BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Structural plasticity in *Mycobacterium tuberculosis* uracil-DNA glycosylase (*Mt*Ung) and its functional implications

S. M. Arif,[a] K. Geethanandan,[a] P. Mishra,[a] A. Surolia,[a] U. 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 13 March 2015
Accepted 15 May 2015

Edited by Z. Dauter, Argonne National Laboratory, USA

Keywords: DNA repair; excision repair; ligand binding; molecular plasticity; conformational selection.

PDB references: *Mt*Ung, form I, 4wpk; form IV, 4wrw; form V, 4wrx; complex with uracil, form I, 4wpl; form II, 4wru; form III, 4wrv; complex with 5-fluorouracil, form I, 4wry; 4wrz; form II, 4ws0; 4ws1; complex with 6-aminouracil, form I, 4ws2; form IV, 4ws3; complex with 5-nitrouracil, form I, 4ws4; form III, 4ws5; complex with 5-aminouracil, form I, 4ws6; complex with 5-chlorouracil, form II, 4ws7; complex with 2-thiouracil, form V, 4ws8

Supporting information: this article has supporting information at journals.iucr.org/d

17 independent crystal structures of family I uracil-DNA glycosylase from *Mycobacterium tuberculosis* (*Mt*Ung) and its complexes with uracil and its derivatives, distributed among five distinct crystal forms, have been determined. Thermodynamic parameters of binding in the complexes have been measured using isothermal titration calorimetry. The two-domain protein exhibits open and closed conformations, suggesting that the closure of the domain on DNA binding involves conformational selection. Segmental mobility in the enzyme molecule is confined to a 32-residue stretch which plays a major role in DNA binding. Uracil and its derivatives can bind to the protein in two possible orientations. Only one of them is possible when there is a bulky substituent at the 5′ position. The crystal structures of the complexes provide a reasonable rationale for the observed thermodynamic parameters. In addition to providing fresh insights into the structure, plasticity and interactions of the protein molecule, the results of the present investigation provide a platform for structure-based inhibitor design.

1. Introduction

Uracil-DNA glycosylase is the first enzyme in the base-excision repair pathway that removes uracil from both single-stranded and double-stranded DNA by cleaving the bond between uracil and deoxyribose (Lindahl, 1993). Uracil can arise in DNA either by spontaneous deamination of cytosine, resulting in the appearance of promutagenic G:U mismatches, or by an erroneous incorporation of dUMP opposite adenosine by DNA polymerase(s), giving rise to A:U base pairs during replication (Lindahl, 1982; Friedberg *et al.*, 1995). If not repaired, the promutagenic G:U mismatch may lead to G:C to A:T transition mutations, while the A:U base pair may interrupt protein–DNA transactions. Uracil from both these lesions can specifically be recognized and removed by uracil-DNA glycosylases (Parikh *et al.*, 2000). Of the five known families of uracil-DNA glycosylases, family I, also known as UNG/Ung, has been the most extensively studied. Ung from *Escherichia coli* (*Ec*Ung), encoded by the *ung* gene, was the first representative of this family to be identified (Lindahl *et al.*, 1977; Varshney *et al.*, 1988). Gene-knockout and mutation studies of *ung* from various sources have shown a manyfold increase in the G:C to A:T mutation frequency, leading to decay of the genetic information in the DNA (Duncan & Weiss, 1982; Impellizzeri *et al.*, 1991).

UNG/Ung proteins are inhibited by a well known proteinaceous inhibitor, uracil-DNA glycosylase inhibitor (Ugi), which is encoded by the phages PBS-I and PBS-II as part of

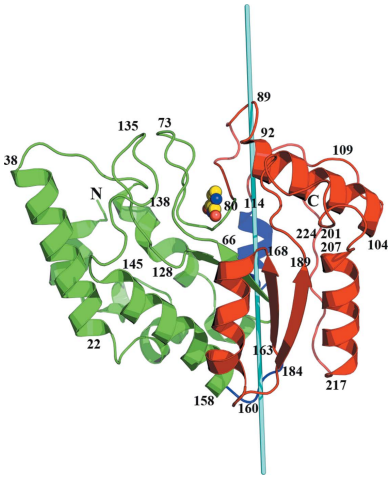

Table 1
Basic data on different crystals.

A and *B* refer to the two orientations of the ligand. Condition refers to the crystallization conditions outlined in the footnote. Please see text for details.

	Ligand	Method	Condition†
Form I			
1	Native		1
2	Uracil *A*	Co-crystallization	2
3	5-Fluorouracil *B*	Soaking	1
4	5-Fluorouracil *AB*	Soaking	1
5	5-Nitrouracil *B*	Soaking	1
6	5-Aminouracil *B*	Soaking	1
7	6-Aminouracil *A*	Soaking	1
Form II			
8	Uracil *A*	Co-crystallization	2
9	5-Fluorouracil *A*	Co-crystallization	3
10	5-Fluorouracil *AB*	Co-crystallization	3
11	5-Chlorouracil *B*	Co-crystallization	4
Form III			
12	5-Nitrouracil *B*	Co-crystallization	5
13	Uracil *B*	Co-crystallization	5
Form IV			
14	Native		1
15	6-Aminouracil *A*	Soaking	1
Form V			
16	Native		6
17	2-Thiouracil *AB*	Soaking	6

† Condition 1, 1.6 *M* sodium citrate tribasic dihydrate pH 6.5. Condition 2, 0.2 *M* ammonium acetate, 0.1 *M* sodium acetate trihydrate pH 4.6, 30%(*w*/*v*) PEG 4000, 4%(*v*/*v*) 1,1,1,3,3,3-hexafluoro-2-propanol. Condition 3, 0.2 *M* sodium acetate trihydrate, 0.1 *M* Tris–HCl pH 8.5, 30%(*w*/*v*) PEG 4000, 4%(*v*/*v*) 1,3-butanediol. Condition 4, 0.15 *M* potassium bromide, 30%(*w*/*v*) polyethylene glycol monomethyl ether 2000. Condition 5, 0.2 *M* sodium acetate trihydrate, 0.1 *M* sodium cacodylate trihydrate pH 6.5, 30%(*w*/*v*) PEG 8000, 5%(*v*/*v*) Jeffamine M-600 pH 7.0. Condition 6, 0.2 *M* ammonium acetate, 0.1 *M* sodium citrate tribasic dihydrate pH 5.6, 30%(*w*/*v*) PEG 4000.

their defence mechanism against host Ung (Cone *et al.*, 1980; Warner *et al.*, 1980; Wang & Mosbaugh, 1988). Ugi has been well characterized biochemically and structurally and it has been extensively used in structural studies of UNG/Ung as it mimics the DNA bound to the enzyme. Apart from Ugi, UNG/Ungs are also inhibited to various extents by uracil, one of the products of the enzymatic reaction, and some of its analogues and derivatives (Krokan & Wittwer, 1981; Blaisdell & Warner, 1983; Focher *et al.*, 1993; Jiang *et al.*, 2005; Krosky *et al.*, 2006; Chung *et al.*, 2009). Although inhibition by these small molecules has been well studied biochemically, the modes of their binding and interactions with the enzyme have not been extensively explored. Structures of free UNG/Ung from many sources and their complexes with Ugi have been reported (Mol, Arvai, Sanderson *et al.*, 1995; Mol, Arvai, Slupphaug *et al.*, 1995; Savva & Pearl, 1995; Savva *et al.*, 1995; Ravishankar *et al.*, 1998; Xiao *et al.*, 1999; Putnam *et al.*, 1999; Saikrishnan *et al.*, 2002; Leiros *et al.*, 2003, 2005; Moe *et al.*, 2004; Géoui *et al.*, 2007; Raeder *et al.*, 2010; Assefa *et al.*, 2012, 2014). The structure of the complex with oligonucleotides, however, is known only for the human (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998, 2000; Bianchet *et al.*, 2003), *Escherichia coli* (Werner *et al.*, 2000) and *Herpes simplex virus 1* (HSV1; Savva *et al.*, 1995) enzymes. Of these, the oligonucleotide bound to HSV1 does not contain uracil. The structures of the complexes of UNG/Ung with uracil, uracil analogues and some of 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; Schormann *et al.*, 2013). Earlier analyses of the relevant structures revealed concerted conformational changes in the enzyme, leading to closure of the active-site cleft consequent to the binding of DNA containing uracil (Slupphaug *et al.*, 1996; Werner *et al.*, 2000). Subsequently, it was demonstrated that Ung is a two-domain enzyme and that the domains close in on the bound DNA containing uracil (Saikrishnan *et al.*, 2002). The mere presence of free uracil in the active site of the enzyme does not lead to the closure of the active site (Savva *et al.*, 1995; Xiao *et al.*, 1999).

The genome of *Mycobacterium tuberculosis*, being rich in G+C content, is highly prone to damage owing to reactive oxygen species (ROS) and reactive nitrogen intermediates (RNI) present inside the macrophages during the latency of the bacteria. The probability of the deamination of cytosine to uracil is high owing to both the capacity of RNI for the deamination of cytosine and the greater availability of cytosine. *Mt*Ung plays an important role in restoring the genomic integrity of the cell by removing deaminated cytosines from DNA, avoiding the detrimental consequences. This makes *Mt*Ung an important target for the search for and design of inhibitors. The native structure of *Mt*Ung with a citrate molecule bound in the active site and the structure of its complex with Ugi have previously been reported from this laboratory (Kaushal *et al.*, 2008, 2010) as part of national and international structural biology efforts on mycobacterial proteins (Terwilliger *et al.*, 2003; Murillo *et al.*, 2007; Arora *et al.*, 2011). In the present study, we report the near-atomic resolution structures of and thermodynamic data on complexes of *Mt*Ung with uracil and its derivatives and new crystal forms of the free enzyme. A detailed structural examination of these high-resolution structures and the thermodynamic data have, amongst other things, led to striking insights into conformational selection in DNA binding and the modes of *Mt*Ung–ligand interactions.

