STRUCTURAL ANALYSIS OF DEOXYURIDINE TRIPHOSPHATASE FROM *ARABIDOPSIS THALIANA*

by

Mamta Bajaj

A THESIS

Presented to Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Biological Sciences

Under the Supervision of Professor Etsuko Moriyama and Professor Hideaki Moriyama

Lincoln, Nebraska

May, 2007

STRUCTURAL ANALYSIS OF DEOXYURIDINE TRIPHOSPHATASE FROM *ARABIDOPSIS THALIANA*

Mamta Bajaj, M.S.

University of Nebraska, 2007

Advisers: Etsuko Moriyama and Hideaki Moriyama

In order to analyze the 3D-structure of a plant deoxyuridine triphosphate nucleotidohydrolase (dUTPase) and investigate its structural differences from a plant viral dUTPase, I solved the first plant dUTPase structure from *Arabidopsis thaliana*. The dUTPase gene from *A. thaliana* was expressed and the gene product was purified. Crystallization was performed by the hanging drop-vapor diffusion method at 298 K using 2 M ammonium sulfate as the precipitant. X-ray diffraction data were collected upto a 2.2 Å resolution using CuKα radiation. The crystal belongs to the orthorhombic space group, $P2_12_12_1$, with unit-cell dimensions of a = 69.90 Å, b = 70.86 Å, and c = 75.55 Å. Assuming the presence of a molecule consist of three homo-subunits in an asymmetric unit, the solvent content was 30% with a V_M of 1.8 $Å^3Da^{-1}$. The structural analysis of dUTPase from *A. thaliana* was performed using the molecular replacement method. The refined structure that resulted values of R factor = 18% and R_{free} = 25% has been deposited in the Protein Data Bank with ID 2PC5. This structure had one possible magnesium site which has been coordinated with three glutamic acids, GLU 138 of each of three chains. The crystal quality was improved after the long-term soaking method.

The structure analyzed from the crystal soaked in magnesium chloride had R factor = 19% and R_{free} = 23% (PDB ID: 2P9O). This structure had two possible magnesium sites coordinated with aforementioned GLU 138 as well as three arginines, ARG 64 of each of three chains. The current structure contained 128, 126, and 127 visible residues from the each chain A, B, and C, respectively. The dUTPase structure from *Arabidopsis* was compared to the structures from human (PDB ID: 1Q5H) and chlorella virus IL-3A (unpublished). Small but important changes that were able to explain kinetic differences.

ACKNOWLEDGEMENTS

I want to thank Dr. Etsuko Moriyama and Dr. Hideaki Moriyama for advising me throughout this thesis and the MS program at School of Biological Sciences. I would also like to thank Dr. James Ven Etten for his continuous guidance and support throughout this thesis. Thanks to Dr. Catherine P. Chia and Dr. John Osterman for their support and guidance. I thank Drs. Yuanzheng Zhang, Kohei Homma, and Mia Nakachi for technical support. I am grateful to DNA sequencing lab in the Center for Biotechnology at University of Nebraska – Lincoln and to the Arabidopsis Biological Resource Center at Ohio State University.

I would like to thank my friends Dr. Nagrama Kottopalli, Aravind Visvanathan, and Sreedevi Madhusoodhanan for supporting and sharing their ideas. Thanks to my brother, Prashant Bajaj, for supporting me throughout the thesis. I want to thank my parents for their patience, encouragement, and support throughout my work. This thesis would not have been possible without the blessings and inspiration from the Art of Living Founder, H. H. Sri Sri Ravi Shankar ji.

Table of Contents

Chapter 1

Introduction

The ubiquitous enzyme, deoxyuridine triphosphate nucleotidohydrolase (dUTPase; EC 3.6.1.23; Figure 1), catalyzes the hydrolysis of deoxyuridine triphosphate (dUTP) to deoxyuridine monophosphate (dUMP) and diphosphate (Mol *et al.*, 1996). This is a housekeeping enzyme. As shown in Figure 2, its roles are to maintain a low dUTP level 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 from *Escherichia coli* (PDB ID 1EUW; Cedergren-Zeppezauer *et al.*, 1992). Due to its clinical importance, more structures of dUTPases were reported mainly 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) (Figure 3; Table 1). All these enzymes take the same homo-trimer structure and their optimum temperatures are at around 310 K, with amino acid lengths ranging between 134 and 172. The dUTPases from human parasitic protozoan trypanosome (1OGK; Harkiolaki *et al.*, 2004) and *Campylobacter jejuni* (1W2Y; Moroz *et al.*, 2004) are homo-dimers with 283 and 229 amino acids per monomer, respectively. An extreme optimum temperature is found in the dUTPase from Archaeon *Methanococcus jannaschii*, whose optium temperature exhibits 343-368 K (Li *et al.*, 2003). This enzyme takes a hexamer formation with 204 amino acids (1PKK;

Huffman *et al.*, 2003), and is bifunctional as deoxycytidine triphosphate (dCTP) deaminase.

In plants, although meristem-localized expression of dUTPase has been reported (Pri-Hadash *et al.*, 1992), there was no crystal structure of dUTPase reported. The optimal 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.

The objective of this project was to analyze the 3D-structure of dUTPase from *A. thaliana* and investigate differences between the plant and plant viral dUTPases. I used the 3D-structure of chlorella virus IL-3A (unpublished data from our lab) for structural comparison. In this thesis, I describe expression, purification, crystallization, and structure analysis of the dUTPase from *A. thaliana*. This is the first plant dUTPase structure from *A. thaliana*. Upon structural comparison of dUTPases between *A. thaliana* and chlorella virus IL-3A, small but significant differences were observed in the size and shape of the active sites.

The outlines of this thesis are organized as follows. Chapter 2 describes the cloning, expression and purification of the dUTPase gene from *Arabidopsis thaliana*. This chapter also explains the method used to confirm the enzyme activity. In Chapter 3, the protein concentration and the crystallization process of the *Arabidopsis* dUTPase is explained. The crystallographic structural analysis on the dUTPase protein crystal and the molecular replacement method used for the analysis are discussed in Chapter 4. Finally Chapter 5 concludes this thesis with overall description of the *Arabidopsis* dUTPase 3D-structure and the discussion on the differences between its structures from

Arabidopsis, human, and chlorella virus IL-3A. The majority of the thesis content has been published in Acta Crystallographica Section F (Bajaj & Moriyama, 2007). The reprint of this paper is attached in Appendix A.

Table 1. A list of sequences used in Figure 3.

Sequences	PDB ID	Loci from NCBI	Description
Chlorella			http://greengene.uml.edu gene_98|GeneMark.hmm|231_aa
Ostreococcus		CAL57986	putative dUTP pyrophosphatase from *Ostreococcus tauri*, a smallest algae
Arabidopsis	2pc5, 2p9o	NP_190278	dUTP diphosphatase / hydrolase from *Arabidopsis thaliana*
Tomato		P32518	DUT_SOLLC Deoxyuridine 5'-triphosphate nucleotidohydrolase (P18)
Dictyostelium		XP_629169	dUTP diphosphatase from Dictyostelium discoideum AX4
Chimpanzee		XP_001165929	PREDICTED: dUTP pyrophosphatase isoform 5 from Pan troglodytes
Human	1q5h	NP_001020420	dUTP pyrophosphatase isoform 3 from Homo sapiens
PBCV-1		O41033	PBCV-1, Paramecium bursaria chlorella virus 1
SH-6A		AAW51453	SH-6A, Paramecium bursaria chlorella virus
IL-3A		AAW51452	IL-3A, Paramecium bursaria chlorella virus
Avian adenovirus		Q9YYS0	Avian adenovirus, Avian adenovirus type 8 (Fowl adenovirus 8)
Vaccinia virus		P21035	Vaccinia virus
Horse IAV	1dun	P11204	Equine Infectious Anemia virus (EIAV)
Cat IDV	1f7d	P16088	Feline Immunodeficiency virus (FIV)
Fruit fly		Q8IPB1	*Drosophila melanogaster* (174 total residues, C-terminal sequence not shown in alignment: NGEKAAEPEGAAPAPVAT)
Yeast		P33317	*Saccharomyces cerevisiae*
Mycobacterium	1six	O07199	*Mycobacterium tuberculosis*
E. coli	1euw	P06968	*Escherichia coli*
T. tengcongensis		Q8RA46	*Thermoanaerobacter tengcongensis*, a thermophile

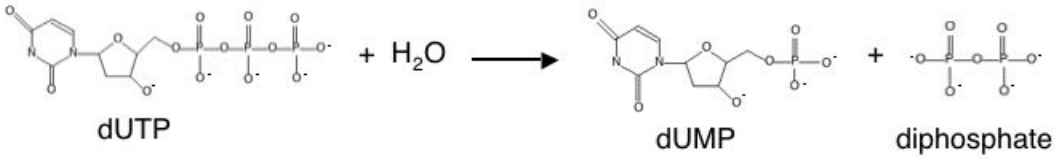


Figure 1. The enzyme reaction by deoxyuridine triphosphate nucleotidohydrolase (EC 3.6.1.23).

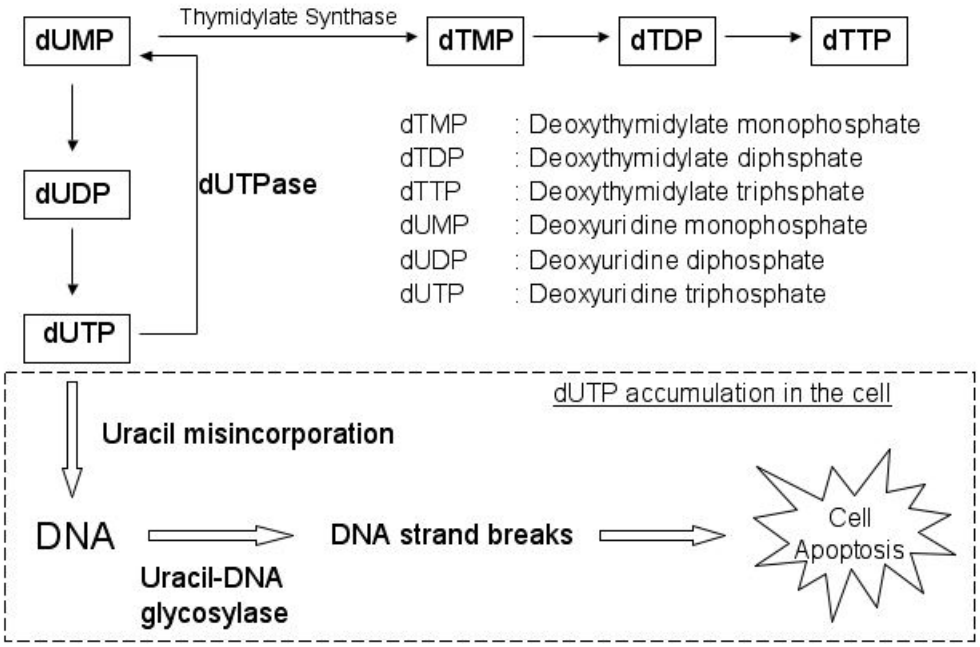


Figure 2. Physiological roles of dUTPase.

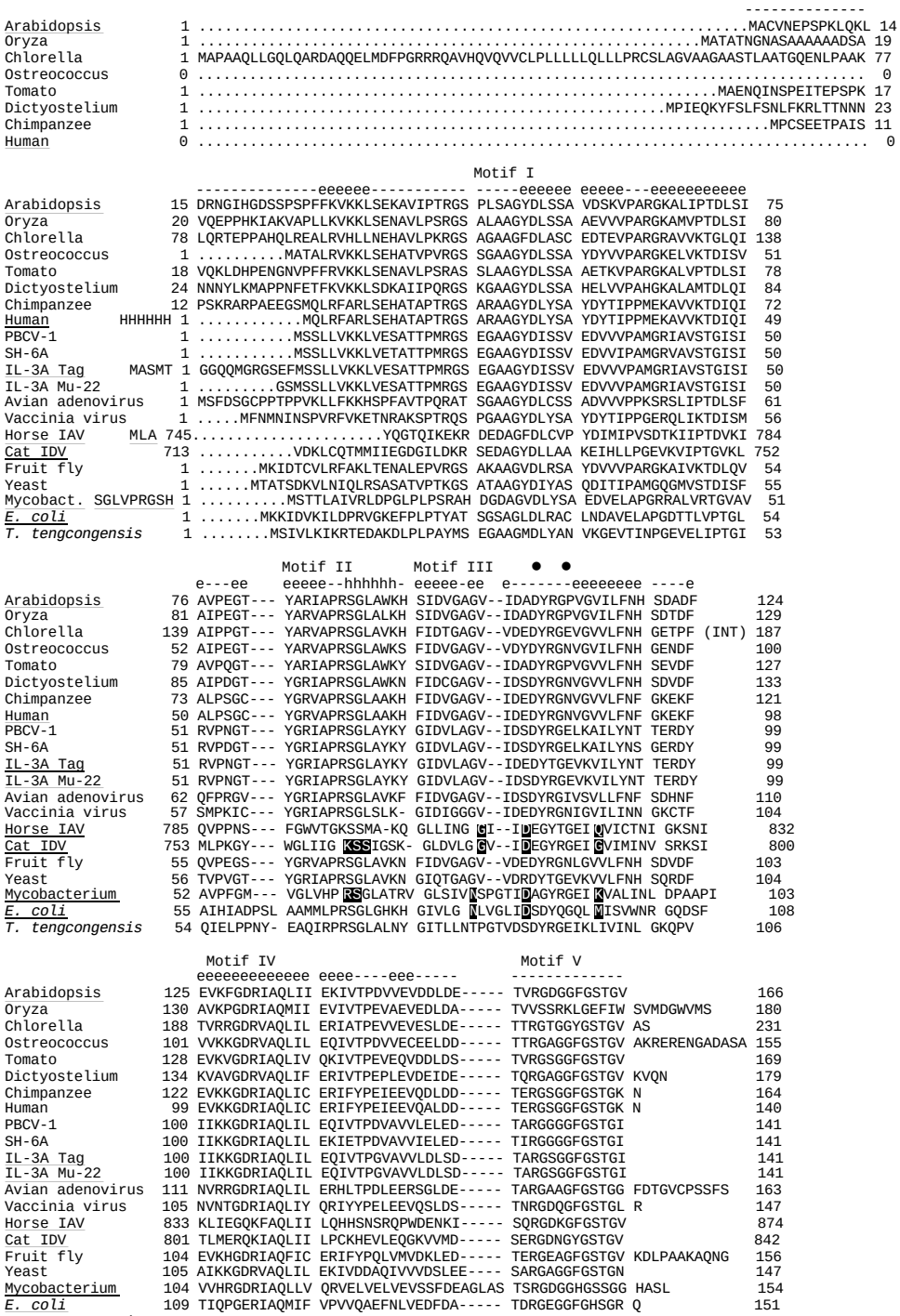


Figure 3. The multiple sequence alignment of selected dUTPases. The left-most column represents source organisms. The details of sequence source were shown in Table 1.

Underlines in the sequence names indicate that their 3D-structures are available. Motifs I through V are conserved domains. Large dots (•) indicate key residues that determine the reaction optimal temperature. Characters with black background indicate the residues that are interacting with substrates revealed by the 3D-structure analysis.

Chapter 2

Cloning, expression, and purification of the dUTPase from *Arabidopsis thaliana*

In this chapter, I will describe the designing of expression system, cloning, expression, and purification of the dUTPase from *Arabidopsis thaliana*. This section also includes the activity assay.

To construct the expression system in *Escherichia coli*, I used the pET-15b expression vector. This vector carries an N-terminal His-tag followed by a thrombin site and three cloning sites. The dUTPase cDNA amplified from the pUNI51 plasmid was inserted into the pET-15b expression vector (restriction sites: *Nde*I and *Xho*I) as shown in Figure 4. The protein sequence after expression contains four amino acids at the N-terminus of His-tag and thrombin site (Table 2). The His-tag was removed by thrombin digestion. The resulting protein carried three amino acids (Glycine, Serine, and Histidine) at the N-terminal region of the protein.

All procedure used in this study followed the general manual (Sambrook & Russell, 2001).

2.1 Single colony isolation

The dUTPase cDNA from *A. thaliana* (Figure 4, Table 2) was obtained from Arabidopsis Biological Resource Center, Ohio State University (Rhee *et al.*, 2003). Single colonies were isolated on the solid Luria-Bertani (LB) broth (Sigma-ALDRICH, St. Louis, MO) media containing 25 µg/ml kanamycin (Sigma-ALDRICH) using the

streak plate method. The streaked plates were incubated for 16 hours at 310 K. I repeated this process twice to have pure colonies. The single colonies were transferred to new LB-kanamycin agar plates and stored at 277 K. Glycerol stock was prepared and stored at 193 K.

