STRUCTURAL BIOLOGY
COMMUNICATIONS

ISSN 2053-230X

Received 3 November 2015

Accepted 26 January 2016

Edited by N. Sträter, University of Leipzig,
Germany

Keywords: X-ray analysis; protein crystallography; nucleoside phosphorylases; nucleosides; substrate specificity; *Salmonella typhimurium*; thymidine phosphorylase.

PDB references: thymidine phosphorylase, unliganded, 4xr5; complex with thymidine, 4yek; complex with uridine, 4yyy

Supporting information: this article has supporting information at journals.iucr.org/f

Structural investigation of the thymidine phosphorylase from *Salmonella typhimurium* in the unliganded state and its complexes with thymidine and uridine

Vladislav V. Balaev,^a Alexander A. Lashkov,^a Azat G. Gabdulkhakov,^a Maria V. Dontsova,^a Tatiana A. Seregina,^b Alexander S. Mironov,^b Christian Betzel^c and Al'bert M. Mikhailov^{a*}

^aA. V. Shubnikov Institute of Crystallography, Leninsky Prospekt 59, Moscow 119333, Russian Federation, ^bState Research Institute of Genetics and Selection of Industrial Microorganisms, 1-st Dorozhny Proezd 1, Moscow 117545, Russian Federation, and ^cLaboratory for Structural Biology of Infection and Inflammation, University of Hamburg, Institute of Biochemistry and Molecular Biology, c/o DESY, Building 22a, Notkestrasse 85, 22603 Hamburg, Germany. *Correspondence e-mail: amm@ns.crys.ras.ru

Highly specific thymidine phosphorylases catalyze the phosphorolytic cleavage of thymidine, with the help of a phosphate ion, resulting in thymine and 2-deoxy- α -D-ribose 1-phosphate. Thymidine phosphorylases do not catalyze the phosphorolysis of uridine, in contrast to nonspecific pyrimidine nucleoside phosphorylases and uridine phosphorylases. Understanding the mechanism of substrate specificity on the basis of the nucleoside is essential to support rational drug-discovery investigations of new antitumour and anti-infective drugs which are metabolized by thymidine phosphorylases. For this reason, X-ray structures of the thymidine phosphorylase from *Salmonella typhimurium* were solved and refined: the unliganded structure at 2.05 Å resolution (PDB entry 4xr5), the structure of the complex with thymidine at 2.55 Å resolution (PDB entry 4yek) and that of the complex with uridine at 2.43 Å resolution (PDB entry 4yyy). The various structural features of the enzyme which might be responsible for the specificity for thymidine and not for uridine were identified. The presence of the 2'-hydroxyl group in uridine results in a different position of the uridine furanose moiety compared with that of thymidine. This feature may be the key element of the substrate specificity. The specificity might also be associated with the opening/closure mechanism of the two-domain subunit structure of the enzyme.

1. Introduction

Pyrimidines and pyrimidine nucleosides are already widely used in the treatment of cancer and infectious diseases (Bronckaers *et al.*, 2009). Thymidine-specific nucleoside phosphorylase (thymidine phosphorylase; TP; EC 2.4.2.4) is one of the enzymes which participate in the metabolism of these drugs. The drugs which are metabolized by thymidine phosphorylases include 5-(*E*)-(2-bromovinyl)-2'-deoxyuridine, 5-trifluorothymidine, 5-fluorouracil and 5-fluoro-5'-deoxyuridine (an intermediate metabolite of capecitabine). All of these drugs are clinically used against metastatic breast and colon cancers (Bronckaers *et al.*, 2009). Detailed structural information about their substrate specificity is essential to understand the mechanisms of binding and cleavage of pharmacologically important ligands and to support rational drug design.

Thymidine phosphorylase (TP) catalyzes the thymidine cleavage reaction in the presence of phosphate ion to give thymine and 2-deoxy- α -D-ribose 1-phosphate (Fig. 1). TP is a

© 2016 International Union of Crystallography

member of the NP-II family of nucleoside phosphorylases and is further characterized as a nonspecific pyrimidine nucleoside phosphorylase (PyNP; Pugmire *et al.*, 1998; Pugmire & Ealick, 2002). The members of the NP-II family of nucleoside phosphorylases differ from the members of the NP-I family by the presence of a two-domain subunit structure in contrast to a one-domain structure (NP-I family) and the location of the active site within the interface between the domains but not that between the subunits (NP-I family). The homology of the bacterial TP to human TP is 39% (Pugmire *et al.*, 1998), but the active-site geometry of the enzyme from prokaryotic and eukaryotic cells is identical. Thus, the results obtained for the bacterial TP might be extrapolatable to human TP. The usage of the bacterial enzyme is also supported by the fact that it is more difficult to crystallize well diffracting crystals of human TP than of the bacterial TP.

Mutagenesis studies (Mitsiki *et al.*, 2009) and chemo-kinetic studies (Panova *et al.*, 2007) of the cleavage of different nucleoside analogues by thymidine phosphorylases have been performed to investigate the kinetic parameters of TP–ligand interactions and of TP substrate specificity. As a result, the amino-acid residues which take part in the binding and/or the cleavage reaction and in the affinity for different thymidine modifications have been identified. Modelling experiments were performed (Pugmire *et al.*, 1998; Rick *et al.*, 1999) in order to determine the spatial organization and to investigate enzyme–substrate complexes. However, these experiments did not provide reliable structural information to understand the structural origin of the specificity of TP towards pyrimidine nucleosides. To date, only one TP structure (from *Escherichia coli*) in complex with pyrimidine or a pyrimidine nucleoside is present in the PDB. It was solved by Walter and coworkers and has a thymine in the active site (Walter *et al.*, 1990). However, this structure has only C^α atoms deposited and therefore it is not possible to investigate the structural aspects of the binding of pyrimidine bases and nucleosides to the enzyme.

