Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structures of human cytosolic and mitochondrial nucleotidases: implications for structure-based design of selective inhibitors

Petr Pachl,[a,b] Milan Fábry,[a] Ivan Rosenberg,[b] Ondřej Šimák,[b] Pavlína Řezáčová[a,b] and Jiří Brynda[a,b]*

[a]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nám. 2, 166 37 Prague 6, Czech Republic, and [b]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nám. 2, 166 37 Prague 6, Czech Republic

Correspondence e-mail: brynda@img.cas.cz

Received 29 August 2013
Accepted 6 November 2013

PDB references: cytosolic deoxyribonucleotidase, 4l57; mitochondrial deoxyribonucleotidase, 4l6a

The human 5′(3′)-deoxyribonucleotidases catalyze the dephosphorylation of deoxyribonucleoside monophosphates to the corresponding deoxyribonucleosides and thus help to maintain the balance between pools of nucleosides and nucleotides. Here, the structures of human cytosolic deoxyribonucleotidase (cdN) at atomic resolution (1.08 Å) and mitochondrial deoxyribonucleotidase (mdN) at near-atomic resolution (1.4 Å) are reported. The attainment of an atomic resolution structure allowed interatomic distances to be used to assess the probable protonation state of the phosphate anion and the side chains in the enzyme active site. A detailed comparison of the cdN and mdN active sites allowed the design of a cdN-specific inhibitor.

1. Introduction

Monophosphate 5′-nucleotidases are ubiquitous enzymes that catalyze the dephosphorylation of nucleoside monophosphates and thus help to regulate cellular pools of nucleotides and nucleosides (Hunsucker *et al.*, 2005; Bianchi *et al.*, 1986). In addition to their physiological role, the human orthologues of these enzymes also participate in deactivation of nucleoside analogues used as antiviral and anticancer agents. Nucleoside analogues require nucleoside kinases for phosphorylation to their active form. Dephosphorylation of nucleoside analogues by human 5′-nucleotidases decreases the concentration of the active drug and can result in drug resistance (Mazzon *et al.*, 2003).

To date, seven human 5′-nucleotidases, which differ in substrate specificity, cellular location and tissue-specific expression, have been cloned and characterized (reviewed in Hunsucker *et al.*, 2005).

Two human 5′-nucleotidases, cytosolic 5′(3′)-deoxyribonucleotidase (cdN) and mitochondrial 5′(3′)-deoxyribonucleotidase (mdN), specifically dephosphorylate the deoxyribo form of nucleoside monophosphates (with dUMP and dTMP as the preferred substrates). A unique feature of these 5′(3′)-deoxyribonucleotidases is their ability to dephosphorylate the 5′- or 3′-phosphates of pyrimidine ribonucleotides. The two enzymes share 61% sequence identity and, as their names indicate, they differ in subcellular localization.

Within the cell, cdN, together with nucleoside kinases, regulates deoxynucleotide levels. The substrate specificity of mdN suggests that it plays a role in protecting mitochondria from excessive dTTP levels, which can contribute to mutagenesis during mtDNA replication (Rampazzo *et al.*, 2000), whereas molecules that specifically inhibit cdN could

potentially protect nucleoside-based drugs in their active form as nucleotides (Mazzon *et al.*, 2003).

Human cdN and mdN have been structurally characterized. The crystal structure of cdN at 1.2 Å resolution was determined in complex with the reaction product 2′-deoxyuridine and a tetrafluoroaluminate ion that mimics the transition state of phosphate-ion hydrolysis (PDB entry 2i7d; Walldén *et al.*, 2007).

The structure of wild-type mdN has been determined in the free form (PDB entry 1mh9; Rinaldo-Matthis *et al.*, 2002) and in complex with two inhibitors (PDB entries 1q91 and 1q92; Rinaldo-Matthis *et al.*, 2004). In addition, nine structures of an mdN active-site mutant (D41N) in complex with different substrates (Rinaldo-Matthis *et al.*, 2002; Walldén *et al.*, 2005) and substrate analogues (AZTMP and BVdUMP) have also been solved (Walldén *et al.*, 2007). Based on these mdN structures, a mechanism for the dephosphorylation of deoxyuridine 5′-monophosphate was proposed (Rinaldo-Matthis *et al.*, 2002), and this model was later corroborated by quantum-chemistry calculations (Himo *et al.*, 2005).

Here, we report the crystal structures of active human mdN and cdN determined at 1.4 and 1.08 Å resolution, respectively. Both structures contain a phosphate ion in the enzyme active site, which was acquired from the crystallization conditions. These high-resolution and atomic resolution structures thus represent the first relevant pair of mdN and cdN structures containing an identical ligand in the active site and could be used to perform a detailed comparison of the active-site conformations of these two enzymes. We used this comparison to propose a structural basis for the development of specific inhibitors. In addition, the structure of cdN at atomic resolution enabled us to determine the protonation states of catalytic site residues and led to validation of the proposed catalytic mechanism.

2. Experimental

2.1. Gene cloning

To obtain the coding sequence for cdN, a cDNA library was prepared from total RNA isolated from the MDA-MB-453 human breast carcinoma cell line. The cdN coding sequence was amplified from this library using the primers 5′-CATACATATGGCGCGGAGCGTGCG and 5′-CCTTGAATTCCAGCTGCTGCCCGC, resulting in a 712 bp DNA fragment that corresponded to the cdN sequence deposited in GenBank (accession No. NM014595). This fragment was re-amplified with primers designed to introduce *Nde*I and *Eco*RI restriction sites, enabling cloning into the pET-22b expression vector.

The mdN coding sequence was amplified from the Clontech cDNA QuickClone (Universal II) human cDNA library using the primers 5′-GATACATATGGGAGGCCGCGCCCTACG and 5′-CCCCACAGAGGAGCCCGAATTCCAGTC, which contain *Nde*I and *Eco*RI restriction sites, respectively. The sequence of the amplified fragment corresponded to the mdN sequence deposited in GenBank (accession No. AF210652). To obtain a construct for the expression of hexahistidine-tagged mdN, the vector pMCSG7 (T7 promoter-driven, originally designed for LIC cloning; Stols *et al.*, 2002) was modified. The modified vector encodes an N-terminal hexahistidine tag followed by a TEV protease recognition site, which is followed by a Ser-Asn-Ala-Ala-Ser linker, where Ala-Ser corresponds to an *Nhe*I site. The mdN coding sequence, re-amplified to introduce an *Nhe*I site at the 5′-terminus, was cloned into this modified vector *via* *Nhe*I–*Eco*RI sites. Cleavage with TEV protease thus yields mature mdN with an SNAAS sequence at its N-terminus.

2.2. Expression and purification

The proteins were overexpressed in *Escherichia coli* BL21 (DE3) cells grown in LB broth (Sigma) supplemented with 0.8%(*v*/*v*) glycerol and 100 µg ml^{-1} ampicillin. The bacteria were cultivated at 310 K and protein expression was induced with 400 m*M* isopropyl β-D-1-thiogalactopyranoside when the culture reached an OD_{600} of approximately 0.6. The cells were cultivated for an additional 4 h at 293 and 288 K for cdN and mdN, respectively.

