Role of sequence evolution and conformational dynamics in the substrate specificity and oligomerization mode of thymidylate kinases

Ansuman Biswas, Subbarao Jasti, Jeyaraman Jeyakanthan and Kanagaraj Sekar*

*Department of Physics, Indian Institute of Science, Bangalore 560012, India; Centre for Chemical Biology and Therapeutics, Institute for Stem Cell Biology and Regenerative Medicine, Bangalore 560065, India; Department of Bioinformatics, Alagappa University, Karaikudi 630003, India; Department of Computational and Data Sciences, Indian Institute of Science, Bangalore 560012, India

Communicated by Ramaswamy H. Sarma

(Received 4 April 2016; accepted 15 June 2016)

Thymidylate kinase (TMK) is a key enzyme for the synthesis of DNA, making it an important target for the development of anticancer, antibacterial, and antiparasitic drugs. TMK homologs exhibit significant variations in sequence, residue conformation, substrate specificity, and oligomerization mode. However, the influence of sequence evolution and conformational dynamics on its quaternary structure and function has not been studied before. Based on extensive sequence and structure analyses, our study detected several non-conserved residues which are linked by co-evolution and are involved in various interactions in flexibility, oligomeric assembly, and substrate specificity among the homologs. These lead to differences in the pattern of interactions at the active site in TMKs of different specificity. The method was further tested on TMK from Sulfolobus tokodaii (STTMK) which has substantial differences in sequence and structure compared to other TMKs. Our analyses pointed to a more flexible dTMP-binding site in STTMK compared to the other homologs. Binding assays proved that the protein can accommodate both purine and pyrimidine nucleotides at the dTMP binding site with comparable affinity. Additionally, the residues responsible for the narrow specificity of Brugia malayi TMK, whose three-dimensional structure is unavailable, were detected. Our study provides a residue-level understanding of the differences observed among TMK homologs in previous experiments. It also illustrates the correlation among sequence evolution, conformational dynamics, oligomerization mode, and substrate recognition in TMKs and detects co-evolving residues that affect binding, which should be taken into account while designing novel inhibitors.

Keywords: sequence evolution; molecular dynamics; substrate specificity; mutual information; co-evolution

Introduction

Thymidylate kinase (TMK, EC 2.7.4.9), a member of nucleoside monophosphate kinase (NMP) family of enzymes, occurs at the junction of de novo and salvage pathways for deoxothymidine triphosphate (dTTP) synthesis (Jong & Campbell, 1984; Reichard, 1988). It catalyzes the reversible transfer of γ phosphate from adenosine triphosphate (ATP) to deoxothymidine monophosphate (dTMP) in the presence of Mg2+. The reaction yields deoxothymidine diphosphate (dTDP) which is further utilized during DNA synthesis. The protein is found in all organisms and till date, around 73-dimensional structures (including both native and substrate-bound forms) have been submitted to the Protein Data Bank (PDB) (Westbrook et al., 2002). Owing to its involvement in DNA synthesis, TMK is a recognized drug target for several pathogens.

TMK is a homodimer with a β-core of five parallel strands surrounded by 7–11 α helices in each subunit (Figure 1(A)). It contains three signature motifs: a P-loop with sequence GX,X,X,X,GKS/T at the ATP binding site, the DRX motif in β3/α4 loop of which the Arg side chain coordinates the phosphate group transfer from ATP to dTMP (Ostermann et al., 2000), and the flexible ‘Lid’ region which accommodates the phosphate donor and plays a role in catalysis. Additional conserved regions detected from the structure-based sequence alignment (Figure 1(B)) include an Arg at the N-terminus of the Lid region, whose guanidine group interacts via cation–π interaction with the adenine base of ATP bound to the P-loop region (Whittingham et al., 2010), the (E/F)P cis peptide motif in β2/α2 loop which forms the base of dTMP-binding cavity (Lavie et al., 1998; Ostermann et al., 2000) and residues F and R in α3 and Y in α4 which coordinate the thymine base of dTMP at the pocket (Whittingham et al., 2010). Based on the presence or absence of an additional positively charged Arg residue in the P-loop, other than the already present
Figure 1. Sequence and structural variations in TMKs. (A) Overall structure of TMK, (B) structure-based sequence alignment depicting the conserved and variable residues. Loops are denoted by brown arrows, the EP cis peptide motif in β2/α2 loop by red triangles, DRX motif by blue triangles, Arg at the N-terminus of the Lid region by inverted cyan triangle and residues F, R, and Y which coordinate the thymine base of dTMP at the pocket by inverted brown triangles and (C) structural variations in loop regions across homologous TMKs (PtTMK: PDB-id 2wwf, HsTMK: PDB-id 1nn5, ETMK: PDB-id 4tmk, StTMK: PDB-id 1n5i, SaTMK: PDB-id 2ccj, VcTMK: PDB-id 3lv8, yTMK: PDB-id 3tmk, MtTMK: PDB-id 1n5i and VTMK: PDB-id 2v54) (Colour online).
lysine of GKS/T, the protein has been classified, respectively, into type 1 (e.g. human, yeast, *Plasmodium*) and type 2 (e.g. *E. coli*) (Lavie et al., 1998).

While homologous proteins conserve the features essential for function, they also exhibit signatures of evolution like variations in sequence, structure, and quaternary state (Marsh & Teichmann, 2014). These signatures are evident in TMKs from different organisms (Figure 1(B) and (C)). Several loops at or near the active site like the $\alpha_2/\alpha_3$ loop, the $\alpha_4/\alpha_5$ loop, and the Lid region ($\alpha_6/\alpha_7$ loop) contain insertions in some TMKs, while residue substitutions are observed in the P-loop, $\beta_2/\alpha_2$ loop, and $\alpha_2$ helix. On the other hand, structural variability is observed primarily in the P-loop, Lid region, $\beta_2/\alpha_2$ loop, and $\alpha_2$ helix regions from the structure alignment of TMKs (Figure 1(C)). The Lid region is dynamic and found to be disordered in some of the crystal structures. Additionally, there are differences in the mode of dimerization as revealed from the orthogonally stacked monomers in Vaccinia virus (VTMK, PDB-id: 2v54), the parallel dimer in *Mycobacterium tuberculosis* (MtTMK, PDB-id: 1w2h), and the anti-parallel dimer in other TMK structures (Figure 2). The parallel MtTMK dimer (mid-temperature of thermal denaturation $T_m$ > 65°C) is more stable compared to the anti-parallel Human TMK ($T_m$ = 52°C), while the VTMK dimer ($T_m$ = 42.5°C) is the least stable among the three (Caillat et al., 2008).

Extensive structural and biochemical studies performed on homologous TMKs revealed diversity in substrate specificity. *Brugia malayi* TMK (BmTMK, three-dimensional structure unavailable) is highly specific to ATP as phosphate donor (Doharey et al., 2014), while *Streptococcus pneumonia* SpTMK (Petit & Korette, 2002), *Plasmodium falciparum* PfTMK (Kandeel & Kitade, 1997) showed broader specificity, the activity with ATP being predominant. *E. coli* ETMK (Nelson & Carter, 1969), MtTMK (Munier-Lehmann, Chaffotte, Pochet, & Labesse, 2001), VTMK (Topalis et al., 2005), and yeast yTMK (Jong & Campbell, 1984) showed activity with both purine (ATP) and pyrimidine (CTP) nucleotides (Table 1A). In case of phosphate group acceptor, most of the TMKs showed high specificity toward dTMP (Table 1B), except PfTMK (Kandeel & Kitade, 2008) and VTMK (Caillat et al., 2008; Topalis et al., 2005) which showed considerable activity with both pyrimidine (dTMP, dUMP) and the larger purine (dGMP) nucleotides and MtTMK (Munier-Lehmann et al., 2001) which showed activity with different pyrimidine nucleotides (dTMP, dUMP).

