Crystal structure of dUTP pyrophosphatase from feline immunodeficiency virus

G. SRIDHAR PRASAD, E.A. STURA, D.E. McREE, G.S. LACO, C. HASSELKUS-LIGHT, J.H. ELDER, AND C.D. STOUT

Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California 92037-1093

Abstract

We have determined the crystal structure of dUTP pyrophosphatase (dUTPase) from feline immunodeficiency virus (FIV) at 1.9 Å resolution. The structure has been solved by the multiple isomorphous replacement (MIR) method using a $P6_3$ crystal form. The results show that the enzyme is a trimer of 14.3 kDa subunits with marked structural similarity to *E. coli* dUTPase. In both enzymes the C-terminal strand of an anti-parallel β-barrel participates in the β-sheet of an adjacent subunit to form an interdigitated, biologically functional trimer. In the $P6_3$ crystal form one trimer packs on the 6_3 screw-axis and another on the threefold axis so that there are two independent monomers per asymmetric unit. A Mg^{2+} ion is coordinated by three asparate residues on the threefold axis of each trimer. Alignment of 17 viral, prokaryotic, and eukaryotic dUTPase sequences reveals five conserved motifs. Four of these map onto the interface between pairs of subunits, defining a putative active site region; the fifth resides in the C-terminal 16 residues, which is disordered in the crystals. Conserved motifs from all three subunits are required to create a given active site. With respect to viral protein expression, it is particularly interesting that the gene for dUTPase (DU) resides in the middle of the Pol gene, the enzyme cassette of the retroviral genome. Other enzymes encoded in the Pol polyprotein, including protease (PR), reverse transcriptase (RT), and most likely integrase (IN), are dimeric enzymes, which implies that the stoichiometry of expression of active trimeric dUTPase is distinct from the other Pol-encoded enzymes. Additionally, due to structural constraints, it is unlikely that dUTPase can attain an active form prior to cleavage from the polyprotein.

Keywords: dUTP pyrophosphatase; enzyme structure; Feline Immunodeficiency Virus; subunit assembly

FIV is a lentivirus associated with an AIDS-like disease in domestic cats (Pedersen et al., 1987; Yamamoto et al., 1988; Ackely et al., 1990). FIV, as well as other non-primate lentiviruses including equine infectious anemia virus (EIAV) and visna/maedi viruses of sheep, encodes dUTPase (DU) within its pol gene in addition to the protease (PR), reverse transcriptase (RT), and integrase (IN) enzymes common to all retroviruses (McClure et al., 1988; McGeoch, 1990; Elder et al., 1992). Interestingly, the primate lentiviruses, including HIV-1, HIV-2, and the SIVs do not encode dUTPase, nor do type C retroviruses (Elder et al., 1992). However, both type B and type D retroviruses (McClure et al., 1988) encode a dUTPase in a distinct location between Gag and Pol (Elder et al., 1992). Additionally, other viruses such as the herpesviruses (McGeoch, 1990) and certain pox viruses (Slabaugh and Roseman, 1989) encode dUTPase.

The enzyme dUTPase (E.C. 3.6.1.23) converts dUTP into dUMP, producing a pivotal substrate for the synthesis of dTTP. Perhaps more importantly, it maintains the cellular dUTP concentration at a low level, which limits the incorporation of uracil into DNA.

Actively dividing and activated cells have relatively high levels of endogenous dUTPase activity to facilitate accurate replication of cellular DNA. Thus, uracil incorporation does not occur to a significant level during the replication of an invading virus in such an environment, even if the virus does not encode dUTPase (Lerner et al., 1995). However, in particular with FIV, EIAV, and visna/maedi lentiviruses, the presence of an active dUTPase is necessary for productive infection of primary macrophages (Threadgill et al., 1993; Lerner et al., 1995; Steagall et al., 1995; Turelli et al., 1996), where the endogenous level of dUTPase is low (Lerner et al., 1995). Infection of cats with DU^- FIV results in a fivefold increase in the number of mutations in the genome of virus produced in primary macrophages, primarily consisting of G→A transitions (Lerner et al., 1995). DU^- FIV is still able to propagate with the same efficiency as wild-type FIV in T lymphocytes (Lerner et al., 1995). The ultimate outcome on the *in vivo* infection and pathogenesis is still under investigation. However, it is clear the DU gene stabilizes the replication of the virus in certain cell populations and thus directly influences viral host cell range.

The purpose of the present study was to define the three-dimensional structure of dUTPase from FIV. The structure provides a basis for comparing the retroviral enzyme to the known structure of *E. coli* dUTPase (Cedergren-Zeppezauer et al., 1992)

and serves as a starting point for substrate binding and mutagenesis experiments that can be used to study the enzyme mechanism. We wish to better understand the catalytic properties of the enzyme to facilitate the development of specific and general dUTPase inhibitors that may prove useful for treatment of certain viral diseases as well as proliferative disorders. Further, the enzyme mechanism is of interest because other well-studied nucleotide pyrophosphatases, such is adenylate kinase (Müller et al., 1996), cleave the β-γ phosphodiester bond, whereas dUTPase cleaves the α-β phosphodiester bond.

Results and discussion

The experimentally derived electron density map for FIV dUTPase is shown in Figure 1. Using this map a model was derived for each of two independent 14.3 kDa molecules in the asymmetric unit. The structure has been refined to 1.9 Å resolution (Table 1). The enzyme forms a trimer of closely packed subunits in which the β-strand at the N-terminus of one molecule hydrogen bonds with a β-strand at the C-terminus of a neighboring molecule (Fig. 2). Consequently, each subunit not only shares side-to-side contacts with its neighbors, but also contributes and receives an element of secondary structure. The threefold symmetric trimers containing the two molecules within the asymmetric unit pack onto the 6_3-screw and threefold axes of the unit cell. These trimers alternate contacts in the *ab* plane and in the *c* direction, creating intervening voids along the threefold axis and a tightly packed arrangement along the 6_3-screw axis. The maximum dimensions of the trimer are 42 Å along the threefold axis of the particle and 55 Å in the plane normal to the threefold where the N- and C-terminal strands interdigitate; the trimer tapers to 30 Å at the opposite surface. Overall, the crystal structure indicates that FIV dUTPase is a trimer in solution and that this is the functional state of the enzyme.

The polypeptide fold of the individual subunits is comprised of 13 β-strand segments and a single two-turn α-helix (Fig. 3a). The β-strands form two sheets of a β-barrel; two additional strands are inserted orthogonal to the β-barrel. Strands 2, 3, 12, 7, and 9 form an anti-parallel "front" sheet; strands 8, 10, 5/6, 1, and 13 form the "back" sheet and are anti-parallel with each other with the exception that strands 1 and 13 are parallel, and strand 13 is the C-terminal segment provided by a neighboring subunit (Fig. 2). Strands 4 and 11, inserted between strands of the front and back sheets, are anti-parallel. Although residues 110–114 of strand 13 are hydrogen bonded with strand 1, 16 residues beyond residue 117 are not visible in the electron density in this crystal form in either independent trimer and are presumed to be disordered. The two independent molecules in the asymmetric unit are very similar with an rms deviation between 117 Cα atoms of 0.37 Å.

The subunit fold is closely related to the eight-stranded canonical jelly-roll fold of β-barrel proteins (Fig. 3b). However, it differs in three ways: strands β1 and β2 are not directly interacting; strand β1 is parallel with strands β2 and β8 rather than being anti-parallel; and strand β8 is provided by a separate subunit, as noted. With respect to strand β3, β1 replaces β8 of the idealized jelly-roll fold. This overall topology is very similar to that observed for *E. coli* dUTPase both for the monomer and the trimer (Cedergren-Zeppezauer et al., 1992). For 103 homologous Cα atoms in all 13 β-strand segments the rms deviation between the FIV and *E. coli* structures is 1.01 Å. Given the high degree of sequence identity between FIV dUTPase and other dUTPases (below) it is expected that all viral, prokaryotic, and eukaryotic dUTPases share this common fold and exhibit a trimeric assembly.