2.2 Isolation of plasmid DNA by alkaline lysis

The single colony was selected and inoculated into a 3 ml culture of LB-kanamycin (25 µg/ml) liquid medium. The tube was cultivated at 305 K with vigorously shaking at 250 rpm for 16 hours. The cells were collected using centrifugation at 14000 x g for 1 min under 277 K. Plasmid DNA was extracted using QIAprep Spin Miniprep kit (QIAGEN, Valencia, CA). DNA concentration and purity was measured using UV spectrophotometer DU-800 (Beckman Coulter Inc., Fullerton, CA). The spectrophotometer was first blanked with sterile water. The UV-spectrum of DNA samples with 50 times dilution by water was recorded. The quality of DNA was judged by the ratio between the absorbance at 260 nm and 280 nm. In this project, DNA with A_{260}/A_{280} greater than 1.8 was considered pure. The plasmid DNA was then stored in 293 K.

To confirm the plasmid structure, I performed a restriction analysis using *EcoR*I and *Hind*III (Figure 4). The reaction solution contained 1µl of *EcoR*I (10 unit/µl), 1 µl of *Hind*III (10 unit/µl), 5 µl of Buffer R (10x concentrated; Fermentas Inc., Hanover, MD; 10 mM Tris-HCl pH 8.5, 10 mM magnesium chloride, 100 mM potassium chloride, and 0.1 mg/ml bovine serum albumin), 2 µg of plasmid, and sterile deionized water making the final volume to be 50 µl. The reaction solution was incubated for two hours at 310 K.

After the reaction, thermal inactivation of the enzymes was carried out for 20 min at 338 K. Two fragments, 1239 bp and 1848 bp, were expected after the restriction enzyme digestion. Electrophoresis was done using 0.8% agarose gel, as shown in Figure 5. The gel was then stained with 8 nM ethidium bromide solution and the bands were observed under the UV light. To verify the dUTPase gene, the DNA sequencing was performed in the Center for Biotechnology at University of Nebraska – Lincoln. The sequencing was carried out with the following 5’ and 3’-primers: T3, 5’-CAATTAACCCTCACTAAAGG and T7 terminator 5’-GCTAGTTATTGCTCAGCGG. The sequences obtained from the facility were verified to be consistent with the DNA sequence available from the Arabidopsis Biological Resource Center (Figure 5).

2.3 Subcloning

2.3a Polymerase chain reaction

In order to subclone the dUTPase gene, the target gene was amplified using polymerase chain reaction (PCR) and the resulted DNA fragment was cloned by the expression vector. To do the PCR, a large scale DNA preparation was done using 12 ml of LB-kanamycin culture.

The dUTPase was cloned from PCR-amplified cDNA with the following oligonucleotide primers: 5'-primer, 5'-AAAACATATGGCTTGCGTAAACGAACC-3' and 3'-primer, 5'-AAAACTCGAGTTAGACACCAGTAGAACCAAAACCAC-3'. The 5' and 3'-primers contained an *Nde*I and *Xho*I restriction sites, respectively. The dUTPase gene was amplified with the Platinum P*fx* DNA polymerase solution (Invitrogen, Carlsbad, CA) that contained 5 µl of 10x P*fx* amplification buffer provided by the vendor,

200 ng of DNA, 0.3 µM concentrations of each primer, 1.5 µl of 10 mM dNTP mixture (0.3 mM each of dATP, dGTP, dCTP, and dTT), 1 mM of $MgSO_4$, 2.5 units Platinum P*fx* DNA polymerase, and sterile water obtaining the final volume of 50 µl. The PCR process was carried out in 35 cycles by heating and cooling the reaction tubes as follows: denaturing at 367 K for 15 s, annealing at 323 K for 30 s, and elongation at 341 K for 32 s. In addition to the 35 cycles, the reaction was run under 341 K for the last 10 min to ensure the elongation step. To prevent evaporation of the reaction mixture, a layer of mineral oil (25 µl) was added on the surface of the mixture.

After the reaction, the mineral oil was removed and the aqueous phase was separated by adding 150 µl of chloroform to each reaction tube. The PCR products were isolated from 1.2% agarose gels (Figure 6) by using Qiaquick gel extraction kit (QIAGEN, Valencia, CA) and stored at 253 K.

To concentrate the DNA in the PCR product, the isopropanol precipitation method was used. To one volume of DNA solution, 0.3 volumes of 3 M sodium acetate pH 5.2 was added and mixed gently. Additional one volume of isopropanol was added at the room temperature and mixed gently. The solution was then centrifuged at 17000 x g for 20 min at 277 K and the supernatant was removed. The precipitant was washed with a 1 ml of 70% ethanol at the room temperature and centrifuged at 17000 x g for 20 min at 277 K. After removing the supernatant, the precipitant was vacuum dried for 10 min and sterile water was added. The concentrated PCR product was 95 ng/µl.

2.3b Restriction enzyme digestion

The fragments and the expression vector were digested separately with restriction enzymes *Nde*I and *Xho*I in two steps. The restriction enzyme reaction solution for the pET-15b expression vector contained 0.6 µl *Nde*I (10 unit/µl), 5 µl of Buffer R, 2 µg of pET-15b vector. Sterile water was added in order to obtain the final volume of 50 µl. Similarly, the restriction enzyme reaction solution for the PCR product contained 3.6 µl *Nde*I (10 unit/µl), 5 µl of Buffer R, 2 µg of PCR products, and sterile water was added in order to obtain the final volume of 50 µl. The solutions were incubated for 16 hours at 310 K and heated for 20 min at 338 K for inactivation of the enzymes.

After cleaning DNA by Qiagen purification protocol, 40 µl of Buffer EB (10 mM Tris-Hcl, pH 8.5) from the Qiaquick gel extraction kit was added to the restriction enzyme solutions. The solution for each of the pET-15b expression vector and the PCR products contained 40 µl of solutions obtained after DNA cleanup, 4.5 µl and 28.7 µl of *Xho*I (10 unit/µl), respectively, 5 µl of Buffer R (Fermentas Inc., Hanover, MD), and sterile water to obtain the final volume of 50 µl. The solutions were incubated for 16 hours at 310 K and incubated for 20 min 353 K for inactivation of the enzymes. The digested DNA was cleaned up and concentrated by isopropanol precipitation. The resulted DNA had concentrations at 78.6 ng/µl for pET-15b and 30.5 ng/µl for PCR products.

2.3c Ligation and transformation

The digested fragments of dUTPase gene were inserted into the *Nde*I/*Xho*I sites of the pET-15b expression vector (Figure 4). Ligation reaction was carried in 20 µl solution

under the following conditions: molar ratio of vector to insert as one to three, 10% ligase buffer provided by the vendor, 10 units of T4 DNA ligase, and incubation for 16 hours at 289 K. Transformation of *E. coli* DH5α-T1^{R} competent cells by the ligated vector was done by adding 5 µl of ligated solution into the 50 µl of the competent cell suspension. The mixture was incubated on ice for 1 hour and then for 45 s in 318 K. After the heat shock, the solution was again incubated on ice for 3 min and then 945 µl of SOC medium was added to it. The mixture was then incubated for 20 min at 310 K in still followed by shaking vigorously for 40 min at 250 rpm and the temperature was set to 310 K. The culture up to 100 µl was transferred to agar plates with 100 mg/ml ampicillin. The rest of culture was subjected to the centrifugation at 5000 rpm for 10 min. The remaining precipitated cells were removed, about 100 µl supernatant was mixed, and it was transferred to another agar plate. The plates were stored at the room temperature until the liquid had been absorbed. Then the plates were inverted and incubated at 310 K for 16 hours.

2.3d Identification of clones

The colonies appeared on the plates were examined by colony PCR (Dieffenbach & Dveksler, 1995). Under this method, the colonies selected were streaked on to a new plate and let it grow for 16 hours at 310 K. Then the conditions described in the previous section were used to perform PCR. The PCR reaction solution contained 5 µl of 10x Taq amplification buffer, 1 µl of 10 mM dNTP mixture (0.2 mM each of dATP, dGTP, dCTP, and dTT), 0.5 µM concentrations of each primer, 2 mM of $MgCl_2$, 1.5 units Taq DNA polymerase (Invitrogen, Carlsbad, CA). The colony was added with a sterile wooden

stick, and sterile water was filled up to 50 µl. The PCR process was carried out in 35 cycles by heating and cooling the reaction tubes as follows: denaturing at 368 K for 15 s, annealing at 223 K for 30 s, elongation at 245 K for 32 s, and final extension for 30 min at 245 K. One of the colonies had a band appeared at 506 bp in 1.2% agarose gel electrophoresis (Figure 7).

The selected colony was grown in 10 ml LB culture medium with 100 µg/ml ampicillin and the presence of insert in the clone was verified with the restriction analysis using *Nde*I and *Xho*I. The presence of the dUTPase gene in the expression vector was also confirmed by DNA sequencing.

2.3e Host change

The expression of the target protein was carried out in *E. coli* BL21 Star (DE3). The expression vector was prepared from the previous host and *E. coli* BL21 Star (DE3) was transformed with the expression vector. The plasmid DNA was extracted from the transformants and subjected to the restriction analysis with *Nde*I and *Xho*I. The colonies that had correct vector DNA were stored as glycerol stocks at 193 K.

2.4 Protein expression and purification

Cells were precultured overnight in 300 ml LB medium at 310 K, which contained 100 µg ampicillin/ml and were vigorously shaken at 250 rpm. Two one-liter flasks containing 500 ml of LB medium and 100 µg ampicillin/ml were inoculated with 1/20 volume of preculture and incubated at 310 K, shaking vigorously at 250 rpm, until absorbance at 500 nm reaching 0.6 to 0.8. Isopropyl-ß-D-thiogalactopyranosid (IPTG) was added to a final concentration of 0.5 mM, and the solution was incubated for 3 hours

at 303 K with shaking at 240 rpm. Cells were harvested by centrifugation at 5,000 x g for 5 min and resuspended in 50 ml of NPI-10 buffer (50 mM Sodium phosphate pH 8.0, 0.3 M Sodium chloride, 10 mM Imidazole). The cells were disrupted by a sonicator (Misonix Inc., Farmingdale, NY) for 960 s at the maximum amplitude in 2 s pulses under chilled conditions on ice. The lysate was centrifuged at 12,000 x g for 1 hour to separate soluble from insoluble fractions. In order to purify the target protein, Ni-nitrilotriacetic acid (NTA) His-Bind resins (Novagen, San Diego, CA) equilibrated with NPI-10 buffer were added to the soluble fraction. After mixing for 3600 s at 277 K, the resin was loaded in a column and washed with 20 column volumes of NPI-20 (50 mM Sodium phosphate pH 8.0, 0.3 M Sodium chloride, 20 mM Imidazole). The dUTPase was eluted from the column by NPI-250 (50 mM Sodium phosphate pH 8.0, 0.3 M Sodium chloride, 250 mM Imidazole) in 7 column-volumes. The protein concentration was measured using UV spectrophotometer at the wavelength 280 nm. The spectrophotometer was first blanked with 100 µl of sterile water followed by the buffer and dUTPase protein. The protein concentration was calculated with the extinction coefficient of 9970 M^{-1} cm^{-1} at 280 nm, which corresponds to the absorbance of 0.515 at 0.1 % (or 1 g/L) of dUTPase. The calculation was done on ProtParam server (Gasteiger *et al.*, 2005). The resulted protein concentration at this point was 1.2 mg/ml.

In order to remove the His_6-tag, the recombinant protein was digested by thrombin. Prior to thrombin digestion, the buffer was exchanged to 1 x PBS (140 mM Sodium chloride, 2.7 mM Potasium chloride, 10.1 mM Sodium phosphate dibasic, 1.8 mM Potasium phosphate monobasic, pH 7.3) using sterilized dialysis bags soaked in 1mM EDTA solution. Thrombin reaction was carried out overnight at the room

temperature. The digested protein, shown in Figure 8, was 1 kDa smaller than the recombinant dUTPase protein as expected. In order to purify the dUTPase protein, imidazole (final concentration 10 mM) and Ni-NTA resin equilibrated with 1x PBS-10 buffer (140 mM Sodium chloride, 2.7 mM Potasium chloride, 10.1 mM Sodium phosphate dibasic, 1.8 mM Potasium phosphate monobasic pH 7.3, 10 mM Imidazole) was added to the solution. After mixing for 1 hour at 277 K, the resin was loaded in a column and eluted with 6 column-volumes of 1x PBS-10 buffer. The protein without the His_6-tag was again dialyzed in DBS buffer (50 mM Tris-HCl pH 7.4, 0.5 M Sodium chloride). Bengamidine Sepharose Fast Flow (Amersham Biosciences, Pittsburgh, PA) was used to remove thrombin from the protein solution. The solution was loaded in the column, pre-equilibrated with DBS buffer. The column was washed with 6 column-volumes of DBS buffer to elute the target protein (Figure 8).

The protein concentration was measured by using UV spectrophotometer to observe absorption at the wavelength 280 nm. The spectrophotometer was first blanked with 100 µl of sterile water followed by the buffer and dUTPase protein. In order to calculate the protein concentration, the difference in OD_{280} between the buffer and protein divided by the extinction coefficient after thrombin digestion was 9970 M^{-1} cm^{-1} (absorbance at 280 nm for 1 g/L dUTPase was at 0.559). The protein concentration calculated at this point was 0.5 mg/ml. The total yield was calculated by dividing the total amount of the purified protein with the total amount of the protein containing His-tag and thrombin site at the N-terminus of the dUTPase protein. The yield was about 73% and approximately 90 mg of the protein was purified from a 5 L culture. The protein was stored in ice bath in a refrigerator at 277 K.

2.5 Kinetic assay

To identify if the enzyme is active, I measured the activity by monitoring the change of pH due to the release of diphosphate with cresol red as an indicator on the UV spectrophotometer at 573 nm at pH 8.0 (Hidalgo-Zarco *et al.*, 2001, Nord *et al.*, 1997). The reaction mixture contained 25 µM cresol red, 100 µM dUTP, 5 mM $MgCl_2$, and 50 mM of sodium diphosphate pH 8.0. At 310 K, the enzyme reaction was started with adding 2 µl of 4.5 nM dUTPase (in 1 x DBS buffer) into 1 ml of the assay buffer. After 5 s, the decreasing of absorbance was monitored for 60 s with 3 s intervals. The enzymatic activity was measured with varying enzyme concentrations (Figure 9). The reaction without the enzyme resulted no significant change in the absorbance at 573 nm (Figure 9, diamonds, ♦). The reaction with the intact enzyme showed a significant decrease in the absorbance (Figure 9, stars, *). As the enzyme concentration was decreased by 10-fold, the slope of the change in absorbance was also decreased accordingly (Figure 9, solid squares, ■). The enzyme concentration is proportionally related to the decrease of the absorbance. This observation evidenced that the enzyme was active. The activity of the enzyme was also measured with varying the substrate (dUTP) concentration. The calculated K_m was 0.4 µM. Previously reported K_m values were at 0.2 µM from the *E. coli* dUTPase and 1.1 µM from the equine infectious anemia virus dUTPase (Nord *et al.*, 1997). The herpes simplex virus dUTPase had the K_m value of 0.3 µM (Wohlrab & Francke, 1980).

Table 2. Materials used in the gene construction.

