J Mol Model (2013) 19:407–419
DOI 10.1007/s00894-012-1566-1

ORIGINAL PAPER

In silico structural and functional analysis of the human TOPK protein by structure modeling and molecular dynamics studies

Palani Kirubakaran · Muthusamy Karthikeyan · Kh. Dhanachandra Singh · Selvaraman Nagamani · Kumpati Premkumar

Received: 10 September 2011 / Accepted: 9 August 2012 / Published online: 1 September 2012

Abstract Over expression of T-lymphokine–activated killer cell–originated protein kinase (TOPK) has been associated with leukemia, myeloma tumors and various other cancers. The function and regulatory mechanism of TOPK in tumor cells remains unclear. Structural studies that could reveal the regulatory mechanism have been a challenge because of the unavailabity of TOPK's crystal structure. Hence, in this study, the 3D structure of TOPK protein has been constructed by using multiple templates. The quality and reliability of the generated model was checked and the molecular dynamics method was utilized to refine the model. APBS method was employed to know the electrostatic potential surface of the modeled protein and it was found that the optimum pH for protein stability is 3.4 which will further help in mechanistic hypothesis of TOPK protein. Active site of TOPK was identified from available literature and HTVS was employed to identify the lead molecules. The expected binding modes of protein-ligand complexes were reproduced in the MD simulation which indicates that the complex is relatively stable. The pharmacokinetic properties of the lead molecules are also under acceptable range. TOPK act as a substrate for CDK1 and the protein-protein docking and dynamics studies were carried out to analyze the effect of Thr^9Ala mutation of TOPK in the two protein complex formation. It shows that the wild type complex is more stable when compared with the mutant type. Such structural information at atomic level not only exhibits the action modes of TOPK inhibitors but also furnishes a novel starting point for structure based drug design of TOPK inhibitors.

Keywords Cancer · Drug design · Electrostatic potential surface · Homology modeling · Molecular dynamics · Protein-protein interaction · TOPK

Electronic supplementary material The online version of this article (doi:10.1007/s00894-012-1566-1) contains supplementary material, which is available to authorized users.

P. Kirubakaran · M. Karthikeyan (✉) · K. D. Singh · S. Nagamani
Department of Bioinformatics, Science Block, Alagappa University,
Karaikudi 630 004 Tamil Nadu, India
e-mail: mkbioinformatics@gmail.com

K. Premkumar
Department of Biomedical Science, School of Basic Medical Sciences, Bharathidasan University,
Tiruchirappalli 620 024 Tamil Nadu, India

Introduction

T-lymphokine–activated killer cell–originated protein kinase (TOPK) is a novel mitotic protein kinase which is over expressed in many malignant tumors. The expression of TOPK has been correlated with malignant potentials of various tumors such as leukemia, myeloma (Kahler's disease) and lymphomas (lymphatic cells) [1–4]. It has been suggested that TOPK comes under mitogen-activated protein kinase like (MAPKK) family [1]. Phosphorylation and activation of TOPK take place during mitosis stage of cell division [2]. TOPK has been shown to phosphorylate the p38 protein kinase but not the extracellular signal-regulated kinases or c-Jun NH_2-terminal kinases [4].

Interleukin-6-mediated interaction between TOPK and Raf-A were up regulated in murine myeloma cells [4, 5]. TOPK act as a marker as well as involved in the proliferation of neuronal progenitor cells [6]. In cell cycle regulation TOPK plays a major role by acting as a substrate for cdc2/cyclin B (CDK1) [2, 7]. Thr^9 is an important phosphorylation site of TOPK because, when Thr^9 was substituted to Ala, the binding ability of TOPK to CDK1 was decreased [7].

Two critical regulators are mainly involved in TOPK expression during growth arrest in leukemia cells called cell cycle-specific transcription factors E2F and cyclic AMP-responsive

element binding protein/activating transcription factor [8]. TOPK has been considered as a molecular target in breast cancer cells since TOPK plays a key role in phosphorylation [9] and regulation of stimuli-specific signaling pathways in colon cancer cells [10, 11]. The mechanism behind TOPK-mediated regulation of cell proliferation and its own functions remain indistinct, and that hinders its use a potential target for cancer.

The present study unveils the structural insights and possible functions of TOPK. TOPK protein structure has been constructed based on the crystal structures of 2NRU, 2Y4I and 3LXK as templates. The constructed model was further validated by various structure analysis tools and the backbone atoms were checked using molecular dynamics (MD) studies. The electrostatic potential of the modeled protein has also been calculated with the help of pKa value. The detailed interactions between protein and ligand complexes were elucidated with the help of a series of docking studies. From these analyses we also found apart from reported amino acids some other amino acids also play a role in the interaction between protein ligand complexes. Further we studied the stability of the model as well as the active site conformation using MD simulation studies. TOPK act as a substrate of cdk1/cyclin B which has a phosphorylation site at N-terminus [7]. We performed protein-protein docking and dynamic studies of TOPK-CDK1 and Thr^9 mutated TOPK-CDK1 complexes to know the binding affinity of each complex. Pharmacokinetic properties of the drug like molecules were also calculated. Results of this study can assist further characterization of the biological roles of TOPK and designing inhibitors against it as drug candidates.

