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

Overexpression, purification, crystallization and preliminary X-ray analysis of uracil *N*-glycosylase from *Mycobacterium tuberculosis* in complex with a proteinaceous inhibitor

Prem Singh,[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 9 October 2006
Accepted 31 October 2006

Uracil *N*-glycosylase is an enzyme which initiates the pathway of uracil-excision repair of DNA. The enzyme from *Mycobacterium tuberculosis* was co-expressed with a proteinaceous inhibitor from *Bacillus subtilis* phage and was crystallized in monoclinic space group *C*2, with unit-cell parameters a = 201.14, b = 64.27, c = 203.68 Å, β = 109.7°. X-ray data from the crystal have been collected for structure analysis.

1. Introduction

The exocyclic amino group of cytosine is highly susceptible to deamination, a process that occurs in response to normal physiological reactions or environmental pollutants. Such a chemical change in DNA results in the conversion of a G–C base pair to a promutagenic G–U wobble pair. If the uracils from such wobble pairs are not repaired prior to replication, an inevitable consequence is that G–C to A–T mutations will accumulate in the progeny DNA. DNA polymerases may also erroneously incorporate dUMP against adenosine (as an A–U base pair) and hamper the recognition of the regulatory sequences by the proteins. To maintain genomic integrity, cells possess uracil-DNA glycosylases (UDGs), which initiate the uracil-excision repair pathway (Lindahl, 1974). Among these, the uracil *N*-glycosylases (Ungs), belonging to family 1 of the UDGs, are a highly efficient, ubiquitous and conserved class of proteins. *Escherichia coli* Ung (*Ec*Ung), the founder member of the family 1 UDGs, is encoded by the *ung* gene (Lindahl *et al.*, 1977; Varshney *et al.*, 1988). *Mycobacterium tuberculosis* Ung (*Mt*Ung; GenBank accession No. Rv2976c) is a 227-amino-acid protein with a molecular weight of 24.5 kDa. X-ray crystal structures of the Ung protein from *E. coli* and from a number of other sources are available (Savva *et al.*, 1995; Mol, Arvai, Slupphaug *et al.*, 1995; Ravishankar *et al.*, 1998; Saikrishnan *et al.*, 2002; Moe *et al.*, 2004; Leiros *et al.*, 2005)

Ung proteins are inhibited by free uracil and some of its derivatives (Krokan & Wittwer, 1981; Blaisdell & Warner, 1983; Focher *et al.*, 1993; Jiang *et al.*, 2005). Another category of Ung inhibitors is represented by the *Bacillus subtilis* phage PBS-1/PBS-2-encoded early gene product Ugi (Cone *et al.*, 1980; Warner *et al.*, 1980; Wang & Mosbaugh, 1988). Ugi (GenBank accession No. J04434) is a heat-stable acidic 84-amino-acid protein with a molecular weight of 9.5 kDa, which interacts with Ung with a 1:1 molar stoichiometry. Biochemical characterization of complexes of Ugi with the Ung proteins from *M. tuberculosis* and *M. smegmatis* shows that unlike the *Ec*Ung–Ugi complex, which is stable to treatment with 8 *M* urea, the mycobacterial Ung–Ugi complexes dissociate in 5–6 *M* urea (Acharya *et al.*, 2003). However, Ugi is still a potent inhibitor of Ung activity in mycobacteria.

Pathogenic mycobacteria which multiply inside the host macrophages are exposed to reactive oxygen species (ROS) and reactive nitrogen intermediates (RNI), agents which are known to inflict major damage to the genome (Wink *et al.*, 1991). Further, because of the high G+C richness of their genomes, mycobacteria (*e.g. M. tuberculosis*) are naturally at high risk of cytosine deamination, making UDGs a crucial group of DNA-repair enzymes. We showed that *ung*$^-$ strains of *M. smegmatis* were indistinguishable from the

Table 1
Crystal data and data-collection statistics.

Values in parentheses are for the highest resolution shell.