The purification of cdN was performed as described previously (Rinaldo-Matthis *et al.*, 2002). Briefly, the bacteria were lysed by three cycles of 30 s sonication in TBS (20 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl). The lysate was clarified by centrifugation (48 000*g* for 30 min) and was precipitated with ammonium sulfate at 20–60% saturation. The precipitate was dissolved in water and dialyzed against 30 m*M* Tris–HCl pH 7.5. After centrifugation, the supernatant was loaded onto a Mono Q HR 5/5 column and eluted with a linear gradient of 0–500 m*M* NaCl. The cdN-rich fractions eluted around 95 m*M* NaCl. The collected fractions were combined and purified using size-exclusion chromatography on a Superdex 200 10/30 GL column (Amersham Bioscience) in TBS. The purified protein was dialyzed into buffer consisting of 20 m*M* Tris–HCl pH 7.5, 20 m*M* $MgCl_2$, 2 m*M* DTT, 1 m*M* EDTA and stored at a concentration of 1 mg ml^{-1} at 253 K.

For the purification of His_6-tagged mdN, the bacteria were lysed by three cycles of 30 s sonication in TBS. The lysate was clarified by centrifugation (48 000*g* for 30 min) and subjected to affinity chromatography using a His-Select Nickel Affinity Gel (Sigma) column equilibrated with TBS. The His_6-tagged mdN protein was eluted with 250 m*M* imidazole. The affinity tag was removed by incubation with recombinant His_6-tagged TEV protease overnight at 298 K while dialyzing against buffer consisting of 20 m*M* Tris–HCl pH 7.5, 2 m*M* DTT, 20 m*M* $MgCl_2$. The sample was subjected to a second round of Ni-affinity chromatography to remove the TEV protease and the cleaved His_6 tag. The purified mdN was dialyzed into buffer consisting of 20 m*M* Tris–HCl pH 7.5, 20 m*M* $MgCl_2$, 2 m*M* DTT, 1 m*M* EDTA and stored at a concentration of 1 mg ml^{-1} at 253 K.

The proteins were >95% pure as judged by silver-stained SDS–PAGE. Enzymatic activity was confirmed by an activity assay assessing the formation of inorganic phosphate from dUMP (Mazzon *et al.*, 2003; D'Angelo *et al.*, 2001). For

crystallization, the proteins were concentrated by ultrafiltration using Centricon YM-10 filters (Millipore).

2.3. Crystallization

For the crystallization of cdN and mdN, we optimized previously described crystallization conditions (Walldén *et al.*, 2007; Rinaldo-Matthis *et al.*, 2004).

Crystals of cdN were obtained using the hanging-drop vapour-diffusion technique at 291 K by mixing 1 µl protein solution (11.8 mg ml^{-1}) with 1 µl reservoir solution consisting of 40 m*M* KH_2PO_4 pH 4.5, 5%(*w*/*v*) PEG 8000, 20%(*v*/*v*) glycerol. Crystals appeared in 3 d and grew to final dimensions of 300 × 80 × 40 µm within two weeks. The crystals were flash-cooled by plunging them into liquid nitrogen and were stored in liquid nitrogen until use for X-ray diffraction experiments.

Microcrystals of mdN were obtained using the hanging-drop vapour-diffusion technique at 291 K. Crystallization drops consisting of 1.33 µl protein solution (8.4 mg ml^{-1}) and 0.66 µl precipitant solution consisting of 24 m*M* KH_2PO_4, 9.6%(*w*/*v*) PEG 8000, 12%(*v*/*v*) glycerol were equilibrated over a reservoir containing 500 µl precipitant solution. Nuclei from this drop were streak-seeded using a cat whisker into a drop consisting of protein and 20 m*M* KH_2PO_4, 8%(*w*/*v*) PEG 8000, 10%(*v*/*v*) glycerol. Crystals grew to final dimensions of 250 × 160 × 160 µm within 3 d. For cryoprotection, the crystals were soaked for 10 s in reservoir solution supplemented with 30%(*v*/*v*) glycerol. The crystals were flash-cooled by plunging them into liquid nitrogen and were stored in liquid nitrogen until use for X-ray diffraction experiments.

Table 1
Crystal parameters and data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Protein	cdN	mdN
Data-collection statistics		
Space group	*P*1	$P4_32_12$
Unit-cell parameters (Å, °)	*a* = 39.13, *b* = 46.49, *c* = 61.41, α = 68.16, β = 81.47, γ = 75.28	*a* = 73.41, *b* = 73.41, *c* = 106.34, α = β = γ = 90.00
No. of molecules in asymmetric unit	2	1
Wavelength (Å)	0.918	0.915
Resolution (Å)	22.48–1.08 (1.13–1.08)	25.95–1.40 (1.47–1.40)
No. of unique reflections	144108 (17466)	57598 (7565)
Multiplicity	3.1 (1.9)	5.0 (4.9)
Completeness (%)	86.3 (76.8)	97.6 (98.5)
R_{merge}† (%)	6.4 (23.5)	6.3 (47.2)
Average $I/\sigma(I)$	31.7 (1.9)	11.9 (2.8)
Wilson *B*‡ (Å^2)	12.4	19.7
Refinement statistics		
Resolution range (Å)	20.0–1.08 (1.13–1.08)	25.50–1.40 (1.44–1.40)
No. of reflections in working set	142636	54696
No. of reflections in test set	1463	2902
R§ (%)	14.4 (23.5)	14.1 (25.5)
R_{free}¶ (%)	18.3	15.4 (28.4)
R.m.s.d., bond lengths (Å)	0.014	0.017
R.m.s.d., angles (°)	2.4††	2.1
Contents of asymmetric unit		
Total No. of atoms	3845	2118
No. of protein atoms	3353	1636
No. of water molecules	443	312
Mean *B* value (Å^2)	16.1	18.5
Ramachandran plot statistics‡‡		
Residues in favoured regions (%)	99.0	97.0
Residues in allowed regions (%)	1.0	3.0
PDB code	4l57	4l6a

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the individual intensity of the *i*th observation of reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity of reflection *hkl* with summation over all data. ‡ The Wilson *B* was calculated using *SFCHECK* from the *CCP*4 suite (Winn *et al.*, 2011). § $R = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. ¶ R_{free} is equivalent to the *R* value but is calculated for 5% of the reflections that were chosen at random and omitted from the refinement process (Brünger, 1992). †† Estimated from the r.m.s.d. of the DANG parameter used for bonding-angle restraint during refinement in *SHELX*. ‡‡ As determined by *MolProbity* (Chen *et al.*, 2010).

2.4. Data collection, structure determination and analyses

Diffraction data were collected at 100 K on beamlines 14.1 and 14.2 of BESSY, Berlin, Germany (Mueller *et al.*, 2012) at wavelengths of 0.918 and 0.915 Å for cdN and mdN, respectively.