A search of the National Center for Biological Information data base using the primary sequence of FIV dUTPase yielded the sequences for 16 dUTPases from other viral, prokaryotic, and eukaryotic sources. In all pairwise comparisons of these 17 sequences the
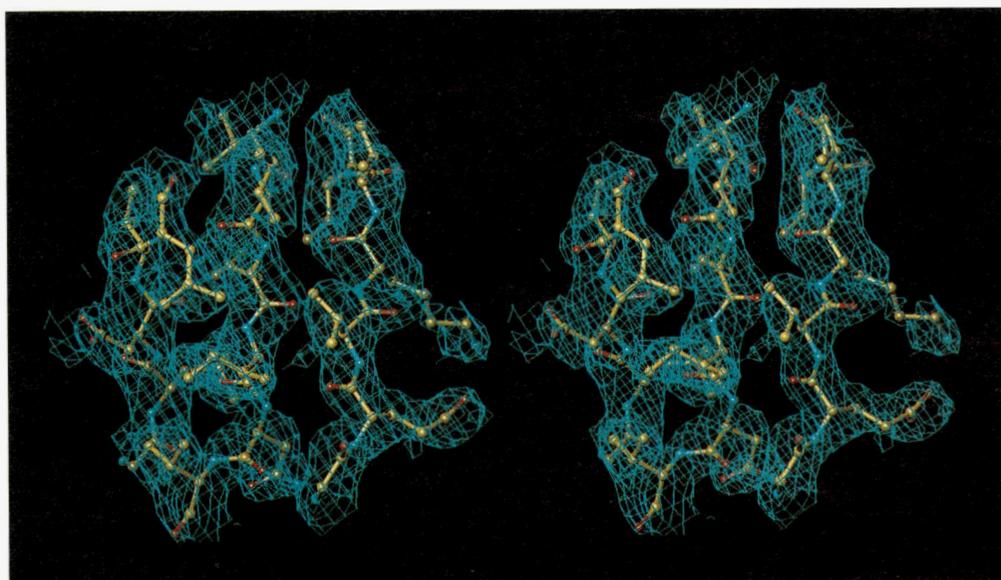
Fig. 1. The experimental electron density for FIV dUTPase. The map is computed using all data in the resolution range 20.0–2.2 Å and phases extended from 2.7 Å resolution MIR-derived phases by solvent flattening and twofold non-crystallographic symmetry averaging. The density shown is contoured at 1.2σ and corresponds to β strands 8, 10, and 5 (left to right in the stereo image): strand 8, residues 62–66, bottom to top, GlyLeuAspValLeu; strand 10, residues 80–85, top to bottom, ValIleMetIleAsnVal; strand 5, residues 31–36, bottom to top, GlyGluValLysValIle.

Table 1. *Data collection and refinement parameters*

Data sets	Resolution (Å)	R_{symm}[a] (%)	No. of unique reflections	Total no. of reflections	Completeness (%)	R_{iso}[b] (%)	R_c[c] or R_k[d] (%)	Phasing power[e] (Resolution, Å)	No. of sites
Native	1.9	6.4	23,522	116,665	96.3	—	—	—	—
Native	2.1	4.3	18,011	63,076	98.1	—	—	—	—
EMTS[f]	2.4	6.9	10,915	69,068	86.8	18.0	53.6	2.90 (2.7)	2
EMP[g]	2.8	7.4	6,237	23,770	77.8	18.6	37.2	2.97 (2.7)	2
$SmCl_3 \cdot 6H_2O$[h]	2.7	7.4	7,641	78,526	85.5	12.9	60.0	1.08 (2.7)	1
$Gd_2(SO_4)_3 \cdot 8H_2O$[i]	2.7	6.9	7,612	40,222	86.7	12.3	50.2	1.11 (2.7)	1
$Yb_2(SO_4)_3 \cdot 8H_2O$[j]	2.7	4.3	7,885	39,537	90.9	9.1	61.1	1.08 (2.7)	1
EMTS (ano)	2.4						8.3	1.91 (2.7)	2

Overall figure of merit 0.671 for 7,421 phased reflections at 2.7 Å

R-factor (for all data 7.0-1.9 Å) 20.9%
R-free 24.9%
RMS deviations from ideality:
  Bond lengths 0.015 Å
  Bond angles 2.4°
Average isotropic B-values:
  Molecule A (888 atoms) 25.0 Å²
  Molecule B (888 atoms) 46.4 Å²
  H_2O molecules (123 atoms) 42.0 Å²
  $Mg^{2+}(A)$, $Mg^{2+}(B)$ 11.0, 54.8 Å²

[a] $R_{symm} = \sum_{hkl}\sum_i |I_{i(hkl)} - \langle I_{(hkl)}\rangle| / \sum_{hkl}\sum_i I_{i(hkl)}$.
[b] $R_{iso} = \sum_{hkl} ||F_{pH}| - |F_p|| / \sum_{hkl} |F_p|$, where F_{pH} and F_p are the derivative and native structure factor amplitudes, respectively.
[c] $R_c = \sum_{hkl} |(|F_{pH}| - |F_p|) - |F_{Hc}|| / \sum_{hkl} ||F_{pH}| - |F_p||$, for centric reflections of isomorphous derivatives where F_{Hc} is the calculated heavy atom structure factor amplitude.
[d] $R_k = \sum_{hkl} ||F_{pH^+}| - |F_{pHc^+}|| + ||F_{pH^-}| - |F_{pHc^-}|| / \sum_{hkl} |F_{pH^+}| + |F_{pH^-}|$ for anomalous scattering.
[e] Phasing power $= \langle F_{Hc}\rangle$/residual; where $\langle F_{Hc}\rangle$ is the rms mean heavy atom contribution; and the residual is defined as $[|F_{pHc}|^2 - |F_{pH}|^2/N]^{1/2}$, the difference between the calculated and observed derivative structure factor amplitudes.
[f] Ethyl mercuric thiosalicylic acid (EMTS).
[g] Ethyl mercuric phosphate (EMP).
[h] Samarium chloride hexahydrate.
[i] Gadolinium sulphate octahydrate.
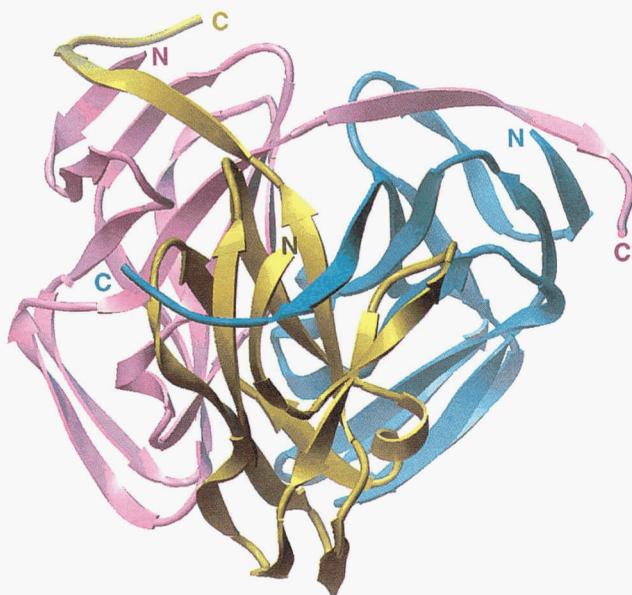[j] Ytterbium sulphate octahydrate.