Materials and methods

Multiple sequence analysis of target and template proteins

The TOPK amino acid sequence was retrieved from the SWISS-PROT and TrEMBL databases of the ExPASy molecular biology server [12] (accession number: Q96KB5) which contains 322 amino acids. The crystal structures of the IRAK-4 kinase (PDB ID: 2NRU), KSR2-MEK1 hetero dimer (PDB ID: 2Y4I) and TYK2-JAK3 kinase domains (PDB ID: 3LXK) were used as templates for homology modeling [13–15]. IRAK–4 (PDBid: 2NRU) protein which has 29 % of sequence identities with 2.0 Å of protein resolution, 3LXK is a complex of TYK2 and JAK3 protein which also has 29 % of identities with the same 2.0 Å of resolution and the 2Y4I is a KSR2-MEK1 hetero dimer which has 34 % of sequence identity with the resolution of 3.46 Å respectively. The sequences were initially aligned using ClustalW server with default parameters. The aligned sequences were manually adjusted using the software BIOEDIT [16]. In BIOEDIT the target sequence were aligned with the three templates. The conserved residues were subjected to verify and adjust the primary alignment leading to correct sequence alignment. Appropriate gaps were introduced wherever required to optimize the alignment for the coverage of whole template sequences.

Construction of the homology model

The multiple sequence alignment has been built by MODELER 9v8 (an automated homology modeling program) [17] and used to construct the 3D structure of TOPK protein. Ten 3D models were built and analyzed by PyMol software [18]. Models were ranked on the basis of internal scoring function, and those with the least internal scores were selected and used for further model validation process. The best model was ranked based on discrete optimized protein energy (DOPE) score generated by MODELER software. Hydrogen atoms were added to the model using the "protein preparation wizard" module within Maestro 9.1 (Schrödinger, LLC, 2010, New York, NY) software. Visual examination of the side chain orientations was done using information from the templates to ensure proper orientation.

Assessment of the homology model

In order to assess the reliability of the TOPK modeled structure, structural analysis and verification server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES) has been used. The backbone conformation and overall stereochemical quality of the protein structure has been calculated by analyzing the phi (Φ) and psi (ψ) torsion angles using PROCHECK program, as determined by Ramachandran plot statistics. The non-bonded atomic interactions were analyzed by using ERRAT program [19]. In order to check the native protein folding energy of the model, ProSa II analysis was done [20, 21] which compares the energy criteria with the potential mean force derived from a large set of known protein structures. Finally, the model quality assessment program ProQ has been used to evaluate the output candidate model and select the most reliable one for final model selection and further refinements [22].

MD simulations of the homology model and protein-ligand complexes

MD simulations of the modeled protein and the protein-ligand complexes were carried out using Desmond version 2.2 (D. E. Shaw Research, New York, NY) with the optimized potentials for liquid simulations (OPLS 2005) force field [23, 24]. Tautomerization, ionization and hydrogen states were modified using protein preparation wizard. The model was set up using TIP4P water [25] as a solvent in orthorhombic box with a physiological salt concentration of 0.15 M Nacl in 10 Å buffer using the system builder module

in maestro. This has resulted in simulation systems of 47,637 atoms respectively. The modeled protein was stabilized via a step by step manner starting with restrained minimization force constant (50 kcalmol^{-1} $Å^{-2}$) and it was applied on solute heavy atoms with maximum iterations of 2000 steps. For maintaining a constant temperature and pressure, short time simulation has been carried out with 500 ps run at 300 K using Martyna – Tobias – Klein barostat method and LBFGS vectors. Interaction with coulombic short range method has been used with the cutoff of 9.0 Å and long range method of smooth particle mesh Ewald tolerance of 1E-5 implemented in SHAKE algorithm. After all the excluded restraints from the system the main simulation runs were carried out. Full simulation was conducted through NPT ensemble and Nose – Hoover for thermostat method and Martyna – Tobias – Klein for Barostat method each energy and trajectory atomic coordinates data were recorded every 1.2 ps. The van der Waals, short (9 Å) and long range electrostatic interactions were computed using the particle mesh Ewald method [26]. Finally, a RESPA integrator has been used with a time step of 2 fs for near and 6 fs for distant bonded distances [27]. We followed the same parameters as mentioned above for protein-ligand complexes, except the energy and atomic coordinate recording intervals. The intervals were recorded every 1.2 and 2.1 ps, respectively. A total of 60 ns MD simulation were carried out for all the protein and ligand complexes.

Function and binding pocket prediction

The function of the TOPK protein was predicted using 3d2GO server. It predicts the protein function using a number of criteria like overall topology, geometric and residue similarity of functional to structures with known functions. Finally all this information is processed by a support vector machine by using true and false positive functional assignments (http://www.sbg.bio.ic.ac.uk/phyre/pfd/). The binding pocket or active site is usually a pocket (site) present in the protein that consists of residues responsible for ligand specificity, which act as proton donors or acceptors. The binding pocket identification plays a major role in structure based drug design. This process has been done based on literature reports and various computational tools. Through literature studies it has been proven that amino acids 170–179 play a vital role in the ligand binding [28]. The binding pocket region of the protein has also been identified by SiteMap program [29] which identifies one or more regions suitable for ligand binding. Next, the hydrophobic and hydrophilic maps (donor, acceptor and metal-binding regions) were produced using various contour maps. The score was generated using default parameters implemented in SiteMap (version 2.4, 2009, Schrödinger, LLC, New York, USA) program to generate more than two sites [30].