2. Materials and methods

2.1. Cloning and expression of the *deoA* thymidine phosphorylase gene from *Salmonella typhimurium*

Amplification of the structural domain of the *deoA* gene was performed from the chromosome of *S. typhimurium* strain

LT2 (from the collection of the State Research Institute of Genetics and Selection of Industrial Microorganisms) using the primers Tpp1, 5'-CGCCATATGTTTCTCGCACAAG-AAATTATTC-3', and Tpp2, 5'-CGCCTCGAGCTATTCA-GTAATTCGACGATAGA-3'. The first oligonucleotide is complementary to the proximal region and the second is complementary to the distal region of the *deoA* gene. PCR was performed on a MyCycler thermal cycler (Bio-Rad) according to the following scheme. Firstly, cells were disrupted at 95°C for 3 min. 25 cycles of amplification were then carried out that included DNA denaturation at 95°C for 30 s, primer annealing at 60°C for 30 s and DNA extension at 72°C for 30 s. In the last stage, DNA extension was performed at 72°C for 2 min. As a result, a fragment of 1321 bp was synthesized that contained the structural region of the *deoA* gene flanked by NdeI and XhoI restriction-enzyme recognition sites. After electrophoretic separation of the PCR products, the desired DNA fragment was eluted from the gel using a GeneClean kit (Fermentas). The *deoA* gene was cloned into the pET-15b high-copy expression vector containing the T7 phage promoter, which is inducible by isopropyl β -D-1-thiogalactopyranoside (IPTG), at the NdeI and XhoI restriction endonuclease sites. The pET-15b vector contains an ampicillin-resistance marker (Amp^R) and nucleotide sequences for a His tag before the NdeI restriction-enzyme recognition site and for the target site of thrombin. *E. coli* strain TGI cells were transformed with the resulting ligase mixture. Selection of transformants was performed in LB agar medium containing ampicillin (100 $\mu\text{g ml}^{-1}$) as a selection marker. Screening for recombinant clones was performed by PCR using the plasmid primers pT7P, 5'-TAATACGACTCACTATAGGGG-3', and pT7T, 5'-GCTAGTTATTGCTCAGCGGT-3'. Plasmid DNA was isolated from the selected transformants, and the presence of the insertion in the hybrid plasmids was determined using restriction analysis. This plasmid, which was designated pET-15b/*deoA*, was used to transform *E. coli* strain BL21 (DE3) cells containing the inducible T7 bacteriophage RNA polymerase gene.

Induction of *S. typhimurium* TP (SrTP) expression was performed by the addition of IPTG to the growth medium to a final concentration of 1 mM. The expression of the *deoA* gene was analyzed by SDS-PAGE of the total protein from the *E. coli* BL21 (DE3) cells containing the pET-15b/*deoA* plasmid on a polyacrylamide gel under denaturing conditions. In the lanes of the gel which correspond to the strains

Figure 1
The reaction catalyzed by thymidine phosphorylase.

containing constructions with a functional *deoA* gene, an additional fraction of the protein monomer was observed. It has an approximate molecular weight of 48 kDa, which is in accordance with the molecular weight of *SrTP* of 47 kDa calculated from the *deoA* gene sequence. Data on the protein source and cloning procedures are listed in Table 1.

2.2. Biomass preparation, isolation and purification of the protein

The *E. coli* BL21 (DE3) cells containing the pET-15b/*deoA* plasmid were diluted 100 times and grown in a thermostatic shaker for 2 h at 37°C, followed by addition of IPTG and incubation for 4 h at 30°C. The biomass was gathered from a total culture volume of 1.2 l by centrifugation for 3 min at 5000g. The cell precipitate was suspended in binding (elution) buffer solution with the following composition: 20 mM Tris-HCl pH 7.5, 500 mM KCl, 30 mM imidazole. The cells were disrupted by sonication using a Fisher Scientific Model 550 dismembrator (Thermo Fisher Scientific, Waltham, USA). The debris was precipitated by centrifugation at 13 000g for 30 min at 4°C.

Chromatographic purification was performed on an ÄKTApurifier plus system (GE Healthcare, UK). The total protein obtained was loaded onto a 1 ml HisTrap FF column (packed with Ni Sepharose 6 Fast Flow; GE Healthcare) and equilibrated with the same buffer solution. A buffer solution consisting of 20 mM Tris-HCl pH 7.5, 500 mM KCl, 300 mM imidazole was used for elution. As a result, His-tagged *SrTP* protein was isolated.

The isolated *SrTP* protein fraction was dialyzed against buffer solution consisting of 20 mM Tris-HCl pH 8.4, 150 mM KCl, 2.5 mM CaCl₂ overnight at 4°C. The solution was treated with thrombin to remove the His tag in the same buffer solution for 4 h at room temperature. Thrombin was added at 1 U of enzyme per milligram of protein.