2. Materials and methods

2.1. Crystallization

*Mt*Ung was expressed and purified as reported previously (Singh *et al.*, 2006). The purified protein was dialyzed against a buffer consisting of 500 m*M* NaCl, 10 m*M* Tris–HCl pH 7.5 before crystallization. Uracil, 5-fluorouracil, 5-chlorouracil, 5-bromouracil, 5-iodouracil, 5-nitrouracil, 5-aminouracil, 6-aminouracil and 2-thiouracil, all obtained from Sigma–Aldrich, were used in attempts to prepare crystalline complexes with the protein. These compounds, except for 5-fluorouracil, were sparsely soluble in water or the buffer. They were therefore dissolved in a mixture of dimethyl sulfoxide (DMSO) and 2-propanol to make up 100 m*M* stock solutions. Drops of these solutions were then added to the protein solution. A molar ratio of 1:10 to 1:30 between the protein and the ligand was used in co-crystallization experiments. Crystallization of the native protein and the complexes

Table 2
Details of data collection and refinement.

All crystals contained one molecule in the asymmetric unit. Values in parentheses are for the highest resolution shell.

	Form I							Form II	
Crystal No.	1	2	3	4	5	6	7	8	9
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$C2$	$C2$
Unit-cell parameters									
a (Å)	38.92	38.91	37.99	38.91	39.02	38.96	38.96	68.50	69.82
b (Å)	63.47	63.90	63.39	63.63	64.12	63.98	64.10	43.65	43.67
c (Å)	45.19	45.08	45.23	45.09	45.13	45.22	45.25	67.46	67.65
β (°)	112.9	112.2	114.1	112.8	112.5	112.4	112.6	98.9	99.1
Resolution range (Å)	23.74–0.98 (1.03–0.98)	19.83–1.15 (1.21–1.15)	41.30–1.43 (1.51–1.43)	19.41–1.19 (1.26–1.19)	32.07–1.18 (1.24–1.18)	25.42–1.10 (1.16–1.10)	17.50–1.13 (1.19–1.13)	36.68–1.24 (1.31–1.24)	36.89–1.97 (2.08–1.97)
V_M (Å^3 Da^{-1})	2.10	2.12	2.03	2.10	2.13	2.13	2.13	2.03	2.08
Solvent content (%)	41.4	41.9	39.4	41.5	42.3	42.2	42.3	39.5	40.9
Unique reflections	115505 (16844)	70859 (9428)	35414 (4622)	62231 (8728)	67385 (9694)	81890 (11287)	71969 (10062)	55335 (7547)	14239 (1901)
Multiplicity	4.9 (4.5)	3.0 (2.5)	4.1 (3.3)	3.8 (3.5)	3.7 (3.4)	3.5 (3.0)	3.9 (3.8)	5.6 (4.4)	3.4 (3.2)
Completeness (%)	100.0 (100.0)	98.0 (90.0)	97.8 (87.5)	96.7 (92.7)	99.8 (98.8)	98.4 (93.3)	94.2 (90.7)	99.9 (93.5)	98.7 (91.2)
$\langle I/\sigma(I)\rangle$	9.4 (2.4)	12.2 (3.6)	13.7 (3.6)	9.7 (3.9)	7.9 (2.4)	10.5 (3.6)	9.8 (3.3)	9.1 (2.6)	10.3 (4.9)
R_{merge}† (%)	11.5 (53.8)	4.9 (25.5)	6.1 (29.4)	8.1 (29.9)	10.1 (48.1)	6.7 (28.1)	7.6 (39.9)	10.7 (57.1)	8.9 (23.1)
Refinement and model statistics									
R factor (%)	13.7 (22.2)	10.4 (19.9)	10.8 (20.0)	11.9 (20.2)	12.9 (26.1)	12.8 (20.8)	12.8 (21.8)	16.6 (25.9)	16.5 (17.1)
R_{free} (%)	15.4 (23.7)	12.4 (21.4)	15.8 (21.9)	14.8 (22.9)	16.4 (28.7)	14.6 (24.2)	15.6 (23.3)	20.9 (29.3)	21.8 (24.1)
R.m.s. deviation from ideal									
Bond lengths (Å)	0.020	0.018	0.020	0.017	0.013	0.010	0.013	0.018	0.018
Bond angles (°)	1.9	1.9	2.0	2.0	1.8	1.7	1.8	1.9	1.8
Ramachandran plot statistics (% of residues)									
Favoured region	91.9	91.4	92.3	92.5	91.4	91.4	92.5	93.4	91.8
Allowed region	7.5	8.0	7.1	7.0	8.1	8.1	7.0	6.0	7.1
Generously allowed region	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1.1
Disallowed region	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PDB code	4wpk	4wpl	4wry	4wrz	4ws4	4ws6	4ws2	4wru	4ws0

	Form II		Form III		Form IV		Form V	
Crystal No.	10	11	12	13	14	15	16	17
Space group	$C2$	$C2$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters								
a (Å)	69.71	69.37	46.52	46.45	45.71	46.10	36.89	37.06
b (Å)	43.64	43.78	60.02	59.92	63.91	64.25	45.00	45.07
c (Å)	67.60	67.23	74.88	73.36	86.47	85.45	122.61	122.06
β (°)	99.1	98.5						
Resolution range (Å)	36.86–1.40 (1.48–1.40)	36.91–1.88 (1.98–1.88)	39.52–1.40 (1.48–1.40)	32.83–1.44 (1.52–1.44)	43.24–1.90 (2.00–1.90)	42.73–1.40 (1.48–1.40)	40.87–1.40 (1.48–1.40)	40.69–1.40 (1.48–1.40)
V_M (Å^3 Da^{-1})	2.07	2.06	2.13	2.08	2.58	2.58	2.08	2.08
Solvent content (%)	40.7	40.4	42.4	41.0	52.3	52.4	40.8	40.9
Unique reflections	39017 (5089)	16412 (2375)	41836 (5813)	37819 (5431)	20588 (2933)	50776 (7264)	40968 (5862)	41291 (5934)
Multiplicity	7.1 (5.1)	3.0 (3.0)	10.6 (7.0)	7.2 (7.2)	12.5 (12.2)	6.4 (4.4)	5.0 (4.9)	6.0 (6.0)
Completeness (%)	98.3 (88.9)	99.9 (99.6)	99.6 (97.2)	100.0 (100.0)	99.7 (99.2)	100.0 (100.0)	99.5 (99.0)	100.0 (100.0)
$\langle I/\sigma(I)\rangle$	11.0 (2.2)	9.6 (2.8)	20.3 (3.8)	10.8 (3.1)	7.4 (2.7)	10.1 (1.8)	10.3 (2.4)	9.3 (1.9)
R_{merge}† (%)	11.5 (68.6)	9.3 (39.0)	6.7 (47.9)	11.0 (103.4)	26.3 (93.1)	11.2 (75.3)	8.1 (62.2)	10.3 (92.3)
Refinement and model statistics								
R factor (%)	13.3 (25.4)	14.2 (31.7)	12.0 (19.2)	12.1 (17.6)	16.3 (24.4)	13.4 (28.3)	12.4 (22.2)	13.2 (26.1)
R_{free} (%)	17.6 (28.5)	18.2 (32.9)	16.0 (22.1)	14.0 (20.5)	20.2 (28.1)	16.2 (29.0)	14.8 (25.8)	16.8 (31.2)
R.m.s. deviation from ideal								
Bond lengths (Å)	0.019	0.018	0.019	0.009	0.019	0.012	0.019	0.012
Bond angles (°)	1.9	2.0	2.0	1.5	2.0	1.6	1.9	1.6
Ramachandran plot statistics (% of residues)								
Favoured region	92.9	91.8	92.9	90.1	93.4	91.8	91.5	93.1
Allowed region	6.6	7.7	6.6	8.8	5.5	7.1	7.9	6.3
Generously allowed region	0.5	0.5	0.5	1.1	0.5	1.1	0.5	0.5
Disallowed region	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0
PDB code	4ws1	4ws7	4ws5	4wrv	4wrw	4ws3	4wrx	4ws8

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the weighted average intensity for all i observations of reflection hkl.

was simultaneously attempted by the microbatch-under-oil method. The same protein solution was used in all crystallization experiments. Drops containing equal volumes of the protein solution and the precipitant solution were used for

crystallization. Crystals were obtained under different conditions involving the precipitants indicated in Table 1. One of these conditions yielded two crystal forms of the native protein and another condition yielded a further crystal form of the native protein. Soaking experiments were carried out using the native crystals in an attempt to generate more complexes. The concentration of the ligand was kept at 40 m*M* in all of the soaking experiments. The extensive crystallization and soaking efforts led to crystals in five different forms. The unique set of 17 native and complex crystals obtained from the experiments is listed in Table 1. The orientation of the ligand with respect to the molecule (see later) was also taken into account in determining uniqueness.

2.2. X-ray data collection and processing

Intensity data were collected from a total of 17 crystals. Data from three of them (crystals 9, 11 and 14) were collected using a home source with a MAR345 detector mounted on a Bruker-AXS Microstar Ultra II Cu $K\alpha$ rotating-anode X-ray generator. Nine data sets (crystals 1, 2, 3, 8, 10, 12, 13, 16 and 17) were collected using a CCD detector on the synchrotron X-ray beamline BM14 at the European Synchrotron Radiation Facility, Grenoble, France and the remaining five data sets (crystals 4, 5, 6, 7 and 15) were collected using the same type of detector on beamline PX-BL21 at the Indus-2 synchrotron facility (1.5 T bending magnet, 2.5 GeV) at RRCAT, Indore, India. All of the data sets were collected at a temperature of 100 K. Glycerol or ethylene glycol was used as the cryoprotectant. The diffraction images were processed and merged using *iMosflm* (Battye *et al.*, 2011) and *SCALA* (Evans, 2006) in the *CCP*4 program suite (Winn *et al.*, 2011). Intensities were converted to structure-factor amplitudes using *TRUNCATE* (French & Wilson, 1978) from the *CCP*4 program suite. The crystal specifications, data-collection parameters and processing statistics are summarized in Table 2. All data sets except for that from crystal 14 were of high quality. The comparatively poor quality of crystal 14 led to a high $R_{\rm merge}$. However, analysis of the structure using the data proceeded smoothly.

