doi:10.1016/j.jmb.2004.06.028 J. Mol. Biol. (2004) 341, 503–517

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Crystal Structure of the *Mycobacterium tuberculosis* dUTPase: Insights into the Catalytic Mechanism

Sum Chan[1], Brent Segelke[2], Timothy Lekin[2], Heike Krupka[2] Uhn Soo Cho[3], Min-young Kim[3], Minyoung So[3], Chang-Yub Kim[3] Cleo M. Naranjo[3], Yvonne C. Rogers[3], Min S. Park[3], Geoffrey S. Waldo[3] Inna Pashkov[1], Duilio Cascio[1], Jeanne L. Perry[4] and Michael R. Sawaya[1]*

[1]*UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, 206 Boyer Hall, Box 951570, Los Angeles CA 90095-1570, USA*

[2]*Lawrence Livermore National Laboratory, Livermore, CA 94551, USA*

[3]*Bioscience Division, Mail Stop M888, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

[4]*Department of Molecular, Cell and Developmental Biology Molecular Biology Institute University of California, Los Angeles, Los Angeles, CA 90095-1570, USA*

**Corresponding author*

The structure of *Mycobacterium tuberculosis* dUTP nucleotidohydrolase (dUTPase) has been determined at 1.3 Å resolution in complex with magnesium ion and the non-hydrolyzable substrate analog, α,β-imido dUTP. dUTPase is an enzyme essential for depleting potentially toxic concentrations of dUTP in the cell. Given the importance of its biological role, it has been proposed that inhibiting *M. tuberculosis* dUTPase might be an effective means to treat tuberculosis infection in humans. The crystal structure presented here offers some insight into the potential for designing a specific inhibitor of the *M. tuberculosis* dUTPase enzyme. The structure also offers new insights into the mechanism of dUTP hydrolysis by providing an accurate representation of the enzyme–substrate complex in which both the metal ion and dUTP analog are included. The structure suggests that inclusion of a magnesium ion is important for stabilizing the position of the α-phosphorus for an in-line nucleophilic attack. In the absence of magnesium, the α-phosphate of dUTP can have either of the two positions which differ by 4.5 Å. A transiently ordered C-terminal loop further assists catalysis by shielding the general base, Asp83, from solvent thus elevating its pK_a so that it might in turn activate a tightly bound water molecule for nucleophilic attack. The metal ion coordinates α, β, and γ phosphate groups with tridentate geometry identical with that observed in the crystal structure of DNA polymerase β complexed with magnesium and dNTP analog, revealing some common features in catalytic mechanism.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: dUTPase; tuberculosis; drug design; nucleotidohydrolase; mechanism

Introduction

Depletion of cellular levels of 2′-deoxyuridine 5′-triphosphate (dUTP) is crucial for maintaining genome integrity. Because dUTP differs from dTTP by only a single methyl group it can be readily accepted as a substrate by DNA polymerases, leading to the substitution of uracil for thymidine in the genetic material. Incorporation of uracil into DNA would overwhelm the DNA repair system, leading to multiple DNA strand breaks and eventually, cell death. Nearly all prokaryotes, eukaryotes and a majority of viruses encode an enzyme called dUTPase to deplete the potentially

Abbreviations used: EIAV, equine infectious anemia virus; FIV, feline immunodeficiency virus; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; IPTG, isopropyl-thio-β-D-galactoside; dUTP, 2′-deoxyuridine 5′-triphosphate; dUMP, 2′-deoxyuridine 5′-monophosphate; dUDP, 2′-deoxyuridine 5′-diphosphate; dUTPαS, 2′-deoxyuridine-5′-(α-thio)-triphosphate.

E-mail address of the corresponding author: sawaya@mbi.ucla.edu

0022-2836/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.

toxic concentration of dUTP in the cell.[1] dUTPases are primarily homotrimeric and catalyze the metal ion dependent hydrolysis of dUTP to 2′-deoxyuridine 5′-monophosphate (dUMP) and pyrophosphate. Clearly, inhibition of dUTPase would be detrimental to the cell and has been recognized as a potential means of slowing viral replication and cancer growth.[2–4]

The crystal structure of *Mycobacterium tuberculosis* dUTPase presented here is one facet of a much larger effort undertaken by the *M. tuberculosis* Structural Genomics Consortium to develop leads for the treatment of tuberculosis infections in the human population. Approximately one third of the world's human population is infected with *M. tuberculosis* and nearly two million people die every year from the infection.[5] The emergence of drug resistant strains of *M. tuberculosis* has made the search for new drugs more urgent. Because dUTPase is essential for cell viability,[6] inhibition of dUTPase might be a means to slow or stop the growth of *M. tuberculosis* in humans. Obtaining the crystal structure of *M. tuberculosis* dUTPase is a necessary first step in the process of designing a drug specific for inhibiting this target.

One important concern in designing an inhibitor of *M. tuberculosis* dUTPase is the possibility of cross-reaction with the human dUTPase. The human dUTPase shares 34% sequence identity with the *M. tuberculosis* enzyme and so conceivably could be inhibited by the same drug designed to inhibit the *M. tuberculosis* enzyme. Ideally, cross-reaction could be prevented by engineering species specificity into the drug, i.e. designing a drug molecule that is exquisitely complementary to the active site of the *M. tuberculosis* enzyme but poorly complementary to the human enzyme. Conceptually, specificity could be more readily achieved if large structural differences exist between the active sites of the host and pathogen enzymes. Structural comparisons between *M. tuberculosis* and human[7] dUTPase suggest a means of exploiting these differences for selectivity in drug design.

The structure of *M. tuberculosis* dUTPase complexed with Mg^{2+}·α,β-imido dUTP offers new insights into the mechanism of dUTP hydrolysis. Previous proposals for a catalytic mechanism[8,9] have been lacking in detail despite the availability of a wealth of biochemical data and a large gallery of crystal structures from *Escherichia coli*,[10–13] human,[7] EIAV,[14] and FIV[8] sources. The structures, though numerous, have provided only a distorted view of the enzyme–substrate (E·S) complex. They offer glimpses of either dUTP or divalent metal ion in the active site, but not in combination. Indeed, the true E·S complex would be unstable over the course of a conventional crystallographic experiment. We offer a more accurate view of the E·S complex by presenting the structure of dUTPase complexed with Mg^{2+} and the non-hydrolyzable substrate analog, α,β-imido dUTP.[15] The analog differs from dUTP solely by the replacement of an imido group for the oxygen bridging the α–β phosphate groups. The decreased electrophilicity of the imido group renders the α-phosphate less electropositive and thus less susceptible to nucleophilic attack.[16] Thus, both metal ion and nucleoside triphosphate can bind stably to the enzyme allowing the capture of a crystallographic snapshot.

The catalytic roles of the metal ion and the 5′-triphosphate tail of the substrate become evident by comparison of the Mg^{2+}·α,β-imido dUTP complex with structures of three other complexes (Table 1): (1) the Mg^{2+}·dUDP complex, which contains divalent metal but lacks the γ-phosphate; (2) the dUTP complex, which contains a 5′-triphosphate tail but lacks a metal ion; and (3) the Cr^{3+}·dUTP complex, which contains both a 5′-triphosphate tail and a metal ion, but also an extra positive charge. The last complex is bound stably to the enzyme because chromium has a 10^{13} times slower exchange rate with its ligands than does magnesium.[17,18] We show that both the metal ion and triphosphate tail are required to achieve a catalytically productive conformation of the substrate and suggest a modified mechanism for nucleotide hydrolysis that appears more consistent with biochemical data.

Results and Discussion

The dUTPase fold: structural comparisons among five homologs

Crystal structures have been determined of dUTPases from four other sources: *E. coli*,[10–13] human,[7] EIAV,[14] and FIV[8] (Table 2). All have been determined in the unliganded form and in complex with the inhibitor dUDP. Both human and FIV dUTPase have been determined in the presence of substrate, dUTP. FIV dUTPase has additionally been determined in the presence of product. All these homologs share five conserved sequence motifs that form the active site (Figure 1(c)). All share a common distorted β-barrel fold and are domain swapped trimers (Figure 1(b)). Structural variation caused by ligand binding is small (rms deviation in α-carbon positions = 0.5(±0.1) Å) and limited to local regions of the structure (see below). All homolog structures can be readily superimposed on the *M. tuberculosis* structure (Figure 1(a)). However, the fit to *M. tuberculosis* dUTPase is somewhat better for two chromosomal dUTPases ($rms^{MTb\text{-}E.\ coli} = 1.1$ Å for 122 α-carbon atoms; $rms^{MTb\text{-human}} = 1.0$ Å for 120 α-carbon atoms) than the viral dUTPases ($rms^{MTb\text{-FIV}} = 1.0$ Å for 106 α-carbon atoms; $rms^{MTb\text{-EIAV}} = 1.2$ Å for 105 α-carbon atoms). An all *versus* all comparison of dUTPase homolog structures using the program MAPS[19] confirms that the chromosomal dUTPases (*M. tuberculosis*, *E. coli*, and human) form a group somewhat set apart from the viral dUTPases (FIV