Temparature (K)	100
Space group	*C*2
Unit-cell parameters	
a (Å)	201.14
b (Å)	64.27
c (Å)	203.68
β (°)	109.7
Packing density (V_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)
Observed reflections	426407
No. of unique reflections	43871 (4095)
Completeness (%)	97.6 (91.8)
Multiplicity	9.7
$\langle I/\sigma(I)\rangle$	8.2 (2.4)
R_{sym}† (%)	14.9 (42.1)

† $R_{sym} = \sum_{\mathbf{h}}\sum_l |I_{\mathbf{h}l} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}}\sum_l \langle I_{\mathbf{h}}\rangle$, where I_l is the lth observation of reflection **h** and $\langle I_{\mathbf{h}}\rangle$ is the weighted average intensity for all observations l of reflection **h**.

wild-type strain from their growth under standard culture conditions. Interestingly, in contrast to there being no overt consequences (for at least 1 y) of the *ung*$^-$/*ung*$^-$ genotype in mice (Nilsen *et al.*, 2000), *ung*$^-$ strains of *M. smegmatis* showed increased mutator phenotype and poor endurance in mouse macrophages, especially under conditions of increased RNI production (Venkatesh *et al.*, 2003). Another study, in which a library of randomly integrated transposons containing *M. tuberculosis* were tested for survival in mouse macrophages, the bacteria that survived failed to reveal the presence of transposons within the *ung* locus, indicating that Ung is highly crucial for the survival of mycobacteria in animal models (Sassetti & Rubin, 2003).

As a part of our structural genomics program on mycobacterial proteins (Datta *et al.*, 2000; Roy *et al.*, 2004; Saikrishnan *et al.*, 2003, 2005; Das *et al.*, 2006; Krishna *et al.*, 2006; Selvaraj *et al.*, 2006), we report here the overexpression, purification and crystallization of the *Mt*Ung–Ugi complex with the aim of understanding the specific features of *M. tuberculosis* Ung and the mechanism of Ugi–inhibitor interaction. Distinctive properties of the mycobacterial Ung may allow the design of selective inhibitors against it.

2. Materials and methods

2.1. Expression and purification of *M. tuberculosis* Ung–Ugi complex

To purify the *Mt*Ung–Ugi complex, a pRSETB-based bicistronic construct (Acharya *et al.*, 2003) was modified as follows to shorten the N-terminal tag attached to *Mt*Ung. The pRSETB *Mt*Ung–Ugi construct was digested with *Nhe*I and *Nco*I and the resulting staggered ends were filled in using T4-DNA polymerase and self-ligated (Sambrook *et al.*, 1989). This procedure resulted in the removal of 26 amino acids from pre-Ung sequence, leaving behind Met-His-His-His-His-His-His-Gly-Met-Ala-Ser as presequence to *Mt*Ung. The above construct was transformed into *E. coli* BL21(DE3) cells and a loopful of transformants were inoculated into 1.8 l Luria–Bertani (LB) medium containing 100 µg ml^{-1} ampicillin and grown at 310 K until they reached the late exponential phase. Cells were harvested by centrifugation and resuspended in 15 ml 10 m*M* Tris–HCl pH 7.5 containing 10%(*v*/*v*) glycerol and 500 m*M* NaCl and disrupted by sonication. The lysate was spun in an SS-34 rotor at 15 000 rev min^{-1} using a Sorvall centrifuge for 30 min at 277 K. The supernatant was filtered through a 0.45 µm filter (Sartorious), loaded onto a pre-equilibrated Ni–NTA column (5 ml capacity) and eluted with a 0–500 m*M* imidazole gradient in the above buffer. Fractions containing near-homogenous *Mt*Ung–Ugi complex were pooled, dialyzed against 10 m*M* Tris–HCl pH 7.5 and estimated by Bradford's method using bovine serum albumin as standard (Sedmak & Grossberg, 1977). The final preparation was concentrated to 10 mg ml^{-1} and used for crystallization.