Electrostatic potential calculations

The electrostatic potential calculations were done by adaptive Poisson-Boltzmann solver (APBS) method [31]. The calculation was based on grid-base method to solve the linearized Poisson-Boltzmann equation. The modeled TOPK protein was then subjected to conversion of partial charges and atomic radii using PDB2PQR tool with PARSE force field [32, 33]. For better understanding of the electrostatic potential surface area protein and solvent dielectric coefficients of 2 and 78.5 were used respectively. The contact surface selection was mapped using a radius of 1.4 Å where as the ion accessibility surface was defined with the probe radius of 2.0 Å. The molecular surface, solvent accessibility, solvent excluded area and van der Waals surfaces have been build with the Jmol visualization tool [34].

Virtual screening protocol

Virtual screening process has been done using Glide in the high throughput virtual screening (HTVS), standard precision (SP) and extra precision (XP) modes with May Bridge and TOSLab databases [35]. Glide v5.5 has been used for all the docking protocol. Glide provides three different level of docking processes (HTVS, SP and XP). The selected databases were screened in HTVS mode; here all the ligands will be processed which includes the ligand preparation using LigPrep, filtering using ligfilter on QikProp and other structural properties. The filtered compounds will further enter into SP mode of screening. The poses generated from Glide SP were again refined in the final process called Glide XP method. This resulted in Glide scoring function using force field OPLS 2005. Followed by docking, entire compounds were ranked according to the Glide score.

Protein-protein docking studies

The ClusPro server was used for protein-protein interaction studies [36–39]. TOPK-CDK1 and Thr^{9} mutated TOPK-CDK1 were used for docking studies. The CDK1 was obtained from PDB database (PDB I.D 1LC9). The Thr^{9} amino acid has been mutated to Ala in TOPK protein by Schrödinger 2011. The mutated structure was optimized and energy minimized with the help of protein preparation wizard. ClusPro is the fully automated web-based program for docking proteins and it includes two main steps. First a rigid body docking program called PIPER was run based on fast fourier transform (FFT) correlation approach. The second step includes 1000 best energy conformations which are clustered and the 30 largest clusters are retained for refinement. We selected one of the returned models after considering the energy and the size of the cluster, preferring lower energies and larger cluster sizes. The obtained protein-protein complexes were prepared and energy minimized

with protein preparation wizard implemented in Schrödinger 2011.

Molecular dynamics simulations of protein–protein complexes

The best conformer of TOPK-CDK1 and Thr^9 mutated TOPK-CDK1 complexes obtained from protein-protein interaction studies were subjected to MD simulation using the previously defined parameters for a total of 10 ns production MD run for each complex. The lowest potential energy conformations (average structure) were selected from the 10 ns MD simulations. All the MD simulation analyses were carried using Desmond package programs. Molecular structure has been visualized in PyMOL software [18].

ADME and biological activity spectrum (BAS) predictions of lead molecules

The best three lead molecules were checked for their drug-like properties using QikProp module from Schrödinger software suite [40]. It predicts physically significant descriptors and pharmaceutically relevant properties. All the screened compounds were neutralized before being used in QikProp. The neutralization is essential since no properties will be generated in the normal mode. The program predicts 44 properties for a single molecule. The major properties include principal descriptors and physiochemical properties, along with detailed analysis of QP%, log HERG and log P (octanol/water). The acceptability of the compounds was evaluated by Lipinski's rule of five. The set of pharmacological effects, mechanisms of action, and specific toxicities that might be exhibited by a particular compound in its interaction with biological entities are predicted by PASS and it is termed as "biological activity spectrum" of the compound [41].

Results and discussion

Multiple sequence alignment and secondary structure analysis

Given that there are no x-ray crystals and NMR structures are available for TOPK protein, we developed a protein model by comparative homology modeling to understand the structure and function of active site residues for structure based drug design. The amino acid sequence of TOPK was retrieved from UniProt database (Q96KB5). The template structures were searched based on blastp analysis against PDB in NCBI database. This search resulted in a large number of sequence similarity, mainly in serine/threonine protein kinase families. Multiple sequence alignment has been constructed by using BIOEDIT. Finally we have taken three protein sequences which almost cover the target protein sequence by means of conserved residues. The PDB structures of 2NRU-B, 3LXK-A and 2Y4I-C were used as template structures to model TOPK protein. The sequence alignment in ESPript and secondary structure of the TOPK protein has been shown in Fig. 1. The TOPK protein has 11 alpha helix and 11 beta sheets. Active site residues were found positioned in the beta sheet 6 to beta sheet 7 positions and few loop regions also play a role in protein ligand interactions. The beta sheets 6 and 7 are made up of three amino acids each. These active site residues were taken from available literature.