Previous studies have revealed diversity in the sequence, structure, oligomerization mode, and substrate preference of TMK homologs. However, no comprehensive study has been performed till date, to the best of our knowledge, to detect the residues responsible for this diversity. This paper reports detailed sequence and structure-based analyses to detect the ‘variable functional residues’ (VFRs) which modulate substrate specificity and oligomerization mode in TMKs. Such residues are usually found to co-evolve with each other (Chakrabarti & Panchenko, 2009; Gloor, Martin, Wahl, & Dunn, 2005; Lee, Mick, Furdui, & Beamer, 2012; Martin, Gloor, Dunn, & Wahl, 2005), that is, a mutation at one position is accompanied by a compensatory mutation of another residue while conserving the function. Residue variability at the substrate-binding sites and oligomerization interfaces is expected to influence the interaction network, thereby modulating the conformational flexibility of the sites (Liu & Bahar, 2012). Thus, sequence evolution coupled with conformational dynamics plays an important role in determining the substrate specificity and oligomerization mode of TMKs.
Table 1A. Percentage enzyme activity of TMKs with different phosphate donors (percentages calculated considering the reaction rate with ATP as 100%).

| Organism                      | Purine nucleotide | Pyrimidine nucleotide |
|-------------------------------|-------------------|-----------------------|
|                               | ATP (%) | GTP (%) | CTP (%) | dTTP (%) | UTP (%) | ATP (%) | GTP (%) | CTP (%) | dTTP (%) | UTP (%) |
| Yeast (Jong & Campbell, 1984) | 100     | 35      | 26       | 3        | 36      | 100     | 35      | 26       | 3        | 36      |
| S. pneumoniae (Petit & Korete, 2002) | 100   | 0       | 20       | 0        | 0       | 100     | 0       | 20       | 0        | 0       |
| Human (Lee & Cheng, 1977)     | 100     | 38      | 5        | NA       | 4       | 100     | 60      | 42       | NA       | 26      |
| P. falciparum (Kandeel & Kitade, 2008) | 100   | <10     | <10      | <10      | <10     | 100     | 1–12    | 1–12     | 1–12     | 1–12    |
| E. coli (Nelson & Carter, 1969) | 100     | 1–12    | 31.6     | 1–12     | 1–12    | 100     | 60      | 42       | NA       | 26      |
| M. tuberculosis (Munier-Lehmann et al., 2001) | 100   | 60      | 42       | NA       | 26      | 100     | 0       | 0        | 0        | 0       |
| Brugia malayi (Doharey et al., 2014) | 100     | 0       | 0        | 0        | 0       | 100     | 0       | 0        | 0        | 0       |

Notes: NA means data not available. High activities with multiple nucleotides denote broad specificity.

*These data are from experiments published in literature.

Table 1B. Percentage enzyme activity of TMKs with different phosphate acceptors (percentages calculated considering the reaction rate with dTMP as 100%).

| Organism                      | Purine nucleotide | Pyrimidine nucleotide |
|-------------------------------|-------------------|-----------------------|
|                               | dAMP (%) | dGMP (%) | dIMP (%) | dCMP (%) | dTTP (%) | dUMP (%) | dAMP (%) | dGMP (%) | dIMP (%) | dCMP (%) | dTTP (%) | dUMP (%) |
| Yeast (Jong & Campbell, 1984) | 0       | 0       | NA       | 0        | 100      | 31       | 0       | 0       | NA       | 0        | 15       | 0       |
| S. pneumoniae (Petit & Korete, 2002) | 0     | 0       | NA       | 0        | 100      | 15       | 0       | 0       | NA       | 0        | 15       | 0       |
| Human (Lee & Cheng, 1977)     | 0       | 0       | 0        | 0        | 100      | 0        | 0       | 0       | 0        | 100      | 0        | 0       |
| P. falciparum (Kandeel & Kitade, 2008) | 0     | 86      | 15       | 0        | 100      | 43       | 0       | 0       | 0        | 100      | 14.7     | 0       |
| E. coli (Nelson & Carter, 1969) | 0       | 0       | 0        | 0        | 100      | 14.7     | 0       | 0       | 0        | 100      | 14.7     | 0       |
| M. tuberculosis (Munier-Lehmann et al., 2001) | 0     | 0       | 0        | 0        | 100      | 33       | 0       | 0       | 0        | 100      | 12.10    | 0       |
| Brugia malayi (Doharey et al., 2014) | 0       | 0       | 0        | 0        | 100      | 12.10    | 0       | 0       | 0        | 100      | 12.10    | 0       |

Notes: NA means data not available. High activities with multiple nucleotides denote broad specificity.

*These data are from experiments published in literature.

role in substrate binding, specificity, oligomerization mode, and evolution of new functions (Carlqvist et al., 2005; Gournas, Evangelidis, Athanasopoulos, Mikros, & Sophianopoulos, 2015; Marsh & Teichmann, 2014; Tokuriki & Tawfik, 2009). In this study, TMK sequences were analyzed to detect the presence of co-evolutionary relationship among the VFRs and identify additional VFRs not detected from the analyses of substrate-bound structures. The interaction network and dynamics of these residues were studied through network analysis and molecular dynamics (MD) simulations. Our analyses were validated by two case studies, namely, explanation of the narrow specificity of Brugia malayi TMK (BmTMK) whose three-dimensional structure is unknown and the prediction of a flexible dTMP-binding cavity that can possibly accommodate larger nucleotides in Sulfolobus tokodaii TMK (StTMK), which was validated using biochemical studies. Our study provides a residue-level understanding of the differences among TMK homologs observed in previous experimental findings and may help in predicting the specificity of uncharacterized TMKs, design of TMKs with altered specificity, and design of specific inhibitors for TMKs from different organisms.

Materials and methods

Sequence analysis

Full-length TMK sequences containing the GX1X2X3X4GK motif were obtained from the UniProt database (Magrane & Consortium, 2011). Sequences with less than 25% identity to the TMKs of known structure were eliminated using a BLAST search (Altschul, Gish, Miller, Myers, & Lipman, 1990), which yielded 1635 sequences. The multiple sequence alignment (MSA) was performed using Clustalw2 (Larkin et al., 2007). Mutual information (MI) analysis (Dunn, Wahl, & Gloor, 2008; Martin et al., 2005) was carried out using the MISTIC server (Simonetti, Teppa, Chernomoretz, Nielsen, & Marino Buslje, 2013) with default parameters to determine co-evolving residues. The MI between two positions in the MSA was given by:

$$MI(i, j) = \sum_{x,y} P(x, y) \log \left( \frac{P(x, y)}{P(x)P(y)} \right)$$

where $P(x)$, $P(y)$, and $P(x,y)$ denote, respectively, the frequencies of amino acid $x$ at position $i$, amino acid $y$ at position $j$ and the frequency of their simultaneous occurrence. Further, the server incorporates corrections to remove the influence of phylogeny (Dunn et al., 2008) and sequence redundancy (Buslje, Santos, Delfino, & Nielsen, 2009).
on MI calculation. Finally, the MI score is transformed into MI z-score, which represents how many standard deviations apart the observed MI value is from the mean MI of a randomized MSA. Higher MI z-scores imply that the observed correlation is unlikely to have occurred by chance and are, therefore, significant. Our study focused on residue pairs with MI z-scores >0.5 which have been proposed to be significant (Buslje et al., 2009; Simonetti et al., 2013) as the predictions at this threshold have a high specificity (.95).

Structure analysis

The structure-based multiple sequence alignment was generated using the RaptorX Structure Alignment server (Wang, Ma, Peng, & Xu, 2013). PDBsum (Laskowski, 2009) and PDBPISA (Krissinel & Henrick, 2007) servers were used for dimeric interface analyses. Interactions among the residues were calculated using MolBridge (Kumar, Kailasam, Chakraborty, & Bansal, 2014). PLIP server was used for calculating and visualizing protein–ligand interactions (Salentin, Schreiber, Haupt, Adasme, & Schroeder, 2015). ASSP server (Kumar & Bansal, 2015) was used to detect the π-helix in MtTMK, while HELANAL-Plus (Kumar & Bansal, 2012) was used to calculate helix radius. The depth of the active site pocket was calculated using the DoGSitescoring server (Volkamer, Kuhn, Grombacher, Rippmann, & Rarey, 2012). Residue interaction networks (RINs) were generated using the RINERATOR software (Dongcheva, Klein, Domingues, & Albrecht, 2011). For this purpose, hydrogen atoms were added to the protein structure by REDUCE (Word, Lovell, Richardson, & Richardson, 1999). Non-covalent interactions were identified and scored using PROBE (Word, Lovell, Labean, et al., 1999). The RIN is an undirected network of nodes (representing residues) connected by edges (representing non-covalent interactions between residues) whose weights are proportional to the interaction score $I_{ij}$. CYTOSCAPE (Shannon et al., 2003) was used to visualize the RINs, while the network parameters were calculated using the NETWORK ANALYZER plugin (Assenov, Ramirez, Schelhorn, Lengauer, & Albrecht, 2008) of CYTOSCAPE.

