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

Purification, crystallization and preliminary crystallographic analysis of deoxyuridine triphosphate nucleotidohydrolase from *Arabidopsis thaliana*

Mamta Bajaj[a] and Hideaki Moriyama[b]*

[a]School of Biological Sciences, University of Nebraska–Lincoln, Manter Hall, Lincoln, Nebraska 68588-0304, USA, and [b]Department of Chemistry, e-Toxicology and Biotechnology, University of Nebraska–Lincoln, Hamilton Hall, Lincoln, Nebraska 68588-0304, USA

Correspondence e-mail: hmoriyama2@unl.edu

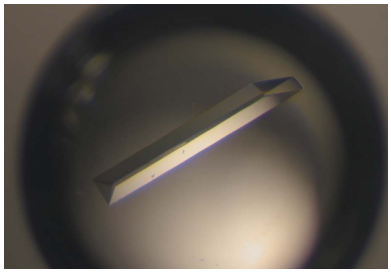
Received 23 February 2007
Accepted 31 March 2007

The deoxyuridine triphosphate nucleotidohydrolase gene from *Arabidopsis thaliana* was expressed and the gene product was purified. Crystallization was performed by the hanging-drop vapour-diffusion method at 298 K using 2 *M* ammonium sulfate as the precipitant. X-ray diffraction data were collected to 2.2 Å resolution using Cu $K\alpha$ radiation. The crystal belongs to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 69.90$, $b = 70.86$ Å, $c = 75.55$ Å. Assuming the presence of a trimer in the asymmetric unit, the solvent content was 30%, with a V_M of 1.8 Å^3 Da^{-1}.

1. Introduction

The ubiquitous enzyme deoxyuridine triphosphate nucleotidohydrolase (dUTPase; EC 3.6.1.23) catalyzes the hydrolysis of deoxyuridine triphosphate (dUTP) to deoxyuridine monophosphate (dUMP) and diphosphate (Mol *et al.*, 1996). This is a housekeeping protein and its roles are to maintain a low dUTP level in order to avoid the incorporation of uracil into DNA (Kornberg & Baker, 1991) and to provide dUMP as a substrate for deoxythymidine triphosphate (dTTP) biosynthesis (Zhang *et al.*, 2005).

The first structure of dUTPase solved was that from *Escherichia coli* (PDB code 1euw; Cedergren-Zeppezauer *et al.*, 1992). Because of the clinical importance of the enzyme, further structures of dUTPases from human (1q5h; Mol *et al.*, 1996), human pathogenic bacteria (1slh; Chan *et al.*, 2004) and mammalian viruses (1dun, Dauter *et al.*, 1999; 1f7d, Prasad *et al.*, 2000) have also been reported. All these enzymes have the same homotrimeric structure and their optimum temperatures are around 310 K; they contain between 134 and 172 amino acids. The dUTPases from human parasitic protozoan trypanosome (PDB code 1ogk; Harkiolaki *et al.*, 2004) and *Campylobacter jejuni* (1w2y; Moroz *et al.*, 2004) are homodimers with 283 and 229 amino acids per monomer, respectively. An extreme optimum temperature is found for the dUTPase from the archaeon *Methanococcus jannaschii*; its optimum temperature is 343–368 K (Li *et al.*, 2003). This enzyme has a hexameric structure and contains 204 amino acids (PDB code 1pkk; Huffman *et al.*, 2003); it is bifunctional as a deoxycytidine triphosphate (dCTP) deaminase. In plants, although meristem-localized expression of dUTPase has been reported (Pri-Hadash *et al.*, 1992), no crystal structure of dUTPase has been reported. The optimum growth temperature of the model plant *Arabidopsis thaliana* is 295 K (Gray *et al.*, 1998). Therefore, we chose *Arabidopsis* dUTPase as a medium/lower-temperature model. The dUTPase from *A. thaliana* contains 166 amino-acid residues.



2. Protein expression and purification

The dUTPase cDNA from *A. thaliana* was cloned in the *Escherichia coli* vector pUNI51 and was obtained from the Arabidopsis Biological Resource Center, Ohio State University (Rhee *et al.*, 2003). The dUTPase gene was amplified by PCR using the following oligonucleotide primers: forward primer 5′-AAAACATATG-GCTTGCGTAAACGAACC-3′ and reverse primer 5′-AAAACTC-GAGTTAGACACCAGTAGAACCAAAACCAC-3′. The forward and reverse primers contained *Nde*I and *Xho*I restriction sites,

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	
a	69.90
b	70.86
c	75.55
Wavelength (Å)	1.542 (Cu $K\alpha$)
Resolution range (Å)	50.0–2.2 (2.3–2.2)
Total observations	199376
Unique reflections	18428
Completeness (%)	99.3 (98.4)
Mean $I/\sigma(I)$	57.1 (37.9)
$R_{\rm merge}$† (%)	7.5 (12.3)

† $R_{\rm merge} = \sum |I_{\rm obs} - \langle I \rangle| / \sum I_{\rm obs}$, where $I_{\rm obs}$ and $\langle I \rangle$ are the observed intensity and the mean intensity of the reflection, respectively.

respectively. The PCR products were purified from 1.2%(*w*/*v*) agarose gels using a QIAquick gel-extraction kit (Qiagen, Valencia, CA, USA). After PCR amplification and purification, the fragments were digested by *Nde*I and *Xho*I and inserted into the corresponding sites of the pET15b vector (Novagen, San Diego, CA, USA).

