Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structures of vaccinia virus dUTPase and its nucleotide complexes

Alexandra Samal,[a]‡ Norbert Schormann,[a]‡ William J. Cook,[b] Lawrence J. DeLucas[a] and Debasish Chattopadhyay[a,c]*

[a]Center for Biophysical Sciences and Engineering, University of Alabama at Birmingham, Birmingham, AL 35294, USA, [b]Department of Pathology, University of Alabama at Birmingham, Birmingham, AL 35294, USA, and [c]Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA

‡ These authors contributed equally.

Correspondence e-mail: debasish@cbse.uab.edu

Received 5 December 2006
Accepted 15 February 2007

PDB References: vaccinia virus dUTPase, 2okb, r2okbsf; 2okd, r2okdsf; 2oke, r2okesf; 2ol0, r2ol0sf; 2ol1, r2ol1sf.

Deoxyuridine triphosphate nucleotidohydrolase (dUTPase) catalyzes the hydrolysis of dUTP to dUMP and pyrophosphate in the presence of Mg^{2+} ions. The enzyme plays multiple cellular roles by maintaining a low dUTP:dTTP ratio and by synthesizing the substrate for thymidylate synthase in the biosynthesis of dTTP. Although dUTPase is an essential enzyme and has been established as a valid target for drug design, the high degree of homology of vaccinia virus dUTPase to the human enzyme makes the identification of selective inhibitors difficult. The crystal structure of vaccinia virus dUTPase has been solved and the active site has been mapped by crystallographic analysis of the apo enzyme and of complexes with the substrate-analog dUMPNPP, with the product dUMP and with dUDP, which acts as an inhibitor. Analyses of these structures reveal subtle differences between the viral and human enzymes. In particular, the much larger size of the central channel at the trimer interface suggests new possibilities for structure-based drug design. Vaccinia virus is a prototype of the poxviruses.

1. Introduction

Smallpox, which is caused by the variola virus, is one of the most devastating contagious diseases that mankind has ever faced. For centuries, repeated epidemics and the horrible aftereffects of this disease changed the course of history. The massive worldwide immunization campaign of the World Health Organization (WHO) successfully eradicated the disease in 1977. However, in the absence of vaccine-induced immunity, humans are universally susceptible to smallpox infection. As a result of the termination of routine vaccination programs, a large section of the US population has no immunity against smallpox. Despite the global impact of smallpox, it is striking that no effective treatment for smallpox has ever been developed. Undisclosed stocks of the virus are likely to exist in some parts of the world and therefore the potential use of smallpox as a biological weapon poses a serious threat to public health. Formal classification of smallpox as a category A (high priority) agent in the CDC's list of pathogens underscores the danger associated with smallpox and the need to expand our understanding of the virus and its life cycle. As part of a new research initiative for developing vaccines and therapeutics, we have undertaken a structural investigation of a number of proteins involved in the replication of vaccinia virus, a prototype of the smallpox virus.

Enzymes regulating DNA-repair mechanisms play a pivotal role in the survival of the virus and can therefore serve as potential drug targets. Two DNA-repair enzymes, deoxy-

uridine triphosphatase (dUTPase) and uracil DNA glycosylase (UDG or UNG), prevent the incorporation and retention of uracil in DNA, respectively. Uracil is added to DNA as a consequence of misincorporation by DNA polymerase, which cannot discriminate between deoxyuridine triphosphate (dUTP) and its natural substrate deoxythymidine triphosphate (dTTP). Spontaneous deamination of deoxycytidine bases also gives rise to uracil in DNA. Unless repaired, the existence of uracil bases in DNA can have detrimental effects, eventually leading to cell death. Therefore, both prevention of DNA incorporation and removal of misincorporated uracil from DNA are essential for the genomic integrity. dUTPase serves two major biological functions. By hydrolyzing dUTP it helps to maintain a low dUTP:dTTP ratio, thereby reducing the rate of dUMP incorporation into DNA. The second role of dUTPase is to generate the precursor dUMP for the biosynthesis of thymidine nucleotides.

Based on their molecular assembly, dUTPases can be grouped into three families. Mammalian and avian herpes viruses contain the monomeric form of the enzyme. Protozoan parasites and the bacterium *Campylobacter jejuni* encode a dimeric form of dUTPase. The third and most studied family of dUTPases contains the trimeric form found in eukaryotes, prokaryotes and viruses, including poxviruses. The monomeric form is believed to have evolved *via* gene duplication from a standard dUTPase-coding sequence of the trimeric form followed by a subsequent loss of one copy of each motif from the double-length chain (McGeehan *et al.*, 2001). These two forms have similar enzymatic properties. In contrast, the dimeric enzymes possess no similarity to members of the other classes in sequence, structure or enzymatic characteristics.

In general, the sequence homology among trimeric dUTPases is high. For poxviruses, the average sequence identity among dUTPases is more than 70% (the similarity is greater than 80%). The viral enzymes show 65% sequence identity with the human enzyme. These sequences are characterized by five well conserved motifs, particularly those responsible for sugar and nucleotide binding (McGeoch, 1990; Prasad *et al.*, 2000). In order to identify exploitable differences between the poxvirus dUTPase and its human counterpart, we have expressed, purified and crystallized vaccinia virus dUTPase (vvdUTPase) and mapped the active site of the enzyme by crystallizing it in a variety of states. Here, we describe the crystal structures of various protein–ligand complexes of full-length dUTPase (apo enzyme, a complex with the substrate analog dUMPNPP, a complex with the product dUMP and a complex with the inhibitor dUDP) and N-terminally truncated (ΔN11) dUTPase (apo enzyme with sulfate bound).

2. Materials and methods

2.1. Cloning and expression

We have cloned the coding sequence (AAO89320; Western Reserve 041) for full-length vvdUTPase and a shorter version (ΔN11) from which 11 N-terminal residues were deleted. The coding sequences for the full-length (1–147 residues) and ΔN11 (12–147 residues) versions of the dUTPase gene were amplified by PCR and engineered into the *Nde*I/*Xho*I restriction sites of pET15b (Novagen) vector. A single amino-acid substitution (Y28G) was identified in our constructs from comparison of the DNA sequence of the plasmid with that in the database (AAO89320; Western Reserve 041). The recombinant dUTPase proteins were expressed in *Escherichia coli* BL21(DE3)pLysS cells. Bacterial cultures were grown at 310 K in 1 l Luria–Bertani broth containing 0.02% glucose, 50 μg ml^{-1} ampicillin and 34 μg ml^{-1} chloramphenicol to an OD_{600} of 0.8 and dUTPase expression was induced by addition of isopropyl β-D-thiogalactopyranoside to a final concentration of 0.4 m*M*. The bacterial culture of full-length dUTPase was grown at 310 K for 4 h post-induction, while the culture of ΔN11 dUTPase was grown at room temperature for 12 h. Both the full-length and ΔN11 proteins contained an N-terminal hexahistidine tag and a thrombin cleavage site for removal of the tag.