MD analysis

MD simulations were performed on HsTMK (PDB-id: 1e2d) and StTMK (PDB-id: 2plr) at their ambient temperatures (27 and 80°C, respectively) to study the functionally important motions in the protein. Water molecules and hetero atoms were removed from the PDBs, while the missing atoms and loops were modeled using SWISS-MODEL (Arnold, Bordoli, Kopp, & Schwede, 2006). MD simulation was run using the GROMACS 4.5.3 package (Hess, Kutzner, van der Spoel, & Lindahl, 2008) with AMBER03 (Duan et al., 2003; Sorin & Pande, 2005) all atom force field and SPC/E water model (Berendsen, Grigera, & Straatsma, 1987). The protein was simulated in a dodecahedron box with a distance of 10 Å between the box and the protein. CI- ions were added to neutralize unbalanced charges on the protein. Particle Mesh Ewald method (Darden, York, & Pedersen, 1993) was used for treating the electrostatic interactions with a Coulomb cut-off 12 Å, while a switch potential with a cutoff of 10 Å nm and with the switch function applied from 9 Å was used to treat the van der Waals interactions. The bond lengths were constrained using the LINCS algorithm (Hess, Bekker, Berendsen, & Fraaije, 1997). The system was energy minimized using the conjugate gradient algorithm with convergence criteria of 1 kJ mol$^{-1}$ nm$^{-1}$ to eradicate all short contacts. Thereafter, temperature and pressure equilibration were carried out by applying position restraints on the protein. The system was coupled to a temperature bath using the Nosé–Hoover thermostat (Hoover, 1985; Nosé, 1984). Initial velocities were generated from the Maxwell–Boltzmann distribution using a random number generator. The pressure of the system was isotropically (compressibility of $4.5 \times 10^{-5}$ bar$^{-1}$) coupled to a barostat at 1 bar using the Parrinello–Rahman method (Parrinello & Rahman, 1981). The position restrained dynamics was run for 100 ps during both temperature and pressure equilibration. Simulation was run with a time step of 2 fs, while the atomic coordinates and velocities were saved for every 2 ps. Production simulation was run for a period of 40 ns. Additionally, dTMP and dGMP-bound StTMK simulations were run at 80°C for 40 ns using the same protocol. The nucleotides were docked into the binding pocket in StTMK using AutoDock (Morris et al., 2009). CCPN-NMR Acpype (Sousa da Silva & Vranken, 2012) was used to generate the dGMP and dTMP topologies.

The g_rms module of GROMACS was used to obtain the distance between selected residues. The variation in volume of the active site pocket over the trajectory was observed using MDpocket (Schmidtke, Bidon-Chanal, Luque, & Barril, 2011). The dynamic behavior of only the active site residues was studied in detail using distance-based parameters which do not require structural superposition, as described by Münz and co-workers (Münz, Hein, & Biggin, 2012). The selected residues in each snapshot in the trajectory were described by a distance matrix where the element $d_{ij}$ represents the distance between the Ca atoms of residues $i$ and $j$. The extent of pair-wise fluctuations of each residue pair over the trajectory was described by the fluctuation matrix. The element $F_{ij}$ contains the variance of Ca distances of residues $i$ and $j$ calculated over $K$ snapshots from the trajectory.
Cloning, expression, and purification

Based on insights from the large-scale sequence analysis, substrate binding studies were carried out using StTMK as a test case. Full length DNA sequence of StTMK (protein sequence length: 213) was amplified using primers (5’-GGCATATGATGAAAAAGGGAGTTCTAATAGC-3’) and (5’-CCGGATCCTCAAAAGCTATTATCA-ATTAACTCACC-3’) by PCR (polymerase chain reaction). The amplified fragment was cloned in the Escherichia coli expression vetorpET-21a. The resulting StTMK plasmid was transformed into E. coli BL21RIL. A single colony was inoculated for primary culture and 1% of primary inoculums was used for secondary inoculation. The culture was grown overnight at 310 K in 2 L of LB (Luria Broth) media having 100 μg/mL ampicilin concentration. The culture was harvested by centrifugation (12,000 rpm for 2 min) and the cell pellet resuspended in a buffer of 20 mM Tris–HCl pH 8 and 50 mM NaCl and then lysed by sonication. The supernatant was obtained through centrifugation (4°C, 15,000 rpm, 30 min) and heat treated for 13 min at 70°C. The soup with denatured proteins was subjected to centrifugation (4°C, 15,000 rpm, 30 min). The supernatant was desalted using a Sephadex G-25 (GE Healthcare Biosciences) column equilibrated with 20 mM Tris–HCl atpH 8.0. Then the fraction was passed through a Sepharose S cation-exchange column (GE Healthcare Biosciences) equilibrated with 20 mM Tris–HCl buffer pH 8 and the protein was eluted using a linear gradient of 0–500 mM NaCl in 20 mM Tris–HCl pH 8.0. The final concentration of the protein was around 40 mg/mL.

Results and discussion

Sequences of TMKs of known three-dimensional structures were analyzed to determine the VFRs responsible for the variations in specificity and oligomerization mode among the homologs. A structure-based sequence alignment was built using TMKs from Plasmodium falciparum (PfTMK, PDB-id: 2wwf), human (HsTMK, PDB-id: 1e2d), E. coli (ETMK, PDB-id: 4tmn), Sulfolobus tokodaii (StTMK, PDB-id: 2plr), Staphylococcus aureus (SaTMK, PDB-id: 2cgj), Vibrio cholerae (VcTMK, PDB-id: 3l8v), yeast (yTMK, PDB-id: 3tnk), Mycobacterium tuberculosis (MtTMK, PDB-id: 1n5i) and Vaccinia virus (VTMK, PDB-id: 2v54). To detect if the VFRs are related by co-evolution, a larger alignment was constructed using 1635 TMK sequences from Uniprot (including the sequences with known three-dimensional structure) and MI analysis was performed (see Materials and Methods). Residue pairs associated with high MI scores (>6.5, Supplementary Figure S1) are believed to have co-evolved and are of functional importance to the protein. A flowchart depicting the method for determining VFRs is provided in Supplementary Figure S2.

ATP-binding pocket in TMKs

The TMK active site is comprised of two different sub-sites for ATP (phosphate donor) and dTMP (phosphate acceptor) binding, respectively (Figure 3(A)). There are differences between the two sub-sites in terms of geometry, conservation of interactions, and specificity. The ATP-binding pocket is found to be shallower than the dTMP-binding pocket. In HsTMK, the former pocket has a depth of 11 Å compared to a depth of 28 Å of the dTMP-binding pocket. A similar trend can be seen in other TMKs. The ATP-binding cavity is comprised of the β5/α8 loop, the conserved P-loop, and a section from the N-terminus of the Lid region. The P-loop is engaged
in interactions with the β and γ-phosphates of the donor nucleotide. The base moiety of the phosphate donor is held in place by a conserved cation–π interaction with an Arg from the Lid region (Figure 3(B)), except in case of yTMK where the Arg is replaced by a Glu (E142) which interacts with adenine via C–H…π interaction (Lavie et al., 1998). In some cases, the base additionally interacts with main chain atoms from the β5/α8 loop. For example, the amino group on C6 of adenine interacts with the backbone carbonyl group of K182 in HsTMK, but not in ETMK where the β5/α8 loop has moved away from the adenine moiety. This interaction would be disrupted in case of guanine-based nucleotides, which contain a carbonyl group on C6. Overall, there is only one conserved interaction between the Lid region (Arg or Glu residue) and the base of the phosphate donor group in case of both purine and pyrimidine nucleotides. Thus, the absence of strong directed interactions between the base and the protein residues is responsible for the broad specificity in the phosphate donor (Table 1A).

dTMP binding cavity in TMKs

Comparison of the percentage enzyme activities of TMK toward the phosphate donor and acceptor (Tables 1A and 1B) shows that, overall, the phosphate acceptor pocket has higher specificity compared to the donor pocket. This may be attributed to the increased depth of the dTMP-binding pocket compared to the ATP-binding pocket, similar to a previous study which attributed the broad specificity in vaccinia H1-related phosphatase to a shallow pocket (Yuvaniyama, Denu, Dixon, & Saper, 1996). However, the observed differences in substrate specificity for the phosphate acceptor (Table 1B) can be explained by residue substitutions at the active site pocket which, in turn, lead to differences in specific