The complete cdN data set, recorded to 1.08 Å resolution, was comprised of three data sets collected from different parts of the same needle-shaped crystal. Diffraction data were processed using the *HKL*-3000 suite of programs (Minor *et al.*, 2006). The crystals exhibited *P*1 symmetry and contained two molecules in the asymmetric unit, with a solvent content of approximately 45%.

The crystal of mdN diffracted to a resolution of up to 1.40 Å. Diffraction data were integrated and reduced using *MOSFLM* (Leslie, 1999) and were scaled using *SCALA* (Evans, 2006) from the *CCP*4 suite (Winn *et al.*, 2011). The crystals exhibited $P4_32_12$ symmetry and contained one molecule in the asymmetric unit, with a solvent content of approximately 60%. Crystal parameters and data-collection statistics are given in Table 1.

The structure of cdN was determined by molecular replacement with *MOLREP* (Vagin & Teplyakov, 2000) using the structure of an identical protein as the template (PDB entry 2i7d; Walldén *et al.*, 2007). The initial model refinement was carried out using *REFMAC*5.2 (Murshudov *et al.*, 2011) from the *CCP*4 package (Winn *et al.*, 2011) interspersed with manual adjustments using *Coot* (Emsley & Cowtan, 2004). Anisotropic atomic displacement parameter refinement at atomic resolution was performed using *SHELX* (Sheldrick, 2008).

The structure of mdN was determined by difference Fourier techniques using the coordinates of the isomorphous structure of an identical protein (PDB entry 1mh9; Rinaldo-Matthis *et al.*, 2002). Refinement was carried out using *REFMAC*5.2

(Murshudov *et al.*, 2011). The final steps included TLS refinement (Winn *et al.*, 2001).

The quality of the final models was validated with *MolProbity* (Chen *et al.*, 2010). Refinement statistics are given in Table 1. All figures showing structural representations were prepared with *PyMOL* (DeLano, 2002). Atomic coordinates and experimental structure factors have been deposited in the Protein Data Bank under codes 4l57 and 4l6a for cdN and mdN, respectively.

2.5. Docking study

Complexes of human cdN or mdN with the inhibitor [5-benzyl-2-(thymin-1-yl)-tetrahydrofuran-4-yl]methylphosphonic acid (BTP-T) were prepared for docking in the *YASARA* modelling package (Krieger *et al.*, 2009) and were based on the structures solved in this work (PDB entries 4l57 and 4l6a). H atoms were added to the protein to mimic a pH value of 4.2 and their positions were optimized. The glycerol and water molecules were removed from the models and the charge of the Mg ion was set to +2. The parameter set used for the protein was AMBER ff03 (Duan *et al.*, 2003). The ligand was optimized in a vacuum and partial charges on its atoms were obtained by a restrained fit to the electrostatic potential (RESP) at the AM1BCC level (Jakalian *et al.*, 2002). The ligand was then docked into the protein using *AutoDock* (Morris *et al.*, 2009). A total of 1000 poses was obtained using a local search protocol. These were subsequently clustered based on similarity (poses with r.m.s.d. < 5 Å were clustered as a distinct complex) and clusters were then scored according their binding energy (Table 2).

3. Results and discussion

3.1. Description of the cdN and mdN structures

The structure of human cdN was refined to 1.08 Å resolution. In the final crystallographic model, each asymmetric unit consists of two cdN molecules encompassing residues 4–195. The three N-terminal residues could not be located in the

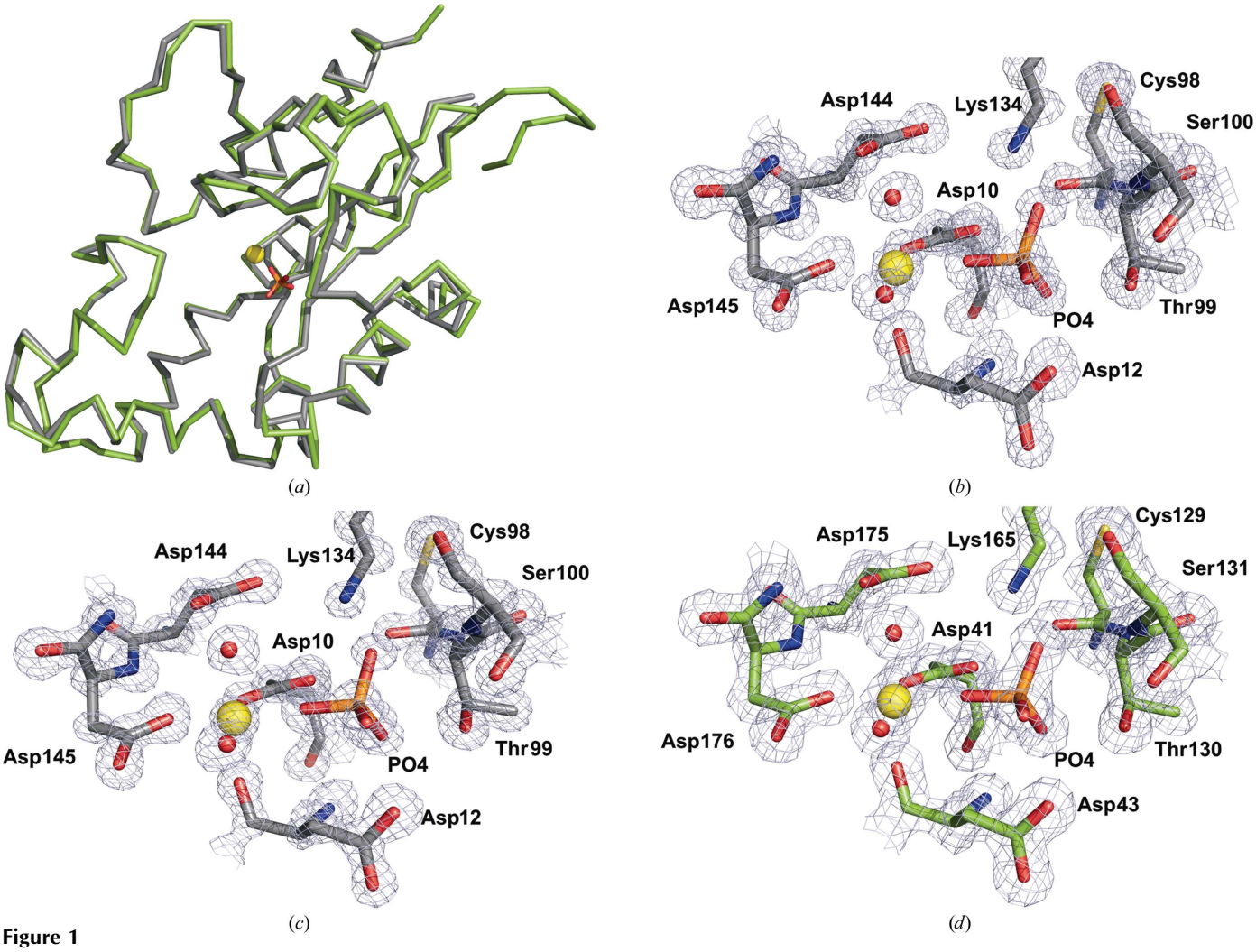


Figure 1
Structures of cdN and mdN. (*a*) The backbones of superimposed cdN and mdN structures are shown in grey and green, respectively. Magnesium ion (yellow sphere) and phosphate anion (in stick representation) are shown in the enzyme active site. (*b–d*) Detailed views of the active site with $2F_o - F_c$ electron-density maps contoured at 2σ are shown for cdN molecule *A* (*b*), cdN molecule *B* (*c*) and mdN (*c*).

