Crystal Structure of Yeast Thymidylate Kinase Complexed with the Bisubstrate Inhibitor P^1-(5′-Adenosyl) P^5-(5′-Thymidyl) Pentaphosphate (TP_5A) at 2.0 Å Resolution: Implications for Catalysis and AZT Activation[†,‡]

Arnon Lavie,[§] Manfred Konrad,[∥] Ralf Brundiers,[∥] Roger S. Goody,[§] Ilme Schlichting,[§] and Jochen Reinstein*[,§]

Department of Physical Biochemistry, Max Planck Institute for Molecular Physiology, Rheinlanddamm 201, 44139 Dortmund, Germany, and Department of Molecular Genetics, Max Planck Institute for Biophysical Chemistry, 37018 Göttingen, Germany

Received August 21, 1997; Revised Manuscript Received November 24, 1997

ABSTRACT: The crystal structure of yeast thymidylate kinase (TmpK) complexed with the bisubstrate inhibitor P^1-(5′-adenosyl) P^5-(5′-thymidyl) pentaphosphate (TP_5A) was determined at 2.0 Å resolution. In this complex, TmpK adopts a closed conformation with a region (LID) of the protein closing upon the substrate and forming a helix. The interactions of TmpK and TP_5A strongly suggest that arginine 15, which is located in the phosphate binding loop (P-loop) sequence, plays a catalytic role by interacting with an oxygen atom of the transferred phosphoryl group. Unlike other nucleoside monophosphate kinases where basic residues from the LID region participate in stabilizing the transition state, TmpK lacks such residues in the LID region. We attribute this function to Arg15 of the P-loop. TmpK plays an important role in the phosphorylation of the AIDS prodrug AZT. The structures of TmpK with dTMP and with AZT-MP [Lavie, A., et al. (1997) *Nat. Struct. Biol. 4*, 601−604] implicate the movement of Arg15 in response to AZT-MP binding as an important factor in the 200-fold reduced catalytic rate with AZT-MP. TmpK from *Escherichia coli* lacks this arginine in its P-loop while having basic residues in the LID region. This suggested that, if such a P-loop movement were to occur in the *E. coli* TmpK upon AZT-MP binding, it should not have such a detrimental effect on catalysis. This hypothesis was tested, and as postulated, *E. coli* TmpK phosphorylates AZT-MP only 2.5 times slower than dTMP.

Thymidylate kinase (EC 2.7.4.9, ATP:dTMP phosphotransferase) catalyzes the phosphorylation of thymidine monophosphate (dTMP) to thymidine diphosphate (dTDP) utilizing ATP as its preferred phosphoryl donor (*1*) according to the scheme

$$\text{dTMP} + \text{ATP}\cdot\text{Mg} \Leftrightarrow \text{dTDP} + \text{ATP}\cdot\text{Mg}$$

Its location at the junction of the *de novo* and salvage pathways for thymidine triphosphate (dTTP) synthesis makes thymidylate kinase (TmpK)[1] an essential enzyme for cell proliferation, and thus an attractive target for the development of drugs against cancer. In addition to its physiological role, TmpK is also involved in the activation of the AIDS drug 3′-azido-3′-deoxythymidine (AZT). AZT is a prodrug that must be phosphorylated three times to its triphosphate form (AZT-TP), since it is AZT-TP that inhibits viral replication by DNA chain termination. TmpK, which catalyzes the second phosphorylation step, from the monophosphate (AZT-MP) to the diphosphate (AZT-DP), has been shown to be the rate-limiting enzyme in the AZT activation pathway (*2*, *3*). This results in a toxic accumulation of AZT-MP to millimolar concentrations in cells exposed to AZT (*4*−*6*), and in a low concentration of the active compound AZT-TP. Slow activation rates of prodrugs have been implicated in allowing the replicating virus to select for resistant mutants.

Understanding the mechanism of phosphoryl transfer and pinpointing the amino acid residues involved could play an important role in attempts to develop or to improve strategies for the treatment of cancer and AIDS. First, the design of a mechanism-based inhibitor of TmpK resulting in the halt of dTTP synthesis, and thus cell proliferation, could play a role in chemotherapy of cancers. Second, revealing the reasons for the poor activation of AZT-MP by TmpK is the first step in the attempt to solve this problem. Therefore, we have solved the three-dimensional structure of TmpK complexed with the bisubstrate inhibitor P^1-(5′-adenosyl) P^5-(5′-thymidyl) pentaphosphate (TP_5A).

This structure, taken together with those of TmpK complexed with dTMP and with AZT-MP (*7*), allows us to identify the residues important for catalysis, and to propose an explanation for the slow phosphorylation of AZT-MP. In addition, on the basis of that knowledge, it allowed us to

[†] A.L. was supported by a fellowship from the Human Frontier Science Program.

[‡] The coordinates were deposited in the Brookhaven Protein Data Bank under accession code 3TMK.

[§] Max Planck Institute for Molecular Physiology.

[∥] Max Planck Institute for Biophysical Chemistry.

[1] Abbreviations: TmpK, thymidylate kinase; UmpK, uridylate kinase; AK, adenylate kinase; HSV1-TK, herpes simplex virus thymidine kinase, type 1; AZT, 3′-deoxy-3′-azidothymidine; AZT-MP, 3′-deoxy-3′-azidothymidine monophosphate; AZT-DP, 3′-deoxy-3′-azidothymidine diphosphate; AZT-TP, 3′-deoxy-3′-azidothymidine triphosphate; TP_5A, P^1-(5′-adenosyl) P^5-(5′-thymidyl) pentaphosphate; UP_5A, P^1-(5′-adenosyl) P^5-(5′-uridyl) pentaphosphate; PA, PB, PC, PD, and PE, phosphate groups of TP_5A; NMP, nucleoside monophosphate; ncs, noncrystallographic symmetry; rmsd, root-mean-square deviation.

S0006-2960(97)02078-3 CCC: $15.00
Published on Web 02/27/1998

make the prediction that the rate of AZT-MP phosphorylation by the *Escherichia coli* TmpK would be similar to the phosphorylation rate of the physiological substrate dTMP. In fact, *E. coli* TmpK discriminates poorly against AZT-MP, supporting our hypothesis that Arg15 of yeast and human thymidylate kinases is responsible for discrimination against AZT-MP as a substrate (the human thymidylate kinase has 44% amino acid identity and >63% similarity with the yeast enzyme with all catalytically important residues conserved between the two species).

MATERIALS AND METHODS

Crystallization and Data Collection. Crystals of the complex between the bisubstrate inhibitor TP_5A with yeast thymidylate kinase were obtained by the hanging drop method. A 12 mg/mL enzyme solution was premixed with a TP_5A solution to a final concentration of 10 mg/mL enzyme and 2 mM TP_5A. Equal volumes of the protein−nucleotide solution and a solution composed of 20% PEG 2000 monomethyl ether, 100 mM sodium acetate (pH 4.6), and 200 mM ammonium sulfate were mixed, and left to equilibrate at room temperature against a reservoir composed of the latter solution. Crystals with typical dimensions (micrometers) of 400 × 200 × 100 grew within days in space groups $P1$, $P2_1$, or $P2_12_12$ with very similar unit cell dimensions. This space group polymorphism made these crystals unsuitable for the initial structure determination (*7*). The structure reported here was solved by molecular replacement using the dTMP−TmpK complex structure as the starting model (see below). A data set was collected at 100 K from a crystal soaked shortly in mother liquor with 25% glycerol as a cryoprotectant, using a Siemens multiwire area detector mounted on a Mac Science rotating anode operating at 45 kV and 100 mA. Processing of the data was carried out with XDS (*8*). The structure reported here is from a crystal which reduced in space group $P2_1$ and diffracted to 2.0 Å resolution (see Table 1).