Table 2A. Interactions at the dTMP-binding site that influence substrate specificity in TMKs (all interactions mentioned below are hydrogen bonds).

| Location                      | Interacting partners                              | Description                                                                 | Effect on specificity |
|-------------------------------|---------------------------------------------------|-----------------------------------------------------------------------------|-----------------------|
| Rear of the phosphate         | Interaction among pocket residues                  | Residue in α4 helix interacts with α3 helix and closes rear side of pocket  | Narrow                |
| acceptor cavity               |                                                   | Residue in α4 helix doesn’t interact with nucleobase or α3 helix, pocket is wider | Broad                 |
| Roof of the phosphate         | Nucleobase of phosphate acceptor interacts directly with protein | Residue in α4 helix interacts with base                                     | Narrow                |
| acceptor cavity               | Water-mediated interaction between nucleobase and protein | Strong water network in presence of a Lys/Arg in α4/α5 loop                 | Narrow                |
| Base of the phosphate         | Nucleobase of phosphate acceptor interacts directly with protein | Weak water network and can be displaced by larger base                       | Broad                 |
| acceptor cavity               |                                                   | Conserved Arg in α3 helix interacts with base                               | Narrow                |
interactions. Firstly, the number and strength of specific interactions between the protein and the acceptor nucleotide determine whether the protein is highly specific or not. Secondly, substitutions of pocket residues interacting with each other modulate the size of the pocket and hence control the size of nucleotide that can bind. Table 2A lists the different types of interactions at the dTMP-binding site that influence substrate specificity in TMKs, while Table 2B shows the combination of the aforementioned interactions influencing substrate specificity in individual TMKs.

**Co-evolving VFRs modulate specificity of the dTMP-binding pocket through changes in protein–substrate interactions**

In general, the nucleobase of the phosphate acceptor is found to interact with a conserved aromatic residue in α3 helix via π–π stacking. Additionally, a conserved pattern of directed hydrogen bond interactions (at sites marked by ‘a’, ‘b’, and ‘c’ in Figure 4(A)) is observed between the nucleobase and the VFRs in dTMP-bound TMK structures. The interactions at sites ‘a’ and ‘b’ occur with the roof of the cavity, while the site ‘c’ interacts with the cavity base. The sites ‘a’ and ‘b’ can interact either directly with the side chains of the active site residues [e.g. SaTMK, Pseudomonas TMK (PDB-id: 4e5u), ETMK (Figure 4(B))] or via a network of hydrogen-bonded water molecules [e.g. HsTMK(for PDB-id: 1E2D, HOH 2276 [O]...Tyr 151 [OH] = 2.91 Å), PřTMK (for PDB-id: 2WWF, HOH 2119 [O]...Tyr 153 [OH] = 2.69 Å)]. In the former case, the cavity is highly specific to dTMP and the residue succeeding the conserved Tyr in α4 helix (Figure 1(B)) is predominantly responsible for interactions with sites ‘a’ and ‘b.’ For water-mediated interactions at sites ‘a’ and ‘b,’ a strong water network formed due to a large number of hydrogen-bonded interactions among each other and with the nearby residues leads to a highly specific pocket (e.g. HsTMK, Figure 4(C), Supplementary Table S1). On the other hand, if the water molecules are not interconnected by a large number of hydrogen bonds (a weak network), they may be displaced to accommodate larger nucleotides (e.g. PřTMK, Figure 4(D) and (E), Supplementary Table S2). In HsTMK, K109 in α4/α5 loop forms a triad of interacting residues via polar interactions with Y56 (α2 helix) and Y151 (Lid region). Thus, K109 restricts the flexibility of Y151 which interacts with the site ‘a’ on the nucleobase. Additionally, the network of water molecules at the dTMP-binding site is held in place by this interacting triad (Figure 4(C)). In case of PřTMK, the Lys has been substituted by Ala111. The interacting water molecules exhibit less number of interactions with the cavity residues (Figure 4(D)), and can be substituted to accommodate dGMP (Figure 4(E)).

A previous study on PřTMK (Kandeel & Kitade, 2008) noted a large reduction in dGMP phosphorylation in the A111 K mutant. However, the role of the Lys in maintaining the water network and thereby imparting specificity has not been analyzed before. Co-evolution analysis showed that K109 (in HsTMK) in the roof of the dTMP-binding cavity has high MI signal with R150 and F156 (Table 3). The bulky residues at positions 150 and 156 interact with Y151 in HsTMK, holding it rigidly and maintaining the interactions with K109-mediated water network. This imparts rigidity to the pocket as a whole, making HsTMK highly specific. In PřTMK, all the positions have been substituted by smaller residues (Table 3), leading to lesser interactions, weaker water network, and a more flexible pocket that can accept purine nucleotides.

Analysis of TMK structures shows that the site ‘c’ in the acceptor nucleotide (Figure 4(A)) interacts with a conserved Arg (in α3 helix) in the cavity base, which is held in appropriate conformation via hydrogen bonds with a residue in either α4 (e.g. ETMK) or α5 (e.g. HsTMK, SaTMK, PřTMK) helix (Figure 4(F)). In StTMK, this Arg has been replaced by F71. F71 is part of a cluster of aromatic residues (Figure 4(G)), composed of W39 in β2/α2 loop, Y116 and F118 in α5/β4 loop, which interact via C–H–O, C–H–π, and π–π interactions. MI analysis (Table 3) pointed to these residue substitutions to be a result of co-evolution. A previous study had observed high flexibility at active sites that lack strong directed interactions (Gunasekaran & Nussinov, 2007) and the high probability of occurrence of Trp at
such sites. The substitution of polar Arg by the aromatic residue Phe and the formation of an aromatic cluster remove the strong polar interactions at the base of the dTMP-binding pocket in StTMK and are expected to increase the pocket flexibility.

Substrate binding is also regulated by changes in the interactions among co-evolving residues lining the dTMP-binding cavity

The size of the dTMP-binding pocket is regulated by interactions among the residues lining the pocket. In
some proteins, the residue in α4 helix may not interact with the nucleobase but with another residue in α3 helix, thereby imparting specificity by closing the rear of the pocket and preventing larger nucleotides from binding. For example, in HsTMK, the rear of the acceptor binding pocket is closed by co-evolving residues H69 (α3) and T106 (α4) interacting via a hydrogen bond and is further covered by the bulky F112 and W116, resulting in a rigid pocket (Table 3 and Figure 5(A)). On the other hand, the residue may interact neither with the base nor with the α3 helix, resulting in a larger pocket and broader specificity as observed in VTMK (Caillat et al., 2008). The residues at the rear of the dTMP-binding pocket in VTMK are N65, A102, A107, and L111 (Figure 5(B)), of which the co-evolving pair N65 and A102 do not interact. The lower contact area among the α3 and α5 helices in each monomer of the orthogonal dimer results in a larger acceptor binding pocket in VTMK. As a result, nucleotides with bulkier pyrimidine (with large groups attached at the C5 position) and purine groups can also bind in place of the conventional pyrimidine nucleotides at the acceptor site (Caillat et al., 2008). In case of ETMK, the corresponding residues V71 and Q109 do not interact with each other. However, Q109 directly interacts with the nucleobase and its side chain is oriented such that the rear of the pocket is closed, thereby leading to high specificity in ETMK (Figure 4(B)).

