

# The structure of the genomic *Bacillus subtilis* dUTPase: novel features in the Phe-lid

Javier García-Nafria, Lynn Burchell, Mine Takezawa, Neil J. Rzechorzek, ‡ Mark J. Fogg and Keith S. Wilson\*

Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, England

‡ Present address: Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, England.

Correspondence e-mail: keith@ysbl.york.ac.uk

Received 24 May 2010

Accepted 4 July 2010

**PDB References:** genomic *B. subtilis* dUTPase, 2xcd; complex with dUpNHpp and calcium, 2xce.

dUTPases are a ubiquitous family of enzymes that are essential for all organisms and catalyse the breakdown of 2-deoxyuridine triphosphate (dUTP). In *Bacillus subtilis* there are two homotrimeric dUTPases: a genomic and a prophage form. Here, the structures of the genomic dUTPase and of its complex with the substrate analogue dUpNHpp and calcium are described, both at 1.85 Å resolution. The overall fold resembles that of previously solved trimeric dUTPases. The C-terminus, which contains one of the conserved sequence motifs, is disordered in both structures. The crystal of the complex contains six independent protomers which accommodate six dUpNHpp molecules, with three triphosphates in the *trans* conformation and the other three in the active *gauche* conformation. The structure of the complex confirms the role of several key residues that are involved in ligand binding and the position of the catalytic water. Asp82, which has previously been proposed to act as a general base, points away from the active site. In the complex Ser64 reorients in order to hydrogen bond the phosphate chain of the substrate. A novel feature has been identified: the position in the sequence of the 'Phe-lid', which packs against the uracil moiety, is adjacent to motif III, whereas in all other dUTPase structures the lid is in a conserved position in motif V of the flexible C-terminal arm. This requires a reconsideration of some aspects of the accepted mechanism.

## 1. Introduction

Deoxyuridine 5'-triphosphate nucleotidohydrolases (dUTPases) catalyze the hydrolysis of dUTP to dUMP and inorganic pyrophosphate in an ion-dependent manner (Shlomai & Kornberg, 1978; Bertani *et al.*, 1963). They are ubiquitous enzymes that are found in eukaryotes, prokaryotes and viruses (Baldo & McClure, 1999) and are essential for viability, as proven in *Escherichia coli* (el-Hajj *et al.*, 1988) and *Saccharomyces cerevisiae* (Gadsden *et al.*, 1993). Their role in ensuring the fidelity of DNA replication, transcription and uracil base-excision repair is achieved by reducing the dUTP concentration while simultaneously increasing the concentration of the dTTP precursor dUMP. The dTTP:dUTP ratio is thereby increased, reducing the likelihood of dUTP misincorporation. dUTPases are classified into three different families based on their oligomeric assembly: monomeric, homodimeric (termed dUTPase/dUDPase) and homotrimeric. While the homotrimeric and monomeric dUTPases share a common fold, with a gene duplication giving rise to the monomeric form, dUTPase/dUDPases show no sequence or structural similarity to the other two families.

The homotrimeric dUTPases are the most studied family and all share a common fold (Vértessy & Tóth, 2009). The first

crystal structure of a dUTPase was the homotrimeric *E. coli* enzyme (Cedergren-Zeppezauer *et al.*, 1992), but several structures of orthologues from a number of organisms and complexes with dUMP, dUDP, dUTP and dUpNHpp and Mg<sup>2+</sup> have subsequently been reported (Varga *et al.*, 2008; Chan *et al.*, 2004; González *et al.*, 2001; Prasad *et al.*, 1996, 2000; Dauter *et al.*, 1998, 1999; Larsson *et al.*, 1996; Mol *et al.*, 1996). They form trimers around a threefold axis and have a central channel with extensive subunit interactions.

These structures, when combined with kinetic data, have provided detailed insights into the active-site geometry and composition as well as the catalytic mechanism. There are three active sites per trimer, formed by five conserved motifs (McGeoch, 1990), and all three subunits contribute to the formation of each active site. Four of the motifs are contributed by two adjacent subunits. Motif V comes from the third subunit and is located in the C-terminal arm, which is disordered in most of the published structures. However, the arm takes up an ordered conformation in the structures of *Mycobacterium tuberculosis* dUTPase in complex with Mg<sup>2+</sup>-dUpNHpp, the feline immunodeficiency virus (FIV) dUTPase in complex with dUDP and the human dUTPase complexes

with dUMP, dUDP and dUTP (Prasad *et al.*, 2000; Chan *et al.*, 2004; Tóth *et al.*, 2007; Varga *et al.*, 2008).

*Bacillus subtilis* has two distinct trimeric dUTPases, a pro-phage enzyme (YosS) and a genomic enzyme (YncF), with 93% sequence identity (Fig. 1). The structure of YosS has been reported previously (PDB code 2baz; J. Wang, Y.-H. Liang & X.-D. Su, unpublished work) and crystals diffracting to 2.7 Å resolution have been reported for YncF (Li *et al.*, 2009). In the present work, the genomic YncF has been expressed, purified and crystallized and structures have been solved at 1.85 Å spacing for the apoenzyme and the complex with dUpNHpp, a nonhydrolysable analogue of dUTP. While the two structures are very similar to known homotrimeric dUTPases, the complex with dUpNHpp reveals novel features in the active site.

## 2. Methods

### 2.1. Expression and purification

The DNA containing the coding sequence for the *B. subtilis* genomic dUTPase was cloned into the pET-26(+) (Novagen) vector. The plasmid was transformed into B834 (DE3) com-