Table 1. Data collection and refinement statistics

Ligand	None	dUMP	dUTP	dUTP	dUDP	α,β-Imido dUTP	α,β-Imido dUTP
Metal	None	Mg^{2+}	None	Cr^{3+}	Mg^{2+}	Mg^{2+}	Mg^{2+}
Data collection statistics							
Resolution (Å)	1.95	1.85	2.05	2.90	3.00	1.30	1.80
Space group	$I23$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P3_121$	$P6_3$	$P2_12_12_1$
Unit cell dimensions							
a (Å)	98.82	58.23	58.60	57.95	108.27	54.82	57.52
b (Å)	98.82	77.50	77.70	77.72	108.27	54.82	78.51
c (Å)	98.82	93.95	94.78	91.61	100.79	84.21	94.51
Radiation source	ALS 5.0.1	NSLS X8C	Rigaku FR-D	Rigaku FR-D	Rigaku FR-D	ALS 8.2.2	ALS 8.2.2
Radiation wavelength (Å)	1.000	1.100	1.542	1.542	1.542	0.979	1.127
Measured reflections	208,809	125,532	364,403	82,930	119,311	426,063	197,772
Unique reflections	11,592	36,211	27,903	10,004	14,054	35,121	39,997
Overall completeness (%)	97.1	97.9	100.0	98.3	99.9	99.8	98.8
Last shell completeness (%)	100.0	98.7	100.0	98.5	100.0	100.0	99.7
Overall R_{sym}[a]	0.078	0.070	0.109	0.143	0.161	0.125	0.069
Last shell R_{sym}	0.393	0.462	0.453	0.473	0.524	0.319	0.325
Overall $I/\sigma(I)$	62.7	14.7	22.7	15.5	14.8	18.1	20.0
Last shell $I/\sigma(I)$	13.5	2.9	5.8	5.1	3.9	7.3	5.7
Refinement statistics							
R_{work}[b]	0.197	0.169	0.166	0.235	0.224	0.120	0.168
R_{free}[c]	0.226	0.204	0.207	0.282	0.265	0.138	0.198
rmsd bond lengths (Å)	0.02	0.02	0.02	0.01	0.01	0.01	0.02
rmsd bond angles (deg.)	2.0	2.7	2.2	1.5	1.8	2.0	1.9
PDB ID code	1mq7	1snf	1smc	1sm8	1slh	1six	1sjn

[a] $R_{sym}(I) = \sum_{hkl}((\sum_i |I_{hkl,i} - \langle I_{hkl}\rangle|)/\sum_i I_{hkl,i})$.
[b] $R_{work} = \sum_{hkl}|F_{obs} - F_{calc}|/\sum_{hkl}F_{obs}$.
[c] $R_{free} = \sum_{hkl}|F_{obs} - F_{calc}|/\sum_{hkl}F_{obs}$, where all reflections belong to a test set of 5% randomly selected data.

and EIAV). The larger structural difference between viral and chromosomal dUTPases is consistent with evidence from sequence analysis.[20] Some of the differences between the viral and chromosomal dUTPases can be readily seen in the outlying loops in Figure 1(a).

Within the chromosomal dUTPases there is conflicting evidence as to whether *M. tuberculosis* dUTPase more closely resembles *E. coli* or human dUTPases. Analysis of conserved dUTPase sequence motifs[20] suggests that *M. tuberculosis* dUTPase is evolutionarily closer to the *E. coli* enzyme than the human enzyme. This analysis is consistent with what would be expected based on phylogeny (i.e. *M. tuberculosis* and *E. coli* are both prokaryotic, whereas humans are eukaryotic). But, further comparison of overall sequence identity among three homologs (using default parameters within the BLAST alignment tool[21]) reveals that *M. tuberculosis* dUTPase shares more sequence identity with human dUTPase (43/108 identity, $E = 2e^{-12}$) than *E. coli* dUTPase (38/107 identity, $E = 4e^{-09}$) (Figure 1(c)). Furthermore, *M. tuberculosis* and human dUTPases share the same number of residues in three of their surface loops (Figure 1(a) and (c)), whereas the *E. coli* enzyme contains short insertions in these loops. One interpretation of this apparent conflict is that *M. tuberculosis* dUTPase is closer to neither *E. coli* nor human dUTPases, but is instead

Table 2. Accession codes of dUTPase crystal structures deposited in PDB

	Unliganded	dUMP	dUDP	dUTP	α,β-Imido dUTP
M. tuberculosis	1mq7	1snf (Mg^{2+})	1slh (Mg^{2+})	1smc 1sm8 (Cr^{3+})	1six (Mg^{2+}) 1sjn (Mg^{2+})
E. coli	1dup 1eu5 1euw		1dud		
Human	1q5u	Not in PDB	1q5h	Not in PDB	
EIAV	1dun		1duc (Sr^{2+})		
FIV	1dut 1f7d 1f7o	1f7k 1f7n	1f7p 1f7r	1f7q	

When metals are included in the crystal structure, they are indicated in parenthesis next to the PDB ID code.

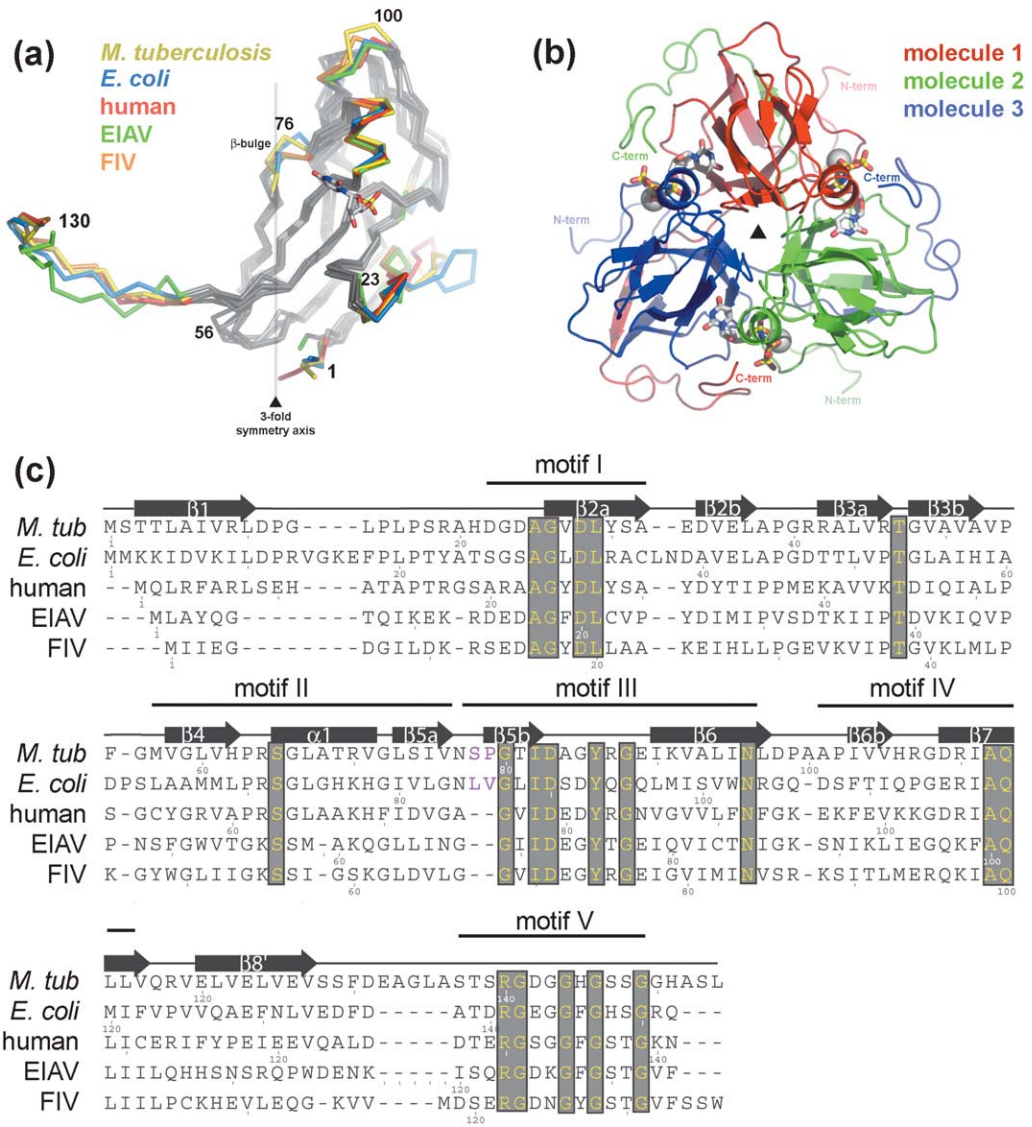

Figure 1. (a) Structure of the dUTPase monomer. α-Carbon traces of five dUTPase homologs are shown superimposed. Regions that are structurally equivalent among all five homologs are colored gray. Regions of structural variation are colored by source according to the key indicated in the panel. (b) Structure of the *M. tuberculosis* dUTPase trimer complexed with magnesium and α,β-imido dUTP. The view is related to (a) by a 90° rotation about the horizontal axis. The cyclic domain swap of C-terminal β-strands can be readily seen. (c) Sequence alignment for the five dUTPase homologs. Boxed residues are absolutely conserved among the five dUTPases. The secondary structure of the *M. tuberculosis* dUTPase is indicated above the sequence. Locations of the five conserved sequence motifs are indicated above the secondary structure. The prokaryote specific two-residue insertion in the uracil-binding pocket is highlighted in magenta.

approximately evolutionarily equidistant from the two. Structural comparisons among the three dUTPases support this interpretation ($rms^{MTb\text{-}E.\ coli} = 1.1$ Å for 122 α-carbon atoms; $rms^{MTb\text{-human}} = 1.0$ Å for 120 α-carbon atoms, $rms^{E.\ coli\text{-human}} = 1.32$ Å for 115 α-carbon atoms). Interestingly, the structural differences between human and *E. coli* dUTPases are significantly greater than in the other two pairwise comparisons suggesting that *M. tuberculosis* dUTPase may somehow represent a composite (or chimera) between *E. coli* and human "extremes".