Items	**Genotype or Description**
Host	
E. coli PIR1	F- ?*lac*169 *rpo*S(am) *rob*A1 *cre*C510 *hsd*R514 *end*A *rec*A1 *uid*A(?Mlu I)::pir-116
E. coli DH5α-T1^{R}	F^{-}φ80*lac*Z?M15 ?(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(rk^{-}, mk^{+}) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 *ton*A (confers resistance to phage T1)
E. coli BL21 Star DE3	F^{-} *omp*T *hsd*S$_{B}$ (r$_{B}^{-}$ m$_{B}^{-}$) *gal dcm rne*131 (DE3)
Expression Vector	
pET-15b	*lacI* pBR322 *bla* High copy vector for N-terminal His-tag hyper expression (Novagen, San Diego, CA)
pUNI51	*dUTPase* R6Kγ *kan* Size of the plasmid was 3087 bp dUTPase gene was cloned as pUNI51 in *E. coli* PIR1
Cloning site	
Nde I	CA^TATG
Xho I	C^TCGAG
Resulting protein sequence	MGSSHHHHHHSSGLVPRGSH-[dUTPase] Number of amino acids: 182 Molecular weight: 19357.8 Theoretical pI: 6.37 Absorption coefficients (1 mg ml^{-1} cm^{-1}; 280nm): 0.515 Thrombin cut---> GSH-[dUTPase] Number of amino acids: 169 Molecular weight: 17838.1 Theoretical pI: 5.53 Absorption coefficients (1 mg ml^{-1} cm^{-1}; 280nm): 0.559

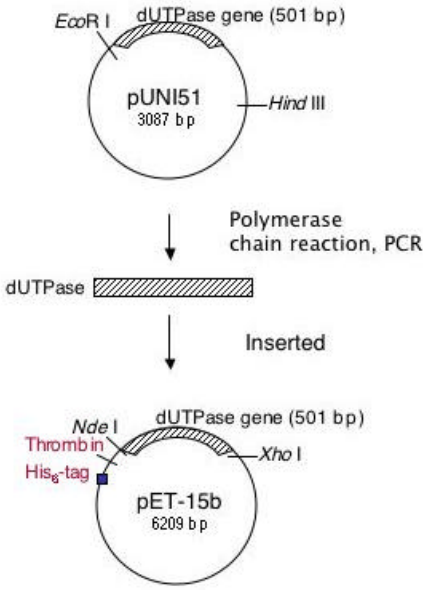
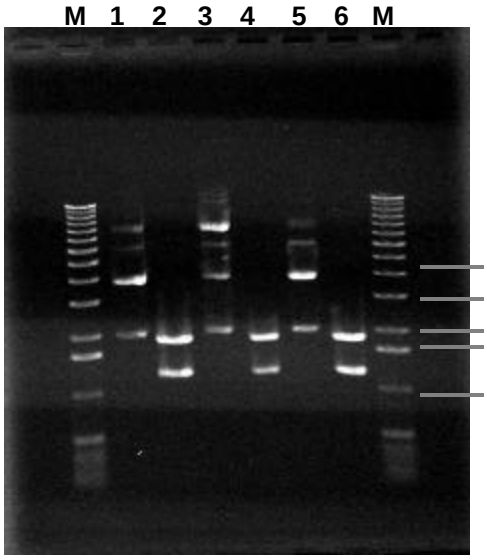
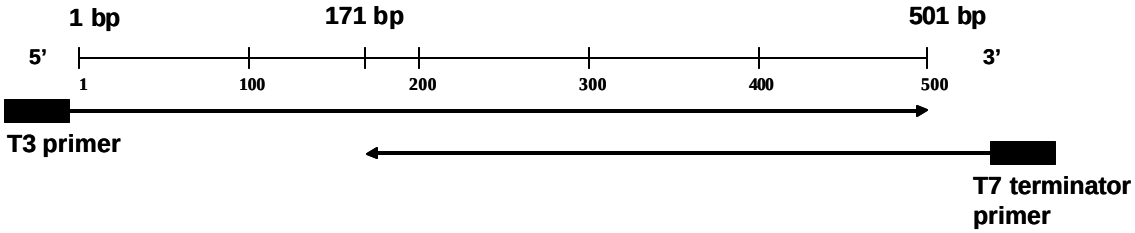


Figure 4. Cloning protocol of *Arabidopsis* dUTPase.



A



B

```
1    ATGGCTTGCG TAAACGAACC ATCACCTAAA CTCCAAAAGC TCGACCGAAA CGGAATCCAT    60
61   GGCGATTCAT CACCATCACC ATTCTTCAAA GTGAAGAAAC TCTCCGAGAA AGCTGTTATA   120
121  CCAACACGCG GCTCACCACT CTCCGCTGGA TACGATCTCT CAAGTGCGGT GGATTCAAAA   180
181  GTTCCAGCGA GAGGTAAAGC GCTTATCCCA ACGGATCTAA GTATCGCTGT TCCCGAAGGA   240
241  ACTTACGCGA GAATTGCGCC GAGATCTGGT TTGGCTTGGA AACATTCGAT TGATGTTGGT   300
301  GCTGGTGTGA TTGATGCTGA TTATAGAGGA CCTGTTGGTG TGATTTTGTT TAATCATTCT   360
361  GATGCTGATT TTGAGGTTAA GTTCGGTGAT CGAATCGCGC AGTTGATTAT CGAGAAGATT   420
421  GTGACTCCTG ATGTTGTTGA GGTTGATGAT TTGGATGAGA CTGTTCGTGG TGATGGTGGT   480
481  TTTGGTTCTA CTGGTGTCTA A                                             501
```

C

Figure 5. Confirmation of dUTPase gene presence in the plasmid, pUNI51, selected from colony isolation (A). "M" represents the DNA marker (1 Kb DNA ladder); 1, 3,

and 5 represent the plasmid before restriction enzyme reaction; and 2, 4, and 6 represent the two fragments, one of which contained the dUTPase gene (expected fragment lengths: 1848 bp and 1239 bp, with the latter containing the dUTPase gene). (B) DNA sequence strategy. The top line indicates dUTPase gene with marks at approximately 100 bp intervals. the T3 (5'-CAATTAACCCTCACTAAAGG) and T7 terminator (5'-GCTAGTTATTGCTCAGCGG) primers were provided by the DNA sequencing facility in the Center for Biotechnology at University of Nebraska – Lincoln. (C) The nucleotide sequence of dUTPase gene.

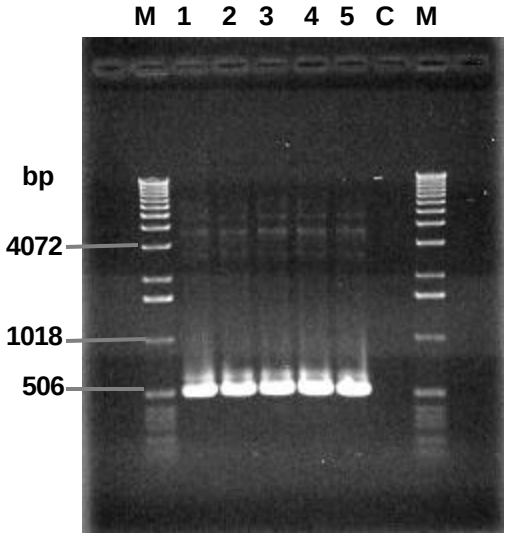


Figure 6. Polymerase chain reaction products of 503 bp. All the five bands (1-5) are from the same sample and the last lane (C) was control with no template DNA loaded. "M" represents the lane for the DNA marker.

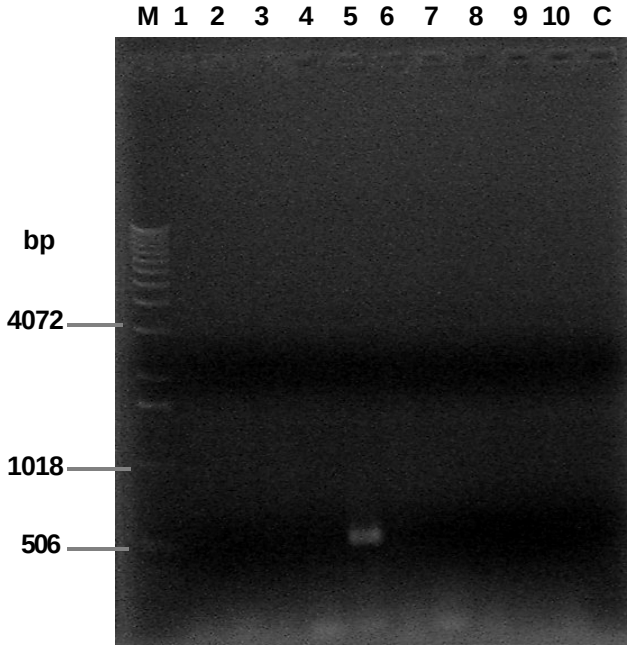


Figure 7. Polymerase chain reaction products from colony screening. "M" represents the DNA marker, samples 1-10 are from different colonies that appeared after transformation, and the last lane (C) was control with no template DNA loaded. The colony with dUTPase gene (sample number 5) had a band of about 503 bp.

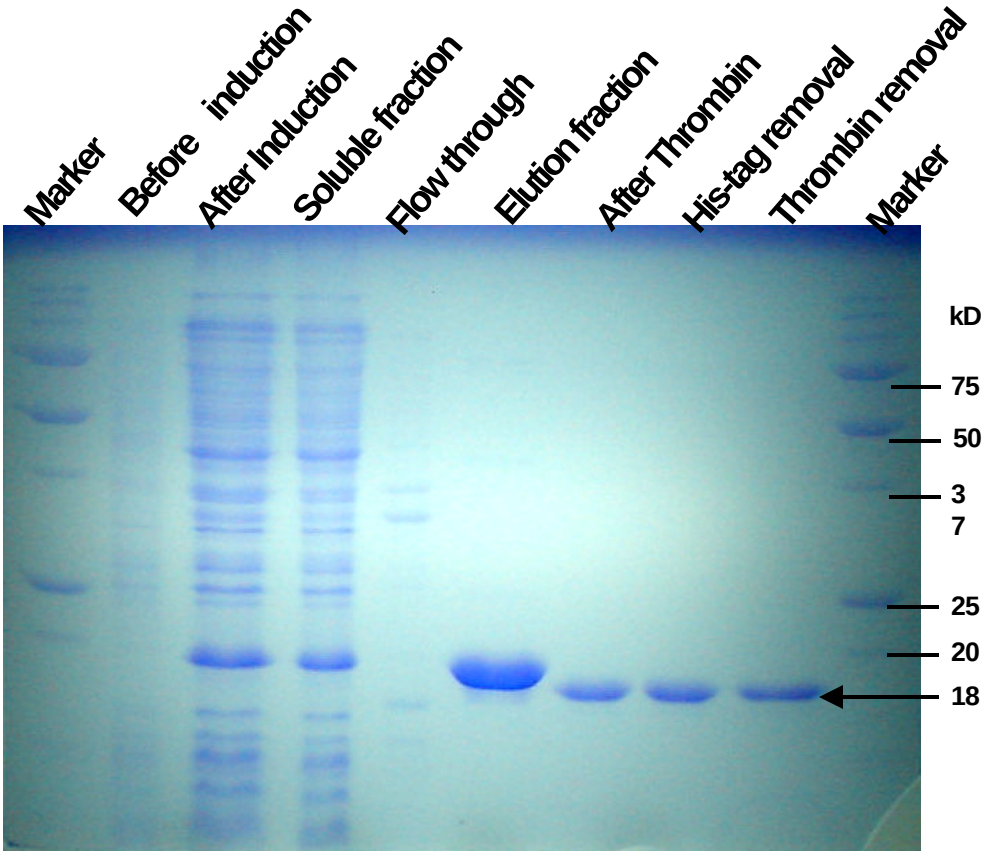


Figure 8. SDS-PAGE gel with samples loaded from the expression and purification. The molecular weight of the protein was about 1 kD less than the protein before thrombin digestion.

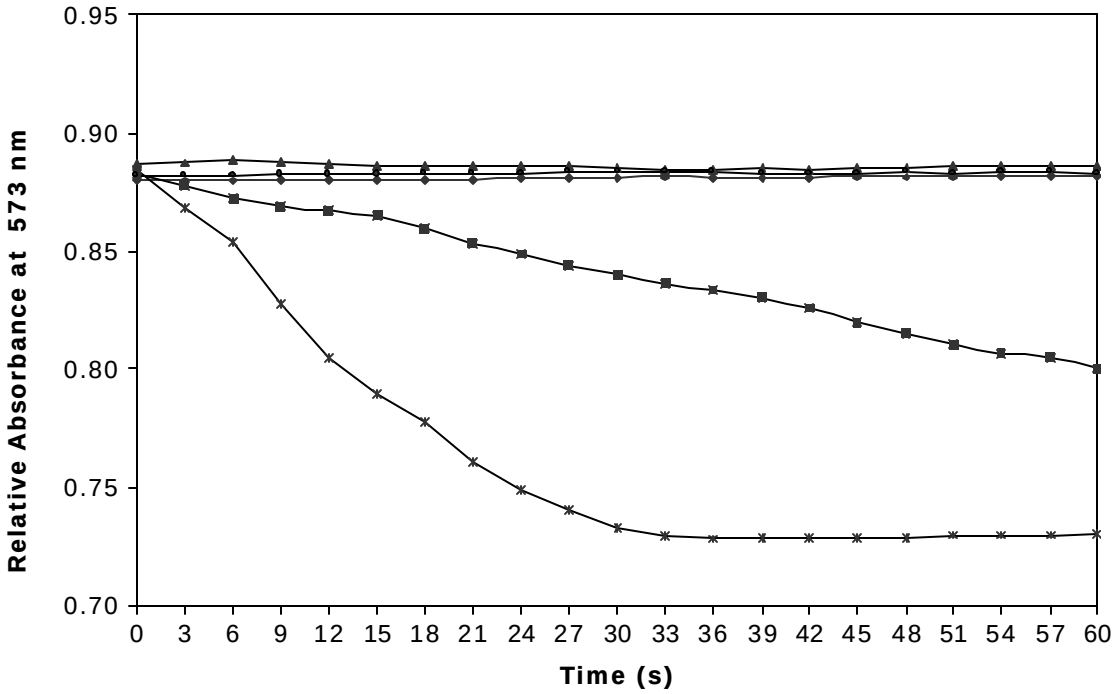


Figure 9. Enzymetic activity assay with varying dUTPase concentration. The symbol diamonds (♦) represent the 1 x DBS buffer, stars (*) represent the intact enzyme solution (the final concentration at 88 µM), solid squares (■) represent $x10^{-1}$ enzyme solution, solid triangles (▲) represent $x10^{-2}$ enzyme solution, and solid circles (●) represent $x10^{-3}$ enzyme solution.

Chapter 3

Crystallization of the *Arabidopsis* dUTPase

In this chapter, I will describe the crystallization of the dUTPase from *Arabidopsis thaliana*. First I report the method used to concentrate the protein. After the primary screening, the crystal quality was improved by refining the solvent conditions.

3.1 Primary screening

The protein was concentrated to 10 mg/ml using Microcon centrifugal filter (YM-30; Millipore Corp, Billerica, MA). The filter unit (Centriplus; Millipore Corp, Billerica, MA) was first washed with reverse osmosis water without touching the membrane with the pipette. The assembly was centrifuged with a fixed angle rotor at 10,000 x g for 10 min, and again centrifuged at 1,000 x g for 3 min by placing the sample reservoir upside down to remove water from the sample reservoir. The purified protein solution was loaded and repeatedly centrifuged at 10,000 x g for 10 min until the sample volume became less than 50 µl. In order to exchange the buffer solution of the protein from DBS buffer (50 mM Tris-HCl pH 7.4, 0.5 M Sodium chloride) to 50 mM Tris-HCl pH 7.4, the Tris buffer was added to the sample reservoir to have the final volume up to 500 µl. This solution was centrifuged at 10,000 x g until the sample volume became less than 20 µl. The volume of concentrated sample was calculated as follows: the volume of the solution was measured in the vial and subtracted from the initial volume of the solution in the reservoir. Once the volume of the sample became about 20 µl, the reservoir was inverted into the vial and spun at 1,000 x g for 3 min. The protein concentration was measured by

the UV spectrophotometer. The concentration of the dUTPase was adjusted to be 10 mg/ml by adding the appropriate amount of 50 mM Tris-HCl pH 7.4.

The initial crystallization conditions were obtained with a screening kit from Hampton Research (Aliso Viejo, CA) using the hanging drop-vapor diffusion method. Using the EasyXtal Tool (Valencia, CA), 1 µl of the screening solution was mixed with 1 µl of the protein solution (10 mg/ml protein and 50 mM 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, 1 drop with phase separation, and 3 drops with an indeterminate number of oil drops (Figure 10; Tables 3 and 4). Only one condition out of 50 (the condition number 4 in Table 3) had many small, plate-like, colorless protein crystals that were stacked upon one another, detected using a screen solution composed of 2 M ammonium sulfate and 0.1 M Tris-HCl, pH 8.5 (Figure 11). Observed crystals were confirmed to be protein crystals by stained with Izit dye (Hampton Research).

3.2 Refinement of crystallization condition

Using additional screening kits (Table 5, Hampton Research) and varying pHs, the primary crystallization condition was refined. After a period of one month, 36 out of 72 different conditions had crystals, 5 had precipitation, 7 had several small oil drops, 2 drops with phase separation, and 22 had clear drops (Table 6). Condition numbers 1 and 50 had crystals along with precipitation (Table 5). Most of the rod shaped crystals had another crystal protruding from it. Only one condition out of 72 had single rod shaped crystals that appeared after two weeks, and these crystals grew to approximate

dimensions of 04 × 01 × 01 mm within one month. The rod-shaped crystals were obtained using the following condition: pH 9.0 with 0.5 µl of 0.1 M taurine present in 2 µl of 1µl protein and 1 µl of 2 M ammonium sulfate and 0.1 M Tris-HCl (Figure 12).

Table 3. 50 initial crystallization conditions.