Structure Determination. The TP_5A complex structure (Figure 1) was solved by the molecular replacement method with the partially refined dimeric dTMP complex structure as the starting search model. As mentioned above, the TP_5A cocrystals exhibit space group polymorphism; the data set used has $P2_1$ symmetry with four dimers in the asymmetric unit. The pseudo-symmetry with the orthorhombic space group suggested the presence of two tetramers in the monoclinic asymmetric unit. This implied that, when searching with the TmpK dimer, one should expect two rotation function solutions, where each solution has two different translation function solutions, thus yielding the two tetramers. Both AMoRe (*9*) and X-PLOR (*10*) yielded the two expected rotation function solutions, but it was not possible to find the two appropriate translation function solutions. The self-Patterson map showed one clear peak that was interpreted as the vector relating the two tetramers. Thus, each of the two rotation function results was first applied on the search dimer, and then the translation vector from the self-Patterson was applied on the rotated dimers. The two resulting pairs of tetramers were used independently as search models for the translation function and yielded clear solutions. In $P2_1$, the origin of the unique axis is not defined so the relative position of one tetramer to the other was found by arbitrarily keeping one tetramer at $y = 0$, which also limits the translational search to two dimensions (x and z). At first, the refinement was carried out using strict noncrystallographic symmetry (ncs), thus allowing us to work with a dimer instead of the octamer. The refinement proceeded smoothly with clear electron density for the TP_5A bisubstrate inhibitor. When the R-factor reached 27%, the ncs constraints were totally relaxed. The R-factor converged at 20 and 29% for the work and test set (5% of total reflections), respectively. Ncs restraints were reapplied at relatively low weight, causing a slight decrease in R_{free} and a bigger increase in R_{work}. Consequently, the final model was refined with ncs restraints grouping monomers 1, 3, 5, and 7 in one group and 2, 4, 6, and 8 in the other (the LID sequence and nucleotides were not included). The numbering used is residues 1−216 and 501−716 for monomers 1 and 2, respectively, with TP_5A numbered 217 in monomer 1 and 717 in monomer 2. Monomers 3 and 4 are numbered as above with the addition of 1000, 5 and 6 with the addition of 2000, and 7 and 8 with the addition of 3000 (thus, TP_5A 3217 is the TP_5A bound to monomer 7). Model building was done with the program O (*11*). The present model consists of 8 monomers with some residues omitted because of poor electron density (residues 1, 2, 1137−1148, 2001, 3001, 3501, and 3502) and some residues modeled as alanines, 8 TP_5A molecules, and 1198 water molecules (see Table 1). A simulated annealing omit map (calculated without the TP_5A and Arg15) with the current TP_5A model superimposed is shown in Figure 2a.

Table 1: Data Collection and Refinement Statistics

Data Collection	
temperature (K)	100
resolution range (Å)	43.8−2.0
observed reflections	270 178
unique reflections	116 451
completeness (%, overall/last shell)	82.0/56.9
R_{sym}[a] (%, overall/last shell)	5.9/26.6
space group	$P2_1$
unit cell dimensions	$a = 72.6$ Å, $b = 87.3$ Å, $c = 155.0$ Å, $\beta = 90.1°$
molecules/asymmetric unit	8
Refinement Statistics	
resolution range (Å)	43.8−2.0
R_{factor}[b]/R_{free} (%)	20.9/27.9
rms deviations	
bond lengths (Å)	0.012
bond angles (Å)	1.64
dihedral angles (deg)	24.7
improper angles (deg)	1.34
reflections with $F > 0\sigma$ (working/test)	101 931/5401
no. of protein atoms	13 791
no. of nucleotide atoms	440
no. of water molecules	1198
average B-factor (Å^2)	
main chain	20.5
side chain	22.1
water	28.6
nucleotide	16.3

[a] $R_{sym} = \sum|I-\langle I\rangle|/I$. [b] R-factor $= \sum||F_{obs}|-|F_{calc}||/\sum|F_{obs}|$; 5% of the reflections were used for R_{free}.

Steady-State Kinetics. The catalytic activity of TmpK was measured with a coupled colorimetric assay essentially as described (*12*) with the following assay buffer: 100 mM Tris/HCl (pH 7.5), 200 μM NADH, 400 μM phosphoenolpyruvate, 80 mM KCl, and 1.4 mM $MgCl_2$ at 25 °C. In