2.2. Crystallization

Initial crystallization screening using Hampton Crystal Screens I and II failed to give any positive results. Therefore, other commercially available screening kits, *i.e.* Index, SaltRx and PEG/Ion Screen from Hampton Research, and Wizard I and II from Emerald Biostructures, were used. Additionally, screening was also performed with commonly used precipitating agents (McPherson, 2004). Hanging-drop vapour-diffusion as well as microbatch methods were used in these experiments. Hanging drops were prepared using 4 µl protein solution and 1 µl reservoir solution and were equilibrated against 500 µl reservoir solution using 24-well Laxbro plates. In the case of the microbatch method, 3 µl protein solution and 3 µl precipitant solution were mixed in a Microtestplate (Sigma). Hampton paraffin oil and silicon oil were used in a 1:1 ratio. The protein solution contained 10 mg ml^{-1} *Mt*Ung–Ugi complex in 10 m*M* Tris–HCl pH 7.5. Crystallization plates were stored at 294 K. Eventually, after ten months, a single crystal appeared in a drop with reservoir containing 10%(*w*/*v*) PEG 8000 and 0.2 *M* NaCl in 0.1 *M* phosphate buffer pH 6.2. Experiments designed to improve the quality using trials around the conditions mentioned above did not yield better crystals.

2.3. X-ray data collection and processing

Diffraction data were collected at low temperature (100 K) from a single crystal using a MAR Research image-plate system (diameter 345 mm) with Osmic mirrors and a Rigaku RU-200 rotating-anode