Reagent No.	Crystallization conditions
1	0.02 M Calcium chloride dihydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
2	0.4 M Potassium sodium tartrate tetrahydrate
3	0.4 M Ammonium phosphate monobasic
4 *	0.1 M Tris hydrochloride pH 8.5, 2.0 M Ammonium sulfate
5	0.2 M Sodium citrate tribasic dihydrate, 0.1 M HEPES sodium pH 7.5, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
6	0.2 M Magnesium chloride hexahydrate, 0.1 M Tris hydrochloride pH 8.5, 30% w/v Polyethylene glycol 4000
7	.0.1 M Sodium cacodylate pH 6.5, 1.4 M Sodium acetate trihydrate
8	0.2 M Sodium citrate tribasic dihydrate, 0.1 M Sodium cacodylate pH 6.5, 30% v/v 2-Propanol
9	0.2 M Ammonium acetate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 30% w/v Polyethylene glycol 4000
10	0.2 M Ammonium acetate, 0.1 M Sodium acetate trihydrate pH 4.6, 30% w/v Polyethylene glycol 4000
11	0.1 M Sodium citrate tribasic dihydrate pH 5.6, 1.0 M Ammonium phosphate monobasic
12	0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES sodium pH 7.5, 30% v/v 2-Propanol
13	0.2 M Sodium citrate tribasic dihydrate, 0.1 M Tris hydrochloride pH 8.5, 30% v/v Polyethylene glycol 400
14	0.2 M Calcium chloride dihydrate, 0.1 M HEPES sodium pH 7.5, 28% v/v Polyethylene glycol 400
15	0.2 M Ammonium sulfate, 0.1 M Sodium cacodylate pH 6.5, 30% w/v Polyethylene glycol 8000
16	0.1 M HEPES sodium pH 7.5, 1.5 M Lithium sulfate monohydrate
17	0.2 M Lithium sulfate monohydrate, 0.1 M Tris hydrochloride pH 8.5, 30% w/v Polyethylene glycol 4000
18	0.2 M Magnesium acetate tetrahydrate, 0.1 M Sodium cacodylate pH 6.5, 20% w/v Polyethylene glycol 8000
19	0.2 M Ammonium acetate, 0.1 M Tris hydrochloride pH 8.5, 30% v/v 2-Propanol
20	0.2 M Ammonium sulfate, 0.1 M Sodium acetate trihydrate pH 4.6, 25% w/v Polyethylene glycol 4000
21	0.2 M Magnesium acetate tetrahydrate, 0.1 M Sodium cacodylate pH 6.5, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
22	0.2 M Sodium acetate trihydrate, 0.1 M Tris hydrochloride pH 8.5, 30% w/v Polyethylene glycol 4000
23	0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES sodium pH 7.5, 30% v/v Polyethylene glycol 400
24	0.2 M Calcium chloride dihydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 20% v/v 2-Propanol
25	0.1 M Imidazole pH 6.5, 10 M Sodium acetate trihydrate
26	0.2 M Ammonium acetate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
27	0.2 M Sodium citrate tribasic dihydrate, 0.1 M HEPES sodium pH 7.5, 20% v/v 2-Propanol
28	0.2 M Sodium acetate trihydrate, 0.1 M Sodium cacodylate pH 6.5, 30% w/v Polyethylene glycol 8000
29	0.1 M HEPES sodium pH 7.5, 0.8 M Potassium sodium tartrate tetrahydrate
30	0.2 M Ammonium sulfate, 30% w/v Polyethylene glycol 8000
31	0.2 M Ammonium sulfate, 30% w/v Polyethylene glycol 4000
32	2.0 M Ammonium sulfate
33	4.0 M Sodium formate
34	0.1 M Sodium acetate trihydrate pH 4.6, 2.0 M Sodium formate

35	0.1 M HEPES sodium pH 7.5, 0.8 M Sodium phosphate monobasic monohydrate, 0.8 M Potassium phosphate monobasic
36	0.1 M Tris hydrochloride pH 8.5, 8% w/v Polyethylene glycol 8000
37	0.1 M Sodium acetate trihydrate pH 4.6, 8% w/v Polyethylene glycol 4000
38	0.1 M HEPES sodium pH 7.5, 1.4 M Sodium citrate tribasic dihydrate
39	0.1 M HEPES sodium pH 7.5, 2% v/v Polyethylene glycol 400, 2.0 M Ammonium sulfate
40	0.1 M Sodium citrate tribasic dihydrate pH 5.6, 20% v/v 2-Propanol, 20% w/v Polyethylene glycol 4000
41	0.1 M HEPES sodium pH 7.5, 10% v/v 2-Propanol, 20% w/v Polyethylene glycol 4000
42	0.05 M Potassium phosphate monobasic, 20% w/v Polyethylene glycol 8000
43	30% w/v Polyethylene glycol 1500
44	0.2 M Magnesium formate dihydrate
45	0.2 M Zinc acetate dihydrate, 0.1 M Sodium cacodylate pH 6.5, 18% w/v Polyethylene glycol 8000
46	0.2 M Calcium acetate hydrate, 0.1 M Sodium cacodylate pH 6.5 , 18% w/v Polyethylene glycol 8000
47	0.1 M Sodium acetate trihydrate pH 4.6, 2.0 M Ammonium sulfate
48	0.1 M Tris hydrochloride pH 8.5, 2.0 M Ammonium phosphate monobasic
49	1.0 M Lithium sulfate monohydrate, 2% w/v Polyethylene glycol 8000
50	0.5 M Lithium sulfate monohydrate, 15% w/v Polyethylene glycol 8000

* The condition where the protein crystals were observed.

Table 4. Summary of the primary screening with 50 conditions.

Drop characteristics	Number of conditions	Numbers corresponding conditions
Clear	23	2, 7, 19, 21, 23, 25, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 44, 47, 48, 49, 50
Precipitation	22	1, 3, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 22, 24, 26, 28, 45, 46
Phase	1	40
Oil	3	41, 42, 43
Crystal	1	4

Table 5. 72 additive screening conditions.

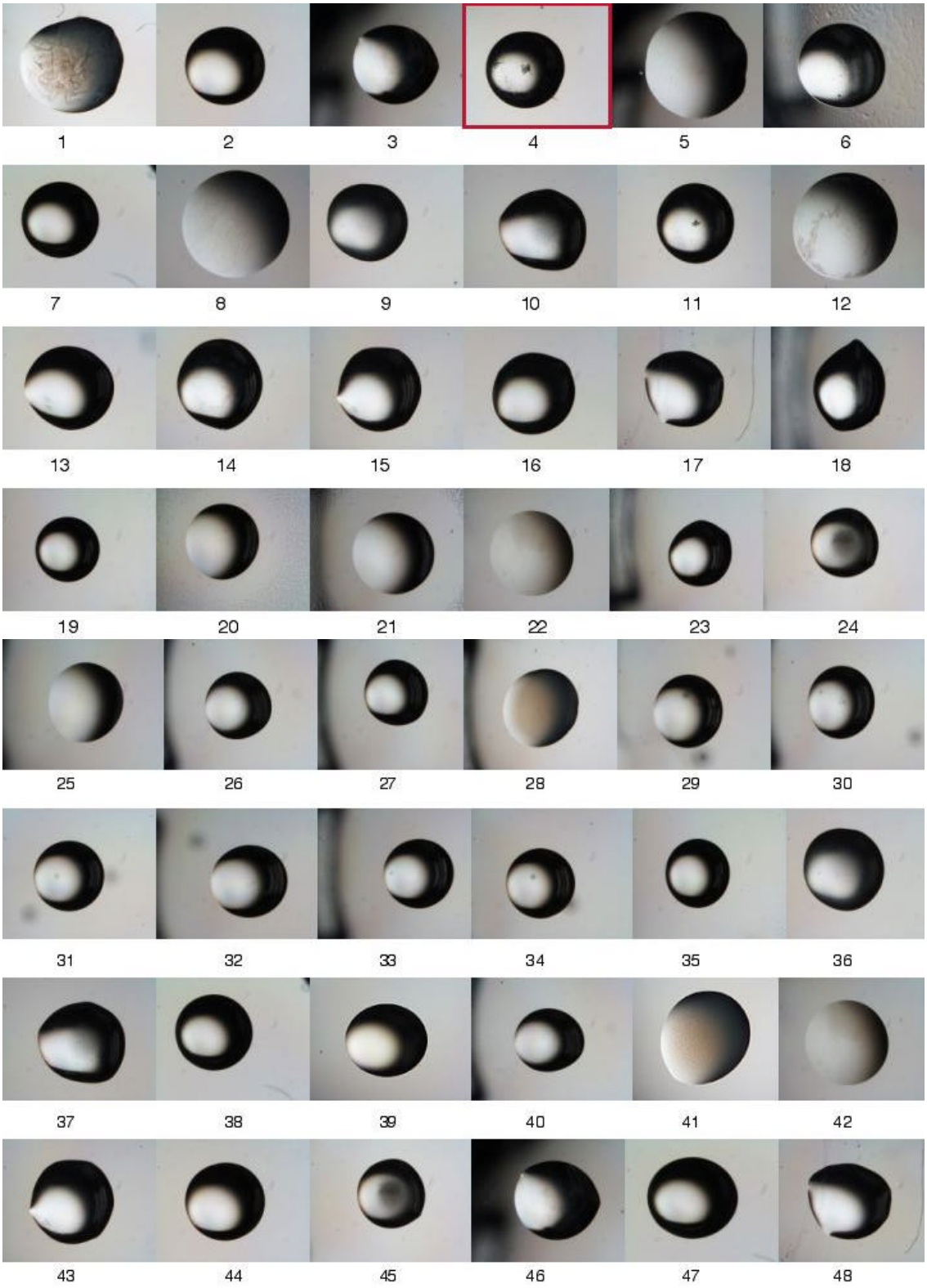
Reagent No.	Crystallization conditions
1 *	0.1 M Barium chloride dihydrate
2 *	0.1 M Cadmium chloride hydrate
3	0.1 M Calcium chloride dihydrate
4	0.1 M Cobalt(II) chloride hexahydrate
5	0.1 M Copper(II) chloride dihydrate
6 *	0.1 M Magnesium chloride hexahydrate
7	0.1 M Manganese(II) chloride tetrahydrate
8	0.1 M Strontium chloride hexahydrate
9	0.1 M Yttrium(III) chloride hexahydrate
10 *	0.1 M Zinc chloride
11 *	30% v/v Ethylene Glycol
12	30% v/v Glycerol anhyderous
13	30% w/v 1,6-Hexanediol
14 *	30% v/v 2-Methyl-2,4-pentanediol
15	50% w/v Polyethylene glycol 400
16 *	0.1 M Trimethylamine hydrochloride
17	1.0 M Guanidine hydrochloride
18 *	0.1 M Urea
19	15% w/v 1,2,3-Heptanetriol
20	20% w/v Benzamidine hydrochloride
21	30% v/v 1,4-Dioxane
22 *	30% v/v Ethanol
23 *	30% v/v 2-Propanol
24 *	30% v/v Methanol
25 *	1.0 M Sodium iodide
26	0.1 M L-Cysteine
27 *	0.1 M EDTA sodium salt
28	0.1 M NAD
29 *	0.1 M ATP disodium salt
30 *	30% w/v D(+)-Glucose
31	30% w/v D(+)-Sucrose
32	30% w/v Xylitol
33	0.1 M Spermidine
34 *	0.1 M Spermine tetrahydrochloride
35 *	30% w/v 6-Aminohexanoic acid
36	30% w/v 1,5-Diaminopentane dihydrochloride
37 *	30% w/v 1,6-Diaminohexane
38 *	30% w/v 1,8-Diaminooctane
39	1.0 M Glycine
40 *	0.3 M Glycyl-glycyl-glycine
41	0.1 M Hexammine cobalt(III) chloride
42 *	0.1 M Taurine
43 *	0.1 M Betaine hydrochloride
44 *	5% w/v Polyvinylpyrrolidone K15
45 *	3.0 M Non-Detergent Sulfo-betaine 195
46 *	2.0 M Non-Detergent Sulfo-betaine 201
47 *	0.1 M Phenol
48	30% v/v Dimethyl sulfoxide

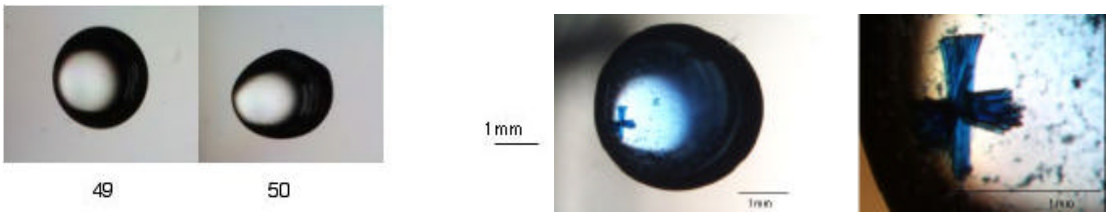
49	1.0 M Ammonium sulfate
50 *	1.0 M Cesium chloride
51 *	1.0 M Potassium chloride
52	1.0 M Lithium chloride
53 *	2.0 M Sodium chloride
54	0.5 M Sodium fl uoride
55	2.0 M Sodium thiocyanate
56	30% w/v Dextran sulfate sodium
57 *	50% v/v Jeffamine M-600 ® pH 7.0
58 *	40% v/v 2,5-Hexanediol
59	40% v/v (±)-1,3-Butanediol
60	40% v/v Polypropylene glycol P 400
61	40% v/v 1,4-Butanediol
62	40% v/v tert-Butanol
63 *	40% v/v 1,3-Propanediol
64 *	40% v/v Acetonitrile
65	40% v/v γ-Butyrolactone
66 *	40% v/v n-Propanol
67 *	5% v/v Ethyl acetate
68	40% v/v Acetone
69	0.25% v/v Dichloromethane
70	7% v/v n-Butanol
71 *	40% v/v 2,2,2-Trifl uoroethanol
72	0.1 M 1,4-Dithio-DL-threitol (DTT)

* The condition where the protein crystals were observed.

Table 6. Summary of the additive screening with 72 conditions.

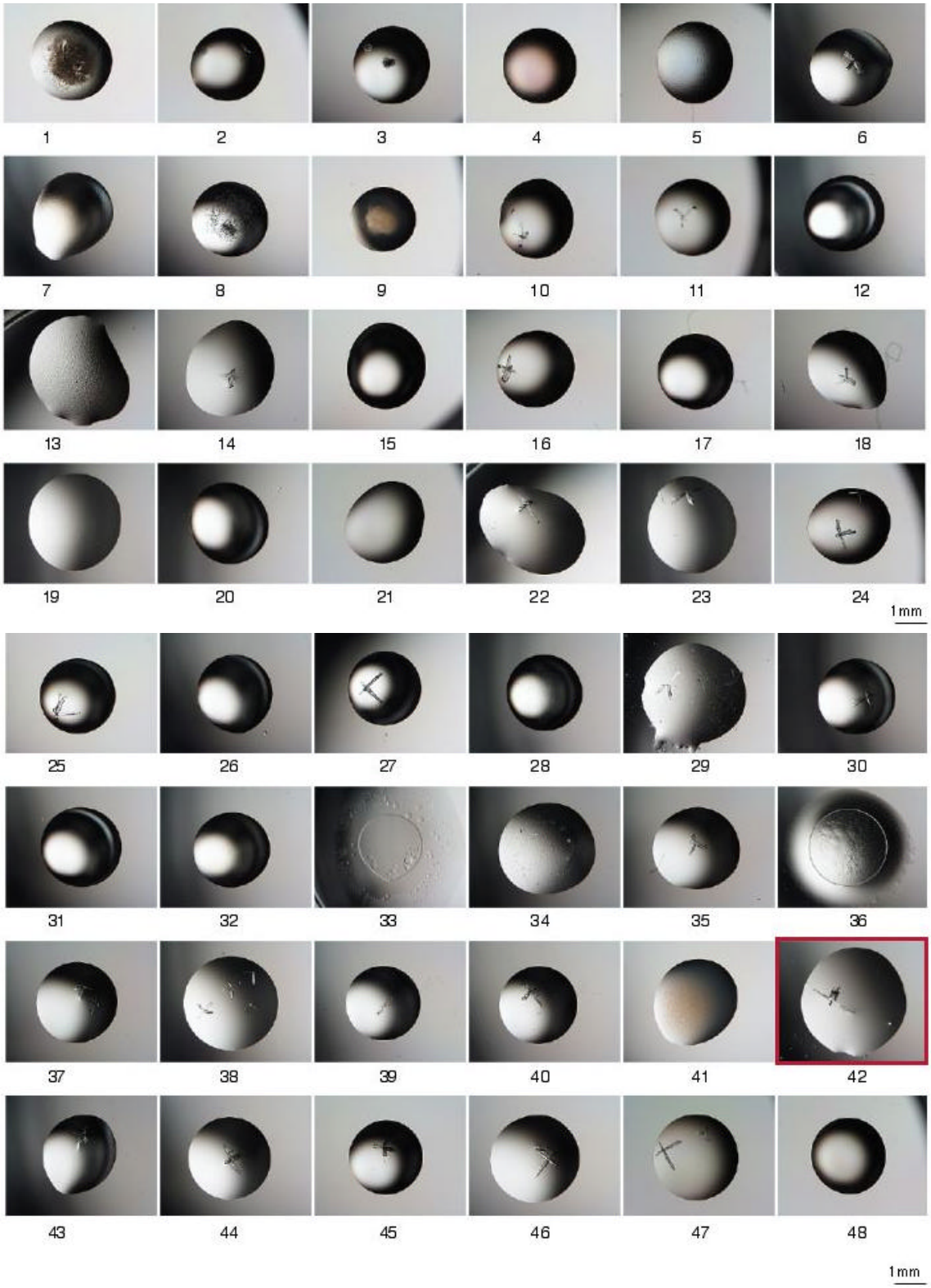
Drop characteristics	Number of conditions	Numbers corresponding conditions
Clear	22	12, 15, 17, 19, 20, 26, 28, 31, 39, 48, 49, 52, 54, 55, 56, 59, 60, 61, 62, 69, 70, 72
Precipitation	5	3, 7, 8, 9, 41
Phase	2	33, 36
Oil	7	4, 5, 13, 21, 32, 65, 68
Crystal	36	1, 2, 6, 10, 11, 14, 16, 18, 22, 23, 24, 25, 27, 29, 30, 34, 35, 37, 38, 40, 42, 43, 44, 45, 46, 47, 50, 51, 53, 57, 58, 63, 64, 66, 67, 71


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48



(A) **(B)**

Figure 10. (A) The results of the primary crystallization screening. The detailed description of 50 crystallization conditions is listed in Table 3. The condition that had crystals is highlighted with red box. (B) The crystal observed in the number 4 condition. The blue color is from the Izit dye.