Previously, an aromatic cluster at the base of the dTMP-binding cavity in StTMK has been implicated to result in a flexible dTMP-binding pocket. Comparison of TMK structures depicts an additional feature that may result in a larger phosphate acceptor pocket in StTMK. The C-terminus of the α2 helix in StTMK is oriented away from the core of the protein (Figure 5(C)), whereas it is oriented toward the core in the other TMKs (Figure 5(D)). This can be understood in terms of the interactions of the α2 helix with the neighboring secondary structures, namely the Lid region and the β2/α2 loop. In HsTMK, K109 has co-evolved (Table 3) with Y56 (α2 helix), R150 (Lid), and F156 (α7 helix). Y56 interacts with both K109 and R150 via hydrogen bonds. This interaction brings the α2 helix closer to the Lid (Figures 4(C) and 5(D)) and closes the roof of the dTMP-binding pocket. In StTMK, the equivalent residues are R105, K54, Q149, and Y168. K54 does not interact with R105 and Q149. Instead, R105 interacts with Y168. As a result, α2 helix moves away from the Lid and the roof of the dTMP-binding cavity is open. Moreover, a positively charged residue is present either

![Table 3. Co-evolving residues involved in functional roles in TMK detected by MI analysis (MI scores >6.5 are significant (Simonetti et al., 2013)).](image-url)

| Sr. no. | Co-evolving residue pair | TMK       | Residue 1 location | Residue 2 location | MI   | Description                                                                 |
|--------|--------------------------|-----------|--------------------|--------------------|------|-----------------------------------------------------------------------------|
| 1      | K109 – R150             | HsTMK     | Roof of the dTMP-binding cavity | Roof of the dTMP-binding cavity | 10.0 | These residue pairs modulate specificity by controlling the flexibility of the pocket and strength of the water-mediated interaction network |
| 2      | F42 – R76               | HsTMK     | Base of dTMP-binding cavity | Base of dTMP-binding cavity | 8.43 | Presence of aromatic cluster in StTMK instead of the residues interacting through polar interactions in HsTMK may lead to increased flexibility in StTMK |
| 3      | H69 – T106              | VTMK     | Rear of the dTMP-binding cavity | Rear of the dTMP-binding cavity | 14.2 | The presence or absence of interactions between the residues in each pair controls the size of the dTMP-binding pocket and hence specificity. |
| 4      | K109 – Y56              | StTMK     | Roof of dTMP-binding cavity | α2 helix           | 16   | Interactions between these residues regulate the orientation of α2 helix and the size of the TMP-binding pocket. |
| 5      | R45 – S54               | StTMK     | Roof of dTMP-binding cavity | α7 helix           | 12.4 | Interactions between these two residues regulate the size of the dTMP-binding pocket. |
| 6      | W77 – P120              | MtTMK     | Dimeric interface | Dimeric interface | 8.04 | Modulated the type of dimeric assembly (anti-parallel in HsTMK vs. parallel in MtTMK) |
| 7      | W116 – P120              | VTMK     | Dimeric interface | Dimeric interface | 7.3  | Modulated the type of dimeric assembly (anti-parallel in HsTMK vs. orthogonal in VTMK) |
in the α2 helix or the preceding β2/α2 loop and the two positions are found to have co-evolved (Table 3). In ETMK, G43 (EPGG43) in β2/α2 loop has co-evolved and interacts with R51 in α2 helix, while in HsTMK there is an Arg in the β2/α2 loop (FPER45), and not in α2 helix, that has co-evolved and interacts with S54 in α2 helix. The positively charged residue is responsible for hydrogen bonded interactions between the β2/α2 loop and α2 helix. In StTMK, there is no Arg in either the α2 helix or the preceding loop. The absence of directed interactions of the α2 helix with the Lid and β2/α2 loop causes the α2 helix (colored magenta in Figure 1(C)) to

Figure 5. Interactions among amino acid residues modulate the size of the dTMP-binding pocket. Hydrogen bond interactions (black dotted line) at the rear of the pocket lead to its closure in (A) HsTMK but not in (B) VTMK. Orientation of the α2-helix (C) away from the protein core in StTMK and (D) toward the protein core in HsTMK also controls the pocket size. The angle between the α2 helix and the preceding β2/α2 loop is 89.2° in StTMK and 60.2° in HsTMK, leading to an increase in distance (depicted by a double-headed arrow) between the α2-helix and the Lid in StTMK compared to HsTMK (Colour online).
orient away from the core of the protein, resulting in a larger dTMP-binding pocket.

**Co-evolving residues at the dimeric interface regulate the type of dimer arrangement in TMKs**

Three-dimensional structures of TMK exhibit differences in the mode of dimerization. Most TMKs were crystallized as an anti-parallel dimer (Figure 2(A)), for example, PfTMK, StTMK, ETMK, and HsTMK. Here, the N-terminus of the α3 helix of a monomer interacts with the C-terminus of the α3 helix from the other monomer. The two monomers are related by a rotation about X-axis. Among these, the dimeric interface in StTMK is very strong with 13 hydrogen bonds and a total interface area of 2204 Å². The second mode is the parallel stacking observed in MtTMK, with 12 hydrogen bonds at the interface and a total interface area of 1628 Å². The N-terminus of α3 helix from both the monomers is found to interact with each other. The monomers are related by a rotation along Z-axis (Figure 2(B)). The third type of arrangement is observed in VTMK, where the monomers are packed in an orthogonal manner (4 hydrogen bonds at the interface and 1763 Å² interface area). The angle between the α3 helices of the two monomers is approximately 80° (Caillat et al., 2008). Here, the monomers are related by two rotations, first around X-axis and then around Y-axis (Figure 2(C)). The orthogonal packing mode in 2v54 displays an appreciable decrease in the surface area and the number of hydrogen bonds at the dimer interface compared to the anti-parallel arrangement. MI analysis showed that co-evolving residues also play a role in the dimer arrangement, as described below.

**Parallel dimer in MtTMK results from a π-helix at the interface**

The parallel packing in MtTMK can be attributed to the presence of the segment 126ARL128 (Figure 6(A)), which is part of a π-helix (122RIEFARL128, (Kumar & Bansal, 2015)) and has a larger radius (2.78 Å) than the preceding α5 helix (2.28 Å) (Kumar & Bansal, 2012). A126 and R127 show a high MI signal with A75 in the C-terminus of α3 (Table 3). The equivalent residues in the anti-parallel HsTMK dimer are P120, D121, and W77. The broadened π-helix in MtTMK leads to the correlated substitution of the bulky W77 in the C-terminus of α3 in HsTMK by A75 in MtTMK to avoid steric hindrance. A75 in MtTMK does not interact with the other monomer of the parallel dimer, while W77 in anti-parallel dimer interacts with H66 from the other dimer.

Structure-based sequence alignment showed that another Trp (W119 in MtTMK) is present in the N-terminus of α5 helix at the dimer interface and interacts with the other chain in HsTMK, StTMK, PfTMK, and MtTMK. If MtTMK were to form an anti-parallel dimer, there would be steric clashes between W119 (from one monomer) and R127 (from the other monomer), whose large side chain points outward from the core of the monomer. In the parallel packing mode, the side chain of R127 (in one monomer) forms bi-dentate hydrogen bonds with E50 in the other monomer while

![Figure 6](https://example.com/figure6.png)

Figure 6. Variations in residues at the dimeric interface which lead to parallel and orthogonal dimers in MtTMK and VTMK respectively. (A) The π-helix (in red) and R127 (a part of π-helix) which protrudes out of the helix in MtTMK. (B) Y115 pointing out from the protein core which results in orthogonal dimer in VTMK (Colour online).
H56 (in the first monomer) interacts via hydrogen bonding with the W119 side chain in the second monomer.

**Orthogonal dimer in VTMK results from a bulky Tyr at the interface**

The orthogonal dimer in VTMK is attributed to Y115 (Figure 6(B)), present in the middle of α5 helix at the interface, whose side chain points outward from the core of the VTMK molecule and would cause steric clashes in a conventional anti-parallel dimer (Caillat et al., 2008). Y115 is found to have high MI signal with L111. The corresponding co-evolving pair in HsTMK is P120-W116. The side chain of the smaller P120 (in HsTMK) points away from the interface, while W116 (present in the interface at the N-terminus of α5 helix in HsTMK, StTMK, MtTMK and PfTMK) interacts with the other monomer. These interactions result in the more stable orthogonal dimer in VTMK (Tm = 42.5°C) (Caillat et al., 2008). The orthogonal dimer in VTMK is attributed to Y115 (Figure 6(B)), present in the middle of α5 helix at the interface, whose side chain points outward from the core of the VTMK molecule and would cause steric clashes in a conventional anti-parallel dimer (Caillat et al., 2008). Y115 is found to have high MI signal with L111. The corresponding co-evolving pair in HsTMK is P120-W116. The side chain of the smaller P120 (in HsTMK) points away from the interface, while W116 (present in the interface at the N-terminus of α5 helix in HsTMK, StTMK, MtTMK and PfTMK) interacts with the other monomer. These interactions result in the more stable anti-parallel (Tm = 52°C) and parallel dimers (Tm > 65°C) compared to the orthogonal dimer in VTMK (Tm = 42.5°C) (Caillat et al., 2008).