2.2. Purification

The bacterial cell pellet was suspended in 50 ml 50 m*M* HEPES buffer pH 7.8, 300 m*M* NaCl, 5 m*M* β-mercaptoethanol (β-ME), 0.1 m*M* phenylmethylsulfonyl fluoride and 2 m*M* benzamidine hydrochloride. The cells were lysed by two freeze–thaw cycles. The lysate was then treated with 1 mg DNase I for 30 min at 277 K. All subsequent steps were carried out at 277 K. The cell extracts were centrifuged at 39 000*g* for 30 min to separate soluble and insoluble fractions. The supernatant was loaded onto a 7 ml Talon column (BD Biosciences) containing immobilized Co^{2+} ions and pre-equilibrated with the lysis buffer. To remove nonspecifically bound proteins, the column was washed with 20 column volumes of 50 m*M* HEPES buffer pH 7.8, 300 m*M* NaCl, 5 m*M* β-ME, 10 m*M* imidazole. The enzyme was eluted with a linear imidazole gradient (10–300 m*M*) in 15 column volumes of elution buffer (50 m*M* HEPES pH 7.8, 300 m*M* NaCl, 5 m*M* β-ME). Fractions containing dUTPase were pooled and dialyzed overnight at 277 K against 50 m*M* HEPES buffer pH 7.8, 100 m*M* NaCl and 5 m*M* β-ME. The hexahistidine tag was cleaved using thrombin (1 unit per milligram of dUTPase). For further purification, the digestion mixture was loaded onto a 15 ml SP Sepharose ion-exchange column (GE Healthcare) equilibrated with 50 m*M* HEPES buffer pH 7.5, 10 m*M* NaCl, 5 m*M* β-ME. The protein without the hexahistidine tag was obtained in the column flowthrough and, after concentration, was subjected to size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column (GE Healthcare) using 20 m*M* HEPES buffer pH 7.5, 100 m*M* NaCl and 10 m*M* dithiothreitol. The protein eluted in a single peak as a trimer. SDS–PAGE analysis revealed that the peak fractions contained a homogeneous preparation of dUTPase. Fractions containing dUTPase were pooled and the protein was concentrated to 15 mg ml^{-1}. Aliquots of enzyme were stored at 193 K until further use.

2.3. Crystallization

Crystallization conditions for both forms of protein were identified by screening commercial crystallization kits using the hanging-drop vapor-diffusion technique. Crystals of the ΔN11 protein (6–15 mg ml^{-1}) were initially obtained in 0.1 *M* phosphate–citrate buffer pH 4.2, 0.2 *M* lithium sulfate with 20% PEG 1000 as precipitant (reagent No. 39 from Wizard I; Emerald Biosciences). After optimization, the best crystals were grown in phosphate–citrate buffer pH 4.2, 0.1 *M* lithium sulfate with 12% PEG 1000 as precipitant. These crystals belong to the hexagonal space group $P6_3$, with unit-cell parameters $a = b = 105.2$, $c = 59.3$ Å. The asymmetric unit contains a homotrimer ($V_M \simeq 2.1$ Å^3 Da^{-1}, corresponding to a solvent content of 41%).

Crystals of full-length protein (15 mg ml^{-1}) were grown in 0.1 *M* Tris buffer pH 7.0 using 20% PEG 2000 monomethyl ether as precipitant (reagent No. 10 from Wizard I). Prior to crystallization, 5–10 m*M* $MgCl_2$ was added to the protein solution. These crystals belong to the hexagonal space group $P6_5$, with unit-cell parameters $a = b = 120.0$, $c = 50.3$ Å. The asymmetric unit contains a homotrimer ($V_M \simeq 2.1$ Å^3 Da^{-1}, corresponding to a solvent content of 41%).

Crystals of the full-length protein with the non-hydrolyzable substrate analog α,β-imido-dUTP (dUMPNPP) were obtained by soaking the apo-enzyme crystals in the presence of 1.7 m*M* dUMPNPP. The product complex with dUMP was prepared by allowing the enzyme to react with its substrate dUTP (7 m*M*) in the presence of 5 m*M* $MgCl_2$ at room temperature prior to crystallization. Crystals of the dUTPase–inhibitor complex were grown by pre-incubating the protein with 2 m*M* dUDP in the presence of 5 m*M* $MgCl_2$.

For cryoprotection, crystals were soaked for 1 min each in five steps in solutions containing increasing concentrations (5–27.5%) of ethylene glycol (EDO) in reservoir solutions before freezing. No ligand was added to any cryosolution.

2.4. Data collection and processing

Intensity data for were collected at 100 K on a Rigaku X-ray generator using R-AXIS IV and R-AXIS IV^{++} image-plate detectors. For data processing we used the *d*TREK* (Pflugrath, 1999) and *HKL*-2000 (Otwinowski & Minor, 1997) program packages. Data-collection statistics are listed in Table 1.

2.5. Structure solution and refinement

The structure of vaccinia virus ΔN11 dUTPase was solved by molecular replacement with *MOLREP* (Vagin & Teplyakov, 1997) and *Phaser* (Storoni *et al.*, 2004) using apo human dUTPase (PDB code 1q5u; Mol *et al.*, 1996) as the search model. Subsequently, we used the structure of the ΔN11 version to solve the structure of the full-length protein. The coordinates of the full-length apo enzyme were used as a starting model in rigid-body refinement of the complexes. The asymmetric unit for both forms of vvdUTPase contains three molecules related by noncrystallographic symmetry.

Model building was performed using *QUANTA* (Accelrys Inc.) and *Coot* (Emsley & Cowtan, 2004). Initial rounds of refinement for all five structures were carried out applying restrained noncrystallographic symmetry using the *CNS* program package (Brünger *et al.*, 1998). Finally, restrained refinement by maximum likelihood in *REFMAC* (Murshudov *et al.*, 1997) was combined with TLS (Winn *et al.*, 2001) refinement (one TLS group per chain). No NCS was used in the final stages of refinement. Structural superimposition of individual chains using *TOPP* (Lu, 1996) reveals an r.m.s. deviation of 0.3–0.4 Å for each of the five homotrimeric vvdUTPase structures. Table 2 provides the refinement statistics.

The surface area and volume of the channel were calculated (Liang *et al.*, 1998) using the *CastP* server (http://sts.bioengr.uic.edu/castp). Analysis of protein–protein interactions (Saha *et al.*, 2006) for the individual subunit–subunit interfaces was performed using the *ProFace* server (http://202.141.148.29/resources/bioinfo/interface). Table 3 shows a comparison of interface area and channel size for dUTPases from different species.

3. Results and discussion

3.1. Quality of structures

The resolution limit of the five refined structures described in this paper ranges from 1.8 to 2.5 Å. *R* values for the refinement are in the range 19.9–22.0% and R_{free} values lie between 23.1 and 27.4% (Table 2). The crystal structures of vvdUTPase contain a trimer in the asymmetric unit. The refined models for the full-length protein (apo enzyme and complexes) lack six to eight residues at the N-terminus and residues 132–147 at the C-terminus of each subunit. Electron density for eight additional C-terminal residues was missing from the structure of the ΔN11 form. Therefore, the model for the truncated version contains residues 11–123; the first residue (Met11) was introduced as a cloning artifact. In all structures, more than 92% of all residues lie in the core region of the Ramachandran plot and none of the six structures has any residues in disallowed or generously allowed regions. The primary, secondary, tertiary and quaternary structures of vvdUTPase are shown in Fig. 1. The quality of the electron density for all structures was excellent and allowed unambiguous positioning of ligands (Fig. 2).

3.2. General features of vvdUTPase structure

In all five structures of vvdUTPase, a chloride ion (Cl^-) is located at the local threefold of the trimer interface. In the full-length protein, this ion is at a distance of 3.0–3.4 Å from the NE2 atom of the Gln118 residues of all three chains, while in the ΔN11 form it is coordinated to the NH2 atoms of the three Arg65 residues (3.4–3.7 Å).

