Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of dNTP-inducible dNTP triphosphohydrolase: insight into broad specificity for dNTPs and triphosphohydrolase-type hydrolysis

Naoyuki Kondo,[a] Noriko Nakagawa,[a,b] Akio Ebihara,[b] Lirong Chen,[c] Zhi-Jie Liu,[c] Bi-Cheng Wang,[c] Shigeyuki Yokoyama,[b,d,e] Seiki Kuramitsu[a,b] and Ryoji Masui[a,b]*

[a]Department of Biological Sciences, Graduate School of Science, Osaka University, 1-1 Machikaneyama-cho, Toyonaka, Osaka 560-0043, Japan, [b]RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan, [c]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA, [d]RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, and [e]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Correspondence e-mail: rmasui@bio.sci.osaka-u.ac.jp

Received 25 July 2006
Accepted 17 November 2006

PDB Reference: dNTP triphosphohydrolase, 2dqb, r2dqbsf.

Deoxyribonucleoside triphosphate triphosphohydrolase from *Thermus thermophilus* (Tt-dNTPase) has a unique regulatory mechanism for the degradation of deoxyribonucleoside triphosphates (dNTPs). Whereas the *Escherichia coli* homologue specifically hydrolyzes dGTP alone, dNTPs act as both substrate and activator for Tt-dNTPase. Here, the crystal structure of Tt-dNTPase has been determined at 2.2 Å resolution, representing the first report of the tertiary structure of a dNTPase homologue belonging to the HD superfamily, a diverse group of metal-dependent phosphohydrolases that includes a variety of uncharacterized proteins. This enzyme forms a homohexamer as a double ring of trimers. The subunit is composed of 19 α-helices; the inner six helices include the region annotated as the catalytic domain of the HD superfamily. Structural comparison with other HD-superfamily proteins indicates that a pocket at the centre of the inner six helices, formed from highly conserved charged residues clustered around a bound magnesium ion, constitutes the catalytic site. Tt-dNTPase also hydrolyzed noncanonical dNTPs, but hardly hydrolyzed dNDP and dNMP. The broad substrate specificity for different dNTPs might be rationalized by the involvement of a flexible loop during molecular recognition of the base moiety. Recognition of the triphosphate moiety crucial for the activity might be attained by highly conserved positively charged residues. The possible mode of dNTP binding is discussed in light of the structure.

1. Introduction

To maintain high-fidelity DNA replication, the level of deoxyribonucleoside triphosphates (dNTPs), precursors of DNA synthesis, must be strictly controlled within the cell. Indeed, an imbalanced pool of dNTPs causes an increased frequency of frameshift mutations (Bebenek & Kunkel, 1990). The intracellular level of dNTPs is mainly regulated by enzymes involved in *de novo* dNTP synthesis, as represented by ribonucleotide reductase (RNR). RNR catalyzes the reduction of the 2′-hydroxyl group of nucleoside diphosphate (NDP) to deoxyribonucleoside diphosphate (dNDP). Most RNRs are allosterically regulated by dNTPs and ATP, which determine the substrate specificity and activity (Eklund *et al.*, 2001; Jordan & Reichard, 1998). Mutations of RNR in the region involved in allosteric regulation perturb the dNTP level, which drastically increases the mutation rate during DNA replication (Chabes *et al.*, 2003). Therefore, in order to maintain genetic stability, the concentration of dNTPs must be strictly controlled by the relevant biosynthetic pathways.

Several enzymes are known to hydrolyze dNTPs. Non-canonical dNTP hydrolases such as dUTPase and MutT/MTH1 specifically hydrolyze dUTP and 8-oxo-dGTP, respectively, in the manner of a pyrophosphohydrolase, thereby preventing misincorporation of mutation-inducible dNTPs (Warner *et al.*, 1981; Sekiguchi & Tsuzuki, 2002; Moroz *et al.*, 2005). However, we recently identified a novel dNTP triphosphohydrolase (dNTPase) that hydrolyzes all canonical dNTPs in *Thermus thermophilus* HB8 (Kondo *et al.*, 2004). This enzyme, encoded by the TTHA0412 gene (project code TT1383), has triphosphohydrolase activity and hydrolyzes dNTP to deoxyribonucleoside and inorganic triphosphate in a divalent cation-dependent manner. We refer to this enzyme as Tt-dNTPase. Interestingly, the activity of Tt-dNTPase is regulated in a complex fashion by dNTPs. Although the enzyme has no activity for single dNTPs, its activity for canonical dNTPs is induced by the presence of mixed dNTPs involving dATP or dTTP (Kondo *et al.*, 2004). dATP and dTTP are also hydrolyzed in the presence of the other dNTPs, suggesting that dATP and dTTP act as both the activator and the substrate (Kondo *et al.*, 2004). dADP, dTDP, dAMP and dTMP, which themselves were not hydrolyzed, were nevertheless able to stimulate the hydrolysis of dCTP. These results suggest the existence of binding sites specific for dATP and dTTP as positive modulators, distinct from the dNTPase catalytic site. Based on further analysis of steady-state kinetic data, we proposed a model in which Tt-dNTPase has one dNTP-binding site for catalysis and two sites for recognition of modulators (Kondo *et al.*, 2004).

Tt-dNTPase belongs to the HD superfamily (Aravind & Koonin, 1998). This superfamily is constituted of a diverse group of metal-dependent phosphohydrolases and is named after their conserved His and Asp residues. It includes a variety of uncharacterized proteins and domains associated with nucleotidyltransferases and helicases from bacteria, archaea and eukaryotes. Sequence homologues of Tt-dNTPase are widely found in eubacteria. Of these, *Escherichia coli* deoxyguanosine triphosphohydrolase (Ec-dGTPase) has been most extensively studied. Ec-dGTPase was originally discovered as a contaminant of a DNA polymerase I preparation (Kornberg *et al.*, 1958). Although Ec-dGTPase shows triphosphohydrolase activity, it has a 20-fold higher specificity for dGTP than for the other canonical dNTPs and does not require mixed dNTPs for dGTPase activity (Seto *et al.*, 1988). Homologues possessing dGTPase activity are restricted to *Enterobacteriaceae* species, which are closely related to *E. coli* (Quirk & Bessman, 1991). There is relatively low sequence similarity among Tt-dNTPase/Ec-dGTPase homologues: for example, Ec-dGTPase has 20% sequence identity to Tt-dNTPase. The enzymes share only the sequence motif characteristic of hydrolases belonging to the HD superfamily (Aravind & Koonin, 1998). Such sequence diversity seems to explain the different specificity and regulatory mechanisms between Tt-dNTPase and Ec-dGTPase. It was also reported that Ec-dGTPase binds to single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) (Wurgler & Richardson, 1993). These data imply a linkage of these enzymes to DNA metabolism. However, the molecular basis of the catalysis, regulation and DNA-binding mechanisms has not been clarified, since no structure of a Tt-dNTPase/Ec-dGTPase family member has previously been determined