Homology model of TOPK protein

The TOPK protein has been modeled based on results obtained from secondary structure analysis. The coordinates of 2NRU-B, 3LXK-A and 2Y4I-C were used as templates to construct the 3D structural model of target protein because of their high sequence similarity and identities with target protein. Recently it has been reported that multiple template selection plays a major role in quality improvement of the modeled protein [42–44]. The homology model of TOPK has been built using MODELLER 9v8 program. The input parameters of MODELLER were set to generate ten models with the level of high optimization and satisfying spatial restraints in terms of protein density function. The final model has been selected based on low MODELLER objective function and the high value of negative DOPE potential. The final model has DOPE score profiles similar to that of the templates used, which indicates folds and loops were similar in target and template structures. Figure 2 shows the modeled protein with the same secondary structure predicted from ESPript. The protein shows helix-helix interaction which stabilizes the protein structure, since the two interacting helices are involved in residue-residue contact. The model also folded into α/β domain and comprises 11-stranded anti parallel β-sheets with helices packed on either side.

Model validation

Validation of the model including the geometric properties of the backbone conformations, compatibility of residue interactions and energy profiles were analyzed using various structure evaluation programs. Ramachandran plot calculations were computed with PROCHECK program. The Phi/Psi distributions of the Ramachandran plot for TOPK has shown 81.3 % of the residues in the core region or most favourable regions, 14.4 % residues in the additionally allowed regions, 3.9 % in the generously allowed regions and 0.4 % in disallowed regions (Supplementary Fig. 1). Different residue properties were analyzed using the G-factor present in PROCHECK that includes maximum deviation of 6.4 Å, bad contacts of 0 Å and bond length/angle

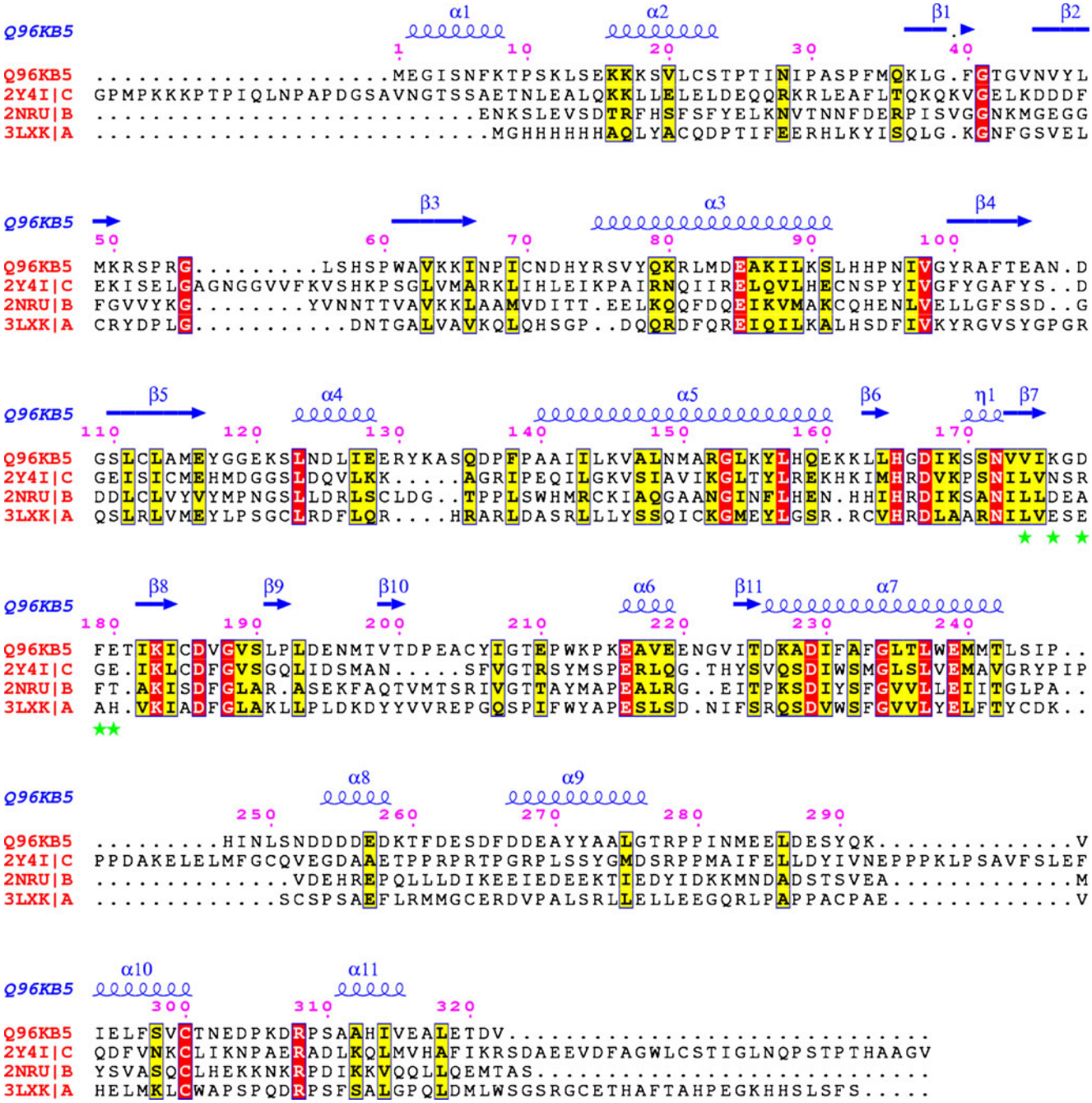

Fig. 1 Multiple sequence alignment using ESPript. Residues highlighted in red correspond to identical/conserved residues, while residues in red text are similar in the three proteins. Secondary structural elements are shown for TOPK protein