The functional unit of vvdUTPase is a trimer. Interactions between the monomers can be divided into three types: subunit–subunit interface contacts, interactions along the threefold axis and contacts through C-terminal 'arm' exchange

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

	Full-length apo	ΔN11 apo + sulfate	Full length + dUMPNPP	Full length + dUMP	Full length + dUDP
Space group	$P6_5$	$P6_3$	$P6_5$	$P6_5$	$P6_5$
Unit-cell parameters (Å)	$a = b = 120.01$, $c = 50.30$	$a = b = 105.26$, $c = 59.29$	$a = b = 120.46$, $c = 50.11$	$a = b = 119.97$, $c = 50.13$	$a = b = 119.66$, $c = 50.16$
Resolution range (Å)	50.00–2.40 (2.49–2.40)	49.70–2.15 (2.23–2.15)	19.72–2.50 (2.59–2.50)	20.00–1.80 (1.86–1.80)	19.94–2.10 (2.18–2.10)
Total reflections	49446	126242	39580	132077	86801
Unique reflections	16130	20238	14156	36946	24014
Average redundancy	3.1 (2.7)	6.2 (2.6)	2.8 (2.6)	3.6 (3.0)	3.6 (3.5)
Completeness (%)	98.3 (97.3)	98.6 (92.1)	97.3 (95.0)	96.2 (87.9)	99.5 (99.0)
R_{merge} (%)	13.4 (42.3)	3.3 (14.4)	13.2 (34.6)	4.6 (19.6)	7.8 (34.7)
Average $I/\sigma(I)$	6.8 (1.2)	29.9 (6.0)	5.7 (2.7)	18.6 (3.0)	10.2 (3.4)

Table 2
Summary of refinement statistics.

Values in parentheses are for the highest resolution shell.

	Full-length apo	ΔN11 apo + sulfate	Full length + dUMPNPP	Full length + dUMP	Full length + dUDP
PDB code	2okd	2okb	2oke	2ol1	2ol0
Resolution range (Å)	19.64–2.40 (2.46–2.40)	18.13–2.15 (2.21–2.15)	13.47–2.50 (2.56–2.50)	19.23–1.80 (1.85–1.80)	19.58–2.10 (2.16–2.10)
No. of reflections	14683 (1053)	19126 (1350)	13164 (962)	34538 (2258)	22718 (1732)
Completeness (%)	97.5 (95.0)	98.4 (90.6)	95.8 (92.9)	98.5 (93.5)	99.6 (99.2)
R_{all} (%)	20.9	21.3	22.2	20.1	20.9
R_{work} (%)	20.6 (29.5)	21.0 (22.4)	22.0 (29.5)	19.9 (27.5)	20.7 (24.7)
R_{free} (%)	26.8 (41.9)	25.7 (29.1)	27.4 (37.9)	23.1 (30.9)	26.1 (31.3)
No. atoms					
Total	3082	2928	3102	3298	3239
Protein	2947	2663	2892	2967	2947
Ligand/EDO	0/16	15/12	84/4	60/8	72/12
Mg^{2+}/Cl^-	0/1	0/2	1/1	0/1	1/1
Water	118	236	120	262	206
Wilson B factor (Å^2)	31.0	27.0	27.0	22.0	23.0
Average B factor (Å^2)					
Overall	31.7	30.8	20.4	25.5	24.6
Protein	31.7	30.1	19.6	25.0	24.0
Anion (Cl^-)	30.3	63.7, 60.0	23.0	17.7	16.5
Metal (Mg^{2+})	N/A	N/A	32.0	N/A	21.3
Ligand	N/A	41.7	43.0	24.6	28.7
EDO	43.5	41.4	30.4	42.0	43.6
Water	29.1	37.6	22.8	31.0	29.6
R.m.s. deviations					
Bond lengths (Å)	0.007	0.007	0.006	0.007	0.008
Bond angles (°)	1.011	0.902	1.058	1.031	1.188
Ramachandran plot					
Core (%)	92.5	92.4	92.4	92.8	93.8
Allowed (%)	7.5	7.6	7.6	7.2	6.2

Table 3
Surface area for subunit–subunit interfaces and surface area and volume of channel at the trimer interface in dUTPases from different species.

	Subunit–subunit interface surface (Å^2)			Channel	
	A–B	*B–C*	*C–A*	Surface (Å^2)	Volume (Å^3)
Vaccinia virus	2771	2734	2690	908 (922)†	1140 (1230)†
Human	3362	3241	3217	537	653
FIV	2878	2878	2878	619	805
M. tuberculosis	3071	3000	3002	271	304
E. coli	3708	3708	3708	51	32
EIAV	2677	2679	2673	159	185

† Surface area and volume of the channel are for vaccinia virus full-length and ΔN11 proteins (the number in parentheses refers to the ΔN11 protein).

(Persson *et al.*, 2001). Each dUTPase monomer folds into a jelly-roll β-barrel structure. The structure of each monomer (Figs. 1*a* and 1*b*) is composed of a total of 11 β-strands grouped into four antiparallel β-sheets (β-sheet 1, residues 11–15 and 53–57; β-sheet 2, residues 32–36, 110–120, 61–67 and 82–84; β-sheet 3, residues 40–42 and 104–106; β-sheet 4, residues 46–52, 92–99 and 76–80) in addition to a single short α-helix (residues 69–74). The C-terminal β-strands extend outside the corresponding subunit domain and contribute a β-strand from the crossing arm to the β-barrel of a neighboring subunit (see Fig. 1*c*). This interchange provides a large portion of the contacts at the subunit–subunit interface. Other contacts are between loop regions and through interactions of residues from a neighboring subunit with the ligand in the diphosphate and triphosphate state. Each active site is located at a cleft between adjacent subunits and the substrate-binding pockets are composed of the five conserved sequence motifs, with motifs I (residues 27–37; strand 2a), II (residues 62–80; strand 4, helix 1 and strand 5a) and IV (residues 102–121; strand 6b and strand 7) contributed by the first subunit, motif III (residues 82–98; strand 5b and strand 6) by the second subunit and motif V (residues 133–145; C-terminal loop) by the third subunit (Fig. 1*d*). In Fig. 2(*a*) superimposition of the active-site residues in the three nucleotide complexes of vvdUTPase shows the common features and subtle variations resulting from substrate binding. Like other trimeric dUTPases, the structure of vvdUTPase is characterized by the presence of a central cavity filled with water molecules with access to bulk solvent; thus, this cavity can be classified as a channel.

Motif V forms the C-terminal arm, which is highly flexible in the majority of the crystal structures. This stretch is glycine-rich and shows similarities to phosphate-binding motifs of other nucleotide-binding proteins; it is therefore termed the 'P-loop'. The sequence homology of this loop in dUTPases from different species is extensive (Prasad, 2001). Although an ordering of this loop is required for catalysis in many dUTPase structures, including those presented here, an extended portion of the C-terminal arm is flexible and not observed in the electron density (Persson *et al.*, 2001).

3.3. Apo enzyme

The model for the apo-enzyme structure of the full-length protein contains residues 9–132 of chain *A*, 9–131 of chain *B* and 7–131 of chain *C*, as well as one chloride ion, 118 water molecules and four EDO molecules. Alternate conformations are observed for Arg11 (chain *A*), Cys62 (chain *B*), Asp85 (chain *B*) and Cys102 (chains *B* and *C*). The structure displays the chloride ion at the center of the local threefold. Distances to the NE2 amide side-chain atoms of all three Gln118 residues are 3.2–3.3 Å. Four molecules of EDO (used as cryoprotectant) are found in the structure, with three of these close to the active sites displaying hydrogen bonds to active-site residues Ser70, Ile79, Gly82, Gly92 and Gly93. The fourth EDO binds in the vicinity of the chloride ion at the threefold and forms hydrogen bonds to the side-chain amide N atom of two Gln118 residues. Superimposition with the human apo-enzyme structure (PDB code 1q5u; Mol *et al.*, 1996) using *TOPP* (Lu, 1996) demonstrates a high degree of structural homology (r.m.s. deviation of only 0.8 Å for 372 C^{α} atoms with a match rate of 99.7%; sequence identity 64%).

