in the water-activating loop referred to above remains a histidine in *Hs*UNG. Therefore, the substitution of this residue by proline in *Mt*Ung could be expected to result in a lower substrate affinity in *Mt*Ung, as in the case of the H67P mutant of *Ec*Ung. However, as mentioned earlier, *Mt*Ung forms additional hydrogen bonds to DNA compared with *Hs*UNG. Also, the surface area buried on complexation is larger in the *Mt*Ung–DNA complex than in the *Hs*UNG–DNA complex. The additional hydrogen bonds and the increased burial of surface area in the *Mt*Ung–DNA complex could well compensate for the loss of affinity caused by the substitution of histidine by proline.

Kinetic parameters pertaining to *Mt*Ung are not available. However, those for the homologous enzyme (*Ms*Ung) from *M. smegmatis*, which is often used as a model for *M. tuberculosis*, are available. *Mt*Ung and *Ms*Ung have a sequence identity of 83%. Furthermore, the sequences of the catalytic loops are identical in the two enzymes. Therefore, the kinetic parameters for *Ms*Ung should provide a reasonably good indication of those for *Mt*Ung. It turns out that the $K_{\rm m}$ of *Ms*Ung is comparable to that of *Ec*Ung when the same substrate (SSU9) is used (Purnapatre & Varshney, 1998). Thus, as expected from the structural considerations outlined above, the reduction in substrate affinity caused by the substitution of histidine by proline in the water-activating loop is compensated by the additional interactions present in the mycobacterial enzyme.

4. Summary and conclusions

The presence of seven crystallographically independent copies of the Ung–Ugi complex in the crystal structure provides a reasonably detailed description of the geometry of the *Mt*Ung molecule, although the resolution of the structure is only 3.1 Å. The sequences in the catalytic loops in *Mt*Ung differ substantially from those in UNGs of known structure. A detailed comparison of *Mt*Ung with UNGs from other sources leads to the delineation of a relatively rigid central region made up of about 60% of the sequence starting from four residues before the first β-strand and ending at the sixth residue after the last strand of the β-sheet at the core of the molecule. More then 85% of the residues which are identical in the sequences of seven UNGs of known three-dimensional structure are in the core region, which encompasses almost all residues in the catalytic loops. The variable N- and C-terminal stretches appear to be important for stability and folding.

The results of studies on a complex of a mutated Ugi with *Ec*Ung appear to suggest that the substitution of a histidine in the water-activating loop by proline in *Mt*Ung could lead to reduced stability of its complex with Ugi. Additional interactions present in the complex substantially restore its stability. Detailed modelling of *Mt*Ung–DNA interactions and comparison with those in the known structure of a *Hs*UNG–DNA complex confirm that unlike in the case of Ugi binding, DNA binding involves substantial closure of the two domains in the molecule. The DNA-binding region of *Mt*Ung is rich in arginyl residues compared with the human and *E. coli* enzymes. The substitution of histidine by proline, referred to earlier, also appears to weaken complexation with DNA. However, the surface area of the enzyme buried on complexation with DNA is larger in *Mt*Ung compared with that in *Hs*UNG and *Ec*Ung. Furthermore, additional hydrogen-bonding interactions exist in the *Mt*Ung–DNA complex on account of the substitution of several key residues in *Hs*UNG by arginine. A comparison of the $K_{\rm m}$ values of *Ec*Ung and the Ung from *M. smegmatis*, which is closely related to *Mt*Ung, appears to indicate that the additional interactions in the *Mt*Ung–DNA complex compensate for the decrease in affinity caused by the substitution of histidine in the water-activating loop by proline.

The known sequences of mycobacterial Ungs exhibits a high degree of amino-acid conservation. In particular, the amino-acid residues involved in the distinctly different intramolecular, Ung–Ugi and Ung–DNA interactions in *Mt*Ung described previously remain unchanged in other mycobacterial Ungs. Thus, the results presented here represent the unique features of the structure and interactions of mycobacterial Ungs.