Table 2
Results of docking the BTP-T inhibitor to mdN and cdN.

(*a*) Complexes of mdN and BTP-T.

Cluster	Binding energy (kcal mol^{-1})	Binding constant (n*M*)
1	9.70	77.5
2	8.34	771.4
3	8.30	828.4
4	8.08	1280.0
5	6.00	39700.0
6	0.90	218000000.0

(*b*) Complexes of cdN and BTP-T.

Cluster	Binding energy (kcal mol^{-1})	Binding constant (n*M*)
1	10.38	24.5
2	8.91	295.2
3	7.78	1990.0
4	7.21	5220.0
5	6.72	11800.0
6	5.85	51500.0
7	5.34	122000.0
8	4.00	1170000.0
9	3.86	1480000.0
10	3.83	1560000.0
11	3.71	1900000.0
12	3.59	2330000.0
13	3.17	4710000.0
14	3.15	4950000.0

electron-density map and thus were not included in the final model. The total r.m.s.d. for the superposition of 1251 atoms of the two molecules of cdN in the asymmetric unit is 0.683 Å. This value is below the values typically observed for identical structures (Betts & Sternberg, 1999). Most atoms were refined with anisotropic atomic displacement parameters (ADPs). However, spatially overlapped atoms in regions where alternative conformations were modelled and O atoms of water molecules with an unrealistic ratio of ellipsoid axes (*i.e.* a ratio of any two axes higher than 3:1) were refined with isotropic ADPs.

Like mdN, cdN possesses a fold characteristic of the haloacid dehalogenase (HAD) enzyme superfamily (Fig. 1*a*), in which the active site is located in the cleft between the core (α/β Rossman fold) and the cap (short four-helix bundle) domains. Catalytic residues are located in the core domain and the residues that determine substrate specificity are located in the cap domain.

The active sites of both of the cdN molecules present in the asymmetric unit contained an octahedrally coordinated Mg^{2+} cation and a PO_4^{3-} anion originating from the crystallization solution. A detailed view of the active site with a $2F_o - F_c$ map is shown in Fig. 1(*b*).

Our cdN structure is the second structure of this enzyme to be reported to date. The first cdN structure was determined at 1.2 Å resolution and contains 2′-deoxyuridine and a tetrafluoroaluminate anion in the active site (PDB entry 2id7; Walldén *et al.*, 2007). Both structures were determined in the same space group. The r.m.s.d. for superposition of C^{α} atoms of the two structures is 0.31 Å, which is below the values typically observed for identical structures (Sternberg & Betts, 1999). Nevertheless, our atomic resolution cdN structure without any substrate analogue in the active site provides additional structural information on the structure of the active site, including the protonation states of individual residues (as described in §3.3).

The structure of human mdN was refined to 1.4 Å resolution. The final crystallographic model consists of 201 amino-acid residues (residues 27–227) in one polyprotein chain. The residue numbering is based on the sequence of the native protein. Amino-acid residues 1–31 belong to the signal peptide, which is cleaved off during transport to the mitochondria. Amino-acid residues 27–31 (SNAAS) are a cloning artifact following TEV protease cleavage. The overall fold and location of the active site is very similar to that of cdN (Fig. 1*a*). The active site of mdN also contained an Mg^{2+} cation and a PO_4^{3-} anion originating from the crystallization solution. A detailed view of the active site with a $2F_o - F_c$ map is shown in Fig. 1(*c*).

Our mdN structure is isostructural to the structure deposited in the PDB with accession code 1mh9 (Rinaldo-Matthis *et al.*, 2002), which was determined at 1.8 Å resolution. This structure also contains an Mg^{2+} cation and a PO_4^{3-} anion in the enzyme active site. The r.m.s.d. for the superposition of 195 equivalent C^{α} atoms is 0.20 Å. However, there are some minor differences originating from the different resolution, crystallization conditions and cryoprotection procedure: (i) in our structure, the side chains of 13 amino-acid residues (Arg49, Asp57, Leu102, Lys110, Gln116, Lys128, Lys143, Glu151, Leu155, Arg157, Ser163, Asp173 and Leu196) acquire double conformations that were not observed in the lower resolution structure, (ii) the tip of the solvent-exposed loop (residues Pro200 and Pro201) was traced differently in the two models

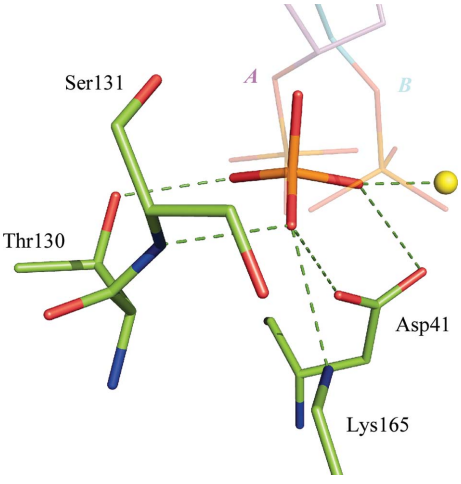


Figure 2
Comparison of the position of the phosphate group in mdN structures. (*a*) Superposition of two substrates bound to the D41N mdN mutant in positions denoted *A* (substrate containing 3′-phosphate is bound) and *B* (substrate containing 5′-phosphate is bound) with the phosphate ion present in our mdN structure (PDB entry 4l6a). The 2′-deoxythymidine 3′-monophosphate from PDB entry 1z4k is shown with azure C atoms; the 2′-deoxythymidine 5′-monophosphate from PDB entry 1z4l is shown in magenta. Polar interactions with active-site residues and Mg^{2+} (yellow sphere) are represented by dotted lines.

and (iii) three molecules of glycerol were found in our structure. The r.m.s.d.s of the active-site ions in the two mdN structures are 0.45 and 0.13 Å for Mg^{2+} and PO_4^{3-}, respectively.

3.2. Phosphate anion mimics the position of substrate 3′-phosphate in the active site

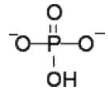
The structures of cdN and mdN reported here both contain a phosphate ion coordinated by catalytic residues and an Mg^{2+} ion in the active site. The phosphate ion originated from the crystallization solutions, as 20 and 40 m*M* KH_2PO_4 were used to prepare the crystals of mdN and cdN, respectively.