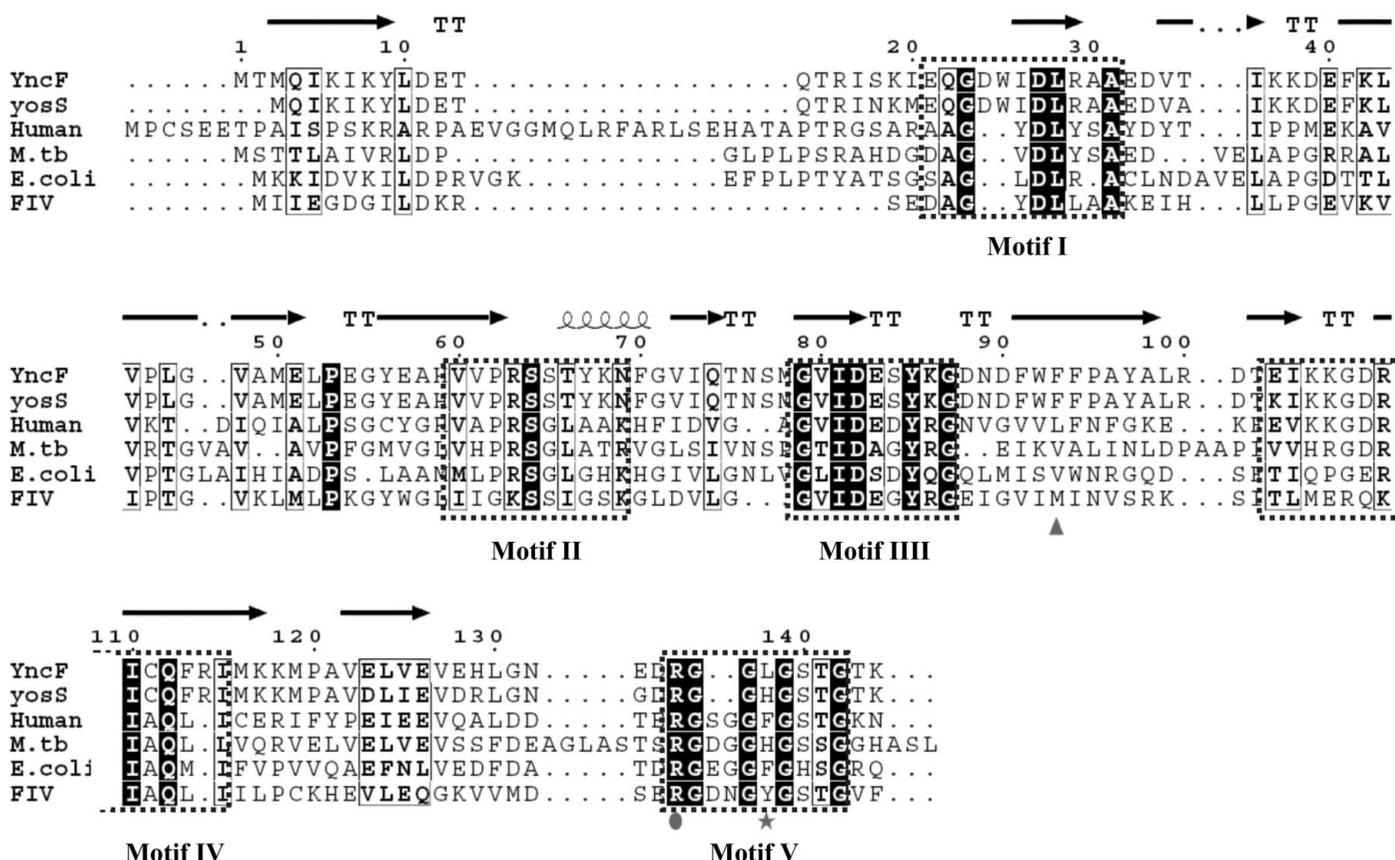


Figure 1

Sequence alignment of a set of representative trimeric dUTPases for which structures are known. Conserved residues are shown on a black background, while boxes indicate the conservation of residues with similar physicochemical properties. The secondary-structure depiction and residue numbering above the sequences corresponds to YncF. The five conserved functional motifs are labelled below the sequences and outlined by dashed boxes. YncF is the only dUTPase that lacks an aromatic residue at position 138 (indicated by a star) and YncF and YosS are the only dUTPases that have an aromatic residue at position 93 (indicated by a triangle), the main chain of which is responsible for uracil binding. The conserved arginine of motif V is indicated by a circle.

**Table 1**

X-ray data and refinement statistics for the crystals of the *B. subtilis* genomic dUTPase (YncF) and its complex with the non-hydrolysable substrate analogue dUpNHpp.

For data collection, the values in parentheses correspond to the outer shell of reflections. For refinement deviations, the values in parentheses are the target values.

	Apo YncF	Complex with dUpNHpp
<b>Data collection</b>		
Source	ESRF	Diamond Light Source
Beamline	ID14-2	MX306104
Wavelength (Å)	0.933	0.9704
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	$a = 98.81, b = 98.84, c = 99.11$	$a = 99.12, b = 99.25, c = 99.35$
Resolution (Å)	31.33–1.85 (1.95–1.85)	40.54–1.85 (1.95–1.85)
$R_{\text{sym}}$ or $R_{\text{merge}}$ (%)	7.4 (88.1)	11 (94.4)
$I/\sigma(I)$	12.8 (1.5)	15.6 (2.4)
Completeness (%)	97.9 (99.2)	99.9 (100)
Redundancy	5.5 (5.0)	9.4 (7.4)
<b>Refinement</b>		
No. of reflections	77321	78918
$R_{\text{work}}/R_{\text{free}}$ (%)	15.4/18.4	15.1/18.7
Twinning operators	$h, k, l, 0.51; k, -l, -h, 0.31; -l, -h, k, 0.18$	Untwinned
Protein atoms	6402	6397
Ligand/ion atoms	17	212
Water molecules	737	793
<b>B factors (Å<sup>2</sup>)</b>		
Protein	28.5	36.1
Ligand/ions	30.1	24.8
Waters	34.4	35.1
<b>R.m.s. deviations</b>		
Bond lengths (Å)	0.013	0.023
Bond angles (°)	1.424	2.007
<b>Ramachandran plot (%)</b>		
Most favoured	90.9	91.8
Additionally allowed	8.2	7.3
Generously allowed	0	0
Disallowed	0.9	0.9

petent cells and expression was carried out by auto-induction at 303 K overnight (Studier, 2005). The cells were harvested by centrifugation and the resulting pellet was resuspended in buffer A (50 mM NaCl, 20 mM MES, 1 mM DTT pH 5.5) and lysed by sonication. The supernatant was loaded onto a phosphocellulose cation-exchange column and the protein was eluted with a gradient of buffer B (20 mM MES pH 5.5, 1 M NaCl, 1 mM DTT). The resulting sample was dialysed against 50 mM Tris-HCl pH 8, 150 mM NaCl, concentrated and stored at 193 K.

