Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Tying down the arm in *Bacillus* dUTPase: structure and mechanism

Javier García-Nafría,‡ Jennifer Timm, Charlotte Harrison,§ Johan P. Turkenburg and Keith S. Wilson*

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

‡ Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England.
§ Present address: Department of Biochemistry, University of Bath, Bath BA2 7AY, England.

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

Received 17 December 2012
Accepted 18 March 2013

PDB References: YncF–dU–PP_i–Mg^{2+} complex, space group $P1$, 4apz; space group $H32$, 4aoo; space group $P2_12_12_1$, 4aoz; 4b0h; YosS–dUMP complex, 4ao5

Homotrimeric dUTPases contain three active sites, each formed by five conserved sequence motifs originating from all three subunits. The essential fifth motif lies in a flexible C-terminal arm which becomes ordered during catalysis and is disordered in most crystal structures. Previously, it has been shown that the two *Bacillus subtilis* dUTPases, YncF and YosS, differ from their orthologues in the position in the sequence of the essential Phe-lid residue, which stacks against the uracil base, and in the conformation of the general base aspartate, which points away from the active site. Here, three structures of the complex of YncF with dU–PP_i–Mg^{2+} and the structure of YosS complexed with dUMP are reported. dU–PP_i–Mg^{2+} triggers the ordering of both the C-terminal arm and a loop (residues 18–26) which is uniquely disordered in the *Bacillus* dUTPases. The dUMP complex suggests two stages in substrate release. Limited proteolysis experiments allowed those complexes in which C-terminal cleavage is hindered and those in which it can be assumed to be ordered to be identified. The results lead to the suggestion that dUpNHpp is not a perfect substrate mimic, at least for the *B. subtilis* enzymes, and provide new insights into the mechanism of these two dUTPases in comparison to their orthologues. The enzyme mechanism is reviewed using the present and previous crystal structures as snapshots along the reaction coordinate.

1. Introduction

Living organisms have their genetic information encoded by a four-letter code in the DNA and it is of the utmost priority to maintain its integrity at all stages of the life cycle. DNA integrity is threatened by harmful compounds such as reactive oxygen species (ROS) or by the incorporation of modified nucleotides (Slupphaug *et al.*, 2003), which can lead to point mutations or even major defects in DNA. Special attention must be paid to the so-called pyrimidine problem (Pearl & Savva, 1996), in which uracil and thymine are not distinguished by DNA polymerase, with the dTTP:dUTP concentration ratio being the sole source of discrimination *in vivo*. Massive incorporation of dUTP into DNA leads to saturation of the repair system and fragmentation of the DNA (thymineless cell death; Ephrati-Elizur *et al.*, 1974). In order to avoid this, dUTPases (EC 3.6.1.23) act as a prevention mechanism, hydrolysing dUTP to dUMP and PP_i in a metal-dependent manner and thereby increasing the dTTP:dUTP ratio. dUTPases have been shown to be essential in *Escherichia coli* (el-Hajj *et al.*, 1988), yeast (Gadsden *et al.*, 1993), mycobacteria (Pécsi *et al.*, 2012) and the early development stages of *Drosophila* (Muha *et al.*, 2012), and are present in all

kingdoms of life (Baldo & McClure, 1999). They are classified into three families based on oligomeric state: homotrimers, homodimers (dUTPase/dUDPase) and monomers. While the homotrimeric and monomeric dUTPases share a common fold, with a gene duplication giving rise to the monomeric form, the dimeric enzymes show no sequence or structural similarity to the other two families and instead are members of the all-α nucleotide pyrophosphatase superfamily (Moroz *et al.*, 2004).

The first crystal structure of a dUTPase was that of the homotrimeric *E. coli* enzyme (Cedergren-Zeppezauer *et al.*, 1992), and several structures of orthologues from a number of organisms [including feline immunodeficiency virus (FIV; Prasad *et al.*, 1996), equine infectious anaemia virus (EIAV; Dauter *et al.*, 1999), *Mycobacterium tuberculosis* (Chan *et al.*, 2004), human (Mol *et al.*, 1996) and *Plasmodium falciparum* (Whittingham *et al.*, 2005)] and a variety of ligand complexes (with dUMP, dUDP, dUTP and dUpNHpp–Mg^{2+}) have subsequently been reported. These structures, when combined with kinetic data, have provided insights into the active-site geometry as well as the catalytic mechanism (Tóth *et al.*, 2007). In contrast to the dimeric enzymes, the trimeric enzymes can only hydrolyse dUTP and not dUDP, which acts as an inhibitor (Larsson *et al.*, 1996).

The trimeric dUTPases have *C*3 symmetry and a central channel along the threefold axis with extensive inter-subunit interactions. There are three active sites per trimer that are located at the interface of the subunits and formed by five conserved sequence motifs (McGeoch, 1990), with all three subunits contributing to the formation of each active site (Fig. 1*a*). Four of the motifs are contributed by two adjacent subunits, while motif V, which is located in a flexible C-terminal arm, originates from the third subunit and caps the active site, alternating between closed and open states along the reaction cycle. This C-terminal arm is disordered in most of the published structures, only adopting an ordered conformation in *M. tuberculosis* dUTPase in complex with Mg^{2+}–dUpNHpp (Varga *et al.*, 2008), FIV dUTPase in complex with dUDP (Prasad *et al.*, 2000), the human dUTPase complexes with dUMP, dUDP and dUTP (Varga *et al.*, 2007) and a recent structure of the yeast dUTPase–dUMP complex (Tchigvintsev *et al.*, 2011). Motif V is the least understood motif and has been the subject of several recent studies (Pécsi *et al.*, 2010; Freeman *et al.*, 2009).

Bacillus subtilis is a model organism for Gram-positive bacteria and plays a role in industrial processes through its ability to produce and secrete large amounts of proteins such as amylases and proteases (Harwood, 1992). It contains two distinct trimeric dUTPases: a genomic dUTPase (*yncF*-encoded) and a dUTPase (*yosS*-encoded) from the SPβ prophage (Ghim *et al.*, 1998), which share 93% sequence identity (Fig. 1*b*). Sequence analyses show that one of the essential residues in motif V, the so-called Phe-lid which stacks over the uracil once the ligand is bound, is missing in YncF. The Phe-lid residue in orthologous dUTPases has been shown to have an effect not only on ligand binding but also in the catalytic step (Pécsi *et al.*, 2010). We have previously reported structures showing two unique features of the *B. subtilis* dUTPases: the position in the sequence of the Phe-lid and the conformation of the aspartate general base. A residue with an equivalent lid function is located in a different position in both *B. subtilis* dUTPases and its main chain also forms hydrogen bonds to the uracil of the ligand (García-Nafría *et al.*, 2010, 2011). The conformation of the catalytic aspartate was unexpectedly in an 'inactive' conformation pointing away from the active site in the free and the substrate-bound structures,

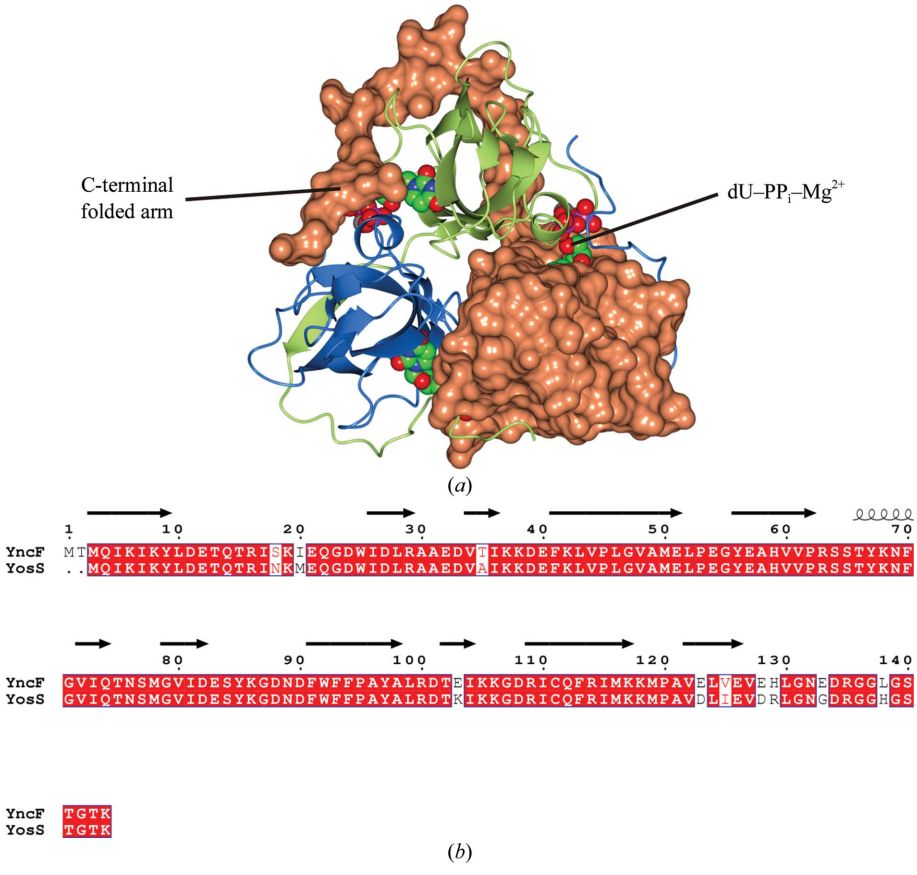


Figure 1
(*a*) The overall fold of YncF. The protein is coloured by chain and represented as ribbons. One of the chains is represented as a surface, facilitating visualization of the contribution of each subunit to the fold in the adjacent subunit and the opposite active site. Ligands are shown as spheres positioned at the active sites. (*b*) Sequence alignment of the *B. subtilis* dUTPases. Conserved residues are highlighted in red. All molecular figures were created using *CCP4mg* (McNicholas *et al.*, 2011).

in marked contrast to its orientation in the structures of homologues in both apoenzyme and ligand-bound forms, in which it points towards the ligand.