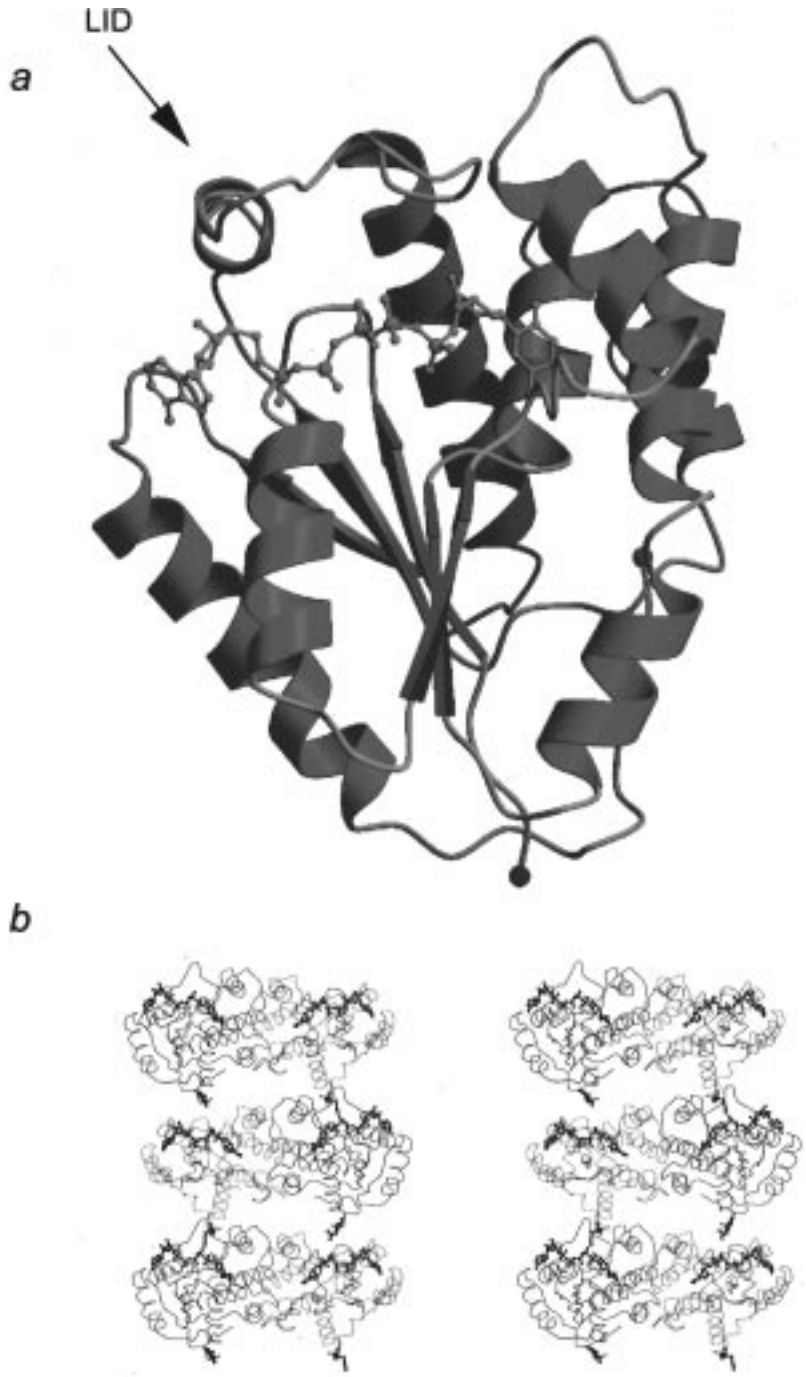

FIGURE 1: TP_5A−TmpK complex structure. (a) A ribbon diagram of a monomer (TmpK is a homodimer) with the helixes drawn in red and strands in green. The TP_5A bound at the active site is in cyan with the five phosphorus atoms in purple. The LID region adopts a helical conformation in the TP_5A-bound complex in contrast to a coil with no secondary structure when only dTMP is present. (b) Stereoview of the TP_5A−TmpK crystal packing (not shown is the asymmetric unit, which would be perpendicular to the page). The monomers can be designated active (green) or inactive (red) according to the presence of a symmetry-related arginine (depicted as ball and stick) in the active site. The TP_5A molecules are shown in cyan. Note that the arginines in green penetrate the active site of the red monomers (unlike the red-colored arginines that are further away) and thus hinder the formation of an active conformation by those monomers (see more in the text).

addition, the assay contained different concentrations of ATP and dTMP and 10−50 nM TmpK as indicated in Figure 3. The data were analyzed with the Michaelis−Menten equation and the nonlinear regression program Grafit (Erithacus Software).

Fluorescence Measurements. Equilibrium fluorescence measurements were performed as described (*13*) using a SLM 8100 spectrofluorimeter with an excitation wavelength of 360 nm and an emission wavelength of 440 nm. The experiments were carried out at pH 7.5 in a solution containing 50 mM Tris/HCl, 5 mM $MgCl_2$, 2 mM EDTA, and 100 mM KCl with the temperature of the cuvette being held constant at 25 °C. The fluorescent *N*-methylanthraniloyl group (MANT) joined to the ribose of the adenosine moiety of the bisubstrate inhibitor TP_5A (TP_5A-MANT) was used as a probe (*13*).

Time-resolved binding studies were performed with the same buffer as described above with a Hi-Tech Scientific

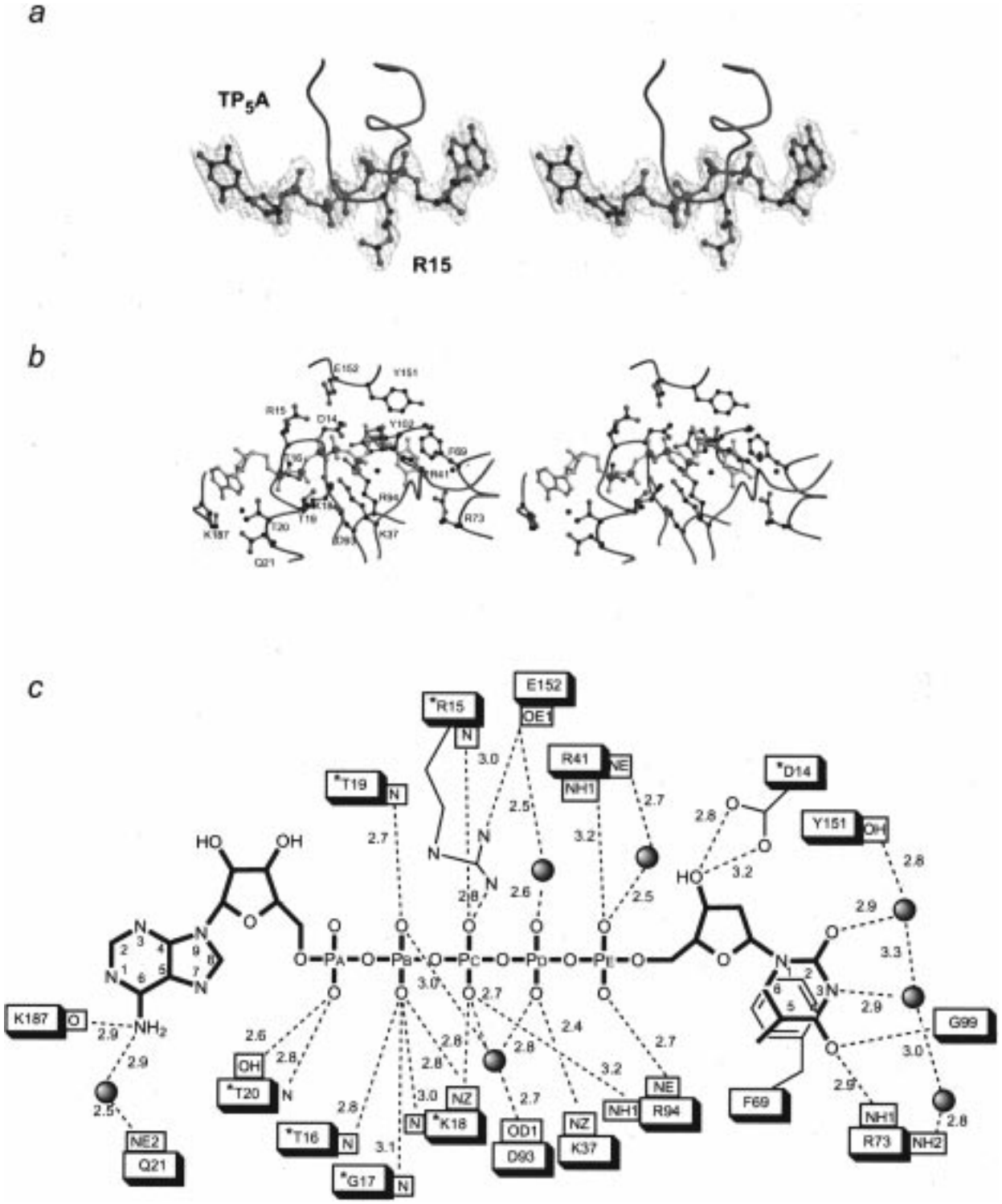

FIGURE 2: (a) Stereoview of a simulated annealing omit map of TP_5A and arginine 15 from an active monomer contoured at 3σ. Arginine 15, which is located in the P-loop (shown as a coil running from left to right), interacts with the middle phosphate of TP_5A. (b) Stereoview of the active site. The TP_5A nucleotide is drawn in yellow except for the five phosphorus atoms that are in purple. (c) A distance map with the same view as above. P-Loop residues are additionally marked with an asterisk.