The model for the apo-enzyme structure of the ΔN11 protein contains residues 11–123 in all three chains, three sulfate ions, two chloride ions, 236 water molecules and three EDO molecules. Alternate conformations are observed for two residues (Glu46 in chain *B* and Asn91 in chain *C*). The chloride ion is at the local threefold as seen in the full-length protein, but is coordinated to the NH2 atoms of all three Arg65 residues with slightly longer distances (3.4–

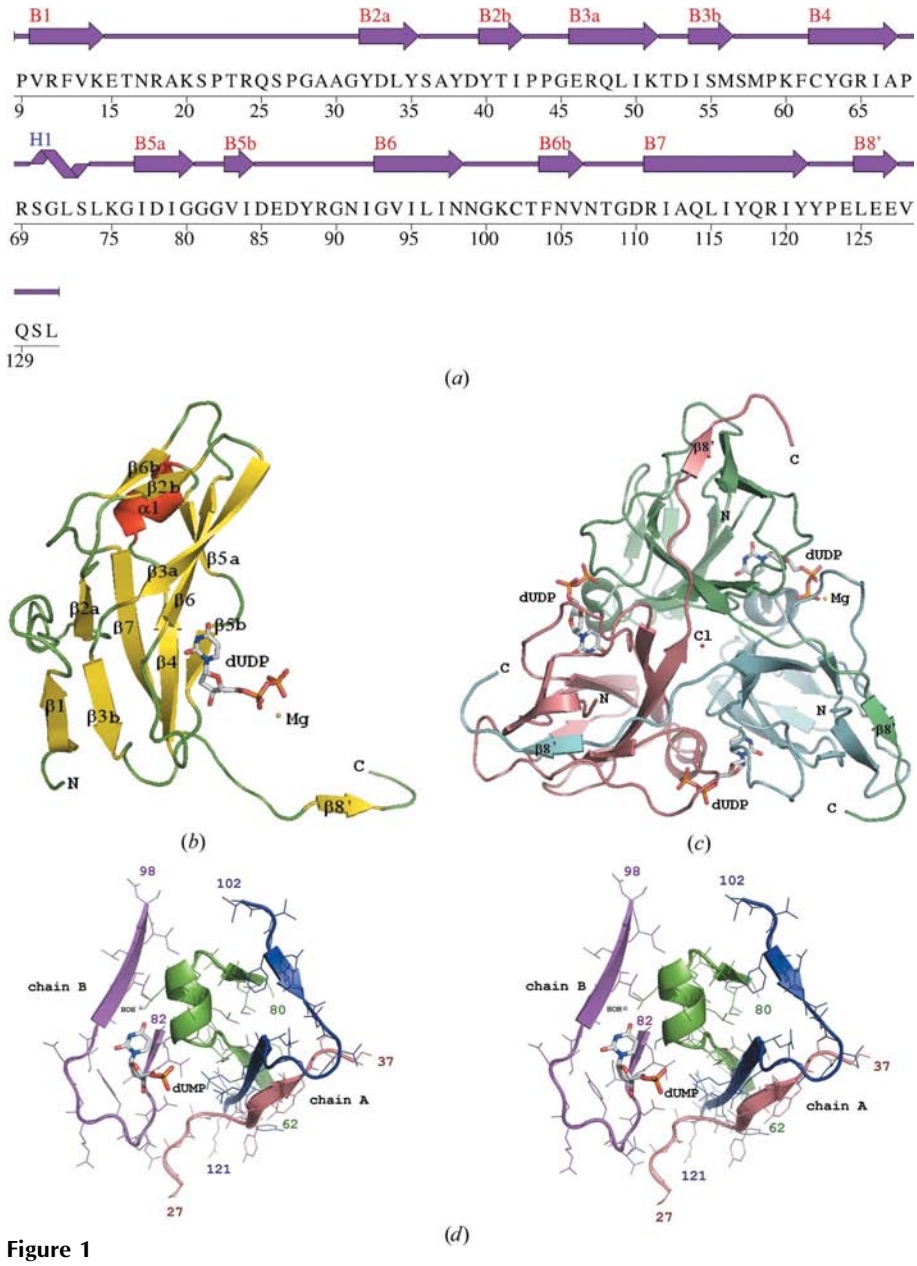

Figure 1
Primary, secondary, tertiary and quaternary structures of vvdUTPase. This figure shows the vaccinia virus protein sequence (for the observed coordinates) with secondary-structure assignments together with a ribbon model for the monomer and the trimer of vvdUTPase. (*a*) Wiring plot (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum; Laskowski, 2001) of secondary-structure assignments with vaccinia virus protein sequence. The secondary-structure assignment (in red) for the β-strands (B1–B8′ for β-strands 1–8′) and the single α-helix (H1 for α-helix 1) is according to Chan *et al.* (2004). Notice that strand 8′ is inserted into the first β-sheet of the adjacent subunit before strand 1. (*b*) The tertiary structure of the monomeric subunit in the vvdUTPase–dUDP complex. The protein is represented as a ribbon model, while the ligand is represented as a stick model. The Mg ion is visible as a small yellow sphere. Strands and the lone helix are labeled as in (*a*). (*c*) Trimer of the vvdUTPase–dUDP complex. The protein is represented by a ribbon model colored by chain, while the ligands are represented as a stick model. The chloride ion is visible in orange at the center on the local threefold. The Mg ion can be seen as a small yellow sphere in one active site. (*d*) Observed motifs in the active site of the vvdUTPase–dUMP complex (stereo figure). Four of the five motifs (in ribbon presentation with line drawing of residues) that form the active site are shown. Notice that motif V is disordered in our structures. Color codes are salmon, motif I; green, motif II; magenta, motif III; blue, motif IV. The first and last residues of each motif are labeled. This figure was prepared with *PyMOL* (DeLano Scientific LLC).

3.7 Å). One sulfate ion is present in each active site (presumably from the lithium sulfate used in crystallization). In all three active sites, the sulfate ions occupy approximately the same position as the β-phosphate in the dUDP and dUMPNPP complexes and form numerous contacts with active-site residues Arg69, Gly71 and Arg111, as well as with several water molecules. One of the three EDO molecules is in the vicinity of the active site, displaying hydrogen bonds to active-site residues Ser70, Gly82 and Gly93, while a second EDO connects to a sulfate ion in the active site through a water molecule (W36). The second chloride ion was found to be coordinated to Arg111 (3.6 Å) and two water molecules (2.9 and 3.6 Å).

Interestingly, superimposition of the apo-enzyme structures of vvdUTPase (ΔN11 and full length) shows a larger r.m.s. deviation (0.9 Å for all C^{α} atoms) than that observed between the full-length vvdUTPase and human dUTPase (0.8 Å). In the ΔN11 form we observe a shift of some secondary-structure elements and loops, which results in an increase in the volume of the central channel and more open active sites. The truncation at the N-terminus causes additional disordering of the C-terminus and as a result the C-terminal 24 residues are not visible in the electron density. In the full-length version Pro9 and Val10 form hydrogen bonds to Glu124 and Glu126, respectively, and these interactions restrict the movement of the C-terminal arm and result in some ordering of the C-terminus in the full-length protein. All other features seen in the full-length protein are conserved in the ΔN11 form.

As described for human (Mol *et al.*, 1996) and *E. coli* (Barabás *et al.*, 2004) dUTPases, the catalytic site in the vaccinia virus apo enzyme shows no major conformational change on ligand binding. The average r.m.s. deviation between the vaccinia virus apo-enzyme structure and the corresponding nucleotide complexes ranges between 0.3 and 0.4 Å for all C^{α} atoms of the trimer.