## 2.2. Crystallization

Protein at 15 mg ml<sup>-1</sup> in 50 mM Tris pH 7.5, 150 mM NaCl was screened for crystallization. The commercial screens PACT (Molecular Dimensions), Index, Crystal Screen and Crystal Screen 2 (Hampton Research) were used to set up 96-well screens with 150 + 150 nl drops in MRC plates using Hydra 96 (Robbins Scientific) and Mosquito (TTP LabTech Ltd, UK) robots. This protocol was followed for both the apoenzyme and the enzyme complexed with dUpNHpp at a 5:1 ratio (ligand:protein) and 10 mM MgCl<sub>2</sub>.

Crystals of the apoenzyme that were suitable for data collection were obtained using protein at 8 mg ml<sup>-1</sup> from

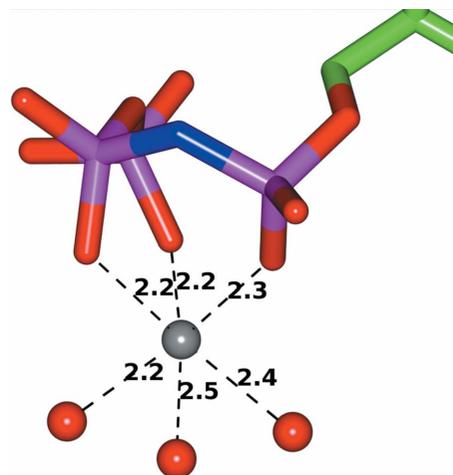
0.2 M MgCl<sub>2</sub>, 28% PEG 550 and 0.1 M HEPES pH 7.5 and were cryoprotected with 30% PEG 550 prior to exposure. Crystals of the complex were optimized at 12 mg ml<sup>-1</sup> using 0.2 M CaCl<sub>2</sub>, 16% PEG 6K and 0.1 M HEPES pH 7 and were cryoprotected with 20% glycerol.

## 2.3. Data collection

For the apoenzyme, data were collected on ESRF beamline ID14-2 ( $\lambda = 0.933$  Å) to a  $d$ -spacing of 1.85 Å. For the complex, data to 1.85 Å resolution were collected on Diamond Light Source beamline I04 ( $\lambda = 0.9704$  Å). The data are summarized in Table 1.

## 2.4. Structure solution and refinement

Crystallization experiments were performed using full-length YncF (residues 2–144; the mature enzyme lacking the N-terminal methionine). The crystals of the apoenzyme and of the complex were isomorphous, belonging to space group  $P2_12_12_1$  with six protomers in the asymmetric unit arranged as two trimers. The structure of the apoenzyme was solved by molecular replacement using the program *Phaser* (McCoy *et al.*, 2007) with the *B. subtilis* prophage YosS structure (PDB entry 2baz) as the search model. The structure of the complex was solved using the native enzyme. For both structures, manual corrections were performed with *Coot* (Emsley & Cowtan, 2004) and refinement was carried out using *REFMAC* (Murshudov *et al.*, 1997). The crystals of the apoenzyme showed pseudo-merohedral twinning and the twinning factor was refined in *REFMAC*. The structure of the complex was refined using TLS refinement. A calcium ion was bound to the dUpNHpp in all six active sites, though at two different positions reflecting the *gauche* or *trans* conformation of the trinucleotide (see below). When the ion was refined as magnesium there was positive difference density at the metal

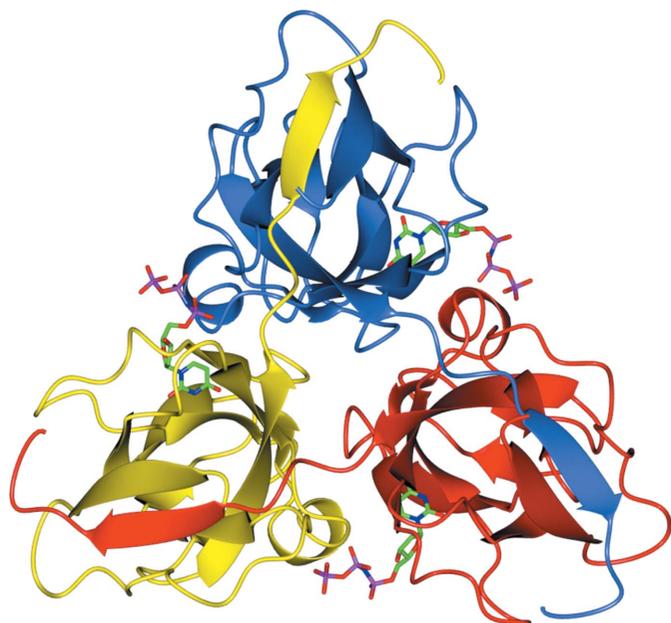
**Figure 2**

The triphosphate chain of dUpNHpp in coordination with a calcium ion. The triphosphate moiety in the *gauche* conformation coordinating a calcium ion is shown as cylinders. The metal is coordinated by the triphosphate and by waters, with distances (shown in Å) that are typical for a calcium ion and too long for a magnesium ion. Figs. 2–5 were produced using *CCP4MG* (Potterton *et al.*, 2004).

sites; the difference peaks disappeared when the ion was refined as calcium. Furthermore, the coordination distances (2.2–2.5 Å) were found to be more typical of calcium than of magnesium (Fig. 2), with the typical ranges of distances being 2.3–2.5 Å for calcium and 2.0–2.15 Å for magnesium (Harding, 1999). The preferred coordination number for both atoms is six. This result is in keeping with the presence of 200 mM calcium in the crystallization drop.

There was a significant positive difference peak ( $17\sigma$ ) on the threefold axes of both trimers. A Tris buffer molecule was reported at this position in other dUTPases, but attempts to model a Tris molecule in YncF did not provide a chemically satisfying depiction of the density. A calcium ion provided a good model for the central peak, leaving some weaker positive difference peaks around the metal. The parameters of the refined models are presented in Table 1.