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
1mm
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
1mm

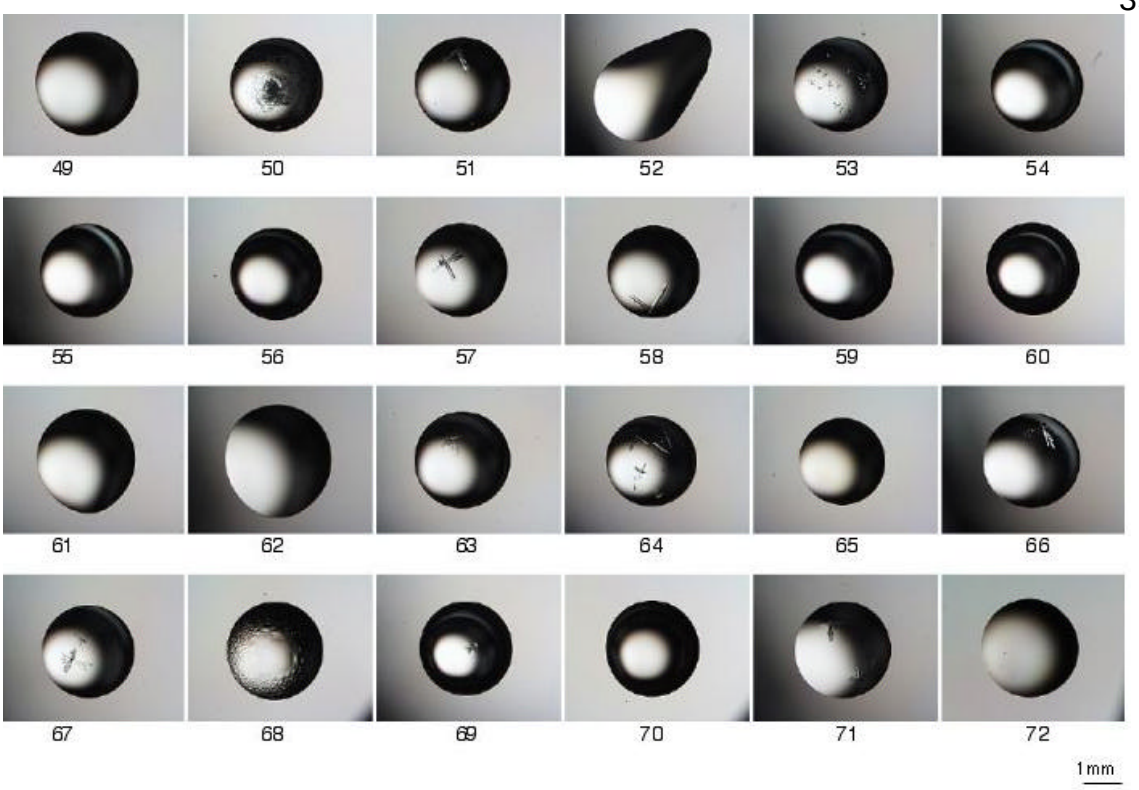


Figure 11. Refinement of the crystallization condition using the additive screening kit from Hampton Research. The details of the additives used are listed in Table 4. The condition that had single crystal is highlighted with red box.

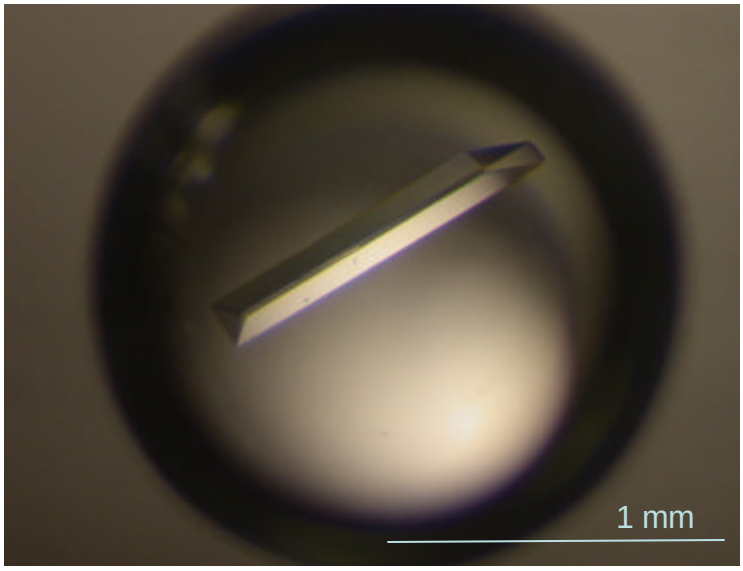


Figure 12. A crystal of the dUTPase from *Arabidopsis thaliana*. This crystal was obtained under the condition number 42 in Table 5 (see also Figure 11).

Chapter 4

Structural analysis of the *Arabidopsis* dUTPase

In this chapter, I will describe the crystallographic structural analysis on the dUTPase protein crystal. The molecular replacement method used is also described.

4.1 Data collection

The crystal shown in Figure 12 in the previous chapter was subjected to the x-ray diffraction experiment. For the mounting, a crystal was 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, in which final concentrations of trehalose and glycerol were 0.5 M and 1.2 M, respectively. After soaking for less than 10 s, the crystal was flash-frozen in a nitrogen stream at 93 K.

A complete data set was collected from a single crystal using a CuKα X-ray of the wavelength 1.542 Å with a generator operating at 40 kV and 20 mA. The native diffraction data consisted of 172 images in total, each exposed for 30 min with a 1.5º oscillation at a crystal-to-detector distance of 150 mm (Figure 13). The data were indexed, integrated, and scaled using *HKL*-2000 (Otwinowski & Minor, 1997).

The crystals diffracted up to a Bragg spacing of 2.2 Å. Observations of systematic absences along the three crystallographic axes were consistent with the space group $P2_12_12_1$ under the orthorhombic crystal system. Crystal parameters and data collection statistics are summarized in Table 7. The unit cell dimensions a, b, and c of the crystal in

the native crystal condition (without additional magnesium and inhibitors such as dUDP, dUMP, and analog dUTP) was 69.9, 70.9, and 75.5 Å, respectively. The number of observed reflections was 199,376 and the total number of unique reflections was 18,428. The ratio of the intensity (I) to σ(I) was 57.1 in overall (50.0 – 2.2 Å) and 37.9 in the outer shell 2.3 – 2.2 Å. This term represents an estimate of the accuracy of an individual measurement to the accuracy of the data set as a whole (McRee, 1993). In practice, if this ratio is more than approximately 5, then a reflection is often considered reliable. The percentage completeness in the data collection was 99.3% in overall and 98.4% in the outer shell. This term was the number of observed reflections divided by the number of possible reflections. The amount of completeness required in practice is above 80% (McRee, 1993). The R_{merge} of this data set was 7.5% in overall, and 12.3% in the outer shell (Table 7).

I tried to improve my crystal quality by using the long-term soaking method. After 136 days of soaking into high salt concentration (reservoir from the crystallized well), the native crystal was further soaked into 10 mM magnesium chloride for 4.5 hrs. A complete data set was then collected. The crystal lattice was similar to the data collected from the native crystal condition. The crystal diffracted up to a Bragg spacing of 2.0 Å. The number of observed reflections was 209,897 and the total number of unique reflections was 25,874. The ratio of the intensity (I) to σ(I) was 41.6 in overall (50.0 – 2.0 Å) and 18.5 in the outer shell (2.1 – 2.0 Å). The quality of the crystal has been improved slightly.

4.2 Molecular replacement

Assuming the presence of three monomeric subunits of the dUTPase from *A. thaliana* in the asymmetric unit of the orthorhombic crystal, the Matthews coefficient of 1.8 $Å^3$ Da^{-1} and a solvent content of 30% were calculated (Matthews, 1968). The structural analysis was performed using the molecular replacement method. Among structurally known dUTPases in the Protein Data Bank, that from human showed the highest sequence similarity with that of *A. thaliana* sharing the sequence identity at 56%. Therefore, I chose the human dUTPase as the structural template. A molecular model of 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 the chain A of the human dUTPase as a template (1Q5H; Leu3 – Phe115). The molecular replacement calculations were performed for three subunits using Molrep (Collaborative Computational Project Number 4, 1994). The rotation function and translation search were performed with all data in the resolution range of 50.0–2.2 Å. The correctly oriented model was subjected to the rigid-body refinement of the trimer using Refmac5 (CCP4). This model was used to calculate the $2\mathbf{F}_{obs} - \mathbf{F}_{calc}$ electron-density maps that were visually inspected with the interactive molecular graphics program, COOT (Emsley & Cowtan, 2004). The current structure, after repetitive calculations and visual inspections, contained 127, 126, and 127 visible residues for each of the three monomers, A, B, and C, respectively, out of 166 residues. The N and C-terminal residues were not visible in the electron density map. The refined structure yielded values of the crystallographic R factor was 0.18 and R_{free} was 0.25 for data from

50.0–2.2 Å (Table 10). Ideally the difference between R factor and R_{free} should be zero, and the difference 0.07 is considered to be acceptable.

There were approximately 25 amino acids in the N-terminus and 13 amino acids in the C-terminus of the protein molecule that were not visible. It may have caused a high R_{free} even after the addition of 212 water molecules. This structure has been deposited in the PDB with ID 2PC5. Previous studies have shown that a glycine-rich C-terminal tail of one subunit interacts with the active site of the second subunit of the molecule (Mol *et al.*, 1996). Three active site regions in the current structure were observed. This structure also had one possible magnesium site at GLU 138 of each chain.

The structure analyzed from the crystal soaked in magnesium chloride (PDB entry: 2P9O) had R factor = 19% and R_{free} = 23% (Table 10). This structure had two possible magnesium sites at GLU 138 of each chain and at ARG 64 of each chain.

Table 7. Data collection statistics.*

Parameters		Native	Mg^{2+}
PDB ID		2PC5	2P9O
System		Orthorhombic	Orthorhombic
Space Group		$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions / Å	a	69.9	69.9
	b	70.9	70.6
	c	75.5	75.0
Resolution range / Å		50.0 - 2.2	50.0 - 1.7
Resolution (scaling) range / Å		50.0 - 2.2 (2.3 - 2.2)	50.0 - 2.0 (2.1 - 2.0)
No. of observed reflections		199,376	209,897
No. of unique reflections used		18,428	25,874
I/s(I)		57.1 (37.9)	41.6 (18.5)
Completeness %		99.3 (98.4)	99.2 (98.5)
$^{\dagger}R_{merge}$ %		7.5 (12.3)	7.6 (17.0)

*Values in parentheses correspond to the highest resolution shell.

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

Table 8. Crystallographic refinement statistics.

Parameters		Native	Mg^{2+}
PDB ID		2PC5	2P9O
Resolution range / Å		50.0 – 2.2 (2.3 – 2.2)	50.0 – 2.0 (2.1 – 2.0)
$^{\dagger}R$ %		18	19
$^{*}R_{free}$ %		25	23
No. of nonhydrogen atoms		3,078	3,098
No. of water molecules		212	222
RMS deviations from ideal values			
Bond lengths / Å		0.02	0.02
Bond angles / º		2.07	2.00
Mean B value / $Å^2$		21.0	21.2
Visible residues:	Chain A	127 out of 166	128 out of 166
	Chain B	126 out of 166	126 out of 166
	Chain C	127 out of 166	127 out of 166

$^{\dagger}R = \sum||F_{obs}| - |F_{calc}|| / \sum|F_{obs}|$

* R_{free} values are collected for a randomly selected 5% of the data that was excluded from the refinement.

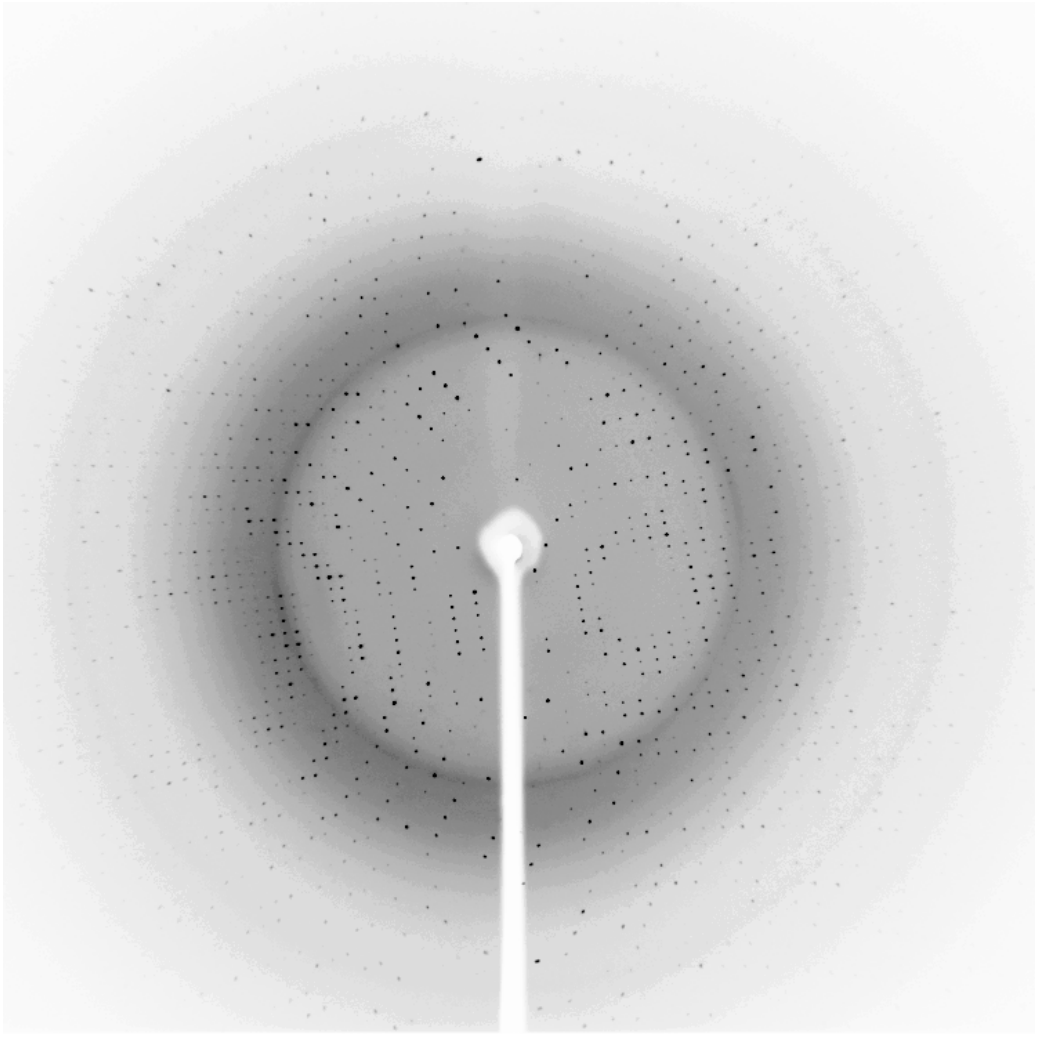


Figure 13. A diffraction image of a crystal shown in Figure 12 (Chapter 3). The image was recorded with Cu K_α radiation (λ = 1.542 Å) at 40 kV 20 mA. The crystal-to-detector distance was at 150 mm; and the imaging plate size was 300 x 300 mm. The resolution at the end of the imaging plate was 2.2 Å in Bragg spacing.