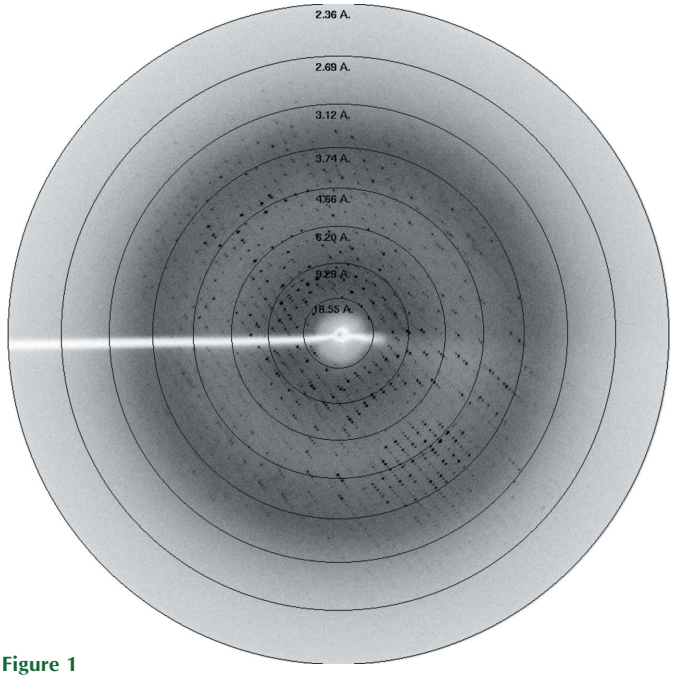

Figure 1
A typical X-ray diffraction image of the *Mt*Ung–Ugi complex using an oscillation of 1°.

```
MtUng    ------------------MTARPLSELVERGWAAALEP--VADQVAHMGQFLRAEIAAGR 40
HsUng    ------------------------MEFFGESWKKHLSGEFGKPYFIKLMGFVAEERK-HY 35
EcUng    -------------------------MANELTWHDVLAEEKQQPHFLNTLQTVASERQSGV 35
HsvUng   MDLTNGGVSPAATSAPLDWTTFRRVFLIDDAWRPLMEPELANPLTAHLLAEYN-RRCQTE 59
                                        *   :          :       .
MtUng    RYLPAGSNVLRAFTFP-FDNVRVLIVGQDPYPTPGHAVGLSFSVAPDVRPWPRSLANIFD 99
HsUng    TVYPPPHQVFTWTQMCDIKDVKVVILGQDPYHGPNQAHGLCFSVQRPVPPPP-SLENIYK 94
EcUng    TIYPPQKDVFNAFRFTELGDVKVVILGQDPYHGPGQAHGLAFSVRPGIAIPP-SLLNMYK 94
HsvUng   EVLPPREDVFSWTRYCTPDEVRVVIIGQDPYHHPGQAHGLAFSVRANVPPPP-SLRNVLA 118
            *.  :*:          :*:*:*:*****  *.:* **.***   :   * ** *:
MtUng    EYTADLG-YPLPSNGDLTPWAQRGVLLLNRVLTVRPSNPASHRGKGWEAVTECAIRALAA 158
HsUng    ELSTDIEDFVHPGHGDLSGWAKQGVLLLNAVLTVRAHQANSHKERGWEQFTDAVVSWLNQ 154
EcUng    ELENTIPGFTRPNHGYLESWARQGVLLLNTVLTVRAGQAHSHASLGWETFTDKVISLINQ 154
HsvUng   AVKNCYPEARMSGHGCLEKWARDGVLLLNTTLTVKRGAAASHSRIGWDRFVGGVIRRLAA 178
                    ..:* *  **: ****** .***:   . **   **: ..  .:  :
MtUng    RAAPLVAILWGRDASTLKPMLAAGNCVAIESPHPSPLSASRGFFGSRPFSRANELLVGMG 218
HsUng    NSNGLVFLLWGSYAQKKGSAIDRKRHHVLQTAHPSPLSVYRGFFGCRHFSKTNELLQKSG 214
EcUng    HREGVVFLLWGSHAQKKGAIIDKQRHHVLKAPHPSPLSAHRGFFGCNHFVLANQWLEQHG 214
HsvUng   RRPGLVFMLWGTHAQN-AIRPDPRVHCVLKFSHPSPLSKVP-FGTCQHFLVANRYLETRS 236
         .   :* :***  *..           .:: .******    *  .. *  :*. *   .
MtUng    AEPIDWRLP------ 227
HsUng    KKPIDWKEL------ 223
EcUng    ETPIDWMPVLPAESE 229
HsvUng   ISPIDWSV------- 244
            ****
```

Figure 2
Multiple sequence alignment of *Mt*Ung and Ungs of known three-dimensional structure using *ClustalW* (Thompson *et al.*, 1994).

X-ray generator. The crystal was soaked in 20%(*v*/*v*) ethylene glycol and reservoir solution for 5 min prior to data collection. The crystal-to-detector distance was kept at 220 mm. Intensity data were processed and scaled using *DENZO* and *SCALEPACK* from the *HKL* program package (Otwinowski & Minor, 1997). The crystal belongs to space group *C*2, with unit-cell parameters a = 201.14, b = 64.27, c = 203.68 Å, β = 109.7°. The crystal diffracted to a resolution of 3.1 Å (Fig. 1). Intensities were converted to structure-factor amplitudes using *TRUNCATE* (Collaborative Computational Project, Number 4, 1994). Data-collection statistics are given in Table 1.

3. Results and discussion

The *Mt*Ung–Ugi complex was crystallized and X-ray data were collected. On the basis of the solvent content (Matthews, 1968), the crystal contains seven molecules in the asymmetric unit. *Mt*Ung has a sequence identity in the range 41–44% with *Ec*Ung, human Ung (*Hs*Ung) and Ung from herpes simplex virus (HsvUng) (Fig. 2). The crystal structures of these three proteins in complex with Ugi are available (Mol, Arvai, Sanderson *et al.*, 1995; Ravishankar *et al.*, 1998; Putnam *et al.*, 1999; Savva & Pearl, 1995). They were used for molecular-replacement calculations using the program *Phaser* (Storoni *et al.*, 2004) from *CCP*4 (Collaborative Computational Project, Number 4, 1994). The calculations led to an acceptable solution with a Z score of 22.9 and a log-likelihood gain of 1743.9 for seven crystallographically independent molecules. Further structural analysis is in progress. The presence of multiple independent copies of the molecules in the crystal is expected to partially compensate for the limited resolution of the data. It also provides a handle for exploring the structural plasticity of the molecule and its implications for eventual inhibitor design.