Both mdN and cdN recognize deoxyribonucleoside monophosphates with a 3′- or 5′-phosphate group as their physiological substrates. The structures of the catalytically inactive D41N mutant of mdN in complex with 3′-dTMP and 5′-dTMP revealed two different binding modes of the phosphate group in the enzyme active site (Fig. 2). In both positions (denoted *A* and *B* in Fig. 2), the O atoms of the phosphate ion coordinate Mg^{2+} and interact with the main-chain amide of Ser131 and with $O^{\delta2}$ of Asn41. In position *A*, additional interactions are provided by interaction with $O^{\gamma1}$ of Thr130, whereas in position *B* the phosphate ion interacts with N^{ζ} of Lys165. The phosphate group occupies position *A* when substrate with a 3′-phosphate is bound and position *B* when substrate with a 5′-phosphate is bound.

The position and interactions of the phosphate anion in our structure are similar to those of the phosphate group in position *A* (Fig. 2). The anion also interacts with N^{ζ} of Lys165, which changes position slightly towards the phosphate group. The r.m.s.d. for the superposition of phosphate ion and the 3′-phosphate group of dTMP is 0.52 Å, while the r.m.s.d. for the superposition of phosphate ion and the 5′-phosphate group of dTMP is 2.13 Å.

We therefore concluded that the phosphate anion bound in our cdN and mdN structures mimics the position of the substrate 3′-phosphate group and consequently the position of the reaction product after dephosphorylation of deoxyribonucleoside monophosphate.

Table 3
Bond lengths in phosphate-ion structures from the Cambridge Structural Database.



Bond	Number	Average length (Å)	Standard deviation (Å)
P—OH	111	1.567	0.066

3.3. Interaction of ions in the cdN active site

We determined the structure of cdN at a resolution close to 1 Å. Owing to this atomic level resolution, we were able to use the interatomic distances to assess the probable protonation states of the phosphate anion and amino-acid side chains in the enzyme active site. Fig. 3 shows a detailed view of the cdN active-site structure and a corresponding scheme with interatomic distances for ions and coordinating amino-acid residues. All distances are average values derived from the two molecules present in the asymmetric unit of the cdN structure.

The interatomic distances in amino acids coordinating the active-site ions were compared with standard distances for

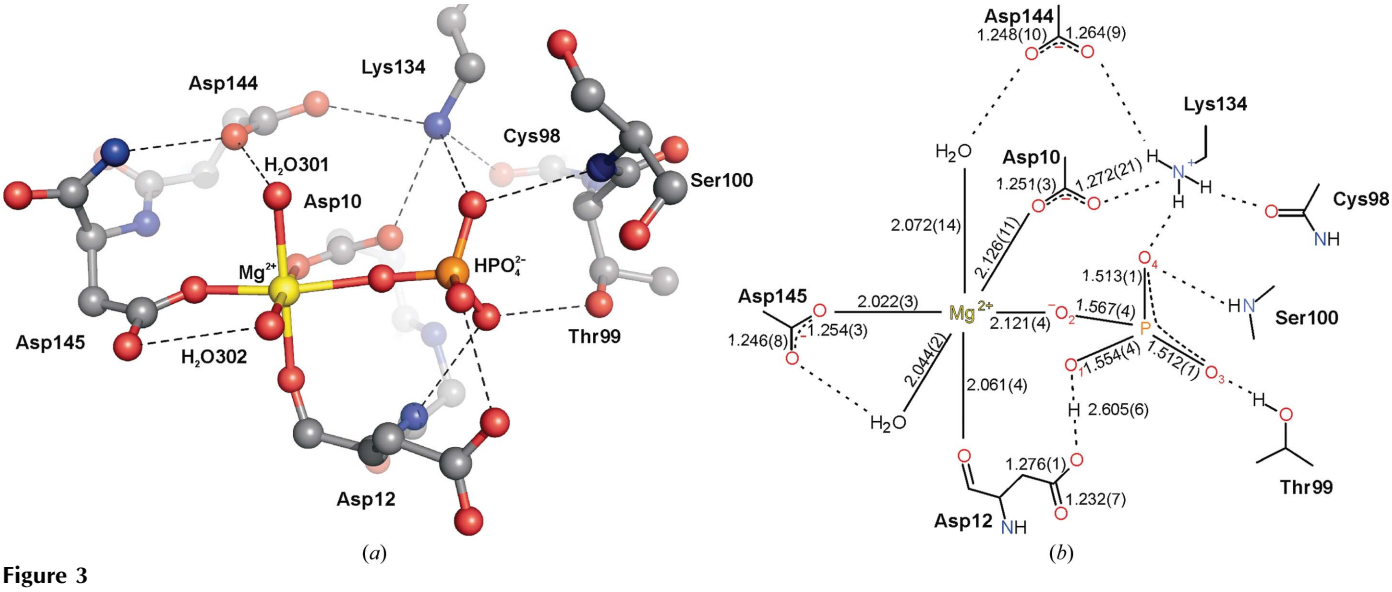


Figure 3
Polar interactions in the active site of the cdN structure determined at atomic resolution. (*a*) Detailed structure of the active site in ball-and-stick representation; hydrogen bonds are shown as dotted lines. (*b*) Schematic diagram with bonding distances for ion-coordination bonds and for carboxylic groups of amino acids in the active site. Distances are average values derived from both molecules in the asymmetric unit (values are in Å; standard deviations of distances are in parentheses). Color coding: carbon, grey in (*a*) and black in (*b*); oxygen, red; nitrogen, blue; magnesium ion, yellow; phosphorus, orange. The distance of 2.605 Å below the phosphate group indicates the hydrogen bond between O1 of the phosphate and the O atom of Asp12.

single and double bonds in carboxylic groups (Engh & Huber, 2001). In the protonated form of a typical carboxylic group, one O atom forms a double bond to carbon with a bond length of 1.20 Å and the second oxygen, carrying a hydrogen, forms a single bond with a bond length of 1.34 Å. In the deprotonated carboxylate, both O atoms are equivalent owing to the resonance effect and are equidistant from the C atom, with a bond length of 1.25 Å.

We deduced the protonation states of the aspartate side chains coordinating the catalytic site ions from the bond distances found in our atomic resolution cdN structure using a previously described method (Wlodawer *et al.*, 2001). Because the crystal was grown at pH 4.5, a value greater than the pK_a of the aspartate side-chain carboxyl group, we expected aspartates 10, 144 and 145 to be deprotonated. This assumption was indeed supported by C—O interatomic distances of close to 1.25 Å found in the crystal structure (Fig. 3*b*).