Chapter 5

Discussion

In this chapter, I will discuss in detail the structure of the dUTPase from *Arabidopsis*. I will compare the structures of *Arabidopsis* dUTPase to those of human (PDB: 1Q5H) and chlorella virus IL-3A (unpublished data from our lab). All the figures of the dUTPase structures were prepared using Pymol (DeLano, 2002).

5.1 Structure of *Arabidopsis* dUTPase

The asymmetric unit contains one molecule and each subunit comprises 166 amino acid residues. Residues from the chain A (26 to 153), chain B (25 to 150), and chain C (26 to 152) were traced on the electron density map. The density map shows that three subunits interact with each other to form one molecule. The residues from N-terminal (chain A: 1 to 25, B: 1 to 24, and C: 1 to 25) and C-terminal (chain A: 154 to 166, B: 151 to 166, and C: 153 to 166) were not visible in the electron density maps (Figure 14). Heavy electron density was observed at the center of the molecule near GLU 138 and at the top of the molecule near ARG 64 of the three chains (Figure 15). This implies the presence of two possible magnesium sites.

The molecule has a dull triangular pyramid shape with the base and the height of a triangle be 52 Å and 39 Å, respectively. The 3-fold axis is at the center of the trimer molecule. At the center of the trimer, there is a cylindrical empty space of 13 Å in diameter and 31 Å in length. There are ten β-strands (E1, 30 – 33; E2, 51 – 54; E3, 59 – 61; E4, 66 – 76; E5, 81 – 86; E6, 103 – 105; E7, 112 – 118; E8, 124 – 126; E9, 131 –

141; and E10, 145 – 147) and one α-helix (H1, 89 – 94) present in each subunit (Figure 16). The C-terminal region of the protein structure has no secondary structure.

There are three clefts present in the molecule. The entrance of each cleft has different sizes (Table 9 and Figure 17). The residues involved in the cleft are ARG 88, SER 89, and GLY 90 from one chain, and GLY 102, ASP 105, and GLY 113 from another chain. These residues were corresponding to the designated residues that have interaction with the substrate. The sizes of cleft and van der Waals volume (Zhao *et al.*, 2003) of the substrate dUTP (356 $Å^3$) were calculated (Table 9). The results show that the cleft size can accommodate the substrate.

5.2 Structural comparisons

The structures of dUTPases from human and chlorella virus IL-3A were compared to that from *Arabidopsis*. The primary structures of these dUTPases were 54% identical to the *Arabidopsis* dUTPase. These sequences were aligned with respect to the secondary structures of the *Arabidopsis* dUTPase (Figure 18). The area of the entrance of the active sites from *Arabidopsis*, human, and chlorella virus IL-3A were also measured using the three amino acids as the gate poles (Figure 17, Table 9). The entrance areas of the active sites involving the chains AB, BC, and CA were 70 $Å^2$, 71 $Å^2$, and 70 $Å^2$ in *Arabidopsis*; and 69 $Å^2$, 67 $Å^2$, and 69 $Å^2$ in human, respectively. The entrance area in the IL-3A dUTPase was the smallest, 65 $Å^2$ for all the three active sites. The enzymatic activity of the chlorella virus IL-3A dUTPase was confirmed (unpublished data from our lab).

The tertiary structures of dUTPases from human (PDB ID: 1Q5H) and chlorella virus IL-3A were then superimposed to that from *Arabidopsis* (PBD ID: 2P9O). The root mean square deviation (RMSD) of corresponding Cα atoms was calculated (Tables 10 and 11). The aligned 3D-structures of the entire dUTPases (including all three chains) from human and *Arabidopsis* had the RMSD of 0.8 Å. After excluding 14 residues from N-terminal and about 11 residues from C-terminal ends, the RMSD became 0.7 Å. The nine different combinations of chains from the two structures were superimposed (Figure 19). The RMSDs calculated from all combinations were all very similar (0.7 ± 0.04 Å). Thus, the chain A – chain A combinations was used as a representative for further comparison.

The selected structures, the dUTPase A chains from human and *Arabidopsis* were structurally aligned and the RMSD for each residue was calculated (Figure 20). Amino acid residues that have RMSDs greater than twice the average, 0.5 Å, were: GLY 44, LYS 60, VAL 61, PRO 62, PRO 78, GLU 79, GLY 80, TRP 93, LYS 94, HIS 119, SER 120, and ASP 121; and corresponding residues in the human dUTPase were: GLY 18, THR 34, ILE 35, PRO 36, PRO 52, SER 53, GLY 54, ALA 67, LYS 68, PHE 93, GLY 94, and LYS 95.

The tertiary structures of dUTPase in *Arabidopsis* (PDB ID: 2P9O) and chlorella virus IL-3A (the structural information was provided by our lab) were also compared in the same manner as *Arabidopsis* and human. The RMSDs were 0.8 Å from the entire structures and 0.6 Å excluding the tail regions (Table 11). The RMSDs calculated from nine different chain combinations were very similar (0.6 ± 0.02 Å). Again the chain A – chain A combination was selected as a representative of the group. The two selected

structures were structurally aligned and the RMSD for each residue was calculated (Figure 21). The amino acids that have RMSDs greater than twice the average, 0.5 Å, in *Arabidopsis* were: LEU 47, VAL 77, Pro 78, GLU 79, LYS 94, ALA 106, and ASP 107; and corresponding residues in IL-3A were: GLY 22, VAL 52, Pro 53, ASN 54, LYS 69, GLU 81, and ASP 82.

The RMSD from the dUTPase structures of human and chlorella virus IL-3A were 0.9 Å and 0.6 Å from the entire structure and after excluding the tail regions, respectively (Table 12). The RMSDs calculated from nine different chain combinations were all very similar (0.6 ± 0.02 Å), and the chain A – chain A combination was selected as a representative of the group. The two selected structures were structurally aligned and RMSD for each residue was calculated (Figure 22). The amino acids that have RMSDs greater than twice the average, 0.5 Å, in human were: THR 16, ARG 21, THR 34, ILE 35, PRO 36, ASP 46, GLY 64, ARG 83 and PHE 93; and corresponding residues in IL-3A were: MET 17, GLY 22, VAL 35, VAL 36, PRO 37, GLY 47, GLY 65, THR 84, and THR 94.

The amino acid difference at p1 was between a strong basic LYS 60 in *Arabidopsis* and an uncharged amino acid THR 34 in human. Alteration at p2 was between a strong acidic GLU 79 in *Arabidopsis* and an uncharged SER 53 in human (Figure 20). Both these residues are located at the surface of the dUTPase structure, which could affect the interaction between the dUTPase and the solvent, water. The change between amino acid residues TRP 93 in *Arabidopsis* and ALA 67 in human dUTPase was observed in the α-helix region (Figure 20, p3). Upon studying the dUTPase structure from *Arabidopsis*, I found that the TRP 93 on the α-helix in each

subunit interacts with LEU 68 in the β-strand E4 in another subunit (Figure 23). The p4 region had SER 120 and ASP 121 in *Arabidopsis*, and corresponding GLY 94 and LYS 95 in human dUTPase were located near the β-strand, E7 towards C-terminus (Figure 20).

Upon comparing the dUTPase structures from *Arabidopsis* and chlorella virus IL-3A, the amino acid differences in the two regions, p5 (GLU 79 in *Arabidopsis* and ASN 54 in IL-3A) and p6 (ALA 106 in *Arabidopsis* and GLU 81 in IL-3A) were observed (Figure 21). The acidic amino acid GLU 79 in *Arabidopsis* was likely to alter the solvent interaction. A negatively charged amino acid GLU 81 located in the cleft of the dUTPase structure from IL-3A (Figure 21, p6) would increase the probability of the substrate to retain at the entrance of cleft, as the size of clefts are small compared to that of *Arabidopsis* and human dUTPases (Table 9).

The comparison of 3D-structures between human and IL-3A dUTPases showed us significant differences in regions p7 (THR 16 in human and MET 17 in IL-3A), p8 (ARG 21 in human and GLY 22 in IL-3A), and p9 (THR 34 in human and VAL 35 in IL-3A) (Figure 22). Though no significant change in the folding of the structures were observed, the N-terminus of IL-3A dUTPase has more hydrophobic surface areas than that of human dUTPase.

In summary, the differences between the residues at the surface of the dUTPase protein structure from *Arabidopsis* are more hydrophilic than that of human dUTPase. On the other hand, surface amino acids of the IL-3A dUTPase are more hydrophobic compared to the human dUTPase. The comparison of the cleft sizes among *Arabidopsis*, human, and chlorella virus IL-3A dUTPases showed that the cleft size of the IL-3A

dUTPase is smaller than those found in the other two dUTPases. Note that with such a small cleft size, the enzymatic activity was still observed with the IL-3A dUTPase (unpublished data). The small entrance area of the active site and the elevation of the surface in the cleft of IL-3A would retain the substrate entrance.

Even though the dUTPases from the three organisms were structurally similar, small but important changes were observed. These changes are presumably due to the consequences of the functional adaptation.

In the future, to observe the invisible N and C-terminal residues the dUTPase structure needs to be solved after soaking the crystals in the inhibitors such as non-hydrolyzable deoxyuridine triphospate (dUTP) analog, deoxyuridine diphosphate (dUDP), and deoxyuridine monophosphate (dUMP).

Table 9. Entrance areas of active sites of the dUTPases from *Arabidopsis* (2P9O), human (1Q5H), and chlorella virus IL-3A (unpublished data from our lab).

Organism	Active site entrance area (Å^2)			Active site size (Å^3)		
	AB	**BC**	**CA**	**AB**	**BC**	**CA**
A. thaliana	70	71	70	426	433	424
Human	69	67	69	423	406	427
IL-3A	65	65	65	376	376	376

AB: chains A and B, BC: chains B and C, and CA: chain C and A.

Area = $bc[\sin(q)/2]$, where $q = \cos^{-1}\left[b^2 + c^2 - a^2/(2bc)\right]$, and a, b, and c each are the distances between the three points shown in Figure 17.

Volume = Bh, where B is the base area of the prism and h is the height.

Table 10. Root mean square deviations (RMSDs) of Cα-atoms in the superimposed dUTPase structures between human and *Arabidopsis*.

Comparison number	Arabidopsis		Human		Overall RMSD /Å
	Chain ID	Chain range	Chain ID	Chain range	
1	A	27 to 153	A	1 to 127	
	B	27 to 150	B	1 to 124	0.77*
	C	27 to 152	C	1 to 126	
2	A	41 to 140	A	15 to 114	
	B	41 to 140	B	15 to 114	0.68*
	C	41 to 140	C	15 to 114	
3	A	41 to 140	A	15 to 114	0.68
4	A	41 to 140	B	15 to 114	0.72
5	A	41 to 140	C	15 to 114	0.69
6	B	41 to 140	A	15 to 114	0.72
7	B	41 to 140	B	15 to 114	0.77
8	B	41 to 140	C	15 to 114	0.74
9	C	41 to 140	A	15 to 114	0.64
10	C	41 to 140	B	15 to 114	0.77
11	C	41 to 140	C	15 to 114	0.70

* Chain A (from 41 to 140 amino acid positions) from *Arabidopsis* was used as a reference for aligning the 3D-structures.

Table 11. Root mean square deviations (RMSDs) of Cα-atoms in the superimposed dUTPase structures between chlorella virus IL-3A and *Arabidopsis*.

Comparison number	Arabidopsis		IL3A		Overall RMSD /Å
	Chain ID	Chain range	Chain ID	Chain range	
1	A	27 to 150	A	2 to 125	
	B	27 to 150	B	2 to 125	0.75*
	C	27 to 150	C	2 to 125	
2	A	41 to 140	A	16 to 115	
	B	41 to 140	B	16 to 115	0.62*
	C	41 to 140	C	16 to 115	
3	A	41 to 140	A	16 to 115	0.62
4	A	41 to 140	B	16 to 115	0.62
5	A	41 to 140	C	16 to 115	0.62
6	B	41 to 140	A	16 to 115	0.63
7	B	41 to 140	B	16 to 115	0.63
8	B	41 to 140	C	16 to 115	0.63
9	C	41 to 140	A	16 to 115	0.66
10	C	41 to 140	B	16 to 115	0.66
11	C	41 to 140	C	16 to 115	0.66

* Chain A (from 41 to 140 amino acid positions) from *Arabidopsis* was used as a reference for aligning the 3D-structures.

Table 12. Root mean square deviations (RMSDs) of Cα-atoms in the superimposed dUTPase structures between human and chlorella virus IL-3A.

Comparison number	Human		IL3A		Overall RMSD /Å
	Chain ID	Chain range	Chain ID	Chain range	
1	A	1 to 124	A	3 to 125	
	B	1 to 124	B	3 to 125	0.92*
	C	1 to 124	C	3 to 125	
2	A	15 to 114	A	16 to 115	
	B	15 to 114	B	16 to 115	0.63*
	C	15 to 114	C	16 to 115	
3	A	15 to 114	A	16 to 115	0.63
4	A	15 to 114	B	16 to 115	0.63
5	A	15 to 114	C	16 to 115	0.63
6	B	15 to 114	A	16 to 115	0.63
7	B	15 to 114	B	16 to 115	0.67
8	B	15 to 114	C	16 to 115	0.67
9	C	15 to 114	A	16 to 115	0.68
10	C	15 to 114	B	16 to 115	0.68
11	C	15 to 114	C	16 to 115	0.68

* Chain A (from 41 to 140 amino acid positions) from *Arabidopsis* was used as a reference for aligning the 3D-structures.

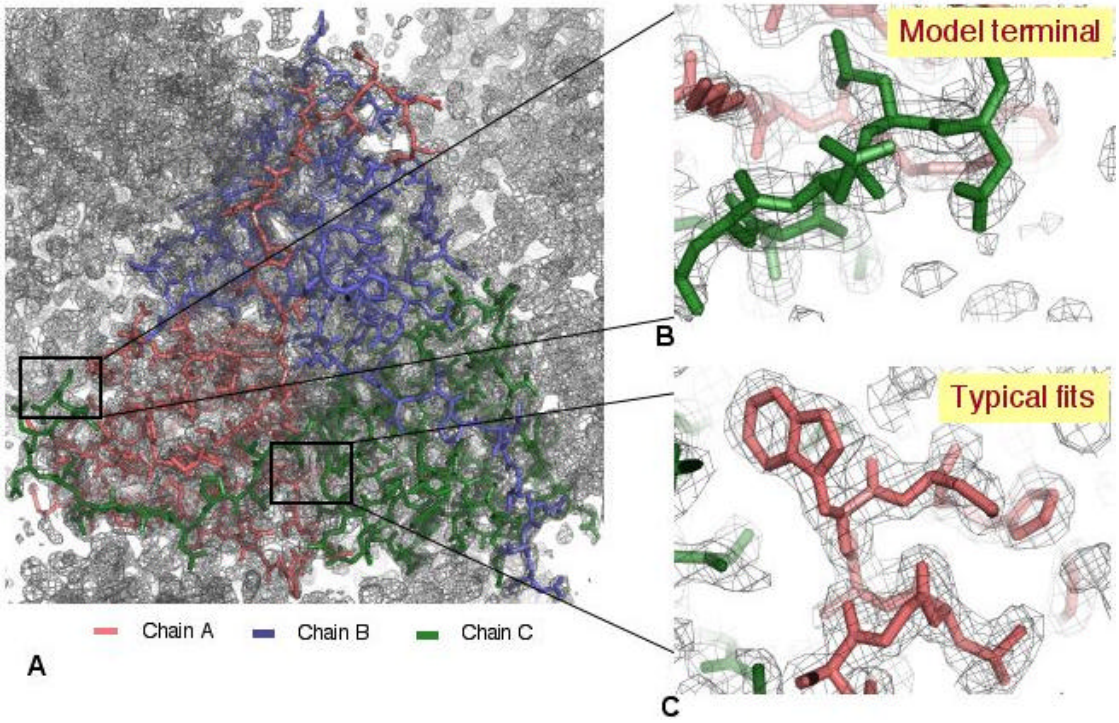


Figure 14. The 2**F**o – **F**c electron density map (A). The electron density at the terminal ends of a subunit (B) and the typical model fits (C).