3.4. Protein–dUMPNPP (substrate analog) complex

The model for the dUMPNPP complex of vvdUTPase contains residues 9–131 in all three chains, as well as three

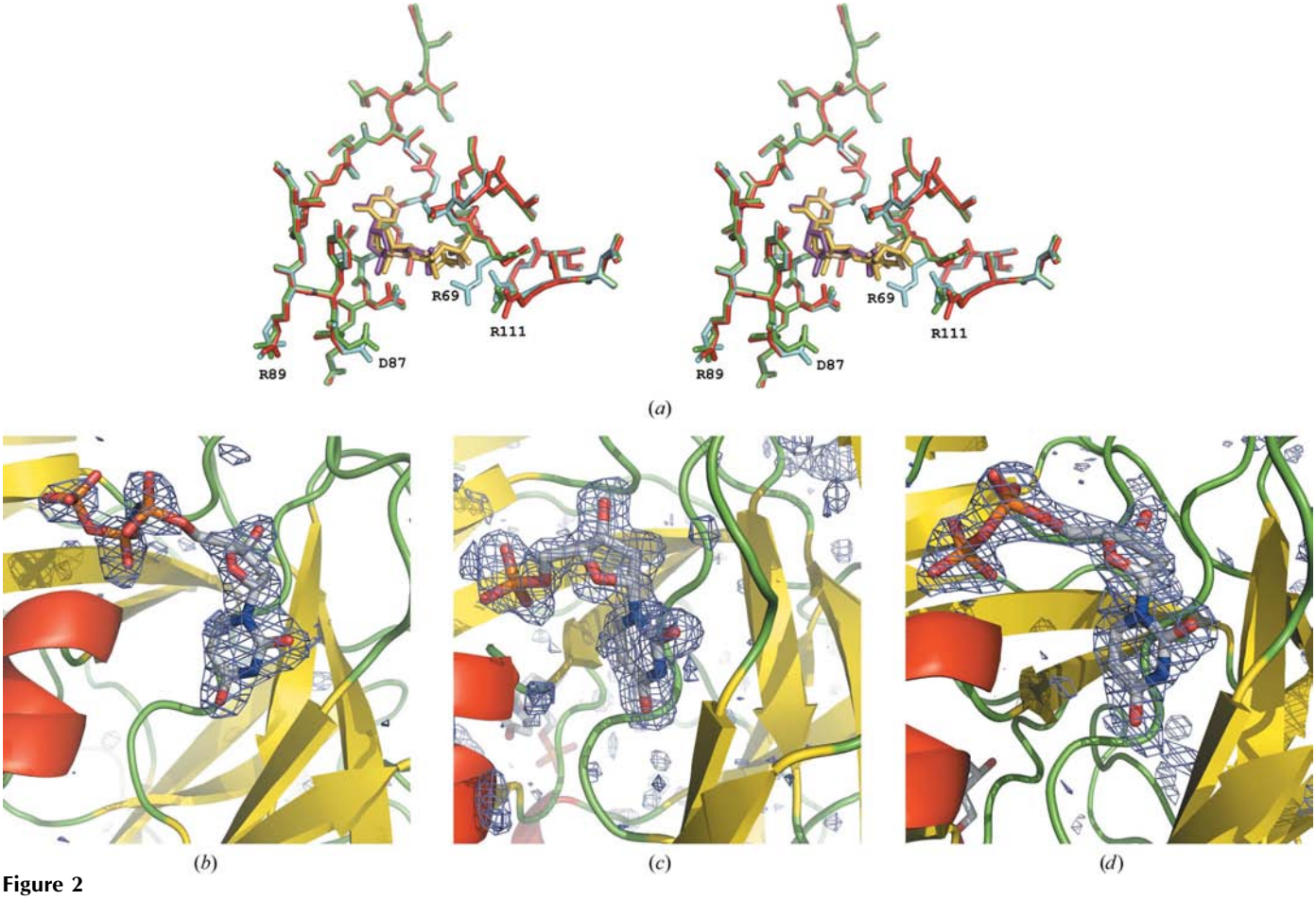

Figure 2
Close-up view of the active sites in nucleotide complexes of vvdUTPase. The protein residues and ligands in the active sites of the protein–nucleotide complexes in full-length vvdUTPase are shown. (*a*) Superimposition of active sites in protein–nucleotide complexes (stereo figure). Color codes for the protein residues are red for the dUMPNPP complex (PDB code 2oke), green for the dUMP complex (2ol1) and aquamarine for the dUDP complex (2ol0). Color codes for the ligands are orange for dUMPNPP, magenta for dUMP and wheat for dUDP. A few residues that show different rotamer conformations in the three nucleotide complexes are labeled. (*b*) Protein–dUMPNPP complex ($F_o - F_c$ OMIT map contoured at 3σ). (*c*) Protein–dUMP complex ($F_o - F_c$ OMIT map contoured at 3σ). (*d*) Protein–dUDP complex ($F_o - F_c$ OMIT map contoured at 3σ). This figure was prepared with *PyMOL* (DeLano Scientific LLC).

dUMPNPP ligands, one magnesium ion, one chloride ion, 120 water molecules and two EDO molecules. The nonhydrolyzable substrate analog contains an imido group instead of the oxygen bridging the α-phosphate and β-phosphate groups, which renders the α-phosphate less susceptible to nucleophilic attack while maintaining geometry similar to the natural substrate. Electron density for the entire ligand was visible in the $F_o - F_c$ map contoured at 3σ (Fig. 2*b*). In contrast to the structures of the dUMPNPP complexes of human and *E. coli* dUTPases, the γ-phosphate does not stabilize the C-terminal loop, which remains disordered. The ligand is bound at the subunit–subunit interface in a pocket created by two antiparallel β-strands (strands 5b and 6) that are connected by a β-hairpin loop (motif III). The protein interactions with the ligand in each of the three active sites fall into two groups: the first group consists of residues belonging to one subunit (motif III), while the second group contains residues from the adjacent subunit (motifs II and IV). The active-site residues of the first group that interact with the deoxyuridine moiety of the substrate analog include Gly82, Ile84, Asp85, Tyr88 and Gly93. Of these, Ile84 and Tyr88 show only hydrophobic interactions, while Gly82 and Gly93 are hydrogen bonded through the backbone N atoms of the amide groups to the carbonyl O atoms (O2 and O4) of the uracil base. In each active site the uracil O4 atom displays a hydrogen bond to a conserved structural water molecule (Mol *et al.*, 1996; Barabás *et al.*, 2004; Chan *et al.*, 2004; Dauter *et al.*, 1998; Prasad *et al.*, 2000). The highly conserved Tyr residue at the bottom of the β-hairpin loop discriminates against a potential ribose sugar by packing against the deoxyribose moiety. The catalytic Asp85 residue forms a hydrogen bond with the hydroxyl group of the deoxyribose sugar, but only in one active site. As observed in other dUTPase structures, the orientation of the triphosphate moiety of dUMPNPP differs in the three active sites. In two active sites the geometry is *gauche*, while the third site displays a *trans* geometry. As discussed by Chan *et al.* (2004), the geometry remains flexible in the absence of metal ions, which seem to stabilize the *gauche* conformation. Indeed, in one of the active sites in the vvdUTPase complex, where a magnesium ion is found, dUMPNPP remains in the *gauche* conformation. The metal ion is coordinated with the β-phosphate (distance 2.77 Å) and γ-phosphate (distance 2.35 Å) groups of dUMPNPP. Additional coordinations involving Gln25 OE1 and a water molecule are also noticed. Three amino-acid residues from the adjacent subunit are hydrogen bonded to the triphosphate moiety: Arg69 and Ser70 belonging to motif II and Arg111 from motif IV. Ser70 from the adjacent subunit interacts through main-chain and side-chain atoms with the γ-phosphate in two active sites, while in the third site the Ser residue forms hydrogen bonds to the α-phosphate and β-phosphate group of the ligand. In this site, the side chains of Arg111 and Arg69 from the adjacent subunit are hydrogen-bonded to an O atom of the γ-phosphate, while in the other two active sites these arginine residues are too far from the γ-phosphate for hydrogen bonding. Several water molecules are also involved in hydrogen bonds to the phosphate groups. The single EDO molecule displays hydrogen bonds to two Asn99 residues, Arg47 and a water molecule.