The negative charges of the deprotonated carboxylic groups are compensated by the Mg^{2+} ion and the protonated N^{ζ} of Lys134. The coordination sphere of Mg^{2+} is formed by $O^{\delta2}$ of

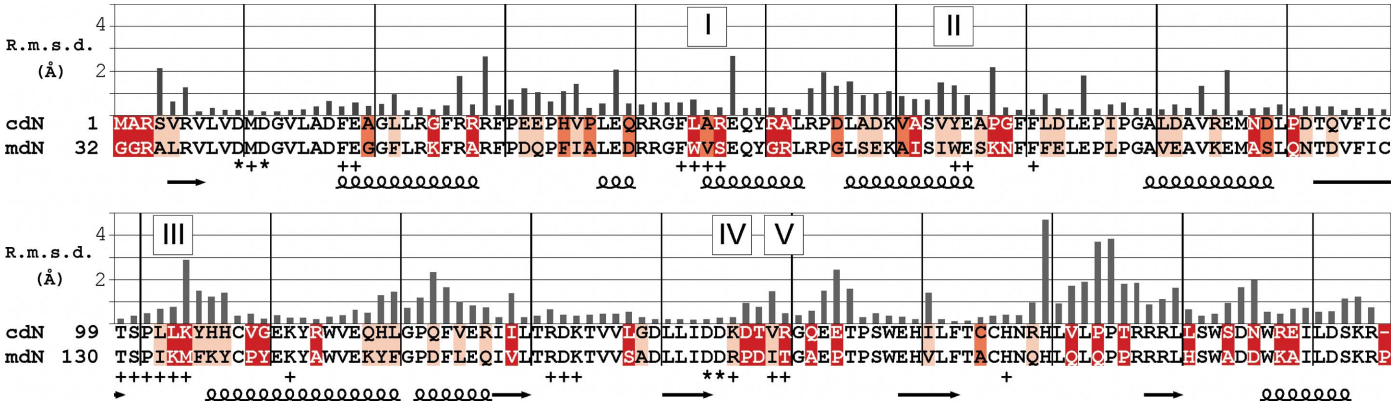


Figure 4
Comparison of the sequences and the three-dimensional structures of cdN and mdN. Histogram bars indicate the r.m.s.d. for the superposition of atoms in the corresponding amino-acid residues. Very similar amino acids are coloured light pink, similar residues are coloured pink and different residues are highlighted in red. Amino-acid residues forming the active site are indicated by crosses; stars indicate the four aspartates that are crucial for catalytic activity. Secondary-structure elements (arrows for β-sheets, spirals for α-helices) are shown below the sequence alignment. Regions with variable sequences incorporated in the surface of the active site are indicated by the Roman numerals I to V corresponding to Fig. 5. The difference in residue numbering stems from the presence of a 31-amino-acid mitochondrial transport leading sequence at the N-terminus of mdN.

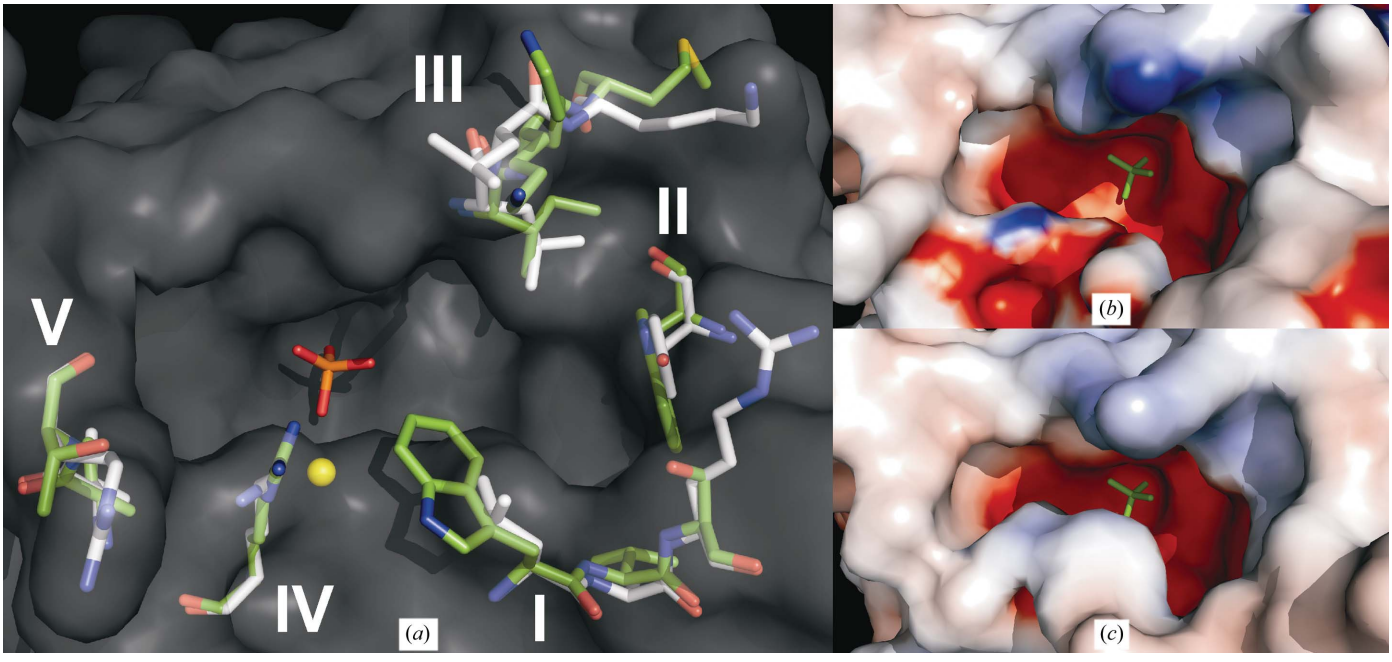


Figure 5
Structural comparison of the cdN and mdN active sites. (*a*) The solvent-accessible surface of the cdN active site is shown in grey. The Mg^{2+} ion is represented by a yellow sphere. The bound phosphate ion and five variable regions (I–V) are shown in stick representation, with C atoms coloured white and green for cdN and mdN, respectively. (*b*) Detail of the active site of cdN with the phosphate ion coloured green. (*c*) Detail of the active site of mdN with the phosphate ion coloured green. (*b*) and (*c*) show the solvent-accessible surface coloured according to its calculated electrostatic potential, where red represents $-5kT$/e and blue represents $+5kT$/e.

Asp10, $O^{\delta1}$ of Asp145, the carbonyl O atom of Asp12 and two water molecules (numbered 301 and 302). The O2 atom of the phosphate anion completes the hexacoordination of the magnesium cation.

To assign the protonation state of the phosphate anion, we analyzed the bond lengths in the crystal structures of organic molecules containing phosphate ions. A survey of the Cambridge Structural Database (Allen, 2002) using *Mogul* 1.3 (Bruno *et al.*, 2004) yielded a total of 111 structures containing a phosphate ion. The average values of the distances between the P and O atoms are summarized in Table 3.

Using these bonding distances, we determined that the phosphate in the cdN active site is a monoprotonated HPO_4^{2-} ion. The P—O1 bond, with a distance of 1.554 Å, is clearly a single bond with O1 protonated. The P—O3 (1.512 Å) and P—O4 (1.513 Å) bonding distances are the shortest and these two bonds have evident multiple bond character. The P—O2 bond distance (1.567 Å) is much longer than the average value for $P—O^-$ bonds. This bond elongation is likely to be caused by the engagement of O2 in a coordinate covalent bond with Mg^{2+}.

Based on our model of the cdN active site (Fig. 3*b*), we propose that the $C^{\gamma}—O^{\delta1}$ bond in the carboxyl group of Asp12 has double-bond character, whereas the $C^{\gamma}—O^{\delta2}$ bond has single-bond character. The $O^{\delta2}$ atom of Asp12 and the O1 atom of the phosphate anion are only 2.6 Å apart, indicating that they share an H atom and form a strong hydrogen-bonding interaction. The $O^{\delta1}$ atom of the deprotonated residue Asp10 is located on the opposite side of the phosphate anion in a position in which it can perform a nucleophilic attack on a substrate P atom.