It might seem peculiar that the similarity between dUTPase homologs within the same kingdom (i.e. *M. tuberculosis* and *E. coli*) is no stronger than the similarity across kingdoms (i.e. *M. tuberculosis* and human). However, this peculiarity is not limited to dUTPase, but may be true of many enzymes encoded by *M. tuberculosis*. Three other metabolic enzymes have been structurally determined from *M. tuberculosis*, *E. coli* and human sources. When the structures from *M. tuberculosis* encoded genes are compared to their *E. coli* and human homologs, only one of these enzymes, dihydrofolate reductase, shows a clear adherence to phylogeny ($rms^{MTb\text{-}E.\ coli} = 0.7$ Å for 126 α-carbon atoms; $rms^{MTb\text{-human}} = 1.0$ Å for 128 α-carbon atoms).[22] Thymidylate kinase, as in

the case of dUTPase, is structurally equidistant between *E. coli* and human ($rms^{MTb\text{-}E.\ coli} = 1.5$ Å for 160 α-carbon atoms; $rms^{MTb\text{-}human} = 1.6$ Å for 137 α-carbon atoms)[23] with sequence and kinetic properties closer to eukaryotic than bacterial homologs.[24] Lastly, purine nucleoside phosphorylase clearly resembles the human enzyme more closely than the *E. coli* enzyme ($rms^{MTb\text{-}E.\ coli} = 1.6$ Å for 140 α-carbon atoms; $rms^{MTb\text{-}human} = 1.2$ Å for 225 α-carbon atoms).[25] The unexpected blurring of phylogenetic relationships might cause one to suspect the involvement of horizontal gene transfer.[26]

Structural conservation of the uracil-binding pocket: prospects for drug design

The organization of the ligand-binding pocket is conserved among all structurally determined trimeric dUTPases to date, including *M. tuberculosis* dUTPase. Trimeric dUTPases contain three identical ligand binding sites located at clefts between adjacent molecules (Figure 1(b)). A single pocket is composed of five conserved sequence motifs. Motifs I, II, and IV from one molecule contribute to binding the triphosphate moiety (colored blue, green, and orange, respectively, in Figure 2(a) and (c)). Motif III from the adjacent molecule contributes primarily to binding the nucleoside moiety (colored yellow in Figure 2(a) and (c)). Motif V from the remaining molecule is a domain swapped element that contributes an arginine residue involved in a bifurcated hydrogen bond with the ligand's phosphate moiety (colored red in Figure 2(a) and (c)). Motif V has been reported to contain sequence similarity to the P-loop motifs commonly employed in nucleotide binding enzymes to bind phosphate groups.[9]

The deoxyuridine moiety binds in a gap between β-strands 5 and 6 (motif III) (Figure 2(a)). Backbone nitrogen and oxygen atoms of β-strand 6 (Lys91 in *M. tuberculosis*) provide hydrogen bond partners for O2 and N3 of uracil (Figure 2(a)). Similarly, Asn77($N^{\delta 2}$), which resides on a bulge in β-strand 5, donates a hydrogen bond to O4 (Figures 2(a), and 3(a) and (b)). The loop which connects β-strands 5 and 6 further contributes to deoxyuridine binding and selectivity. The side-chain of Tyr86 provides van der Waals contacts with the deoxyribose pentose ring and at the same time it excludes UTP binding by steric interference with a potential 2′-OH of ribose (Figure 2(a)). The side-chain of Asp83 accepts a hydrogen bond from the deoxyribose 3′-OH. These interactions are conserved whether dUMP, dUDP, or dUTP is bound.

Because the residues in the ligand-binding pocket are well conserved, most of the details of deoxyuridine binding geometry are analogous to what has been reported for dUTPases homologs in the literature. However, there is one significant difference in the architecture of the uracil-binding pocket of prokaryotic dUTPases that sets them apart from dUTPases from other sources. It was mentioned above that a residue from β-strand 5 provides a hydrogen bond donor for O4 of uracil. But this donor can be a backbone atom or side-chain depending on the enzyme source (Figure 3). In the non-prokaryotic dUTPases, the hydrogen bond donor is a backbone amide nitrogen atom (e.g. Gly76 in human). But, in the prokaryotic dUTPases there is a two-residue β-bulge insertion in β-strand 5 (Ser78 and Pro79 in *M. tuberculosis*) which places the $N^{\delta 2}$ atom of an asparagine side-chain (Asn77 in *M. tuberculosis*) in the position of hydrogen bond donor to O4. It might be expected that the insertion of two residues would significantly change the volume of the nucleotide-binding pocket. But, because the β-bulge is directed toward the trimer axis and away from the uracil-binding pocket, the volume of the pocket changes only slightly (Figure 3). Hence, the structural complementarity between ligand and enzyme is remarkably similar in both dUTPase homologs.

Exploiting the small, species-specific differences between binding pockets of human and *M. tuberculosis* dUTPases would present a challenging hurdle to overcome in drug design. Instead, it might be simpler and equally effective to exploit the larger species-specific differences in the trimer interface channel caused by the Ser78, Pro79 insertion. The insertion constricts the diameter of the trimer interface channel relative to the human enzyme (Figure 3), reducing the volume of the channel from 673 $Å^3$ in the human enzyme to 309 $Å^3$ in the *M. tuberculosis* enzyme.[27] Furthermore, side-chains lining the channel differ significantly between species. A non-nucleoside inhibitor targeted to bind in this channel would likely interfere with dUTP binding by distorting the conformation of residues directly contacting the uracil moiety (Asn77, Gly80). In addition, proximity of the drug to the 3-fold symmetry axis of the trimer might allow it to disrupt the catalytic activity by interfering with the geometry of trimer assembly. A well ordered molecule of Tris buffer was found in this channel, only 5 Å from the β-bulge in the crystal structure of the unliganded *M. tuberculosis* dUTPase as well as complexes with dUMP and dUDP (Figure 3(b)). The Tris molecule might even provide an initial lead for the design of bulkier inhibitors that would disrupt the active site.

Triphosphate–metal interactions: α,β,γ tridentate chelation

Electron density maps of the *M. tuberculosis* dUTPase·Mg^{2+}·α,β-imido–dUTP complex ($2F_{obs} - F_{calc}$ as well as omit maps) reveal the triphosphate moiety chelating the metal ion as an α,β,γ tridentate; each phosphate contributes one oxygen atom to magnesium's coordination sphere (Figure 2(b)). The structure has been determined at high resolution in two different crystal packings (1.3 Å resolution in space group $P6_3$ and 1.8 Å resolution in space group $P2_12_12_1$), lending accuracy to our observations and allowing us to

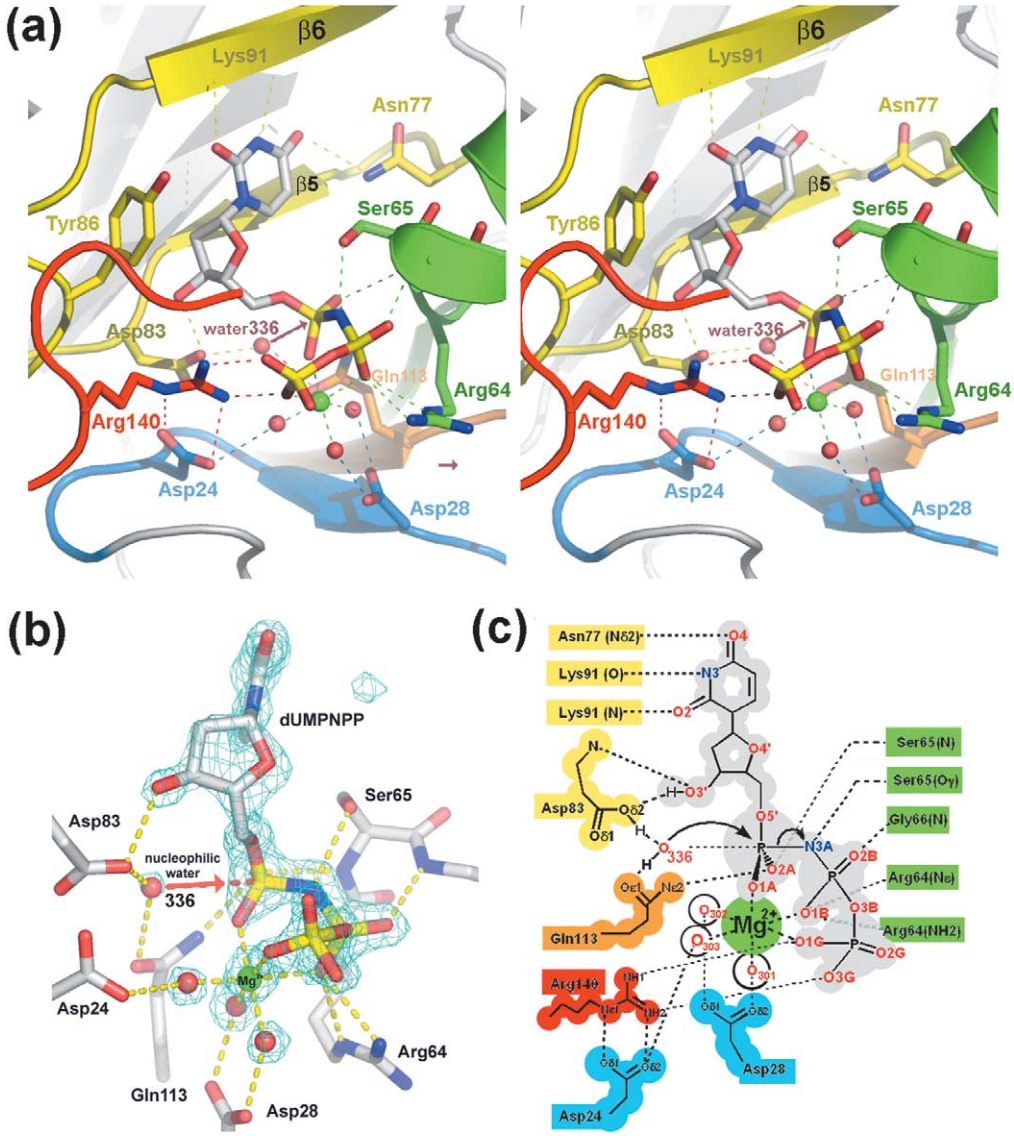

Figure 2. dUTPase·Mg^{2+}·α,β-imido dUTP complex (PDB ID 1six). (a) Stereo view of the active site. The α,β-imido dUTP atoms are colored by atom type. Enzyme residues are colored according to conserved motif. Colors are as follows: motif I (blue), motif II (green), motif III (yellow), motif IV (orange), motif V (red). The active site is a composite of all three molecules. Motifs I, II, and IV are contributed by molecule A. Motif III is contributed by molecule B. Motif V is contributed by molecule C. (b) Simulated annealing omit map showing the density for the ligand, magnesium, and coordinated water molecules (1.6 Å resolution, contoured at 4.2σ). Density is also shown for the proposed nucleophilic water molecule, 336. (c) Schematic of the dUTPase reaction mechanism.

conclude that crystal packing forces play no significant role in altering the geometry of ligand binding (Table 1).