**Analysis of interaction networks of VFRs**

Previous analyses have revealed several variable but important residues that modulate the specificity and oligomerization mode in TMKs. The former occur either at the roof, rear, or base of the dTMP-binding cavity, while the latter occur at the dimer interface. Sub-networks for the VFRs and their interaction score (Is), which is proportional to the strength of the interaction between the concerned residues. Sub-networks for the VFRs and their first nearest neighbors at each of the relevant sites (roof, base, and rear of dTMP-binding pocket) were extracted from the RINs. The corresponding sub-networks from different TMKs were compared through three network parameters (Table 4) namely, the number of edges (E), the ratio (E/N) of the number of edges (E) to the number of nodes (N) in the sub-network, and the average number of neighbors (<N>) of a node. Higher values of E/N indicate enhanced interactions among the nodes (residues) in the network, while higher value of <N> is an indicator of an increase in the number of interacting partners of any node (residue) in the network. Evidently, higher values of E, E/N, and <N> are indicative of a stronger network. Comparison of the network parameters for interacting residues at the roof of the dTMP-binding cavity showed the presence of a stronger sub-network in HsTMK than in PfTMK. Similar observations were made for the base and rear of the dTMP binding cavity network in HsTMK when compared with StTMK and VTMK (Table 4). This clearly shows that the previously described compensatory mutations in VFRs result in higher flexibility of the dTMP binding site in StTMK, VTMK, and PfTMK compared to HsTMK.

**Case studies**

Based on the sequence and structure analyses of TMKs, VFRs which influence the specificity and dimerization mode of well-characterized TMKs have been detected. Several of these VFRs are related by co-evolution. As a further extension of our analysis, two case studies are reported. Firstly, our study provides a residue-level explanation of the narrow specificity of BmTMK (Doharey et al., 2014) for which no three-dimensional structure is available. Secondly, binding assays have been carried out for StTMK, since sequence and structure analyses show several residue substitutions and changes in secondary structure alignment that are expected to increase the flexibility of the dTMP-binding pocket in StTMK which may, in turn, influence its substrate binding (Bone, Silen, & Agard, 1989).

**Case study 1: BmTMK**

Previous experiments have shown that BmTMK is highly specific to dTMP as phosphate acceptor (Doharey et al., 2014). BmTMK shares 43% sequence identity with HsTMK. The sequence alignment of BmTMK with HsTMK (Figure 7) depicts the residues at the base (blue triangle), roof (brown triangle), and rear (black inverted

| Location of sub-network | TMK | E (Edge) | E/N  | <N> | Relative strength of the network |
|-------------------------|-----|----------|------|-----|---------------------------------|
| dTMP pocket roof        | HsTMK | 95       | 3.8  | 3.60 | HsTMK has stronger network at roof of dTMP pocket |
|                         | PfTMK | 62       | 3.44 | 3.33 |                                  |
| dTMP pocket base        | HsTMK | 87       | 3.62 | 3.41 | HsTMK has stronger network at base of dTMP pocket |
|                         | StTMK | 72       | 3.13 | 3.13 |                                  |
| dTMP pocket rear        | HsTMK | 53       | 3.78 | 3.57 | HsTMK has stronger network at rear of dTMP pocket |
|                         | VTMK  | 19       | 2.37 | 2.00 |                                  |
triangle) of the dTMP-binding cavity which impart narrow specificity to HsTMK through appropriate interactions at the dTMP-binding site. The equivalent sites in BmTMK are occupied by either identical or similar residues in most of the cases. This was also verified by superposing the structure of HsTMK on a modeled structure of BmTMK. Hence, it may be argued that the residues responsible for the narrow specificity in HsTMK also lead to the narrow specificity in BmTMK. Moreover, the modeled structure of BmTMK neither has any π-helix or bulky residues at the end of the α5 helix as observed in MtTMK, nor does it have any Tyr (similar to VTMK) at the middle of the α5 helix. Hence, BmTMK possibly forms an anti-parallel dimer, which is also observed in protein–protein docking studies.

Case study 2: StTMK

StTMK is found to be different from other TMKs of known structure. It has approximately 30% sequence identity to other characterized TMKs, as observed from a BLAST run against the PDB. Polar interactions at the base of the dTMP-binding pocket have been substituted by hydrophobic interactions among an aromatic cluster, which are expected to increase the pocket flexibility. Moreover, the co-evolving residues that engage the α2 helix in interactions with the preceding β2/α2 loop and the Lid region in other TMKs have been substituted by non-interacting residues in StTMK. These lead to a larger dTMP-binding pocket in StTMK compared to the other TMKs of known structures and the pocket is predicted to accommodate different nucleotides. To check if there were prior examples of proteins which bind to different nucleotides and have aromatic residues lining their binding site, a search was conducted in the PLIC database (Anand, Nagarajan, Mukherjee, & Chandra, 2014). Our search revealed that the binding site of Ectonucleotide pyrophosphate 1 (Enpp1) protein from mouse (Kato et al., 2012) contains a cluster of aromatic residues and can accommodate both purine and pyrimidine nucleotides (PDB-ids: 4gtw, 4gtx, 4gty and 4gtz). To study how the VFRs affect the flexibility of the dTMP-binding site in StTMK, 40 ns MD simulations of apo-StTMK were carried out and compared to the dynamics of the well-characterized, highly specific HsTMK (apo-form). The simulations were performed at their respective ambient temperatures (80°C for StTMK and 27°C for HsTMK).

Comparison of MD simulations of HsTMK and StTMK indicates a flexible dTMP-binding pocket in StTMK

Our sequence analysis has shown that the ATP-binding site is relatively more conserved, while there are variations in the roof and base of the dTMP-binding cavity between HsTMK and StTMK. To understand their influence on protein dynamics, MD simulations were performed for 40 ns each on the apo forms of HsTMK and StTMK. The percentage coverage of the Ramachandran map was obtained for 10,000 snapshots (extracted every 4 ps from the trajectory, Supplementary Figure S3) to ascertain the convergence of conformational sampling, similar to a previous study (Scott, Alonso, Sato, Fersht, & Daggett, 2007). In all cases, the coverage increased steadily for the first 1000 frames (corresponding to 4 ns) and showed a small increase beyond that point, signifying sufficient conformational sampling in 40 ns. Subsequently, fluctuation matrices were calculated (see Materials and Methods) from the trajectories. These matrices represent the variation in the Cα–Cα distances between the residues in these sites over the MD trajectory. Calculations for the ATP sub-site involved residues from the P-loop and the conserved Arg from the
Figure 8. Comparison of the active site in StTMK and HsTMK from MD simulations shows a larger and more flexible dTMP-binding pocket in StTMK. (A) Overall fluctuation of the ATP and dTMP-binding sites are denoted by bars while the CC values denote the extent of similarity in dynamics at the corresponding sites, (B) distribution of the distance between the Lid region and the α2 helix in StTMK (red curve) and HsTMK (red curve) depicts a larger dTMP-binding pocket in StTMK, (C) distribution of the volume of the dTMP-binding pocket (in black) in StTMK throughout the MD trajectory (the red curve is the smoothened plot of the same), (D) distribution of the volume of the dTMP-binding pocket in HsTMK throughout the trajectory. The values are smaller than those observed for StTMK. (The jump in the distribution for a short duration after the 10,000th snapshot is due to the flexibility of the side chains of residues lining the active site pocket in the absence of substrate), (E) distribution of the distance between the ND1 atom of H69 and the OG1 atom of T106, depicting closure of the rear of the dTMP-binding pocket, (F) distribution of the overall RMSD of apo (black), dTMP-bound (blue) and dGMP-bound (red) StTMK during MD. The apo-protein has a broader distribution with the peak around 2.5 Å, depicting higher flexibility. The peaks are shifted to lower RMSD values for the holo structures, implying reduction in flexibility upon ligand binding. Binding affinities of (G) dTMP and (H) dGMP to StTMK obtained from MST demonstrate that the larger pocket in StTMK can accept both dTMP and dGMP (Colour online).
Lid region that interacts with the nucleobase of ATP, while co-evolving residues in the base and roof were included for the dTMP-binding site (see Materials and Methods). The overall fluctuation (OF, see Materials and Methods) was calculated for each of these sub-sites. The OF at the ATP sub-site, roof and base of the dTMP-binding cavity were .37, .64, and .41 Å in HsTMK and .51, 1.03, and .66 Å in StTMK, respectively (Figure 8(A)). The correlation coefficient (CC) of the sub-site fluctuations between HsTMK and StTMK were also calculated to get an idea of the similarity in their dynamics. The CC values for the dynamics at the ATP sub-site, roof and base of the dTMP-binding cavity were, respectively, .95, .02, and .55. The above results imply that, overall, the StTMK active site is more flexible (higher OF) than HsTMK. The ATP-binding site is more rigid and its dynamics is similar in both the organisms (high CC and low overall OF). The base of the dTMP-binding cavity is more flexible in StTMK and there is a lower match in dynamics with HsTMK (CC .55). This can be attributed to the presence of the aromatic cluster at the base of the pocket. Finally, the roof of the dTMP-binding cavity is the most flexible in StTMK (1.03 Å) and there is hardly any correlation in dynamics with HsTMK (CC .02). This is because the co-evolving residues in the α2 helix and the Lid region do not interact in StTMK, which leads to a wider dTMP-cavity.