Fig. 2. Ribbon representation of the FIV dUTPase trimer. The threefold axis is tilted forward from the plane of figure.

percent identities range from 22–97%. The sequences range in length from 133 to 196 residues with FIV dUTPase being the shortest. All 17 can be aligned in their central regions; the results of this alignment are summarized in Figure 4. The alignment reveals five conserved segments, or motifs, each of which contain residues conserved in the 17 dUTPases, as originally observed by McGeoch (1990). Adjacent to these positions are semi-conserved residues, which are the same in at least 15 of the 17 available sequences, or are conserved for the type of amino acid (Leu, Ile, Val; Phe, Tyr; Arg, Lys), or contain glycine and just one other type of amino acid (Asp; Ala). This strong conservation of amino acid sequence implies functional significance and provides a means to identify the active site of the enzyme.

Three of the motifs (I, II, IV) cluster together on the "front" surface of the β-barrel, while the motif III lies on the "back" surface (Fig. 3b). However, in the trimer this arrangement places motif III opposite motifs I, II, IV at the interface between subunits and causes the conserved residues to be clustered around a shallow depression (Fig. 5a). Within this depression is a distinct pocket lined with conserved residues, indicating that this is the active site (Fig. 5b) (below). This region has recently been identified as the active site of *E. coli* dUTPase on the basis of three sigma difference electron density observed in a crystal soaked with dUDP (Larsson et al., 1996).
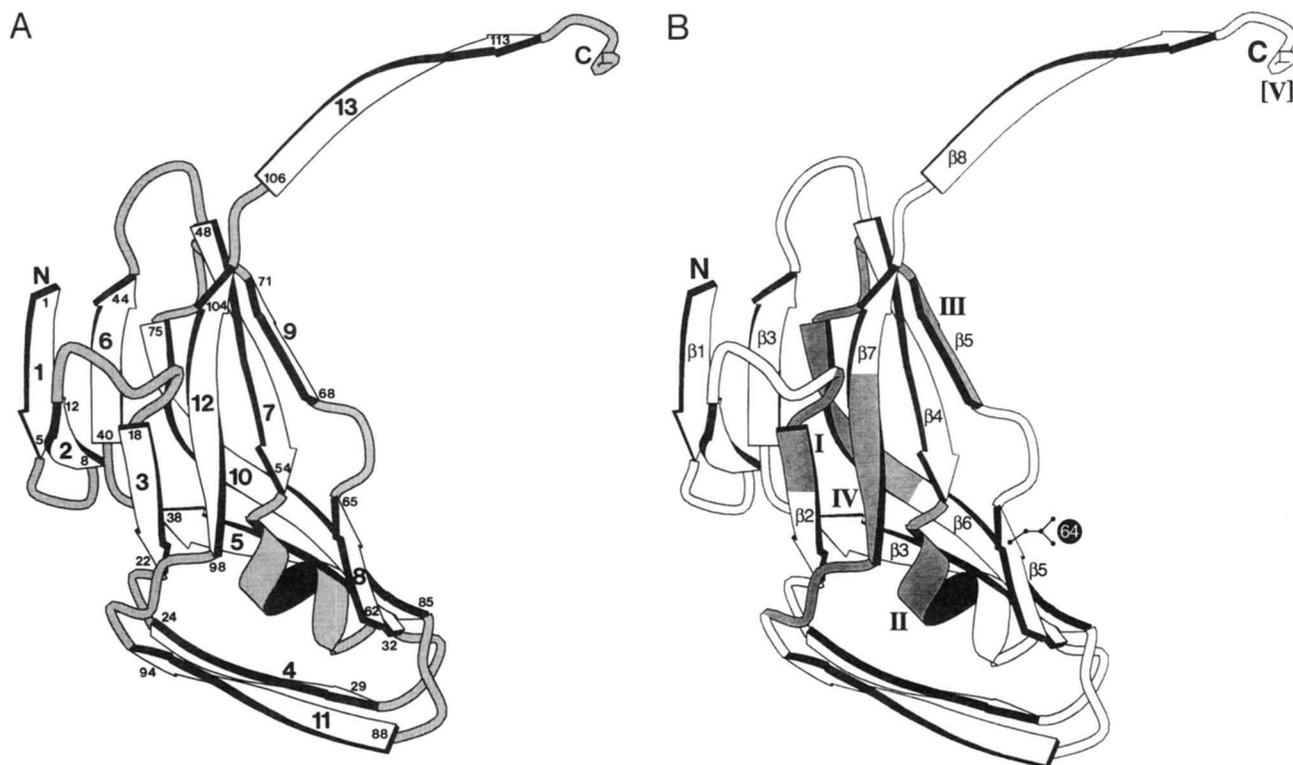
Fig. 3. Ribbon representation of the FIV dUTPase monomer. (**A**) The 13 β strand segments and corresponding residue numbers. (**B**) The 8 β strand segments of the canonical jelly-roll β-barrel fold and the five motifs of conserved residues superposed on the subunit (shaded areas). In both (**A**) and (**B**) the C-terminal strand (13/β8) of another subunit is parallel and adjacent to the N-terminal strand (1/β1).

Residues of motif V are not visible in the electron density, but the chain terminus at residue 117 is oriented such that residues 118–133 could easily interact with those of motifs I–IV, or participate in binding substrate. It appears that motif V may interact with dUTP because it contains a conserved ArgGlyxxGlyxGly sequence with flanking Ser and Thr residues reminiscent of the conserved P-loop sequence, GlyxxGlyxGlyLysThr, found in many ATP and GTP binding enzymes (Saraste et al., 1990). For example, in adenylate kinase the P-loop forms an anion hole (Dreusicke and Schulz, 1986) involved in binding the triphosphate moiety of ATP (Müller and Schulz, 1992; Berry et al., 1994). Another feature of the P-loop and the anion hole in adenylate kinase is an electropositive helical dipole. In FIV dUTPase the N-terminal end of the only helix contains motif II and is oriented toward the active site region (Fig. 3b). Therefore, it appears that all five conserved motifs participate in the formation of the enzyme active site. Furthermore, motifs from all three subunits of the trimer are required to form a given active site. For example, in Figure 2, motifs I, II, and IV of the pink subunit (whose orientation is the same as in Fig. 3b) converge with motif III of the yellow subunit, while motif V from the blue subunit is directed toward this same interface region.

The structure of the FIV dUTPase subunit is compatible with other aspects of the sequence alignment. Semi-conserved residues outside of the five motifs (Fig. 4) are involved in hydrophobic packing interactions in the core of the β-barrel (Ile 36, Gly 50, Ile 52, Leu 63, Leu 92), or hydrogen bonds stabilizing the tertiary structure (Thr 38, Asn 84). The semi-conserved hydrophobic residues have contacts with each other as well as with conserved residues, including Leu 20, Ile 78, Val 80, Leu 101, and Ile 102 (e.g., Fig. 1). There are seven sites where deletions or insertions occur in FIV dUTPase with respect to other dUTPases (Fig. 4). Six of these correspond to turns on the surface of the subunit between elements