In the second stage of the purification, the protein was loaded onto a HiLoad 16/60 Superdex 200 pg (GE Healthcare) column for gel filtration equilibrated with the following buffer solution: 20 mM Tris-HCl pH 7.5, 20 mM KCl. The *SrTP* protein fractions were analyzed by SDS-PAGE.

2.3. Crystallization

Screening of crystallization conditions was performed using The PACT Suite, The pHClear Suite, Stura Footprint, Morpheus and JCSG crystallization kits (Qiagen, Germany; Molecular Dimensions, UK). Finally, crystals were grown by the vapour-diffusion method using the sitting-drop technique at 19°C. Crystallization and protein solution compositions are listed in Table 2. Crystals grew in a week (Fig. 2) and were used for X-ray diffraction data collection. Crystals of *SrTP* complexed with thymidine and uridine were obtained by soaking in a 0.1 M solution of the corresponding ligand. The approximate dimensions of the crystals were 400 × 50 × 50 µm (Fig. 2).

Table 1

Macromolecule-production information.

Source	<i>S. enterica</i> subsp. <i>enterica</i> serovar Typhimurium (strain LT2/SGSC1412/ATCC 700720)
GenBank ID	CKH45935
Gene	<i>deoA</i>
Forward primer	5'-CGCCATATGTTTCTCGACAAGAAATTAT-TC-3'
Reverse primer	5'-CGCCTCGAGCTATTAGTAATTCGACGAT-AGA-3'
Cloning vector	pET-15b
Expression vector	pET-15b
Expression host	<i>E. coli</i>
Complete amino-acid sequence of the construct produced	MFLAQEIIRKKRDGHALSDEEIRFFINGIRDNTI-SEGQIAALAMTIFFFHDMTPERVSLTMAMRDS-GTVLDWKSLLNLNGPIVDKSTGGVGDVTSMLL-GPMVAACGGYVPMISGRGLGHTGGTLDKLEAI-PGFDIFPDDNRFREIIQDVGVAIIGQTSSSLAP-ADKRFYATRDITATVDSIPLITGSILAKKLAE-GLDALVMDVKVSGAFMPTYELSEALAEATVG-VANGAGVRTTALLTDMNQVLASSAGNAVEVRE-AVQFLTGEYRNPRFLDVTMALCVEMLSGQLA-KDDAEARAKLQAVLDNGKAAEVFGRMVAQKG-PSDFVENYDKYLPATMLSKAVYADTEGFISAM-DTRALGMVAVSMGGRRQASDTIDYSVGFTDM-ARLGDSIDGQRPLAVIHAKDEASWQEAQAVK-AAIILDDKAPASTPSVYRRITE

2.4. Data collection and processing

Diffraction data sets for unliganded *SrTP* and its complexes were collected at 100 K on beamline 14.1 at BESSY II (Helmholtz-Zentrum, Berlin, Germany; Mueller *et al.*, 2012). Paraffin oil was used as a cryoprotectant. All data were indexed, merged and processed using the *XDS* package (Kabsch, 2010) and *SCALA* (Evans, 2011) within the *CCP4* suite (Winn *et al.*, 2011). The data statistics are summarized in Table 3.

2.5. Structure solution and refinement

The structure of *SrTP* was solved by molecular replacement using *Phaser* (McCoy, 2007) from the *CCP4* suite. The structure of TP from *E. coli* (PDB entry 4eaf; V. I. Timofeev, Y. A.

Figure 2
Crystals of unliganded thymidine phosphorylase from *S. typhimurium*.

Table 2
Crystallization.

	Unliganded <i>St</i> TP	<i>St</i> TP–THM	<i>St</i> TP–URI
Method	Sitting-drop vapour diffusion	Sitting-drop vapour diffusion	Sitting-drop vapour diffusion
Plate type	24-well plate	24-well plate	24-well plate
Temperature (K)	292	292	292
Protein concentration (mg ml ⁻¹)	19	19	19
Buffer composition of protein solution	20 mM Tris–HCl, 20 mM NaCl	20 mM Tris–HCl, 20 mM NaCl	20 mM Tris–HCl, 20 mM NaCl
Composition of reservoir solution	0.2 M potassium thiocyanate, 0.1 M bis-tris propane, 20% (w/v) PEG 3350 pH 7.5	0.1 M Bicine, 1.6 M (NH ₄) ₂ SO ₄ pH 9.0	0.1 M citric acid, 10% (w/v) PEG 6000 pH 5.0
Volume and ratio of drop	2 µl protein solution + 2 µl reservoir solution	2 µl protein solution + 2 µl reservoir solution	2 µl protein solution + 2 µl reservoir solution
Volume of reservoir (µl)	300	300	300