To get an idea about the relative separation between the Lid region and α2 helix in StTMK and HsTMK, the Ca–Ca distance between a residue each on the Lid region and α2 helix was calculated over the entire trajectory and the distribution obtained (Figure 8(B)). Residue pairs Q149-K54 in StTMK and R150-Y56 in HsTMK were chosen. The distance exhibits a narrow distribution with peak about .76 nm for HsTMK (black curve) and a wider distribution about 1.86 nm for StTMK (red curve). Moreover, the average volume of the dTMP-binding pocket calculated from 20,000 snapshots in the StTMK trajectory is 1280 Å³ compared to 666 Å³ in HsTMK and the distributions of their pocket volumes (Figure 8(C) and (D)) show that the dTMP-binding pocket in StTMK is larger than in HsTMK. Incidentally, the distance between the ND1 atom of H69 and the OG1 atom of T106 in HsTMK varies between 2.6 and 4 Å over the trajectory with the peak of the distribution at 2.9 Å (Figure 8(E)). The two residues interact through an average of one hydrogen bond over the trajectory, thereby closing the rear of the dTMP-binding pocket which has also been noted in the crystal structure.

Subsequently, dTMP and dGMP-bound StTMK simulations were performed to obtain insights into the binding of pyrimidine and purine nucleotides in the larger pocket in StTMK. Overall, RMSD of dTMP and dGMP-bound StTMK is less than the apo-protein (Figure 8(F)), signifying reduction in the active site dynamics upon ligand binding. Moreover, the OF of the dTMP-binding site is also reduced in the dTMP (OF: .56 Å) and dGMP-bound (OF: .56 Å) simulations compared to the apo-protein (OF: .66 Å). dTMP and dGMP form an average of 5.4 and 6 hydrogen bonds, respectively, to the protein over the course of simulations. These indicate that the dTMP-binding pocket in StTMK can accommodate both the nucleotides.

**Binding assays of recombinant StTMK with dTMP (pyrimidine) and dGMP (purine) nucleotides**

The enhanced structural plasticity of the dTMP-binding pocket is predicted to influence substrate binding in StTMK. To identify whether the dTMP-binding pocket can accommodate both purine and pyrimidine nucleotides, the binding affinity ($K_d$) of different nucleotides (dTMP and dGMP) to StTMK was studied using MST (see Materials and Methods). The experiments yielded $K_d$ values of 11.6 ± 1.4 μM for dTMP (Figure 8(G)) and 66.0 ± 5.4 μM for dGMP (Figure 8(H)). This indicates that the more flexible dTMP-binding pocket in StTMK can accommodate nucleotides of different sizes. Supplementary Figure S4 depicts the purity of the StTMK sample used for our studies.

**Conclusion**

Naturally occurring TMKs exhibit variations in substrate specificity and oligomerization mode. Our study essentially provides an understanding of these variations on the basis of observed sequence insertions and mutations, a number of which occur at or near the active site and at the dimeric interface. It has been demonstrated that firstly, these mutations are not random, but have been selected through evolution and often occur in conjunction with other mutations. These co-evolved residues differ in their bulk (small alkyl side chains vs. bulky aromatics) or the type of interactions they can form (hydrogen bonds, van der Waals interactions), which affect their mobility. This, in turn, plays a role in the flexibility and size of the active site pocket, thereby influencing substrate specificity and function or the type of oligomeric assembly that can occur without any steric clashes. Secondly, these VFRs can be targeted to change substrate specificity in a given TMK or design molecules which bind better to particular TMK homolog. The former has applications in industry (Takahashi, Liu, & Liu, 2006) where enzymes with altered specificity will be of great use (Piotukh et al., 2011), while the latter has implications in the design of novel inhibitors. A previous study on PfTMK depicted the narrowing of substrate specificity upon mutating a residue (Kandeel & Kitade, 2008) which our study has detected as a VFR. Thirdly, our method was used to explain the residues responsible
for the high specificity in BmTMK and understand the substrate binding preference of StTMK, whose active site is substantially different from other TMKs in terms of residue composition and three-dimensional structure. Biochemical characterization of the protein revealed that the protein has high binding affinity to its cognate substrate dTMP and lower but appreciable affinity to the larger dGMP nucleotide. The enzyme can be a valued addition to the TMK repertoire. Finally, our study can be leveraged to obtain some idea about the flexibility of the binding site and hence the substrate binding preference in uncharacterized TMKs based on the residue composition at the active site.

Acknowledgements
The authors gratefully acknowledge DBT (Department of Biotechnology, Govt. of India, DST (Department of Science and Technology, Govt. of India and the facilities offered by the Centre of Excellence and Interactive Graphics Facilities (IISc), Bangalore. AB thanks IISc for a fellowship and SPing-8, Japan, for high resolution data collection and StTMK plasmid.

Disclosure statement
No potential conflict of interest was reported by the authors.

Supplemental data
The supplementary material for this article is available online at http://dx.doi.org/10.1080/07391102.2016.1207563.

References
Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215, 403–410. doi:10.1016/S0022-2836(05)80360-2
Anand, P., Nagarajan, D., Mukherjee, S., & Chandra, N. (2014). PLIC: Protein–ligand interaction clusters. Database (Oxford), 2014, 1–9. doi:10.1093/database/bau029
Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22, 195–201. doi:10.1093/bioinformatics/btl770
Assenov, Y., Ramirez, F., Schelhorn, S.-E., Lengauer, T., & Albrecht, M. (2008). Computing topological parameters of biological networks. Bioinformatics, 24, 282–284. doi:10.1093/bioinformatics/btn554
Berendsen, H. J. C., Grigera, J. R., & Straatsma, T. P. (1987). The missing term in effective pair potentials. The Journal of Physical Chemistry, 91, 6269–6271. doi:10.1021/j100308a038
Bone, R., Silen, J. L., & Agard, D. A. (1989). Structural plasticity broadens the specificity of an engineered protease. Nature, 339, 191–195. doi:10.1038/339191a0
Buslje, C. M., Santos, J., Delfino, J. M., & Nielsen, M. (2009). Correction for phylogeny, small number of observations and data redundancy improves the identification of coevolving amino acid pairs using mutual information. Bioinformatics, 25, 1125–1131. doi:10.1093/bioinformatics/btp135
Caillat, C., Topalis, D., Agrofoglio, L. A., Pochet, S., Balzarini, J., Deville-Bonne, D., & Meyer, P. (2008). Crystal structure of poxvirus thymidylate kinase: An unexpected dimerization has implications for antiviral therapy. Proceedings of the National Academy of Sciences, 105, 16900–16905. doi:10.1073/pnas.0804525105
Carlgvist, P., Svedendahl, M., Branneby, C., Hult, K., Brinck, T., & Berghlund, P. (2005). Exploring the active-site of a rationally redesigned lipase for catalysis of michael-type additions. ChemBioChem, 6, 331–336. doi:10.1002/cbic.200400213
Chakrabarti, S., & Panchenko, A. R. (2009). Coevolution in defining the functional specificity. Proteins: Structure, Function and Bioinformatics, 75, 231–240. doi:10.1002/prot.22239
Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics, 98, 10089–10092. doi:10.1063/1.464397
Doharey, P. K., Suthar, M. K., Verma, A., Kumar, V., Yadav, S., Balaramnavar, V. M., … Saxena, J. K. (2014). Molecular cloning and characterization of Brugia malayi thymidylate kinase. Acta Tropica, 133, 83–92. doi:10.1016/j.actatropica.2014.02.003
Doncheva, N. T., Klein, K., Domingues, F. S., & Albrecht, M. (2011). Analyzing and visualizing residue networks of protein structures. Trends in Biochemical Sciences, 36, 179–182. doi:10.1016/j.tibs.2011.01.002
Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., … Kollman, P. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. Journal of Computational Chemistry, 24, 1999–2012. doi:10.1002/jcc.10349
Dunn, S. D., Wahl, L. M., & Gloo, G. B. (2008). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. Bioinformatics, 24, 333–340. doi:10.1093/bioinformatics/btm604
Gloo, G. B., Martin, L. C., Wahl, L. M., & Dunn, S. D. (2005). Mutual information in protein multiple sequence alignments reveals two classes of coevolving positions. Biochemistry, 44, 7156–7165. doi:10.1021/bi050293e
Gourmas, C., Evangelidis, T., Athanasopoulos, A., Mikros, E., & Sophianopoulou, V. (2015). The Aspergillus nidulans pyrimidine permease as a model for understanding the factors determining substrate binding and specificity of fungal amino acid transporters. Journal of Biological Chemistry, 290, 6141–6155. doi:10.1074/jbc.M114.612069
Gunasekaran, K., & Nussinov, R. (2007). How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. Journal of Molecular Biology, 365, 257–273. doi:10.1016/j.jmb.2006.09.062
Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry, 18, 1463–1472. doi:10.1002/(SICI)1096-987X(19970918)18:12<1463::AID-JCC4>3.0.CO;2
Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of Chemical Theory and Computation, 4, 435–447. doi:10.1021/ct070030q
Hoover, W. G. (1985). Canonical dynamics: Equilibrium phase-space distributions. Physical Review A, 31, 1695–1697. doi:10.1103/PhysRevA.31.1695
Data mining studies using thymidylate kinases 2153