Fig. 4 (*facing page*). Alignment of the primary sequences of 17 dUTPases. Residue numbers above the lines correspond to FIV; numbers at the end of the lines correspond to the full length proteins. Black boxes are conserved residues in all 17 sequences; open boxes are semi-conserved residues; together these cluster into five motifs (Roman numerals) (McGeoch, 1990). An asterisk indicates the Mg²⁺ ligand in FIV. The sequences and their National Center for Biological Information accession codes are as follows: FIV, feline immunodeficiency virus, P16088; FIV-puma, puma lentivirus, U03982; EIAV, equine infectious anemia virus, P03371; VISNA, visna virus, JQ1162; CAEV, caprine arthritis-encephalitis virus, M33677; VACCV, vaccinia virus, P17374; MMTV, mouse mammary tumor virus, M15122; ADENO-avian, avian adenovirus, Z17216; HAEIN, hemophilus influenzae virus, P43792; E. coli, *Escherichia coli*, P06968; Coxiella, *Coxiella burnetii*, S44300; Myco-leprae, *Mycobacterium leprae*, U15181; Yeast-sacc, *Saccharomyces cerevisiae*, P33317; Yeast-mito, *Saccharomyces cerevisiae* mitochondrion, S38189; Yeast-candida, *Candida albicans*, P43058; Tomato, *Lycopersicon esculentam*, P32518; Human, *Homo sapiens*, P33316.
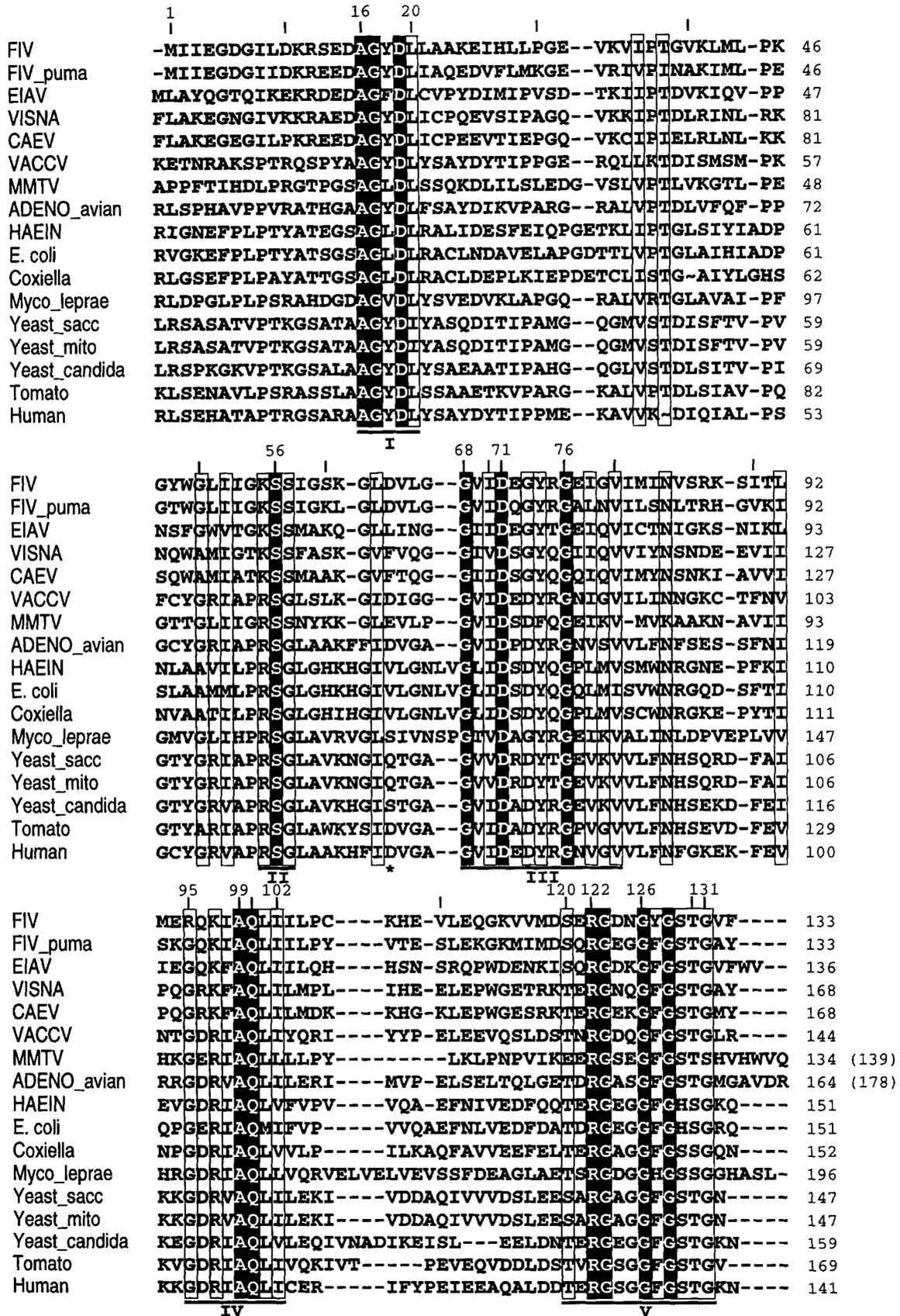
```
                   1                16   20
                   |                |    |              |              |
FIV           -MIIEGDGILDKRSEDAGYDLLAAKEIHLLPGE--VKVIPTGVKLML-PK  46
FIV_puma      -MIIEGDGIIDKREEDAGYDLIAQEDVFLMKGE--VRIVPINAKIML-PE  46
EIAV          MLAYQGTQIKEKRDEDAGFDLCVPYDIMIPVSD--TKIIPTDVKIQV-PP  47
VISNA         FLAKEGNGIVKKRAEDAGYDLICPQEVSIPAGQ--VKKIPTDLRINL-RK  81
CAEV          FLAKEGEGILPKREEDAGYDLICPEEVTIEPGQ--VKCIPIELRLNL-KK  81
VACCV         KETNRAKSPTRQSPYAAGYDLYSAYDYTIPPGE--RQLLKTDISMSM-PK  57
MMTV          APPFTIHDLPRGTPGSAGLDLSSQKDLILSLEDG-VSLVPTLVKGTL-PE  48
ADENO_avian   RLSPHAVPPVRATHGAAGYDLFSAYDIKVPARG--RALVPTDLVFQF-PP  72
HAEIN         RIGNEFPLPTYATEGSAGLDIRALIDESFEIQPGETKLIPTGLSIYIADP  61
E. coli       RVGKEFPLPTYATSGSAGLDLRACLNDAVELAPGDTTLVPTGLAIHIADP  61
Coxiella      RLGSEFPLPAYATTGSAGLDLRACLDEPLKIEPDETCLISTG-AIYLGHS  62
Myco_leprae   RLDPGLPLPSRAHDGDAGVDLYSVEDVKLAPGQ--RALVRTGLAVAI-PF  97
Yeast_sacc    LRSASATVPTKGSATAAGYDIYASQDITIPAMG--QGMVSTDISFTV-PV  59
Yeast_mito    LRSASATVPTKGSATAAGYDIYASQDITIPAMG--QGMVSTDISFTV-PV  59
Yeast_candida LRSPKGKVPTKGSALAAGYDLYSAEAATIPAHG--QGLVSTDLSITV-PI  69
Tomato        KLSENAVLPSRASSLAAGYDLSSAAETKVPARG--KALVPTDLSIAV-PQ  82
Human         RLSEHATAPTRGSARAAGYDLYSAYDYTIPPME--KAVVK-DIQIAL-PS  53
                                       I

                    56        68  71      76
                    |         |   |       |               |
FIV           GYWGLIIGKSSIGSK-GLDVLG--GVIDEGYRGEIGVIMINVSRK-SITL  92
FIV_puma      GTWGLIIGKSSIGKL-GLDVLG--GVIDQGYRGALNVILSNLTRH-GVKI  92
EIAV          NSFGWVTGKSSMAKQ-GLLING--GIIDEGYTGEIQVICTNIGKS-NIKL  93
VISNA         NQWAMIGTKSSFASK-GVFVQG--GIIVDSGYQGIIQVVIYNSNDE-EVII 127
CAEV          SQWAMIATKSSMAAK-GVFTQG--GIIDSGYQGQIQVIMYNSNKI-AVVI 127
VACCV         FCYGRIAPRSGLSLK-GIDIGG--GVIDEDYRGNIGVILINNGKC-TFNV 103
MMTV          GTTGLIIGRSSNYKK-GLEVLP--GVIDSDFQGEIKV-MVKAAKN-AVII  93
ADENO_avian   GCYGRIAPRSGLAAKFFIDVGA--GVIDPDYRGNVSVVLFNFSES-SFNI 119
HAEIN         NLAAVILPRSGLGHKHGIVLGNLVGLIDSDYQGPLMVSMWNRGNE-PFKI 110
E. coli       SLAAMMLPRSGLGHKHGIVLGNLVGLIDSDYQGQLMISVWNRGQD-SFTI 110
Coxiella      NVAATILPRSGLGHIHGIVLGNLVGLIDSDYQGPLMVSCWNRGKE-PYTI 111
Myco_leprae   GMVGLIHPRSGLAVRVGLSIVNSPGIVDAGYRGEIKVALINLDPVEPLVV 147
Yeast_sacc    GTYGRIAPRSGLAVKNGIQTGA--GVVDRDYTGEVKVVLFNHSQRD-FAI 106
Yeast_mito    GTYGRIAPRSGLAVKNGIQTGA--GVVDRDYTGEVKVVLFNHSQRD-FAI 106
Yeast_candida GTYGRVAPRSGLAVKHGISTGA--GVIDADYRGEVKVVLFNHSEKD-FEI 116
Tomato        GTYARIAPRSGLAWKYSIDVGA--GVIDADYRGPVGVVLFNHSEVD-FEV 129
Human         GCYGRVAPRSGLAAKHFIDVGA--GVIDEDYRGNVGVVLFNFGKEK-FEV 100
                    II                *         III

               95   99 102             120 122  126    131
               |    |  |                |   |    |      |
FIV           MERQKIAQLIILPC----KHE-VLEQGKVVMDSERGDNGYGSTGVF---- 133
FIV_puma      SKGQKIAQLIILPY----VTE-SLEKGKMIMDSQRGEGGFGSTGAY---- 133
EIAV          IEGQKFAQLIILQH----HSN-SRQPWDENKISQRGDKGFGSTGVFWV-- 136
VISNA         PQGRKFAQLILMPL----IHE-ELEPWGETRKTERCNQGFGSTGAY---- 168
CAEV          PQGRKFAQLILMDK----KHG-KLEPWGESRKTERGEKGFGSTGMY---- 168
VACCV         NTGDRIAQLIYQRI----YYP-ELEEVQSLDSTNRGDQGFGSTGLR---- 144
MMTV          HKGERIAQLLLPY--------LKLPNPVIKEERGSEGFGSTSHVHWVQ 134 (139)
ADENO_avian   RRGDRVAQLILERI----MVP-ELSELTQLGETDRGASGFGSTGMGAVDR 164 (178)
HAEIN         EVGDRIAQLVFVPV----VQA-EFNIVEDFQQTERGEGGFGHSGKQ---- 151
E. coli       QPGERIAQMIFVP-----VVQAEFNLVEDFDATDRGEGGFGHSGRQ---- 151
Coxiella      NPGDRIAQLVVLP-----ILKAQFAVVEEFELTERGAGGFGSSGQN---- 152
Myco_leprae   HRGDRIAQLLVQRVELVELVELVEVSSFDEAGLAETSRCDGGHGSSGGHASL- 196
Yeast_sacc    KKGDRVAQLILEKI----VDDAQIVVVDSLEESARGAGGFGSTGN----- 147
Yeast_mito    KKGDRVAQLILEKI----VDDAQIVVVDSLEESARGAGGFGSTGN----- 147
Yeast_candida KEGDRIAQLVLEQIVNADIKEISL---EELDNTERGEGGFGSTGKN---- 159
Tomato        KVGDRIAQLIVQKIVT-----PEVEQVDDLDSTVRGSGGFGSTGV----- 169
Human         KKGDRIAQLICER-----IFYPEIEEAQALDDTERGSGGFGSTGKN---- 141
                   IV                              V
```