The r.m.s. deviation between the vvdUTPase–dUMPNPP structure and the *E. coli* structure in complex with dUMPNPP is 1.2 Å for 354 C^{α} atoms (sequence identity 31%; match rate 84%) of the trimer. The coordinates of human dUTPase with this ligand are not available for comparison.

3.5. Protein–dUMP (product) complex

The model for the dUMP complex of full-length vvdUTPase contains residues 8–131 of chain *A*, 8–131 of chain *B* and 7–131 of chain *C*, three dUMP ligands, one chloride ion, 262 water molecules and two EDO molecules. Alternate conformations are observed for a total of 14 residues (Gln25, Met58, Cys62, Cys102 and Gln124 in chain *A*, Glu15, Cys62, Cys102 and Arg111 in chain *B*, and Arg89, Cys102, Tyr121, Gln124 and Gln129 in chain *C*). Although this complex was prepared in the presence of Mg^{2+}, no metal ion could be located in this structure. Excellent electron density for the ligand was visible in the $F_o - F_c$ map (Fig. 2*c*). The deoxyuridine moiety of the ligand displays the same interactions with active-site residues from motif III as described for the dUMPNPP complex. As seen in the human dUTPase complex (Mol *et al.*, 1996), the side-chain hydroxyl of Ser70 from an adjacent subunit donates a hydrogen bond to the α-phosphate group of dUMP in each of the active sites of vvdUTPase. Additional hydrogen bonds between the α-phosphate group of the ligand and water molecules are observed in this complex. Numerous water molecules

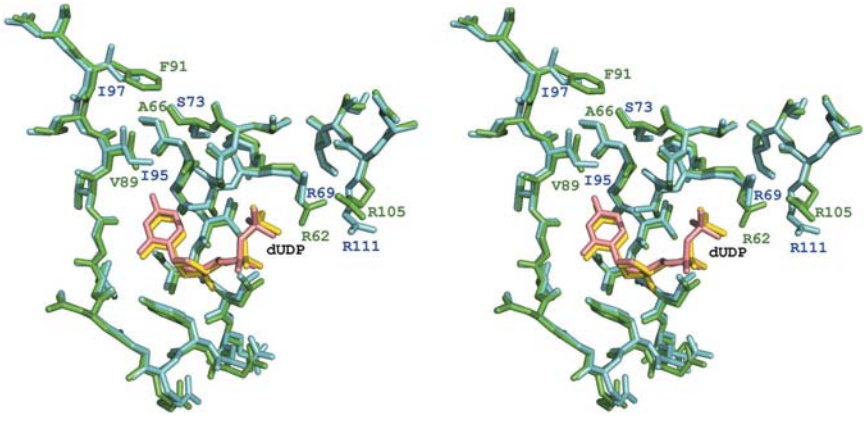

Figure 3
Superimposition of the active sites in dUDP complexes of human (PDB code 1q5h) and vaccinia virus (PDB code 2ol0) dUTPase. The protein residues and the ligand(s) in the active site(s) of the dUDP complexes for vvdUTPase and human dUTPase are shown (stereo figure). Three residues in the active site that differ between 1q5h and 2ol0 are labeled. In addition, two Arg residues with different rotamer conformations are also labeled. Color codes for the protein residues are cyan, vvdUTPase; green, human dUTPase. Color codes for the ligands are salmon, vvdUTPase; yellow, human dUTPase. This figure was prepared with *PyMOL* (DeLano Scientific LLC).

can be seen in the vicinity of the active-site pocket, but none are close enough to the side chain of Asp85 and the α-phosphate group of dUMP to qualify as a 'catalytic' water.

Coordinates for the dUMP complex of human dUTPase are not available for superimposition and detailed structural comparison. The r.m.s. deviation between the vvdUTPase–dUMP and the *E. coli* dUTPase–dUMP structures is 1.2 Å for the matching C^{α} atoms in the trimer.

3.6. Protein–dUDP (inhibitor) complex

The model for the dUDP complex of full-length vvdUTPase contains residues 8–131 of chain *A*, 8–131 of chain *B* and 7–131 of chain *C*, three dUDP ligands, one magnesium ion, one chloride ion, 206 water molecules and three EDO molecules. Alternate conformations are observed for a total of ten residues in the three chains (Arg11, Glu15, Cys62 and Cys102 in chain *A*, Cys62 and Cys102 in chain *B,* and Arg11, Cys62, Cys102 and Gln129 in chain *C*). dUDP is a weak inhibitor of dUTPase (K_i for various dUTPases is in the low micromolar range; Larsson *et al.*, 1996; Kovári *et al.*, 2004). The ligand could be placed unambiguously into the electron density (see Fig. 2*d*). Interactions between the protein and the deoxyuridine moiety of dUDP are the same as observed in the dUMPNPP and dUMP complexes. The diphosphate moiety is surrounded by residues from the adjacent subunit. Here, Ser70 interacts with the β-phosphate group through main-chain atoms. Gly71 also provides hydrogen bonds to the β-phosphate, while Arg111 is hydrogen bonded to the O atoms of the β-phosphate group through side-chain atoms in two of the three active sites. In one of these sites, there is also a hydrogen bond between Arg69 and the β-phosphate group through a side-chain atom. In the third active site, we observe an Mg^{2+} ion that coordinates to both the α-phosphate (distance 2.63 Å) and β-phosphate (distance 2.31 Å) groups of dUDP. The octahedral coordination is completed through hydrogen bonds to three water molecules and the OE1 side-chain O atom of Gln25. To allow this interaction, the side-chain orientation of both arginine residues (Arg69 and Arg111) is changed compared with the other two active sites in this complex and in the apo enzyme and the dUMP complex. In addition, several water molecules are within hydrogen-bonding distance of the diphosphate moiety. In the second active site, a large cylinder-shaped unexplained positive density is observed in the vicinity of the diphosphate moiety. One of the three EDO molecules is found close to the active site. It forms hydrogen bonds to O2 of the uracil base of dUDP and the side-chain hydroxyl of Ser70, while another EDO close to the threefold interacts with the main-chain carbonyl O atom of Arg119 and the side-chain NE2 atom of a Gln118 residue.

In the human enzyme, similar residues are involved in binding of dUDP (see Fig. 3). Superimposition of the dUDP complexes of human and vvdUTPase shows an r.m.s. deviation of 0.7–0.8 Å for 369 C^{α} atoms of the trimer.

3.7. Structural comparison with other dUTPases

vvdUTPase is highly homologous to its human counterpart in terms of sequence, structure and enzymatic properties (Roseman *et al.*, 1996). The active-site residues as well as water molecules that make contacts with the ligand(s) are well conserved. Nonetheless, subtle differences in the nature of residues forming intersubunit contacts give rise to distinguishable features in the contact area (see Fig. 3 for a comparison between vvdUTPase and human dUTPase). The interface areas of the interacting subunits in the trimer for human and vaccinia virus dUTPase are approximately 3300 and 2700 $Å^2$, respectively (see Table 3). The number of residues and atoms contributing to the subunit–subunit interface (core plus rim) in the crystal structures also differs considerably among the known dUTPases [70 residues and 293 atoms for vaccinia virus, 80 residues and 345 atoms for human, 74 residues and 294 atoms for feline immunodeficiency virus (FIV), 84 residues and 300 atoms for *Mycobacterium tuberculosis*, 90 residues and 372 atoms for *E. coli*, and 71 residues and 281 atoms for equine infectious anemia virus (EIAV)].