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

	Unliganded <i>St</i> TP	<i>St</i> TP–THM	<i>St</i> TP–URI
Diffraction source	Beamline 14.1, BESSY	Beamline 14.1, BESSY	Beamline 14.1, BESSY
Wavelength (Å)	0.918	0.918	0.918
Temperature (K)	100	100	100
Detector	Pilatus 6M	Pilatus 6M	Pilatus 6M
Crystal-to-detector distance (mm)	350	550	526
Rotation range per image (°)	0.1	0.1	0.2
Total rotation range (°)	200	120	120
Exposure time per image (s)	1	1	1
Space group	<i>P</i> 2 ₁ 2 ₁ 2	<i>I</i> 4	<i>I</i> 4
<i>a</i> , <i>b</i> , <i>c</i> (Å)	114.94, 171.21, 45.34	190.92, 190.92, 57.72	193.93, 193.93, 57.73
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Mosaicity (°)	0.23	0.11	0.16
Resolution range (Å)	47.72–2.05 (2.09–2.05)	39.02–2.55 (2.59–2.55)	48.06–2.43 (2.57–2.43)
Total No. of reflections	207489	142010	173860
No. of unique reflections	58911	64691	37930
Completeness (%)	95.4 (99.6)	97.4 (94.1)	92.8 (92.1)
Multiplicity	2.8 (2.6)	4.5 (4.5)	4.6 (4.7)
$\langle I/\sigma(I) \rangle$	15.70 (1.55)	15.72 (1.91)	11.97 (1.77)
CC _{1/2} (%)	99.8 (65.6)	99.9 (67.4)	99.9 (91.9)
<i>R</i> _{free}	0.057 (0.969)	0.071 (0.863)	0.077 (0.557)
Overall <i>B</i> factor from Wilson plot (Å ²)	44.5	68.2	62.6

Abramchik, R. S. Esipov & I. P. Kuranova, unpublished work) was used as a starting model to solve the unliganded structure. This structure, refined to 2.05 Å resolution (PDB entry 4xr5), was then used to solve the structures of *St*TP complexed with uridine and thymidine.

The *St*TP structures were refined using *PHENIX* (Adams *et al.*, 2010) and *REFMAC5* (Murshudov *et al.*, 2011). A visual monitoring of structure quality and its manual correction were performed using *Coot* (Emsley & Cowtan, 2004; Emsley *et al.*, 2010). The reasonableness of the refinement results was monitored using *Coot* (Emsley & Cowtan, 2004; Emsley *et al.*, 2010) and *MolProbity* (Chen *et al.*, 2010) at the *PDB Validation Server* (<http://validate.rcsb.org/>). All parameters of the refined *St*TP models are shown in Table 4. The final refinement cycle was performed in *phenix.refine* (Adams *et al.*, 2010).

Table 4
Structure solution and refinement.

Values in parentheses are for the outer shell.

	Unliganded <i>St</i> TP	<i>St</i> TP–THM	<i>St</i> TP–URI
PDB entry	4xr5	4yek	4yyy
Resolution range (Å)	47.72–2.05 (2.09–2.05)	39.02–2.55 (2.59–2.55)	48.06–2.43 (2.49–2.43)
Completeness (%)	95.5	97.4	92.5
No. of reflections			
Working set	51931 (2609)	61417 (2735)	36033 (2371)
Test set	2735 (138)	3274 (138)	1897 (125)
Final <i>R</i> _{cryst}	0.221 (0.336)	0.176 (0.362)	0.204 (0.356)
Final <i>R</i> _{free}	0.247 (0.386)	0.215 (0.436)	0.272 (0.448)
Cruickshank DPI (Å)	0.17	0.25	0.28
No. of non-H atoms			
Protein	6582	6580	6580
Ion	1	—	—
Ligand	37	96	83
Water	89	86	44
Total	6709	6812	6707
R.m.s. deviations			
Bonds (Å)	0.009	0.010	0.008
Angles (°)	1.275	1.325	1.157
Average <i>B</i> factors (Å ²)			
Protein	51.9	67.4	74.0
Ion	60.5	—	—
Ligand	61.3	93.9	81.0
Water	47.1	60.1	67.2
Ramachandran plot			
Most favoured (%)	97.83	96.58	96.01
Allowed (%)	0.11	0.00	0.00

3. Results and discussion

3.1. Three-dimensional structure of TP

The monomer of thymidine phosphorylase from *S. typhimurium* (*St*TP) consists of 440 residues (Schwartz, 1971; Fig. 3). 22 residues differ in the analogous enzyme from *E. coli* (*Ec*TP). These bacterial thymidine phosphorylases have 39% homology to human thymidine phosphorylase (*Hu*TP). An alignment of the primary structures of *Hu*TP and *St*TP is shown in Fig. 3. The *St*TP subunit as well as the other bacterial TP subunits (Walter *et al.*, 1990) consist of two domains. The first domain includes residues 1–65 and 163–193 and contains only α -helices (α -domain), while the other domain includes both α -helices and β -strands (α/β -domain; Fig. 4). The authors of the first structural investigation of a TP assumed that movement of the domains towards each other is essential for the catalytic reaction to proceed (Walter *et al.*,

1990). The α -helical fragments of *St*TP comprise 46% of the residues and the β -strands comprise 12.7% of the residues according to *DSSP* (Touw *et al.*, 2015). It is worth mentioning that despite the low homology of the primary structures of the bacterial thymidine phosphorylases (*St*TP and *Ec*TP) to human TP (HuTP), the secondary-structure elements are mostly conserved (based on PDB entries 4xr5, 4eaf and 1uou; Norman *et al.*, 2004). In *St*TP, as well as in *Ec*TP and HuTP, the domains are connected by three loops, L3 (66–81), L8 (156–162) and L10 (193–197) according to the *St*TP numbering (see

Figs. 3 and 4). The residues of the H1, H3 and H8 α -helices of both subunits of the *St*TP homodimer participate in inter-subunit interactions. The surface of the interaction interface covers about 5% of the total subunit surface (915 of 17 653 Å²). The calculated Gibbs free energy for homodimer formation is -11.2 kcal mol⁻¹ according to *PDBEPIA* (Krissinel & Henrick, 2007). The interaction between subunits is provided mainly by hydrogen bonds. The parameters of these bonds are similar to those listed in Walter *et al.* (1990). Hydrophobic interactions between subunits were not found.