Only residues 2–130 of each chain were included in the final model; there was no observable density for the C-terminal 14 residues in either structure. Superposition of the six independent protomers in the asymmetric unit of the apoenzyme using *SSM* (Krissinel & Henrick, 2004) demonstrated that they were very similar, with an overall r.m.s.d. in  $C^\alpha$  positions of 0.4 Å. A similar analysis of the complex yielded an overall r.m.s.d. of 0.5 Å. The main differences were found to be in a flexible surface segment between residues 14 and 22, well away from the active site, which appears to have no functional significance. All residues were within favoured regions of the Ramachandran plot, except for the 8.2 and 7.3% of residues that were in additionally allowed regions for the native and complexed enzyme, respectively. Only Ser77 in each chain is in a disallowed region.



**Figure 3**  
Overall fold of the YncF trimer with the ligands shown. The YncF trimer is depicted as a ribbon, with the three subunits coloured blue, yellow and red. Each subunit contributes one  $\beta$ -strand to the adjacent subunit. There are three active sites positioned at the interfaces of the different subunits; the dUpNHpp ligands are depicted as cylinders.

### 3. Results and discussion

#### 3.1. Structure of YncF: the *B. subtilis* genomic dUTPase

The overall fold is similar to those of previously published dUTPase structures and is extremely similar to that of the *B. subtilis* prophage YosS, with an r.m.s.d. of 0.39 Å (PDB entry 2baz). The subunit has a conical shape formed by 12  $\beta$ -strands and a single  $\alpha$ -helix.  $\beta 5$  is split into two sections by the insertion of Ser77, the single Ramachandran outlier, which causes a bulge that is essential for the formation of the active site. This feature is present in other dUTPases, such as Ala75 in the human enzyme (Mol *et al.*, 1996).

YncF is a trimer with a pyramidal shape, with a central channel running along the threefold axis. The trimer shows the characteristic subunit interactions of homotrimeric dUTPases, with extensive binary protein–protein interactions between adjacent subunits formed by the C-terminal strand  $\beta 12$  of each subunit extending into the adjacent subunit and quaternary contacts around the threefold axis. Most of the trimer interface comes from  $\beta 4$ ,  $\beta 5a$ ,  $\beta 6$  or  $\beta 8$ , and equivalent residues from each of the three subunits form symmetrical triads along the channel. The intersubunit interactions are species-specific as the residues in the axial interface are not highly conserved, but the central channel interactions do show some trends: eukaryotic/DNA viral sequences always contain charged or polar residues, while prokaryotic and retroviral sequences mainly contain apolar residues (Fiser & Vértessy, 2000) which may give them additional structural stability (Takács *et al.*, 2004). The central channel of the YncF trimer is essentially identical to that of YosS and otherwise is most similar to that seen in FIV dUTPase (Prasad *et al.*, 1996). It is formed by alternating hydrophobic and hydrophilic residues with ordered water molecules through the entire channel. Some dUTPases, such as those from FIV and *Chlorella* virus, have been shown to contain a metal ion in the threefold-axis channel (Homma & Moriyama, 2009; Prasad *et al.*, 1996), but this metal is trapped inside the channel, in contrast to that in YncF which is positioned on the surface of the trimer.

Within the trimer, three hydrophobic clefts symmetrically positioned in the interface of the subunits correspond to the three active sites, each composed of the five conserved motifs (McGeoch, 1990). Motifs I, II and IV from one subunit interact with motif III of the adjacent subunit and motif V of the third subunit, which lies in the flexible C-terminal arm. In apo YncF the arm is disordered, a feature that has been observed in several other dUTPase structures (Prasad *et al.*, 1996; Larsson *et al.*, 1996; Cedergren-Zeppezauer *et al.*, 1992). This is also true for the complex with dUpNHpp and  $Ca^{2+}$ , which has been proposed to induce the arm to close down over and complete the active site as a prerequisite for hydrolysis (Vértessy *et al.*, 1998).

#### 3.2. The dUpNHpp/ $Ca^{2+}$ complex: the three *gauche* active sites

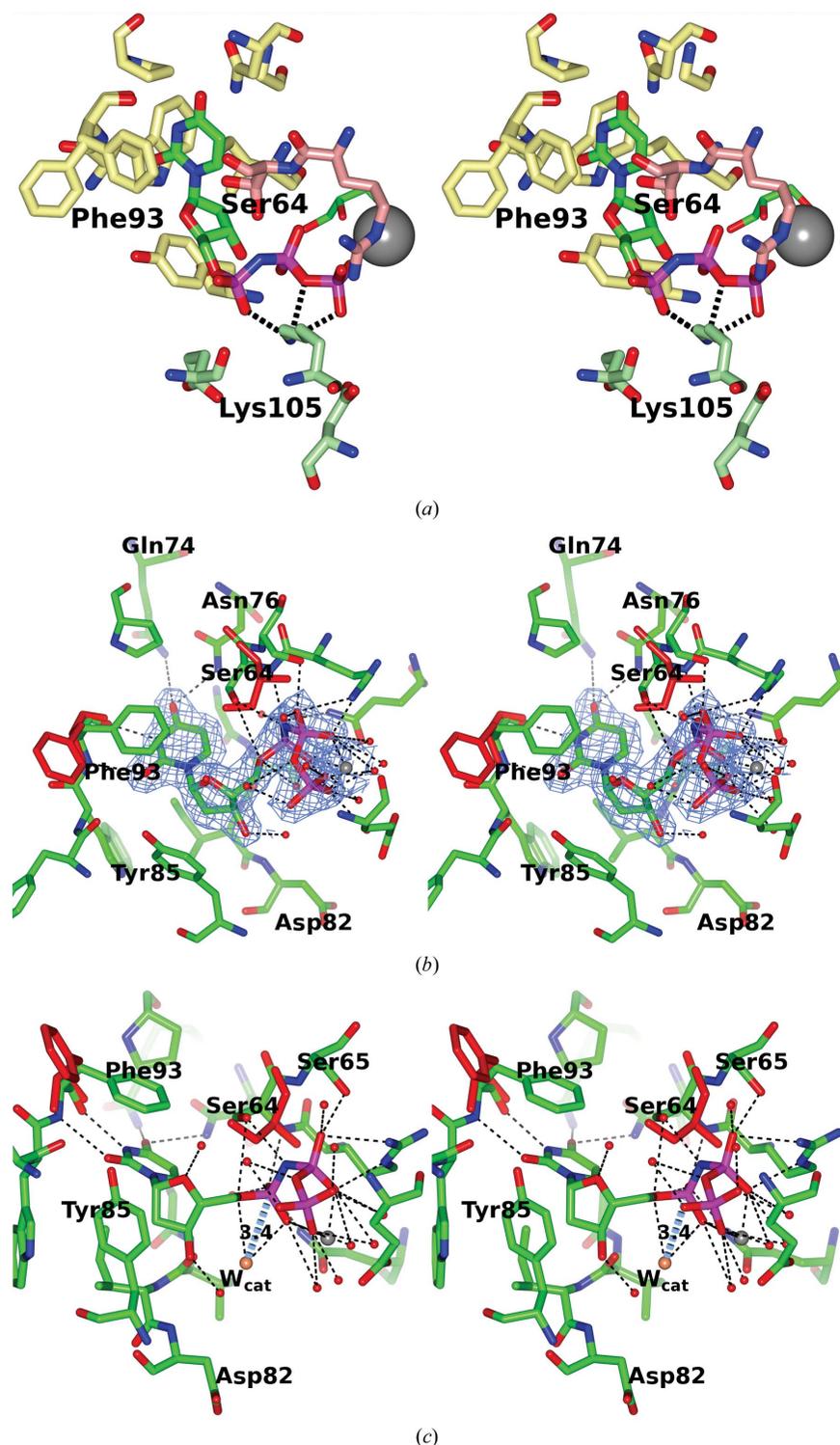
The conformation and position of the protein residues in the complex are the same as in the apoenzyme for both the subunit and for the trimer (Fig. 3), as evidenced by the low