2.3. Structure solution and refinement

The structures of the 17 crystals were determined by the molecular-replacement method using *Phaser* (McCoy *et al.*, 2007) from the *CCP*4 program suite, with the coordinates of *Mt*Ung from its citrate complex (PDB entry 3a7n; Kaushal *et al.*, 2010) as the initial search model. In consonance with the indication from the Matthews coefficients of the crystals (ranging between 2.03 and 2.58 Å^3 Da^{-1}; Matthews, 1968), *Phaser* gave the best solution for one molecule in the asymmetric unit in all 17 cases. The solutions obtained from *Phaser* were then built manually using *Coot* (Emsley *et al.*, 2010) before refinement. An initial rigid-body refinement followed by positional refinement and individual *B*-factor refinement was performed on each structure using *REFMAC*5 (Murshudov *et al.*, 2011) from the *CCP*4 program suite. Alternative conformations with different occupancies were assigned to the side chains of a few amino-acid residues during refinement cycles. The stretches comprising residues 180–184 in crystal 3, 171–173 in crystal 8 and 182–184 in crystal 16 were also found to possess two possible conformations each. $2F_{\rm o} - F_{\rm c}$ and $F_{\rm o} - F_{\rm c}$ maps computed at stages where R and $R_{\rm free}$ had acceptable values showed unambiguous density for the respective ligands in the active site of the enzyme in 14 crystals. The presence of the ligands was further confirmed by computing simulated-annealing $F_{\rm o} - F_{\rm c}$ OMIT maps using *CNS* v.1.3 (Brünger *et al.*, 1998) to avoid model bias. Electron densities and the results of refinement, particularly temperature factors, indicated two possible orientations of the ligand in crystal 4 (form I), crystal 10 (form II) and crystal 17 (form V). Water O atoms were added successively to the structures in the final cycles of refinement using peaks with heights of greater than 3.0σ in $F_{\rm o} - F_{\rm c}$ maps and 1.0σ in $2F_{\rm o} - F_{\rm c}$ maps. Different components of the mother liquor were also modelled into the electron density when appropriate. In the final stage of refinement, anisotropic *B* factors were used in cases where the data set had a resolution better than 1.5 Å (all except those from crystals 9, 11 and 14). The final refinement parameters are given in Table 2.

2.4. Analysis of the structures

The refined models were evaluated using *PROCHECK* (Laskowski *et al.*, 1993). Structural superpositions were performed using *ALIGN* (Cohen, 1997). Sausage plots were generated using *MOLMOL* (Koradi *et al.*, 1996). Figures were generated using *PyMOL* (DeLano, 2002) and *CHIMERA* (Pettersen *et al.*, 2004). The domain movement in one molecule with respect to another was calculated by superposing domain 1 of the two molecules and then computing the angle of rotation required to superpose domain 2 of the two molecules.

2.5. Isothermal titration calorimetry

All titrations were conducted using a MicroCal VP-ITC calorimeter equilibrated at various temperatures ranging from 278 to 298 K. The protein solution was exhaustively dialyzed against a buffer consisting of 500 m*M* NaCl, 10 m*M* Tris–HCl pH 7.5 before each titration. The molarity of the salt was kept at a high level to prevent precipitation of the protein at high concentrations. The ligand solutions were prepared in the same dialysate as used for the dialysis of protein to avoid any artifacts from the heat of mixing. A concentration range of 0.3–0.5 m*M* for protein solutions and 5–20 m*M* for various ligand solutions were used in the titrations. The corresponding *C* values for the titrations ranged between 0.21 to 10.78 ($n \times K_{\rm b} \times M_{\rm t}$, where n is the number of binding sites of the macromolecule, $K_{\rm b}$ is the binding constant and $M_{\rm t}$ is the total protein concentration used in titration). The lower end of the *C*-value range was on account of the weak binding of a few of the ligands and the limitation on protein concentration (Turnbull & Daranas, 2003). 1.36 ml of the protein solution was added to the sample cell, ensuring that no air bubbles were trapped, and was then equilibrated to the appropriate pre-set temperature. Solutions of uracil, 6-aminouracil, 2-thio-

uracil, 5-nitrouracil, 5-flurouracil, 5-chlorouracil and 5-aminouracil were added as a series of injections (aliquots of 7–10 µl) separated by means of a software-controlled syringe with constant stirring at 307 rev min^{-1}. The volume of the first injection in each titration was set to 3 µl to avoid any experimental artifacts associated with the loading of the ligand with the syringe. The time spacing between successive injections was kept at 180 s to ensure a steady baseline of titration. Titrations under identical conditions using ligand solution and dialysate (protein buffer) only were also performed for dilution correction. Data were analyzed using *Origin* 7.0 Pro (OriginLab) and the thermodynamic parameters, namely the change in enthalpy (ΔH), the binding constant (K_b) and the number of binding sites (n), were determined employing nonlinear least-square curve fitting of the data. Dilution corrections were applied during the determination of the thermodynamic parameters. The free energy (ΔG) was calculated using the equation $\Delta G = -RT\ln K_b$, where R and T are the universal gas constant and the absolute temperature (in K), respectively. The change in entropy (ΔS) was then calculated using the equation $\Delta G = \Delta H - T\Delta S$.

3. Results and discussion

3.1. Overview of crystal forms

Among the ligands used in the crystallization and soaking experiments, bromouracil failed to bind to the protein. Iodouracil also failed to bind, but a co-crystallization experiment involving it yielded a complex with uracil. A mass-spectrometric analysis of the ligand sample showed the presence of a small amount of uracil. Presumably, this uracil interacted with the protein, leading to complex formation. Native crystals occurred in three different forms (Table 1). The *Mt*Ung–uracil complex occurred in three different forms. Four distinct 5-fluorouracil complexes were distributed between two different crystal forms. Complexes with 5-nitrouracil and 6-aminouracil occurred in two different crystal forms each. The 5-aminouracil, 5-chlorouracil and 2-thiouracil complexes occurred only in one crystal form each.

As in the case of all of the known structures of Ung from different species, including *M. tuberculosis*, the molecule has an $\alpha/\beta/\alpha$ fold with a four-stranded parallel β-sheet (residues 62–65, 123–127, 163–168 and 184–189) at the centre flanked by several helices (Kaushal *et al.*, 2008). The helices include four long helices involving residues 20–35, 92–104, 145–158 and 206–216. Nearly 50% of the residues are in loops, most of which are involved in interconnecting strands or helices. Four of the five stretches indentified as being important for catalytic activity, namely the water-activating loop (residues 66–72), the proline-rich motif (89–93), the Gly-Ser loop (169–170) and the leucine loop (191–199), are in the loop regions. The stretch described as the uracil-recognition loop (124–128) encompasses a strand of the central sheet (Kaushal *et al.*, 2008).

Seven of the 17 crystals considered in the present study belonged to form I. The r.m.s. deviations on pairwise superposition of C$^\alpha$ positions of their molecules range between 0.09 and 0.50 Å (Supplementary Table S1*a*). The corresponding range in the four form II crystals is 0.08–0.37 Å. The r.m.s. deviation in C$^\alpha$ positions when the two form III crystals are superposed is somewhat high at 1.52 Å. The corresponding values when the two form IV and form V crystals are superposed are low at 0.61 and 0.20 Å, respectively. Thus, except perhaps in the case of form III crystals, molecules in crystals belonging to the same form appear to have a very similar geometry. Close examination shows that crystals belonging to forms I and II cluster together, while the remaining three forms tend to form another cluster. The highest r.m.s. deviation on pairwise superposition within the first cluster is 0.73 Å. The corresponding value in the second cluster is 1.52 Å. When a structure from one cluster is superposed on another from the other cluster, the r.m.s. deviation in C$^\alpha$ positions increases to as high as 1.77 Å. This clustering, and indeed the structural mobility discussed below, appears to be independent of the composition, including the pH, of the medium of crystallization. For instance, structures in both clusters are found among crystals grown under condition 1. Likewise, crystals in the first cluster include those grown under conditions 1, 2, 3 and 4.

3.2. Domain movement and conformational selection

The clustering of structures mentioned above can be readily understood if the domain movement within the molecule is

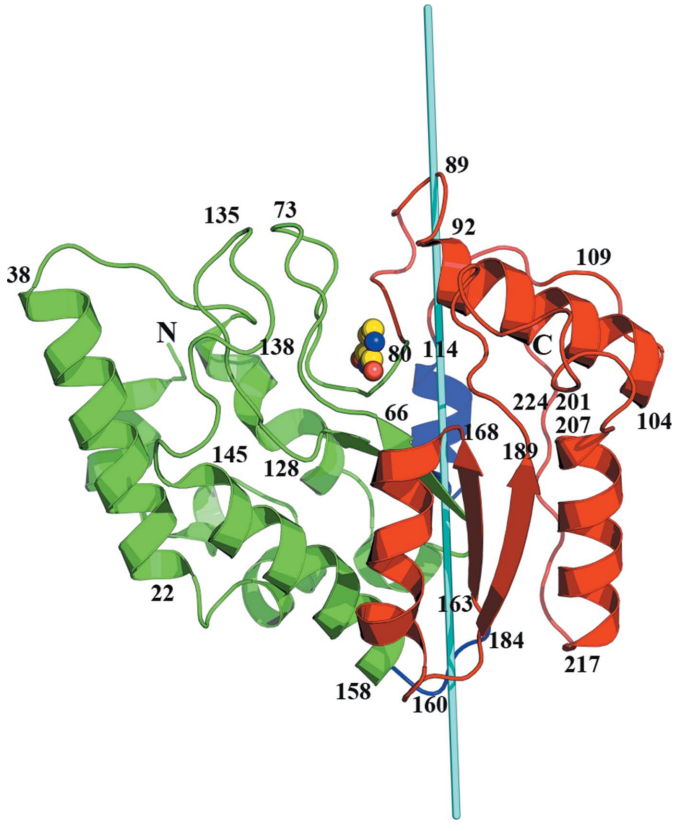

Figure 1
Structure of *Mt*Ung. Domain 1 is in green and domain 2 in red. The hinge regions are in blue. The thick line represents the axis about which the domains move. Bound uracil is shown as a van der Waals representation.