Expression of the target protein was carried out in *E. coli* BL21 Star (DE3) cells. A 1 l culture was harvested by centrifugation and the cells were disrupted by sonication (Misonix Inc., Farmingdale, NY, USA) for 960 s at maximum amplitude in 2 s pulses under chilled conditions. The clarified lysate was subjected to Ni–nitrilotriacetic acid (NTA) His-Bind batch column chromatography (Novagen, San Diego, CA, USA). The target dUTPase protein was eluted from the column using 50 m*M* sodium phosphate pH 8.0, 0.3 *M* NaCl and 250 m*M* imidazole. In order to remove the His_6 tag, the recombinant protein was digested with thrombin for 16 h at 298 K and the enzyme was purified by Ni–NTA batch column chromatography followed by Benzamidine Sepharose Fast Flow (Amersham Biosciences, Pittsburgh, PA, USA). Approximately 90 mg of protein was purified from a 5 l culture at 277 K.

3. Crystallization

The initial crystallization conditions were obtained using a screening kit from Hampton Research (Aliso Viejo, CA, USA) by the hanging-drop vapour-diffusion method. Using the EasyXtal Tool (Qiagen, Valencia, CA, USA), 1 µl screening solution was mixed with 1 µl protein solution (10 mg ml^{-1} protein and 50 m*M* Tris–HCl pH 7.4) and equilibrated against 1 ml of the same screening solution at 298 K. Two weeks after the initial screening, we found 23 clear drops, 22 drops with heavy precipitation, one drop with phase separation and three drops with an indeterminate number of oil drops. Only one condition out of 50 yielded crystals. These were small plate-like colorless protein crystals stacked upon one another. The reservoir was composed of 2 *M* ammonium sulfate and 0.1 *M* Tris–HCl pH 8.5. Using additional screening kits (Hampton Research) and varying the pH, the primary crystallization conditions were refined. Rod-shaped crystals appeared after mixing 1 µl primary screening solution at pH 9.0 with 0.5 µl 0.1 *M* taurine and 1 µl protein solution. The rod-shaped crystals appeared after two weeks and grew to approximate dimensions of 0.4 × 0.1 × 0.1 mm within one month (Fig. 1).

4. Data-collection and structure solution

For mounting, the crystals were transferred from the crystallization drop into a cryoprotectant solution using a clean nylon loop. The cryoprotectant solution was composed of 20 mg trehalose, 10 µl glycerol and 90 µl reservoir; the final concentrations of trehalose and glycerol were 0.5 and 1.2 *M*, respectively. After soaking for less than 10 s, the crystal was flash-cooled in a nitrogen stream at 93 K. A complete data set was collected from a single crystal using Cu $K\alpha$ X-rays of wavelength 1.542 Å from a generator operating at 40 kV and 20 mA. The native diffraction data consisted of a total of 172 images, each exposed for 1800 s with 1.5° oscillation at a crystal-to-detector distance of 150 mm. The data were indexed, integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997).

The crystals diffracted to a Bragg spacing of 2.2 Å. Observations of systematic absences along the three crystallographic axes are consistent with space group $P2_12_12_1$. Crystal parameters and data-collection statistics are summarized in Table 1. Assuming the presence of three monomeric subunits of dUTPase from *A. thaliana* in the asymmetric unit of the orthorhombic crystal, a Matthews coefficient of 1.8 Å^3 Da^{-1} and a solvent content of 30% were calculated (Matthews, 1968). A preliminary structural analysis was performed using the molecular-replacement method. We chose human dUTPase as the structural template. Of the structurally known dUTPases in the Protein Data Bank, the human enzyme showed the highest similarity to that from *A. thaliana*, with a sequence identity of 56%. A molecular model for the *A. thaliana* dUTPase was generated with *MODELLER* (Marti-Renom *et al.*, 2000) using the *A. thaliana* dUTPase core sequence (Phe29–Val141; 113 amino acids) with chain *A* of the human dUTPase as a template (PDB code 1q5h; Leu3–Phe115). The molecular-replacement calculations were performed for three subunits using *CCP*4 (Collaborative Computational Project, Number 4, 1994). The rotation-function and translation-search calculations were performed with all data in the resolution range 50.0–2.2 Å. The correctly oriented model was subjected to rigid-body refinement of the trimer using *REFMAC*5 (Collaborative Computational Project, Number 4, 1994). This model was used to calculate $2F_{\rm obs} - F_{\rm calc}$ electron-density maps, which were visually inspected using the interactive molecular-graphics program *Coot* (Emsley & Cowtan, 2004). After repetitive calculations and visual inspections, the current structure contained 123, 125 and 126 visible residues out of 166 residues for each of the three monomers *A*, *B* and *C*, respectively. The N-terminal and C-terminal residues were not visible in the electron-density map. The refined structure yielded an *R* factor of 0.20 and an $R_{\rm free}$ of 0.28 for data in the resolution range 50.0–2.2 Å.