Trimeric dUTPase enzymes have been classified into two groups based on the character of the residues at the subunit interface around the threefold axis (Fiser & Vértessy, 2000). In the crystal structures of dUTPases, the threefold axis is either crystallographic (one monomer per asymmetric unit) or noncrystallographic (one trimer per asymmetric unit). Eukaryotic enzymes, such as human dUTPase, display a polar triad constructed of nine polar residues (three from each subunit; Tyr56, Arg58 and Glu112). Occasionally, this threefold is occupied by a metal ion. Poxvirus dUTPases contain a similar but slightly different polar triad (Tyr63, Arg65 and Gln118). In vvdUTPase, replacement of Glu by Gln and the

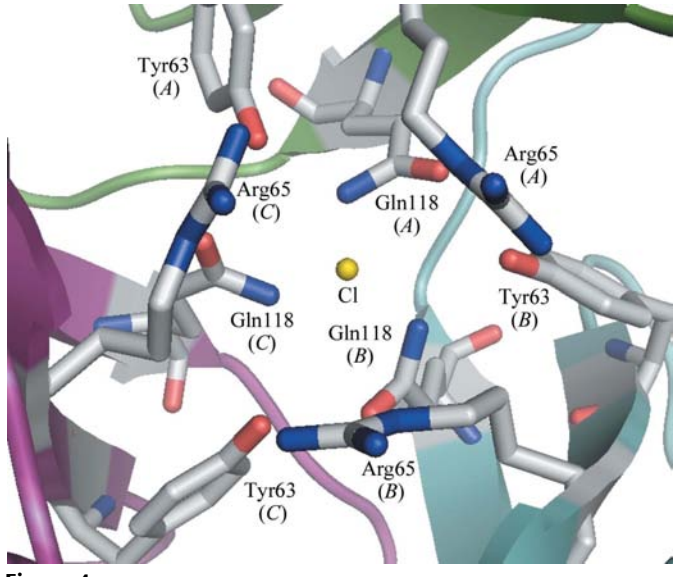

Figure 4
Close-up view of the trimer interface with the polar triad in vaccinia virus full-length dUTPase. This figure shows the trimer interface with the polar triad. The chloride ion sits on the local threefold and is surrounded by residues Tyr63, Arg65 and Gln118 from all three subunits. The distances between Cl^- and the NE2 atoms of the three Gln118 residues are 3.2 Å. The chloride ion is represented as a yellow sphere, the residues of the triad are shown as stick models and the remainder of the protein is displayed as ribbon model colored by chains. Residues and chloride ion are labeled. This figure was prepared with *PyMOL* (DeLano Scientific LLC).