were 6.5 Å. The quality of overall G-factor is -0.51 Å for dihedrals, -0.15 Å for covalent and overall -0.35 Å. The overall quality factor of the modeled protein has been calculated based on the non-bonded atomic interactions using ERRAT tool. The usual accepted score ranges for high quality model is >50 [19]. The ERRAT score for the modeled protein was 72.70 % which is within the range of a high quality model. Therefore, the PROCHECK and ERRAT results confirm the quality of predicted 3D structure as more reliable and within an acceptable range. The interaction energy and Z-score of each residue were calculated by ProSA II energy plot. The ProSA interaction energy calculations for each residue were done using distance-based pair potential method. In this study, the modeled protein shows maximum residues with negative interaction energy, very few residues with positive interaction energy that conform the reliability of the model. The overall Z-score of the model was -6.00 kcalmol^{-1}, which has been found quite similar to the templates 2NRU (-7.12 kcal

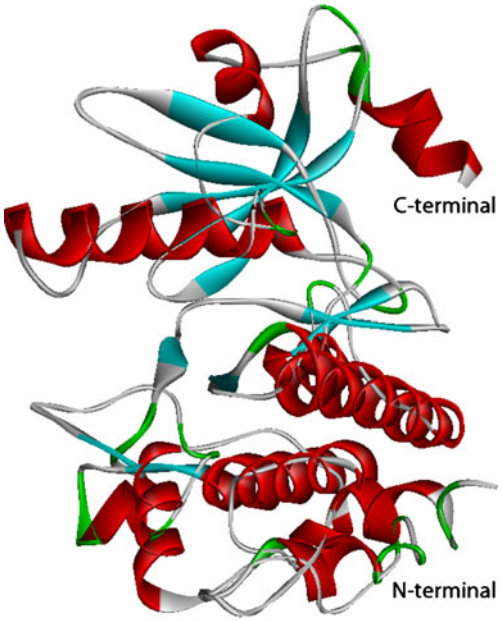

Fig. 2 The 3D structure of TOPK protein is shown as schematic diagram and colored according to secondary structures. Helices are red, beta sheets are cyan, turns are green, and coils are beige

mol^{-1}), 2Y4I (-6.66 kcalmol^{-1}) and 3LXK (-7.62 kcalmol^{-1}) (Fig. 3). The results obtained from ProSA suggest that the generated TOPK protein structure was energetically consistent compared to the template structures. The modeled structure has been checked for its quality by using ProQ. The result shows that predicted LGscore: 2.880 (>2.5: very good model) and predicted MaxSub score: 0.637 (>0.5: very good model) were in acceptable range of a good model. These validation studies concluded that backbone conformations, non-bonded interactions and energy profiles were within the limits established for reliable structure. The constructed model is reasonable and can be used for further refinement with energy minimization and MD simulation studies with constrained backbone in order to relieve unfavorable steric clashes.

TOPK protein RMSD and flexibility using MD simulations

To investigate the stability and conformational changes of the modeled TOPK protein, MD simulation was performed in the fully solvated water molecule. Analysis of the flexibility of TOPK protein provides detailed information on the structural characterization of the modeled protein [45]. The total atoms (TOPK) soaked with water molecules were calculated and found to have 47,637 particles. The pressure, temperature, volume, potential and total energy of the modeled protein with improved steric parameters play major role in stabilizing the target protein. The analyses of the simulation parameters revealed that the pressure for the modeled protein remains constant with the average of 1.18 bar and overall standard deviation of 57.67 bar. The average temperature was maintained constant with 300 K and the total volume of 365,209.3 Å^3 with standard deviation of 378.8 Å^3. The potential and total energy of the protein maintains stable from starting time period to ending 10 ns time period with -103,198.4 kJmol^{-1} total energy and -127,148.8 kJmol^{-1} potential energy. These results show that the model is highly stable in nature with all the simulation event analysis parameters mentioned above and the energy conservation was satisfied in MD simulation. The RMSD for the backbone atoms are shown in Fig. 4a which explains the constructed model is a stable and consistent model throughout the 10 ns simulation time with 0.98 Å standard deviation value. The RMSD was stable about 5–5.5 Å after 5 ns of simulation and was not increased significantly after 5 ns of simulation time which indicates that the systems evolved into a stable state and have reasonably converged over the production MD studies. The radius of gyration has been calculated through the intrinsic dynamics, improved relaxation and structural stability of the model protein shows 0.40 Å of standard deviation (Fig. 4b). This indicates that the level of compaction in the structure (fold or unfolded nature of the modeled protein) as a stable one after 5.5 ns of simulation process. The final model passed all the quality assurance tests and hence it is proposed that this model has a good estimation of TOPK structure and can be used to describe various receptor-ligand complex interaction studies and to study the relationship between the structure function.