SF61 stopped flow apparatus. As a signal for complex formation, either the extrinsic signal of TP_5A-MANT with excitation at 360 nm and a cutoff at 420 nm or the intrinsic tryptophan signal with excitation at 295 nm and a cutoff at 320 nm was used. These experiments were performed essentially as described (*14*).

Cloning and Purification of E. coli TmpK. The tmk protein (*cdc8* gene) was overproduced in *E. coli* by applying the phage T7 RNA polymerase-based expression system (*15*). Using standard PCR techniques with Taq polymerase (Boehringer, Mannheim, Germany), the oligonucleotides 5′-GGA ATT CCA TAT GCG CAG TAA GTA TAT CGT C-3′ and 5′-CGC GGA TCC TCA TGC GTC CAA CTC CTT CAC CCA G-3′ were used to introduce an *Nde*I restriction site at the initiating methionine and a *Bam*HI site behind the authentic Stop codon, respectively, of the tmk-encoding sequence (*16*). The 0.65 kb PCR product was cut with *Nde*I and *Bam*HI, ligated into the vector pJC20, and transformed

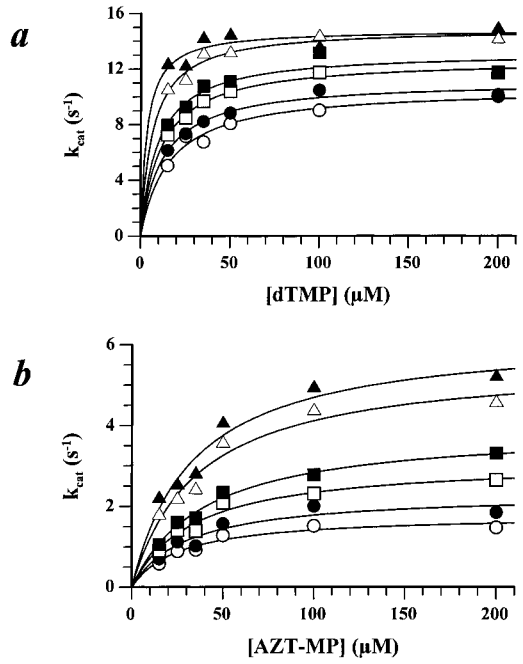

FIGURE 3: Steady-state kinetics of dTMP versus AZT-MP with *E. coli* TmpK. (a) The catalytic activity of *E. coli* TmpK (16.4 nM) with dTMP and ATP was measured as described in Materials and Methods at the dTMP concentrations indicated on the *x*-axis and the following ATP concentrations: 15 μM (○), 25 μM (●), 50 μM (□), 100 μM (■), 300 μM (△), and 1000 μM (▲). The maximal catalytic activity attained (k_{cat}) was 15 s^{-1} with a K_M for dTMP of 2.7 μM (see also Results) with ATP and dTMP binding with a factor for synergism of 6. (b) The catalytic activity of *E. coli* TmpK (15.1 nM) with AZT-MP and ATP. The ATP concentrations were as indicated in panel a. The turnover number (k_{cat}) reached a maximum of 6 s^{-1} at saturating concentrations of both substrates. The binding of dTMP and ATP is synergistic by a factor of 6. The apparent K_M of dTMP is lowered from 17 μM (extrapolated to no ATP) to 2.7 μM at saturating ATP concentrations, and likewise, the apparent K_M value of ATP is lowered from approximately 50 to 8 μM. This synergism of binding is completely absent with ATP and AZT-MP, the K_M value of ATP being around 50 μM independent of the AZT-MP concentration and that of AZT-MP (30 μM) being very close to that of dTMP (17 μM) in the absence of ATP. Therefore, we predict that the 3′-OH group of dTMP may be directly involved in the synergism of ATP and dTMP binding.

into the *E. coli* strain BL21(DE3). *E. coli* TmpK is strongly expressed, comprising more than 50% of the total protein as judged by SDS−PAGE. After ion-exchange chromatography (Q-Sepharose) and gel filtration (S-200), the enzyme was more than 95% pure.

RESULTS AND DISCUSSION

Overall Structure. TmpK from *Saccharomyces cerevisiae* (*cdc8* gene) has 216 amino acid residues and a molecular mass of 25 kDa (see Figure 1). Unlike the known structures of other nucleoside monophosphate (NMP) kinases (adenylate kinase, uridylate kinase, and guanylate kinase) which are all monomeric, TmpK is a homodimer. Despite no amino acid sequence similarity between TmpK and any of the other NMP kinases, TmpK assumes the common fold of the other NMP kinases with five parallel β-strands forming a β-sheet core surrounded by helixes (Figure 1a). The highly hydrophobic dimer interface is composed of three parallel α-helixes provided by each monomer that stack against each other.

The asymmetric unit of the $P2_1$ crystal form consists of eight monomers which pack in four dimers. A homodimer is the basic unit of yeast TmpK, but no cooperativity was detected in our kinetic assays (*17*). Upon overlaying all eight monomers, we found that they can be divided into two groups; mon1, mon3, mon5, and mon7 adopt one conformation, whereas mon2, mon4, mon6, and mon8 adopt another conformation (as neither ncs constraints nor restraints were used in the last stages of refinement, the conformation adopted by each monomer is totally independent of the other monomers; in the final cycle of refinement ncs restraints were reapplied). For reasons that become apparent later, we designate the first group as inactive monomers, and the second group of monomers as active (Figure 1b). The rmsd between the active and inactive monomers excluding the LID sequence is 0.3 Å (on Cα atoms).

Most likely, it is the crystal packing that determines which monomers adopt the active conformation, and which do not. Apparently, an active monomer of one dimer inactivates a monomer of another dimer. Arg173 (a nonconserved residue) from all active monomers interacts with the TP_5A bound to the inactive monomers through an oxygen atom of phosphate PB, whereas Arg173 from the inactive monomers cannot form such an interaction. The presence of Arg173 prevents the formation of an active conformation of those monomers. Thus, it is this interaction of Arg173 from an active monomer with the TP_5A of an inactive monomer which presumably prevents the closure of the LID of those monomers and results in poor density. Since TmpK is a dimer in solution (on the basis of gel filtration and dynamic light scattering experiments) and the fact that binding studies with TP_5A clearly indicate a single class of binding sites in solution, we assume that the partitioning into active and inactive monomers is a pure crystallization artifact with no physiological significance. Therefore, we have limited our discussion to the active conformers.

Comparison with the dTMP−TmpK Complex Structure. Kinases undergo large conformational changes upon the binding of substrates (*18*). At least three different states have been described: an open state in the absence of substrates, a partially closed state when one of the substrates is bound, and a closed state when both substrates are present. The previously reported dTMP−TmpK complex (*7*) represents the partially closed state (ATP is missing), whereas the TP_5A complex represents the fully closed state, which is catalytically competent. Superposition of these two complexes shows that the main difference between them lies in the LID region (the other slight differences are concentrated in loop sequences, which are expected to vary as a result of the different crystal packing between the two complexes). In NMP kinases, the LID region has been observed to change its conformation upon ATP binding. In the dTMP complex, the LID sequence was only traceable in one of the monomers, and the density for the traced sequence was weak, indicating disorder, as expected in the absence of ATP. With TP_5A, we expected to see a much more ordered LID region. In the monomers that we designated as the inactive conformer, the LID region is very hard to trace and lacks secondary structure. In contrast, the LID region of the active conformers was easier to trace and forms a helix (Figure 1a). Upon superposition of TmpK with uridylate kinase (see more later for a detailed comparison with other NMP kinases), this newly formed helix overlays very well with a helix from the UmpK LID region (*19*).