Here, we describe the structures of YncF in a pseudo-transition-state mimic dU–PP_i–Mg^{2+} complex (which is novel for dUTPases) and of the YosS–dUMP product complex. In addition, we have carried out limited proteolysis experiments, which together with the structures revealed new features of the *B. subtilis* dUTPases, providing fresh insights into their mechanism.

2. Materials and methods

2.1. Cloning, expression and purification

A pET-26b(+) (Novagen) plasmid containing the DNA for YncF was transformed into *E. coli* B834 cells, which were grown in LB medium (2 l). The cells were harvested by centrifugation and lysed by sonication, and the supernatant was recovered by centrifugation. The sample was then loaded onto a QFF column (anion-exchange chromatography) in buffer *A* (20 m*M* sodium phosphate pH 7.4, 20 m*M* NaCl) and eluted with a linear gradient of ionic strength to buffer *B* (20 m*M* sodium phosphate pH 7.4, 1 *M* NaCl). The resulting peak was concentrated and loaded onto an S200 size-exclusion chromatography column in buffer *C* (50 m*M* Tris pH 8, 50 m*M* NaCl). The peak corresponding to pure YncF was concentrated and the protein aliquots were cooled to 193 K in 50 m*M* Tris pH 7.5, 150 m*M* NaCl at 32 mg ml^{-1}. *B. subtilis* YosS was cloned, expressed and purified using auto-induction medium as described previously (García-Nafría *et al.*, 2011). The recombinant protein was overexpressed in *E. coli* and purified using nickel-affinity and gel-filtration chromatography. The resulting protein was dialysed against 10 m*M* MOPS pH 7 containing 1 m*M* β-mercaptoethanol.

2.2. Crystal structure determination

Crystallization screening was performed using full-length YncF (residues 2–144; the mature enzyme lacking the N-terminal methionine) and YosS (residues 1–142). The commercial screens PACT (Molecular Dimensions; Newman *et al.*, 2005), Index, Crystal Screen and Crystal Screen 2 (Hampton Research) were used to set up 96-well plates with 150 + 150 nl drops in MRC plates using Hydra 96 (Robbins Scientific) and Mosquito (TTP LabTech) robots.

2.2.1. Experiments designed to crystallize transition-state mimics. Attempts to cocrystallize metallofluorine transition-state mimics (dU–MgF_3–PP_i–Mg^{2+} or dU–AlF_3–PP_i–Mg^{2+}) were set up with protein at 20 mg ml^{-1} (YncF) or 14 mg ml^{-1} (YosS) and final concentrations of 10 m*M* dU, 10 m*M* $MgCl_2$, 10 m*M* PP_i and 50 m*M* NaF either with or without 10 m*M* $AlCl_3$, and were incubated for 30 min at 277 K before setting up the trays. Good crystals were only obtained for YncF in three crystal forms. The initial form I (dU–AlF_3–PP_i–Mg^{2+}-derived) crystals were obtained under several conditions. Optimization performed in 96-well trays using 300 + 300 µl drops and 60 µl well solution led to larger crystals after five weeks, with the best being obtained using 30% MPD, 0.1 *M* sodium acetate pH 4.6, 0.02 *M* $CaCl_2$. Form II (dU–MgF_3–PP_i–Mg^{2+}-derived) crystals grew directly from the 96-well plates, with the best crystals growing from 30% MPD, 0.1 *M* sodium cacodylate pH 6.5, 0.2 *M* magnesium acetate. Forms IIIa and IIIb were obtained with different ligand concentrations to avoid the chelation of metals by PP_i as a possible cause of the lack of metal fluoride binding in the active site. Form IIIa (dU–MgF_3–PP_i–Mg^{2+}-derived) crystals were grown from protein complexed with 5 m*M* dU, 25 m*M* $MgCl_2$, 25 m*M* NaF, 5 m*M* PP_i in 150 nl screens (PACT and Index) and data were collected from a crystal obtained using condition A7 of the PACT screen (0.2 *M* NaCl, 0.1 *M* sodium acetate pH 5, 20% PEG 6K). Form IIIb (dU–AlF_3–PP_i–Mg^{2+}-derived) crystals were obtained from protein complexed with 5 m*M* dU, 3 m*M* $MgCl_2$, 10 m*M* $AlCl_3$, 50 m*M* NaF, 5 m*M* PP_i in 150 nl screens (PACT and Index) and data were collected from a crystal grown in condition A9 of the PACT screen (0.2 *M* LiCl, 0.1 *M* sodium acetate pH 5, 20% PEG 6K). The form IIIa and IIIb crystals were mounted without cryoprotectants.

2.2.2. The YosS–dUMP complex. YosS was incubated at 14 mg ml^{-1} in 50 m*M* Tris–HCl pH 8, 150 m*M* NaCl with 10 m*M* dUMP for 30 min at 277 K before setting up the screens. Diffraction-quality crystals were obtained from optimizations performed using 14 mg ml^{-1} protein in 24-well plates with 1 + 1 µl drops and wells containing 500 µl mother liquor. The best diffracting crystals were obtained using 14% PEG 3350, 0.1 *M* sodium malonate, 0.1 *M* bis-tris propane pH 7.5. Data were collected from a number of crystals without cryoprotectant. Those with the least evidence of ice rings were used for structure solution.

2.3. Data collection, structure solution and refinement

Programs from the *CCP*4 suite (Winn *et al.*, 2011) were used unless otherwise stated, with *Coot* (Emsley & Cowtan, 2004) and *REFMAC* (Murshudov *et al.*, 2011) being used for model building and refinement, respectively. The form I (dU–AlF_3–PP_i–Mg^{2+}-derived) crystals were obtained first and data were collected to 2.06 Å resolution, integrated with *XDS* (Kabsch, 2010) and scaled with *SCALA* in the automatic *xia*2 pipeline (Winter, 2010). The space group was *P*1 and the structure was solved by molecular replacement using the apo YncF trimer (PDB entry 2xcd; García-Nafría *et al.*, 2010) in *MOLREP* (Vagin & Teplyakov, 2010), which positioned 12 trimers. Inspection of the resulting maps showed that extra electron density for further trimers in the asymmetric unit was present, but molecular-replacement programs including *AMoRe* (Trapani & Navaza, 2008) and *Phaser* (McCoy *et al.*, 2007) failed to locate them. However, *FFFear* (Cowtan, 2008) positioned four additional trimers, giving a total of 16. Reorganization with *PISA* (Krissinel & Henrick, 2007) yielded an asymmetric unit with 48 protomers arranged in two cubes each made up of eight trimers. The C-terminal arm and the ligands were built in the first chain and then copied to the other protomers using *Coot*. In the active site three components, dU, PP_i and Mg^{2+}, were clearly visible in the electron density, but

there was no density in the position where AlF_3 was expected. The final model, refined using NCS restraints, contained residues 2–144 with a fully ordered C-terminal arm, dU–PP_i–Mg^{2+} and more than 4000 water molecules.

Data for the form II (dU–MgF_3–PP_i–Mg^{2+}-derived) crystal were collected to 2.3 Å resolution, integrated with *MOSFLM* and merged with *SCALA*. Molecular replacement was performed with the trimer (chains *A*, *B* and *C*) of PDB entry 2xcd using *Phaser*.

The data for the form IIIa (dU–MgF_3–PP_i–Mg^{2+}-derived) crystal were integrated with *XDS* and scaled with *SCALA*. Molecular replacement was carried out with *Phaser* using a trimer from form I as a model. Residues 2–131 were modelled in chain *A* and residues 2–144 were modelled in chains *B* and *C*. In chain *C* the loop 18–20 is missing owing to flexibility. *SSM* gives an overall r.m.s.d. of 0.46 Å for the C^{α} atoms of 124 residues for the three protomers, showing the similarity between the chains in the asymmetric unit. Data for the form IIIb crystal were collected to 1.2 Å resolution and were integrated and scaled with *MOSFLM* (Battye *et al.*, 2011) and *SCALA* (Evans, 2006), respectively. The starting model was that of form IIIa and the structure was refined with individual anisotropic temperature factors and riding H atoms.

Data for the YosS–dUMP complex were collected to 1.6 Å resolution and were integrated and scaled using *MOSFLM* and *SCALA*. The structure was solved by positioning the previously refined YosS–dUpNHpp–Mg^{2+} complex (PDB entry 2xy3; García-Nafría *et al.*, 2011) in the cell. Model refinement was performed applying anisotropic temperature factors. The structure of the trimers is consistent, as shown by the low C^{α}-atom r.m.s.d. calculated by *SSM* (0.49 Å for 129 C^{α} atoms) for the protomers in the asymmetric unit (Krissinel & Henrick, 2004). There is clear electron density for residues 2–132 of each protomer. The C-terminal arm composed of residues 133–144 is disordered and residues 18–26 in all six protomers have high *B* factors.

All five structures had good geometry as indicated by the Ramachandran plots, in which only Ser77 (Ser75 in YosS) is an outlier, as was the equivalent serine in previous dUTPase structures. A summary of the data-processing and refinement statistics is shown in Table 1. Coordinates and structure factors have been deposited in the PDB for the three YncF–dU–PP_i–Mg^{2+} complexes (PDB entries 4apz, 4aoo, 4aoz and 4b0h for forms I, II, IIIa and IIIb, respectively) and for the YosS–dUMP complex (PDB entry 4ao5).

2.4. Limited proteolysis

Four 60 µl reactions at 1 mg ml^{-1} were set up with YncF alone, with a 1:5 molar ratio of dUpNHpp:Mg^{2+}, with a 1:5 molar ratio of dU, PP_i, $AlCl_3$ and Mg^{2+} together with an excess of NaF (1:5 with respect to $AlCl_3$) and with a 1:5 molar ratio of dU, PP_i and Mg^{2+}. The reactions were carried out using trypsin at a 1:10 weight ratio with respect to YncF at room temperature. 8 µl samples were taken at 2 and 30 min and 1, 2 and 3 h. A control mixture without trypsin was prepared for each of the complexes and was left at room temperature until the end of the experiment to check for any degradation owing to instability. The proteolysis reaction was stopped by adding SDS and heating the sample at 368 K. Samples were then loaded onto 15% SDS–PAGE and stained with Coomassie Blue for visualization.