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 at the Supercomputer Education and Research Centre and the Bioinformatics Centre and the Graphics Facility, both of which are supported by the Department of Biotechnology (DBT). This work forms part of a DBT-sponsored research programme. PSK is a CSIR research fellow. MV is supported by a Distinguished Biotechnologist Award from the DBT.

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.

Acharya, N., Talawar, R. K., Purnapatre, K. & Varshney, U. (2004). *Biochem. Biophys. Res. Commun.* **320**, 893–899.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.

Bennett, S. E., Schimerlik, M. I. & Mosbaugh, D. W. (1993). *J. Biol. Chem.* **268**, 26879–26885.

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.

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., Krishna, R., Ganesh, N., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2003). *J. Bacteriol.* **185,** 4280–4284.

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 Visualization System.* http://www.pymol.org.

Dinner, A. R., Blackburn, G. M. & Karplus, M. (2001). *Nature (London)*, **413**, 752–755.

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.

Gille, C. & Frommel, C. (2001). *Bioinformatics*, **17**, 377–378.

Hubbard, S. J. & Thronton, J. M. (1993). *NACCESS Computer Program.* Department of Biochemistry and Molecular Biology, University College, London.

Jiang, Y. L., Krosky, D. J., Seiple, L. & Stivers, J. T. (2005). *J. Am. Chem. Soc.* **127**, 17412–17420.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Krishna, R., Manjunath, G. P., Kumar, P., Surolia, A., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2006). *Nucleic Acids Res.* **34**, 2186–2195.

Krishna, R., Prabu, J. R., Manjunath, G. P., Datta, S., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2007). *J. Mol. Biol.* **367**, 1130–1144.

Krokan, H. & Wittwer, C. U. (1981). *Nucleic Acids Res.* **11**, 2599–2613.

Kubota, Y., Nash, R. A., Klungland, A., Schar, P., Barnes, D. E. & Lindahl, T. (1996). *EMBO J.* **15**, 6662–6670.

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. & Nyberg, B. (1974). *Biochemistry*, **13**, 3405–3410.

McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.

Moe, E., Leiros, I., Riise, E. K., Olufsen, M., Lanes, O., Smalas, A. & Willassen, N. P. (2004). *J. Mol. Biol.* **343**, 1221–1230.

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.

Mrabet, N. T. *et al.* (1992). *Biochemistry*, **31**, 2239–2253.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Nicholl, I. D., Nealon, K. & Kenny, M. K. (1997). *Biochemistry*, **36**, 7557–7566.

Olufsen, M., Brandsdal, B. O. & Smalas, A. O. (2007). *J. Mol. Graph. Model.* **26**, 124–134.

Olufsen, M., Smalas, A. O., Moe, E. & Brandsdal, B. O. (2005). *J. Biol. Chem.* **280**, 18042–18048.

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.

Purnapatre, K. & Varshney, U. (1998). *Eur. J. Biochem.* **256**, 580–858.

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.

Roy, S., Gupta, S., Das, S., Sekar, K., Chatterji, D. & Vijayan, M. (2004). *J. Mol. Biol.* **339**, 1103–1113.

Roy, S., Saraswathi, R., Chatterji, D. & Vijayan, M. (2008). *J. Mol. Biol.* **375**, 948–959.

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.

Saikrishnan, K., Manjunath, G. P., Singh, P., Jeyakanthan, J., Dauter, Z., Sekar, K., Muniyappa, K. & Vijayan, M. (2005). *Acta Cryst.* D**61**, 1140–1148.

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.

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.

Srinath, T., Bharti, S. K. & Varshney, U. (2007). *DNA Repair*, **6**, 1517–1528.

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.

Warner, H. R., Johnson, L. K. & Snustad, D. P. (1980). *J. Virol.* **33**, 535–538.

Wang, Z. & Mosbaugh, D. W. (1988). *J. Bacteriol.* **170**, 1082–1091.

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.

Winn, M. D., Murshdov, 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.