Binding of TP_5A. Titration of TmpK to a cuvette containing the fluorescent bisubstrate inhibitor TP_5A-MANT at 0.09 μM (data not shown) results in a fluorescence increase that indicates complex formation. The titration data were fitted by a quadratic equation (*13*) which yields a K_d of 135 nM for the binding of TP_5A-MANT to TmpK. Competitive displacement of TP_5A-MANT from the so-formed complex by TP_5A and analysis by a cubic equation (*20*) show the K_d for the binding of TP_5A to TmpK to be 95 nM. The affinity of TP_5A to TmpK appears to be rather weak compared to that of *E. coli* adenylate kinase that binds AP_5A with a K_d of 15 nM or to *Dictyostelium discoideum* UmpK that binds UP_5A with a K_d of 3 nM (*13*, *21*). Time-resolved measurements (see Materials and Methods) show that the binding of TP_5A and TP_5A-MANT is relatively fast but perhaps somewhat below the diffusion-controlled limit with a k_{on} of 10×10^6 M^{-1} s^{-1} and a k_{on} of 5.5×10^6 M^{-1} s^{-1}, respectively. This indicates that the conformational changes induced by the bisubstrate inhibitor, e.g. closing of the LID region over substrate, do not constitute the rate-limiting step of the catalytic cycle since k_{cat} of yeast TmpK was shown to be 35 s^{-1} (*17*) whereas the binding of TP_5A did not deviate from a linear relationship for k_{obs} versus [TP_5A] up to 60 s^{-1}.

In the crystal, the bisubstrate inhibitor TP_5A appears to be bound tightly by all monomers (see Figure 2), as can be inferred from the relatively low *B*-factors (see Table 1) and the excellent electron density for the nucleotides. Before we describe in detail the interactions between the TP_5A dinucleotide and the enzyme, it is important to realize that the TP_5A molecule is not a transition-state analogue but is rather a bisubstrate or a biproduct analogue (*19*). In other words, it is not clear whether the observed TP_5A is closer to the ATP−dTMP or the ADP−dTDP state.

The thymidine part of TP_5A, which is completely surrounded by protein atoms, has a lower *B*-factor than the relatively exposed adenine part. The detailed interactions between the thymidine moiety and TmpK have already been described for the dTMP−TmpK complex (*7*), so only the differences in the TP_5A complex will be outlined here. Lysine 37 is observed to interact with the phosphate of dTMP in the dTMP−TmpK complex structure. This phosphate (for phosphate notation, see Figure 2*c*) corresponds to PE in TP_5A. In the TP_5A complex, Lys37 interacts with PD instead; PD can be seen either as the connector between dTMP (PE) and ATP (PA, PB, or PC), and thus as an artificial moiety, or as the β-phosphate of dTDP. In the other NMP kinase structures with such bisubstrate analogues, the position of this connecting phosphate varies (*19*, *22*, *23*). Thus, it is hard to interpret this change in Lys37 interaction from PE to PD, but it is possible that this residue participates in stabilizing the transferred phosphate group in the transition state.

The P-loop motif has been identified in many ATP- and GTP-binding proteins (*24*). The binding of the nucleotide to the P-loop is through main chain nitrogen atoms to the α- and β-phosphates of the nucleotide, and through a strictly conserved lysine. The ATP part of TP_5A interacts in such a manner with the P-loop of TmpK, through the amides of Thr16 and Thr19 to the PB oxygen atoms, and Thr20 to the PA oxygen atoms. In addition to the amide−phosphate interaction of Thr20, its side chain also interacts with PA. The P-loop lysine (Lys18) is observed to interact with the PB and PC oxygen atoms of TP_5A. It is positioned such that it may stabilize the transition state, in agreement with mutational analysis of adenylate kinase (*12*, *13*, *25*). In the nonactive monomers, Arg15 (situated at the tip of the P-loop) is in an extended conformation making no interaction with TP_5A, while in the active monomers, Arg15 bends to interact with a PC oxygen. It is in fact this interaction which leads us to divide the monomers into active and inactive, since we attribute an essential catalytic role to this arginine (see below).

The adenine part of TP_5A makes only two interactions with the enzyme, both through the amino group at C6. One is to the main chain carbonyl group of Lys187, the other through a water molecule to the highly conserved Gln21. This would explain the preference of TmpK for adenine nucleotides over guanine; in guanine-based nucleotides, there is a carbonyl group at C6 instead of the amino group, which would prevent the favorable interaction with the main chain carbonyl of Lys187. However, the partial acceptance of non-adenine-based nucleotides as phosphoryl donors by TmpK (*1*) is consistent with our structure, as there are no strong interactions between the adenine base and the enzyme, and there would be no steric discrimination against the amino group at the C2 position of guanine, which would face toward the solvent.

We do not observe bound magnesium in our structure, which we attribute to the low pH and the presence of ammonium sulfate that were part of the crystallization conditions. Inferring from other NMP kinases (*19*), the magnesium is expected to be octahedrally coordinated to the PB and PC oxygens (corresponding to β- and γ-phosphates of ATP), Thr16, Asp93 (directly or through a water molecule), and two additional water molecules. At the position where magnesium is expected, we see electron density only for a water molecule bridging Asp93 and PC.

Comparison with Other NMP Kinases. In addition to this TP_5A−TmpK complex structure, there are at present three other structures of NMP kinases with bisubstrate analogues: adenylate kinase (yeast and *E. coli*) with AP_5A (*22*, *23*) and uridylate kinase (slime mold) with UP_5A (*19*). The above-mentioned adenylate kinases (AKs) are both of the long LID region type, while uridylate kinase is more like TmpK having a short LID region (for the LID's amino acid sequence, see Figure 4). Therefore, we have chosen to compare TmpK to uridylate kinase (UmpK). The structure of UmpK complexed with UP_5A was superimposed on the TP_5A−TmpK complex structure by aligning the P-loop sequences of both enzymes. While UmpK and AK have a high degree of sequence identity, none exists between TmpK and either UmpK or AK, but structurally, TmpK overlays on UmpK surprisingly well and UP_5A occupies nearly the same position as TP_5A. There is, however, a striking difference. The LID domain of UmpK interacts directly with the phosphates of UP_5A via the basic residues Arg131 and Arg137, which have been shown to stabilize the transition state (*26*). Which residues in TmpK would correspond to those arginines in UmpK? While lacking such basic residues in the LID domain, TmpK has Arg15 in the P-loop, the corresponding amino acid in UmpK being a glycine. And in fact, in what we define as the monomers in the active conformation, Arg15 makes a 2.8 Å hydrogen bond interac-