3. Results

3.1. Correct definition of the ligand conformations of *B. subtilis* dUTPases

The structure of the Michaelis complex obtained using a nonhydrolysable analogue of dUTP (dUpNHpp) in *M. tuberculosis* (Chan *et al.*, 2004) yielded a structure in which the six crystallographically independent sites bound the ligand in two alternate conformations (termed *gauche* and *trans*) with regard to the C3′—C4′—C5′—O5′ dihedral angle. While in the *gauche* conformation the ligand and metal are in a catalytically competent conformation, in the *trans* conformation the α-phosphate is in the wrong position for catalysis to proceed (Kovári *et al.*, 2008). In our previous structures of *B. subtilis* dUTPase–dUpNHpp complexes the two conformations were reported as being *gauche* and *trans* (García-Nafría *et al.*, 2010, 2011). However, in hindsight we see that the inactive conformation is not *trans* but rather *gauche*(+), but still differs from the active *gauche*(−) form observed in complexes of *B. subtilis* dUTPase and orthologues. This is merely a redefinition of the conformation: the experimental data and coordinates remain unchanged. The *gauche*(+) form is functionally inactive, with the α-phosphate positioned far from the catalytic water molecule and in the wrong orientation for catalysis. Its conformation is dependent on crystal contacts with an adjacent trimer and care must be taken when inferring biological information from *gauche*(+) active sites. Fortunately, the change from the *trans* to the correct *gauche*(−) assignment has no consequences for the functional conclusions of the previous studies.

3.2. The protein fold in the dUTPase complexes

The fold of the *B. subtilis* dUTPases in the crystal structures solved here closely resembles that of orthologues and has been extensively described previously. In summary, it is constituted by 12 β-strands and a single α-helix. The structure has a jelly-roll-like fold with some deviations: (i) β1 and β2 do not directly interact, (ii) β8 originates from the adjacent subunit, (iii) β8 is parallel to β1 and β2 and (iv) an extra α-helix and two β-strands (β3 and β10) are present. While the C-terminal arm and the residues in the 18–26 loop (from now on termed the lower loop) are disordered in the YosS–dUMP complex, as observed previously in the apoenzyme and the dUpNHpp–metal^{2+} complexes, these regions become fully ordered in the YncF–dU–PP_i–Mg^{2+} complexes, which has mechanistic implications.

3.3. Attempts to crystallize a transition-state mimic

dUTPases catalyse the hydrolysis of the phosphoanhydride bond by promoting nucleophilic attack by an activated water

Table 1
Data-collection and processing statistics.

Values in parentheses are for the outer shell.

		YncF–dU–PP_i–Mg^{2+}			
	YosS–dUMP	Form I	Form II	Form IIIa	Form IIIb
Asymmetric unit contents	2 trimers	16 trimers	1 trimer + 1 protomer	1 trimer	1 trimer
Data collection					
Source	Diamond	Diamond	ESRF	Diamond	Diamond
Beamline	I04-1	I24	ID23-2	I04-1	I03
Wavelength (Å)	0.9173	0.9704	0.873	0.9173	0.9763
No. of images	2200	1100	360	1100	1100
Oscillation (°)	0.1	0.2	0.5	0.2	0.2
Space group	$P2_12_12_1$	$P1$	$H32$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters					
a (Å)	99.51	97.06	146.63	70.27	70.03
b (Å)	100.01	97.07	146.63	71.53	71.31
c (Å)	99.82	193.85	144.80	92.30	91.88
α (°)	90	89.72	90	90	90
β (°)	90	88.47	90	90	90
γ (°)	90	90.11	120	90	90
Resolution (Å)	28.8–1.60 (1.69–1.60)	36.44–2.00 (2.06–2.00)	70–2.30 (2.34–2.30)	50.13–2.05 (2.16–2.05)	35.01–1.18 (1.24–1.18)
R_{merge} (%)	8.0 (59.5)	9.2 (39.6)	13.3 (75.2)	9.5 (29.7)	15.4 (133.0)
$\langle I/\sigma(I)\rangle$	7.8 (2.1)	5.9 (2.4)	22.6 (3.2)	16.7 (4.4)	8.1 (2.2)
Completeness (%)	99.5 (99.9)	96.7 (95.9)	100 (100)	99.4 (96.9)	92.8 (100.0)
Multiplicity	5.9 (6.4)	2.2 (2.2)	11.3 (11.0)	7.3 (4.4)	7.3 (7.3)
Refinement					
Resolution (Å)	1.60	2.01	2.30	2.05	1.18
No. of unique reflections	123669	434755	25311	28124	132942
R_{work}/R_{free} (%)	15.03/21.36	16.90/21.18	16.32/22.19	15.90/21.29	11.72/15.94
No. of atoms					
Protein	6415	54927	4568	3426	3602
Ligand/ion	125	1258	109	77	83
Water	1144	4726	225	568	942
B factors (Å^2)					
Protein	23.0	25. 90	31.1	15.24	11.68
Ligand/ion	22.6	17.10	23.1	14.60	9.94
Water	36.9	32.19	29.8	26.58	38.56
R.m.s. deviations					
Bond lengths (Å)	0.010	0.018	0.015	0.015	0.0225
Bond angles (°)	1.44	1.310	1.86	1.61	2.29
Ramachandran plot, residues in (%)					
Most favoured regions	98.34	97.6	96.06	97.92	97.51
Additionally allowed regions	0.83	1.68	3.23	1.30	1.39
Disallowed regions	0.83	0.72	0.72	0.78	1.11
PDB code	4ao5	4apz	4aoo	4aoz	4b0h

molecule on the α-phosphate of dUTP, leading to a pentacoordinated phosphoryl transition state, followed by release of the products dUMP and PP_i (Vértessy & Tóth, 2009). A snapshot of this transition state would shed further light on the catalytic mechanism of these enzymes. Pentacoordinated phosphoryl transition states have been successfully mimicked in several enzymes that hydrolyse the γ-phosphate of a trinucleotide by replacing the latter with AlF_3, MgF_3 or BeF_3 (Carlier *et al.*, 1989; Graham *et al.*, 2002; Baxter *et al.*, 2006; Cliff *et al.*, 2010). We aimed to form a transition-state mimic by replacing the proposed pentacoordinate α-phosphate with AlF_3 or MgF_3 in combination with dU, PP_i and Mg^{2+}, a strategy that has not previously been attempted. Crystals were obtained for YncF under several conditions and data were collected for three crystal forms: form I (space group $P1$, AlF_3-derived, diffracting to 2.00 Å resolution), form II (space group $H32$, MgF_3-derived, 2.30 Å resolution) and form III (space group $P2_12_12_1$; two crystals, MgF_3-derived diffracting to 2.05 Å resolution and AlF_3-derived diffracting to 1.20 Å resolution). However, while they all contained well defined density for dU, PP_i and the Mg^{2+} ion, none of the four structures (in different space groups and grown under different crystallization conditions) showed electron density for the expected metal fluorides at the α-phosphate position. Forms I and II have essentially identical structures; therefore, only the former will be described, while form II will only be used for discussion of the crystal packing. The form III crystals were obtained later: the 2.05 Å resolution data set was first obtained and subsequently the 1.2 Å resolution data. The C-terminal arm is ordered in an essentially identical conformation in the active sites of all three forms, with the exception of one form III site which is affected by crystal packing (discussed below). Hence, details of the active site for the dU–PP_i–Mg^{2+} complex focus on the 1.2 Å resolution structure.

3.4. The YncF–dU–PP_i–Mg^{2+} structure

3.4.1. Form I. The $P1$ AlF_3-derived crystal was obtained first and solution of its structure provided a challenge, with 48 protomers in the asymmetric unit arranged as 16 trimers; all 16 have essentially identical structures and therefore that formed by protomers *A*, *B* and *C* will be described. In all three active sites there was clear difference density for dU, PP_i and magnesium ions, but none at the α-phosphate site where AlF_3 was expected to bind. However, the vacant space corresponds closely to that needed to accommodate the α-phosphate (or AlF_3/MgF_3). The ligands bind in the same conformation in all 48 active sites, in contrast to the variation between the *gauche*(−) and *gauche*(+) conformations observed in the structures of the dUpNHpp–$metal^{2+}$ complexes (García-Nafría *et al.*, 2010, 2011). The C-terminal arm of each subunit is ordered, completing the active site (Fig. 1*a*). The lower loop, which in previous structures was positioned far away from the active site and had high *B* factors, is now well ordered, capping the active site from the opposite side to the arm. The existence of two alternate loop conformations for this region is observed here for the first time in *B. subtilis* dUTPases and appears to be unique since all orthologous homotrimeric dUTPases have an ordered loop.

3.4.2. Form II: cubes in crystals of homotrimeric dUTPases. The AlF_3-derived crystals (form I) contained 48 molecules in the asymmetric unit arranged in two 24-mer cubes (Fig. 2*a*)

with a trimer at every corner (Fig. 2*b*). While the MgF_3-derived crystal in form II was in a different space group (*H*32) with only four molecules in the asymmetric unit, it contains a similar cube (Fig. 2*c*) to those observed in form I, with the only difference being the manner in which the cubes pack against one another. Furthermore, in the 24-mer asymmetric unit (in space group *P*1) of yeast apo dUTPase (Tchigvintsev *et al.*, 2011) we identified a similar cube but again with different packing between adjacent cubes within the crystal lattice. The yeast dUTPase cube does not superpose exactly on those of the *B. subtilis* dUTPases, reflecting some distortion of the intra-cube molecular packing (Fig. 2*d*).

PISA (Krissinel & Henrick, 2007) identified both the 24-mer and the 48-mer as potential stable assemblies, with the *B. subtilis* dUTPase 24-mer having an extensive buried surface of 148 630 Å^2. This raises the question as to whether the 24-mer could be an biological unit *in vivo*. However, on searching the PDB no similar cubes were found in any other dUTPase crystals, from which we conclude that it is probably an artefact of the crystallization.