taken into account. Ung was originally considered to be a single-domain molecule (Savva *et al.*, 1995). However, it was subsequently shown that it contains two domains, and their mutual orientation is affected by interaction with uracil-containing or abasic DNA (Saikrishnan *et al.*, 2002). In the *Mt*Ung numbering, domain 1 encompasses the N-terminal stretch up to residue 81 and residues 124–156 and domain 2 is made up of residues 83–114 and 163–224, with residues 115–123 and 157–162 constituting the link region (Fig. 1; Kaushal *et al.*, 2008). The variability in the mutual orientation of the two domains involves a rotation about the axis shown in Fig. 1. Domain closure in UNG/Ung on interaction with uracil-containing or abasic DNA has been demonstrated in the human and *E. coli* enzymes (Saikrishnan *et al.*, 2002). In the enzyme from these two sources, the molecule has a 'closed' structure when it interacts with DNA and an 'open' structure when it is free. The difference between the closed and the open structures involves a rotation of 8.0–11.7° about an axis similar to that shown in Fig. 1.

The closure of the domains in response to interactions with DNA could have resulted from induced fit or conformational selection (Boehr *et al.*, 2009). The interdomain orientations in the 17 *Mt*Ung structures presented here facilitate a choice between the two possibilities. None of them contain DNA. Yet, the molecule in crystals 1–11 belonging to form I and form II has a closed conformation. That in the rest of the crystals, except crystal 14, has an open conformation with a rotation of 7.6–10.1° of domain 2 with respect to the molecule in the closed conformations in crystal 1. The molecule in crystal 14 has a 'half-closed' or 'half-open' conformation with a rotation of 5.5°. On the reasonable assumption that the 17 structures reported here sample the different accessible conformations of the molecule, it would appear that a closed conformation is chosen from the ensemble when the molecule interacts with DNA, in a case of conformational selection. Interestingly, the mutual orientation of the two domains appears to be unaffected by the absence or the nature of the small ligand. One of the native structures has a closed conformation while the other two have an open conformation. The uracil, 5-nitrouracil and 6-aminouracil complexes occur in closed as well as open conformations.

3.3. Segmental mobility within domain 2

It turns out that domain 1 has nearly the same structure in all of the crystals, with the maximum r.m.s. deviation in C^{α} positions on pairwise superposition being 0.45 Å (Supplementary Table S1*b*). Domain 2, however, exhibits substantial structural variability among the crystals. A close examination revealed that most of the differences in this domain are confined to a variable region involving the 168–199 stretch (Fig. 2). R.m.s. deviations in C^{α} positions on pairwise superpositions of the rest of the domain are always less than 0.4 Å (Supplementary Table S1*c*). The deviations in the variable region range between 0.10 and 4.26 Å (Supplementary Table S1*d*). The values are less than 1.57 Å among the form I and form II crystals, all of which exhibit a closed conformation. Large differences exist among the structures of this stretch among molecules with open conformation. The two crystals belonging to form III (crystal 12 and 13) present an interesting

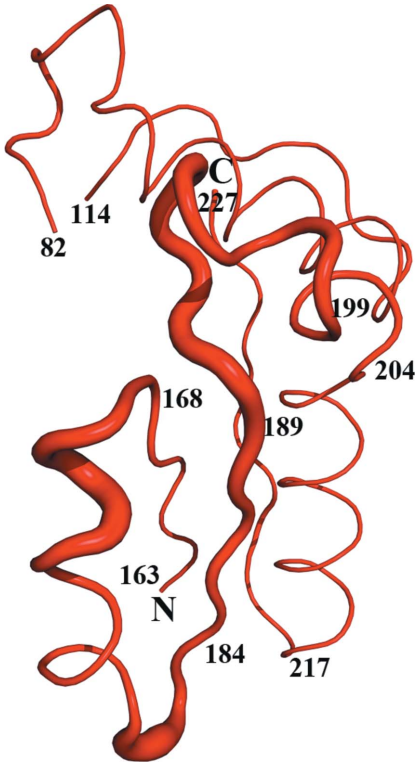

Figure 2
The sausage plot of domain 2, highlighting the variable stretch.

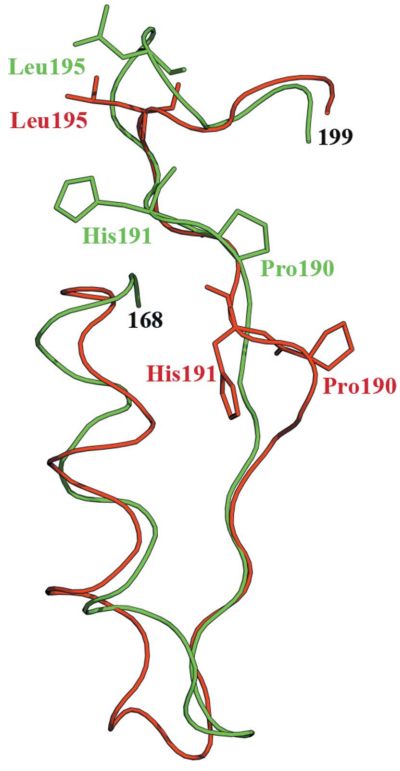

Figure 3
Superposition of the variable stretch in crystal 1 (form I; green line) and crystal 13 (form III; red line).

situation. The molecule in both of them has an open conformation. However, the structure of the variable stretch in crystal 12 is closer to those in form I and form II crystals, while that in crystal 13 is closer to that in the form IV crystals. The flexible region in form V crystals, which have molecules in the open conformation, exhibits structural similarity to that in molecules in the closed conformation. In fact, the stretches in crystals belonging to forms I, II and V superpose reasonably well. Thus, the structural variability in the 168–199 stretch in domain 2 does not appear to be correlated with the variations in the mutual orientation of the domains.

Broadly speaking, the crystals can be divided into two sets in terms of the structure of the 168–199 stretch. One set involves all crystals except for 13, 14 and 15. The other is made up of crystals 13, 14 and 15. The r.m.s. deviations in C^{α} positions in each set are less than 1.85 Å. Those resulting from comparisons of a crystal from one set and another from the other set vary between 3.47 and 4.25 Å. The difference between the two sets is most pronounced in the 189–199 segment. The Ser189–Pro190 peptide bond is *trans* in all of the stretches except in those of crystals 13, 14 and 15, in which this bond has a *cis* conformation. This change in isomerism results in a distance of 7.2 Å between the C^{α} positions of Pro190 in the two sets, as measured between those in crystals 1 and 13 (Fig. 3). The distances between the C^{α} atoms of His191 and Leu195 in the two sets are 7.0 and 4.7 Å, respectively. The differences decrease and converge by the 200th residue.

The differences in the conformation of this stretch within the two large sets of *Mt*Ung structures, delineated on the basis of the conformation about the Ser189–Pro190 peptide bond, can be explained in terms of the Ramachandran angles at Gly169 (Fig. 4*a*) and Gly182 (Fig. 4*b*). Both sets of angles cluster together separately for crystals belonging to form I and form II. The difference between the two is primarily in the conformation at Gly182. The φ, ψ angles at this residue are in the helix region in form I and in the region corresponding to the extended conformation in form II. This difference is reflected by the conformation of the stretch in the two sets of crystals (Fig. 5*a*). The Ramachandran angles at the two residues and hence the conformation of the stretch in crystal 12 are similar to those in the crystals belonging to form I. Crystals 16 and 17 belonging to form V have also φ, ψ angles at Gly169 and Gly182 similar to those in form I crystals. As can be seen

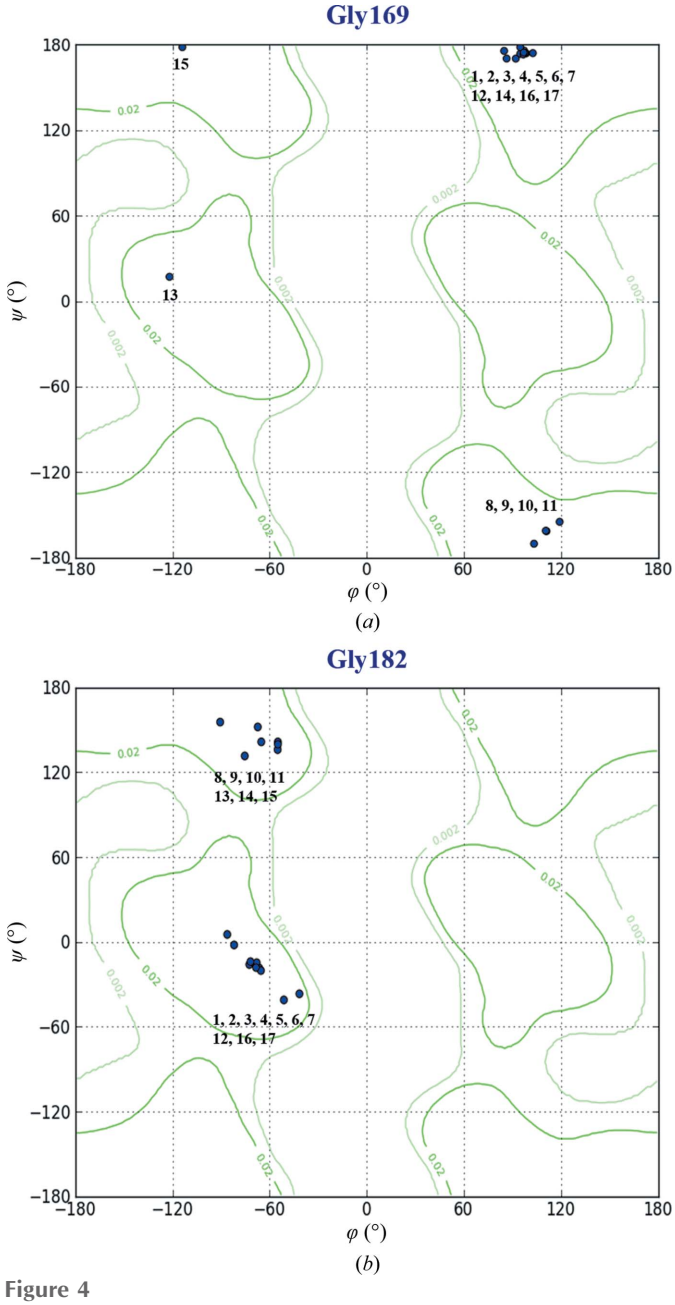

Figure 4
φ, ψ angles in different crystals for (*a*) Gly169 and (*b*) Gly182.