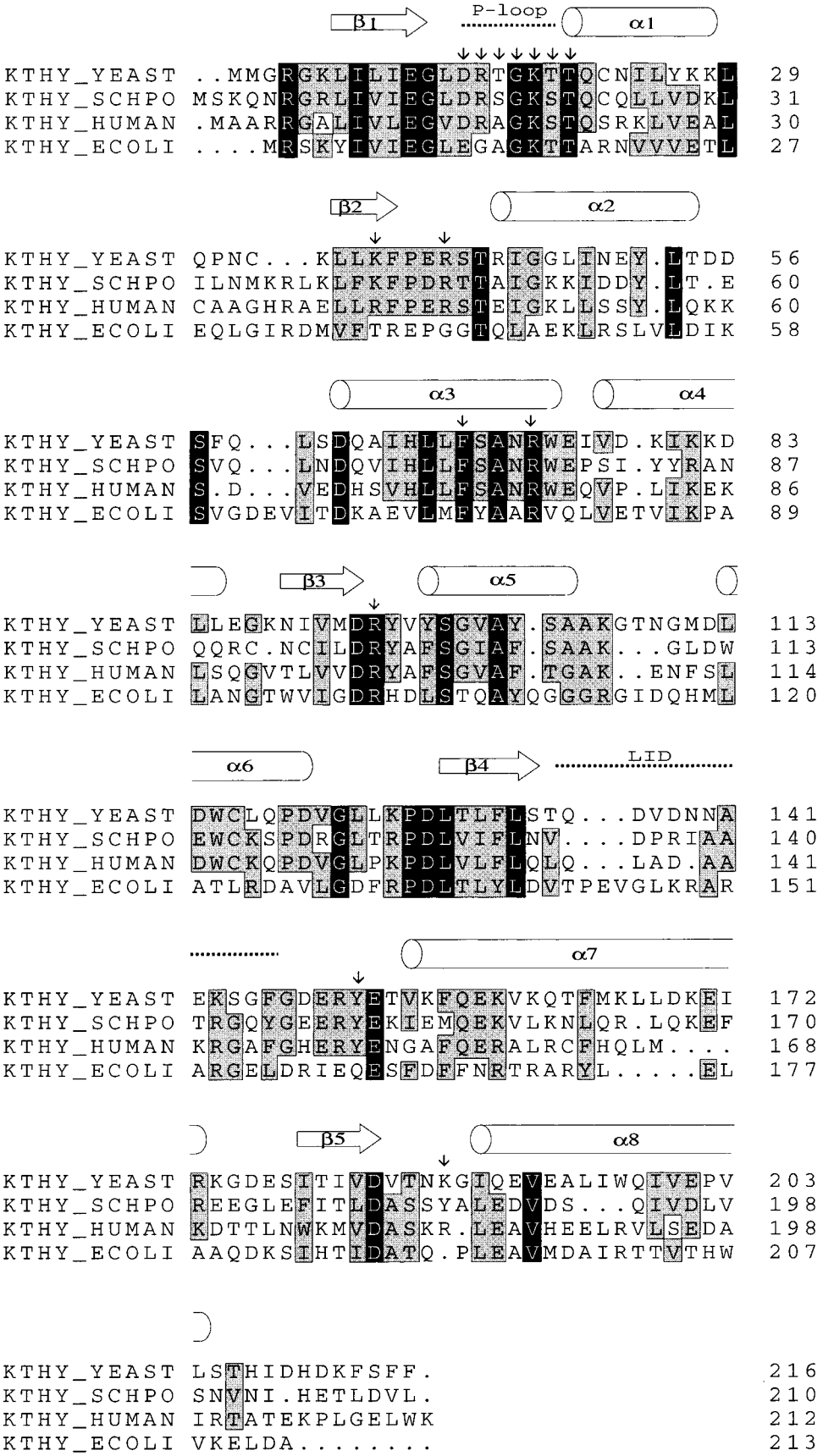

FIGURE 4: Sequence alignment (using GCG) of three eukaryotic (KTHY_SCHPO is TmpK from *Schizosaccharomyces pombe*) and the *E. coli* thymidylate kinase amino acid sequences. Shaded black are residues conserved in all sequences; in gray are similar amino acids found in at least three of the sequences. The secondary structural elements (helixes as tubes and strands as arrows) of the yeast TmpK are also shown. Residues that contact TP_5A directly are marked with an arrow. The LID part of the sequence is marked as a dotted line; note the difference in the P-loop region between the eukaryotic TmpKs and the *E. coli* enzyme (between β1 and α1), and the additional basic residues found only in *E. coli* TmpK (see the text).

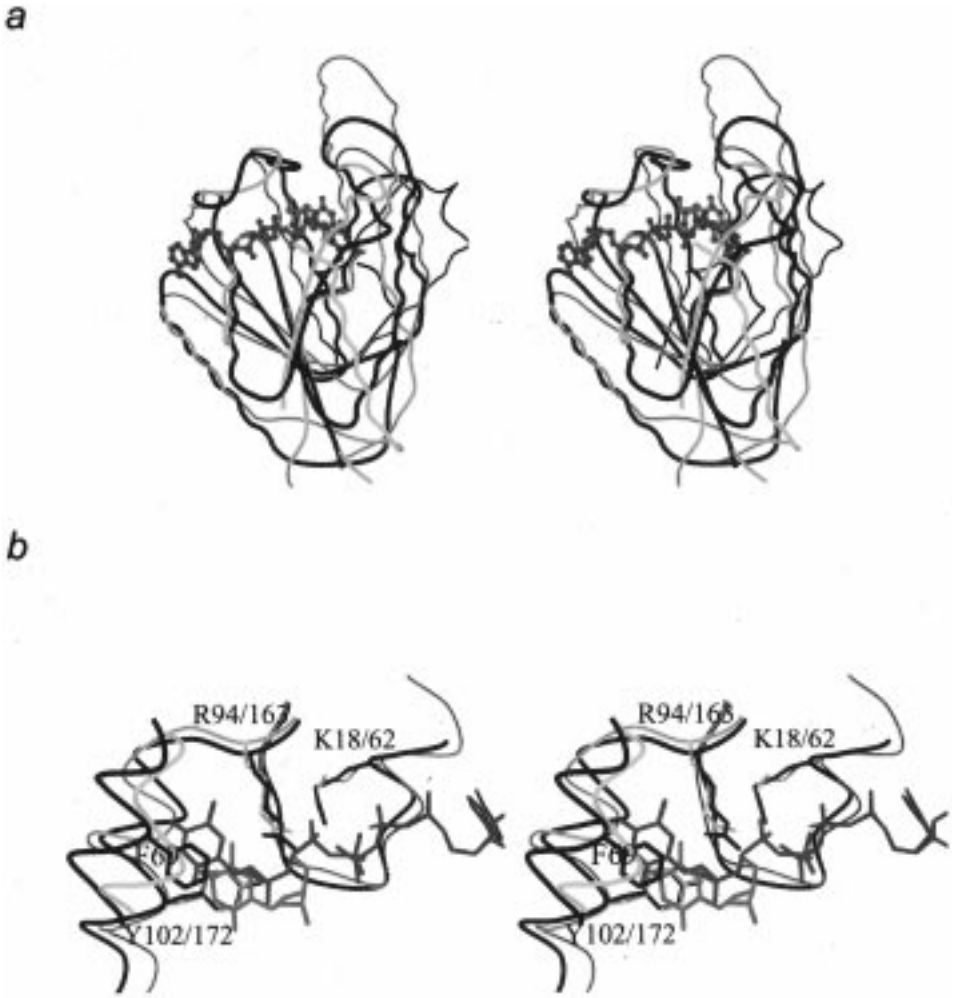

FIGURE 5: Comparison of yeast thymidylate kinase with HSV1 thymidine kinase (Brookhaven Protein Data Bank code 1kin). (a) Superposition of both kinases (both are homodimers; only a monomer is shown) with TmpK in blue and HSV1-TK in yellow. The TP_5A bound to TmpK is also shown (cyan), as is the thymidine bound to HSV1-TK (orange). (b) Stereoview (rotated 180° relative to the figure above) with a closeup of the active sites. Thymidine (orange) is displaced relative to the thymidine moiety of TP_5A (cyan). As a result, where in TmpK Tyr102 discriminates against ribonucleotides, the corresponding Tyr172 in HSV1-TK makes a base-stacking interaction with the thymidine base, a role accomplished by Phe69 in TmpK.

tion with one oxygen atom of PC. This would be an example of the importance of having catalytic residues in the proximity of the reaction center, but their origin, LID or P-loop, is irrelevant for catalysis.