3.4.3. Form III. The structures of the form IIIa and IIIb crystals are essentially identical. The asymmetric unit contains three protomers that form a biological trimer. Two of the active sites (between chains *A* and *B* and between chains *C* and *A*) contain dU–PP_i–Mg^{2+} positioned in an identical manner to that in forms I and II. However, the third active site (between chains *B* and *C*) contains only a dU molecule. The first two sites have a closed conformation with both loops ordered, a closed Phe-lid and the general base Asp82 pointing into the active site, while the third site is open with both loops disordered, an open Phe-lid and the general base aspartate pointing away from the active site. In the open active site between chains *B* and *C* the Phe-lid side chain is mostly in the closed conformation but with some density for a minor open conformation. The latter is clearly the result of crystal packing, with a loop from an adjacent trimer sterically preventing the ordering of the C-terminal arm. This suggests there is only a small free-energy difference between the open and closed conformations of the arm.

We will use the active site between chains *A* and *B* of the high-resolution form IIIb to describe the active-site

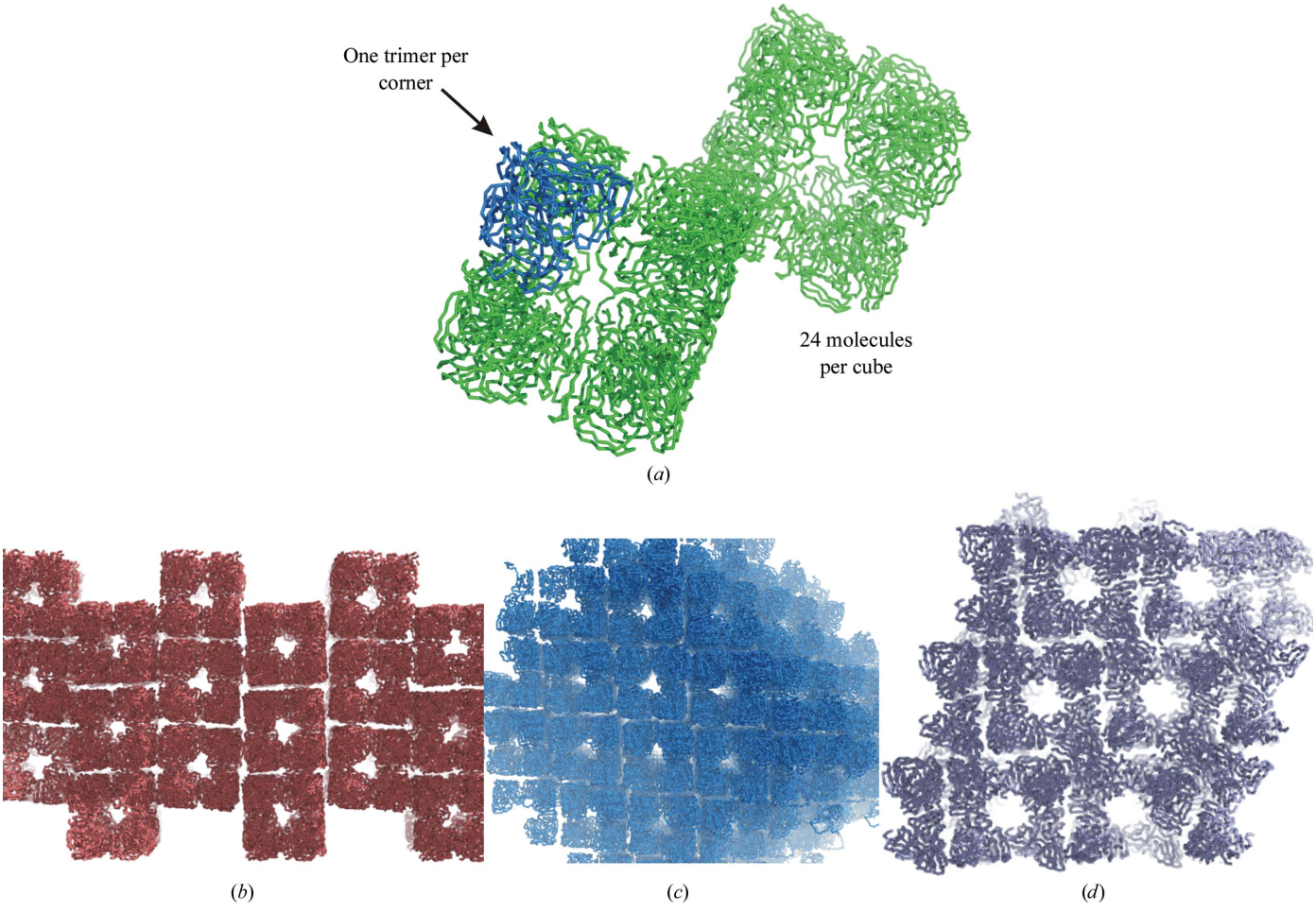
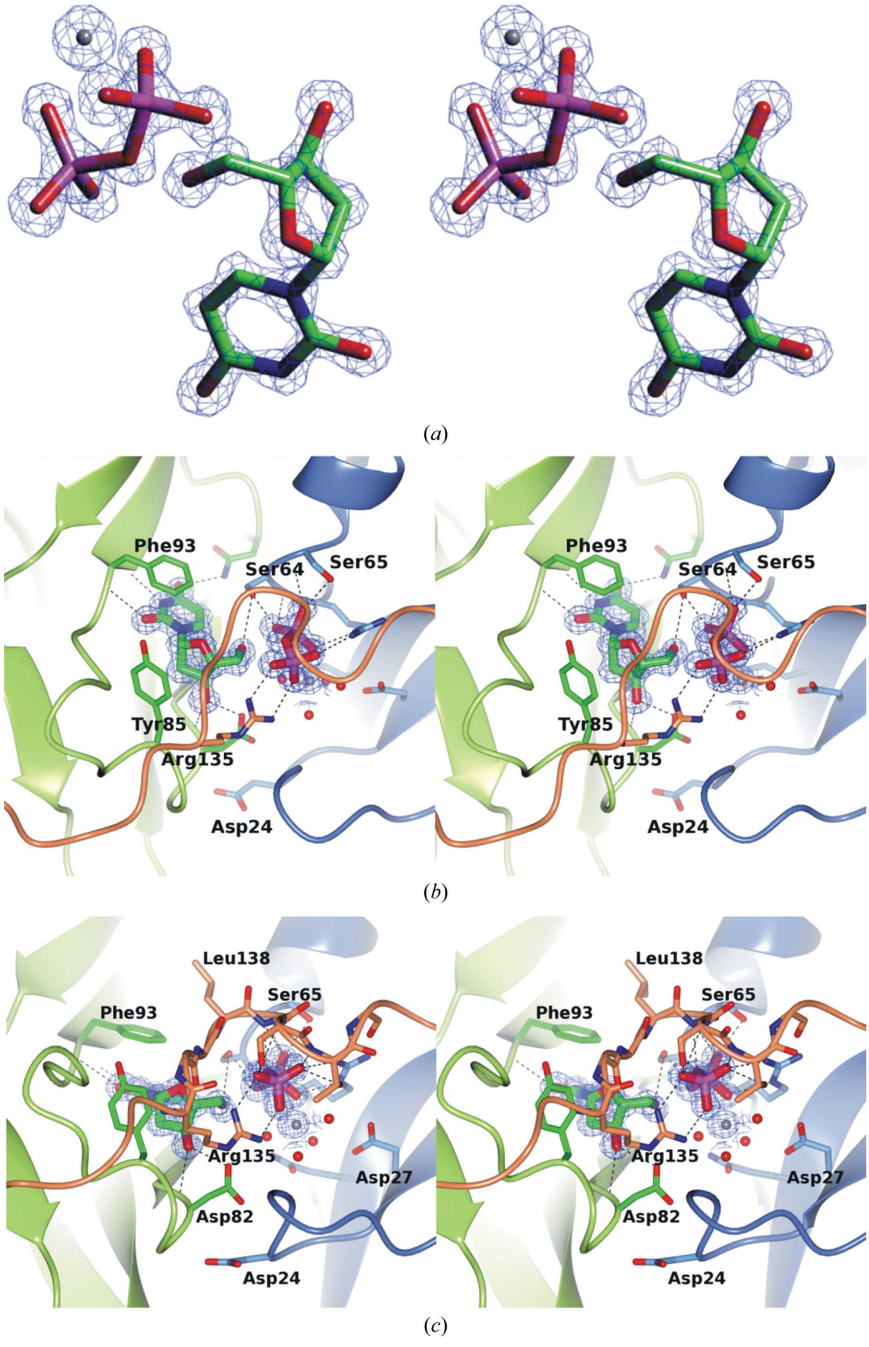


Figure 2
Crystal packing in the YncF–dU–PP_i–Mg^{2+} complexes. (*a*) The asymmetric unit of the form I complex with the protein chains shown as ribbons. The 48 protomers pack into two 24-mer cubes, each composed of eight trimers. (*b*) The crystal lattice of the same complex showing the packing of the cubes. (*c*) The crystal lattice of form II showing similar cubes to those of form I but arranged in a different manner. (*d*) The crystal lattice of yeast apo dUTPase, with similar cubes to those of YncF but with a different packing between cubes from those observed in forms I or II.



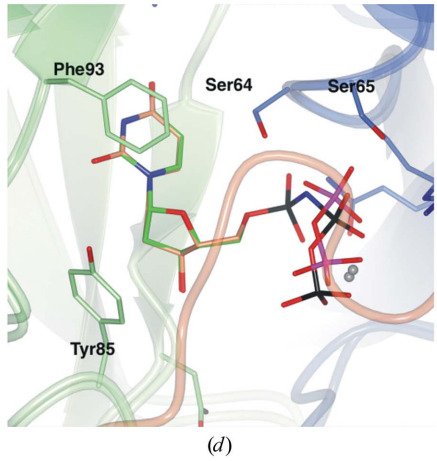


Figure 3
Stereoviews of the closed active site in the form IIIb dU–PP_i–Mg^{2+} complex. (*a*) The electron density in the refined structure contoured at 3σ around the dU, PP_i and Mg^{2+} ion between chains *A* and *B* in the form IIIb crystal. There is very clear density for the dU and PP_i, but none at the position at which the metal fluoride was expected to bind. (*b*) The interactions of the C-terminal arm with the ligands. The protein is in ribbon format coloured by chain, with the active-site residues shown as cylinders. The ligands are shown as thick cylinders, with hydrogen bonds shown as dashed lines. (*c*) The interactions of the adjacent subunits with the ligands. Protein, ligands and hydrogen bonds are shown as in (*b*). (*d*) Superposition of the dUpNHpp–Mg^{2+} and the dU–PP_i–Mg^{2+} ligands in the active site. Neighbouring residues and ligands are shown as cylinders, with green C atoms and magenta phosphates for dU–PP_i and coral C atoms and black phosphates for dUpNHpp.

components in detail. Fig. 3(*a*) shows the electron density around the dU, PP_i and metal between chains *A* and *B* in the form IIIb crystal.