Fig. 4. *See caption on facing page.*

of secondary structure and would not be expected to alter the tertiary folding. A seventh site at Gly 67 contains a two residue insertion in *E. coli* and only three other dUTPases; it is accommodated as a larger bulge between strand segments 8 and 9 (Fig. 3a). None of the deletion or insertion sites occur within the five conserved motifs.

Interactions between subunits stabilizing the formation of the trimer do not in general involve conserved residues. The total accessible surface area buried in the trimer is 5400 Å^2, or 1800 Å^2 per monomer (Connoly, 1983). Additional surface area would be expected to be buried when residues 118–133 are ordered. Strand 1 and residues Lys 11 and Ser 13 on the front side of the subunit (Fig. 3a) have nine hydrogen bonds and two hydrophobic contacts with residues 109–116 (strand 13), including four hydrogen bonds involving Met 1. Motifs I and III are in direct contact at the subunit interface via a hydrogen bond between the amide of Ala 16 and side chain of Asp 71, both conserved residues, and via the methyl group of Ala 16, which stacks on Trp 49, and the side chain of Tyr 18, which hydrogen bonds to amide and carbonyl of Glu 109. Strands 7, 8, 10, and 12 provide residues that line a hydrophobic channel on the threefold axis (Fig. 5a). In particular, there are close contacts of Trp 49 with Ile 102 and Leu 51 with Leu 53. Other residues lining this channel are Pro 30, Gly 31, Val 65, Leu 66, Ile 83, Val 85, Leu 104, and Pro 105. In addition to these hydrophobic residues, Asp 64 on strand 8 of each subunit is buried deeply in the channel and coordinated to a Mg^{2+} ion (Figs. 3b, 5a).

While Mg^{2+} is presumed to be required for binding substrate, the structure reveals the presence of Mg^{2+} on the threefold axis of trimer A as a 9σ peak in a 2.1 Å resolution difference Fourier map. This site is surrounded by three 8σ peaks (H_2O molecules) and the Asp 64 Oδ1 atoms from each subunit making a symmetric octahedral coordination sphere (Fig. 6). Each of the Mg^{2+}-oxygen

distances is 2.0 ± 0.1 Å. Since the crystallization medium contains 2.9 mM $MgCl_2$, Mg^{2+} has been included in the model (Table 1). The Oδ2 atoms of Asp 64 and the amide and carbonyl of Val 65 participate in a hydrogen bonding network (Fig. 6). Two additional H_2O molecules on the threefold axis are hydrogen bonded with the coordinated H_2O molecules and Asp 64 Oδ1 atoms, such that the hydrogen bonding potential of the polar atoms is satisfied. Given the presence of Asp 64 within the hydrophobic environment it is probable that Mg^{2+} is required for stabilization of the FIV dUTPase trimer. An Asp or Glu residue occurs at the homologous position in 7 of 18 dUTPases; in the remaining sequences it is Val, Leu, Phe, Ser, or Gln (Fig. 4).

In trimer B the Mg^{2+} site appears as a 5σ peak in the 2.1 Å resolution difference Fourier map. The lower peak height, or equivalently larger temperature factor (Table 1), is due to crystal packing; trimer A is closely packed along the *c* direction, whereas trimer B has no such contacts. This difference is reflected in the average B-factor for the subunits and the number of tightly bound H_2O waters (seven per subunit in trimer A with B < 20 Å^2, none in trimer B). The crystal packing also accounts for the observed substitution of Sm^{3+}, Gd^{3+}, and Yb^{3+} for Mg^{2+} on the crystallographic threefold axis (trimer B) but not on the 6_3-axis (trimer A) (Table 1). The substitution of lanthanides for Mg^{2+} at the Asp 64 ligands in trimer B is consistent with the interpretation of a Mg^{2+} site in trimer A. In contrast, the Hg compounds bind equally well to Cys 106 of each trimer at a site proximal to the active site (Fig. 5b) and also within 7 Å of the His 108.