Figure 3
Primary-structure alignment of *St*TP and HuTP with the secondary-structure elements of *St*TP (top) and of HuTP (bottom). Identical residues are marked in green and homologous residues in orange. The figure was prepared using the *ESPrpt* 3.0 online service (Gouet *et al.*, 1999). H, helices; S, β -strands; L, loops.

3.2. Phosphate-binding site

The bound sulfate, which is similar to phosphate, the original substrate of TP, was localized only in subunit *B* of the *SrTP*–thymine (*SrTP*–THM) complex. The sulfate was localized in a pocket formed by β -strand S1, the ends of the L6 loop and the turn between S1 and α -helix H5 (Figs. 4 and 5), and was further stabilized by the structural water molecule (HOH619 in *SrTP*–THM subunit *B*). The phosphate and the sulfate are practically identical in terms of the characteristics of the interaction. The atomic radii of phosphorus and sulfur differ by less than 2% (1.07 and 1.05 Å, respectively). The phosphate has a valency of 3. It has two H atoms bound to it at the pH optimum of TP (Panova *et al.*, 2007), while the sulfate has only one. The charges on the sulfate and the phosphate are the same. Structures of TP with bound phosphate are absent from the PDB. However, several structures with bound sulfate have been deposited, for instance the structures of *E. coli* thymidine phosphorylase with PDB codes 2tpt and 4eaf. The authors of the first structure (Pugmire *et al.*, 1998) claim that the binding of the phosphate or sulfate results in domain movement induced by the stabilization of loop L6 (residues 114–122), which results in the formation of a stable hydrogen bond between the NE2 atom of the His119 side chain and the O atom of the Gly208 main chain. In *SrTP* all residues of the L6 loop have relatively high *B*-factor values compared with the mean values for each structure: unliganded *SrTP* and the complexes with thymidine and uridine. For instance, the mean value of the *B* factor for all of the residues of subunit *B* of the *SrTP*–THM complex is 67.4 Å², while the mean value for the L6 loop residues is 113 Å². The distance between the NE2 atom of His119 and the O atom of Gly208 in subunit *B* of the *SrTP*–THM complex has a value of 3.7 Å, despite the presence of sulfate and thymidine in the active site of the subunit.

3.3. The binding of TP to nucleosides

3.3.1. The active site. Thymidine molecules were localized in the active sites (Fig. 5*b*) of both subunits of the *SrTP* homodimer (PDB entry 4yek). In both subunits the bonds formed to the ligand are similar and the difference in the bond length does not exceed the coordinate error estimate (DPI = 0.25 Å). In the structure of *SrTP* complexed with uridine (*SrTP*–URI; PDB entry 4yyy), uridine was only localized in the active site of subunit *B* (Fig. 5*c*). The location of the pyrimidine component of uridine in *SrTP*–URI is similar to that found for thymidine in the *SrTP*–THM complex. The position of the ribose component of uridine in the *SrTP*–URI complex differs from the position of the ribose component of thymidine in *SrTP*–THM. The ribose component of uridine forms two hydrogen bonds to residues of the enzyme (to Leu117 and Thr87), while that of THM forms only one (to Thr87) (Figs. 5*b* and 5*c*). The r.m.s.d. between the coordinates of the furanose components is 4.0 Å.

Thymidine differs from uridine in the presence of a methyl group at the fifth position and the absence of the hydroxyl group at the 2'-position. The absence of the 5-methyl group in thymidine (2'-deoxyuridine) results in a fourfold lower value

of the ratio of the reaction rate constant (k_{cat}) to K_{m} : $k_{\text{cat}}/K_{\text{m}}$ is 0.18 s μM^{-1} for 2'-deoxyuridine compared with 0.66 s μM^{-1} for thymidine (Panova *et al.*, 2007). Thymidine with a hydroxyl group at the 2'-position becomes an inhibitor (Panova *et al.*, 2007). We explain the difference in the position of the ribose component in the ligands (uridine and thymidine) in the active site of TP by the presence of the hydroxyl group at the 2'-position in uridine. This leads to a change of the Leu117 side chain (hydrophobic): it rotates by $\sim 180^\circ$ around the axis passing through its main chain. The shift of the Leu117 side chain results in a shift of its main chain towards the O2' atom of uridine and the formation of a hydrogen bond to it. The formation of this hydrogen bond affects the position of the uridine ribose ring.

However, it is worth mentioning that this difference in the location of the ribose component might be associated with the absence of the first substrate (phosphate) or the pseudo-substrate (sulfate) in the *SrTP*–URI structure. Citrate was localized in the phosphate-binding site of the *SrTP*–URI complex. The C2 atom of the citrate is at a distance of ~ 4 Å from the C5' atom of the thymidine when superposing the *SrTP*–URI and *SrTP*–THM *B* subunits.