3.5. The active site

Binding of the uracil occurs by pseudo-pair formation (Kondo & Westhof, 2011) with stacking between the uracil ring and the main chain of Phe93, as in the dUpNHpp–$metal^{2+}$ complexes. The deoxyribose interacts with the protein through Ile81 and Tyr85, which discriminate against the ribose form of the nucleotide. The Phe-lid residue (Phe93) is closed, making displaced face-to-face π–π interactions with the uracil ring, as observed previously in the *gauche*(−) dUpNHpp–$metal^{2+}$ active sites. The Phe-lid residue is in a fully closed conformation in forms I and II in all active sites, while in form III the two closed active sites have a minor alternative open conformation. The PP_i occupies a similar but not identical position to the active *gauche*(−) dUpNHpp β- and γ-phosphates. For ease of description the PP_i phosphates will be termed β and γ corresponding to the positions in the triphosphate moiety. Residues from the α-helix interact with the phosphates, with the Ser64 side chain having a single conformation pointing towards where the O atom between the α- and β-phosphates would be expected to lie from its location in the dUpNHpp complex.

The O1 atom of the β-phosphate forms hydrogen bonds to both the main-chain NH and the side-chain OH of Ser64. In a similar manner, the O2 atom of the same phosphate forms hydrogen bonds to the main-chain NH and the side-chain OH of Ser65. Arg63 makes two well defined hydrogen bonds to the O3 atoms of the β- and γ-phosphates. Most importantly, the hydrogen-bonding pattern of the phosphates differs from that of the nonhydrolysable dUpNHpp in the active *gauche*(−) conformation. The magnesium ion is octahedrally coordinated by O3 of the β-phosphate and O5 of the γ-phosphate plus four water molecules (two coordinated by Asp27). One water is positioned where the O atom of the α-phosphate is located in the dUpNHpp–metal^{2+} structure and the other three are in the same position as in the previous YncF and YosS structures.

3.6. The glycine-rich loop of the *B. subtilis* genomic dUTPase

Many crystal structures of trimeric dUTPases have been solved and in most of them the C-terminal arm is disordered. In the dU–PP$_i$–Mg^{2+} complex the arm is ordered, bringing motif V to the active site and capping it. The arm forms two consecutive turns over the ligand and interacts with both ligand and protein. Interactions with the ligand are restricted to the γ-phosphate, apart from the main-chain amide group of Gly139, which contacts O2 of what would be the β-phosphate. Arg135 makes two hydrogen bonds to O1 and O2 of the γ-phosphate. The main-chain amide groups of Gly139, Ser140 and Thr141 form hydrogen bonds to the O atom between the β- and γ-phosphates, O1 and O3 of the γ-phosphate, respectively. In addition, the side chains of Ser140 and Thr141 hydrogen bond to O2 and O3 of the γ-phosphate, respectively. Leu138, the position of which corresponds to the Phe-lid in other dUTPases, points towards the YncF Phe-lid but sits above its plane, making hydrophobic interactions with Leu43 and Phe93 (Fig. 3). In YosS there is an aromatic residue in the conserved C-terminal position, but since the Phe-lid is positioned outside the C-terminal arm it is likely that this residue would be positioned over the Phe-lid similar to that in the YncF complex. Apart from the Phe-lid residue being positioned at a different location, the rest of the interactions formed by the C-terminal arm and the ligand are highly conserved when superposed with other available structures (Varga *et al.*, 2007, 2008).

Residues Arg63 and Arg109 serve as an important clamping point for ordering the arm. The Arg63 guanidinium group forms a hydrogen bond to the Gly142 carbonyl O atom, while the Arg109 side chain forms a hydrogen bond to the Thr141 main-chain carbonyl and side-chain hydroxyl groups. The Arg109 main-chain amine group also binds to the carboxyl group of Thr143.

3.7. The lower 18–26 loop is only seen in *B. subtilis* dUTPases

In all of the subunits in all previous *B. subtilis* structures, residues 18–26 (the lower loop) next to motif I are positioned far from the active site and have high *B* factors with weak and diffuse electron density. In contrast, in the dU–PP$_i$–Mg^{2+} complexes this loop adopts an ordered conformation. Unexpectedly, the loop main chain does not follow the trace observed in the previous structures but moves closer to the active site, capping it together with the C-terminal arm. This ordered conformation resembles that observed in orthologues and it is rather the disordered conformation which differs from other dUTPases. The ordered loop interacts with residues in β2 from its own subunit, forming β-strand contacts. Contacts to the adjacent subunit occur through the Gln22 main-chain carbonyl group, which hydrogen bonds to the Ser84 side chain of the adjacent subunit, while the Asp24 side chain hydrogen bonds to the main-chain amine group of Glu83 (Figs. 3*b* and 3*c*). Contacts to the arm are through the Gln22 side chain, which hydrogen bonds to the amino group of Arg135 and through a water molecule to its main-chain carbonyl. There are additional interactions through water molecules between the Gln22 side chain and the side-chain hydroxyl group of Ser140.

For the first time in a *Bacillus* dUTPase complex, Asp82, the side chain of which is proposed to be the general base of the catalytic reaction, points directly towards the active site. Interestingly, Asp24 from the 18–26 lower loop is now very close to Asp82 in a position that would sterically clash with the inactive conformation of Asp82. It appears that the ordering of the lower loop is needed to position the general base aspartate in the active conformation. However, even though the general base is pointing towards the active site, there is no electron density for the catalytic water.

3.8. The YosS–dUMP structure: two stages of product release

This structure provides a snapshot of the final step in the reaction. The crystal form is isomorphous to the structures of the free YncF enzyme (PDB entry 2xcd; García-Nafría *et al.*, 2010), the complexes of YncF and YosS with dUpNHpp–metal^{2+} (PDB entries 2xce and 2xy3, respectively; García-Nafría *et al.*, 2010, 2011) and the YosS–dUDP complex (PDB entry 2y1t; García-Nafría *et al.*, 2011), with two trimers in the asymmetric unit. The general fold is closely similar, with both the C-terminal arm (residues 133–142) and the lower loop being disordered. Note that YosS contains two residues fewer than YncF at the N-terminus.

3.8.1. The active sites. There is a dUMP molecule in all six active sites, with the contacts made by the uridine and the deoxyribose as in previous *Bacillus* dUTPase complexes. The α-phosphate adopts two alternate conformations similar to the phosphate moieties in the dUpNHpp complexes, and superposition of the dUpNHpp and dUMP structures shows that the three *gauche*(+) and the three *gauche*(−) active sites are in equivalent positions in the two trimers. Therefore, the *gauche*(+)/*gauche*(−) terminology will be retained here. These different conformations of the ligands have an impact on key residues in the active site.

The *gauche*(+) active sites between subunits *A*–*B*, *A*–*C* and *F*–*D* have the α-phosphate in an equivalent position to that in the dUpNHpp *gauche*(+) conformation. However, in contrast to the latter the Phe-lid is here in the closed position and stacks over the uracil. The active site between subunits *A* and

C has a fully closed Phe-lid conformation, while those between subunits *A* and *B* and between subunits *F* and *D* show a split between open (occupancy 0.2) and closed (occupancy 0.8) conformations that is related to a degree of disorder in the α-phosphate. Ser64 has two partial conformations suggesting flexibility. The α-phosphate of the ligand forms a hydrogen bond to a lysine in an adjacent trimer.

The *gauche*(−) active sites between protomers *B*–*C*, *D*–*E* and *E*–*F* have the Phe-lid predominantly in the open form, although in the *E*–*F* site it has alternate conformations modelled as 0.2 for the closed form and 0.8 for the open form, again correlated with a degree of disorder in the α-phosphate. This causes the phosphate and the Phe-lid conformations to be in opposite correlation, as in the triphosphate ligands. Ser64 has alternative conformations, but a new feature appears: Ser63 (equivalent to Ser65 in YncF), which until now had always displayed a single ordered conformation, has two alternate conformations. In the second conformation the hydroxyl chain interacts directly with the α-phosphate of the dUMP. Superposition of the *gauche*(−) dUpNHpp and dUMP ligands revealed a rotation of the dihedral C6—N1—C1′—C2′ angle (Fig. 4*a*), with the N1—C1′ bond between the base and the sugar rotating from an angle of −68 to −22° (from dUpNHpp to dUMP). This displaces the deoxyribose out of the active site by 3.4 Å as measured between the C3′ atoms of the two ligands, with a 5 Å displacement of the α-phosphates between the two *gauche*(−) conformations, with the deoxyribose leaning towards the outside of the protein. Furthermore, in this conformation the sugar ring is in steric conflict with the closed form of the Phe-lid. These two active sites show the importance of the light interaction of the deoxyribose with the protein, with it being able to leave after hydrolysis.

Asp80 has its carboxyl group pointing away from both the *gauche*(+) and *gauche*(−) active sites and in this conformation cannot act as the general base in the catalysed reaction. There is density at the expected position for the catalytic water.