In this paper, we have solved the crystal structure of Tt-dNTPase using the single anomalous scattering (SAD) method. Tt-dNTPase possesses hexameric architecture and the active site was identified in the inner region, which includes highly conserved residues from HD-superfamily proteins. Based on structural comparison of the active site between Tt-dNTPase and other proteins belonging to the HD superfamily, we constructed a model of ligand-bound Tt-dNTPase. The broad substrate specificity for different dNTPs and the requirement of the triphosphate moiety for the activity could be rationalized using this model. The results presented in this paper have increased our knowledge of the fundamental molecular function of the Tt-dNTPase/Ec-dGTPase homologues.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data-collection statistics	
Space group	*P*1
Unit-cell parameters (Å)	a = 74.1, b = 84.2, c = 108.0, α = 110.2, β = 109.7, γ = 94.4
Resolution (Å)	50–2.1 (2.18–2.1)
Wavelength (Å)	0.9792
No. of observations†	239729
Unique reflections†	220923
Data completeness	84.9 (72.8)
Average $I/\sigma(I)$	11.6 (3.2)
$R_{\rm merge}$‡ (%)	8.3 (31.9)
Refinement statistics	
Resolution (Å)	20–2.2
Data completeness (%)	90.1
No. of reflections	102946
Mean B value	38.2
R§	0.223
$R_{\rm free}$§	0.286
R.m.s. deviation from ideality	
Bond lengths (Å)	0.009
Bond angles (°)	1.4
Dihedral angles (°)	19.8
Improper angles (°)	0.97

† Friedel pairs were treated as independent reflections. ‡ $R_{\rm merge} = \sum |I_{\rm obs} - \langle I \rangle| / \sum I_{\rm obs}$. § $R = \sum ||F_{\rm obs}| - |F_{\rm calc}|| / \sum |F_{\rm calc}|$. 10% of the data were omitted from the refinement and used in $R_{\rm free}$ calculations.

2. Materials and methods

2.1. Protein expression and purification

Selenomethionine (SeMet) substituted Tt-dNTPase was prepared as follows. *E. coli* B834(DE3) (Novagen, Madison, WI, USA) was transformed with pET-11a/*tt1383* and cultured in 5 ml Luria–Bertani medium containing 50 µg ml^{-1} ampicillin (Wako, Osaka, Japan) at 310 K for 3 h. When the preculture had grown to 1 × 10^8 cells ml^{-1}, it was transferred into 1 l LeMaster medium (Wako; Hendrickson *et al.*, 1990) containing 50 µg ml^{-1} SeMet (Sigma–Aldrich, St Louis, MO, USA), 10 mg ml^{-1} lactose (Wako) and 10 ml Kao and

Michayluk Vitamin Solution (Sigma–Aldrich). After cultivation for 24 h, the cells were harvested by centrifugation and then disrupted by ultrasonication. SeMet-substituted Tt-dNTPase was purified from the cell-free extract as described previously for wild-type Tt-dNTPase (Kondo *et al.*, 2004) and stored at 277 K.

2.2. Crystallization

All crystals used for structural analysis were grown in hanging drops by vapour diffusion. After initial screening using Crystal Screen Cryo (Hampton Research), small crystals were obtained at 293 K by mixing 1 µl protein solution (7.7 mg ml^{-1} protein in 50 m*M* Tris–HCl and 100 m*M* KCl pH 8.0) and 1 µl reservoir solution containing 0.08 *M* Tris–HCl, 0.16 *M* $MgCl_2.6H_2O$, 24% polyethylene glycol 4000 and 20% glycerol pH 8.5. By optimizing the crystallization conditions, flat hexagonal plate-shaped crystals with dimensions of 500 × 50 µm (limiting diameter × thickness) were obtained. These crystals were grown at 293 K by mixing 1 µl protein solution with 1 µl reservoir solution containing 0.08 *M* Tris–HCl, 0.3–0.35 *M* $MgCl_2.6H_2O$, 15–17% polyethylene glycol 4000 and 18–20% glycerol pH 8.6–8.9 for 5–7 d. These optimally diffracting crystals belonged to the triclinic space group *P*1; some of them could be scaled in $P3_1$ or $P3_2$.

2.3. Data collection and processing

Diffraction data for the crystals of the SeMet derivative of the protein were collected using R-AXIS V or Jupiter 210 CCD detectors and synchrotron radiation at beamline BL45XU at SPring-8. Based on the results of a fluorescence scan of the sample to determine the peak wavelength, multiwavelength anomalous dispersion (MAD) or SAD data sets were collected by rotating 1° per frame.

Diffraction images were scaled using the *HKL*-2000 suite (Otwinowski & Minor, 1997). The data sets scaled in $P3_1$ or $P3_2$ were not optimal for the determination of the position of heavy atoms; therefore, the *P*1 data sets were used in the subsequent analysis. Radiation damage to the protein crystals during the collection of the edge and remote data was a major obstacle in determining the position of the heavy atoms and the initial phasing with the MAD method. Initial phases were obtained using the SAD method. The positions of the 20 heavy atoms in the asymmetric unit were determined using the *ISAS* pipe with SAD phasing of scaled data collected at the peak of the absorption edge of Se (Wang, 1985; Liu *et al.*, 2005). Initial phases were determined with *SOLVE* using the ANALYZE_SOLVE option, which refines the position of heavy atoms and calculates phase angles, at 2.1 Å resolution (Terwilliger & Berendzen, 1999).