Within the shallow depression between subunits (Fig. 5a) is a pocket lined with conserved residues (Fig. 5b), including Ala 16, Gly 17, Asp 19, Ser 56, Gly 68, Asp 71, Ala 99, and Gln 100. Lys 55 and Lys 97 contribute to the formation of the pocket as well; these two residues are semi-conserved as Lys or Arg in 17
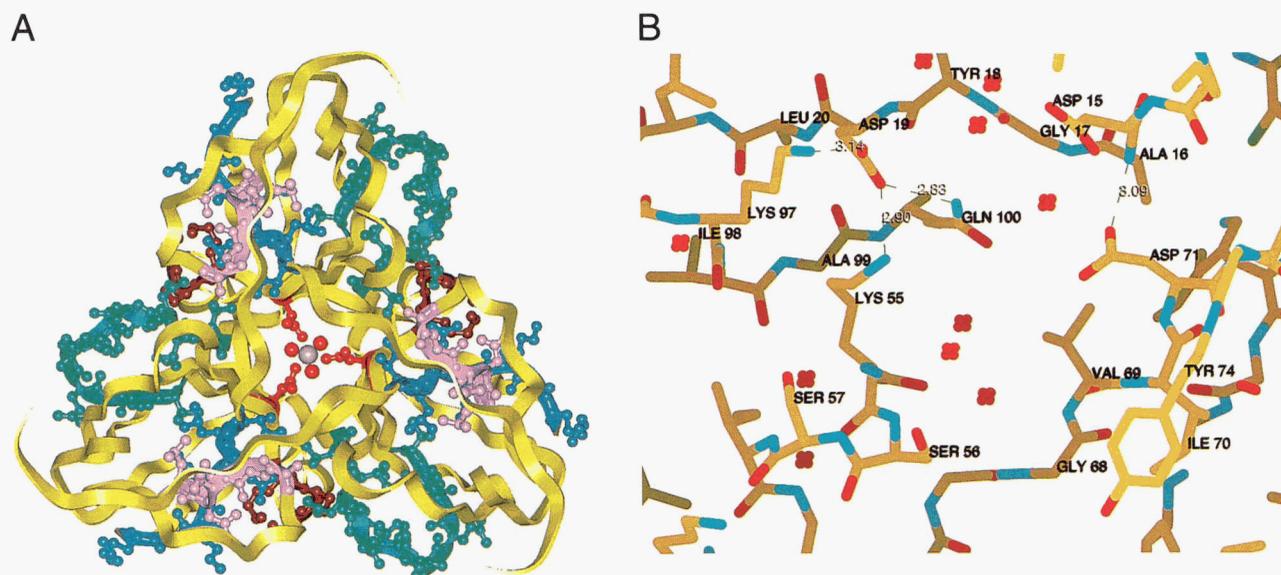

Fig. 5. (**A**) Ribbon representation of the trimer of FIV dUTPase viewed along the threefold axis. Residues within sequence conserved motifs cluster around the active sites at the interfaces between subunits. All atoms within conserved motifs I–IV (Fig. 4) are shown; motif V is not ordered in the structure. The motifs are colored: I, lavender; II, dark red; III, turquoise-green; IV, blue. The Mg^{2+} ion bound on the threefold axis is also shown coordinated by the side chains of Asp 64 and three water molecules. (**B**) Conserved residues surrounding the active site pocket from motifs I–IV. Residues from motif III (lower right) are from a separate subunit of the trimer. Four hydrogen bonds involving conserved residues are indicated. The Hg ligand Cys 106 is in the upper right-hand corner. The atoms are colored: C, yellow; N, blue; O, red; S, green. Crosses indicate bound water molecules.
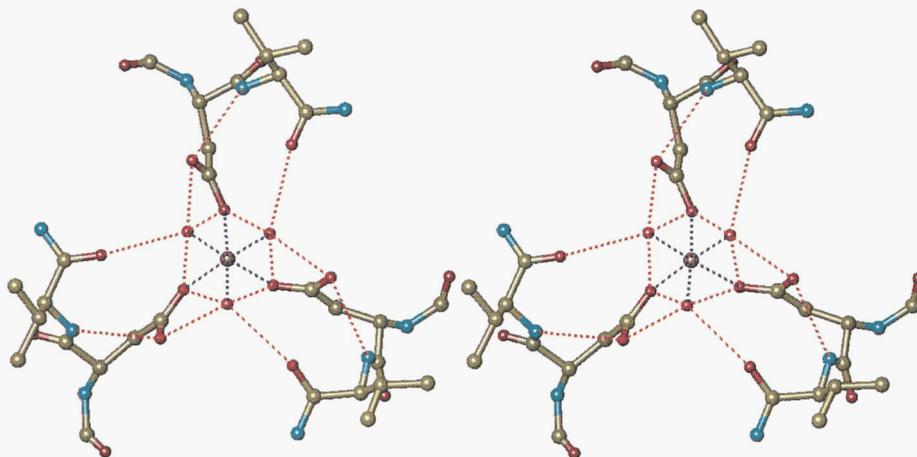
Fig. 6. Stereo figure of the coordination geometry of Mg^{2+} (silver sphere) and the hydrogen bonding network involving bound water molecules (red spheres), and Asp 64 and Val 65 of threefold-related subunits. Hydrogen bonds are depicted with dashed red lines and presumed Mg–O bonds are depicted as dashed blue lines.

dUTPase sequences (Fig. 4). Additional conserved or semi-conserved residues at the perimeter of this pocket are Ser 57, Ile 70, Tyr 74, and Gly 76. Within the pocket there are four hydrogen bonds between conserved residues: Lys 55 and Lys 97 interact with each oxygen of the Asp 19 side chain, and this residue hydrogen bonds to Gln 100 (Fig. 5b); Ala 16 and Asp 71 are also hydrogen bonded, as noted. Together, these interactions are presumed to be important for precisely orienting the side chains of conserved residues in the active site pocket, imparting to the enzyme a high degree of specificity in binding the uracil and deoxyribose moieties of the substrate. Binding of the triphosphate of dUTP may occur closer to the surface of the trimer if the P-loop–like residues of motif V and the helix dipole of motif II are involved.

In view of the observed structure for FIV dUTPase and the presence of sequence conserved hydrogen bonds in the active site, the reported structure for the dUDP complex of *E. coli* dUTPase is difficult to rationalize (Larsson et al., 1996). The model proposes that only two conserved residues make direct hydrogen bonds to dUDP: Gln 119 (Gln 100) to the α-phosphate, and Asp 90 (Asp 71) to deoxyribose. Further, the recognition of uracil is modeled to be provided by Asn 84 (Gly 67) and the main chain atoms of Met 98 (Gly 79). Neither of these residues are conserved in other dUTPases, and as noted Asn 84, is part of a larger bulge between structurally homologous chain segments due to a two residue insertion in *E. coli* dUTPase. Clearly, it is necessary to study the structures of substrate complexes of FIV dUTPase as well.

Materials and methods

Viruses, plasmid, and cell lines

The DU gene encoded by molecularly cloned FIV-34TF10 (Talbott et al., 1989) was utilized in these studies. Primers were prepared to coincide with the N- and C-terminus of DU, as defined previously (Wagaman et al., 1993). A Nde1 site encompassing the initiator methionine at the N-terminus of DU was engineered into the plus strand primer (bases 3988–4014, 5′-GTCAAACACATATGATAAT AGAAGGGG-3′) and the complement of a stop codon as well as an EcoR1 site was inserted immediately following the C-terminus in the negative strand primer (complement of bases 4389–4422, 5′-CCTGAATTCTGTCAACCCAAGACTAGAATACTC-3′). The gene was then amplified by 25 cycles of polymerase chain reaction (PCR) using Amplitaq (Promega), with one minute denaturation at

94 °C, followed by two minute annealing at 60 °C, and two minute elongation at 72 °C per cycle. The PCR product was then digested with Nde1 and Eco R1 and cloned into the pUC112 Nde vector (a gift of S. Hughes). After nucleotide sequence analyses to verify the fidelity of the amplification, the insert was subcloned into the pT7-7 vector (a gift of S. Tabor) and comparisons were performed to determine the relative levels of expression of each construct. DU in pUC112 Nde was expressed in *E. coli* DH5 cells as described (Wagaman et al., 1993) and the pT7-7 construct was expressed in BL21.DE3 cells (Studier et al., 1990). In addition, an Xba1-Bam H1 fragment containing the inserted DU was cleaved from the pT7-7 construct and ligated into the pBlueBac II baculovirus transfer vector (Invitrogen) cut with Nhe1-Bam H1. Usable crystals were obtained from both constructs, but we chose to pursue the baculovirus-produced enzyme, since the level of expression and ease of purification was greater from insect cells.