Jong, A. Y., & Campbell, J. L. (1984). Characterization of Saccharomyces cerevisiae thymidylate kinase, the CDC8 gene product. General properties, kinetic analysis, and subcellular localization. Journal of Biological Chemistry, 259, 14394–14398. Retrieved from http://www.jbc.org/content/259/23/14394.long

Kandeel, M., & Kitade, Y. (2008). Molecular characterization, heterologous expression and kinetic analysis of recombinant Plasmodium falciparum thymidylate kinase. Journal of Biochemistry, 144, 245–250. doi:10.1093/jb/mvn062

Kato, K., Nishimasu, H., Okudaira, S., Mihara, E., Ishitani, R., Takagi, J., … Nureki, O. (2012). Crystal structure of Enpp1, an extracellular glycoprotein involved in bone mineralization and insulin signaling. Proceedings of the National Academy of Sciences, 109, 16876–16881. doi:10.1073/pnas.1208017109

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. Journal of Molecular Biology, 372, 774–797. doi:10.1016/j.jmb.2007.05.022

Kumar, P., & Bansal, M. (2012). HELANAL-Plus: A web server for analysis of helix geometry in protein structures. Journal of Biomolecular Structure and Dynamics, 30, 773–783. doi:10.1080/07391102.2012.689705

Kumar, P., & Bansal, M. (2015). Identification of local variations within secondary structures of proteins. Acta Crystallographica Section D, D71, 1077–1086. doi:10.1107/S1399004715003144

Kumar, P., Kailasam, S., Chakraborty, S., & Bansal, M. (2014). MolBridge: A program for identifying nonbonded interactions in small molecules and biomolecular structures. Journal of Applied Crystallography, 47, 1772–1776. doi:10.1107/S160057671401468X

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., … Higgins, D. G. (2007). Clustal W and clustal X version 2.0. Bioinformatics, 23, 2947–2948. doi:10.1093/bioinformatics/btm404

Laskowski, R. A. (2009). PDBsum new things. Nucleic Acids Research, 37(Database issue), D355–D359. doi:10.1093/nar/gkn860

Lavie, A., Konrad, M., Brundiers, R., Goody, R. S., Schlitchting, I., & Reinstein, J. (1998). Crystal structure of yeast thymidylate kinase complexed with the bisubstrate inhibitor P 1 -(5' -adenosyl) P 5 -(5' -thymidyl) pentaphosphate (TP 5 A) at 2.0 Å resolution: Implications for catalysis and AZT activation. Biochemistry, 37, 3677–3686. doi:10.1021/bi9720787

Lavie, A., Ostermann, N., Brundiers, R., Goody, R. S., Reinstein, J., Konrad, M., & Schlitchting, I. (1998). Structural basis for efficient phosphorylation of 3'-azidothymidine monophosphate by Escherichia coli thymidylate kinase. Proceedings of the National Academy of Sciences, 95, 14045–14050. doi:10.1073/pnas.95.24.14045

Lee, L. S., & Cheng, Y. (1977). Human thymidylate kinase. Purification, characterization, and kinetic behavior of the thymidylate kinase derived from chronic myelocytic leukemia. Journal of Biological Chemistry, 252, 5686–5691. Retrieved from www.jbc.org/content/252/16/5686.long

Lee, Y., Mick, J., Furdai, C., & Beamer, L. J. (2012). A coevolutionary residue network at the site of a functionally important conformational change in a phosphohexomutase enzyme family. PLoS One, 7, e38114. doi:10.1371/journal.pone.0038114

Liu, Y., & Bahar, I. (2012). Sequence evolution correlates with structural dynamics. Molecular Biology and Evolution, 29, 2253–2263. doi:10.1093/molbev/msq097

Magrane, M., & Consortium, U. (2011). UniProt knowledge-base: A hub of integrated protein data. Database (Oxford), 2011, bar009. doi:10.1093/database/bar009

Marsh, J. A., & Teichmann, S. A. (2014). Parallel dynamics and evolution: Protein conformational fluctuations and assembly reflect evolutionary changes in sequence and structure. BioEssays, 36, 209–218. doi:10.1002/bies.201300134

Martin, L. C., Gloor, G. B., Dunn, S. D., & Wahl, L. M. (2005). Using information theory to search for co-evolving residues in proteins. Bioinformatics, 21, 4116–4124. doi:10.1093/bioinformatics/bti467

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry, 30, 2785–2791. doi:10.1002/jcc.21256

Munier-Lehmann, H., Chaflotte, A., Pochet, S., & Labesse, G. (2001). Thymidylate kinase of Mycobacterium tuberculosis: A chimera sharing properties common to eukaryotic and bacterial enzymes. Protein Science, 10, 1195–1205. doi:10.1110/ps.45701

Münz, M., Hein, J., & Biggin, P. C. (2012). The role of flexibility and conformational selection in the binding promiscuity of PDZ domains. PLoS Computational Biology, 8, e1002749. doi:10.1371/journal.pcbi.1002749

Nelson, D. J., & Carter, C. E. (1969). Purification and characterization of Thymidine 5-monophosphate kinase from Escherichia coli B. Journal of Biological Chemistry, 244, 5254–5262. Retrieved from http://www.jbc.org/content/244/19/5254.long

Nosé, S. (1984). A unified formulation of the constant temperature molecular-dynamics methods. The Journal of Chemical Physics, 81, 511–519. doi:10.1063/1.447334

Ostermann, N., Schlitchting, I., Brundiers, R., Konrad, M., Reinstein, J., Veit, T., … Lavie, A. (2000). Insights into the phosphoryl transfer mechanism of human thymidylate kinase gained from crystal structures of enzyme complexes along the reaction coordinate. Structure, 8, 629–642. doi:10.1016/S0969-2126(00)00149-0

Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics, 52, 7182–7190. doi:10.1063/1.33058

Petit, C. M., & Korete, K. K. (2002). Characterization of Streptococcus pneumoniae thymidylate kinase: Steady-state kinetics of the forward reaction and isothermal titration calorimetry. Biochemical Journal, 363, 825–831. doi:10.1042/bj3630825

Piotukh, K., Geltinger, B., Heinrich, N., Gerth, F., Beyermann, M., Freund, C., & Schwarzer, D. (2011). Directed evolution of sortase A mutants with altered substrate selectivity profiles. Journal of the American Chemical Society, 133, 17536–17539. doi:10.1021/ja205630g

Reichard, P. (1988). Interactions between deoxyribonucleotide and DNA synthesis. Annual Review of Biochemistry, 57, 349–374. doi:10.1146/annurev.bi.57.070188.002025

Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F., & Schroeder, M. (2015). PLIP: Fully automated protein–ligand interaction profiler Nucleic Acids Research, 43 (Web Server Issue), W443–W447. doi:10.1093/database/bar009

Schmidtke, P., Bidon-Chanal, A., Luque, F. J., & Barril, X. (2011). MDpocket: Open-source cavity detection and characterization on molecular dynamics trajectories. Bioinformatics, 27, 3276–3285. doi:10.1093/bioinformatics/btr550