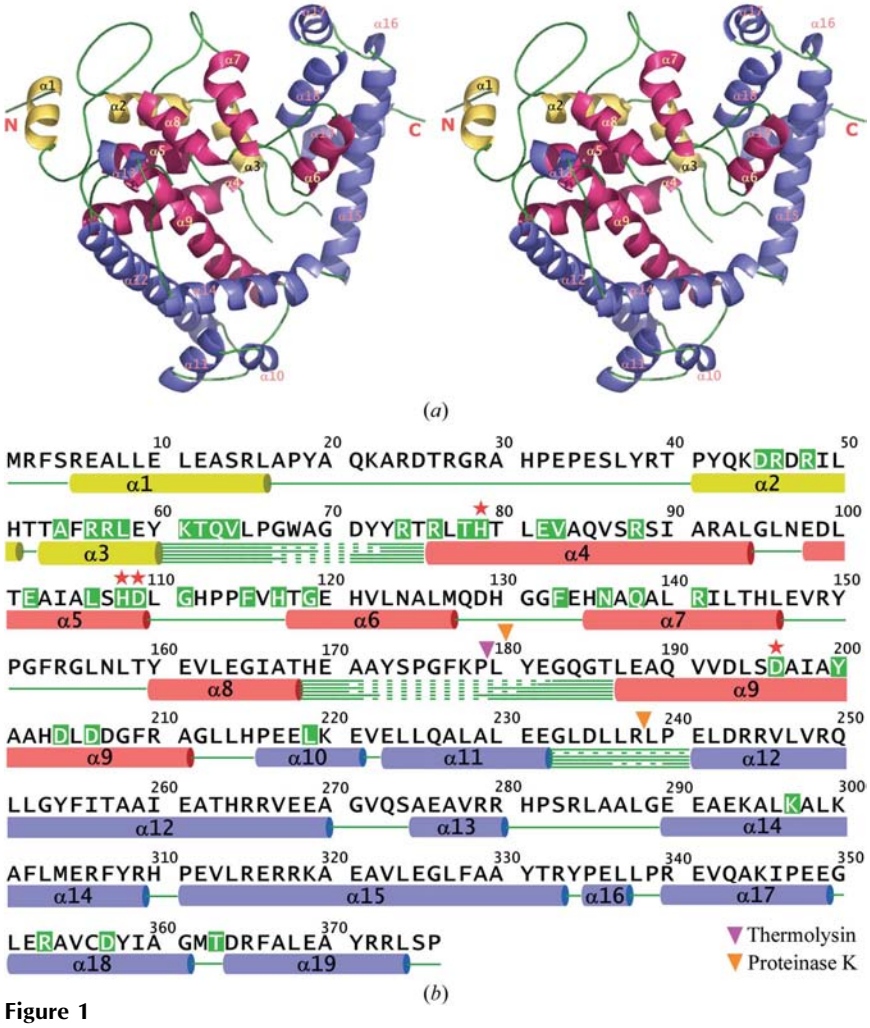

Figure 1
Crystal structure of the subunit. (*a*) Stereoview ribbon diagram of the overall structure of Tt-dNTPase. Chain *A* is represented. Helices α1–α3, α4–α9 and α10–α19 are coloured yellow, red and blue, respectively. (*b*) Secondary-structural elements. Coloured tubes and lines indicate α-helix and loop regions, respectively. Dotted lines represent the regions in which electron density was not observed. The six (dotted) lines in the α3-α4, α8-α9 and α11-α12 regions represent chains *A*–*F* in that order. The colours of the helices are the same as those in (*a*). White letters on green indicate residues that are highly conserved among Tt-dNTPase homologues. Red stars indicate the metal-binding residues in the active site. Magenta and orange triangles indicate the cleavage site on limited proteolysis with thermolysin and proteinase K, respectively.

2.4. Model building and refinement

The initial model was built with *RESOLVE* and *REFMAC*5 using the `resolve_build.csh` script, which carries out iterative model building and refinement (Terwilliger, 2000). Manual model correction was performed with *Xfit* (McRee, 1999). Following rigid-body minimization, simulated annealing, energy minimization and *B*-factor refinement were performed with *CNS* (Brünger *et al.*, 1998). Several rounds of model building using $2F_o - F_c$ maps

allowed us to fit the residue side chains to the model. The progress and validity of the refinement protocol were checked by monitoring the R_{free} value. The oligomeric architecture was determined based on the results of calculation of the self-rotation function (peak observed at χ = 120, 180°). Self-rotation function calculation showed strong peaks indicative of noncrystallographic symmetry elements at (θ, φ, χ) = (90, −120, 180°), (150, 52, 180°), (30, 59, 180°), (88, −35, 120°) and (92, 150, 120°). The positions of the ligand-binding sites were predicted by *WHAT IF* (Vriend, 1990) using a protomer (chain *A*) or an oligomer. Buried surface area calculations were performed using *CNS* (Brünger *et al.*, 1998). The solvent-accessible surface area was calculated with *VOIDOO* (Kleywegt & Jones, 1994). The volumes of the void at the centre of the hexamer and the cavity at the interface of the nearest dimer were also calculated with *VOIDOO* using probe radii of 5.0 and 1.4 Å, respectively. Based on structural comparison with the PDE4–AMP complex (PDB code 1ptw) and the SpoT–ppGp complex (PDB code 1vj7), a putative model of the arrangement of dATP in the HD-domain region of Tt-dNTPase was built with *Xfit* (McRee, 1999). The dATP-bound model structure was refined by Cartesian dynamics annealing with *CNS* (Brünger *et al.*, 1998) within 10 Å of any atom in the dATP molecule. All molecular images were drawn using *PyMOL* (DeLano Scientific; http://www.pymol.org), except for the electrostatic potential, which was drawn using *SPOCK* (Christopher, 1998).

Table 2
Buried surface area at inter-subunit interfaces in the Tt-dNTPase hexamer.

The subunits are labelled with the same letters as in Fig. 2.

Interface	Surface area buried at interface (Å^2)
A–B	1112
A–C	1111
A–D	4708
A–E	0
A–F	370
B–C	1126
B–D	0
B–E	4625
B–F	12
C–D	0
C–E	0
C–F	4797
D–E	1174
D–F	1192
E–F	1250

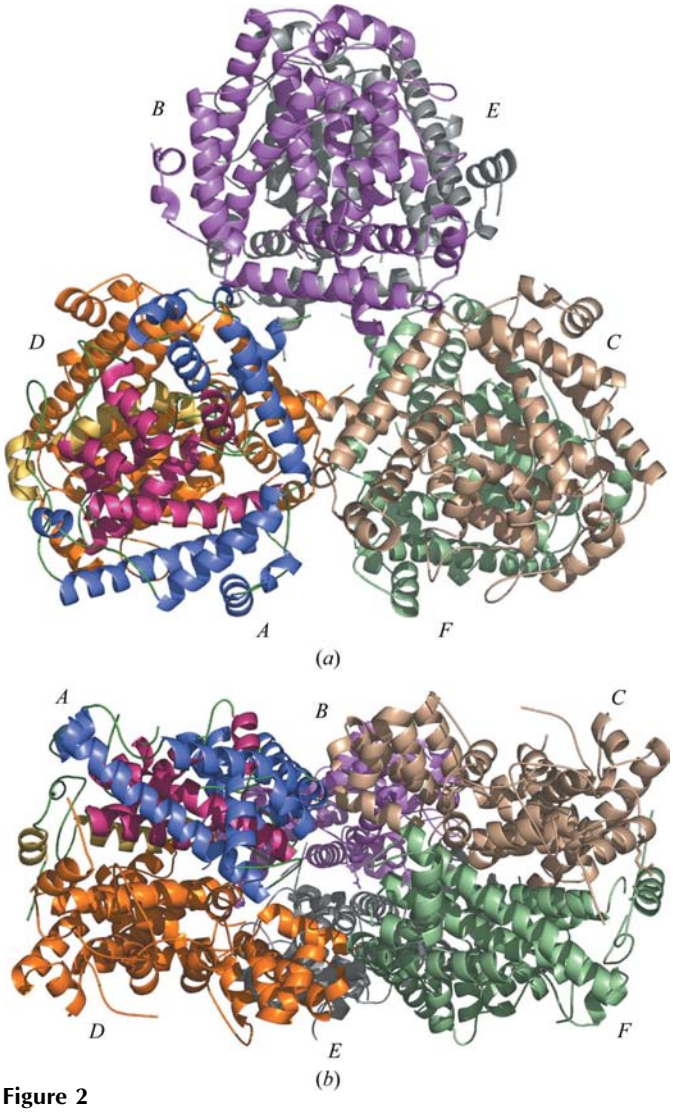

Figure 2
Architecture of the hexameric organization. (*a*) Top view. (*b*) Side view. Chain *A* is represented in the same colours as in Fig. 1. The subunits coloured purple, brown, orange, grey and green indicate chains *B*, *C*, *D*, *E* and *F*, respectively. The view shown in (*b*) represents a 90° rotation from that in (*a*), so that chains *F* and *B* are hidden.

2.5. Limited proteolysis

Tt-dNTPase (20 µ*M*) was treated with two proteases in 50 m*M* Tris–HCl pH 7.5. Digestion with thermolysin (750 n*M*) was performed in the presence of 2 m*M* $CaCl_2$ at 333 K. The reaction was stopped by the addition of SDS–PAGE loading buffer. Digestion with proteinase K (50 n*M*) was performed at 298 K and stopped by the addition of 2 m*M* phenylmethylsulfonyl fluoride (Wako). The digestion products were separated by SDS–PAGE on 12.5% acrylamide gels. The fragments were electroblotted onto a polyvinylidene disulfide membrane and then analyzed using a protein sequencer (Applied Biosystems, Foster City, CA, USA).