Arg171, Ser186 and Tyr168 (Fig. 5*b*) are reported to be significant for the reaction to proceed according to kinetic studies of human TP. Substitution of Arg171 by Glu or Ser, thus changing the charge of the residue to a negative charge, leads to a loss of TP activity (Mitsiki *et al.*, 2009). Substitution of Ser186 by Gly also leads to a loss of enzymatic capability. The binding of thymidine and uridine is also supported by a weak perpendicular stacking interaction of the pyrimidine ring with Tyr168. The distance between the centres of the pyrimidine ring of thymidine and the side chain of Tyr168 is 5.2 Å for thymine (in both subunits) and 4.7 Å for uridine. The substitution of Tyr168 by other residues with hydrophobic properties results in a twofold decrease in the activity on substitution with Leu and a threefold decrease on substitution with Phe (Mitsiki *et al.*, 2009). TP loses its activity when Ala is present at the position of Tyr168 (Mitsiki *et al.*, 2009). The amino-acid

Figure 4
SrTP homodimer structure in the unliganded state. The α -domain is marked in yellow and the α/β -domain in blue. The secondary-structure elements are labelled for subunit *A*. The active site containing thymidine (THM₁) and the buffering site containing thymidine (THM₂) are shown for subunit *B*. The figure was produced using PyMOL (Schrödinger).

residues which take part in the interaction with substrates are conserved in the TPs from higher and lower organisms (Fig. 3).

3.3.2. The buffering site of thymidine and uridine. The thymidine in the *SrTP*–THM complex structure (in both subunits) and the uridine in the *SrTP*–URI complex structure (in subunit *B*) were localized in the additional binding site (Fig. 6). This site is located near the H11 α -helix and two loops: L12 and L13 (see Figs. 3 and 4). The amino-acid residues within this binding site of TP are not conserved between the TPs from prokaryotic and eukaryotic cells. In particular, Tyr267 is substituted by Gly298 in HuTP and Gln261 is substituted by Leu292 in HuTP. Only one structure of TP with bound substrates in the active site has been deposited in the

PDB: the structure of *EcTP* with a thymine and a sulfate (PDB entry 1tp1). In the second binding site, the thymidine molecule is absent. However, in the accompanying paper (Walter *et al.*, 1990) the authors mention that a thymine molecule was localized near residues 263–268. Structures of nonspecific pyrimidine nucleoside phosphorylases with substrates from *Bacillus stearothermophilus* (*BsPyNP*; PDB entry 1brw; 42.5% homology to *SrTP*; Pugmire & Ealick, 1998) with bound uracil and from *Staphylococcus aureus* (*SaPyNP*; PDB entry 3h5q); 67.2% homology to *SrTP*; Center for Structural Genomics of Infectious Diseases, unpublished work) with bound thymidine have been deposited and neither of them has a substrate localized in the second binding site.

The binding of the pyrimidine component of the second substrate is provided by the stacking interaction with Tyr267. The distance between the centre of the tyrosine aromatic group and the centre of the pyrimidine component of the thymidine (uridine) is 3.6 Å, *i.e.* these groups are more closely located by 1 Å compared with the distance between them in the active site. Their relative position in this region also differs from that of the active site: the aromatic groups are parallel to each other. In subunit *A* of the *SrTP*–THM complex and in subunit *B* of the *SrTP*–URI complex the following hydrogen bond is formed: THM(URI) O2–Tyr267 OH (3.4 Å in *SrTP*–THM subunit *A* and 2.6 Å in *SrTP*–URI subunit *B*). In subunit *B* of this complex the THM O2–Tyr267 OH bond is absent and the ribose component of the uridine is rotated by $\sim 180^\circ$ around the C–N glycosidic bond compared with its conformation in subunit *A*. In subunit *B* of the *SrTP*–URI complex the position of the uridine in the second binding site differs from the position of thymidine in this binding site in both subunits of the *SrTP*–THM complex. The pyrimidine

Figure 5
The active site of *SrTP* complexed with sulfate (*a*), with thymidine (THM; PDB entry 4yek) (*b*) and with uridine (URI; PDB entry 4yyy) (*c*). The residues which form hydrogen bonds to the nucleoside are marked in orange. The residues forming stacking interaction are marked in blue. The residues of the hydrophobic pocket in the nucleoside-binding site are marked in cyan. The residues which bind the sulfate (SO₄) in the complex of *SrTP* with thymidine are marked in green. The ligands are shown with the corresponding OMIT map (magenta).

component is rotated around the axis passing through its centre in such a manner that the hydrogen bond to Tyr267 OH in subunit *B* is formed by the O4 but not the O2 substrate atom. Compared with thymidine, this component is rotated around the C—N glycosidic bond. It results in the formation of the following hydrogen bond of length 3.3 Å in subunit *B*: URI502 O4—Ser246 OG. We assume that this position of the uridine is more energetically favourable than the position of thymidine in *SrTP*–THM. The thymidine cannot be located in the same position as the uridine as it has an additional methyl group at the fifth position. Placing the thymidine in a position

similar to the uridine would lead to this hydrophobic methyl group being at a distance of ~ 4 – 5 Å from the O atoms: Ser248 O and Glu258 OE1/OE2. As a result the whole uridine molecule has a different orientation compared with the thymidine, and the ribose component of the uridine forms only one hydrogen bond to the enzyme atoms: URI502 O2'–Gln261 OE1 (2.9 Å).