Comparison with Herpes Simplex Virus-1 Thymidine Kinase. From sequence similarity, the herpes simplex virus-1 thymidine kinase (HSV1-TK) has been grouped with the thymidylate kinase family. In fact, HSV1-TK has both a thymidine and a thymidylate kinase activity, suggesting that it evolved from thymidylate kinase and acquired the additional thymidine kinase activity. The three-dimensional X-ray structure of HSV1-TK has been solved recently revealing, as for TmpK, a dimer as the basic unit (*27*). The viral kinase, however, contains 376 amino acids compared to TmpK's 216. It has additional N-terminal (45 residues) and C-terminal sequences (87 residues), as well as a few short inserts in the overlapping sequence region. Nevertheless, the overlapping sequence regions can be relatively easily overlayed structurally, with both kinases displaying the basic kinase fold as previously described (Figure 5).

The low level of substrate specificity of the viral kinase, which is medicinally important in the activation of nucleoside prodrugs, can be rationalized upon comparison with the highly specific TmpK. Using the 14-amino acid stretch of the P-loop regions of both enzymes for the superposition (Cα rms of 0.7 Å), it becomes apparent that the base moiety of thymidine bound to HSV1-TK is about 3.5 Å displaced in relation to the position of the base in TmpK. The displacement is away from the core of the enzyme toward the solvent. This slippage of the nucleotide toward the active site residues (mainly basic residues which would interact with phosphate groups) but without the concomitant contraction of the nucleotide's base binding pocket explains why HSV1-TK can phosphorylate even purines in addition to pyrimidines, and how it can possess both thymidine and thymidylate kinase activity. Presumably, after the first phosphorylation of thymidine to thymidine monophosphate, the nucleotide has room to slip deeper into the binding site, thus placing the phosphate at the position previously occupied by the C5 hydroxyl. Now, the second phosphorylation can take place. Thus, it is the nucleotide that moves relative to the enzyme's catalytic residues. This would suggest that HSV1-TK, while accepting purine nucleosides as substrates, might not accept purine monophosphates (the substrates for the thymidylate kinase activity) as they must be bound deeper in the binding cavity and thus might not fit. And in fact, the guanine based

anti-herpes drug acyclovir is phosphorylated to the monophosphate by HSV1-TK, but the following phosphorylation step is catalyzed by guanylate kinase (*28*), consistent with this interpretation.

In contrast to the base, other residues in the active site overlay very well, the noteworthy ones being Tyr102 (TmpK) and Tyr172 (HSV1-TK) (within 0.3 Å). Tyr102 is responsible for the discrimination against ribonucleotides in TmpK by being located 3.5 Å from the C2 of the ribose. In HSV1-TK, it fulfills a completely different role by base stacking with thymine (in TmpK, Phe69 stacks analogously to this Tyr with the base but from the opposite side of the base; see Figure 5b).

Important for the previous discussion of catalysis are the P-loop and LID regions; in this respect, HSV1-TK is more similar to the *E. coli* TmpK than to the yeast TmpK in lacking the arginine in the P-loop (D14R15 in yeast and E12G13 and H58G59 in *E. coli* and HSV1-TK, respectively) and having basic residues in the LID region. Arg222 of HSV1-TK, which is in the LID region, is situated less than 4.5 Å from the deoxyribose O5′ of thymidine. In analogy to other NMP kinases that utilize basic residues from the LID region to stabilize the negative charge developing in the transition state, we propose that Arg222 of HSV1-TK fulfills this catalytic role.

Implications for Catalysis and AZT Activation. In addition to phosphorylating dTMP to dTDP, thymidylate kinase is also part of the activation pathway of the anti-HIV prodrug 3′-deoxy-3′-azidothymidine (AZT). In fact, TmpK is the rate-limiting enzyme in this activation pathway. Together with the previously reported nucleotide complex structures with dTMP and with AZT-MP (*7*), the TP_5A−TmpK complex structure suggests which residues are needed for catalysis, and why AZT-MP is phosphorylated by TmpK so poorly.

Whether the phosphoryl transfer mechanism is associative in nature, dissociative, or of mixed type, the negative charge developing during the transition state compared to that of the ground state has to be stabilized. This would be achieved by a basic residue (Arg or Lys), and its interaction with the phosphate must be made or strengthened at the transition state. In UmpK, it has been shown (*26*) that the mechanism of phosphoryl transfer is most probably associative, and that the charge developing on the transferred phosphoryl group is stabilized mainly by Arg131 and Arg137 from the LID region, in addition to Lys19, Arg148, and the catalytic magnesium ion. In contrast, yeast TmpK lacks basic residues in its LID region, but does have an arginine residue at the tip of the P-loop. In the TmpK−TP_5A complex, this arginine's side chain is at a position similar to that of an arginine originating from the LID domain, and is located 2.8 Å from PC of TP_5A. Therefore, we attribute to Arg15 a role similar to that of the UmpK's LID arginines.

As pointed out previously (*7*, *17*), the binding of AZT-MP causes the P-loop of TmpK to shift by about 0.5 Å due to the interaction between the azido moiety of AZT and the side chain of Asp14. This shift affects Arg15 as well, causing it to be not optimally located to fulfill its catalytic role. Most but not all thymidylate kinases sequenced to date have an arginine in the P-loop. A noticeable exception is thymidylate kinase from *E. coli* which has a glycine residue at that position (like UmpK). The LID domain of the *E. coli* TmpK contains basic residues, unlike yeast TmpK but like UmpK. We postulate that the *E. coli* TmpK functions in a manner very similar to that of the slime mold UmpK, where arginine residues from the LID domain participate in catalysis (candidates are Arg149, Arg151, Arg153, and Arg158). This would suggest that a P-loop movement of *E. coli* TmpK, like that observed with the yeast enzyme and AZT-MP, would not have such a detrimental effect on catalysis, as the residues important for catalysis (except the conserved lysine) do not originate from the P-loop. To test this hypothesis, we cloned, expressed, and purified thymidylate kinase from *E. coli* and compared its dTMP and AZT-MP phosphorylation rates.