2.6. Hydrolase assay

The reaction mixture contained 50 m*M* Tris–HCl pH 7.5, 100 m*M* KCl, 25 m*M* $MgCl_2$, 120 n*M* enzyme, 10 µ*M* dATP, 290 µ*M* dTTP and various concentrations of dCTP, dGTP, dITP, dCDP, dCMP (Sigma, St. Louis, MO, USA), 8-oxo-dGTP or 8-oxo-dATP (Trilink, San Diego, CA, USA). After a reaction time at 310 K appropriate for the initial velocity measurement, the mixture was analyzed by HPLC as described previously (Kondo *et al.*, 2004). The initial velocity was plotted against substrate concentration and the data were fitted to the Michaelis–Menten equation.

3. Results and discussion

3.1. Overall structure of dNTPase

The crystal belongs to space group $P1$ and contains six Tt-dNTPase molecules per asymmetric unit. The final structure was refined to 2.2 Å with R and $R_{\rm free}$ values of 22.3 and 28.7%, respectively. The data and refinement statistics are listed in Table 1. Tt-dNTPase forms a hexamer, which is consistent with the results of size-exclusion analysis in our previous work (Kondo *et al.*, 2004). The subunit is an all-α structure with 19 α-helices (Fig. 1). The individual chains in the Tt-dNTPase structure were named *A*–*F* as shown in Fig. 2. The structures of the six subunits in the asymmetric unit are almost identical to each other. In comparison to chain *A*, the average root-mean-square deviation (r.m.s.d.) values for 376 C^{α} positions in chains *B*, *C*, *D*, *E* and *F* are 0.209, 0.207, 0.191, 0.315 and 0.179 Å, respectively. The ribbon representation of chain *A* is shown in Fig. 1(*a*). The six helices α4–α9 (Fig. 1*a*, red ribbons) are surrounded by three N-terminal helices, α1–α3 (Fig. 1*a*, yellow ribbons), and ten C-terminal helices, α10–α19 (Fig. 1*a*, blue ribbons). The inner helices include the region annotated as the catalytic domain of the HD superfamily; the conserved His and Asp residues are located in helices α4, α5 and α9. The loop regions show relatively high *B* factors, but are located at almost identical positions in each subunit. In every subunit a relatively long loop between α1 and α2 is visible, whereas the α3-α4 and α8-α9 loops at the centre of the subunit are disordered (Fig. 1*b*). The α8-α9 and α11-α12 loops both contain cleavage sites on limited proteolysis (Fig. 1*b*), suggesting that they are positioned at the solvent-accessible surface.

A structural similarity search using *DALI* (Holm & Sander, 1999) revealed that the local structure surrounding the putative active site shares a common architecture with other HD-superfamily proteins (see §3.3).

3.2. Oligomeric architecture

The overall shape of the hexamer can be viewed as a double ring of trimers in face-to-face contact. As seen in Fig. 2, two subunits in a 'trimer' unit (*e.g.* chains *A* and *B*) are likely to have less interaction than those in a 'dimer' unit (*e.g.* chains *A* and *D*). The α2, α3, α4, α5, α9, α12 helices and the α3-α4 loop form a dimer interface, whereas the α14, α15, α16 and α19 helices form a small interface in the trimer near the central hole. The buried surface areas in the subunit–subunit interfaces of the 'dimer' and 'trimer' are 4600–4800 and 1100–1300 Å^2, respectively (Table 2). These results indicate that contact at the dimer interface is stronger than at the trimer interface, but nevertheless these values are within the range of the area expected for a stable protein complex (Jones & Thornton, 1996). Thus, our structural analysis suggests that these two interfaces are responsible for the formation of a stable hexameric structure. Indeed, fragments produced by limited proteolysis still adopt a higher order structure with an apparent molecular weight similar to that of the native enzyme in addition to exhibiting weak hydrolytic activity (data not shown). Intriguingly, no other member of the HD superfamily of proteins is known to form a hexamer.

A large void is observed in the interior of the hexamer, which is accessible to solvent *via* two types of pore (Fig. 3). At the exposed (upper) surface of the trimer ring plane, the diameter of the central hole (Fig. 3, pore 1) is in the range 13–16 Å [12.7 Å from CB of Ser375 (chain *A*) to the side-chain OH of Tyr371 (chain *C*)]. A gap in the side surface of the ring is located at the centre of the trimer interface; for example,

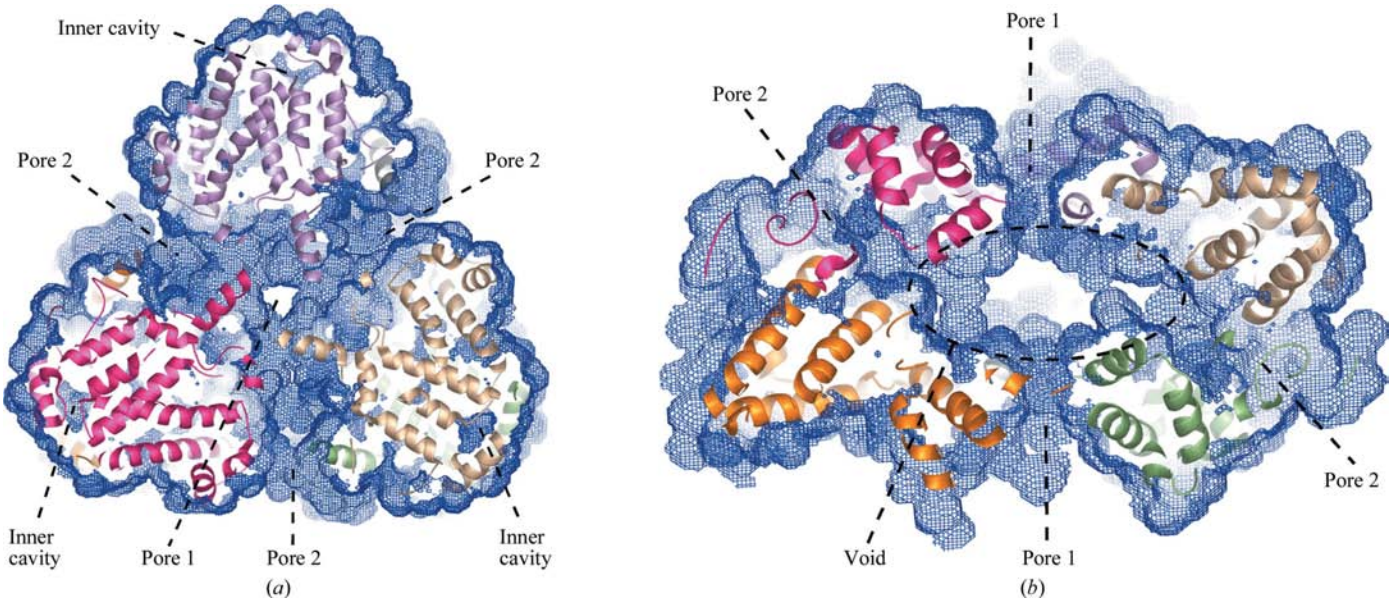

Figure 3
Solvent-accessible surface of Tt-dNTPase. (*a*) Top view. (*b*) Side view. The surface is removed to show the central void and pores. The subunits are represented in the same colours as in Fig. 2, except for chain *A*, which is coloured pink. The surface map was calculated using *VOIDOO* (Vriend, 1990). The images were drawn using *PyMOL* (DeLano Scientific; http://www.pymol.org).