The coordination geometry observed here closely resembles the α,β,γ tridentate coordination geometry observed in the structure of DNA polymerase β complexed with gapped DNA, magnesium, and incoming dNTP analog (Figure 4).[28] This similarity is not surprising given that like dUTPases, DNA polymerases also catalyze the in-line nucleophilic attack on the α-phosphate of a dNTP. The difference is that the nucleophile in the polymerase reaction is the 3′-hydroxyl of the primer strand rather than a water molecule as it is in dUTPase. This tridentate coordination geometry is also consistent with expectations from kinetic experiments that the α-phosphate would be involved in metal ion coordination.[29] In these experiments, the use of racemic mixtures of 2′-deoxyuridine-5′-(α-thio)-triphosphate (dUTPαS), showed that only one enantiomer of the substrate analog was hydrolyzed by the enzyme. Based on the crystal structure we presume that the enantiomer which is unable to serve as a substrate for dUTPase contains the *S* configuration at the α-phosphate moiety. In this configuration the favorable Mg^{2+}–O interaction seen in the crystal structure would be replaced with an unfavorable Mg^{2+}–S interaction, thus preventing the reaction from proceeding.

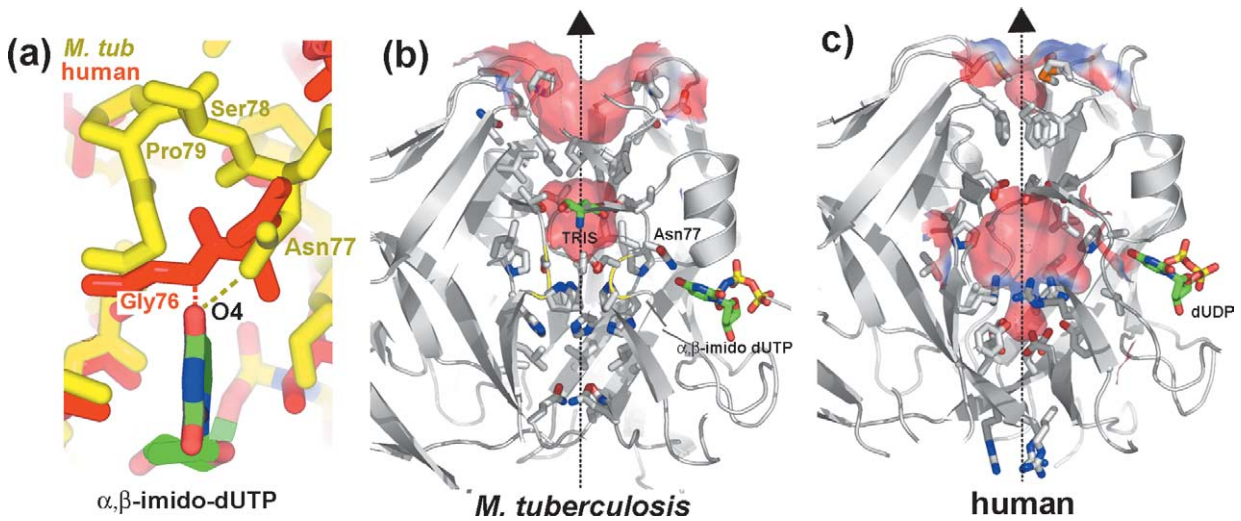

Figure 3. Comparison of uracil binding pockets of *M. tuberculosis* and human dUTPases. (a) The two-residue β-bulge insertion in conserved motif III of *M. tuberculosis* dUTPase (Ser78-Pro79) is one of the largest structural differences in comparison to the human enzyme. The hydrogen bond donor to O4 of uracil is a side-chain residue (Asn77($N^{\delta 2}$)) in the *M. tuberculosis* enzyme, but it is the backbone amide of Gly76 in the human enzyme. (b) Cut-away view of the trimer interface channel of *M. tuberculosis* dUTPase and (c) human dUTPase. The broken line marks the position of the 3-fold symmetry axis. The channel surface is colored by electrostatic potential. Red and blue colors correspond to negative and positive potential, respectively. The insertion in the *M. tuberculosis* enzyme (colored yellow in (b)) produces only a small increase in the breadth and width of the uracil-binding pocket, but it profoundly constricts the diameter of the trimer interface channel relative to the human enzyme. A non-nucleoside inhibitor targeted to bind in this channel would likely interfere with dUTP binding by distorting the conformation of residues lining the uracil-binding pocket and/or the trimer assembly.

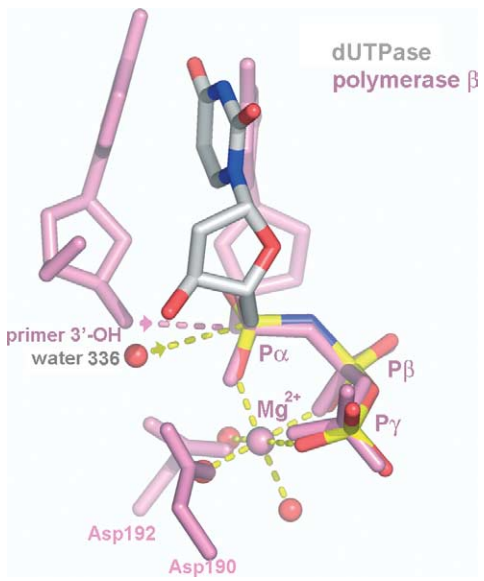

Figure 4. Comparison of conformation of Mg^{2+}·α,β-imido dUTP bound to dUTPase with the conformation of the incoming ddCTP in DNA polymerase β (1bpy). (The 3′-hydroxyl of the primer was modeled based on the position of the dideoxyribose ring.) Both enzymes use a metal ion to coordinate the triphosphate moiety with α,β,γ tridentate geometry. Both enzymes catalyze the in-line nucleophilic attack on the α-phosphorus of a dNTP. Even the nucleophiles in both enzymes have the same position with respect to the α-phosphorus. The difference is that the nucleophile in the polymerase reaction is the 3′-hydroxyl of the primer strand rather than a water molecule as it is in dUTPase.

The coordination geometry of the magnesium ion is nearly perfectly octahedral; half of magnesium's coordination shell is contributed by the α,β,γ phosphate groups as described above and the other half is contributed by three enzyme bound water molecules. These water molecules are hydrogen bonded to conserved active site residues: Asp24, Asp28 (motif I), and Arg140 (motif V). These water molecules have low *B*-factors (under 30 $Å^2$) and are present in both crystal forms of the dUTPase·Mg^{2+}·α,β-imido–dUTP complex (Figure 2).

Interestingly, there is a water molecule situated 3.4 Å from the α-phosphorus (water 336 in the $P6_3$ structure), uniquely positioned for a direct in-line nucleophilic attack (Figure 2).

In addition to the metal ion mediated contacts with the triphosphate, there are several enzyme mediated contacts (primarily from motif II) involving most, but not all of the remaining oxygen atoms on the triphosphate. These are illustrated in Figure 2. Most notable is the appearance of a bifurcated hydrogen bond between the Arg140 side-chain (motif V) and the γ-phosphate of α,β-imido–dUTP. Arg140 is part of the domain swapping loop and is disordered in all the other *M. tuberculosis* complexes.

The structural role of the β and γ phosphate moieties

The β and γ phosphate groups might facilitate nucleophilic attack at the α-phosphate by sterically

crowding the α-phosphorus closer to the nucleophilic water molecule. Crystal structures of Mg^{2+}·dUDP, and Mg^{2+}·dUMP closely resemble the Mg^{2+}·α,β-imido–dUTP geometry described above, but there is a small displacement in the position of the α-phosphate that varies with the number of phosphate groups in the 5′ tail. Water 336 was described earlier as a potential candidate for making a nucleophilic attack on the α-phosphorus in the dUTPase reaction. The distance between this water and the α-phosphate of dUMP, dUDP, and α,β-imido–dUTP was measured after superimposing the Mg^{2+}·dUDP, and Mg^{2+}·dUMP complexes on the Mg^{2+}·α,β-imido–dUTP complex. This distance was found to decrease from 4.0 Å to 3.8 Å to 3.4 Å for the mono, di-, and tri-phosphate ligands, respectively. The observation also suggests that the inability of dUTPase to catalyze the hydrolysis of dUDP[29] might arise in part from the increased distance between the α-phosphate and the nucleophilic water molecule. The γ-phosphate appears to help position the α-phosphate for catalysis, but it likely performs other catalytic roles as well (see below).

The role of the metal ion in ordering the 5′-phosphate tail of dUTP

The structure of the dUTPase·dUTP complex in the absence of metal ion reveals two conformations of the 5′-triphosphate tail (Figure 5(b)). The two conformations differ in the position of the 5′-phosphate tail. The difference can be roughly described by a 120° rotation about the C4′–C5′ bond of the deoxyribose moiety. The predominant conformation of the triphosphate tail closely resembles the α,β,γ-tridentate conformation described above for the Mg^{2+}·α,β-imido–dUTP complex and maintains the same set of interactions with enzyme residues, even though the metal ion is absent. However, *B*-factors are approximately 50 $Å^2$ higher in the absence of metal. The value of the C3′–C4′–C5′–O5′ torsion angle is close to +60° in this conformation and so will be referred to as the *gauche* conformation. In the less occupied conformation of dUTP, the value of this torsion angle is 180° and so will be referred to as the *trans* conformation. The overall appearance of the *trans* conformation is that the α and γ phosphate groups have swapped places with respect to their positions in the *gauche* conformation (Figure 5(b)). The result is that many of the enzyme–phosphate interactions in the *trans* conformation involve the same enzyme residues as in the *gauche* conformation. This flexibility was not observed in any of the liganded complexes containing metal ion, suggesting that the metal ion plays a significant role in positioning the α-phosphorus for nucleophilic attack. Indeed, the dUTPase reaction is known to be metal ion dependent and the K_d for dUTP binding to *E. coli* dUTPase increases by a factor of 100 in the absence of magnesium.[29]

Conformational variability about the C4′–C5′ bond has been reported in crystallographic studies of dUTPases from other organisms as well. The *gauche* conformation is found in the *E. coli*–dUDP complex and FIV–dUMP complexes and the *trans* conformation is found the dUDP complexes of human, FIV, and EIAV dUTPase (Figure 5(a)). Apparently there is a low energy barrier between the two conformations. Given our understanding of how metal ions stabilize the *gauche* conformation of triphosphate binding geometry in *M. tuberculosis* dUTPase, it is our hypothesis that if a non-hydrolyzable dUTP analog with magnesium were included in these complexes, the predominant conformation of the triphosphate would be the *gauche* conformation.