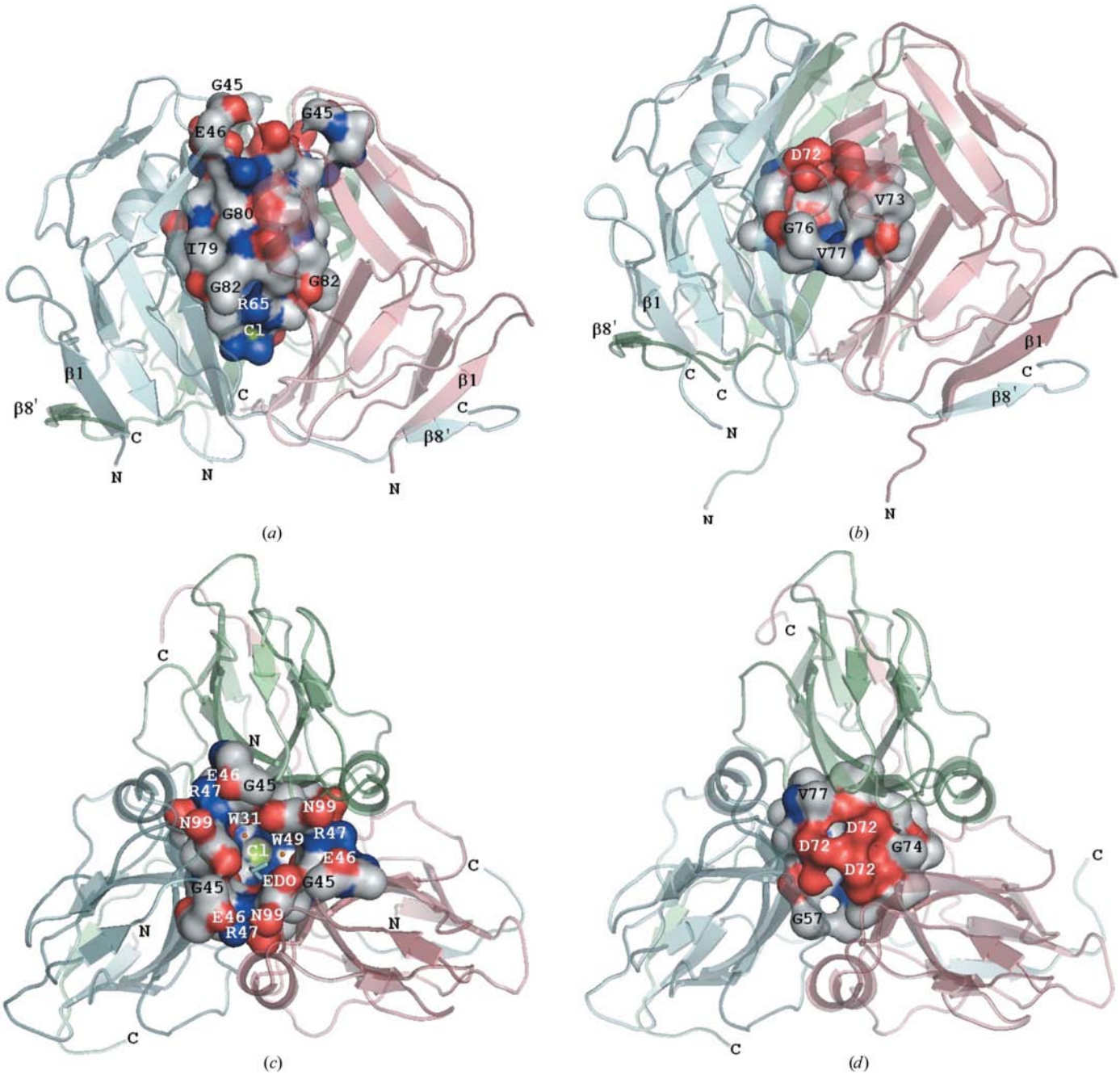

Figure 5
The channel in human and vaccinia virus dUTPase. The structures of human and vaccinia virus dUTPase were superimposed with *TOPP* (Lu, 1996) and calculations for the channel were performed using the *CastP* server (http://sts.bioengr.uic.edu/castp/). The figure shows two different views of human and vaccinia virus dUTPase trimers as ribbon models, together with a surface representation of the channels for direct comparison, emphasizing differences in size and extension of the channels. (*a*) Trimer of vvdUTPase shown as a ribbon model colored by chains and channel in solid surface representation colored by element. Some of the residues whose atoms are part of the channel wall are labeled. Labeling can also be seen for the chloride ion in green at the bottom of the channel as well as the termini and β-strands 1 and 8′. The top-to-bottom distance for the channel is approximately 25 Å, while the diameter is approximately 16 Å. The open mouth of the channel is on top between the Gly45 residues. The view is perpendicular to the threefold. (*b*) The trimer of human dUTPase and the channel are shown in the same representation and in the same orientation as in the vaccinia virus enzyme in (*a*). Labeling of the termini and β-strands 1 and 8′ as well as residues making up the channel wall is shown. In this orientation the channel opening for the human enzyme can be seen at the center between residues Gly76 and Asp72. (*c*) The trimer of vvdUTPase and the channel in the same representation, but in a different orientation (perpendicular to Fig. 5*a*), showing the channel opening. The green sphere at the center is the chloride ion at the back of the channel. This chloride ion sits on the threefold. Two water molecules in red and an EDO molecule in aquamarine are visible inside the channel and are labeled. Residues making up the channel wall are labeled. The distances between the corresponding atoms of the Gly45 residues are 12 Å, while the distances between corresponding atoms of Asn99 residues are 8 Å. (*d*) The trimer of human dUTPase and the channel are shown in the same representation and orientation as the vaccinia virus enzyme in (*c*). Compared with the vaccinia enzyme, the opening of the channel in the human enzyme is tilted and is visible between residues Gly57 and Asp72 in this orientation. Labeling of the termini and residues making up the channel wall is shown. This figure was prepared with *PyMOL* (DeLano Scientific LLC).

distances from other residues in the triad favor an anion at this location over a positively charged metal ion (see Fig. 4). In the bacterial and retroviral enzymes, the polar triad is replaced by hydrophobic residues either exclusively (Met67, Leu69, Ile121 and Val123 in *E. coli* and Phe50, Trp52, Ile103, Leu105 in EIAV) or at least to a large extent (Leu60, His62 and Pro79 in *M. tuberculosis* and Asp64, Leu66 and Val85 in FIV).

One major difference between human dUTPase and vvdUTPase is the size of the channel. Table 3 lists the surface area and volume of the channel at the trimer interface for the known dUTPase structures from different species. This is the largest observed channel in a dUTPase to date, with a surface area and volume almost twice as large as that in the human enzyme (see Table 3). The channel in vvdUTPase is constructed from alternating layers of positively and negatively charged residues and contains numerous water molecules. The channel is closed at one end, while the other end is open and as a result water molecules filling the channel have contact with bulk solvent. The channel in vvdUTPase is not only longer with a larger volume than in human dUTPase, but the opening is also much more extended. Fig. 5 shows two different views of the central channel in human dUTPase and vvdUTPase. The size of the channel seems to correlate with the nature of the residues at the triad. In the enzymes that contain hydrophobic residues at the trimer interface, the channel behind the threefold is smaller (see Table 3). In these cases the site at or around the threefold is empty as in EIAV dUTPase (Dauter *et al.*, 1999) or occupied by a Tris molecule as in *M. tuberculosis* (Chan *et al.*, 2004) and *E. coli* (Dauter *et al.*, 1998) dUTPases. The presence of an Asp residue at the threefold in FIV dUTPase (Prasad *et al.*, 2000) in addition to two hydrophobic residues allows coordination with a metal ion and as a result the size of the channel is much larger than those with exclusively hydrophobic polar triads (see Table 3). In vvdUTPase, owing to the presence of the Cl^- ion, the channel is more open and larger than in any known dUTPase. In the truncated (ΔN11) vvdUTPase protein, the chloride ion is coordinated to the Arg65 residues and the bond distances are even longer; as a result, the volume of the channel is even larger. In contrast to the relatively small species-specific differences in binding pockets, much larger species-specific differences in the trimer interface channel are observed, which can be exploited for drug design. Suitable compounds targeting this central channel could potentially interfere with the binding of dUTP by distorting the conformations of contact residues to the uracil moiety or could affect the catalytic activity by disrupting trimer assembly. The size and nature of this channel would allow discrimination between enzymes of different species and provide the necessary selectivity.

Another difference between vaccinia virus and human dUTPases is the distinctive ordering of the C-terminal arm in the nucleotide complexes of the latter. Upon ordering, the C-terminal arm wraps around and caps the substrate-bound active sites, utilizing a conserved Phe residue (Phe135 in human dUTPase) that stacks over the uracil base of the nucleotide. Mol *et al.* (1996) coined the term 'Phe-lid' for this capping (residue) of the uracil-recognition pocket. Although the sequence homology of the C-terminal residues in the human enzyme compared to the vaccinia virus enzyme is high (only two out of ten residues differ), for reasons that are not clear to us this region remains disordered in the viral enzyme.

We thank Professor Dmitry Vassylyev and his colleagues for assistance in X-ray data collection. We also thank A. Grigorian for technical assistance. The work described here was accomplished as part of the Southeast Regional Center of Excellence for Emerging Infections and Biodefense (SERCEB) initiative and was supported by NIH Grant No. U54 AI 057157.

References

Barabás, O., Pongrácz, V., Kovári, J., Wilmanns, M. & Vértessy, B. G. (2004). *J. Biol. Chem.* **279**, 42907–42915.
Brünger, A. T., Adams, P. D. & Rice, L. M. (1998). *Curr. Opin. Struct. Biol.* **8**, 606–611.
Chan, S., Segelke, B., Lekin, T., Krupka, H., Soo Cho, U., Kim, M., So, M., Kim, C.-Y., Pashkov, I., Cascio, D., Perry, J. L. & Sawaya, M. R. (2004). *J. Mol. Biol.* **341**, 503–517.
Dauter, Z., Persson, R., Rosengren, A. M., Nyman, P. O., Wilson, K. S. & Cedergren-Zeppezauer, E. S. (1999). *J. Mol. Biol.* **285**, 655–673.
Dauter, Z., Wilson, K. S., Larsson, G., Nyman, P. O. & Cedergren-Zeppezauer, E. S. (1998). *Acta Cryst.* D**54**, 735–749.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fiser, A. & Vértessy, B. G. (2000). *Biochem. Biophys. Res. Commun.* **279**, 534–542.
Kovári, J., Barabás, O., Takács, E., Békési, A., Dubrovay, Z., Pongrácz, V., Zagyva, I., Imre, T., Szabó, P. & Vértessy, B. G. (2004). *J. Biol. Chem.* **279**, 17932–17944.
Larsson, G., Nyman, P. O. & Kvassman, J.-O. (1996). *J. Biol. Chem.* **271**, 24010–24016.
Laskowski, R. A. (2001). *Nucleic Acids Res.* **29**, 221–222.
Liang, J., Edelsbrunner, H. & Woodward, C. (1998). *Protein Sci.* **7**, 1884–1897.
Lu, G. (1996). *Protein Data Bank Q. Newsl.* **78**, 10–11.
McGeehan, J. E., Depledge, N. W. & McGeoch, D. J. (2001). *Curr. Protein Pept. Sci.* **2**, 325–333.
McGeoch, D. J. (1990). *Nucleic Acids Res.* **18**, 4105–4110.
Mol, C. D., Harris, J. M., McIntosh, E. M. & Tainer, J. A. (1996). *Structure*, **4**, 1077–1092.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Persson, R., Cedergren-Zeppezauer, E. S. & Wilson, K. S. (2001). *Curr. Protein Pept. Sci.* **2**, 287–300.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Prasad, G. S. (2001). *Curr. Protein Pept. Sci.* **2**, 301–311.
Prasad, G. S., Stura, E. A., Elder, J. H. & Stout, C. D. (2000). *Acta Cryst.* D**56**, 1100–1109.
Roseman, N. A., Evans, R. K., Mayer, E. L., Rossi, M. A. & Slabaugh, M. B. (1996). *J. Biol. Chem.* **271**, 23506–23511.
Saha, R. P., Bahadur, R. P., Pal, A., Mandal, S. & Chakrabarti, P. (2006). *BMC Struct. Biol.* **6**, 11.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.