surrounded by chains *A*, *C*, *D* and *F* (Fig. 3, pore 2). Although the precise dimensions of the hole are difficult to define owing to the disordered nature and the high *B* value of several elements such as α10 and the α11-α12 loop in this region, its average diameter is larger than that of the central hole. The volume of the void at the centre of the hexamer is approximately 24 000 Å^3. Moreover, the electrostatic potential diagram of the hexamer revealed that the central void in the hexamer is basic (Fig. 4). Wurgler & Richardson (1993) reported that Ec-dGTPase binds to DNA, especially ssDNA, in the presence of NaCl. The narrowest diameter of the pore 1 is 13 Å, which is smaller than the diameter of dsDNA but greater than that of ssDNA. The residues comprising the basic void, Arg57, Lys61, Arg74 and Arg76, are highly conserved among Tt-dNTPase/Ec-dGTPase homologues (Fig. 1*b*). As the interior void of the hexamer is solvent-accessible through two types of pore (Fig. 3), it can be suggested that the ssDNA molecule binds within the basic pore/void of the Tt-dNTPase hexamer. Another possibility is that the interior void acts as the dNTP-binding pocket. This will be discussed later.

3.3. Active sites

A *DALI* search (Holm & Sander, 1999) using the full-length Tt-dNTPase as the query sequence gave the highest similarity to three HD-superfamily members. A hypothetical protein from *Archaeoglobus fulgidus* (PDB code 1ynb), dNMP-specific nucleotidase from *E. coli* (PDB code 1wph; Proudfoot *et al.*, 2004) and cyclic GMP-specific 3′,5′-cyclic phosphodiesterase from *Homo sapiens* (PDB code 1rkp; Huai *et al.*, 2004) showed *Z* scores of 7.4, 6.9 and 5.9, respectively. Tt-dNTPase can be structurally aligned with these three proteins with r.m.s.d. values of 5.7, 6.9 and 5.0 Å for the C$^\alpha$ atoms of residues 51–305, 51–299 and 51–304, respectively. This analysis suggests that the active site of Tt-dNTPase has similar architecture to these HD-superfamily proteins.

Some highly conserved charged residues, including His79, His108, Asp109 and Asp196 (Fig. 1, red stars), are clustered in a pocket at the centre of the subunit. The calculated electrostatic potential revealed this pocket to be particularly acidic compared with the other regions (Fig. 4*a*, black arrow), suggesting that the pocket can bind to positively charged species such as metal ions. It has been established that these conserved residues constitute the catalytic centre by chelating metal ions in other HD-superfamily proteins (Proudfoot *et al.*, 2004; Huai *et al.*, 2003, 2004; Hogg *et al.*, 2004).

The probable locations of the active site and ligand-binding site of Tt-dNTPase were determined by structural comparison with two HD-superfamily proteins: phosphodiesterase 4 (PDE4) and (p)ppGpp synthetase/hydrolase (SpoT). The results of the *DALI* search (Holm & Sander, 1999) revealed similarity to PDE4 (PDB code 1ptw) and SpoT (PDB code 1vj7), with *Z* scores of 5.9 and 6.4 and r.m.s.d.s of 4.8 and 4.5 Å for 152 and 135 C$^\alpha$ atoms, respectively. The crystal structures of the PDE4–AMP and SpoT–ppGp complexes have been determined (Huai *et al.*, 2003; Hogg *et al.*, 2004). In these structures, the active site is surrounded by five helices made up of highly conserved residues. Tt-dNTPase also shows a similar arrangement of five α-helices: α4, α5, α7, α8 and α9 (Fig. 5*a*). The length and direction of the elemental helices are not highly conserved in these HD-superfamily proteins, whereas the location and orientation of the conserved catalytic residues, two His and two Asp, are well conserved. In particular, the orientations of the side chains of these residues are important for chelating metal ions. In the $2F_o - F_c$ electron-density map, spherical electron density was observed in the putative active site close to the highly conserved residues His79, His108, Asp109 and Asp196 (Fig. 5*b*). This electron density is in an equivalent location to the metal ions in the structures of PDE4 and SpoT. Because the crystals were grown in a solution containing Mg^{2+}, this spherical density was interpreted as an Mg^{2+} ion.

Further analysis using the program *WHAT IF* (Vriend, 1990) was performed to identify the binding sites for a ligand in the Tt-dNTPase monomer. This program can be employed to detect the largest dent on the surface of a target molecule. Among the residues identified, Ala54, Arg56–Leu58, Lys61, Arg76, Asp109, Gly111, His112, Phe133, Gln138, Arg141, Arg353, Asp357

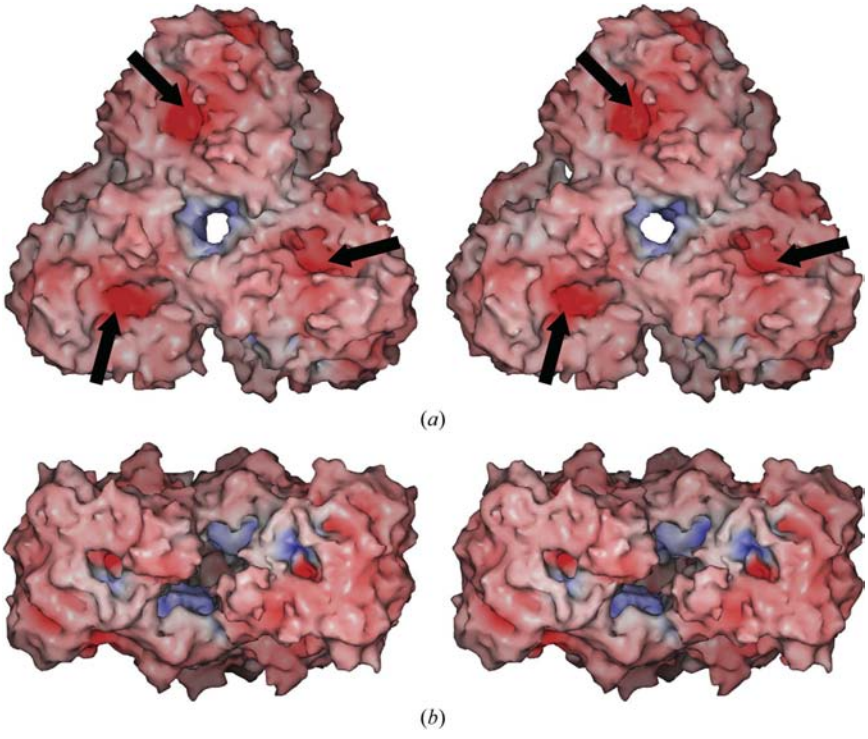

Figure 4
Electrostatic potential map of the Tt-dNTPase hexamer drawn in stereo. Positive potentials are drawn in blue and negative potentials in red. The view in (*b*) represents a 90° rotation from (*a*). Black arrows indicate probable metal-chelating pockets.