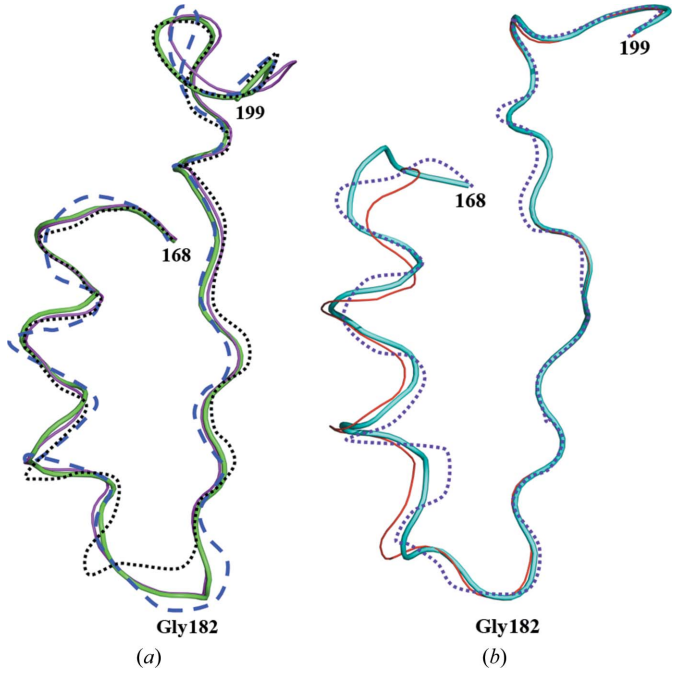

Figure 5
Superposition of the variable stretch in (*a*) crystal 1 (form I; thick green line), crystal 8 (form II; blue dashed line), crystal 12 (form III; thin magenta line) and crystal 16 (form V; black dotted line) and in (*b*) crystal 13 (form III; thin red line), crystal 14 (form IV; purple dotted line) and crystal 15 (form IV; thick cyan line).

in Fig. 5(*a*), the difference between the stretches in form I and form V is confined to the region around Gly182.

Among the smaller set, the φ, ψ angles at Gly182 in crystals 13, 14 and 15 cluster in the region corresponding to the extended chain. However, the angles at Gly169 exhibit substantial differences. They are at the two ends of the region corresponding to the extended chain in crystals 13 and 15. The corresponding angles in crystal 14 are in a region that is only allowed for a glycyl residue. Consequently, the conformations of the stretch in crystals 13 and 15 are closer to each other than they are to that in crystal 14 (Fig. 5*b*).

To sum up, the segmental variability involving the 168–199 stretch can be described to a reasonable extent in terms of *cis–trans* isomerism at Pro190 and conformational variability at Gly169 and Gly182. Among these, the effect of the *cis–trans* isomerism is the most pronounced and forms the basis for the classification into two broad sets. Differences in the φ, ψ angles at Gly169 and Gly182 lead to conformational variability in the stretch within each set. Among the other UNG/Ung molecules of known three-dimensional structure, the residue corresponding to the 190th position in *Mt*Ung is a proline only in the enzymes from *E. coli*, *Vibrio cholerae* and *Coxiella burnetii*. In the structures of the enzyme from these three sources, the peptide bond preceding the proline residue is *trans*. However, the possibility of this bond assuming a *cis* conformation in an appropriate environment cannot be ruled out. This possibility does not exist in cases where the residue at the position corresponding to 190 in *Mt*Ung is not a proline. Among the two glycines referred to above, Gly169 is invariant in all UNGs/Ungs of known three-dimensional structure. Therefore, flexibility arising from the presence of the glycyl residues could occur in these enzymes. However, the residue at position 182 is glycine only in *Mt*Ung. This adds to the variability of the mycobacterial enzyme.

Among the stretches identified as being important for catalytic activity, the water-activating loop (66–72) and the uracil-recognition loop (124–128) belong to domain 1. The remaining three stretches, namely the proline-rich motif (89–93), the Gly-Ser loop (169–170) and the leucine loop (191–199), belong to domain 2. The last two belong to the most variable stretch of the molecule. Interestingly, the leucine loop is the most variable structural element in the molecule. This loop plays a major role in Ung–Ugi and Ung–DNA interactions (Kaushal *et al.*, 2008). The loop also moves on complex formation with Ugi and DNA. In particular, Leu195 exhibits a high degree of burial on complexation. Furthermore, His191, the first residue in the loop, is involved in an interaction with the ligand in almost all of the known complexes of the enzyme with uracil.

3.4. Protein–ligand interactions

14 of the 17 crystals reported here contain uracil or one of its derivatives as a ligand. The bound derivative in each case is 5-fluorouracil, 5-chlorouracil, 5-aminouracil, 5-nitrouracil, 2-thiouracil or 6-aminouracil. The electron density for the ligands are well defined in all cases (Fig. 6). Co-crystallization and soaking involving 5-bromouracil and 5-iodouracil were attempted but were unsuccessful.

In addition to complexes involving Ugi and DNA, complexes of Ung from other sources with uracil, 6-aminouracil and uridine have been reported. The present

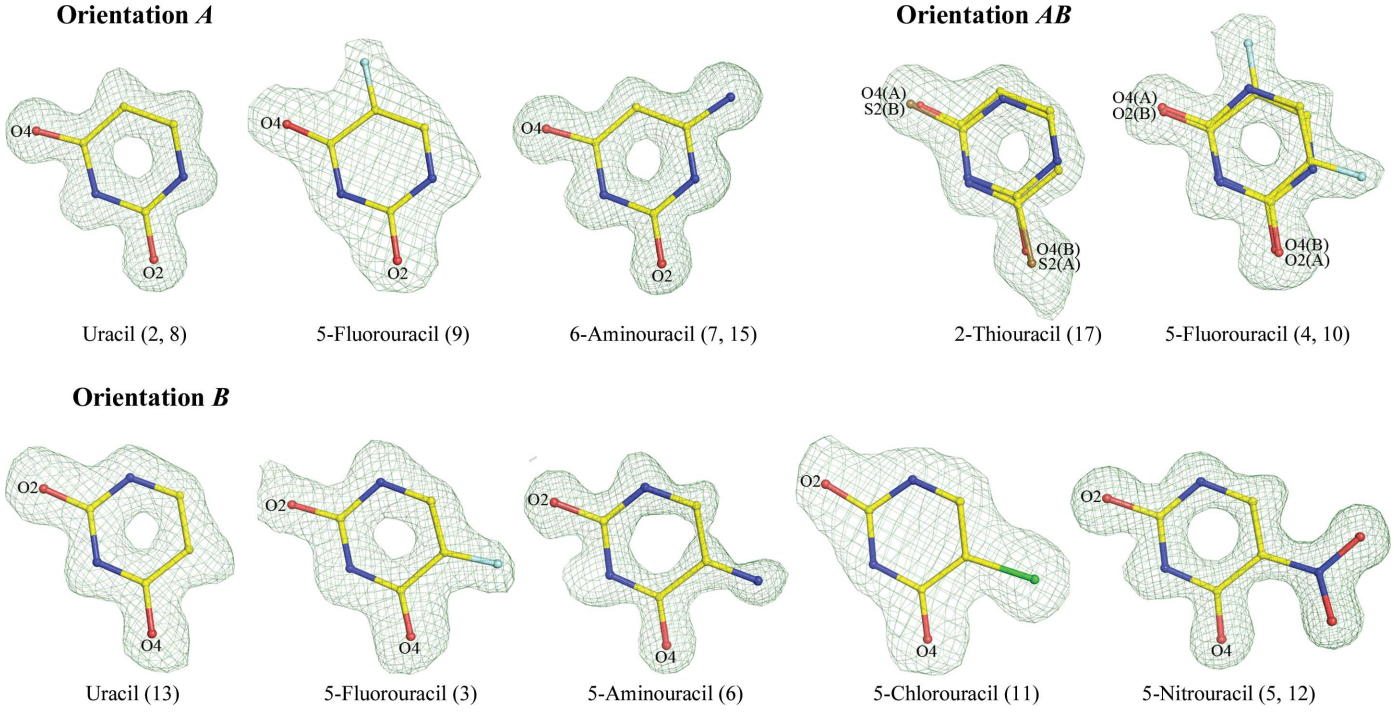

Figure 6
Typical electron densities for ligands in appropriate simulated-annealing $F_o - F_c$ OMIT maps. Contours are at the 3σ level. The numbers in parentheses are crystal numbers.