r.m.s.d.s (0.21 and 0.38 Å, respectively). There are only two significant changes. Firstly, the side chain of Phe93 in those active sites in which the polyphosphate chain is in the *gauche* conformation (below) flips from its position in the apoenzyme to stack over the uracil in the complex. The residues around Phe93 retain their apoenzyme conformations. Secondly, the position and orientation of Ser64 vary. The main chain of Ser64 is slightly displaced from the active site upon substrate

binding in order to accommodate the trinucleotide and the side chain reorients to contact the imino group between the  $\alpha$ - and  $\beta$ -phosphates, while in the apoenzyme it points in the opposite direction (Fig. 4).

The conformation of the nucleotide triphosphate chain is not the same in all six protomers. In the three active sites positioned between subunits *A–B*, *A–C* and *F–D* it is in a *trans* conformation, while for those between subunits *B–C*, *D–E* and *E–F* it is *gauche*. Thus, one trimer contains two *trans* ligands and one *gauche* ligand, while the second trimer contains two *gauche* ligands and one *trans* ligand. This variation in conformation has been observed in previously determined trimeric dUTPase structures, as first described for the *M. tuberculosis* enzyme (Chan *et al.*, 2004). The three active sites with the *trans* conformation are assumed to reflect crystal packing, as the ligand forms lattice contacts to the adjacent trimer, including hydrogen bonds from the triphosphate to Lys105 (Fig. 4*a*). In contrast, there are no crystal contacts for the three *gauche* active sites which are able to assume the catalytically active conformation.

In this section, the description and comparison will focus on the three ligands in the *gauche* conformation (Figs. 4*b* and 4*c*); differences between the *trans* and *gauche* ligands will be summarized later. The uracil and the sugar moiety are recognized by the  $\beta$ -hairpin between  $\beta 5$  and  $\beta 6$ . The uracil O2 and N3 atoms hydrogen bond to the Phe93 main-chain N and O atoms, respectively. The O2 atom of the uracil also makes a weak hydrogen bond (3.5 Å) to the



**Figure 4**

Stereo representation of the YncF active sites with dUpNHpp and calcium. (*a*) The *trans* active site. Residues are depicted as light yellow and pink cylinders for the subunits responsible for the biological accommodation of the ligand, while residues from an adjacent subunit forming crystal contacts are depicted as light green cylinders. (*b*, *c*) The *gauche* active site. Residues are represented as green cylinders for the complex and as red cylinders for the apoenzyme. For the latter, only those residues that change position upon ligand binding are displayed. Interactions of the ligand with enzyme residues are shown as dashed lines. (*b*) The  $2F_o - F_c$  map at  $1\sigma$  is shown for the ligand. Phe93 is seen to flip out of the active site to stack against the uracil ring. Ser64 is displaced to accommodate the substrate and changes its direction so as to make contact with the triphosphate moiety. (*c*) Stereoview of the active site from a different angle. The catalytic water positioned at a distance of 3.4 Å to attack the  $\alpha$ -phosphate is shown. Asp82, the proposed general base for catalysis, is seen to point away from the active site.

carbonyl group of Phe93. The aromatic side chain of Phe93 stacks against the uracil. This interaction keeps the substrate bound to the enzyme and helps to exclude water molecules from the active site. The amido group of the Asn76 side chain forms a stabilizing interaction with the ligand *via* the uracil O4 atom. The uracil-binding pocket excludes other nucleotides by steric hindrance. The side-chain O atom of Asn76 is hydrogen bonded to Gln74, exposing its amide group to the incoming base and discriminating against cytosine. The conserved Tyr85 stacks over the deoxyribose and discriminates against ribose, while its hydroxyl group is hydrogen bonded to the Phe91 main-chain carbonyl.

It has previously been shown that the *gauche* conformation is correctly positioned to allow the catalytic water to perform an in-line nucleophilic attack on the  $\alpha$ -phosphate of the dUTP (Kovári *et al.*, 2008; Varga *et al.*, 2007). In the *gauche* conformation in YncF a calcium ion is bound to the O atoms of the trinucleotide, favouring the active conformation, at essentially the same place as the  $Mg^{2+}$  in similar dUTPase trinucleotide complexes from other organisms (Mol *et al.*, 1996; Chan *et al.*, 2004; Barabás *et al.*, 2004).