Table 3
Substrate dNTP specificity of Tt-dNTPase.

The reaction solution contained 50 m*M* Tris–HCl pH 7.5, 100 m*M* KCl, 25 m*M* $MgCl_2$, 10 µ*M* dATP, 290 µ*M* dTTP and the other dNTPs at various concentrations.

Substrate	K_m (µ*M*)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} M^{-1})
dCTP	400 ± 28	0.86 ± 0.019	2.2×10^3
8-oxo-dGTP	470 ± 90	0.64 ± 0.042	1.4×10^3
dGTP	470 ± 62	0.61 ± 0.026	1.3×10^3
dITP	728 ± 64	0.43 ± 0.015	5.9×10^2
8-oxo-dATP	770 ± 47	0.21 ± 0.0053	2.9×10^2

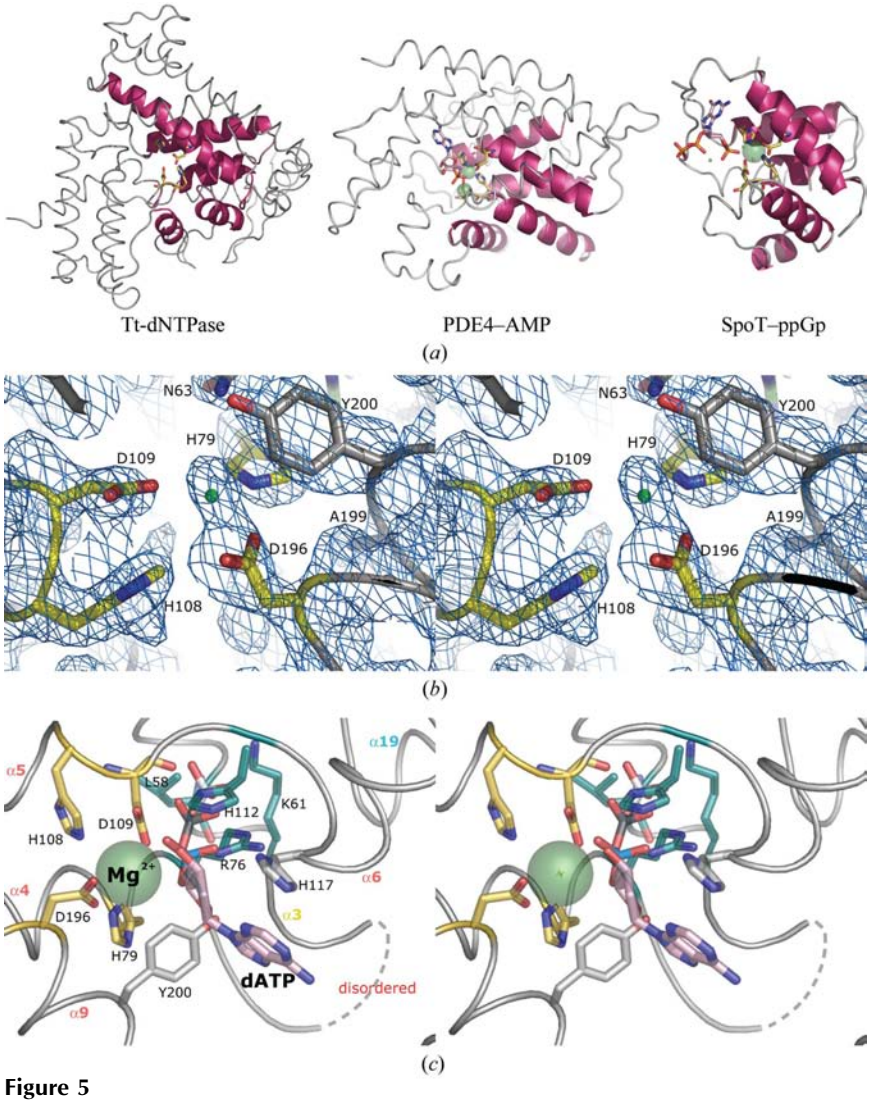

Figure 5
Structure of the putative active site. (*a*) Structural comparison of Tt-dNTPase (subunit) with the PDE4–AMP complex (PDB code 1ptw) and the SpoT–ppGp complex (PDB code 1vj7). Structurally similar elements are displayed as thick red ribbons. The four conserved residues and ligand molecules (AMP and ppGp) are shown as stick representations in yellow and pink, respectively. Zn^{2+} or Mn^{2+} ions (green) and Mg^{2+} ions (grey) are shown as transparent spheres. (*b*) The putative active site of Tt-dNTPase. The difference Fourier map ($2F_o - F_c$ map, $\sigma = 1.0$) is superimposed on the model drawn in stereo. (*c*) The model of the substrate (dATP) bound structure of the active site with Mg^{2+}. The figure is drawn in stereo. The putative key residues are represented as sticks. The catalytic residues that are highly conserved among HD-superfamily proteins are coloured yellow. The residues highly conserved among Tt-dNTPase/Ec-dGTPase homologues that were predicted to be a ligand-binding site by computational search are coloured blue. The residues highly conserved among Tt-dNTPase/Ec-dGTPase homologues that are located near to the ribose moiety of dATP are coloured grey. The disordered region of the α3-α4 loop is represented as a dashed line.

and Thr363 are highly conserved among Tt-dNTPase/Ec-dGTPase homologues (Fig. 1*b*). These residues are situated near to the active site predicted by structural comparison.

Based on our analysis of the likely location of the active site, we constructed a model of ligand-bound Tt-dNTPase. Initially, a dATP molecule was placed as a ligand within the active site of the protein structure containing Mg^{2+} so that the sugar and base were in an equivalent position to AMP in PDE4 and ppGp in SpoT. Molecular-dynamics calculations were performed to calculate a model of the protein–ligand interactions (Fig. 5*c*). The phosphate groups were positioned close to the α5-α6 loop and inserted into the space surrounded by residues Lys61, Arg76, Asp109 CO, Leu58 NH and His112. These residues are highly conserved in Tt-dNTPase/Ec-dGTPase homologues (Fig. 1*b*) and were predicted to constitute the ligand-binding pocket by the binding-site search *WHAT IF*. The ribose moiety of bound dNTP is situated near His117 and Tyr200, which are the highly conserved residues in Tt-dNTPase/Ec-dGTPase homologues (Fig. 5*c*).

In the case of PDE4, His160 and aromatic residues such as Phe340 and Phe372 make van der Waals contacts with the ribose moiety of the bound AMP (Huai *et al.*, 2003). It should be noted here that Tt-dNTPase has a high preference for deoxyribose over ribose (Kondo *et al.*, 2004). In addition, His117 at the tip of the α6 helix has a relatively high *B* factor and is even disordered in several subunits. A fixed conformation of this residue (and the α6 helix) may only occur upon binding of the substrate. Although such structural changes may be associated with specific recognition of the deoxyribose, it is uncertain which residue is responsible for recognition of the sugar moiety.