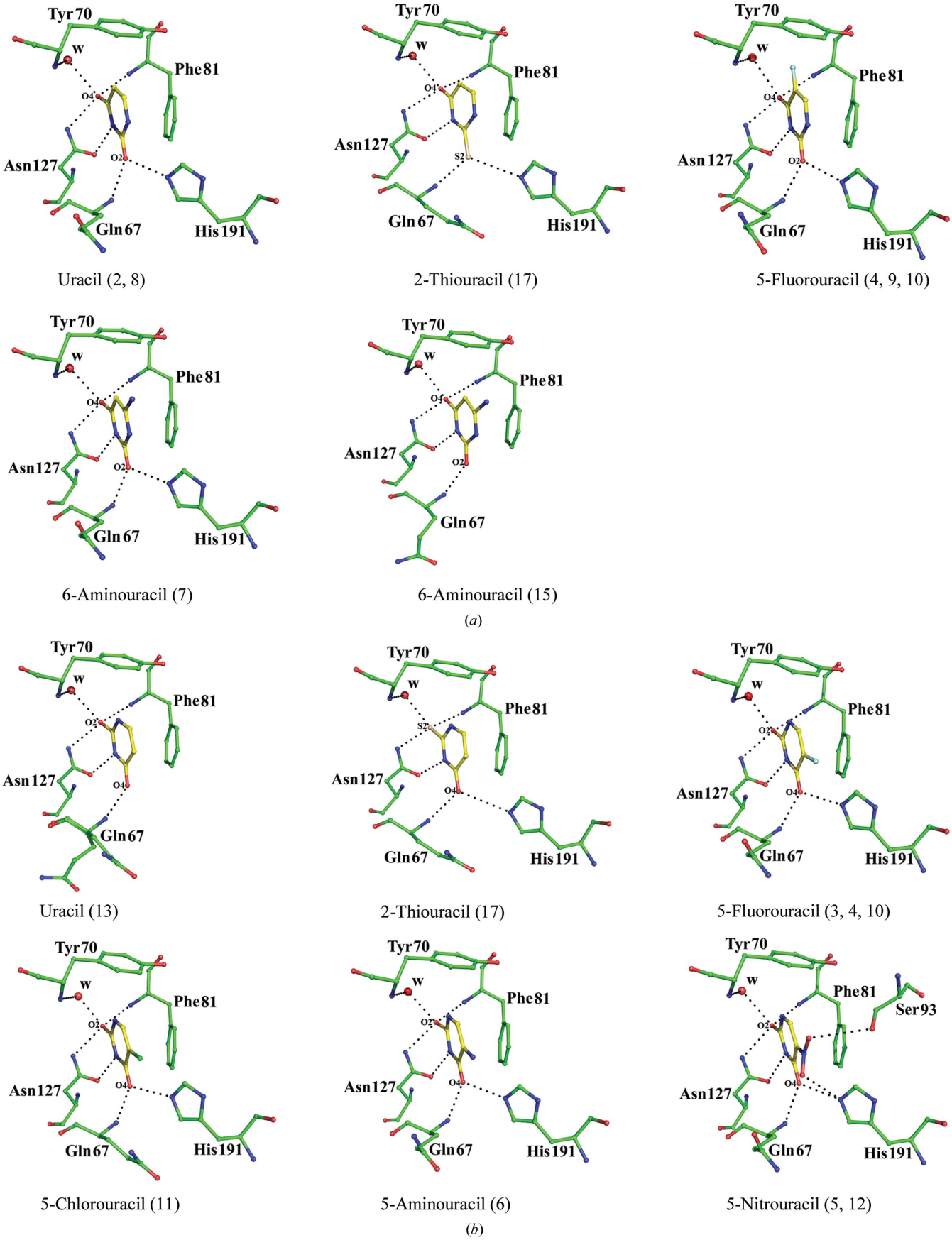

Figure 7
Protein–ligand interactions in different crystals with ligand orientation *A* (*a*) and *B* (*b*). The numbers in parentheses are crystal numbers.

investigation involves a comprehensive study of the interactions of *Mt*Ung with uracil and a variety of representative uracil derivatives. Representative sets of interactions which occur in the 14 structures are illustrated in Fig. 7. Unexpectedly, it was found that the ligands can bind to Ung in two different, but related, orientations, hereafter described as *A* and *B*. Orientations *A* and *B* are related to each other by a 180° rotation about a line joining N3 and C6 in the uracil moiety. This involves an interchange of O2 and O4 and their interactions. In orientation *A*, O2 makes hydrogen-bonding interactions with Gln67 N and His191 NE2 (in most cases), while O4 forms hydrogen bonds to Phe81 N, Asn127 ND2 and a water molecule. The water molecule bound to O4 is hydrogen-bonded to Tyr70 N, thus forming a water bridge. An N3···Asn127 OD1 hydrogen bond occurs in both orientations, as does the stacking interaction between the uracil ring and the side chain of Phe81.

Except in the cases of unsubstituted uracil and 6-aminouracil, the two orientations can be easily distinguished on the basis of the location of the substituents. When uracil or 6-aminouracil is the ligand, the structure was refined in both orientations. The structure which gave nearly the same *B* values for the ring atoms at the 1 and 5 positions was accepted as that corresponding to the correct orientation. The relevant structures have been refined at comparatively higher resolutions ranging from 1.15 to 1.4 Å and the discrimination based on *B* values appeared to be reliable, especially as restrained and unrestrained refinements led to the same conclusion in each case. The conclusion was further confirmed by the examination of OMIT maps contoured at various levels (Supplementary Fig. S1)

Uracil is the ligand in crystal 2 (form I), crystal 8 (form II) and crystal 13 (form III). In the first two the ligand is in orientation *A*, while it is in orientation *B* in crystal 13. 6-Aminouracil, which is bound to the protein in crystal 7 (form I) and crystal 15 (form IV), is in orientation *A*. 5-Chlorouracil (crystal 11, form II), 5-aminouracil (crystal 6, form I) and 5-nitrouracil (crystal 5, form I and crystal 12, form III) bind to the enzyme in orientation *B*. 5-Fluorouracil presents an interesting case. It has orientation *A* in crystal 9 (form III), while it is in orientation *B* in crystal 3 (form I). Both orientations coexist with 50% occupancy each in crystal 4 (form I) and crystal 10 (form II). Similarly, 2-thiouracil exists in both orientations in crystal 17 (form V). These results clearly indicate that the orientation of the ligand is not influenced by the nature of the crystal form and the conditions, including the pH, under which the crystals were grown. Nor is it influenced by the interdomain orientation. Both ligand orientations are found in the open and the closed forms of the molecule.

The above results appear to suggest that both the orientations are nearly equally preferred unless the interactions involving the substituents favour or disfavour one orientation. The majority of the ligands used in the present investigation involve substitution at the 5′ position of the uracil. In unsubstituted uracil, the hydrogen at this position could form a C—H···π interaction with the side chain of Tyr70 in orientation *A* (Handa *et al.*, 2002). This is replaced by an N—H···π interaction in orientation *B*. These interactions are replaced by C—F···π and N—H···π interactions in the complexes of 5-fluorouracil. When the substituent is larger as in 5-chlorouracil, 5-aminouracil and 5-nitrouracil, orientation *A* is disfavoured on account of the steric clash of the substituent with

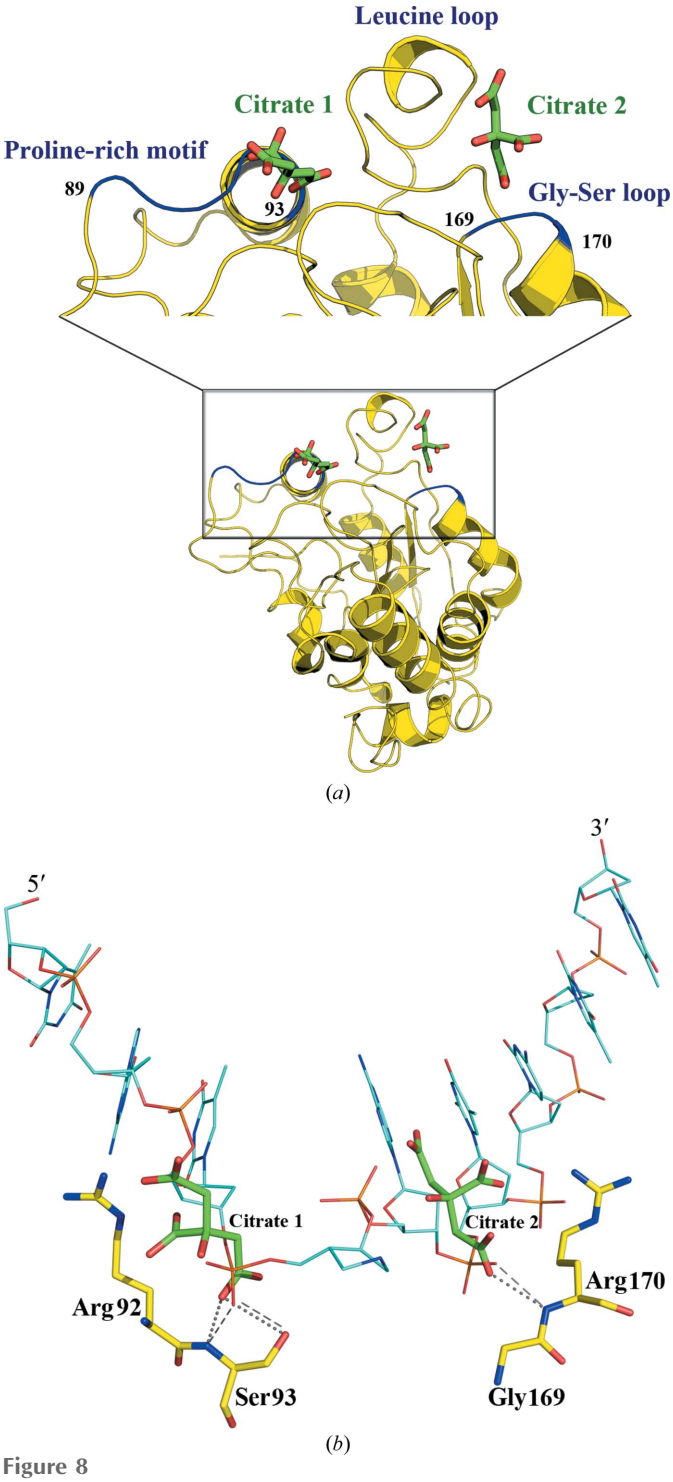

Figure 8
(*a*) Location of citrate ions in crystal 5. (*b*) One strand of a modified DNA molecule in complex with human UNG (PDB entry 1q3f) superposed on the citrate ions and interacting residues in crystal 5. Hydrogen bonds involving citrate and phosphate O atoms are indicated by dotted and dashed lines, respectively.