There were approximately 25 amino acids at the N-terminus and 13 amino acids at the C-terminus of the protein molecule that were not visible that may cause a high $R_{\rm free}$ even after the addition of 210 water molecules. Previous studies have shown that the glycine-rich

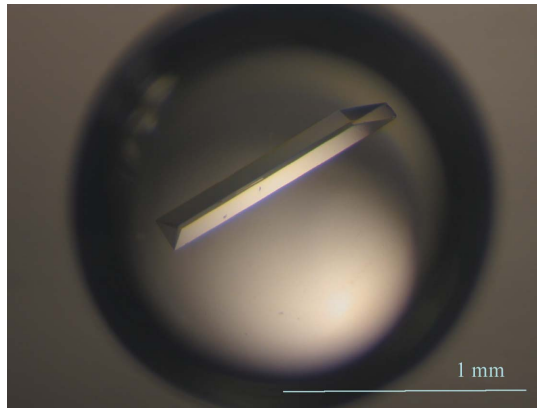


Figure 1
A crystal of dUTPase from *A. thaliana*.

C-terminal tail of one subunit interacts with the active site of a second subunit of the molecule (Mol *et al.*, 1996). Three active-site regions were observed in the current structure. We attempted to fix the C-terminal residues in order to yield better crystals by adding the inhibitor deoxyuridine diphosphate at various molar ratios. However, those efforts resulted in smaller and highly stacked plate crystals. This structure has been deposited in the PDB with code 2pc5.

This work was supported in part by a start-up fund from the University of Nebraska–Lincoln to HM. We would like to thank Professors James Van Etten and Etsuko Moriyama (University of Nebraska–Lincoln) for constant scientific support. We thank Drs Yuanzheng Zhang and Kohei Homma for technical support. We are grateful to the Arabidopsis Biological Resource Center at Ohio State University.

References

Cedergren-Zeppezauer, E. S., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. S. (1992). *Nature (London)*, **355**, 740–743.
Chan, S., Segelke, B., Lekin, T., Krupka, H., Cho, U. S., Kim, M.-Y., So, M., Kim, C.-Y., Naranjo, C. M., Rogers, Y. C., Park, M. S., Waldo, G. S., Pashkov, I., Cascio, D., Perry, J. L. & Sawaya, M. R. (2004). *J. Mol. Biol.* **341**, 503–517.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dauter, Z., Persson, R., Rosengren, A. M., Nyman, P. O., Wilson, K. S. & Cedergren-Zeppezauer, E. S. (1999). *J. Mol. Biol.* **285**, 655–673.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gray, W. M., Ostin, A., Sandberg, G., Romano, C. P. & Estelle, M. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 7197–7202.
Harkiolaki, M., Dodson, E. J., Bernier-Villamor, V., Turkenburg, J. P., Gonzalez-Pacanowska, D. & Wilson, K. S. (2004). *Structure*, **12**, 41–53.
Huffman, J. L., Li, H., White, R. H. & Tainer, J. A. (2003). *J. Mol. Biol.* **331**, 885–896.
Kornberg, A. & Baker, T. A. (1991). *DNA Replication*, 2nd ed. New York: Freeman Press.
Li, H., Xu, H., Graham, D. E. & White, R. H. (2003). *J. Biol. Chem.* **278**, 11100–11106.
Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. & Sali, A. (2000). *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mol, C. D., Harris, J. M., McIntosh, E. M. & Tainer, J. A. (1996). *Structure*, **4**, 1077–1092.
Moroz, O. V., Harkiolaki, M., Galperin, M. Y., Vagin, A. A., González-Pacanowska, D. & Wilson, K. S. (2004). *J. Mol. Biol.* **342**, 1583–1597.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Prasad, G. S., Stura, E. A., Elder, J. H. & Stout, C. D. (2000). *Acta Cryst.* D**56**, 1100–1109.
Pri-Hadash, A., Hareven, D. & Lifschitz, E. (1992). *Plant Cell*, **4**, 149–159.
Rhee, S. Y. *et al.* (2003). *Nucleic Acids Res.* **31**, 224–228.
Zhang, Y., Moriyama, H., Homma, K. & Van Etten, J. L. (2005). *J. Virol.* **79**, 9945–9953.