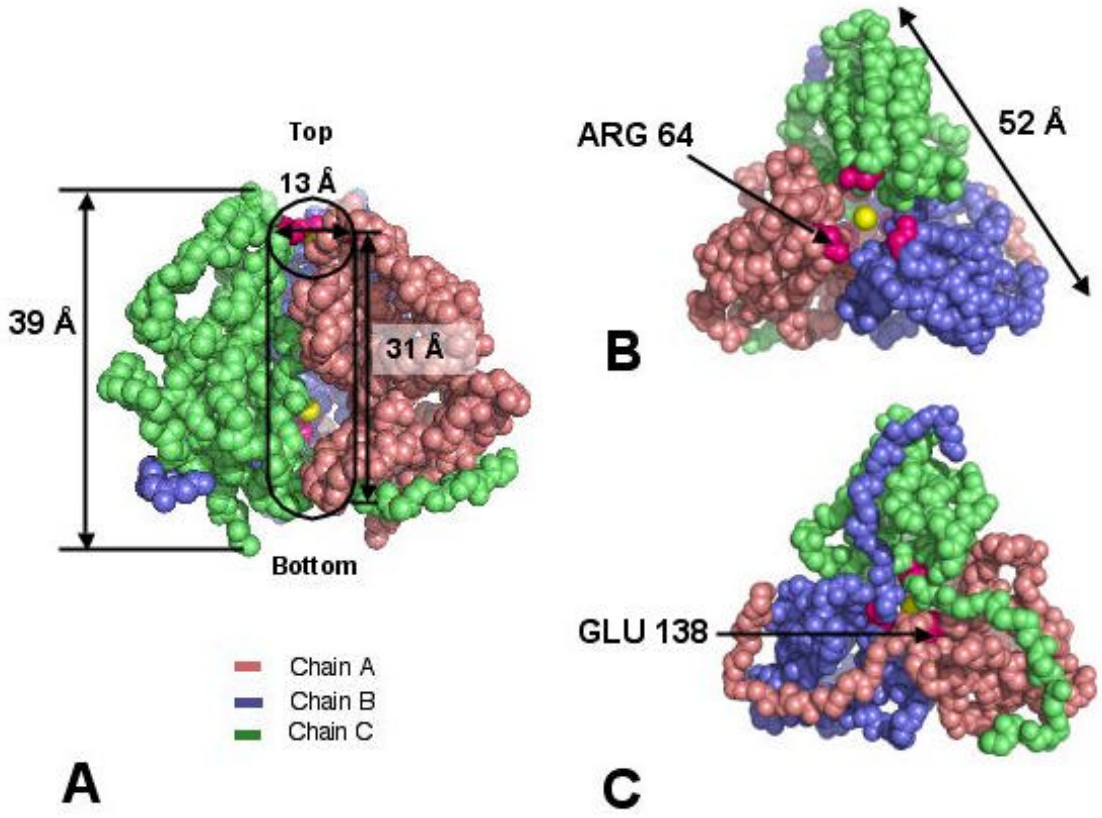


Figure 15. The dUTPase molecule. (A) A side view of the structure. The cylindrical empty space in the center of the molecule has diameter of 13 Å and length of 31 Å. The height of the molecule is 39 Å. The magnesium is colored in yellow and the two possible magnesium binding site residues are colored in magenta. Chains A, B, and C are indicated by red, blue, and green, respectively. (B) The top view of the molecule with the two magnesium ions, overlapped. The length of the molecule is 52 Å. (C) Bottom view of the molecule.

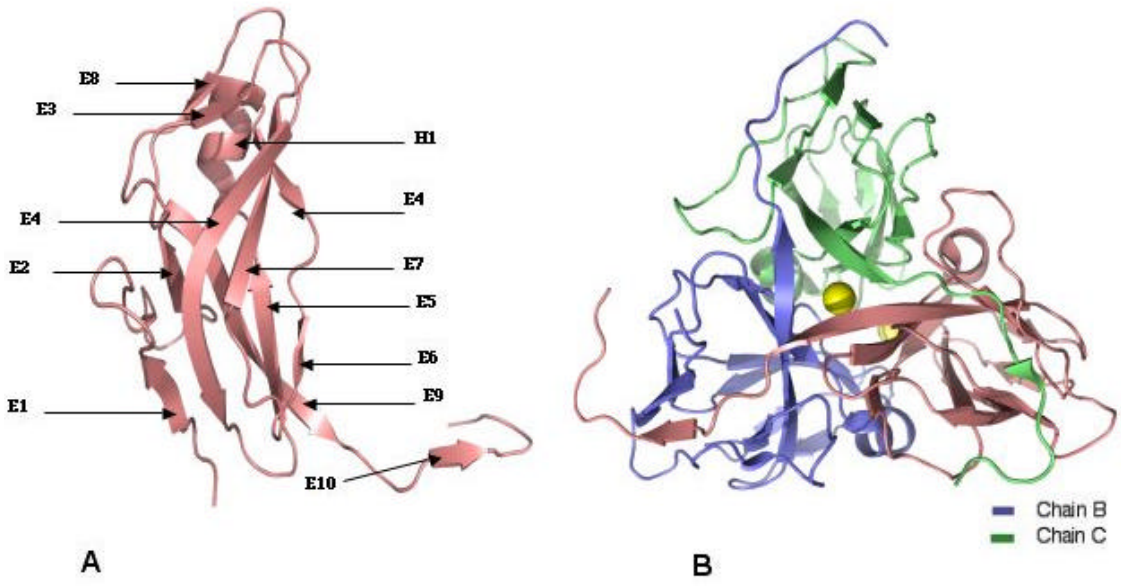


Figure 16. Structure of the *Arabidopsis* dUTPase. (A) A single subunit has ten β-strands (E1–E10) and one α-helix (H1). (B) A trimer molecule formed by the interaction of the three subunits. Three chains are differentiated by color; red for chain A, blue for chain B, and green for chain C.

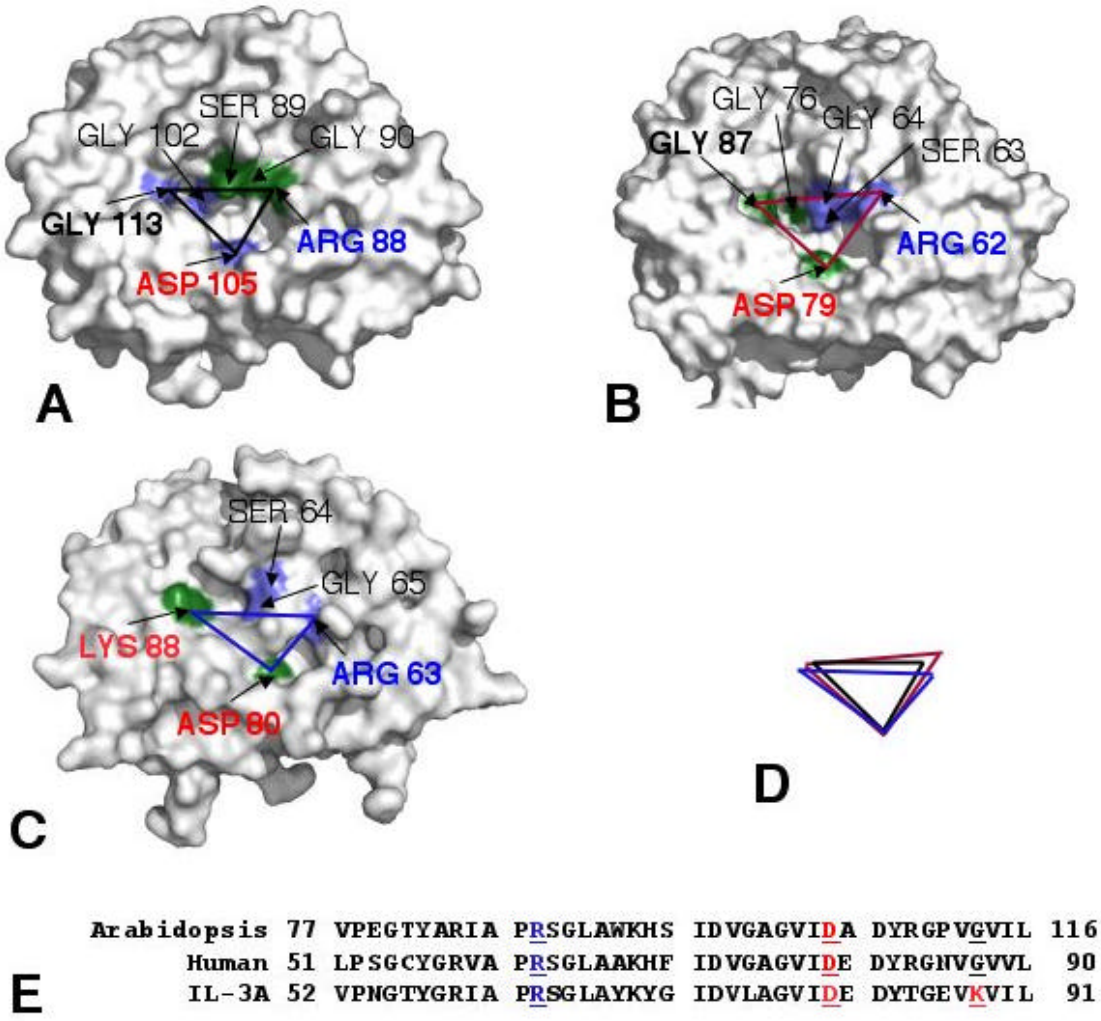


Figure 17. The surface view of the dUTPase of *Arabidopsis* (A), human (B), and IL-3A (C) illustrating one of the three active site regions. Chain A is not visible in this view. The active site residues from chain B is represented by blue and C is represented by green. Grey represents the surface view of the molecule as a whole. The triangles represent the area of the active site entrance (D), where black is from *Arabidopsis*, red from human, and blue from IL-3A. The alignment shows the three active site residues (underlined) used to calculate the area of the active site entrance (E). The acidic amino acids are colored red and basic are colored in blue.

```
                                                            E1                  E2    E3
                  -----------------------------eeee-----------------eeee----ee
Arabidopsis GSH1  MACVNEPSPKLQKLDRNGIHGDSSPSPFFKVKKLSEKAVIPTRGSPLSAGYDLSSAVDSK  60
      Human    1  --------------------------MQLRFARLSEHATAPTRGSARAAGYDLYSAYDYT  34
      IL-3A    1  ---------MASMTGGQQMGRGSEFMSSLLVKKLVESATTPMRGSEGAAGYDISSVEDVV  35

                           E4           E5      H1          E6          E7
                  e----eeeeeeeeeee----eeeeee--hhhhhh--------eee------eeeeeee--
Arabidopsis   61  VPARGKALIPTDLSIAVPEGTYARIAPRSGLAWKHSIDVGAGVIDADYRGPVGVILFNHS  120
      Human   35  IPPMEKAVVKTDIQIALPSGCYGRVAPRSGLAAKHFIDVGAGVIDEDYRGNVGVVLFNFG  94
      IL-3A   36  VPAMGRIAVSTGISIRVPNGTYGRIAPRSGLAYKYGIDVLAGVIDEDYTGEVKVILYNTT  95

                      E8        E9        E10
                  ---eee----eeeeeeeeeee---eee--------------------
Arabidopsis  121  DADFEVKFGDRIAQLIIEKIVTPDVVEVDDLDETVRGDGGFGSTGV-       166
      Human   95  KEKFEVKKGDRIAQLICERIFYPEIEEVQALDDTERGSGGFGSTGKN       141
      IL-3A   96  ERDYIIKKGDRIAQLILEQIVTPGVAVVLDLSDTARGSGGFGSTGI-       141
```

Figure 18. Sequence alignment of Arabidopsis (PDB ID: 2P9O), human (PDB ID: 1Q5H), and chlorella virus IL-3A structure provided by Dr. H. Moriyama's lab. The grey highlighted areas indicate the active site residues. The terminal regions colored in grey are not visible in the 3D-structure. E1-10 and H1 represent the β-strands (e) and the α-helix (h), respectively. The underlined area indicated the tag used during the protein expression.

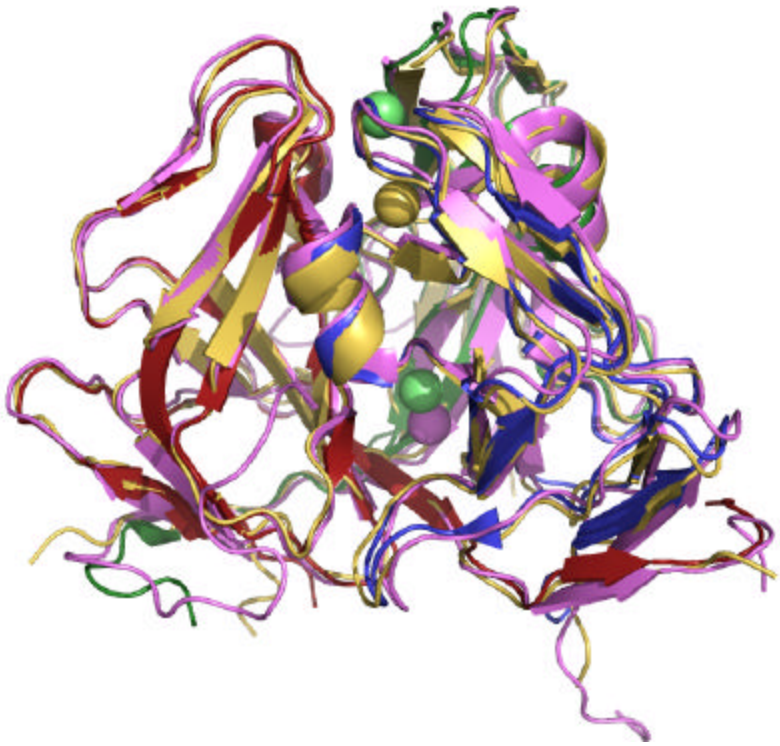
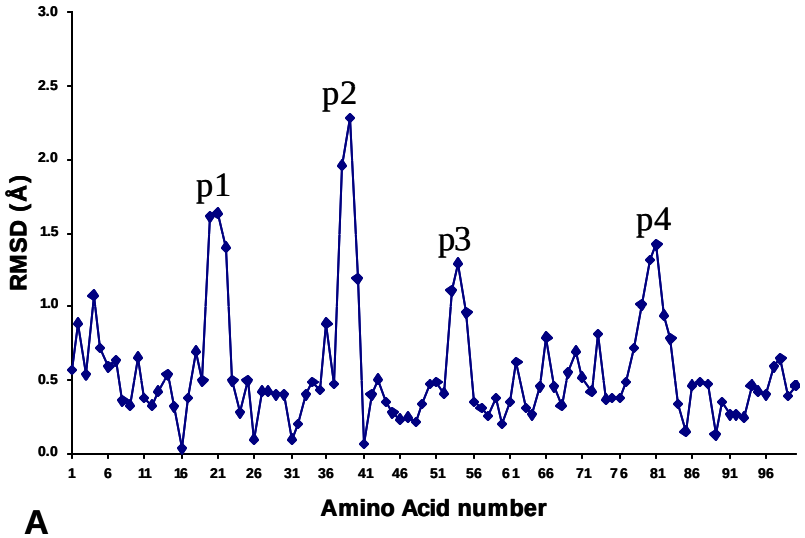
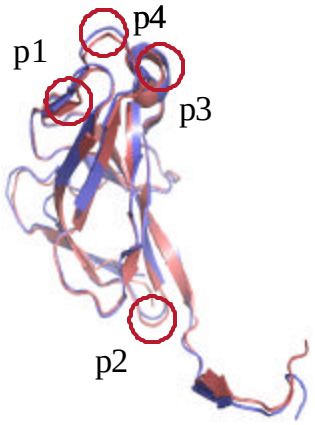


Figure 19. The superimposed dUTPase 3D-structures from *Arabidopsis*, human, and chlorella virus IL-3A. Chain A from *Arabidopsis* is represented by red, B by blue, and C by green. The two magnesium ions are represented in green spheres. The dUTPase structure and magnesium ion from human is indicated in violet, and that from IL-3A is in yellow.