3.9. Limited proteolysis

Limited proteolysis has been widely used to probe the flexibility of segments in a protein (Fontana *et al.*, 2004), as such segments tend to be exposed and subject to cleavage. A number of previous limited proteolysis experiments have been carried out for the dUTPases. In *Drosophila* and EIAV dUTPases limited proteolysis was used to cleave the flexible C-terminal arm and to establish that it is essential for enzyme activity (Nord *et al.*, 2000; Kovári *et al.*, 2004). In addition, it has been shown that the dUMP product protected *Drosophila* dUTPase from proteolysis owing to triggering of the catalytically competent conformation with the C-terminal arm ordered, while an equivalent experiment yielded the opposite result for the *E. coli* enzyme (Kovári *et al.*, 2004). These results indicate that the ordering of the C-terminal arm upon binding of the product varies in a species-specific manner. Since the C-terminal arm is disordered in most dUTPase structures and only adopts an ordered conformation in certain states, we subjected YncF to limited proteolysis using trypsin to probe the flexibility of the arm at different stages of the catalytic cycle, exploiting the complexes seen in the crystal structures. This was aimed in particular at identifying differences between the dU–PP_i–Mg^{2+} and dUpNHpp–metal complexes, with the latter having a disordered arm. YncF alone or in complex with the different ligands (dUpNHpp–Mg^{2+}, dU–AlF_3–PP_i–Mg^{2+}, dU–PP_i–Mg^{2+} without metal fluoride and dUMP) was digested with trypsin while samples were taken at regular intervals, monitoring the speed of proteolysis. For YncF alone some degradation was already visible after 2 min, while after 30 min roughly half of the sample had been digested and after 1 h or longer most of the enzyme was found as a lower molecular-weight band on the SDS–PAGE (Fig. 5*a*). The tryptic digestion was checked using ESI mass spectrometry, showing that the lower band in the SDS gels corresponds exactly to the loss of the C-terminal arm downstream

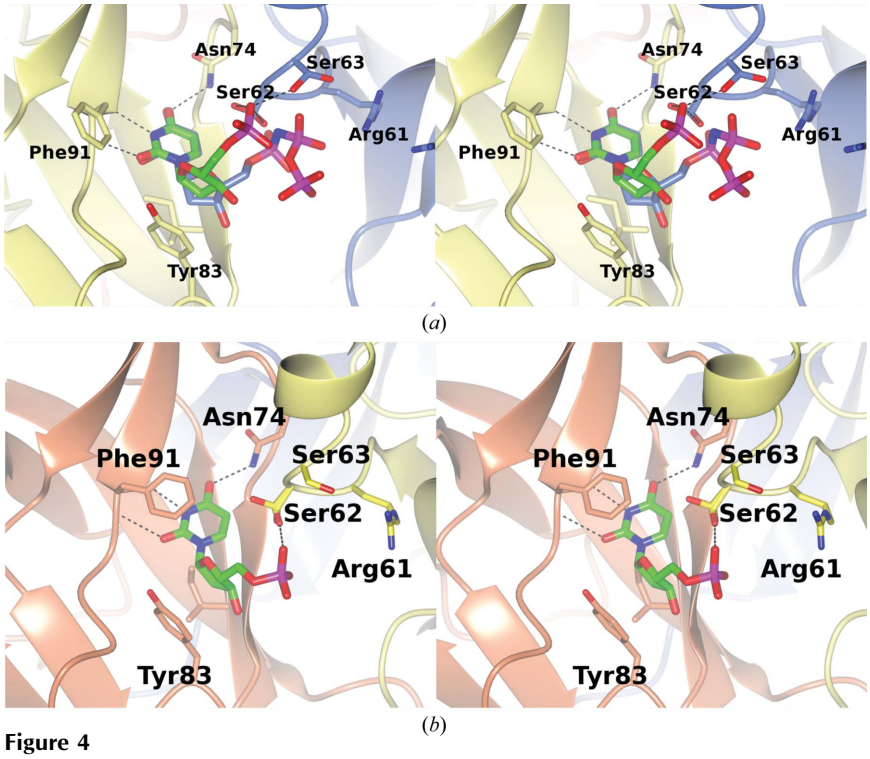


Figure 4
Stereoviews of the active sites in the YosS–dUMP complex. (*a*) The *gauche*(−) active site with the protein ribbons coloured by chain and the active-site residues displayed as cylinders and coloured by atom but with the C atoms coloured by chain. Hydrogen bonds are shown as black dashed lines. The dUMP (green C atoms) is superposed with the dUpNHpp (blue C atoms), showing the displacement of the sugar and the different position of the phosphates. (*b*) Active site with a *gauche*(+) dUMP ligand. The protein, ligand and interactions are represented as in (*a*).

of the conserved Arg135, with an experimental molecular mass of 15.519 kDa, which exactly coincides with the expected value for cleavage at this point. Trypsinolysis in the presence of the product dUMP revealed no protection, with time scales of hydrolysis that were essentially identical to those obtained for the enzyme alone.

In the presence of the substrate analogue (and known inhibitor) dUpNHpp and Mg^{2+}, the rate of the reaction was similar to that for the protein alone, suggesting that the C-terminal arm primarily remains disordered. In contrast, when digestion was performed in the presence of dU–AlF_3–PP_i–Mg^{2+} there was no observable digestion after 2 min and only around 40% of the sample had been digested after 3 h, showing clear protection from proteolysis and suggesting an ordered arm. Interestingly, a similar result was obtained in the absence of AlF_3, suggesting that the presence of the metal fluoride is not required to induce ordering of the C-terminal arm. This raises the question as to what the difference between the dU–PP_i–Mg^{2+} and dUpNHpp–Mg^{2+} complexes is and questions the nonhydrolysable substrate analogue as a truly accurate mimic of the dUTP-binding mode.

4. Discussion

4.1. Significance of the *gauche*(+) and *gauche*(−) ligand conformations in the *B. subtilis* dUTPases

In the *M. tuberculosis* enzyme–substrate complex two conformations of the triphosphate chain of the substrate were identified, termed *gauche* and *trans*, and were shown to be the active and inactive conformations. The structures of *B. subtilis* dUTPases reported previously (YncF and YosS in complex with dUpNHpp–metal^{2+} and YosS in complex with dUMP) were crystallized in isomorphous crystals with two trimers in the asymmetric unit. The six active sites in the asymmetric unit contained two alternative ligand conformations which, although initially termed *gauche* and *trans*, are here more properly defined as *gauche*(−) and *gauche*(+) but still reflect active and inactive conformations, respectively.

This raises the question as to whether the inactive *gauche*(+) conformation has biological relevance or is rather an artefact of crystal packing. The latter is the most likely explanation since there are crystal contacts from the side-chain N atom of a lysine (Lys105 in PDB entry 2xce; García-Nafría *et al.*, 2010) from an adjacent trimer to the α-phosphate in all of the *gauche*(+) active sites in both the dUpNHpp and the dUMP complexes. However, a number of factors raise the possibility that both conformations may have biological relevance. Firstly, the different conformations of the ligands have a significant impact on important active-site residues which have no such crystal contacts. In the active *gauche*(−) active sites of the dUpNHpp–metal^{2+} complex structures of both YosS and YncF the Phe-lid residue flips to stack over the uracil and the residues in the helix move in order to accommodate the ligand. This is not observed in the inactive *gauche*(+) conformation. In addition, the *gauche*(+) or *gauche*(−) nature of the dUMP ligand in the active sites of the YosS complex has a marked effect on the position of the Phe-lid and on the serine residues of the helix. Secondly, there are two alternate conformations in the *M. tuberculosis* dUpNHpp complex structure in the absence of any crystal contacts, which suggests that there are two alternate conformations in solution (PDB entry 1sjn; Chan *et al.*, 2004). Structures in different space groups without the lysine crystal contacts are needed to resolve this issue.

The presence of alternate ligand conformations regularly distributed across the active sites in a series of complexes is a unique feature of the *B. subtilis* dUTPases. In other orthologues the two alternate conformers are present when the metal is present or absent or as multiple conformations in the active sites.

4.2. dU–PP_i–Mg^{2+} *versus* dUpNHpp–Mg^{2+}

The C-terminal arm becomes ordered in the presence of dU–PP_i–Mg^{2+}. The arm makes a key contribution to ligand binding, making seven hydrogen bonds to the γ-phosphate (compared with a total of six from the other conserved motifs)

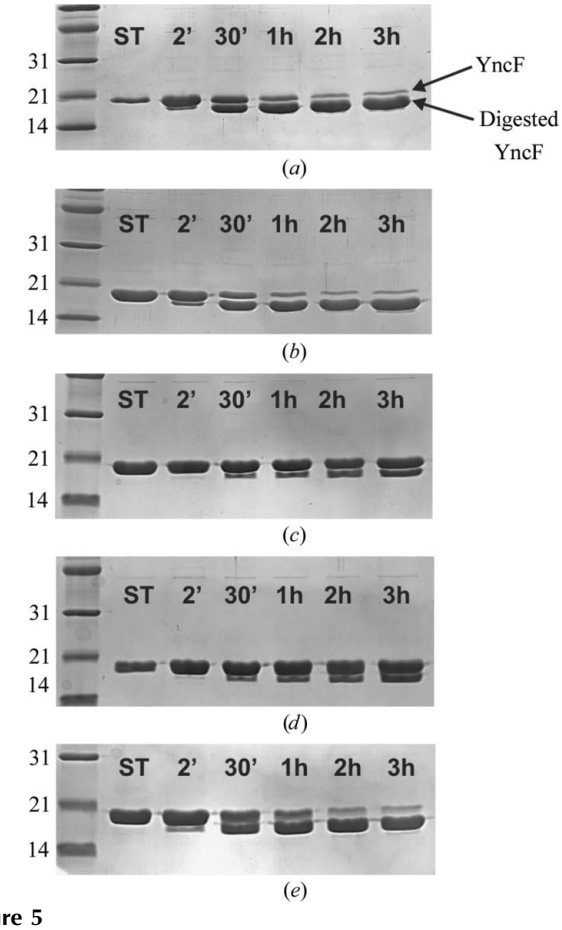


Figure 5
Limited proteolysis (trypsin, 1:10) experiments on the complexes. The SDS–PAGE gels show the results for (*a*) YncF, (*b*) the YncF–dUpNHpp–Mg^{2+} complex, (*c*) the YncF–dU–AlF_3–PP_i–Mg^{2+} complex, (*d*) the YncF–dU–PP_i–Mg^{2+} complex and (*e*) the YncF–dUMP complex. Molecular-mass markers are shown in the leftmost lane with masses in kDa. The second lane contains the standard (ST) for the sample in the absence of protease. Incubation times are indicated on each lane.