A derivative of the TN5 cell line from *Trichoplusia ni* (Summers and Smith, 1987) was prepared that was adapted to growth in shaker culture in serumless medium (unpublished results). Briefly, cells were grown initially in EXCELL 401 medium containing 10% fetal bovine serum (FCS) with constant shaking at 100 rpm at 28 °C. Baffled flasks were employed to prevent the cells from clumping. After several weeks, cells began to proliferate with a doubling time of approximately 17 h. After several passages, the cells were then gradually weaned onto EXCELL 401 without FCS and after a short lag time, recovered to the original growth rate. The cells (TN5-JE) were then employed for high-level protein expression, using high titer stocks of recombinant baculovirus prepared in SF9 insect cells (Wickham et al., 1992). The latter stocks were prepared by co-transfecting with DU-containing transfer vector and linearized viral DNA (BaculoGold, PharMingen) according to the manufacturer's recommendations. Virus recovered from these cells was amplified at low multiplicity of infection in SF9 cells for subsequent infection of TN5-JE cells.

Purification of dUTPase

For baculovirus expression, TN5-JE cells (2×106/mL) were infected with recombinant baculovirus and allowed to incubate at 28 °C for 45 min without shaking; the cells were then returned to shaking and harvested after 48–52 h growth at 28 °C. Typically, 30–50% of the cells failed to exclude trypan blue at this time. The cells were recovered by centrifugation, then resuspended in

500 mL (per liter of cell culture) 20 mM Tris-HCl, pH 8, 2 mM EDTA, 0.5% NP-40, 150 mM NaCl, and 1 mM dithiothreitol (DTT) and frozen at −80 °C. The lysates were stored for future use or re-thawed to facilitate lysis and processed for dUTPase preparation. The lysate was clarified by centrifugation at 8000 × g for 30 min and the pellet was discarded. The supernatant was then diluted with water to the conductivity of the input buffer, adjusted to pH7.5, 5 mM MgCl₂, and 2 mM 2-mercaptoethanol (final volume, 1 liter). The supernatant was then passed over an Affigel Blue Sepharose Fast Flow column (Pharmacia, 2.0 × 20 cm) at a flow rate of 16 mL/min and the column was washed in 50 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, 2 mM 2-mercaptoethanol. dUTPase was eluted using a 0–500 mM NaCl linear gradient in the same buffer. Peak fractions were identified by immunoblot analyses, pooled, and concentrated to 1/10 volume using Centripreps (Amicon). The sample was diluted 10-fold and applied to a Resource Q (Pharmacia) column (1.5 × 10 cm) in the same buffer used for Affigel Blue chromatography. dUTPase was eluted using a 0–300 mM linear NaCl gradient, and peak fractions were identified by immunoblot analyses. Peak fractions were pooled, diluted fivefold with 10 mM Tris-MOPS buffer, pH 6.5, and the pH adjusted by addition of concentrated HCl. The sample was then applied to a Resource S (Pharmacia) column (1.5 × 10 cm) equilibrated in 10 mM Tris-MOPS, pH 6.5, and washed with the same buffer. dUTPase was eluted using a 0–300 mM NaCl linear gradient and peak fractions were identified as above, pooled, and concentrated using centricon concentrators (Amicon). dUTPase was applied to a Superdex-75 column (Pharmacia) equilibrated in 50 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, 2 mM 2-mercaptoethanol, and eluted isocratically in the same buffer. Peak fractions were identified as above, concentrated, and protein concentration was determined by the Lowry method (Lowry et al., 1951). The protein was stored at 15–20 mg/mL at 4 °C. dUTPase was judged to be 98% pure by SDS-PAGE analysis and was readily amenable to crystallization, as detailed below. The yield was approximately 40 mg pure dUTPase per liter of infected cells.

Crystallization of FIV dUTPase

The purified baculovirus expressed enzyme in 50 mM Tris-HCl, 5 mM MgCl₂, 2 mM 2-mercaptoethanol at pH 7.5 was concentrated to 21 mg/mL. Crystals were grown by the sitting drop vapor diffusion method in a constant temperature incubator at 22.5 °C. Drops consisted of 3.5 μL protein solution with 2.5 μL reservoir solution of 13% MPEG 5 K, 100 mM sodium cacodylate, pH 6.5. The crystals appeared in about 50% of drops; crystals were obtained from the remaining drops by macroseeding. The crystals are hexagonal prisms 0.3 × 0.3 mm in size with variable length from 0.1 to 0.5 mm.

Structure Determination

The FIV dUTPase crystals belong to space group P6₃ with unit cell dimensions a = b = 79.93 and c = 87.63 Å. The asymmetric unit contains two molecules with a solvent content of 55% by volume. Intensity data for the 2.1 Å resolution native data were collected using a Mar Research image plate detector and Siemens SRA rotating anode X-ray generator equipped with a graphite monochromator and operated at 50 kV, 100 mA for CuKα radiation. The data were processed with the program Mosflm (Leslie, 1994). The data collection statistics are summarized in Table 1. Attempts to

solve the structure by molecular replacement using the structure of *E. coli* dUTPase (Cedergren-Zeppezauer, 1992) as a model were not successful. The structure was solved by the MIR method in combination with twofold non-crystallographic symmetry (ncs) averaging.

Five isomorphous heavy atom derivatives were prepared (Table 1). All the heavy atom soaking experiments were done using 2 mM heavy atom solutions and the crystals were soaked 16 h at room temperature. The intensity data for all the derivatives were collected using a Rigaku Ru200 rotating anode X-ray generator, monochromatic CuKα radiation, and a Siemens X-1000 area detector mounted on four-circle diffractometer. The data sets were processed using the XENGEN suite of programs (Howard et al., 1985). The Sm, Gd, and Yb derivatives have a single site in common on the crystallographic threefold axis. These derivatives were not included in the final MIR phase calculations because the Fourier maps were poorer in quality than those based on only the two Hg derivatives. However, the binding sites of the lanthanide ions were analyzed with respect to the native structure. All of the heavy atom sites were determined by using difference Patterson maps and self- and cross difference Fourier maps and the Xtalview suite of programs (McRee, 1992). The heavy atom parameters were refined and phase calculations were carried out using the program Phases (W. Furey, pers. comm.). The MIR phases were calculated to 2.7 Å (Table 1) and extended to 2.4 Å using dm (Cowtan, 1994). The solvent-flattened electron density was of sufficient quality for chain tracing and building a poly-alanine model for the molecule on the crystallographic 6₃-axis. The map quality was further improved by twofold ncs averaging (Kleyweyt and Jones, 1994). An initial mask was built using the poly-alanine model of the subunit on the 6₃-axis. The ncs operator relating the two subunits was determined by superposing the heavy atom sites in the asymmetric unit associated with the crystallographic 6₃- and threefold axes. Cycles of density averaging, refinement of the ncs operator, improvement of the mask, and extension of the resolution to 2.2 Å resulted in a map that could be used for unambiguous assignment of residues 1–100 in both independent molecules. The initial model with side chains was built using the program "O" (Jones et al., 1991) for the subunit on the 6₃-axis (molecule A). The ncs-related molecule on the threefold axis (molecule B) was generated by applying the refined ncs operator.