Ligand induced ordering of the active site loop, motif V

There is little overall structural variation caused

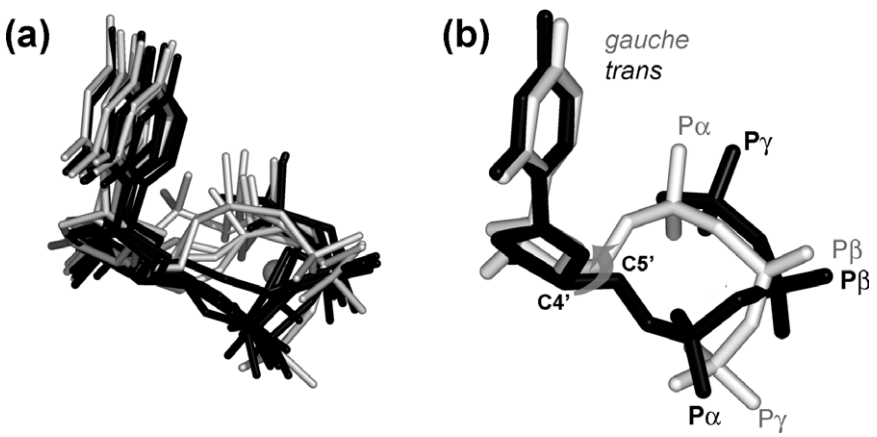

Figure 5. (a) Superposition of ligands bound to dUTPases from all five sources. The structures may be divided into two groups, differing primarily in the torsion angle about the C4′–C5′ bond. One group binds with *trans* geometry (black). The other group binds with *gauche* geometry (gray). (b) *Trans* (black) and *gauche* (gray) geometry of dUTP bound to *M. tuberculosis* dUTPase in the absence of metal ions. The distance between αP positions of the two conformations is approximately 4.5 Å.

by ligand binding in *M. tuberculosis* dUTPase (rms deviation in α-carbon positions $<0.5(\pm 0.1)$ Å for 130 residues). Most of the variation is limited to motif V (131–154), the C-terminal active site loop and is observed in dUTPases from all five sources. The kinetic behavior of motif V, the C-terminal loop is of particular interest because it is known to be essential for catalysis; proteolytic removal of these C-terminal residues[30,31] or point mutations of certain key residues in this motif result in complete loss of activity.[32]

Crystallographic and kinetic evidence suggests that both the divalent metal ion and the triphosphate moiety are required to order the loop. Among the seven reported structures of *M. tuberculosis* dUTPase, motif V appears ordered in only two structures, both containing Mg^{2+}·α,β-imido dUTP. The loop remains disordered in the other enzyme–substrate analog complexes: dUTP, Mg^{2+}·dUDP, and Cr^{3+}·dUTP, as well as the unliganded enzyme and Mg^{2+}·dUMP product complex. Consistent with this observation, far UV circular dichroism and NMR spectral shifts indicate that ordering of the C-terminal loop is strictly Mg^{2+} dependent and is elicited by binding α,β-imido–dUTP but not dUDP.[16]

A possible mechanism for ordering the loop might be that the divalent metal ion first orders the γ-phosphate through direct coordination, and then the γ-phosphate orders the loop through a bifurcated hydrogen bond with motif V residue Arg140. Thus, loop disorder in the Mg^{2+}·dUDP and Mg^{2+}·dUMP complexes can be explained by the lack of one of the participants in this hydrogen bond, the γ-phosphate. Loop disorder in the Cr^{3+}·dUTP complex might be the result of decreased electrostatic attraction between Arg140 and the γ-phosphate imparted by the extra positive charge carried by the chromium ion. Loop disorder in the dUTP complex probably arises from the increased mobility in the γ-phosphate caused by the lack of a chelating metal ion; a larger entropic barrier to ordering the loop is incurred. It should be further noted that even in the Mg^{2+}·α,β-imido dUTP complex structure, the ordered portion of the loop does not involve the full length of motif V. Only the N-terminal half of motif V (residues 131–144) becomes ordered. Residues 145–154 remain disordered.

Crystal structures of four homologous dUTPases largely support our hypothesis that ordering of the C-terminal loop requires both metal and nucleoside triphosphate. However, there are four exceptions. Four crystal structures (human dUMP, dUDP (1q5h), and dUTP complexes; FIV complexes with dUDP (1f7r)) show clearly resolved active site loops yet lack the combination of a metal ion and triphosphate. Furthermore, these crystal structures contain the *trans* conformation of nucleotide binding, which we claim to be non-productive. In the following section, we propose that the ordering of the loop in these complexes is stabilized fortuitously by the non-productive *trans* conformation.

A mechanism for activating the nucleophilic water

The crystal structures of *M. tuberculosis* dUTPase complexed with Mg^{2+}·α,β-imido dUTP suggest the identity of a nucleophilic water molecule and general base that conform well to the accepted geometric requirements of an in-line nucleophilic attack and are consistent with kinetic data available on dUTPases. Water molecule 336 (using nomenclature from the $P6_3$ structure) is located 3.4 Å away from the α-phosphate and deviates only 15° from making a direct in-line attack (Figure 2), i.e. the water molecule, αP and O3 are positioned nearly in a line. This water is well ordered and completely shielded from solvent, with a *B*-factor of 30 Å^2. The shielded environment of water 336 is consistent with steady state kinetic studies of *E. coli* dUTPase which established that k_{cat} is insensitive to pH.[29] It was proposed that the nucleophilic water molecule becomes trapped in the substrate pocket on formation of the enzyme–substrate complex and is shielded from the surrounding medium. Indeed, this appears to be the case in the structure of the *M. tuberculosis* complex.

Anchoring water 336 near the α-phosphate is a pair of hydrogen bonds to the $O^{\varepsilon 1}$ atom of conserved glutamate residue Gln113 (motif IV) and absolutely conserved aspartate side-chain of Asp83 (motif III) (Figure 2(b) and (c)). Asp83 appears to be well suited to play the role of general base for the buried water molecule. Binding of substrate, in addition to burying water 336 also shields Asp83 from solvent, increasing its pK_a significantly. Shielding of Asp83 provides a mechanism by which the carboxylate may abstract a proton from water 336 to generate a nucleophilic hydroxide ion. The role of Asp83 as general base is further supported by mutational studies in which the homologous aspartate in human endogenous retrovirus dUTPase (Asp84 to Asn) was found to abolish activity.[33] The role of Gln113, in addition to positioning the nucleophilic water molecule, may be to direct the lone pair of electrons on the water molecule's oxygen atom toward the phosphorus. Gln113 does this through a hydrogen bond between $O^{\varepsilon 2}$ and water molecule's remaining hydrogen atom.

Further stabilization of the transition state is derived from additional hydrogen bonds formed between conserved enzyme residues and the phosphate oxygen atoms. Negative charges developing on the non-bridging oxygen atoms of the α-phosphate are stabilized by contact with Mg^{2+} and a hydrogen bond to the backbone nitrogen atom of Ser65 (motif II) (Figure 2). The leaving group, pyrophosphate, is stabilized by hydrogen bonds to the backbone nitrogen atom of Gly66 (motif II), the side-chain of Arg64 (motif II), and a bifurcated hydrogen bond with Arg140 (motif V). The largest

change in charge upon reaching the transition state is likely to be in the α–β bridging oxygen.[34] The developing negative charge in this oxygen atom (an imido nitrogen atom in the crystal structure) would appear to be stabilized by a hydrogen bond to the absolutely conserved hydroxyl side-chain of Ser65 (motif II). In G proteins, a class of enzymes that hydrolyzes GTP, a hydrogen bond to the bridging oxygen atom is donated by a backbone amide.

The nearly ideal geometry for nucleotide hydrolysis described above is largely dependent on the *gauche* conformation of nucleotide binding. If the nucleotide bound in the *trans* conformation as hypothesized[8,9] the α-phosphate would be shifted 4.5 Å away and many of the requirements for in-line nucleophilic attack would be lacking. Indeed, in crystal structures where the *trans* conformation is observed (human and FIV dUTPase structures complexed with dUTP and dUDP), there is no suitably located molecule between the aspartate and the α-phosphate. Furthermore, if a water molecule were modeled between the aspartate and α-phosphate, it would be 60° away from the proper angle to make an in-line attack and in steric conflict with the conserved arginine residue from motif V. Lastly, the enzyme's ability to electrostatically stabilize the transition state would appear to be compromised by partial solvent exposure of the α-phosphate.

Catalytic role of conserved sequence motif V, the C-terminal loop

Kinetic and structural studies suggest that motif V plays an important role in activating the nucleophilic water molecule for attack on dUTP. In the crystal structure of *M. tuberculosis* dUTPase complexed with Mg^{2+}·α,β-imido-dUTP, the arginine side-chain (Arg140) is positioned so that it can shield the general base (Asp83) from solvent. Asp83 is solvent accessible in the unliganded enzyme (36 $Å^2$ of accessible surface area). Ordering of Arg140 through its bifurcated hydrogen bonding interaction with the γ-phosphate reduces the solvent accessibility of Asp83 to 0 $Å^2$ thus increasing its pK_a and enhancing its ability to activate the nucleophilic water for attack (Figure 2(a)). The crystal structure of the E·P complex (dUTPase·Mg^{2+}·dUMP) suggests that after the nucleophilic attack, the loop again becomes disordered and the pyrophosphate product diffuses from the active site. Thus, the role of motif V appears to be to act as a gate, facilitating substrate binding and product release when disordered, then transiently altering the environment of the active site to promote the chemical step. Such a role is in keeping with the flexibility of this loop and is typical of active site loops in general. This proposed role is also consistent with kinetic studies indicating that motif V is essential for catalysis; removal of these C-terminal residues[30,31] or point mutations of certain key residues in this motif result in complete loss of activity.[32]