and raises several questions. Firstly, how is the dU–PP_i–Mg^{2+} ligand (lacking a metal fluoride to truly mimic the transition state) but not the dUpNHpp–Mg^{2+} ligand able to order the arm? When superposing the dU–PP_i–Mg^{2+} and dUpNHpp–Mg^{2+} structures it is hard to distinguish between mis-superposition of the structure and biological differences. We have addressed this problem by basing the superposition on the uracil ring and looking at the different hydrogen-bond patterns in each complex, whereupon two features appeared that distinguish the two structures. This shows that the position of the β- and γ-phosphates is different, with the pyrophosphate being buried deeper in the structure by ∼1 Å than the dUpNHpp, which affects the hydrogen-bonding pattern. In those dUTPase structures in the PDB which have dUTP as a substrate instead of dUpNHpp [structures of an inactive mutant (Barabás *et al.*, 2004) or using Cr^{2+} instead of Mg^{2+} (Chan *et al.*, 2004)], the phosphates are at a deeper position than in the dUpNHpp complexes and have hydrogen-bonding patterns similar to that of dU–PP_i–Mg^{2+}. This suggests that dUpNHpp is not able to localize in precisely the same place in the active site as the real substrate. Although there are two different sets of hydrogen-bonding patterns they do not correlate with the ordering of the C-terminal arm, since both open and closed C-terminal arms are observed with each of the hydrogen-bonding patterns. In addition, the position of the Gln112 side chain is invariant and if the α-phosphate was present it would be within hydrogen-bonding distance. This qualifies this residue for transition-state stabilization and it has been shown to be important for activity but with a role that has yet to be assigned.

A second question is at what point in time does the C-terminus take up an ordered conformation. Several crystal structures of the Michaelis complex have been obtained for this family and the ordering of the C-terminus did not correlate with the presence or absence of any ligand, but rather seemed to be species-specific. This might be owing to crystal-packing forces which do not let the C-terminal arms order in some crystal lattices. However, in our proteolysis studies the C-terminus remains cleavable and is presumably flexible in solution in the presence of dUpNHpp–Mg^{2+}. This is in marked contrast to a similar experiment with the *E. coli* dUTPase, in which dUpNHpp–Mg^{2+} protected the enzyme against hydrolysis, suggesting that the ordering of the arm occurs in a different manner. These results point towards differential dynamics of the C-terminal arm in a species-specific manner. This is highly significant since recent work has indicated that viral dUTPases are moonlighting enzymes with a signalling function in the cell as well as an enzymatic role (Tormo-Más *et al.*, 2010). The switch between the on and off states of the signalling function are directly correlated with ordering of the C-terminal arm, with this regulation varying between species. We conclude that although the YncF–dU–PP_i–Mg^{2+} complex does not contain a mimic of the pentacoordinated phosphate, it contains all of the elements positioned as in the transition-state stage and has small but significant differences from the YncF–substrate complex.

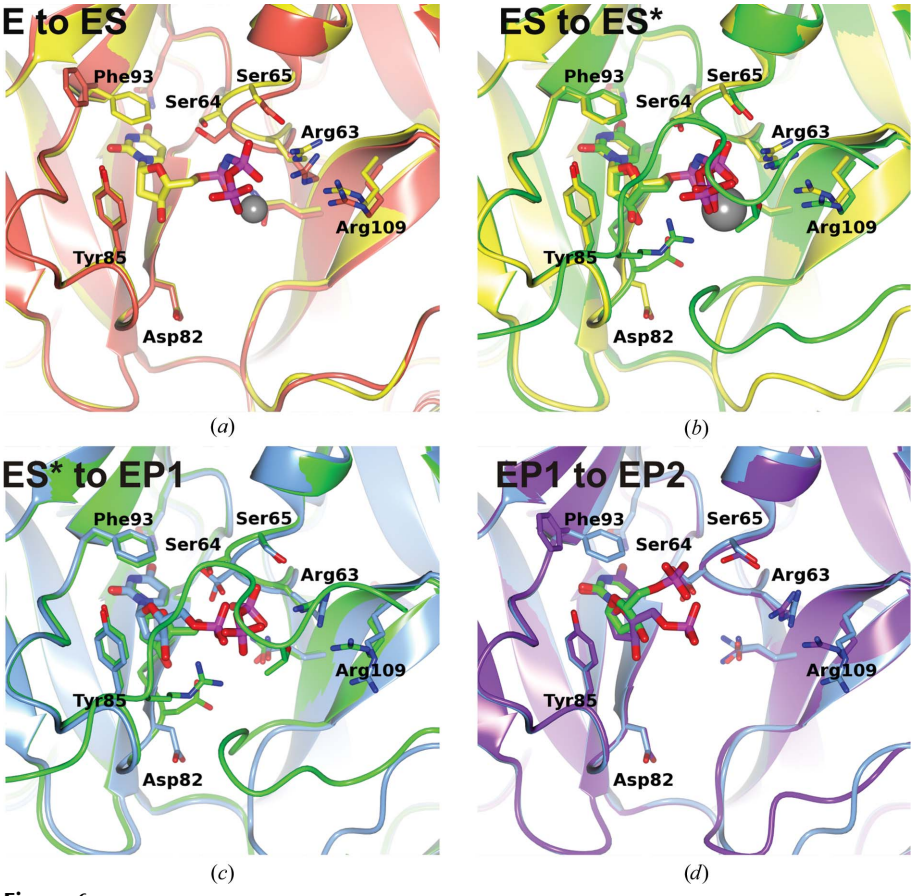


Figure 6
Snapshots at key stages of the *B. subtilis* reaction mechanism. The protein is shown as ribbons, with active-site residues shown as cylinders. The protein is coloured by stage of reaction in each panel, comparing the changes occurring in the protein from one step to the next. (*a*) Superposition of the native (red) and substrate-bound complex (ES; yellow), showing the rearrangement of the Phe-lid and residues in the helix. (*b*) Superposition of the ES complex (yellow) onto the transition-state-like complex (green), showing the ordering of the C-terminal arm, the lower loop and the aspartate general base. (*c*) Superposition of the transition-state-like structure (yellow) onto the first product complex (blue), showing the release of the C-terminal arm, the lower lobe and the 'deactivation' of the aspartate general base. (*d*) Superposition of the first product complex (blue) onto the second product complex (purple), showing that the opening of the Phe-lid together with the initial release of the substrate may be aided by the Ser65 (in YncF) side chain.

4.3. Insights into the *B. subtilis* dUTPase catalytic mechanism

Structures and biochemical data prior to these studies led to a proposed mechanism in which there is (i) fast

diffusion of the substrate into the active site, (ii) a change in the enzyme conformation in the catalytically active form, (iii) the chemical step and (iv) product release. The set of structures that we have determined for the *B. subtilis* dUTPases will now be discussed in this context, providing new insights into the catalytic mechanism as well as revealing features of the *B. subtilis* dUTPases in which they differ from all of their orthologues. For simplicity, the YncF numbering is used throughout. The conformational changes along the key snapshots of the reaction cycle are summarized in Fig. 6.

4.3.1. Free enzyme and diffusion of the ligand into the active site. In apo YncF the C-terminus is disordered and the lower loop is very poorly ordered. The Phe93 side chain (Phe-lid) is in the open conformation and Ser64 is disordered or points away towards Asn76, while the general base Asp82 points away from the active site. The substrate in the *gauche*(−) or *gauche*(+) conformation together with the metal ion can diffuse into the active site.

4.3.2. The ES complex, isomerization to the catalytically active conformation and pathway to the ES* complex. Once the ligand is bound there must be changes in the structure of the enzyme to convert it to the active form. These changes account for completion of the active site by the C-terminal arm carrying motif V and, in *B. subtilis* dUTPases, a change in the conformation of the Phe-lid and the general base Asp82 and the ordering of the 18–26 loop. In YncF and YosS we then propose two substeps. In the first substep the Phe-lid changes from the open to the closed conformation, together with an induced fit of the residues in the helix (Arg62, Ser63 and Ser64) and Arg109 when the substrate is present in the *gauche*(−) conformation. This helix tilts so that the N-terminal microdipole points toward the phosphates, stabilizing the negative charges, and its residues relocate slightly to interact with the phosphates. Ser64 adopts a single conformation pointing to the imino group between the first two phosphates of the dUpNHpp and together with Ser65 hydrogen bonds to the phosphate groups. Arg109 not only plays an important part in catalysis but also in anchoring the C-terminus. The crystal structures of YncF and YosS in complex with the nonhydrolysable substrate analogue dUpNHpp show that the *gauche*(−) conformation with the metal bound is not sufficient to trigger the ordering of the C-terminal arm and the 18–26 loop. It remains unclear why the C-terminal arm folds down in the presence of dUpNHpp and magnesium in some homologues while it does not in others. Two possible explanations are that the crystal packing influences the folding of the arm or that the enzymes from these species vary in the relative energetics of its ordering. The ligand in the *gauche*(+) conformation does not trigger any effect in the protein apart from a double conformation observed for the Phe-lid residue.

In the second substep the arm and the 18–26 loop become ordered. It is not possible to say which orders first or if the ordering is simultaneous or cooperative, but this is an important substep of the reaction. Firstly, it excludes all water molecules from the active site, surrounding the ligand with a more hydrophobic protein environment in which the electrostatic forces are enhanced by the low dielectric constant. Moreover, motif V provides new positively charged residues that interact with the negatively charged β- and γ-phosphates, thus polarizing the ligand as well as stabilizing the transition state. The role of motif V is similar to that of the P-loops in mononucleotide-binding proteins, as it contains glycines which form an anion hole to stabilize the negative charge of the phosphates and adopt dihedral angles that are only possible for this amino acid. The conserved arginine is in motif V but, in contrast to other dUTPases, the Phe-lid residue lies elsewhere and its position in motif V is taken by a leucine. The fact that the arm contacts the γ-phosphate explains why the homotrimeric dUTPases hydrolyse dUTP and not dUDP (Pécsi *et al.*, 2011).