Table 2: Steady-State Kinetic Parameters

	yeast	*E. coli*
K_M for dTMP with ATP (μM)	9	2.7
K_M for AZT-MP with ATP (μM)	6	30
K_M for ATP with dTMP (μM)	190	8
K_M for ATP with AZT-MP (μM)	300	50
k_{cat} with ATP and dTMP (s^{-1})	35	15
k_{cat} with ATP and AZT-MP (s^{-1})	0.175	6
ratio of k_{cat} for dTMP/ k_{cat} for AZT-MP	200	2.5
ratio k_{cat}/K_M for dTMP/AZT-MP	133	27.5

Steady-State Kinetics of E. coli TmpK. As predicted on the basis of the structural rationale explained above, the rates of phosphorylation of AZT-MP and dTMP by *E. coli* TmpK are comparable (Figure 3). While in the case of yeast TmpK AZT-MP is phosphorylated 200-fold slower than dTMP, for the *E. coli* enzyme the factor is only 2.5 (Table 2). If the K_M values are also taken into account, this comparison is still visible; whereas the ratio of k_{cat}/K_M of dTMP versus AZT-MP is 133 in the case of yeast TmpK, it is only 27 in the case of *E. coli* TmpK, illustrating the big difference in acceptance of AZT-MP by these two dTMP kinases from different organisms. It appears from this comparison that NMP kinases have developed different strategies to attain catalysis; one class has catalytic residues in the P-loop and the LID regions (AK, UmpK, etc.), whereas most dTMP kinases (*E. coli* TmpK being an exception) apparently lack the catalytic arginine residues of the LID region. In this respect, the possibility that the mechanism of phosphoryl transfer of the yeast TmpK may not be purely associative as was deduced for UmpK must also be considered (*26*).

CONCLUSIONS

In the presence of the bisubstrate inhibitor TP_5A, the LID region of TmpK forms a helix and adopts a closed conformation. This situation is similar to the LID closure observed in AK and UmpK, despite the fact that these NMP kinases have catalytically essential residues in the LID region in contrast to yeast TmpK which appears to have them mainly in the P-loop. The TmpK−TP_5A complex structure shows that Arg15 from the P-loop interacts with PC, which is most likely of catalytic importance.

We have shown previously that the binding of AZT-MP to yeast TmpK results in a P-loop movement which affects the precise placement of this catalytic arginine (*7*, *17*). The *E. coli* TmpK does not have this arginine in the P-loop but rather contains arginines in the LID region that may fulfill the same role. Thus, we predicted that the presumable AZT-

MP-induced displacement of the P-loop of *E. coli* TmpK should not be detrimental to catalysis. Indeed, the catalytic rates for dTMP and AZT-MP phosphorylation differ by a factor of only 2.5. This observation is of far-reaching medicinal importance since it signals a rational approach to using gene therapy to improve the antiviral impact of AZT, as outlined previously (*17*, *29*, *30*).

ACKNOWLEDGMENT

We thank Ingrid R. Vetter for valuable discussions, Petra Herde and Sonja Hönig for skilled technical assistance, Georg Holtermann for technical support with the X-ray equipment, and Klaus Scheffzek for careful reading of the manuscript.

REFERENCES

1. Jong, A. Y., and Campbell, J. L. (1984) *J. Biol. Chem. 259*, 14394−8.
2. Furman, P. A., Fyfe, J. A., St, C. M., Weinhold, K., Rideout, J. L., Freeman, G. A., Lehrman, S. N., Bolognesi, D. P., Broder, S., Mitsuya, H., et al. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8333−7.
3. Qian, M., Bui, T., Ho, R. J., and Unadkat, J. D. (1994) *Biochem. Pharmacol. 48*, 383−9.
4. Yan, J. P., Ilsley, D. D., Frohlick, C., Steet, R., Hall, E. T., Kuchta, R. D., and Melancon, P. (1995) *J. Biol. Chem. 270*, 22836−41.
5. Tornevik, Y., Ullman, B., Balzarini, J., Wahren, B., and Eriksson, S. (1995) *Biochem. Pharmacol. 49*, 829−37.
6. Bridges, E. G., Faraj, A., and Sommadossi, J. P. (1993) *Biochem. Pharmacol. 45*, 1571−6.
7. Lavie, A., Vetter, I. R., Konrad, M., Goody, R. S., Reinstein, J., and Schlichting, I. (1997) *Nat. Struct. Biol. 4*, 601−4.
8. Kabsch, W. (1993) *J. Appl. Crystallogr. 24*, 795−800.
9. Navaza, J. (1994) *Acta Crystallogr. A 50*, 157−63.
10. Brünger, A. T. (1993) *X-PLOR: a system for X-ray crystallography and NMR*, Yale University Press, New Haven, CT.
11. Jones, T. A., Zhou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A 47*, 110−9.
12. Reinstein, J., Brune, M., and Wittinghofer, A. (1988) *Biochemistry 27*, 4712−20.
13. Reinstein, J., Vetter, I. R., Schlichting, I., Rosch, P., Wittinghofer, A., and Goody, R. S. (1990) *Biochemistry 29*, 7440−50.
14. Packschies, L., Theyssen, H., Buchberger, A., Bukau, B., Goody, R. S., and Reinstein, J. (1997) *Biochemistry 36*, 3417−22.
15. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60−89.
16. Reynes, J. P., Tiraby, M., Baron, M., Drocourt, D., and Tiraby, G. (1996) *J. Bacteriol. 178*, 2804−12.
17. Lavie, A., Schlichting, I., Vetter, I. R., Konrad, M., Reinstein, J., and Goody, R. S. (1997) *Nat. Med. (N.Y.) 3*, 922−4.
18. Vonrhein, C., Schlauderer, G. J., and Schulz, G. E. (1995) *Structure 3*, 483−90.
19. Scheffzek, K., Kliche, W., Wiesmüller, L., and Reinstein, J. (1996) *Biochemistry 35*, 9716−27.
20. Thrall, S. H., Reinstein, J., Wohrl, B. M., and Goody, R. S. (1996) *Biochemistry 35*, 4609−18.
21. Wiesmüller, L., Scheffzek, K., Kliche, W., Goody, R. S., Wittinghofer, A., and Reinstein, J. (1995) *FEBS Lett. 363*, 22−4.
22. Müller, C. W., and Schulz, G. E. (1992) *J. Mol. Biol. 224*, 159−77.
23. Abele, U., and Schulz, G. E. (1995) *Protein Sci. 4*, 1262−71.
24. Saraste, M., Sibbald, P. R., and Wittinghofer, A. (1990) *Trends Biochem. Sci. 15*, 430−4.
25. Tsai, M.-D., and Yan, H. (1991) *Biochemistry 30*, 6806−18.
26. Schlichting, I., and Reinstein, J. (1997) *Biochemistry 36*, 9290−9296.
27. Brown, D. G., Visse, R., Sandhu, G., Davies, A., Rizkallah, P. J., Melitz, C., Summers, W. C., and Sanderson, M. R. (1995) *Nat. Struct. Biol. 2*, 876−81.
28. Elion, G. B. (1982) *Am. J. Med. 73* (1A), 7−13.
29. Balzarini, J., Pauwels, R., Baba, M., Herdewijn, P., De Clerecq, E., Broder, S., and Johns, D. G. (1988) *Biochem. Pharmacol. 37*, 897−903.
30. Guettari, N., Loubiere, L., Brisson, E., and Klatzmann, D. (1997) *Virology 235*, 398−405.

BI9720787