A few residues seem to be located at a suitable distance to interact with the adenine moiety of the modelled dATP. The nearest residue, His117, is located about 3.5 Å away from the base moiety. However, the model of dATP in the crystal structure predicts that the disordered α3-α4 loop is located relatively near the base moiety (Fig. 5*c*). Like the α6 helix described above, the α3-α4 loop may take a fixed conformation upon dNTP binding. Interestingly, this loop includes highly conserved residues (61-KTQV-64), suggesting functional significance (Fig. 1*b*). Conformational flexibility of this loop may be associated with the broad specificity for the bases of different dNTPs. The apparent lack of specific base recogni-

tion in the determined structure is consistent with the observation that Tt-dNTPase hydrolyzes all canonical dNTPs in the presence of dATP and dTTP (Kondo *et al.*, 2004).

3.4. Substrate specificity

Structural features further prompted us to explore the possibility that Tt-dNTPase hydrolyzes noncanonical dNTPs. Tt-dNTPase actually hydrolyzed dITP, 8-oxo-dGTP and 8-oxo-dATP as well as dCTP and dGTP in the presence of dATP and dTTP (Fig. 6*a*). Both dATP and dTTP were also hydrolyzed, but were excluded from the comparison because these also acted as activators. The experimental data for all dNTPs examined fitted well to the Michaelis–Menten equation. The K_m values increased in following order: dCTP > 8-oxo-dGTP $\simeq$ dGTP > dITP > 8-oxo-dATP, whereas the k_{cat} values decreased in the following order: dCTP, 8-oxo-dGTP, dGTP, dITP, dATP, dTTP (Table 3). Although dCTPs were more favourable substrates for Tt-dNTPase under these conditions, the K_m and k_{cat} values for the dNTPs examined differed only slightly. These results confirmed that the broad specificity for the base moiety of dNTP is an inherent feature of Tt-dNTPase.

In contrast to the base moiety, the phosphate moiety of the bound dNTP is likely to be recognized by several residues. Thus, we assessed the hydrolytic activity for deoxyribonucleoside diphosphate and monophosphate (Fig. 6*b*). dCDP was hydrolyzed to deoxycytidine and pyrophosphate, but its reaction rate (k_{app}) was lower by more than two orders of magnitude compared with that of dCTP: the k_{app} value was at most 0.007 s^{-1}. The data for dCDP hydrolysis could not be fitted to Michaelis–Menten equation (Fig. 6*b*, inset), which might partly be because of an inhibitory effect of dCDP on the binding of dATP or dTTP. The hydrolysis of dCMP was scarcely detected: its reaction rate was below 0.0001 s^{-1}. These results suggest that Tt-dNTPase efficiently recognizes the phosphate moiety of dNTP. The relatively high specificity for the triphosphate group, which is crucial for triphosphohydrolase-type reaction, might be rationalized by putative involvement of highly conserved residues with positive charge.

3.5. Insights into the activation mechanism

Our previous characterization of Tt-dNTPase activity suggested the presence of more than two binding sites for dNTPs (Kondo *et al.*, 2004). As shown above, we identified one plausible dNTP-binding site per monomer of protein. This site is comprised of the region annotated as the catalytic domain of the HD superfamily, strongly suggesting that it is a catalytic site for the hydrolysis of dNTP. As the previous study suggested the presence of an activator-recognition site separate from the catalytic site, we further investigated the presence of additional ligand-binding sites in the Tt-dNTPase structure. Although our search analysis of ligand-binding sites using *WHAT IF* (Vriend, 1990) revealed only one plausible site, the analysis was only applied to the Tt-dNTPase monomer. Therefore, we performed a similar analysis by *WHAT IF* for the Tt-dNTPase hexamer and found a distinct region separate from the predicted catalytic site. This site is situated at the trimer–trimer interface (Fig. 7) and has a region primarily composed of eight α-helices (α1, α2, α4 and α5) from two subunits of the dimer unit. The following residues form

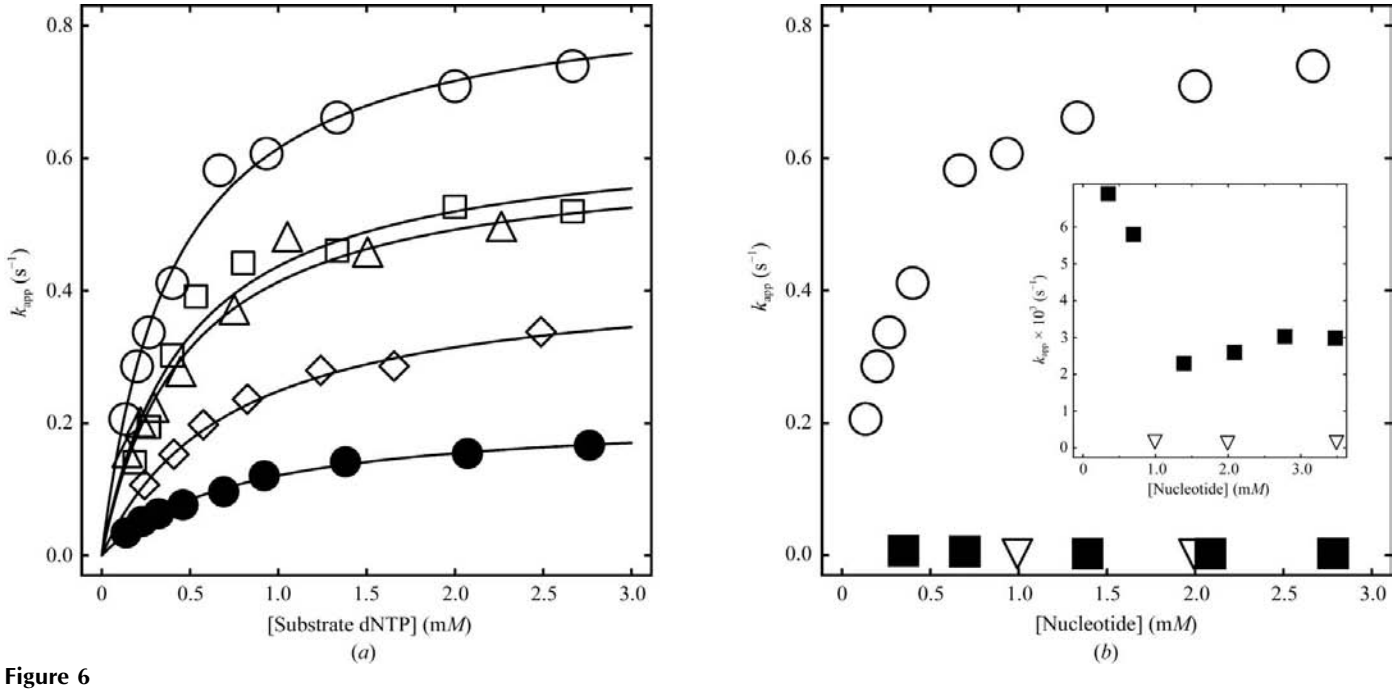

Figure 6
Hydrolase activity for various deoxyribonucleotides. The measurements were performed in 50 m*M* Tris–HCl pH 7.5, 100 m*M* KCl, 25 m*M* $MgCl_2$, 10 μ*M* dATP, 290 μ*M* dTTP with various concentrations of dNTPs. (*a*) The activities for canonical or noncanonical dNTPs. The substrates were dCTP (open circles), dGTP (open triangles), 8-oxo-dGTP (open squares), dITP (open diamonds) and 8-oxo-dATP (filled circles). (*b*) The activities for dCTP (open circles), dCDP (filled squares) and dCMP (open inverted triangles). The inset shows an expanded view of the data for dCDP and dCMP.