Although important for catalytic activity, motif V appears to contribute little to dUTP-binding affinity. Kinetic studies on *E. coli* and EIAV dUTPase report little change in K_M after the loop is proteolyzed.[30,31] Consistent with this observation, the structure of the *M. tuberculosis* dUTPase·Mg^{2+}·α,β-imido–dUTP complex reveals few contacts between motif V and the dUTP analog. The strongest of these few contacts is provided by Arg140, through a pair of hydrogen bonds with the γ-phosphate (Figure 2(a) and (c)). The observation that motif V interacts most strongly with the γ-phosphate is supported by NMR studies which have shown that the γ-phosphate is required for ordering the C-terminal tail of *E. coli* dUTPase in solution.[35] There are virtually no contacts involving conserved glycine residues of motif V (RGXXGϕGXXG). The first two glycine residues (G141 and G144) make a series of hairpin turns that meander over the active site and position a conserved aromatic side-chain over the uracil base. The aromatic side-chain is proposed to act like a lid, providing additional stability to the "closed loop" conformation during the chemical step. (In *M. tuberculosis* dUTPase, the aromatic residue is His145, but it is invisible in the crystal structure due to disorder. In crystal structures of the FIV and human dUTPases, the aromatic residue is a phenylalanine and is somewhat more ordered. The increased order observed in the latter structures might be due to improved stacking interactions between the uracil base and the larger surface area of the phenyl group of Phe compared to the imidazole group of His.) The lack of interactions observed between these conserved glycine residues and α,β-imido–dUTP conflicts with the role implied by motif V's sequence resemblance to a P-loop motif (e.g. GXXGXGKS/T).[9] P-loops are glycine-rich sequence motifs known to provide the majority of binding energy between a nucleotide binding enzyme and the phosphate portions of its ligand. In this particular case, the functional role implied by sequence similarity appears to be misleading. The structural and kinetic evidence presented here suggests that the role of these glycine residues is to enable formation of the hairpin structure while providing the overall flexibility needed to facilitate substrate binding and product release.

Conclusions

The crystal structure of *M. tuberculosis* dUTPase has been determined as part of a Structural Genomics Consortium effort to provide leads for the development of new anti-tuberculosis drugs. As dUTPase is also encoded by the human genome and provides an essential housekeeping function for host as well as pathogen, it would be beneficial for the host if the inhibitor drug were designed to be specific for the *M. tuberculosis* enzyme.

Comparison of *M. tuberculosis* and human dUTPase structures reveals few differences in the active site that can be exploited for engineering species specificity. However, there are large differences in the trimer interface channel, which lies directly behind the active site. A drug molecule binding in the *M. tuberculosis* dUTPase channel might disrupt the dUTPase active site with little or no chance of inhibiting the human enzyme.

In addition to the unliganded *M. tuberculosis* dUTPase, crystal structures have also been determined in complex with Mg^{2+}·dUMP, Mg^{2+}·dUDP, Cr^{3+}·dUTP, dUTP, and Mg^{2+}·α,β-imido–dUTP. These structures imply that both magnesium and the complete triphosphate tail are required for positioning the α-phosphate for nucleophilic attack and for transiently ordering part of the C-terminal loop, a 15 residue segment containing conserved sequence motif V. The metal ion coordinates α, β, and γ phosphate groups with tridentate geometry identical with that observed in the crystal structure of DNA polymerase β complexed with magnesium and dNTP analog, revealing some common features in reaction mechanism. Catalytically important residues identified from kinetic and mutagenesis studies are optimally positioned around the α-phosphate to participate in an in-line nucleophilic attack. The C-terminal loop appears to aid in activating the nucleophilic water for attack by temporarily shielding the general base from solvent and elevating its pK_a.

Materials and Methods

Enzyme preparation

Recombinant *M. tuberculosis* dUTPase was produced and purified from two different constructs: one with an N-terminal His tag (LANL), the other with a C-terminal His tag (UCLA).

Cloning, production, and purification of the N-terminally tagged dUTPase

Cloning. A 0.47 kb DNA fragment containing the *dut* gene (Rv2697c), was amplified by PCR from *M. tuberculosis* H37Rv genomic DNA as the template, using the following oligonucleotide primers: 5′-AGA TATA*CATATG* TCGACCACTCTG GCGATCGTCCGC-3′ and 5′-AATTCGGATCCCAAACTCGCATGTCCGCCG GAGGA-3′. The underlined bases represent the NdeI and BamHI sites, respectively. The amplified DNA fragment was digested with NdeI and BamHI restriction enzymes, and subcloned into the corresponding restriction sites in modified pET28b vector which provided an N-terminal 6-His tag upstream of the NdeI site. Thus, the expressed protein has amino acid residues extension MGSSHHHHHHSSGLVPRGSH N terminus. *E. coli* BL21PRO (Clontech) cells were transformed with the dut-modified pET28b/His vector.

Production. The transformed cells were grown to exponential phase at 37 °C in 5 ml EZMix LB broth medium (Sigma) containing 30 μg/ml kanamycin and 50 μg/ml spectinomycin. This seed culture was transferred to a 0.5 l EZMix Terrific broth medium (Sigma) and the expression was induced with 0.5 mM IPTG at an A_{600} of approximately 0.5. The growth was continued at 20 °C for approximately 21 hours until the A_{600} reached approximately 15 (as inferred from dilutions). The cells were harvested and stored at −80 °C.

Purification. The cell pellet was lysed by sonication in 10 ml of buffer A (20 mM Tris (pH 8.0), 100 mM NaCl) per gram of cells for ten minutes in 30 second pulses at 10 °C. The cell debris was removed by ultra-centrifugation for 30 minutes at 38,000 rpm using a Ti 60 rotor (Beckman). The clear supernatant was filtered through a 0.2 μm pore membrane and loaded onto a 5 ml Talon superflow affinity column equilibrated with buffer A. After washing with 50 ml buffer A, the His-tagged *dut* was eluted from the cobalt affinity column using buffer B (20 mM Tris–HCl (pH 8.0), 500 mM NaCl and 300 mM imidazole). The elutant was dialyzed against buffer C (20 mM Tris (pH 8.0), 100 mM NaCl, 10 mM β-mercaptoethanol) and purified by gel filtration on a Superdex-75 column using buffer C for equilibration and elution (Amersham Pharmacia Biotech). The peak fractions (monitored at A_{280}) were analyzed by SDS-PAGE and the pooled protein fractions were concentrated to 140 mg/ml using a Centriprep YM-3 (Millipore). The 95% pure *dut*, which was estimated by SDS-PAGE and MALDI-TOF mass spectroscopy (Applied Biosystem), was used for crystallization.

Cloning, production, and purification of the C-terminally tagged dUTPase

Cloning. The *dut* gene was amplified by PCR, in which *M. tuberculosis* H37Rv genomic DNA was used as the template. The forward primer: 5′-GATCCATATG GCT GTGTCGACCACTCTGGCGATCGTCCGCCTC-3′ introduced an NdeI site (underlined) and inserted an alanine codon (GCT) immediately following the start codon to enhance protein expression.[36] The reverse primer: 5′-CCACGGTTCCTCCGGCGGACATGCGAGTT TGGGTGTACCACGAGGTGCGGCCGCAAAT-3′ introduced a NotI site (underlined) and a thrombin recognition sequence to the C terminus. The PCR product was cloned into pCR-Blunt II-TOPO (Invitrogen), followed by sequence confirmation. The gene was subcloned into pET22b (Novagen), which appended a hexa-histidine tag to the expressed recombinant protein, trailing the thrombin recognition sequence. The resulting C-terminal extension had the amino acid sequence of GVPRGAAALEHHHHHH, immediately following the natural protein sequence.

Production. The recombinant protein was expressed in BL21-Gold (DE3) (Stratagene) *E. coli* in a 10-l batch fermentation, using a BioFlo 3000 (New Brunswick Scientific) fermentor. Enriched ECPM1 medium[37] was used with 100 μg/ml ampicillin.

Purification. Each gram of cell pellet was resuspended and lysed in 5 ml of lysis buffer (20 mM Tris (pH 8.0), 0.3 M NaCl, 10% (v/v) glycerol, 0.2% (w/v) NP40, 2 μg/ml DNase I, 0.2 mg/ml lysozyme, 2 mM β-mercaptoethanol, and 1 : 100 protease inhibitor cocktail (Sigma, cat. no. P8849)). The lysate was clarified by

centrifugation at 27,000*g* for 30 minutes. The soluble recombinant protein was initially purified using Ni-NTA Superflow resin (Qiagen). The protein was batch-eluted in buffer containing 300 mM imidazole.

Non-reducing Coomassie stained SDS-PAGE and Western blot indicated the presence of monomer, dimer, and trimer. However, only monomer is present upon treatment with approximately 20% (2.8 M) β-mercaptoethanol. While SDS-PAGE illustrated different polymeric forms of the protein, native-PAGE showed a homogeneous sample, suggesting that the protein existed in only one polymeric form in solution. Liquid chromatography/mass spectrometry (LC/MS) was carried out to confirm the molecular mass of the recombinant protein. It was found that the first methionine residue of the recombinant protein was proteolytically removed.

Gel filtration chromatography was subsequently performed on a Superdex 75 column (Amersham Biosciences) equilibrated with 20 mM Tris (pH 8.0), 0.3 M NaCl, 10% glycerol, and 0.2% NP40. According to the standard curve created using BioRad Gel Filtration Protein Standard (cat. no. 151-1901), the extrapolated molecular mass of the purified recombinant protein was roughly three times (3.3 times) the size of a monomer. The recombinant protein was purified to >99% purity after the two steps of chromatography, which was confirmed by SDS-PAGE and LC/MS.

Crystallization

Recombinant *M. tuberculosis* dUTPase was concentrated to approximately 60–100 mg/ml concentration in a Centricon YM10 (Millipore). The concentration of dUTPase was quantified by measurement of the optical density at 280 nm wavelength and use of an extinction coefficient of 2560 M^{-1} cm^{-1}. Buffer was exchanged with 10 mM NaCl, 20 mM Tris (pH 8.0), 5 mM β-mercaptoethanol with three consecutive washes in the Centricon using 2 ml volumes of buffer. Both His-tagged versions of the protein (N-terminal His-tagged version = 17.9 kDa; C-terminal His-tagged version = 17.7 kDa) were used in screening for crystals of unliganded enzyme and liganded complexes. Frequently, results differed depending on the location of the His-tag. The particular His-tagged version used for crystallization of each crystal form is noted below. Both versions were stored at 4 °C. Crystals were grown using either the hanging or sitting drop vapor diffusion method. Reservoirs contained 0.5 ml of solution. All crystals were grown at room temperature. Unless otherwise noted, crystals were cryo-protected by swiping the crystal quickly through a solution containing 65% (v/v) reservoir solution and 35% glycerol.