Data were collected at the X-ray Facility for Structural Biology supported by the Department of Science and Technology. The work forms part of a structural genomics programme supported by the Department of Biotechnology (DBT). PS 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.
Blaisdell, P. & Warner, H. (1983). *J. Biol. Chem.* **258**, 1603–1609.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cone, R., Bonura, T. & Friedberg, E. C. (1980). *J. Biol. Chem.* **255**, 10354–10358.
Das, S., Kumar, P., Bhor, V., Surolia, A. & Vijayan, M. (2006). *Acta Cryst.* D**62**, 628–638.
Datta, S., Prabu, M. M., Vaze, M. B., Ganesh, N., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2000). *Nucleic Acids Res.* **28**, 4964–4973.
Focher, F., Verri, A., Spadari, S., Manservigi, R., Gambino, J. & Wright, G. E. (1993). *Biochem. J.* **292**, 883–889.
Jiang, Y. L., Krosky, D. J., Seiple, L. & Stivers, J. T. (2005). *J. Am. Chem. Soc.* **127**, 17412–17420.
Krishna, R., Manjunath, G. P., Kumar, P., Surolia, A., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2006). *Nucleic Acids Res.* **34**, 2186–2195.
Krokan, H. E. & Wittwer, C. U. (1981). *Nucleic Acids Res.* **9**, 2599–2613.
Leiros, I., Moe, E., Smalas, A. O. & McSweeney, S. (2005). *Acta Cryst.* D**61**, 1049–1056.
Lindahl, T. (1974). *Proc. Natl Acad. Sci. USA*, **71**, 3649–3653.
Lindahl, T., Ljungquist, S., Siegert, W., Nyberg, B. & Sperens, B. (1977). *J. Biol. Chem.* **252**, 3286–3294.
McPherson, A. (2004). *Methods*, **34**, 254–265.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
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.
Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C. F., Andersen, S., Slupphaug, G., Daly, G., Krokan, H. E., Lindahl, T. & Barnes, D. E. (2000). *Mol. Cell*, **6**, 1059–1065.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
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 Acid Res.* **26**, 4880–4887.

Roy, S., Gupta, S., Das, S., Sekar, K., Chatterji, D. & Vijayan, M. (2004). *J. Mol. Biol.* **339**, 1103–1113.
Saikrishnan, K., Bidya Sagar, M., Ravishankar, R., Roy, S., Purnapatre, K., Handa, P., Varshney, U. & Vijayan, M. (2002). *Acta Cryst.* D**58**, 1269–1276.
Saikrishnan, K., Jeyakanthan, J., Venkatesh, J., Acharya, N., Sekar, K., Varshney, U. & Vijayan, M. (2003). *J. Mol. Biol.* **331**, 385–393.
Saikrishnan, K., Kalapala, S. K., Varshney, U. & Vijayan, M. (2005). *J. Mol. Biol.* **345**, 29–38.
Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.
Sassetti, C. M. & Rubin, E. J. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 12989–12994.
Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995). *Nature (London)*, **373**, 487–493.
Savva, R. & Pearl, L. H. (1995). *Nature Struct. Biol.* **2**, 752–757.
Sedmak, J. J. & Grossberg, S. E. (1977). *Anal. Biochem.* **79**, 544–552.
Selvaraj, M., Singh, N. S., Roy, S., Sangeetha, R., Varshney, U. & Vijayan, M. (2006). *Acta Cryst.* F**62**, 913–915.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Varshney, U., Hutcheon, T. & van de Sande, J. H. (1988). *J. Biol. Chem.* **263**, 7776–7784.
Venkatesh, J., Kumar, P., Krishna, P. S., Manjunath, R. & Varshney, U. (2003). *J. Biol. Chem.* **278**, 24350–24358.
Warner, H. R., Johnson, L. K. & Snustad, D. P. (1980). *J. Virol.* **33**, 535–538.
Wang, Z. & Mosbaugh, D. W. (1988). *J. Bacteriol.* **170**, 1082–1089.
Wink, D. A., Kasprzak, K. S., Maragos, C. M. Elespuru, R. K., Misra, M., Dunams, T. M., Cebula, T. A., Koch, W. H., Andrews, A. W. & Allen, J. S. (1991). *Science*, **254**, 1001–1003.