Electrostatic potential of TOPK protein

Different electrostatic potentials of the TOPK protein are shown in Fig. 5. The molecular electrostatic potential surface of the modeled protein showed that N-domain region of protein is mostly favourable for hydrophobic nature when compared to the C-domain. A part of the C-domain regions were found to be covered with non-polar amino acids. The solvent accessible surface area, solvent excluded regions and van der Waals regions are shown in Fig. 5. The protein pKa prediction result shows that the optimum pH of stability is 3.4 for which the free energy is 53.1 kcalmol^{-1} at the temperature of 298 K. Protein charge for folded nature is 5.40 and unfolded state is 4.98. The stability of the protein has been shown in Supplementary Fig. 2. The electrostatic data were useful to establish mechanistic hypothesis of the modeled protein.

Functional prediction of TOPK protein

The functions of the protein model were predicted using 3d2GO server in terms of gene ontology (GO). The GO term I.D's, descriptions and the confidence level is shown in Table 1. The confidence level should be less than 0 and the confidence level 1 will be considered as most confident

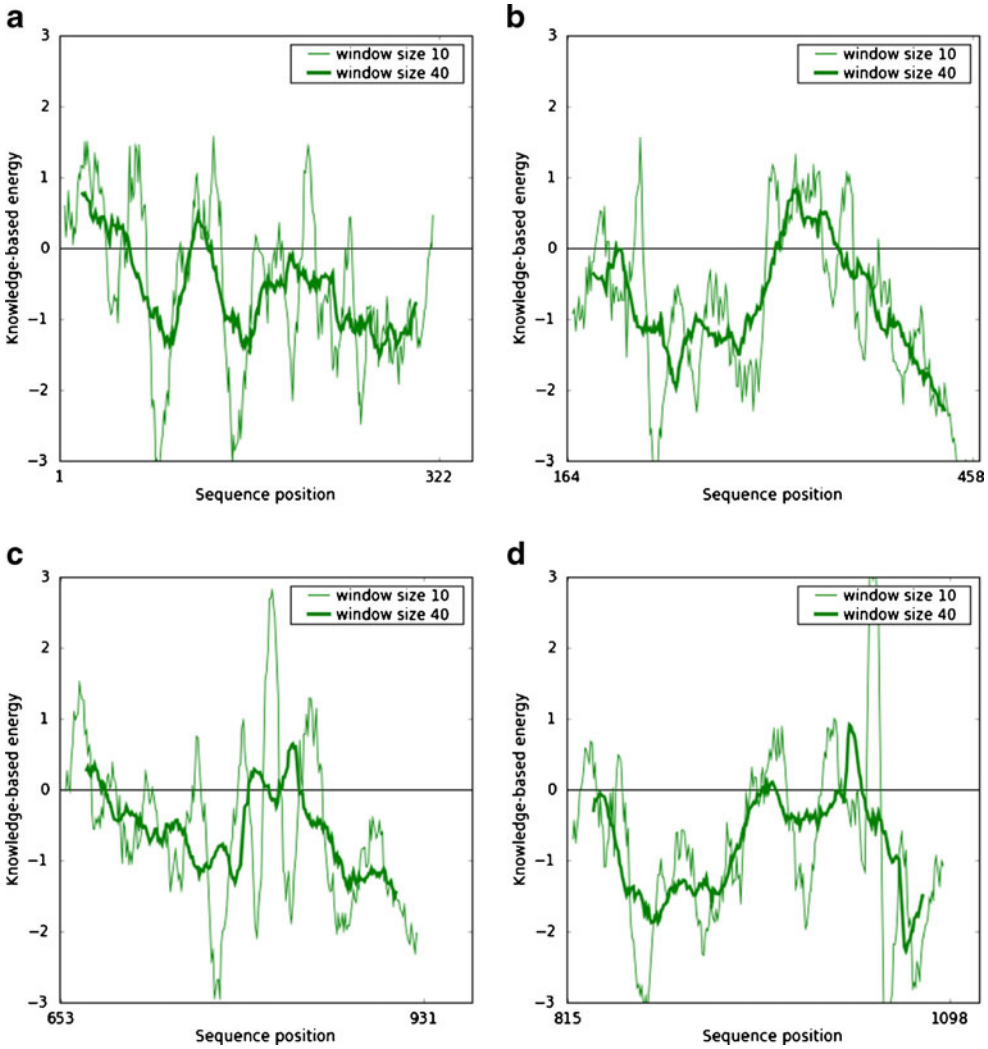

Fig. 3 The energy profile of TOPK protein obtained using ProSA web server analysis. The ProSA analysis of the model showed maximum residues with negative region of energy. (**a**) The ProSA analysis of the model -6.00 kcalmol^{-1}. (**b**) The ProSA analysis of the 2NRU showed z-score of -7.12 kcal mol^{-1}. (**c**) The ProSA analysis of the 2Y4I showed z-score of -6.66 kcalmol^{-1}. (**d**) The ProSA analysis of the 3LXK showed z-score of -7.62 kcalmol^{-1}

prediction. Results show that the TOPK protein has different functions like phosphotransferase activity, protein kinase activity, transferase activity, transferring phosphorus-containing groups, kinase activity, transferase activity and