3.4. Domain configuration in TP complexed with substrates

In the structures of the *SrTP*–THM (PDB entry 4yek) and *SrTP*–URI (PDB entry 4yyy) complexes and in the structure of unliganded *SrTP* the asymmetric unit contains a homodimer and the subunits have a different conformation in each structure. In order to determine the relative positions of the domains in the subunits of the *SrTP*–THM and *SrTP*–URI complexes and in unliganded *SrTP*, the atoms of the main chain of the α/β -domains for all of the subunits from these structures were superimposed in a pairwise manner. For each pair the mean displacement of the α -domain of one subunit from the α -domain of the other subunit was calculated. Relatively small values were found for the superposition of the *A* subunits of the *SrTP*–THM complex and the *SrTP*–URI complex (0.4 Å) and for the superposition of the *B* subunits of the *SrTP*–URI and *SrTP*–THM complexes (0.7 Å), and for the superposition of each of these two subunits with subunit *A* of unliganded *SrTP* (0.9 and 1.0 Å, respectively). All other values of the α -domain displacement were higher than 2 Å. Based on these results, we assume that two subunit conformations exist in the *SrTP* structures complexed with substrates: in each case one subunit of the homodimer is in one conformation and the second subunit is in the other. It is most likely that this occurs owing to the fact that the α -domains, as mentioned above, are responsible for the interactions between subunits of the homodimer. When one subunit moves upon binding substrate, the other is opened. After the reaction in the first subunit is over, it opens and the other subunit closes for the reaction to take place within it. However, the unit cells of *EcTP* complexed with 3'-azido-2'-fluorodeoxyuridine and with 3'-azido-3'-deoxythymidine (PDB entries 4ead and 4lhm, respectively; Timofeev *et al.*, 2014) contain only one subunit, which means that both subunits of the *EcTP* homodimer are identical.

Although the nucleoside is bound in the active sites of the *SrTP* complexes described in this paper, as well as in the abovementioned structures of *EcTP*, the distance between the phosphate-binding site and the second substrate-binding site corresponds to different variants of the open conformation described in Pugmire *et al.* (1998). Thus, we suppose that the mechanism of the opening/closure of the subunit domains might also be responsible for the substrate specificity of the thymidine and pyrimidine nucleoside phosphorylases. Based on quantum-mechanics calculations, Rick *et al.* (1999) claim that the potential barrier of C—N glycosidic bond cleavage in thymidine without subunit closure and thus without nucleophilic attack by the phosphate ion in *E. coli* TP is only 5 kcal mol⁻¹ higher than the barrier to nucleophilic attack.

Figure 6

Superposition of the three-dimensional structures of the second binding sites of subunits *A* and *B* of *SrTP* in complex with thymidine (PDB entry 4yek) (*a*) and of subunit *B* of *SrTP* in complex with thymidine and of *SrTP* in complex with uridine (PDB entry 4yyy) (*b*). The active-site residues of subunit *A* of the *SrTP*–THM complex are in cyan and the thymidine of subunit *A* is in dark blue with the corresponding bonds in black, the active-site residues of subunit *B* of the *SrTP*–THM complex are in green and the thymidine of subunit *B* is in dark green with the corresponding bonds in magenta, and subunit *B* of the *SrTP*–URI complex is in brown and the uridine is in orange with the corresponding bonds in orange.

According to Rick *et al.* (1999) the phosphate does not participate in the nucleophilic attack, but only binds a furanose radical in the second stage of the reaction. This fact is additionally supported by investigations of TP inhibition by ligands mimicking the transition state of the substrates (Kalman & Lai, 2005) and by investigations of the transition state of the substrate by means of quantum mechanics (Schwartz *et al.*, 2010). Schwartz and coworkers assume that nucleophilic attack on the thymidine is performed by a water molecule. Summing up all of the facts mentioned above, we conclude that a small distance between the thymidine and the phosphate might not be required for the reaction to proceed in TP.

3.5. The influence of ligand conformation on substrate specificity

TP is a highly specific nucleoside phosphorylase that catalyzes the phosphorolysis of thymidine and 2'-deoxyuridine and does not catalyze the phosphorolysis of uridine. In the *SrTP*-URI complex (PDB entry 4yyy) the uridine molecule is localized in only one subunit (*B*) of the two. In the *SrTP*-THM complex (PDB entry 4yek) the thymidine is localized both in the *A* and the *B* subunits, and sulfate is localized only in subunit *B*. The thymidine conformations in the active site are the same. In both complexes (*SrTP*-THM and *SrTP*-URI) the nucleoside is in the *syn* conformation, which is less energetically favourable than the *anti* conformation (Saenger, 1984). These conformations depend on the torsion angle of the glucosidic bond O4'-C1'-N1-C2 (χ angle). The *syn* conformation corresponds to values of χ ranging from 0 to 90° and from 270 to 360°. Values of χ between 90 and 270° correspond to the *anti* conformation. Thymidine in the *SrTP*-THM complex has χ values of 13° for subunit *A* and 25° for subunit *B*, and χ is equal to -75° for the *SrTP*-URI complex. The conformation of uridine is also named the high-*anti* conformation, as it corresponds to the potential barrier that appears when the H atoms of C6 and C2' are becoming closer. The distance between these atoms in the thymidine in subunit *A* of the *SrTP*-THM complex is 3.0 Å and that in subunit *B* is 3.2 Å, but for the uridine in the *SrTP*-URI complex this distance is 2.8 Å. The conformation of the uridine in the *SrTP*-URI complex is more stressed than the conformation of the thymidine found in both *SrTP*-THM subunits. Interestingly, the thymidine in *SaPyNP* is in the more energetically favourable and thus less stressed *anti* conformation (-148°), while in *SaPyNP*-THM, *SrTP*-URI and *SrTP*-THM identical residues are located near the nucleoside substrate. In uridine phosphorylases the uridine tends to have a *syn* conformation with high energy (Safonova *et al.*, 2014), which corresponds to the potential barrier that appears when the H atoms of O2 and C2 are becoming closer.