The triphosphate interacts with several residues in the active site as well as with a number of water molecules. Ser64, Ser65, Arg63 and Gln112 are responsible for binding the phosphates. Ser64 O $\gamma$  forms a direct contact with the imino group of the dUpNHpp. This latter residue is strictly conserved amongst monomeric and trimeric dUTPases (Chan *et al.*, 2004) and has been shown to have two roles in the reaction mechanism whereby it electrostatically stabilizes the transition state and destabilizes the reactant ground state (Palmén *et al.*, 2008). The O1 of the  $\alpha$ -phosphate forms hydrogen bonds to the amine group of Ser64 and to the amino group of Gln112. The  $\beta$ -phosphate forms hydrogen bonds to Ser65 O $\gamma$  and Arg63 N $\gamma$  and NH. Ligand binding is further stabilized by

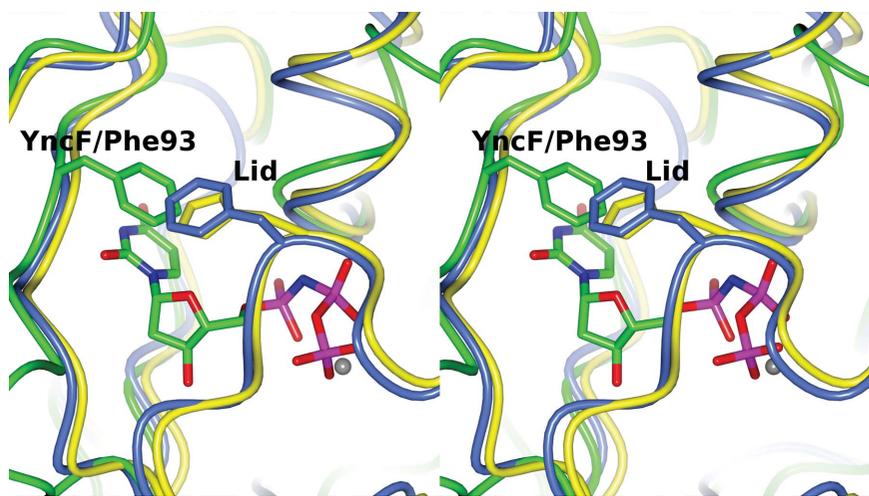
multiple hydrogen bonds to surrounding water molecules and the coordination of the calcium ion to the triphosphate tail. The bivalent metal ion is key to the binding of the triphosphate in the *gauche* geometry (Kovári *et al.*, 2008) *via* tridentate coordination to the three negatively charged O atoms of the triphosphate tail and three water molecules. Thus, calcium can take the place of magnesium and still induce the change to the *gauche* conformation.

The catalytic water can be identified (W233, W506 and W680 in the three *gauche* active sites) lying in an equivalent position to the catalytic water seen in the *M. tuberculosis* enzyme (336 in PDB entry 2py4; Varga *et al.*, 2008) and is in a position to carry out in-line nucleophilic attack on the  $\alpha$ -phosphate (Fig. 4c). This water is hydrogen bonded to the main-chain carbonyl of Val80 and two other water molecules, but not to Asp82, which has previously been proposed to act as a general base in other structures. In the YncF complex Asp82 points away from the active site, implying that it changes its conformation prior to catalysis, assuming that it acts as a general base in this enzyme. In other dUTPase structures the equivalent aspartate points in the opposite direction towards the active site, contacting the catalytic water, with the exception of the monomeric dUTPase from Epstein–Barr virus (Tarbouriech *et al.*, 2005).

### 3.3. The ‘Phe-lid’

In previously determined structures of trimeric dUTPases in which the C-terminus is ordered, an aromatic residue has been shown to stack over the uracil and this was presumed to be necessary for hydrolysis of the  $\alpha$ – $\beta$  phosphate bond. It was seen for the first time in human dUTPase (Mol *et al.*, 1996), where it was termed the ‘Phe-lid’. Subsequent studies have shown that the lid does not need to be a phenylalanine but can also be a histidine as in *M. tuberculosis* dUTPase (Varga *et al.*, 2008). A structural alignment showing the Phe-lid residues in the different species is shown in Fig. 5. The alignment of 19 sequences confirmed the conservation of an aromatic residue at this position (Vertessy & Toth, 2009). We have aligned more than 200 dUTPase sequences (not shown) and confirmed that there is a Phe, Tyr or His at this position in motif V with only one exception, YncF, in which the equivalent residue is Leu138 (Fig. 1).

Phe93 is the residue which shows the most significant conformational change in YncF induced by the binding of dUpNHpp in the *gauche* conformation. The side chain, which lies on the edge of motif III, flips from its position in the apoenzyme (where it points away from the active site) by roughly 150° to stack over the uracil. This completes the uracil-binding pocket in the complex, in which the dUpNHpp is in the *gauche*



**Figure 5**  
Structural comparison of the YncF, *M. tuberculosis* and human dUTPase Phe-lids. The structures were superimposed using *SSM* (Krissinel & Henrick, 2004) and are shown as worms: YncF in green, human in light blue and *M. tuberculosis* in yellow. The ligand of YncF as well as the Phe-lid residues of the three enzymes are represented as cylinders. The Phe-lid residue is Phe158 in the human enzyme and His145 in *M. tuberculosis*. The calcium ion is depicted as a grey sphere.

conformation, and helps to exclude water molecules from the active site. This new feature requires some reconsideration of the catalytic mechanism (discussed below) as until now the Phe-lid was always part of the flexible motif V.

Thus, in YncF Phe93 plays two key functional roles, with the main chain forming most of the interactions with the base and the side chain acting in place of a Phe-lid. It is worthy of note that other dUTPases utilize two different residues to perform these separate functions

### 3.4. Comparison of the *gauche* and *trans* conformations

In addition to the change in the conformation of the triphosphate moiety itself, there are three key differences between the *trans* and *gauche* active sites. Firstly, while the calcium ion is bound to the three phosphates in the *gauche* conformation, it is located in a quite different position and is only bound to one of the O atoms of the  $\gamma$ -phosphate of the *trans* ligand. Secondly, the position of the side chain of Phe93 varies from a totally open form in the apoenzyme to a fully rotated position stacked against the uracil in the *gauche* active sites. It is distributed between the two conformations for the *trans* ligands, with the relative occupancy varying between the three *trans* active sites. Thirdly, Ser64 moves from a position contacting the imino group of the phosphate chain to a position where the hydroxyl group is pointing away from the ligand. In the *trans* conformation, as for Phe93, the Ser64 side chain is distributed between the two conformations, with different occupancies among the three active sites. The occupancies of the Ser64 and Phe93 correlate with one another.