Crystallization of unliganded dUTPase

The unliganded enzyme was crystallized by the LLNL crystallization facility using the N-terminal His-tagged version (60 mg/ml). Protein (0.5 μl) was mixed with 0.5 μl of reservoir in a sitting drop vapor diffusion tray. The reservoir solution was composed of 16% (w/v) PEG 4000, 0.13 mM EDTA, and 0.1 M Tris–HCl (pH 8.5). Cubic shaped crystals were visible approximately three days after setting the drop and reached a full size in approximately one week. The crystals grew to approximately 0.2 mm × 0.2 mm × 0.2 mm and diffracted to 1.95 Å at UC Berkeley's Advanced Light Source (ALS), beamline 5.0.1. They belonged to space group *I*23 with a single molecule in the asymmetric unit. The dUTPase with the C-terminal His-tag was crystallized at the UCLA facility by mixing 1.5 μl of complex with 1.5 μl of reservoir in a sitting drop vapor diffusion tray. The reservoir solution was composed of 20% (w/v) PEG 3350 and 0.2 M KSCN (pH 7.0). Cubic shaped crystals appeared within one week. The crystals grew to approximately 0.2 mm × 0.2 mm × 0.2 mm. The crystals apparently belonged to space group *P*23 with a single molecule in the asymmetric unit. However, the crystals were diagnosed as merohedrally twinned using the UCLA twinning server.[38] The true space group was *R*3 with twinning fraction near 0.5. Because these crystals were twinned and diffracted poorly relative to the N-terminal His-tagged protein (3 Å), refinement was abandoned after receiving the LLNL data.

Crystallization of the dUTPase·Mg^{2+}·α,β-imido dUTP complex

The complex was prepared by mixing 10 μl of 100 mg/ml dUTPase (N-terminal His-tagged version = 17.9 kDa; C-terminal His-tagged version = 17.7 kDa) with 10 μl of 10 mM α,β-imido dUTP (Jena Bioscience), yielding a protein : ligand molar ratio of approximately 1 : 2. Both His-tagged versions of the protein were used in crystallization. Both yielded crystals under the same conditions, though they belonged to different space groups. Complex (1.5 μl) was mixed with 1.5 μl of reservoir in a hanging drop vapor diffusion tray. The reservoir solution was composed of 16% PEG 3350 and 0.33 M $Mg(NO_3)_2$. The dUTPase with the N-terminal His-tag produced a crystal visible approximately 11 days after setting the drop and reached a full size in approximately two weeks. The crystal grew to approximately 0.3 mm × 0.3 mm × 0.3 mm. It belonged to space group $P6_3$ with a single molecule in the asymmetric unit. It diffracted to 1.3 Å using a wavelength of 0.9786 Å at UC Berkeley's Advanced Light Source (ALS), beamline 8.2.2. The crystal was diagnosed as merohedrally twinned using the UCLA twinning server.[38] The approximate twin fraction was 0.16. The dUTPase with the C-terminal His-tag produced a crystal in approximately one week. The crystal grew to approximately 0.3 mm × 0.2 mm × 0.1 mm. It belonged to space group $P2_12_12_1$ with a single molecule in the asymmetric unit.

Crystallization of the dUTPase·Mg^{2+}·dUDP complex

The complex was prepared and crystallized in a manner analogous to the dUTPase-α,β-imido dUTP–Mg^{2+} complex, using the same reservoir conditions. The N-terminal His-tagged version of dUTPase was used. The ligand, dUDP, was purchased from Jena Bioscience as a 10 mM solution. Hexagonal rod shaped crystals grew to approximately 0.4 mm × 0.2 mm × 0.2 mm shape. The crystals belonged to space group $P3_121$.

Crystallization of the dUTPase·dUTP complex

This complex was prepared by soaking crystals of the unliganded enzyme with dUTP. To prepare enzyme crystals, 1.5 μl of dUTPase (60 mg/ml) was mixed with 1.5 μl of reservoir in a hanging drop vapor diffusion tray. The reservoir solution was composed of 10% PEG 3350 and 0.40 M NH_4NO_3. Crystals were visible approximately three days after setting the drop and reached a full size in approximately one week. The crystals grew

to approximately 0.4 mm × 0.3 mm × 0.2 mm having the appearance of a plate with an ellipsoidal face. They belonged to space group $P2_12_12_1$ with a trimer in the asymmetric unit. To soak in the ligand, an artificial mother liquor was prepared containing 10% PEG 3350, 400 mM NH_4NO_3 and 40% of a solution containing 100 mM dUTP. Unliganded enzyme crystals were transferred to the artificial mother liquor and soaked for overnight at room temperature.

Crystallization of the dUTPase·Cr^{3+}·dUTP complex

This complex was prepared by soaking crystals of the unliganded enzyme with dUTP-Cr^{3+}. To prepare enzyme crystals, 1.5 μl of dUTPase (60 mg/ml) was mixed with 1.5 μl of reservoir in a hanging drop vapor diffusion tray. The reservoir solution was composed of 10% PEG 3350 and 0.40 M NH_4NO_3. Crystals were visible approximately three days after setting the drop and reached a full size in approximately one week. The crystals grew to approximately 0.4 mm × 0.3 mm × 0.2 mm having the appearance of a plate with an ellipsoidal face. They belonged to space group $P2_12_12_1$ with a trimer in the asymmetric unit. To soak in the ligand, an artificial mother liquor was prepared containing 10% PEG 3350, 400 mM NH_4NO_3 and 40% of a solution containing 100 mM-Cr^{3+}·dUTP. The later solution was prepared by mixing $CrCl_3{\cdot}6H_2O$ with dUTP (Sigma). Unliganded enzyme crystals were transferred to the artificial mother liquor and soaked for one hour at room temperature. The crystal was cryo-protected by swiping the crystal through a 65% : 35% mixture of the same artificial mother liquor and glycerol.

Crystallization of the dUTPase·Mg^{2+}·dUMP complex

The complex was prepared by mixing dUTPase (N-terminal His-tag) (60 mg/ml) with dUTP (Sigma) in a 1 : 5 molar ratio and allowing the hydrolysis reaction to proceed to completion. Complex (1.5 μl) was mixed with 1.5 μl of reservoir in a sitting drop vapor diffusion tray. The reservoir solution was composed of 16% PEG 3350 and 0.22 M $Mg(NO_3)_2$. A crystal was visible approximately three days after setting the drop and reached a full size in approximately two weeks. The crystal grew to approximately 0.4 mm × 0.4 mm × 0.2 mm. It belonged to space group $P2_12_12_1$ with a trimer in the asymmetric unit. It diffracted to 1.85 Å at National Synchrotron Light Source (NSLS), beamline x8c, Brookhaven National Laboratory.

X-ray data collection and processing

Diffraction data were collected on the unliganded dUTPase enzyme at the Advanced Light Source (ALS), beamline 5.0.1 equipped with an ADSC Quantum 4 CCD detector. Data were collected on the Mg^{2+}·dUMP complex at the National Synchrotron Light Source at Brookhaven National Laboratory using beamline X8C and a Quantum 4 CCD detector. Data were collected on the dUTP, Mg^{2+}·dUDP, and Cr^{3+}·dUTP complexes using a Rigaku FR-D rotating anode generator equipped with R-AXIS IV++ detector. Data were collected on Mg^{2+}·α,β-imido dUTP complexes at ALS beamline 8.2.2 equipped with an ADSC Quantum 315 CCD detector. Data collection was performed at 100 K. Unless stated otherwise above, crystals were cryo-protected in a mixture of 65% reservoir solution and 35% glycerol by a quick dunk, then flash freezing. Data were processed and reduced using Denzo/Scalepack from the HKL suite of programs.[39]

Structure solution and refinement

Molecular replacement operations were performed with the program EPMR.[40] The initial molecular replacement solution for the unliganded *M. tuberculosis* dUTPase was obtained using *E. coli* dUTPase as a model. The refined *M. tuberculosis* dUTPase coordinates were then used as a molecular replacement model to solve the complexes in other space groups. Initial crystallographic refinement was performed for all dUTPase complexes using the CNS suite of programs.[41] Specifically, simulated annealing was performed, followed by conjugate gradient refinement and *B*-factor refinement. To aid in the medium resolution refinement of the Mg^{2+}·dUDP complex, hydrogen bonding restraints were included during refinement.[42] Model building was performed with the aid of the graphics program "O".[43] To model anisotropic displacements of the molecules, TLS parameterization[44] was implemented with the program REFMAC[45] for the following complexes: Mg^{2+}·dUMP, dUTP, and Mg^{2+}·α,β-imido-dUTP (in space group $P2_12_12_1$). Diffraction data collected on the dUTPase·Mg^{2+}·α,β-imido dUTP complex were tested positive for merohedral twinning using the UCLA twinning server.[38] To benefit from twin fraction refinement and individual anisotropic temperature refinement at 1.3 Å resolution, the structure was refined using SHELXL.[46] The twin operator used was $k, h, -l$. The twin fraction was refined to 16% by SHELXL. All structures were validated with ERRAT,[47] PROCHECK,[48] and WHATIF[49] structure validation tools. The program, ERRAT, reported overall quality factors of 97.5% or higher. PROCHECK reported no Ramachandran outliers. All structural superpositions were performed using the program MAPS.[19] Protein structures were illustrated using the program PyMOL.[50]

Protein Data Bank accession numbers

Atomic co-ordinates and structure factors have been deposited in the Protein Data Bank. Codes for the individual complexes are listed in Table 1.

Acknowledgements

For technical assistance in protein preparation we thank Beom-Seop Rho. For assistance in data collection, we thank the staff at the National Synchrotron Light Source at Brookhaven National Laboratory and the staff at the Advanced Light Source at Berkeley Laboratory. We thank David Eisenberg, Todd Yeates, and Thomas Terwilliger for advice and support. This work was funded by grants from the DOE, NIH, and HHMI.