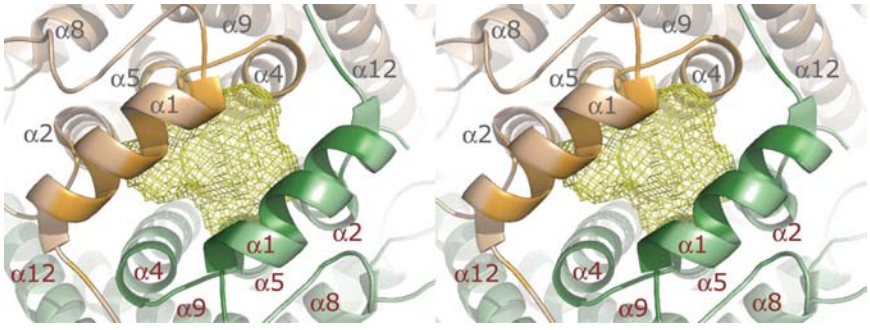

Figure 7
Putative dNTP-binding site. Chains *C* and *F* are coloured orange and green, respectively. The residues comprising the putative binding site predicted by *WHAT IF* (Vriend, 1990) are represented in deep colours. The cavity detected by *VOIDOO* (Kleywegt & Jones, 1994) is represented as a yellow mesh.

the walls of this site: Phe3, Ala7–Leu9, Leu11, Glu12, Arg15–Ala17, Thr40, Gln43, Arg46, Arg88, Ala91–Asp99 and Glu102 from each subunit (for example, chains *C* and *F*), Ala23 from one subunit (chain *F*) and Leu100 from the other subunit (chain *C*) (Fig. 7). The spatial arrangement of these residues in chains *C* and *F* are almost symmetrical within the site. The site is of approximately 19 Å in length, with a maximum width of 7.8 Å. The same cavities in the dimers were also detected with *VOIDOO* (Kleywegt & Jones, 1994). The detected cavities can also be seen as the 'inner cavity' in Fig. 3. Among the contributing residues, Leu16, Ala17, Leu94, Leu96 and Asp99 are highly conserved in dGTPase homologues, with the exception of the *E. coli* enzyme (Kondo *et al.*, 2004). The calculated volume of the cavities between chains *A*–*D*, *B*–*E* and *C*–*F* are 880, 770 and 700 Å^3, respectively. The dimensions of these cavities seem to be slightly too small to accommodate dNTP, with a maximum length of 17 Å. In addition, the cavity is inaccessible from the exterior of the protein (Fig. 7). Therefore, a conformational change of the region near the cavity, such as the long (mobile) loop between α1 and α2, is required to allow a dNTP molecule to bind in this putative binding site. A dNTP molecule may enter the cavity *via* a conformational change from the central void, which is accessible to solvent through pores and provides sufficient room to accommodate a nucleotide.

Although we can infer possible mechanisms of induction of Tt-dNTPase activity *via* allosteric interactions between two plausible dNTP-binding sites, this is at present too speculative to discuss in detail. It should be mentioned, however, that a regulatory mechanism involving multiple dNTPs is known for allosteric enzymes concerned with dNTP metabolism. RNR retains a complicated regulatory mechanism of allosteric coupling between the effector-binding site and the catalytic site (Jordan & Reichard, 1998). The regulatory site for dNTP/ATP is situated at the interface of the α and β domains, whereas the catalytic site for NDP is in the centre of the β domain. The specificity of the enzyme for substrate is determined by the nature of the bound effector. This sophisticated mechanism for modulating enzyme activity is required for controlling the intracellular concentration of dNTPs to ensure precise DNA replication. Ec-dGTPase was reported to retain DNA-binding activity, implying the linkage of Ec-dGTPase/Tt-dNTPase homologues to DNA metabolism. A regulatory mechanism for Tt-dNTPase involving multiple dNTPs suggests that the activity of this enzyme is of physiological importance to the cell.

The authors would like to acknowledge Taisuke Wakamatsu, Hirohito Ishikawa and Dr Kei Wada for helpful advice during data analysis. This work was partly supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, Science and Technology of Japan. This work was partly supported in part by Grant-in-Aid for Scientific Research 17770089 (to NN) from the Ministry of Education, Science, Sports and Culture of Japan. NK is the recipient of a Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists (No. 8759).

References

Aravind, L. & Koonin, E. V. (1998). *Trends Biochem. Sci.* **23**, 469–472.
Bebenek, K. & Kunkel, T. A. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 4946–4950.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chabes, A., Georgieva, B., Domkin, V., Zhao, X., Rothstein, R. & Thelander, L. (2003). *Cell*, **112**, 391–401.
Christopher, J. A. (1998). *SPOCK: The Structural Properties Observation and Calculation Kit.* Texas A&M University, College Station, TX, USA.
Eklund, H., Uhlin, U., Farnegardh, M., Logan, D. T. & Nordlund, P. (2001). *Prog. Biophys. Mol. Biol.* **77**, 177–268.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Hogg, T., Mechold, U., Malke, H., Cashel, M. & Hilgenfeld, R. (2004). *Cell*, **117**, 57–68.
Holm, L. & Sander, C. (1999). *Nucleic Acids Res.* **27**, 244–247.
Huai, Q., Colicelli, J. & Ke, H. (2003). *Biochemistry*, **42**, 13220–13226.
Huai, Q., Lin, Y., Francis, S. H., Corbin, J. D. & Ke, H. (2004). *J. Biol. Chem.* **279**, 13095–13101.
Jones, S. & Thornton, J. M. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
Jordan, A. & Reichard, P. (1998). *Annu. Rev. Biochem.* **67**, 71–98.
Kleywegt, G. J. & Jones, T. A. (1994). *Acta Cryst.* D**50**, 178–185.
Kondo, N., Kuramitsu, S. & Masui, R. (2004). *J. Biochem.* **136**, 221–231.
Kornberg, S. R., Lehman, I. R., Bessman, M. J., Simms, E. S. & Kornberg, A. (1958). *J. Biol. Chem.* **233**, 159–162.
Liu, Z.-J., Lin, D., Tempel, W., Praissman, J. L., Rose, J. P. & Wang, B.-C. (2005). *Acta Cryst.* D**61**, 520–527.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Moroz, O. V., Murzin, A. G., Makarova, K. S., Koonin, E. V., Wilson, K. S. & Galperin, M. Y. (2005). *J. Mol. Biol.* **347**, 243–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Proudfoot, M., Kuznetsova, E., Brown, G., Rao, N. N., Kitagawa, M., Mori, H., Savchenko, A. & Yakunin, A. F. (2004). *J. Biol. Chem.* **279**, 54687–54694.
Quirk, S. & Bessman, M. J. (1991). *J. Bacteriol.* **173**, 6665–6669.

Sekiguchi, M. & Tsuzuki, T. (2002). *Oncogene*, **21**, 8895–8904.
Seto, D., Bhatnagar, S. K. & Bessman, M. J. (1988). *J. Biol. Chem.* **263**, 1494–1499.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Warner, H. R., Duncan, B. K., Garret, C. & Neuhard, J. (1981). *J. Bacteriol.* **145**, 687–695.
Wurgler, S. M. & Richardson, C. C. (1993). *J. Biol. Chem.* **268**, 20046–20054.