## 4. Conclusions

The structure of YncF, the *B. subtilis* genomic dUTPase, was solved in its apoenzyme form and in complex with the non-hydrolysable substrate analogue dUpNHpp. In the crystal of the complex, the trinucleotide was bound in all six active sites in the two trimers. However, two different sets of conformations of the phosphate moieties were seen: three *trans* and three *gauche*. The *gauche* active sites allowed us to confirm the key residues involved in catalysis as well as the catalytic water. The metal ion places the  $\alpha$ -phosphate in a conformation where the nucleophilic water can perform an in-line attack on the phosphorus, breaking dUTP into dUMP and pyrophosphate, as seen in complexes of the human, FIV and *M. tuberculosis* enzymes. The catalytic water lies in-line with the  $\alpha$ -phosphate at a distance of 3.4 Å, which makes it a good candidate for the catalytic nucleophile. The calcium site in YncF lies very close to the magnesium site seen in these enzymes. The Ser64 side chain changes its position from the apoenzyme to the ligand-bound form in order to make contacts with the ligand.

YncF contains a novel Phe-lid residue (Phe93) which packs against the uracil of the substrate in a distinctly different location in the sequence compared with all other known dUTPases. The *gauche* and *trans* conformations and the position of the calcium ion correlate with the flipping of the Phe-lid over the uracil and the change in position of Ser64. It

appears that all these changes, including the Phe-lid, the ion and Ser64, work together to place the substrate and enzyme in the correct conformation for catalysis. The aromatic side chain of the YncF Phe-lid seems to stack even more directly over the uracil ring than do the side-chain lids from the C-terminal extension of the *M. tuberculosis* and human enzymes, which have a somewhat poorer stacking of the two rings.

### 4.1. Mechanistic implications

A number of structures of trimeric dUTPases with different ligands have been solved and, together with kinetic and mutational data, have shed light on the mechanism of catalysis. The currently accepted mechanism involves at least four steps (Toth *et al.*, 2007).

(i) Fast substrate binding. The active site is open with the C-terminal arm disordered. The trinucleotide diffuses rapidly into the active site and the magnesium ion helps to order the triphosphate moiety in the *gauche* conformation.

(ii) Conversion of the initial complex into the catalytically competent conformation. The C-terminal arm becomes ordered and produces the closed form of the active site. The primary contributor to this is the Phe-lid (from motif V), which stacks over the uracil moiety. In addition, a conserved arginine (Arg135 in YncF), as well as other conserved elements of the C-terminus, become ordered. The importance of the Phe-lid has been shown to be critical in EBV dUTPase as mutation to alanine totally abolished the activity (Freeman *et al.*, 2009). In contrast, mutating the conserved arginine left some residual activity.

(iii) The hydrolysis (chemical) step.

(iv) Product release. The C-terminus opens, followed by rapid release of the products. The interaction of the Phe-lid in motif V has been proposed to drag the uracil of the dUMP product with it as the C-terminal arm moves to the open form and becomes disordered (Vertessy *et al.*, 1998; Mol *et al.*, 1996). The YncF complex structure brings new features into consideration. In the complex crystal, the phosphate moieties that are in the catalytically relevant *gauche* conformation have Ser64 in the correct place for catalysis and the catalytic water is in place for in-line attack on the  $\alpha$ -phosphate. However, in YncF, Phe93, the Phe-lid residue, is not part of motif V and can be seen to stack over the uracil ring while the C-terminal arm is still disordered.

In the *gauche* active sites Phe93 is completely flipped to stack against the uracil, but it is partitioned between the two conformations in the *trans* ligands. Ser64 also has two alternate conformations in the *trans* complexes. The observation of the two alternative conformations suggests that the flipping of the Phe93 over the uracil trinucleotide is probably a consequence of the ligand adopting the active *gauche* conformation. This implies that Phe93 is not involved in the initial binding of the ligand, but rather flips over it coupled with the change from *trans* to *gauche*. Ser64 also appears to contact the phosphate chain more directly in the *gauche* conformation, so that the change in the ligand conformation can bring the Ser64 hydroxyl into contact with the phosphate chain. As the

different positions of Phe93 correlate with the positions of Ser64, the change in Ser64 may be the trigger for the Phe-lid to flip over the uracil ring, leaving the sequence of the events as follows.

(i) Diffusion of the ligand to the active site with metal (in the *gauche* conformation) or without metal (in the *trans* conformation). If the latter occurs, then binding of the metal would subsequently induce the *gauche* conformation.

(ii) Flipping of the aromatic side chain of Phe93 to stack over the uracil ring of the ligand.

(iii) Displacement and reorientation of Ser64 to make contact with the phosphate chain.

Events (ii) and (iii) may well happen in the reverse order or indeed be coupled together. Confirmation of this sequence of events will require further experiments.

Nevertheless, we propose that the binding of the *gauche* ligand with the metal ion together with the stacking of Phe93 and positioning of Ser64 appears to be sufficient to place the key elements, including the hydrolytic water, in the appropriate positions for catalysis.

A general problem for dUTPases, which is shared with other DNA-repair enzymes, is product inhibition: the dUMP product is itself a potential inhibitor, with the level of inhibition varying amongst different species. For example, the *Plasmodium falciparum* trimeric dUTPase showed  $K_d$  values of  $77 \pm 12$  and  $1.87 \mu\text{M}$  for dUMP and dUTP, respectively (Quesada-Soriano *et al.*, 2007, 2008), while the human orthologue showed  $K_d$  values of  $366 \pm 72$  and  $1 \mu\text{M}$  for the same compounds, respectively (Quesada-Soriano *et al.*, 2008). In previous studies it has been proposed that removal of the dUMP product is achieved by the opening of the C-terminal arm with the Phe-lid, the latter carrying the product with it as it is stacked against the uracil base (Mol *et al.*, 1996). The situation in YncF is different as the Phe-lid is not in the C-terminal arm and so disordering of the arm is no longer directly linked to the movement of the lid.