References

1. Persson, R., Cedergren-Zeppezauer, E. S. & Wilson, K. S. (2001). Homotrimeric dUTPases; structural

solutions for specific recognition and hydrolysis of dUTP. *Curr. Protein Pept. Sci.* **2**, 287–300.
2. McIntosh, E. M. & Hanes, R. H. (1997). dUTP pyrophosphatase as a potential target for chemotherapeutic drug development. *Acta Biochim. Pol.* **44**, 159–171.
3. Grässer, F. A., Romeike, B. F. M., Niedobitek, G., Nicholls, J. & Kremmer, E. (2001). dUTPase in human neoplastic cells as a potential target for therapeutic intervention. *Curr. Protein Pept. Sci.* **2**, 349–360.
4. Ladner, R. D. (2001). The role of dUTPase and uracil-DNA repair in cancer chemotherapy. *Curr. Protein Pept. Sci.* **2**, 361–370.
5. Corbett, E. L., Watt, C. J., Walker, N., Maher, D., Williams, B. G., Raviglione, M. C. & Dye, C. (2003). The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch. Intern. Med.* **163**, 1009–1021.
6. Gadsden, M. H., McIntosh, E. M., Game, J. C., Wilson, P. J. & Haynes, R. H. (1993). dUTP pyrophosphatase is an essential enzyme in *Saccharomyces cerevisiae*. *EMBO J.* **12**, 4425–4431.
7. Mol, C. D., Harris, J. M., McIntosh, E. M. & Tainer, J. A. (1996). Human dUTP pyrophosphatase: uracil recognition by a β hairpin and active sites formed by three separate subunits. *Structure*, **4**, 1077–1092.
8. Prasad, G. S., Stura, E. A., Elder, J. H. & Stout, C. D. (2000). Structures of feline immunodeficiency virus dUTP pyrophosphatase and it nucleotide complexes in three crystal forms. *Acta Crystallog. sect. D*, **56**, 1100–1109.
9. Prasad, G. S. (2001). Glycine rich P-loop motif in deoxyuridine pyrophosphatase. *Curr. Protein Pept. Sci.* **2**, 301–311.
10. Cedergren-Zeppezauer, E. S., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. S. (1992). Crystal structure of a dUTPase. *Nature*, **355**, 740–743.
11. Larsson, G., Svensson, L. A. & Nyman, P. O. (1996). Crystal structure of the *Escherichia coli* dUTPase in complex with substrate analogue (dUDP). *Nature Struct. Biol.* **3**, 532–538.
12. Dauter, Z., Wilson, K. S., Larsson, G., Nyman, O. P. & Cedergren-Zeppezauer, E. S. (1998). The refined structure of dUTPase from *Escherichia coli*. *Acta Crystallog. sect. D*, **54**, 735–749.
13. Gonzalez, A., Larsson, G., Persson, R. & Cedergren-Zeppezauer, E. S. (2001). Atomic resolution structure of *Escherichia coli* dUTPase determined *ab initio*. *Acta Crystallog. sect. D*, **57**, 767–774.
14. Dauter, Z., Persson, R., Rosengren, A. M., Nyman, O. P., Wilson, K. S. & Cedergren-Zeppezauer, E. S. (1999). Crystal structure of dUTPase from equine infectious anaemia virus; active site metal binding in a substrate analogue complex. *J. Mol. Biol.* **285**, 655–673.
15. Persson, T., Larsson, G. & Nyman, P. O. (1996). Synthesis of 2′-deoxyuridine 5′-(α,β-imido)triphosphate: a substrate analogue and potent inhibitor of dUTPase. *Bioorg. Med. Chem.* **4**, 553–556.
16. Vertessy, B. G., Larsson, G., Perssoc, T., Bergman, A.-C., Persson, R. & Nyman, P. O. (1998). The complete triphosphate moiety of non-hydrolyzable substrate analogues is required for a conformational shift of the flexible C-terminus in *E. coli* dUTP pyrophosphatase. *FEBS Letters*, **421**, 83–88.
17. Kyte, J. (1995). Metalloenzymes. In *Mechanism in Protein Chemistry*, Garland Publisher, New York p. 353.
18. Arndt, J. W., Gong, W., Zhong, X., Showalter, A. K., Liu, J., Dunlap, C. A. *et al.* (2001). Insight into the catalytic mechanism of DNA polymerase β: structures of intermediate complexes. *Biochemistry*, **40**, 5368–5375.
19. Zhang, Z., Lindstam, M., Unge, J., Peterson, C. & Lu, G. (2003). Potential for dramatic improvement in sequence alignment against structures of remote homologous proteins by extracting structural information from multiple structure alignment. *J. Mol. Biol.* **332**, 127–142.
20. McClure, M. A. (2001). Evolution of the DUT Gene: horizontal transfer between host and pathogen in all three domains of life. *Curr. Protein Pept. Sci.* **2**, 313–324.
21. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
22. Li, R., Sirawaraporn, R., Chitnumsub, P., Sirawaraporn, W., Wooden, J., Athappily, F. *et al.* (2000). Three-dimensional structure of *M. tuberculosis* dihydrofolate reductase reveals opportunities for the design of novel tuberculosis drugs. *J. Mol. Biol.* **295**, 307–323.
23. Li de la Sierra, I., Munier-Lehmann, H., Gilles, A. M., Bârzu, O. & Delarue, M. (2001). X-ray structure of TMP kinase from *Mycobacterium tuberculosis* complexed with TMP at 1.95 Å resolution. *J. Mol. Biol.* **311**, 87–100.
24. Munier-Lehmann, H., Chaffotte, A., Pochet, S. & Labesse, G. (2001). Thymidylate kinase of *Mycobacterium tuberculosis*: a chimera sharing properties common to eukaryotic and bacterial enzymes. *Protein Sci.* **10**, 1195–1205.
25. Shi, W., Basso, L. A., Santos, D. S., Tyler, P. C., Furneaux, R. H., Blanchard, J. S. *et al.* (2001). Structures of purine nucleoside phosphorylase from *Mycobacterium tuberculosis* in complexes with immucillin-H and its pieces. *Biochemistry*, **40**, 8204–8215.
26. Gamieldien, J., Ptitsyn, A. & Hide, W. (2002). Eukaryotic genes in *Mycobacterium tuberculosis* could have a role in pathogenesis and immunomodulation. *Trends Genet.* **18**, 5–8.
27. Liang, J., Edelsbrunner, H. & Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7**, 1884–1897.
28. Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J. & Pelletier, H. (1997). Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: evidence for an induced fit mechanism. *Biochemistry*, **36**, 11205–11215.
29. Larsson, G., Nyman, P. O. & Kvassman, J.-O. (1996). Kinetic characterization of dUTPase from *Escherichia coli*. *J. Biol. Chem.* **271**, 24010–24016.
30. Vertessy, B. G. (1997). Flexible glycine rich motif of *Escherichia coli* deoxyurdine triphosphate nucleotidohydrolase is important for functional but not for structural integrity of the enzyme. *Proteins: Struct. Funct. Genet.* **28**, 568–579.
31. Nord, J., Kiefer, M., Adolph, H.-W., Zeppezauer, M. M. & Nyman, P.-O. (2000). Transient kinetics of ligand binding and role of the C-terminus in the dUTPase from equine infectious anemia virus. *FEBS Letters*, **472**, 312–316.
32. Shao, H., Robek, M. D., Threadgill, D. S., Mankowski, L. S., Cameron, C. E., Fuller, F. J. & Payne, S. L. (1997). Characterization and mutational

studies of equine infectious anemia virus dUTPase. *Biochim. Biophys. Acta*, **1339**, 181–191.
33. Harris, J. M., McIntosh, E. M. & Muscat, G. E. O. (1999). Structure/function analysis of a dUTPase: catalytic mechanism of a potential chemotherapeutic target. *J. Mol. Biol.* **288**, 275–287.
34. Maegley, K. A., Admiraal, S. J. & Herschlag, D. (1996). Ras-catalyzed hydrolysis of GTP: a new perspective from model studies. *Proc. Natl Acad. Sci. USA*, **93**, 8160–8166.
35. Nord, J., Nyman, P.-O., Larsson, G. & Drakenberg, T. (2001). The C-terminus of dUTPase: observation on flexibility using NMR. *FEBS Letters*, **492**, 228–232.
36. Looman, A. C., Bodlaender, J., Comstock, L. J., Eaton, D., Jhurani, P., de Boer, H. A. & Knippenberg, P. H. (1987). Influence of the codon following the AUG initiation codon on the expression of a modified lacZ gene in *Escherichia coli*. *EMBO J.* **6**, 2489–2492.
37. Coligan, J. E.; Dunn, B. M.; Ploegh, G. H.; Speicher, D. W.; Wingfield, P. T. (eds) (1995). Production of recombinant proteins. In *Current Protocols in Protein Science*, pp. 5.3.13, Wiley, Brooklyn, NY.
38. Yeates, T. O. (1997). Detecting and overcoming crystal twinning. *Methods Enzymol.* **276**, 344–358.
39. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 620–633.
40. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). Rapid automated molecular replacement by evolutionary search. *Acta Crystallog. sect. D*, **55**, 484–491.
41. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
42. Fabiola, F., Bertram, R., Korostelev, A. & Chapman, M. S. (2002). An improved hydrogen bond potential: impact on medium resolution protein structures. *Protein Sci.* **11**, 1415–1423.
43. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
44. Schomaker, V. & Trueblood, K. N. (1998). Correlation of internal torsional motion with overall molecular motion in crystals. *Acta Crystallog. sect. B*, **54**, 507–514.
45. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallog. sect. D*, **57**, 122–133.
46. Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: high-resolution refinement. *Methods Enzymol.* **277**, 319–343.
47. Colovos, C. & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–1519.
48. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
49. Vriend, G. & Sander, C. (1993). Quality control of protein models: directional atomic contact analysis. *J. Appl. Crystallog.* **26**, 47–60.
50. DeLano, W. L. (2002). *The PyMOL User's Manual*, DeLano Scientific, San Carlos, CA, USA.

Edited by I. Wilson

(Received 23 March 2004; received in revised form 7 June 2004; accepted 8 June 2004)