Loop 18–26 undergoes a significant change at this point in order to complete the active site and, more importantly, its ordering is coupled to the change in conformation of the general base Asp82, which now points towards the active site. Superposition of human apo dUTPase and its dUpNHpp complex with an ordered C-terminal arm shows a structurally invariant equivalent 18–26 loop in which the main chain changes by a maximum of 1.5–2 Å.

4.3.3. The chemical step. dUTP is converted into dUMP and PP_i–metal^{2+} using electrostatic stabilization of the transition state and acid–base catalysis. Electrostatic interactions from protein side chains, the metal ion and the positive charges from the helix dipole, Arg62, Arg63 and Arg135 from motif V, polarize the triphosphate chain, destabilizing the ground state and stabilizing the transition state (Pécsi *et al.*, 2011). The general base removes a proton from the nucleophile water, which in the *gauche*(−) active state is optimally positioned to perform an in-line attack on the α-phosphate, passing through a pentacoordinate transition state and finally leading to a PP_i leaving group.

Obtaining an ordered structure of the C-terminal arm in trimeric dUTPases has been a challenging task since it does not take up an ordered conformation simply upon the binding of trinucleotide substrate analogues or products. Indeed, its degree of order varies somewhat randomly among different species. We propose that aiming for the structure of a similar complex offers a general strategy for trapping the elusive arm in other trimeric dUTPases.

4.3.4. Substrate release. Once the chemical step is complete, the products dUMP and PP_i–Mg^{2+} must leave the active site. This requires that the 18–26 loop and the aspartate general base return to their open positions and that the arm opens and becomes disordered. The PP_i–Mg^{2+} probably leaves immediately, perhaps aided by the interaction with the arm, and finally the dUMP leaves the active site. If the *gauche*(+) and *gauche*(−) conformations have biological relevance, it can be supposed that the two types of active sites represent two stages of product release. The initial *gauche*(+) active site, with the Phe-lid still in the closed conformation and the phosphate still well accommodated in the active site, would be followed by the *gauche*(−) conformation in which the Phe-lid residue has flipped to the open conformation and the phosphate is starting to diffuse away, perhaps aided by the Ser63 side chain.

The role of Ser63 should be probed experimentally through site-directed mutagenesis.

4.4. Special features of the *B. subtilis* dUTPases

Upon structural superposition with their most studied orthologues (from FIV, EIAV, *Vaccinia virus*, *E. coli*, *M. tuberculosis*, *Saccharomyces cerevisiae* and human), it is evident that *B. subtilis* dUTPases contain unique sequence and structural features. Firstly, the Phe-lid residue is positioned next to motif III, while in all others it is in motif V. This residue flips to stack over the uracil when dUTP–Mg^{2+} binds in the *gauche*(+) conformation and contributes to ligand binding and catalysis (Pécsi *et al.*, 2010). The additional step in which the disordered lower loop and the inactive conformation of the aspartate general base move to create an active conformation of the enzyme is also special to *B. subtilis* dUTPases. The lower loop shifts 10 Å from the native to the ordered complex structure, a movement that has not previously been observed in any other trimeric dUTPase upon ligand binding and ordering of the C-terminal arm. In all other orthologues the loop is always ordered and the catalytic aspartate is always in the active conformation.

A remaining question is why there are two dUTPases in *B. subtilis*. Looking at *B. subtilis* mRNA expression profiles (Nicolas *et al.*, 2012), it can be seen that only the genomic YncF is expressed at all times, while YosS is only expressed under certain conditions when the whole prophage is also expressed. The fact that its dUTPase gene has been retained is probably related to the nature of the SPβ prophage, which is nonstandard in the sense that it contains uracil instead of thymine in its DNA. The SPβ phage genome codes for several proteins in addition to dUTPases, the role of which is to inhibit uracil-repair enzymes while enhancing other metabolic pathways leading to the accumulation of dUTP. *B. subtilis* may have retained two dUTPases as a defence mechanism to keep infection by such phages to a minimum. Indeed, the unique features of *B. subtilis* dUTPases may have evolved under pressure to escape from the action of the protein inhibitors present in such bacteriophages.

We thank Glyn Hemsworth for fruitful discussions, Simon Grist for carrying out the mass spectrometry and Eleanor Dodson for her help in the interpretation of the cubes. We thank the staff at Diamond (beamlines I04-1 and I24) and the ESRF (beamline ID23-2) for provision of synchrotron facilities and Sam Hart for assistance in data collection.

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.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Baxter, N. J., Olguin, L. F., Golicnik, M., Feng, G., Hounslow, A. M., Bermel, W., Blackburn, G. M., Hollfelder, F., Waltho, J. P. & Williams, N. H. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 14732–14737.
Carlier, M.-F., Didry, D., Simon, C. & Pantaloni, D. (1989). *Biochemistry*, **28**, 1783–1791.
Cedergren-Zeppezauer, Larsson, G., Nyman, P., Dauter, Z. & Wilson, K. S. (1992). *Nature (London)*, **355**, 740–743.
Chan, S. *et al.* (2004). *J. Mol. Biol.* **341**, 503–517.
Cliff, M. J., Bowler, M. W., Varga, A., Marston, J. P., Szabo, J., Hounslow, A. M., Baxter, N. J., Blackburn, G. M., Vas, M. & Waltho, J. P. (2010). *J. Am. Chem. Soc.* **132**, 6507–6516.
Cowtan, K. (2008). *Acta Cryst.* D**64**, 83–89.
Dauter, Z., Persson, R., Rosengren, A. M., Nyman, P. O., Wilson, K. S. & Cedergren-Zeppezauer, E. S. (1999). *J. Mol. Biol.* **285**, 655–673.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Ephrati-Elizur, E., Yosuv, D., Shmueli, E. & Horowitz, A. (1974). *J. Bacteriol.* **119**, 36–43.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Fontana, A., de Laureto, P. P., Spolaore, B., Frare, E., Picotti, P. & Zambonin, M. (2004). *Acta Biochim. Pol.* **51**, 299–321.
Freeman, L., Buisson, M., Tarbouriech, N., Van der Heyden, A., Labbe, 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.
García-Nafría, J., Burchell, L., Takezawa, M., Rzechorzek, N. J., Fogg, M. J. & Wilson, K. S. (2010). *Acta Cryst.* D**66**, 953–961.
García-Nafría, J., Harkiolaki, M., Persson, R., Fogg, M. J. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 167–175.
Ghim, S.-Y., Choi, S.-K., Shin, B.-S. & Park, S.-H. (1998). *DNA Res.* **5**, 121–126.
Graham, D. L., Lowe, P. N., Grime, G. W., Marsh, M., Rittinger, K., Smerdon, S. J., Gamblin, S. J. & Eccleston, J. F. (2002). *Chem. Biol.* **9**, 375–381.
el-Hajj, H. H., Zhang, H. & Weiss, B. (1988). *J. Bacteriol.* **170**, 1069–1075.
Harwood, C. R. (1992). *Trends Biotechnol.* **10**, 247–256.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kondo, J. & Westhof, E. (2011). *Nucleic Acids Res.* **39**, 8628–8637
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.
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.* D**60**, 2256–2268.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Larsson, G., Svensson, L. A. & Nyman, P. O. (1996). *Nature Struct. Biol.* **3**, 532–538.
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.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* D**67**, 386–394.
Mol, C. D., Harris, J. M., McIntosh, E. M. & Tainer, J. A. (1996). *Structure*, **4**, 1077–1092.
Moroz, O. V., Harkiolaki, M., Galperin, M. Y., Vagin, A. A., Gonzalez-Pacanowska, D. & Wilson, K. S. (2004). *J. Mol. Biol.* **342**, 1583–1597.
Muha, V., Horváth, A., Békési, A., Pukáncsik, M., Hodoscsek, B., Merényi, G., Róna, G., Batki, J., Kiss, I., Jankovics, F., Vilmos, P., Erdélyi, M. & Vértessy, B. G. (2012). *PLoS Genet.* **8**, e1002738.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst.* D**61**, 1426–1431.
Nicolas, P. *et al.* (2012). *Science*, **335**, 1103–1106.
Nord, J., Kiefer, M., Adolph, H. W., Zeppezauer, M. M. & Nyman, P. O. (2000). *FEBS Lett.* **472**, 312–316.
Pearl, L. H. & Savva, R. (1996). *Nature Struct. Biol.* **3**, 485–487.

Pécsi, I., Hirmondo, R., Brown, A. C., Lopata, A., Parish, T., Vértessy, B. G. & Tóth, J. (2012). *PLoS One*, **7**, e37461.
Pécsi, I., Leveles, I., Harmat, V., Vértessy, B. G. & Tóth, J. (2010). *Nucleic Acids Res.* **38**, 7179–7186.
Pécsi, I., Szabó, J. E., Adams, S. D., Simon, I., Sellers, J. R., Vértessy, B. G. & Tóth, J. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 14437–14442.
Prasad, G. S., Stura, E. A., Elder, J. H. & Stout, C. D. (2000). *Acta Cryst.* D**56**, 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.
Slupphaug, G., Kavli, B. & Krokan, H. E. (2003). *Mutat. Res.* **531**, 231–251.
Tchigvintsev, A., Singer, A. U., Flick, R., Petit, P., Brown, G., Evdokimova, E., Savchenko, A. & Yakunin, A. E. (2011). *Biochem. J.* **437**, 243–253.
Tormo-Más, M. A., Mir, I., Shrestha, A., Tallent, S. M., Campoy, S., Lasa, I., Barbé, J., Novick, R. P., Christie, G. E. & Penadés, J. R. (2010). *Nature (London)*, **465**, 779–782.
Tóth, J., Varga, B., Kovács, M., Málnási-Csizmadia, A. & Vértessy, B. G. (2007). *J. Biol. Chem.* **282**, 33572–33582.
Trapani, S. & Navaza, J. (2008). *Acta Cryst.* D**64**, 11–16.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Varga, B., Barabás, O., Kovári, J., Tóth, J., Hunyadi-Gulyás, E., Klement, E., Medzihradszky, 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. & Tóth, J. (2009). *Acc. Chem. Res.* **42**, 97–106.
Whittingham, J. L. *et al.* (2005). *Structure*, **13**, 329–338.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.