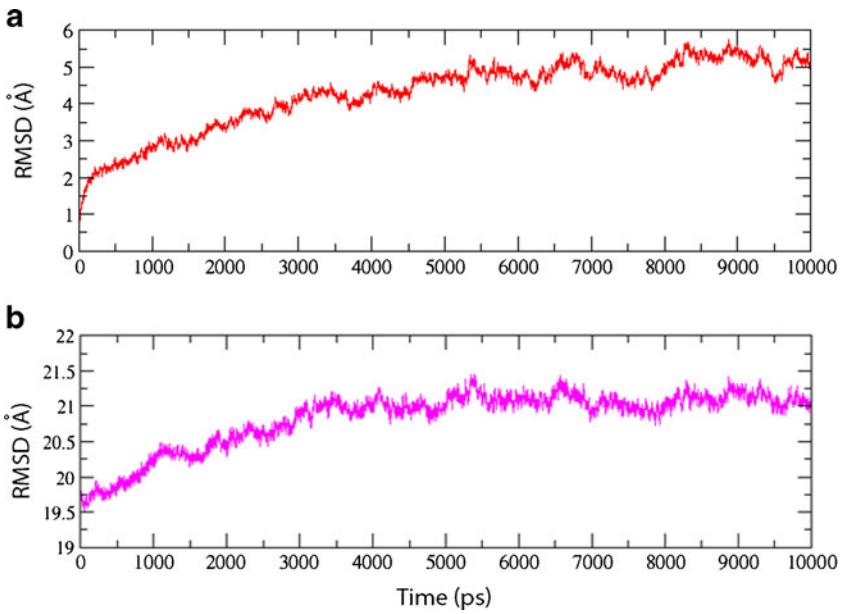

Fig. 4 Time-dependent RMSD (Å) from TOPK homology model in a 10 ns MD simulation. (**a**) The backbone atoms are shown in red color. (**b**) The radius of gyration is shown in magenta color

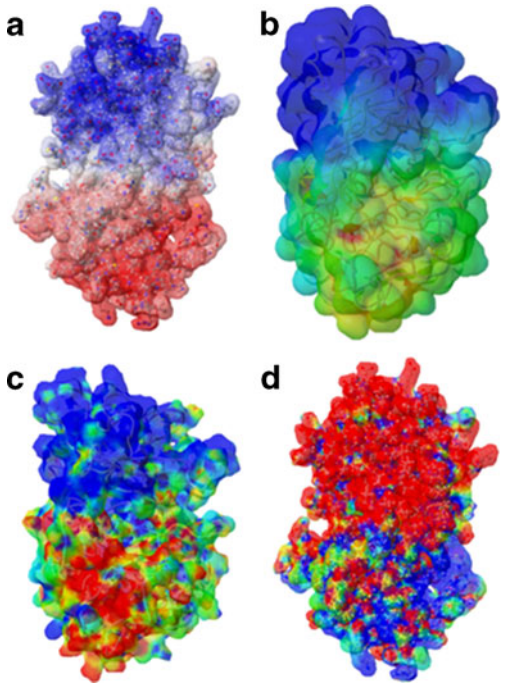

Fig. 5 Electrostatic potential on the molecular surface area of TOPK protein. Red and blue indicates negative and positive potential, respectively. The surface potential was calculated and displayed using the Jmol ABPS tool. (**a**) Molecular electrostatic potential surface. (**b**) Solvent accessible. (**c**) Solvent excluded. (**d**) van der Waals surface

protein serine/threonine kinase activity. The predicted results show that TOPK protein has high confidence level [0.99–0.98].

Active site prediction and molecular docking studies

In order to investigate the interaction between TOPK and the ligands, the active site was defined as a subset that contains residues of any atoms within 5 Å from ligand. The TOPK active site has been calculated by SiteMap module in Schrödinger. The best site was selected based on the site score (1.044) and relatively exposed regions of the active site displays the hydrophobic (yellow) with surface area of 244.2, hydrogen-bond donor (blue) of 1589.1 and acceptor (red) maps of 991.2. The predicted site is comprised of 11 amino acid residues Pro95, Gly119, Val174, Ile175, Lys176, Gly177, Asp178, Phe179, Glu180, Thr181 and Lys183 (Supplementary Fig. 3). Based on the experimental results and our predicted theoretical results, this site has been chosen as the more favorable binding site to dock the ligands. Normal Glide docking was performed for reference ligand (Staurosporine) in the validated binding pocket of TOPK protein. Staurosporine is a broad specificity protein kinase inhibitor [46] which also inhibits TOPK protein (www.reactionbiology.com/webapps/main/Kinases/Invitrogen100114/PBK.pdf). The docking results indicate that binding of the Staurosporine has been facilitated by two binding site atoms were within 5 Å of the ligand. This particular binding site was utilized for HTVS of compounds from Maybridge and TOSLab databases. Glide HTVS has been performed for each database (Maybridge and TOSLab). Finally, 23 compounds from Maybridge and 38 compounds from TOSLab databases were selected based on their docking results and pharmacokinetic properties. Totally 61 drug-like molecules were docked in the active site region of the constructed protein. Fifteen drug-like compounds were selected based on Glide score greater than that of reference ligand (Staurosporine) further these compounds were tested for its binding conformation by docking analysis. Three out of 15 drug-like compounds were identified as potential lead molecules based on their hydrogen bonding interactions and Glide energies (-31 to -37 kcalmol^{-1}) that confirmed strong protein-ligand interactions. The chemical structures of the three lead molecules along with their IUPAC and database names are displayed in Fig. 6. The best conformations and binding poses of the three lead compounds and the reference ligand with the modeled protein are shown in Fig. 7. The protein-ligand complex of three lead molecules posses hydrogen bond interactions, which are more or less similar to that of reference ligand and their Glide scores are comparable to that of Staurosporine (Table 2).