For refinement, a second native data set was collected to 1.9 Å resolution at Stanford Synchrotron Research Laboratory beam line 7-1 using monochromatic radiation (1.08 Å) and a Mar Research image plate detector. The data were collected from one crystal at 4 °C and processed using Mosflm with refined unit cell parameters a = b = 79.80 and c = 86.95 Å (Table 1). The structure was refined using XPLOR version 3.1 (Brünger et al., 1989). The two molecules were treated independently. $2|F_o| - |F_c|$ maps calculated with all the data in the resolution range 20.0–1.9 Å were used for adjusting and correcting the model. These maps revealed unambiguous electron density for an additional 17 C-terminal residues of both independent molecules, but no density was apparent for residues 118–133 of the primary sequence. The C-terminal residues 101–117 of both molecules were included in the model and $2|F_o| - |F_c|$ Fourier maps were used to model the bound Mg²⁺ ions and ordered water molecules (Table 1). For both molecules 95% of the residues are in the most favored regions of the Ramachandran plot; 5% are in additional allowed regions, or are glycine or proline (Laskowski et al., 1993). Coordinates have been deposited with the Protein Data Bank, accession codes 1DUT (molecule A) and 1DUU (molecule B).

Note added in proof

The structures of human dUTPase and its complexes with dUMP, dUDP, and dUTP have recently been published (Mol CD, Harris JM, McIntosh EM, Tainer JA. 1996. Human dUTP pyrophosphatase: Uracil recognition by a β hairpin and active sites formed by three separate subunits. *Structure 4*:1077–1092) in accordance with the results in this paper.

Acknowledgments

References

Ackley CD, Yamamoto JK, Levy N, Pedersen NC, Cooper MD. 1990. Immunologic abnormalities in pathogen-free cats experimentally infected with feline immunodeficiency virus. *J Virol 64*:5652–5655.

Berry MB, Meador B, Bilderback T, Liang P, Glaser M, Phillips GN Jr. 1994. The closed conformation of a highly flexible protein: The structure of *E. coli* adenylate kinase with bound AMP and AMPPNP. *Proteins Struct Funct Gen 19*:183–198.

Brünger AT, Karplus M, Petsko GA. 1989. Crystallographic refinement by simulated annealing: Application to crambin. *Acta Crystallogr A45*:50–61.

Cedergren-Zeppezauer ES, Larsson G, Nyman PO, Dauter Z, Wilson KS. 1992. Crystal structure of a dUTPase. *Nature 355*:740–743.

Connoly, ML. (1983). Solvent accessible surfaces of proteins and nucleic acids. *Science 221*:709–713.

Cowtan K. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D50*:760–763.

Dreusicke D, Schulz GE. 1986. The glycine-rich loop of adenylate kinase forms a giant anion hole. *FEBS J 208*:301–304.

Elder JH, Lerner DL, Hasselkus-Light CS, Fontenot DJ, Hunter E, Luciw PA, Montelaro RC, Phillips TR. 1992. Distinct subsets of retroviruses encode dUTPase. *J Virol 66*:1791–1794.

Howard AJ, Nielsen C, Xuong NH. 1985. Software for a diffractometer with multiwire area detector. *Methods Enzymol 114A*:452–472.

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A47*:110–119.

Kleyweyt GL, Jones TA. 1994. In: Bailey S, Hubbard R, Waller D, eds. *From first map to final model*. SERC Daresbury Laboratory. p 59.

Larsson G, Svensson LA, Nyman PO. 1996. Crystal structure of the *Escherichia coli* dUTPase in complex with a substrate analogue (dUDP). *Nat Struct Biol 3*:532–538.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.

Lerner DL, Wagaman PC, Phillips TR, Prospero-Garcia O, Henriksen SJ, Fox HS, Bloom FE, Elder JH. 1995. Increased mutation frequency of FIV lacking functional deoxyuridine triphosphatase. *Proc Natl Acad Sci USA 92*:7480–7484.

Leslie AGW. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D50*:760–763.

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. 1951. Protein measurement with the Folin phenol reagent. *J Biol Chem 193*:265–275.

McClure MA, Johnson MS, Feng D-F, Doolittle RF. 1988. Sequence comparisons of retroviral proteins: Relative rates of change and general phylogeny. *Proc Natl Acad Sci USA 85*:2469–2473.

McGeoch DJ. 1990. Protein sequence comparisons show that the "pseudoproteases" encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucleic Acids Res 18*:4105–4110.

McRee DE. 1992. A visual protein crystallographic software system for XII/Xview. *J Mol Graphics 10*:44–47.

Müller CW, Schlauderer GJ, Reinstein J, Schulz GE. 1996. Adenylate kinase motions during catalysis: An energetic counterweight balancing substrate binding. *Structure 4*:147–156.

Müller CW, Schulz GE. 1992. Structures of the complex between *Escherichia coli* adenylate kinase and the inhibitor Ap$_5$A refined at 1.9Å resolution. *J Mol Biol 224*:159–177.

Pedersen NC, Ho EW, Brown ML, Yamamoto JK. 1987. Isolation of a T-lymphotropic virus from domestic cats with an immunodeficiency-like syndrome. *Science 235*:790–793.

Saraste M, Sibbald PR, Wittinghofer A. 1990. The P-loop—A common motif in ATP- and GTP-binding proteins. *Trends Biosci 15*:430–434.

Slabaugh MB, Roseman NA. 1989. Retroviral protease-tike gene in the vaccinia virus genome. *Proc Natl Acad Sci USA 86*:4152–4155.

Steagall WK, Robek MD, Perry ST, Fuller FJ, Payne SL. 1995. Incorporation of uracil into viral DNA correlates with reduced replication of EIAV in macrophages. *Virology 210*:302–313.

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol 185*:60–89.

Summers MD, Smith GA. 1987. A manual of methods for baculovirus vectors and insect cell culture preparations. *Tex Agric Exp Stn Bull B1555*:1–56.

Talbott RL, Sparger EE, Lovelace KM, Fitch WM, Pedersen NC, Luciw PA, Elder JH. 1989. Nucleotide sequence and genomic organization of feline immunodeficiency virus. *Proc Natl Acad Sci USA 86*:5743–5747.

Threadgill DS, Steagall WK, Flaherty MT, Fuller FJ, Perry ST, Rushlow KE, Le Grice SFJ, Payne SL. 1993. Characterization of equine infectious anemia virus dUTPase: Growth properties of a dUTPase-deficient mutant. *J Virol 67*:2592–2600.

Turelli P, Pétursson G, Guiguen F, Mornex J-F, Vigne R, Quérat G. 1996. Replication properties of dUTPase-deficient mutants of caprine and ovine lentiviruses. *J Virol 70*:1213–1217.

Wagaman PC, Hasselkus-Light CS, Henson M, Lerner DL, Phillips TR, Elder JH. 1993. Molecular cloning and characterization of deoxyuridine triphosphatase from feline immunodeficiency virus (FIV). *Virology 196*:451–457.

Wickham TJ, Davis T, Granados RR, Shuler ML, Wood HA. 1992. Screening of insect cell lines for the production of recombinant proteins and infectious virus in the baculovirus expression system. *Biotechnol Prog 8*:391–396.

Yamamoto JK, Sparger E, Ho EW, Andersen PR, O'Connor TP, Mandell CP, Lowenstine L, Munn R, Pedersen NC. 1988. Pathogenesis of experimentally induced feline immunodeficiency virus infection in cats. *Am J Vet Res 49*:1246–1258.