	1	6	11	16	21	26	31	36	41	46	51	56	61	66	71	76	81	86	91	96
Arabidopsis	41	46	51	56	61	66	71	76	81	86	91	96	101	106	111	116	121	126	131	136
Human	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	105	110

C

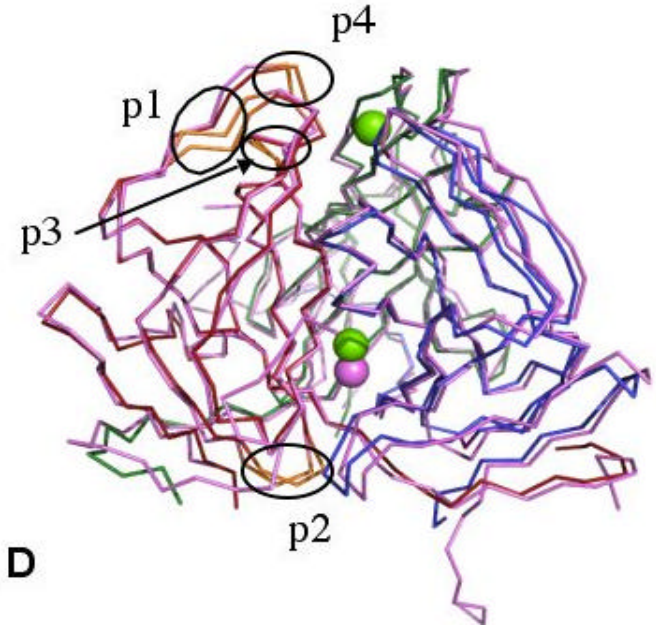
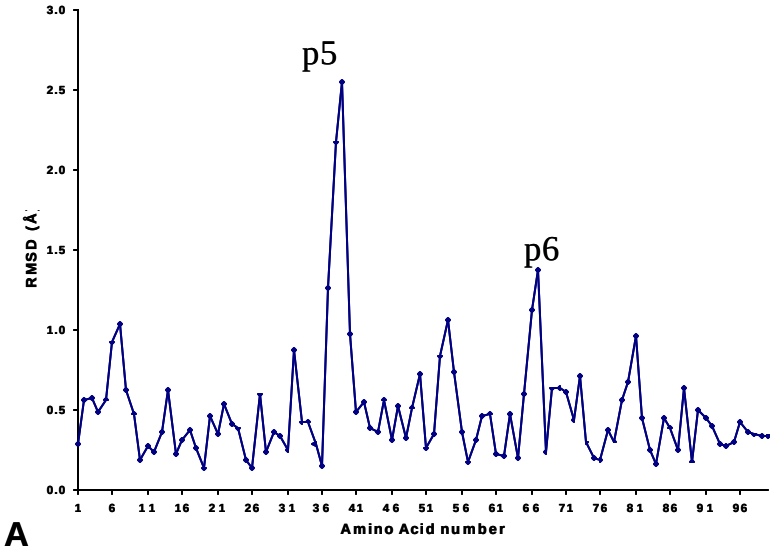
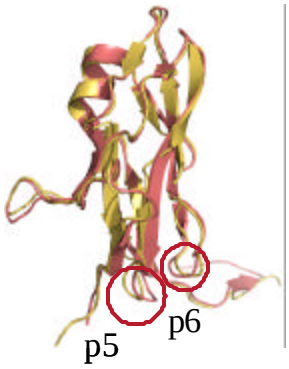


Figure 20. (A) The root mean square deviation (RMSD) of the distance calculated for the superimposed structures of dUTPases from *Arabidopsis* and human, chain A. The high peaks of RMSDs are indicated with p1, p2, p3, and p4. (B) Superimposed structures

of dUTPase chain As. Red represents the dUTPase from *Arabidopsis* and blue represents that of human. Amino acid differences in *Arabidopsis* / human are as follows: p1 (LYS 60 / THR 34, VAL 61 / ILE 35, and PRO 62 / PRO 36), p2 (PRO 78 / PRO 52, GLU 79 / SER 53, and GLY 80 / GLY 54), p3 (TRP 93 / ALA 67 and LYS 94 / LYS 68), and p4 (SER 120 / GLY 94 and ASP 121 / LYS 95). (C) Corresponding amino acid position numbers from the entire dUTPases in *Arabidopsis* and human. Amino acid sequences are presented in Figure 18. The first row shows the amino acid numbers used in A. (D) Superimposed structures of dUTPase trimer from *Arabidopsis* and human. The chains A , B, and C of *Arabidopsis* dUTPase are shown in red, blue, and green, respectively. Green spheres represent the two magnesium ions. The human dUTPase structure and the magnesium ion are shown with violet. The amino acids at the four peaks (p1, p2, p3, and p4) are represented in yellow.





	1	6	11	16	21	26	31	36	41	46	51	56	61	66	71	76	81	86	91	96
Arabidopsis	41	46	51	56	61	66	71	76	81	86	91	96	101	106	111	116	121	126	131	136
IL-3A	16	21	26	31	36	41	46	51	56	61	66	71	76	81	86	91	96	101	106	111

C

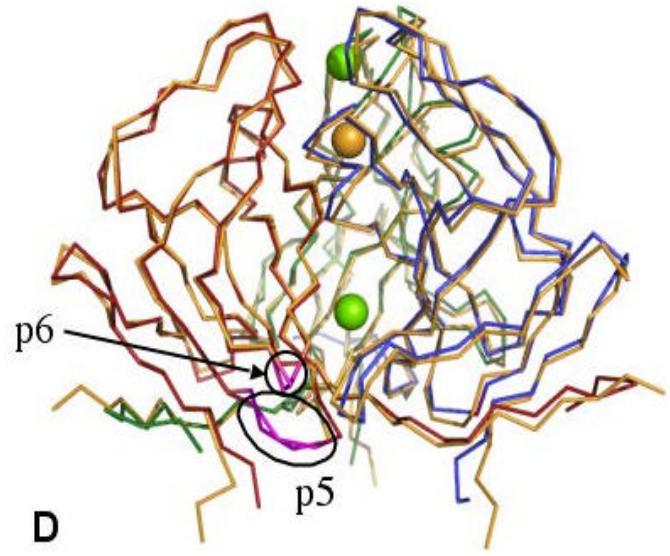
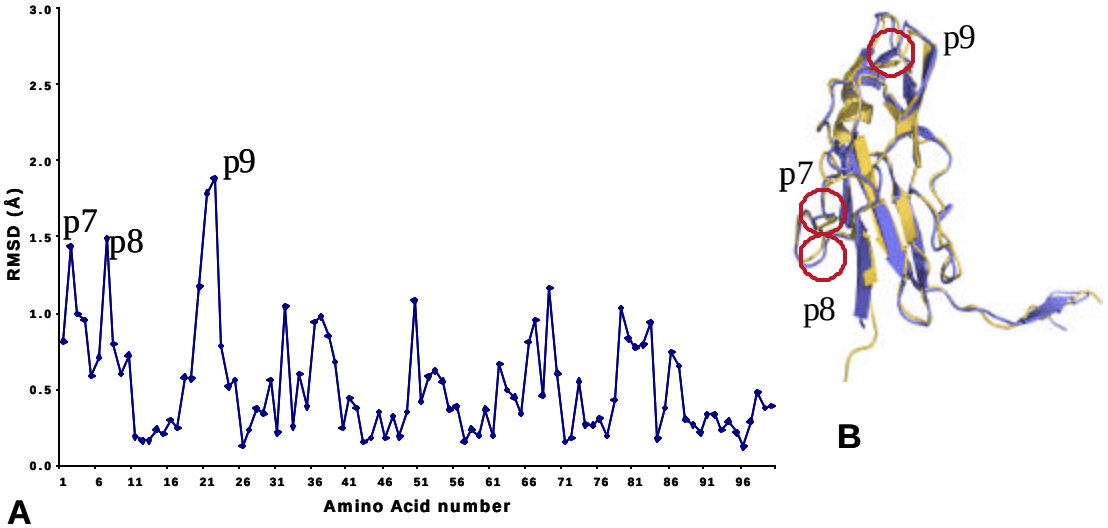


Figure 21. (A) The root mean square deviation (RMSD) of the distance calculated for the superimposed structures of dUTPases from *Arabidopsis* and chlorella virus IL-3A, chain A. The high peaks of RMSDs are indicated with p5 and p6. (B) Superimposed structures of dUTPase chain As. Red represents the dUTPase from *Arabidopsis* and

yellow represents that of IL-3A. Amino acid differences in *Arabidopsis* / IL-3A are as follows: p5 (VAL 77 / VAL 52, PRO 78 / PRO 53, and GLU 79 / ASN 54), and p6 (ALA 106 / GLU 81 and ASP 107 / ASP 82). (C) Corresponding amino acid position numbers from the entire dUTPases in *Arabidopsis* and IL-3A. Amino acid sequences are presented in Figure 18. The first row shows the amino acid numbers used in A. (D) Superimposed structures of dUTPase trimer from *Arabidopsis* and IL-3A. The chains A, B, and C of the *Arabidopsis* dUTPase are shown in red, blue, and green, respectively. Green spheres represent the two magnesium ions. The IL-3A dUTPase structure and the magnesium ion are represented in yellow. The amino acids at the two peaks (p5 and p6) are represented in magenta.



	1	6	11	16	21	26	31	36	41	46	51	56	61	66	71	76	81	86	91	96
Human	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	105	110
IL-3A	16	21	26	31	36	41	46	51	56	61	66	71	76	81	86	91	96	101	106	111

C

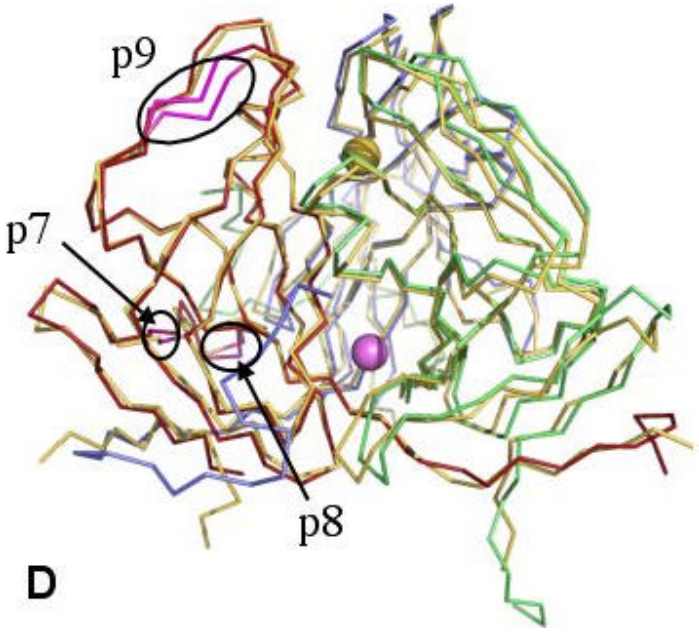


Figure 22. (A) The root mean square deviation (RMSD) of the distance calculated for the superimposed structures of dUTPases from human and chlorella virus IL-3A, chain

A. The high peaks of RMSDs are indicated with p7, p8, and p9. (B) Superimposed structures of dUTPase chain As. Blue represents the dUTPase from human and yellow represents that of IL-3A. Amino acid differences in human / IL-3A are as follows: p7 (THR 16 / MET 17), p8 (ARG 21 / GLY 22), and p9 (THR 34 / VAL 35, ILE 35 / VAL 36, and PRO 36 / PRO 37). (C) Corresponding amino acid position numbers from the entire dUTPases in human and IL-3A. Amino acid sequences are presented in Figure 18. The first row shows the amino acid numbers used in A. (D) Superimposed structures of dUTPase trimer from human and IL-3A. The chains A, B, and C of the human dUTPase are shown in red, blue, and green, respectively. Sphere in magenta represents the magnesium ion. The IL-3A dUTPase structure and the magnesium ion are represented in yellow. The amino acids at the three peaks (p7, p8 and p9) are represented in magenta.

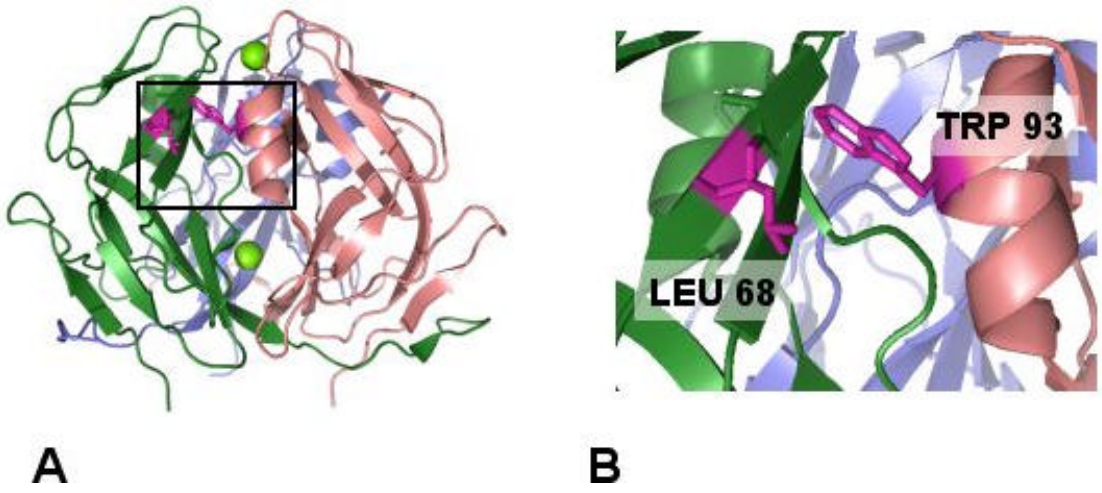


Figure 23. (A) The hydrophobic interaction between two neighboring subunits of the dUTPase trimer from *Arabidopsis*. Chain A from *Arabidopsis* is represented by red, B by blue, and C by green. The two magnesium ions are represented in green spheres. The amino acid residues involved in the hydrophobic interaction between LEU 68 in chain C and TRP 93 in chain A is indicated in magenta. (B) A closer view of the amino acids involved in the hydrophobic interaction between the two subunits.

References

Bajaj, M. & Moriyama, H. (2007). Acta Crystallographica Section F, 63, 409-411.

Cedergren-Zeppezauer, E. S., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. S. (1992). Nature 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). Journal of Molecular Biology 341, 503-517.

Collaborative Computational Project Number 4 (1994). Acta Crystallogr D Biol Crystallogr 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.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System.

Dieffenbach, C. W. & Dveksler, G. S. (1995). PCR primer: a laboratory manual., Second ed. New York: Cold Spring Harbor Laboratory Press.

Emsley, P. & Cowtan, K. (2004). Acta Crystallogr D Biol Crystallogr 60, 2126-2132.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press 571-607.

Gray, W. M., Ostin, A., Sandberg, G., Romano, C. P. & Estelle, M. (1998). Proc Natl Acad Sci U S A 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.

Hidalgo-Zarco, F., Camacho, A. G., Bernier-Villamor, V., Nord, J., Ruiz-PpRez, L. M. & * RQli Lez-Pacanowska, D. (2001). Protein Science 10, 1426–1433.

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). Annual review of biophysics and biomolecular structure 29, 291-325.

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

McRee, D. E. (1993). Practical Protein Crystallography. San Diego, CA: Academic Press.

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). Journal of Molecular Biology 342, 1583–1597.

Nord, J., Larsson, G., Kvassman, J. O., Rosengren, A. M. & Nyman, P. O. (1997). FEBS Lett 414, 271-274.

Otwinowski, Z. & Minor, W. (1997). Methods Enzymol. 276, 307–326.

Prasad, G. S., Stura, E. A., Elder, J. H. & Stout, C. D. (2000). Acta Crystallographica Section D Biological Crystallography 56, 1100-1109.

Pri-Hadash, A., Hareven, D. & Lifschitz, E. (1992). Plant Cell 4, 149-159.

Rhee, S. Y., Beavis, W., Berardini, T. Z., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M., Miller, N., Mueller, L. A., Mundodi, S., Reiser, L., Tacklind, J., Weems, D. C., Wu, Y., Xu, I., Yoo, D., Yoon, J. & Zhang, P. (2003). Nucleic acids research 31, 224-228.

Sambrook, J. & Russell, D. (2001). Molecular Cloning: A Laboratory Manual, 3rd Labmn edition ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Wohlrab, F. & Francke, B. (1980). Proc Natl Acad Sci U S A 77, 1872-1876.

Van Etten, J. L., Burbank, D. E., Joshi, J. & Meints, R. H. (1984). Virology 134, 443-449.

Zhang, Y., Moriyama, H., Homma, K. & Van Etten, J. L. (2005). J Virol 79, 9945-9953.

Zhao, Y. H., Abraham, M. H. & Zissimos, A. M. (2003). Journal of Organic Chemistry 68, 7368-7373.

Appendix A

Reprint of the published material.

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

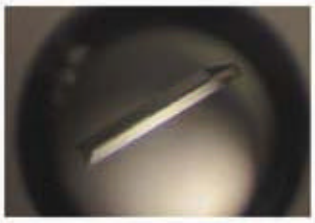
Received 23 February 2007
Accepted 31 March 2007

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 pUN151 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_{merge}† (%)	7.5 (12.3)

† $R_{merge} = \sum|I_{obs} - \langle I\rangle|/\sum I_{obs}$, where I_{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 $2F_{obs} - F_{calc}$ 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_{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_{free} even after the addition of 210 water molecules. Previous studies have shown that the glycine-rich

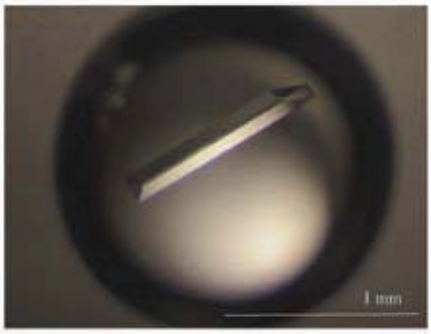


Figure 1
A crystal of dUTPase from *A. thaliana*.

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.