4. Conclusions

The following structures were solved, refined and deposited in the Protein Data Bank (PDB): thymidine phosphorylase from

S. typhimurium in the unliganded state (2.05 Å resolution; PDB entry 4xr5), in complex with thymidine (2.55 Å resolution; PDB entry 4yek) and in complex with uridine (2.43 Å resolution; PDB entry 4yyy). The specificity of TP for thymidine and not uridine might be based on the difference in the location of the furanose component in the TP active site. The presence of the second substrate (thymidine) was shown not to lead to TP subunit closure. The cleavage reaction thus might proceed without subunit closure, *i.e.* without nucleophilic attack by the phosphate. We assume that the role of the phosphate lies in stabilization of the transition state, as has been shown in several investigations.

The buffering site with bound nucleoside was determined during structural investigation of the complexes of *SrTP* with uridine and thymidine. This binding site is situated on the molecule surface and consists of the following residues: Tyr267, Ser248, Ser249, Arg257 and Gln261. An ability to inhibit two binding sites is particularly significant for regulation of the enzyme activity. This is additionally supported by the fact that TP has been referred to as an anticancer target not only owing to its role in pyrimidine metabolism, but also to its role as an angiogenesis factor.

Acknowledgements

The reported study was funded by RFBR according to research project No. 14-04-00952a. The study was also supported by basic financial support from the A. V. Shubnikov Institute of Crystallography, Russian Academy of Sciences.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Bronckaers, A., Gago, F., Balzarini, J. & Liekens, S. (2009). *Med. Res. Rev.* **29**, 903–953.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. R. (2011). *Acta Cryst.* **D67**, 282–292.
- Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kalman, T. I. & Lai, L. (2005). *Nucleosides Nucleotides Nucleic Acids*, **24**, 367–373.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- McCoy, A. J. (2007). *Acta Cryst.* **D63**, 32–41.
- Mitsiki, E., Papageorgiou, A. C., Iyer, S., Thiagarajan, N., Prior, S. H., Sleep, D., Finnis, C. & Acharya, K. R. (2009). *Biochem. Biophys. Res. Commun.* **386**, 666–670.
- Mueller, U., Darowski, N., Fuchs, M. R., Förster, R., Hellmig, M., Paithankar, K. S., Pühringer, S., Steffien, M., Zocher, G. & Weiss, M. S. (2012). *J. Synchrotron Rad.* **19**, 442–449.
- 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.* **D67**, 355–367.
- Norman, R. A., Barry, S. T., Bate, M., Breed, J., Colls, J. G., Ernill, R. J., Luke, R. W. A., Minshull, C. A., McAlister, M. S. B., McCall, E. J., McMicken, H. H. J., Paterson, D. S., Timms, D., Tucker, J. A. & Pauptit, R. A. (2004). *Structure*, **12**, 75–84.

- Panova, N. G., Alexeev, C. S., Kuzmichov, A. S., Shcheveleva, E. V., Gavryushov, S. A., Polyakov, K. M., Kritzyn, A. M., Mikhailov, S. N., Esipov, R. S. & Miroshnikov, A. I. (2007). *Biochemistry*, **72**, 21–28.
- Pugmire, M. J., Cook, W. J., Jasanoff, A., Walter, M. R. & Ealick, S. E. (1998). *J. Mol. Biol.* **281**, 285–299.
- Pugmire, M. J. & Ealick, S. E. (1998). *Structure*, **6**, 1467–1479.
- Pugmire, M. J. & Ealick, S. E. (2002). *Biochem. J.* **361**, 1–25.
- Rick, S. W., Abashkin, Y. G., Hilderbrandt, R. L. & Burt, S. K. (1999). *Proteins*, **37**, 242–252.
- Saenger, W. (1984). *Principles of Nucleic Acid Structure*. New York: Springer-Verlag.
- Safonova, T. N., Mikhailov, S. N., Veiko, V. P., Mordkovich, N. N., Manuvera, V. A., Alekseev, C. S., Kovalchuk, M. V., Popov, V. O. & Polyakov, K. M. (2014). *Acta Cryst.* **D70**, 3310–3319.
- Schwartz, M. (1971). *Eur. J. Biochem.* **21**, 191–198.
- Schwartz, P. A., Vetticatt, M. & Schramm, V. L. (2010). *J. Am. Chem. Soc.* **132**, 13425–13433.
- Timofeev, V., Abramchik, Y., Zhukhlistova, N., Muravieva, T., Fateev, I., Esipov, R. & Kuranova, I. (2014). *Acta Cryst.* **D70**, 1155–1165.
- Touw, W. G., Baakman, C., Black, J., te Beek, T. A. H., Krieger, E., Joosten, R. P. & Vriend, G. (2015). *Nucleic Acids Res.* **43**, D364–D368.
- Walter, M. R., Cook, W. J., Cole, L. B., Short, S. A., Koszalka, G. W., Krenitsky, T. A. & Ealick, S. E. (1990). *J. Biol. Chem.* **265**, 14016–14022.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.