Table 3
Thermodynamic parameters for the interaction of uracil and uracil derivatives with *Mt*Ung.

Standard deviations for the thermodynamic parameters were obtained from three independent ITC experiments.

Ligand	T (K)	n	$K_b \times 10^{-3}$ (M^{-1})	$-\Delta G_b$ (kJ mol^{-1})	$-\Delta H_b$ (kJ mol^{-1})	$-\Delta S$ (J mol^{-1} K^{-1})
6-Aminouracil	278	1.07 ± 0.03	33.6 ± 0.7	24.1	46.4 ± 1.7	81
Uracil	278	1.00 ± 0.01	18.9 ± 0.3	22.7	58.6 ± 2.5	129
2-Thiouracil	278	0.88 ± 0.07	3.1 ± 0.3	18.5	42.7 ± 3.3	88
5-Nitrouracil	278	0.81 ± 0.01	1.64 ± 0.01	17.2	52.7 ± 0.8	128
5-Fluorouracil	278	0.93 ± 0.08	0.45 ± 0.02	14.1	21.0 ± 2.1	25

the aromatic ring of Tyr70. In the case of 2-thiouracil, interactions involving a CO group and a CS group are exchanged when the orientations are interchanged, without any apparent difference in the energy of interaction. Both orientations of 2-thiouracil coexist in crystal 17. The position of the amino group in 6-aminouracil remains the same in both of the possible orientations of the bound ligand. Therefore, both orientations could be equally preferred. Only one of them has been observed so far in the complexes (crystals 7 and 15).

Most of the interactions with the uracil moiety remain the same in all of the complexes. The only exception is the hydrogen bond involving His191. This is the only residue from domain 2 which hydrogen-bonds to the ligand. The movement of domain 2 with respect to domain 1 slightly affects the position of the residue, but it does not affect the hydrogen bond as long as there is no drastic change in the variable stretch referred to earlier. However, in crystals 13 and 15, where there are drastic changes in the variable stretch, the histidyl side chain has very different positions and orientations. The hydrogen bond involving the histidyl residue does not exist in them. However, the location and the orientation of the ligand is unaffected by the absence of this hydrogen bond.

3.5. Protein–citrate interactions

In the course of the refinement, some components of the mother liquor were fitted into the density when appropriate. No biological relevance can be readily attributed to the interactions of these components with the protein, except in the case of citrate. The precipitant used for crystallization contained 1.6 *M* sodium citrate in eight of the 17 experiments which resulted in crystals. Six of these eight crystals contain ordered citrate ions with unambiguous density at one or both of two locations. These locations, as observed in the complex involving 5-nitrouracil (crystal 5), are illustrated in Fig. 8(*a*). Interestingly, the citrate ions interact with the proline-rich motif (89–93) and the Gly-Ser loop (169–170), which are regions of the molecules involved in DNA binding. The ion that interacts with the Gly-Ser loop is also in steric contact with the leucine loop (191–199). This loop is known to play an important role in UNG–DNA interactions.

One strand of a cationic 1-aza-2′-deoxyribose-containing DNA molecule in its complex with human UNG (*Hs*UNG) (PDB entry 1q3f; Bianchet *et al.*, 2003) was superposed on the citrate ions and the residues that interact with them in crystal 5, as illustrated in Fig. 8(*b*). A terminal carboxylate O atom of citrate 1 hydrogen-bonds to the peptide N atom and the side-chain hydroxyl group of Ser93. In the DNA complex, a phosphate O atom located at a distance of only 0.55 Å from the citrate O atom interacts in a similar manner with Ser169 in the human enzyme. A terminal O atom in citrate 2 forms a hydrogen bond to the main-chain N atom of Arg170. The residue in the corresponding position in human enzyme is a serine (Ser247). A phosphate O atom 0.51 Å from the citrate O atom forms a hydrogen bond to the main-chain N atom of Ser247 in *Hs*UNG.

3.6. Thermodynamics of ligand binding

ITC measurements were performed for the binding of all of the ligands found in the crystalline complexes at 278, 298 K and at least one intermediate temperature in most cases. For 5-aminouracil and 5-chlorouracil, binding at all temperatures was too weak to be quantified. As for the rest, sensible thermodynamic parameters could be obtained for 5-fluorouracil only at 278 K, presumably as it is the weakest binder among them. Thus, only the parameters determined at 278 K were available for the binding of all of the ligands and they were used for purposes of comparison (Table 3, Supplementary Fig. S2). Interestingly, 5-aminouracil and 5-chlorouracil, the binding of which to *Mt*Ung was too weak to be quantified, can bind the protein only in orientation *B*. Thus, it would appear that *A* is the preferred orientation. 5-Nitrouracil also binds to the protein in orientation *B*. However, thermodynamic parameters of its binding to *Mt*Ung could be measured. This is probably because of the additional interactions of the nitro group with Ser93 OG and His191 NE (Fig. 7*b*).

The free energy of binding between the various ligands and the enzyme varies in a comparatively small range between −14.1 and −24.1 kJ mol^{-1}. The only derivative which has nearly the same free energy of interaction with the enzyme as that of uracil is 6-aminouracil. The protein–ligand interactions in the two cases are nearly identical. 6-Aminouracil has an additional interaction in the form of a close approach of the amino group to Ser93 OG. The interaction energy is lower in the case of 2-thiouracil than in the case of uracil, as one set of hydrogen-bonded interactions involving a CO group is now replaced by a set involving the CS group. The energy is also lower when the ligand is 5-fluorouracil, as a C—H$\cdots\pi$ or N—H$\cdots\pi$ interaction in the case of uracil is replaced by a C—F$\cdots\pi$ interaction. Thus, a reasonably satisfactory structural rationale exists for the observed thermodynamic parameters.

3.7. Comparison with the human enzyme

A comparison between *Mt*Ung and *Hs*UNG is of particular interest in relation to inhibitor design aimed at drug development. The two enzymes exhibit a sequence identity of 45%.

Three crystal structures of ligand-free *Hs*UNG in the open conformation and six of its complexes in the closed conformation are available (Slupphaug *et al.*, 1996; Parikh *et al.*, 1998, 2000; Bianchet *et al.*, 2003; Moe *et al.*, 2004; Assefa *et al.*, 2012). Calculations involving representative structures show that the domains of the two enzymes superpose well, with an r.m.s. deviation in C^{α} positions ranging from 1.09 to 1.37 Å, except when domain 2 of the molecules in crystal 13, 14 and 15 is involved. Superposition involving domain 2 of these three crystals yield r.m.s.d.s ranging from 1.89 to 2.42 Å. As indicated earlier, the molecule in these three crystals differs from that in all others in containing a *cis* Ser189–Pro190 bond. The residue corresponding to position 190 in *Mt*Ung is an alanine in *Hs*UNG. Therefore, the peptide bond preceding it is always *trans* in the human enzyme, while it is *cis* in crystals 13, 14 and 15 and *trans* in all other crystals of *Mt*Ung. Consequently, the conformation of the stretch in *Hs*UNG corresponding to the 168–199 stretch in *Mt*Ung is similar to that in crystals other than 13, 14 and 15. It is the difference in the conformation of this stretch that makes the superposition of domain 2 of the molecules in crystal 13, 14 and 15 onto that of the molecule in the crystals of *Hs*UNG comparatively poor. Among the two glycyl residues which contribute to the conformational flexibility of the stretch to a lesser extent than the *cis–trans* isomerism of the 189–190 bond, Gly182 is replaced by a lysyl residue in *Hs*UNG. However, the Ramachandran angles of Gly182 are appropriate for an L-amino acid and therefore the substitution does not have serious conformational consequences. Thus, the overall structure of *Mt*Ung and *Hs*UNG is similar except that *Mt*Ung can additionally assume a slightly different structure making use of the segmental mobility facilitated by the *cis–trans* isomerism of the 189–190 peptide bond.

Uracil and its derivatives bind to *Hs*UNG in exactly the same way as they bind to *Mt*Ung. However, there are two critical differences between the two at the mouth of the uracil-binding pocket. The relevant residues are Thr72 and Arg92 in *Mt*Ung. The corresponding residues in *Hs*UNG are a glycine and a proline, respectively. Thus, unlike in *Hs*UNG, an appropriate small molecule can hydrogen-bond to Thr72 and Arg92 and occupy this space, as citrate 1 does in crystal 5 (Fig. 9), as well as interfere with DNA binding. Admittedly, this is only an indication provided by the structure, which needs to be pursued further.

4. Conclusions

The structural analyses of 17 different crystals of *Mt*Ung and its complexes with uracil and its derivatives presented here provide a comprehensive picture of the plasticity of the enzyme. This plasticity involves variability in the interdomain orientation and segmental mobility of a 32-residue stretch. The structures also provide evidence for the limited variability in the orientation of ligands when bound to the enzyme.

A large number of crystallographically independent structures of UNG/Ung from different sources and their complexes with Ugi, DNA fragments and small ligands have been reported earlier. In terms of the orientation between the two domains, the structures can be classified into two broad categories: those interacting with uracil-containing or abasic DNA, with a relatively closed structure, and all others, with an open structure. In the light of the current discussions on conformational selection *versus* induced fit, it is of interest to enquire whether uracil-containing or abasic DNA induces the closed conformation or whether the closed form is among the conformations that are accessible to the native enzyme. In none of the 17 structures reported here does the molecule interact with DNA. However, 11 of them have a closed conformation, five an open conformation and one a half-open conformation. It would thus appear that the molecule can access an ensemble of conformations, of which the closed conformations are chosen when it interacts with the appropriate DNA.