The protonation states in the active site of our atomic resolution cdN structure suggest that cdN follows the catalytic mechanism previously proposed for mdN (Rinaldo-Matthis *et al.*, 2002). We propose that the side chain of Asp10 of cdN performs a nucleophilic attack on the P atom of the substrate phosphate group coordinating Mg^{2+} and the positively charged Lys143. This gives rise to a hypothetical pentavalent phospho-enzyme intermediate, which is immediately hydrolyzed by a water molecule activated by Asp12 (Rinaldo-Matthis *et al.*, 2002; Himo *et al.*, 2005).

3.4. Structural implications for the design of specific cdN inhibitors

Inhibitors of cdN are of therapeutic potential because they could prolong the lifetime of phosphorylated nucleoside analogues used as antiviral and anticancer agents. Such inhibitors must not block mdN activity, which is important for regulation of the mitochondrial triphosphate pool (Mazzon *et al.*, 2003).

The cdN and mdN proteins share 61% sequence identity and 77% sequence similarity, as determined by *BLASTp* (Altschul *et al.*, 2005). This high sequence similarity results in very similar three-dimensional structures (Fig. 4). The r.m.s.d. for the superposition of 192 corresponding C^{α} atoms is 1.87 Å. The sequence and structural similarity for the active-site residues is higher than the average similarity (Fig. 4).

A detailed sequence comparison of cdN and mdN revealed five regions with variable sequences close to the enzyme active site. These regions (labelled in Fig. 4) are (I) the amino-acid sequence LAR 45–47 in cdN and WVS 76–78 in mdN, (II) Tyr65 in cdN and Trp96 in mdN, (III) LLK 102–104 in cdN and IKM 133–135 in mdN, (IV) Lys146 in cdN and Arg177 in mdN and (V) Arg150 in cdN and Thr181 in mdN.

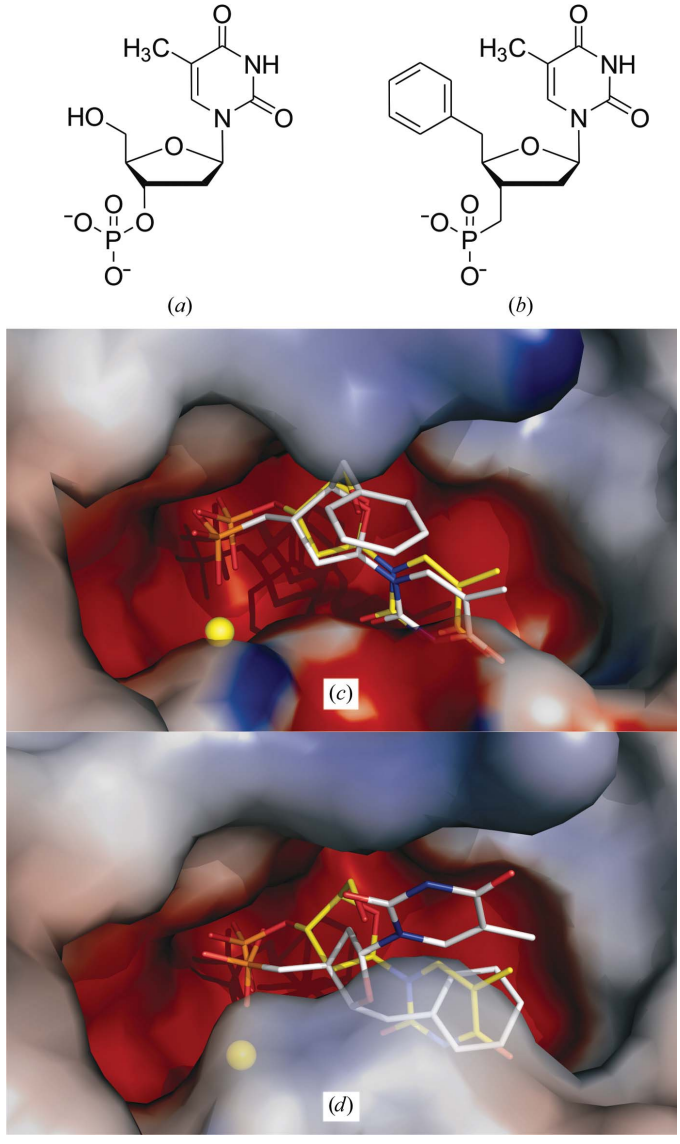


Figure 6
The results of molecular docking. (*a*, *b*) The chemical structures of 2′-deoxythymidine 3′-monophosphate (3′-dTMP) (*a*) and [5-benzyl-2-(thymin-1-yl)-tetrahydrofuran-4-yl]methylphosphonic acid (BTP-T) (*b*). (*c*) Superposition of BTP-T docked into the cdN crystal structure (C atoms in white) with 3′-dTMP from the crystal structure of D41N mdN (C atoms in yellow; PDB entry 1z4k; Walldén *et al.*, 2005). (*d*) Superposition of BTP-T docked into the mdN crystal structure (C atoms in white) with 3′-dTMP from the crystal structure of D41N mdN (C atoms in yellow; PDB entry 1z4k; Walldén *et al.*, 2005). The protein is represented by its solvent-accessible surface coloured according to calculated electrostatic potential, where red represents $-5kT$/e and blue represents $+5kT$/e. The Mg^{2+} ion is shown as a yellow sphere.

These regions contribute to the structural variability of the cdN and mdN active sites (Fig. 5). By comparing the cdN and mdN structures with phosphate ion in the active site, we can decipher the structural impact of the variable regions I–V (Fig. 5*a*). When the shapes of the active sites of cdN and mdN are compared, the effects of the variability in regions I and IV are the most profound. The substitution of Leu45 and Lys146 in cdN by the bulkier amino-acid residues Trp76 and Arg177 in mdN results in a more closed substrate pocket in mdN than in cdN (Figs. 5*b* and 5*c*). The difference in volume of the active sites of mdN and cdN is almost 50 Å^3. The active-site solvent-accessible volume calculated by the *CASTp* server (Dundas *et al.*, 2006) is 217 Å^3 for mdN and 270 and 256 Å^3 for cdN chains *A* and *B*, respectively.

The entry to the mdN active site is tighter, which might explain why the K_m values for hydrolysis of dTMP and dUMP by mdN [K_M(dTMP) = 0.2 m*M*, K_M(dUMP) = 0.1 m*M*] are approximately one order of magnitude lower than those for cdN-catalyzed hydrolysis [K_M(dTMP) = 1.5 m*M*, K_M(dUMP) = 1.5 m*M*] (Mazzon *et al.*, 2003). The differences in the shapes of the active sites could be exploited to design specific inhibitors for cdN or mdN.