In summary, the structure of YncF provides new insight into the dUTPase mechanism since it separates the functions of the Phe-lid and the C-terminal arm and this information may be relevant to other dUTPases. A remaining question in YncF is why the C-terminal arm does not become ordered in the dUpNHpp/Ca<sup>2+</sup> complex, in contrast to what is observed in *M. tuberculosis* dUTPase. It would seem, at least for this *B. subtilis* dUTPase, that there is a second step that is required to achieve catalytic competence that involves the ordering of the arm. This would involve additional features including the glycine-rich loop. In addition, in the YncF complex Asp82 points away from the active site, where it is generally assumed to act as a general base activating the catalytic water. Some reordering upon closure of the C-terminal arm is almost certain to occur, as the proposed general base points away from the active site in the present *gauche* complexes. A second possibility is that the calcium does not completely fulfil the role of the magnesium ion and fails to induce ordering of the arm. Resolution of this question will require a structure with magnesium rather than calcium bound. Enzymatic activity data in the presence of Ca<sup>2+</sup> versus Mg<sup>2+</sup> would provide insight

into the metal-ion preference of this enzyme. Further biochemical and structural data are needed in order to reach to a better understanding of this already well characterized essential enzyme.

We thank the European Commission for funding through the SPINE2-COMPLEXES project LSHG-CT-2006-031220. We thank the staff at the ESRF (beamline ID14-2) and Diamond Light Source (beamline I04) for provision of synchrotron facilities. We thank Rob Byrne and Callum Smits for data collection at Diamond.

## References

- Baldo, A. M. & McClure, M. A. (1999). *J. Virol.* **73**, 7710–7721.
- Barabás, O., Pongrácz, V., Kovári, J., Wilmanns, M. & Vértessy, B. G. (2004). *J. Biol. Chem.* **279**, 42907–42915.
- Bertani, L. E., Haeggmark, A. & Reichard, P. (1963). *J. Biol. Chem.* **238**, 3407–3413.
- Cedergren-Zeppezauer, E. S., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. S. (1992). *Nature (London)*, **355**, 740–743.
- Chan, S. *et al.* (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.* **D54**, 735–749.
- el-Hajj, H. H., Zhang, H. & Weiss, B. (1988). *J. Bacteriol.* **170**, 1069–1075.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Fiser, A. & Vértessy, B. G. (2000). *Biochem. Biophys. Res. Commun.* **279**, 534–542.
- Freeman, L., Buisson, M., Tarbouriech, N., Van der Heyden, A., Labbé, P. & Burmeister, W. P. (2009). *J. Biol. Chem.* **284**, 25280–25289.
- Gadsden, M. H., McIntosh, E. M., Game, J. C., Wilson, P. J. & Haynes, R. H. (1993). *EMBO J.* **12**, 4425–4431.
- González, A., Larsson, G., Persson, R. & Cedergren-Zeppezauer, E. (2001). *Acta Cryst.* **D57**, 767–774.
- Harding, M. M. (1999). *Acta Cryst.* **D55**, 1432–1443.
- Homma, K. & Moriyama, H. (2009). *Acta Cryst.* **F65**, 1030–1034.
- Kovári, J., Barabás, O., Varga, B., Békési, A., Tölgyesi, F., Fidy, J., Nagy, J. & Vértessy, B. G. (2008). *Proteins*, **71**, 308–319.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.
- Larsson, G., Svensson, L. A. & Nyman, P. O. (1996). *Nature Struct. Biol.* **3**, 532–538.
- Li, G.-L., Wang, J., Li, L.-F. & Su, X.-D. (2009). *Acta Cryst.* **F65**, 339–342.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- 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.* **D53**, 240–255.
- Palmén, L. G., Becker, K., Bülow, L. & Kvassman, J. O. (2008). *Biochemistry*, **47**, 7863–7874.
- Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A. & Noble, M. (2004). *Acta Cryst.* **D60**, 2288–2294.
- Prasad, G. S., Stura, E. A., Elder, J. H. & Stout, C. D. (2000). *Acta Cryst.* **D56**, 1100–1109.
- Prasad, G. S., Stura, E. A., McRee, D. E., Laco, G. S., Hasselkus-Light, C., Elder, J. H. & Stout, C. D. (1996). *Protein Sci.* **5**, 2429–2437.
- Quesada-Soriano, I., Leal, I., Casas-Solvas, J. M., Vargas-Berenguel, A., Barón, C., Ruiz-Pérez, L. M., González-Pacanowska, D. &

- García-Fuentes, L. (2008). *Biochim. Biophys. Acta*, **1784**, 1347–1355.
- Quesada-Soriano, I., Musso-Buendia, J. A., Tellez-Sanz, R., Ruíz-Pérez, L. M., Barón, C., González-Pacanowska, D. & García-Fuentes, L. (2007). *Biochim. Biophys. Acta*, **1774**, 936–945.
- Shlomai, J. & Kornberg, A. (1978). *J. Biol. Chem.* **253**, 3305–3312.
- Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
- Takács, E., Grolmusz, V. K. & Vértessy, B. G. (2004). *FEBS Lett.* **566**, 48–54.
- Tarbouriech, N., Buisson, M., Seigneurin, J. M., Cusack, S. & Burmeister, W. P. (2005). *Structure*, **13**, 1299–1310.
- Tóth, J., Varga, B., Kovács, M., Málnási-Csizmadia, A. & Vértessy, B. G. (2007). *J. Biol. Chem.* **282**, 33572–33582.
- Varga, B., Barabás, O., Kovári, J., Tóth, J., Hunyadi-Gulyás, E., Klement, E., Medzihradzky, K. F., Tölgyesi, F., Fidy, J. & Vértessy, B. G. (2007). *FEBS Lett.* **581**, 4783–4788.
- Varga, B., Barabás, O., Takács, E., Nagy, N., Nagy, P. & Vértessy, B. G. (2008). *Biochem. Biophys. Res. Commun.* **373**, 8–13.
- Vértessy, B. G., Larsson, G., Persson, T., Bergman, A. C., Persson, R. & Nyman, P. O. (1998). *FEBS Lett.* **421**, 83–88.
- Vértessy, B. G. & Tóth, J. (2009). *Acc. Chem. Res.* **42**, 97–106.