Domain 1 is relatively rigid. Except for the 168–199 segment, domain 2 is also relatively rigid. The conformational difference is most pronounced in the 189–199 stretch, mainly on account of *cis–trans* isomerization about the Ser189–Pro190 peptide bond. The variability of this segment is accentuated by the differences in the main-chain conformation at Gly169 and Gly182. The leucine loop (191–199), which interacts extensively with DNA, forms part of the most variable stretch of the segment. This flexibility perhaps endows the molecule with a degree of versatility in its interaction with DNA.

A detailed examination of the complexes of *Mt*Ung with uracil and its derivatives reveal that the ligand can assume two orientations that are related to each other by a twofold axis passing through N1 and C6 of the uracil moiety. Both of the orientations appear with equal facility when the ligand is uracil. Only one orientation is possible in the case of some of the uracil derivatives, as the other orientation results in severe steric clashes. Most of the residues which interact with the ligands belong to domain 1. His191 is the only residue from

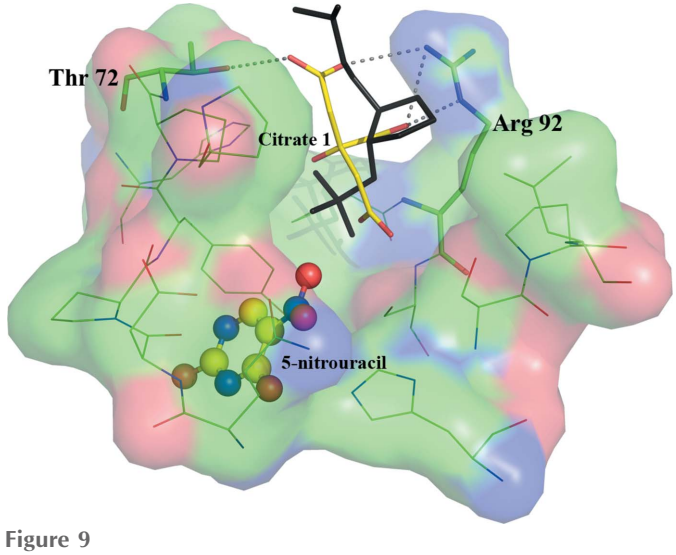

Figure 9
Surface representation of the binding pocket of *Mt*Ung as seen in crystal 5. The bound ligand and citrate ions are shown. Also shown in black is the region of the DNA molecule illustrated in Fig. 8(*b*), which overlaps with the location of the citrate ion in crystal 5.

domain 2 that interacts with the ligand. In the cases where the Ser189–Pro190 peptide bond is *cis*, this interaction is abolished as the histidyl residue moves away from the ligand. This does not appear to cause any dislocation of the ligand, indicating that the enzyme can bind uracil and its derivatives even without an interaction with His191. Thus, residues from domain 1 are almost wholly responsible for binding uracil, whereas most of the interactions with DNA involve domain 2. The structures presented here provide a reasonably comprehensive picture of the interactions of Ung with uracil and its derivatives. Thermodynamic parameters provide a satisfactory rationale for the information provided by the crystal structures. This information and that on the differences between *Mt*Ung and the human enzyme at the mouth of the binding pocket could be useful for structure-based design of inhibitors of the enzyme.

The interactions of the two citrate ions bound to the molecule in a few crystals appear to present a case of molecular mimicry. They overlap at two regions of the bound DNA when the structures are superposed onto those of human UNG–DNA complexes. Furthermore, one carboxylate O atom in each citrate ion interacts with the protein in a very similar way to a phosphate O atom from DNA. It has been observed earlier that a citrate ion from the buffer occupies the location of the phosphate tail of the nucleotide near the P-loop of *M. tuberculosis* pantothenate kinase when nucleotides are not present in the medium (Chetnani *et al.*, 2009). Similar observations have also been made for the structures of human dyanin-1-like GTPase (Wenger *et al.*, 2013) and *M. tuberculosis* nucleoside diphosphate kinase (Georgescauld *et al.*, 2014). Thus, the locations of the citrate ions in the present structures could well be appropriately made use of to interfere with the DNA binding of the enzyme.

Acknowledgements

X-ray data were collected at ESRF, Grenoble through an arrangement made by the Department of Biotechnology (DBT), at Indus-2 at RRCAT, Indore and at the X-ray facility for Protein X-ray Structure Determination and Protein Design at this institute, supported by the Department of Science and Technology (DST). Some of the computations were performed at the Graphics Facility supported by the DBT. PM is a DBT Research Associate. AS is a Bhatnagar fellow of the Council of the Scientific and Industrial Research. UV is a J. C. Bose fellow of the DST. MV is Albert Einstein Professor of the Indian National Science Academy. This work was supported by a research grant from the DBT.

References

Arora, A., Chandra, N. R., Das, A., Gopal, B., Mande, S. C., Prakash, B., Ramachandran, R., Sankaranarayanan, R., Sekar, K., Suguna, K., Tyagi, A. K. & Vijayan, M. (2011). *Tuberculosis*, **91**, 456–468.

Assefa, N. G., Niiranen, L., Johnson, K. A., Leiros, H.-K. S., Smalås, A. O., Willassen, N. P. & Moe, E. (2014). *Acta Cryst.* D**70**, 2093–2100.

Assefa, N. G., Niiranen, L., Willassen, N. P., Smalås, A. & Moe, E. (2012). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **161**, 60–68.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.

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.

Boehr, D. D., Nussinov, R. & Wright, P. E. (2009). *Nature Chem. Biol.* **5**, 789–796.

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. H. (1997). *J. Appl. Cryst.* **30**, 1160–1161.

Cone, R., Bonura, T. & Friedberg, E. C. (1980). *J. Biol. Chem.* **255**, 10354–10358.

DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.

Duncan, B. K. & Weiss, B. (1982). *J. Bacteriol.* **151**, 750–755.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Focher, F., Verri, A., Spadari, S., Manservigi, R., Gambino, J. & Wright, G. E. (1993). *Biochem. J.* **292**, 883–889.

French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.

Friedberg, E. C., Walker, G. C. & Siede, W. (1995). *DNA Repair and Mutagenesis*. Washington: American Society for Microbiology.

Georgescauld, F., Moynié, L., Habersetzer, J. & Dautant, A. (2014). *Acta Cryst.* F**70**, 40–43.

Géoui, T., Buisson, M., Tarbouriech, N. & Burmeister, W. P. (2007). *J. Mol. Biol.* **366**, 117–131.

Handa, P., Acharya, N. & Varshney, U. (2002). *Nucleic Acids Res.* **30**, 3086–3095.

Impellizzeri, K. J., Anderson, B. & Burgers, P. M. (1991). *J. Bacteriol.* **173**, 6807–6810.

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.

Kaushal, P. S., Talawar, R. K., Varshney, U. & Vijayan, M. (2010). *Acta Cryst.* F**66**, 887–892.

Koradi, R., Billeter, M. & Wüthrich, K. (1996). *J. Mol. Graph.* **14**, 51–55.

Krokan, H. & Urs Wittwer, C. (1981). *Nucleic Acids Res.* **9**, 2599–2614.

Krosky, D. J., Bianchet, M. A., Seiple, L., Chung, S., Amzel, L. M. & Stivers, J. T. (2006). *Nucleic Acids Res.* **34**, 5872–5879.

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. (1982). *Annu. Rev. Biochem.* **51**, 61–87.

Lindahl, T. (1993). *Nature (London)*, **362**, 709–715.

Lindahl, T., Ljungquist, S., Siegert, W., Nyberg, B. & Sperens, B. (1977). *J. Biol. Chem.* **252**, 3286–3294.

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

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Moe, E., Leiros, I., Riise, E. K., Olufsen, M., Lanes, O., Smalås, 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.

Murillo, A. C. *et al.* (2007). *Infect. Disord. Drug Targets*, **7**, 127–139.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E. & Tainer, J. A. (1998). *EMBO J.* **17**, 5214–5226.
Parikh, S. S., Putnam, C. D. & Tainer, J. A. (2000). *Mutat. Res.* **460**, 183–199.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
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.
Raeder, I. L. U., Moe, E., Willassen, N. P., Smalås, A. O. & Leiros, I. (2010). *Acta Cryst.* F**66**, 130–136.
Ravishankar, R., Bidya Sagar, M., Roy, S., Purnapatre, K., Handa, P., Varshney, U. & Vijayan, M. (1998). *Nucleic Acids Res.* **26**, 4880–4887.
Saikrishnan, K., Bidya Sagar, M., Ravishankar, R., Roy, S., Purnapatre, K., Handa, P., Varshney, U. & Vijayan, M. (2002). *Acta Cryst.* D**58**, 1269–1276.
Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995). *Nature (London)*, **373**, 487–493.
Savva, R. & Pearl, L. H. (1995). *Nature Struct. Mol. Biol.* **2**, 752–757.
Schormann, N., Banerjee, S., Ricciardi, R. & Chattopadhyay, D. (2013). *Acta Cryst.* F**69**, 1328–1334.
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.
Terwilliger, T. C. *et al.* (2003). *Tuberculosis*, **83**, 223–249.
Turnbull, W. B. & Daranas, A. H. (2003). *J. Am. Chem. Soc.* **125**, 14859–14866.
Varshney, U., Hutcheon, T. & van de Sande, J. H. (1988). *J. Biol. Chem.* **263**, 7776–7784.
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.
Wenger, J., Klinglmayr, E., Fröhlich, C., Eibl, C., Gimeno, A., Hessenberger, M., Puehringer, S., Daumke, O. & Goettig, P. (2013). *PLoS One*, **8**, e71835.
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. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Xiao, G., Tordova, M., Jagadeesh, J., Drohat, A. C., Stivers, J. T. & Gilliland, G. L. (1999). *Proteins*, **35**, 13–24.