To test this hypothesis *in silico*, we designed the molecule [5-benzyl-2-(thymin-1-yl)-tetrahydrofuran-4-yl]methylphosphonic acid (BTP-T), the structure of which was derived from the structure of the substrate 2′-deoxythymidine 3′-monophosphate (3′-dTMP) (Fig. 6*a*). The 3′-dTMP phosphate group was replaced by the uncleavable methylenephosphonate analogue and the 5′-hydroxyl group was replaced by a benzene ring (Fig. 6*a*). The replacement of the 5′-hydroxyl group by a benzene ring was chosen to explore the differences in the fit of a bulky substituent into the differently shaped active sites of cdN and mdN.

BTP-T was docked into the cdN and mdN active sites. The results of the docking analysis (see Table 2) showed that BTP-T fits well into the active site of cdN, with the phosphonate group mimicking the position of the substrate phosphate group (Fig. 6*c*). The most favoured pose of BTP-T in the cdN active site resembles the position of 3′-dTMP in the D41N variant of human mdN (PDB entry 1z4k; Walldén *et al.*, 2005), with an r.m.s.d. for 19 identical atoms of 0.96 Å. Placement of BTP-T into the mdN active site is only possible after rearrangement of free torsions in the molecule (compare Figs. 6*c* and 6*d*). In the most favoured pose of BTP-T in the mdN active site, the phosphonate group is nearly in the position of the substrate phosphate group; however, the entire molecule is rotated almost 180° relative to the substrate, with the positions of the phenyl ring and the sugar base switched (Fig. 6*d*).

Analysis of all of the docking results for mdN did not reveal a BTP-T pose resembling the position of 3′-dTMP in the mdN active site. When we compare the estimated binding constants (Table 2) we have to consider that the effect of selectivity is more pronounced owing to the different catalytic efficiencies of the two enzymes (the K_m value for cleavage of dUMP is 0.1 and 1.5 m*M* for mdN and cdN, respectively; Mazzon *et al.*, 2003). The inhibitor affinity is better than that of the substrate by a factor of 1300 for mdN, and for cdN this factor is 60 000. Based on the results of our docking analysis, we thus conclude that a benzyl modification at the 5′-position of 3′-dTMP might be the key modification that enables selective binding to cdN over mdN. To use phosphonates as cdN inhibitors *in vivo*, prodrug strategies will have to be employed to achieve oral bioavailability and intracellular delivery (Hecker & Erion, 2008; Hostetler *et al.*, 1990; Périgaud *et al.*, 1993).

This work was supported by the Grant Agency of the Czech Republic (research project No. GA 203/09/0820) and by research projects RVO68378050 and RVO61388963 awarded by the Academy of Sciences of the Czech Republic. Diffraction data were collected on beamlines 14.1 and 14.2 operated by the Helmholtz-Zentrum Berlin (HZB) at the BESSY II electron-storage ring (Berlin-Adlershof, Germany). The authors wish to thank P. Bartunek from the Institute of Molecular Genetics AS CR Prague for the generous gift of the cDNA library from human breast carcinoma cell line MDA-MB.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. A. & Yu, Y.-K. (2005). *FEBS J.* **272**, 5101–5109.
Betts, M. J. & Sternberg, M. J. E. (1999). *Protein Eng.* **12**, 271–283.
Bianchi, V., Pontis, E. & Reichard, P. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 986–990.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Bruno, I. J., Cole, J. C., Kessler, M., Luo, J., Motherwell, W. D. S., Purkis, L. H., Smith, B. R., Taylor, R., Cooper, R. I., Harris, S. E. & Orpen, A. G. (2004). *J. Chem. Inf. Comput. Sci.* **44**, 2133–2144.
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.* D**66**, 12–21.
D'Angelo, E., Crutchfield, J. & Vandiviere, M. (2001). *J. Environ. Qual.* **30**, 2206–2209.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. & Kollman, P. (2003). *J. Comput. Chem.* **24**, 1999–2012.
Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. & Liang, J. (2006). *Nucleic Acids Res.* **34**, W116–W118.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Hecker, S. J. & Erion, M. D. (2008). *J. Med. Chem.* **51**, 2328–2345.
Himo, F., Guo, J. D., Rinaldo-Matthis, A. & Nordlund, P. (2005). *J. Phys. Chem. B*, **109**, 20004–20008.
Hostetler, K. Y., Stuhmiller, L. M., Lenting, H. B. M., van den Bosch, H. & Richman, D. D. (1990). *J. Biol. Chem.* **265**, 6112–6117.
Hunsucker, S. A., Mitchell, B. S. & Spychala, J. (2005). *Pharmacol. Ther.* **107**, 1–30.
Jakalian, A., Jack, D. B. & Bayly, C. I. (2002). *J. Comput. Chem.* **23**, 1623–1641.
Krieger, E., Joo, K., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D. & Karplus, K. (2009). *Proteins*, **77** Suppl. 9, 114–122.
Leslie, A. G. W. (1999). *Acta Cryst.* D**55**, 1696–1702.
Mazzon, C., Rampazzo, C., Scaini, M. C., Gallinaro, L., Karlsson, A., Meier, C., Balzarini, J., Reichard, P. & Bianchi, V. (2003). *Biochem. Pharmacol.* **66**, 471–479.
Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). *Acta Cryst.* D**62**, 859–866.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. & Olson, A. J. (2009). *J. Comput. Chem.* **30**, 2785–2791.
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.* D**67**, 355–367.
Périgaud, C., Gosselin, G., Lefebvre, I., Girardet, J.-L, Benzaria, S., Barber, I. & Imbach, J.-L. (1993). *Bioorg. Med. Chem. Lett.* **3**, 2521–2526.
Rampazzo, C., Johansson, M., Gallinaro, L., Ferraro, P., Hellman, U., Karlsson, A., Reichard, P. & Bianchi, V. (2000). *J. Biol. Chem.* **25**, 5409–5415.
Rinaldo-Matthis, A., Rampazzo, C., Balzarini, J., Reichard, P., Bianchi, V. & Nordlund, P. (2004). *Mol. Pharmacol.* **65**, 860–867.
Rinaldo-Matthis, A., Rampazzo, C., Reichard, P., Bianchi, V. & Nordlund, P. (2002). *Nature Struct. Biol.* **9**, 779–787.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sternberg, M. J. E. & Betts, M. J. (1999). *Protein Eng.* **12**, 271–283.
Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F. R. & Donnelly, M. I. (2002). *Protein Expr. Purif.* **25**, 8–15.
Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.
Walldén, K., Rinaldo-Matthis, A., Ruzzenente, B., Rampazzo, C., Bianchi, V. & Nordlund, P. (2007). *Biochemistry*, **46**, 13809–13818.
Walldén, K., Ruzzenente, B., Rinaldo-Matthis, A., Bianchi, V. & Nordlund, P. (2005). *Structure*, **13**, 1081–1088.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.
Wlodawer, A., Li, M., Gustchina, A., Dauter, Z., Uchida, K., Oyama, H., Goldfarb, N. E., Dunn, B. M. & Oda, K. (2001). *Biochemistry (Moscow)*, **40**, 15602–15611.