Table 1 Functions of the TOPK were predicted in terms of gene-ontologies and their respective confidence level

GO term	Description	Confidence
GO:0016773	phosphotransferase activity	0.99
GO:0006468	protein kinase activity	0.99
GO:0016772	transferase activity, transferring phosphorus-containing groups	0.99
GO:0016301	kinase activity	0.99
GO:0016740	transferase activity	0.98
GO:0004674	protein serine/threonine kinase activity	0.98

MD simulations of protein-ligand complex

The behavior of a protein-ligand complex over a specific time period will be identified with the help of MD simulation studies. The expected binding mode of the protein-ligand complex has been reproduced in the MD simulation technique. MD simulations were performed for 40 ns with apo (without ligand) and holo enzymes (with ligand). The changes in the active site region and the RMSD of the ligands are shown in Fig. 8a. All three compounds tested have shown minimal deviation in there conformation or positional changes. Overall RMSD of the compounds were around 3 Å. The Maybridge_120 have shown no noted fluctuation in whole simulation process. TOSLab_351 has

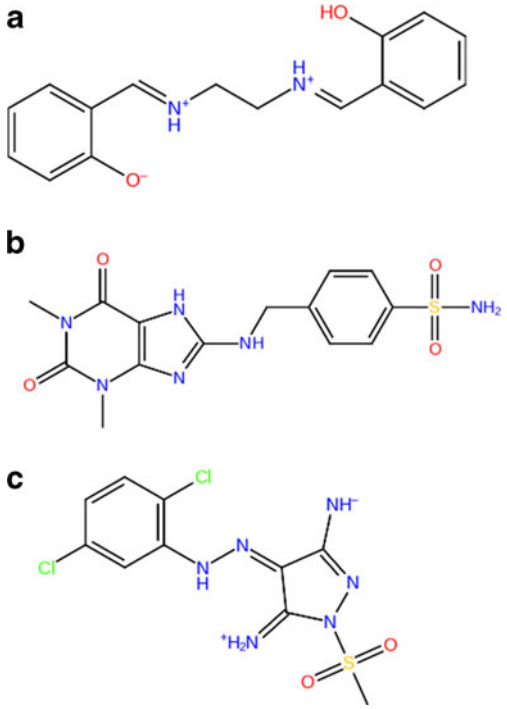

Fig. 6 Structure of the three lead molecules. (**a**) Maybridge 120 and the IUPAC name is 2-[(1E)-({2-[(E)-[(2-hydroxyphenyl)methylidene] azaniumyl]ethyl}iminiumyl)methyl]benzen-1-olate (**b**) TOSLab 351 and the IUPAC name is 4-{[(1,3-dimethyl-2,6-dioxo-2,3,6,7-tetrahydro-1H-purin-8-yl)amino]methyl}benzenesulfonamide (**c**) TOSLab 111 and the IUPAC name is (4Z)-3,5-diimino-1-(methylsulfonyl)pyrazolidin-4-one (2,5-dichlorophenyl)hydrazine

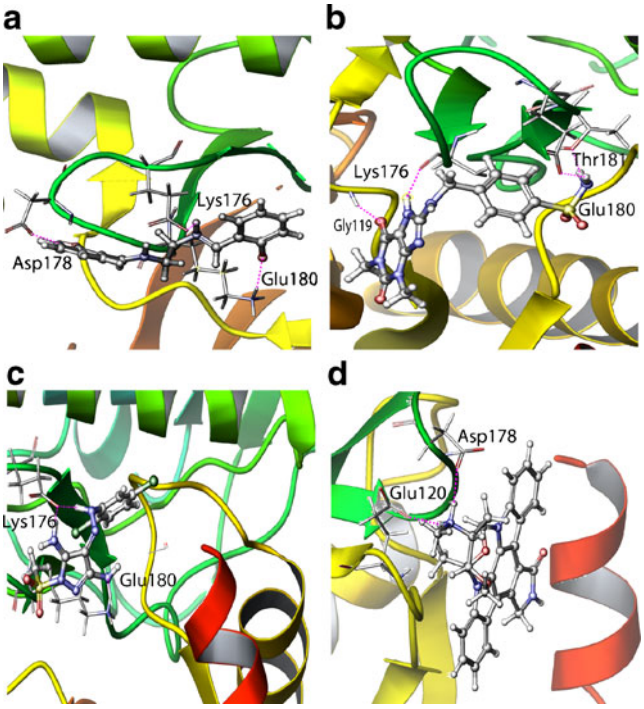

Fig. 7 Binding poses of the three lead molecules. The proposed binding mode of the lead molecules are shown in ball and stick display and non carbon atoms are colored by atom types. Hydrogen bonds are shown as dotted pink lines. Atom type color code: red for oxygen, blue for nitrogen, gray for carbon and yellow for sulfur atoms respectively. (**a**) protein-ligand complex with Maybridge 120. (**b**) protein-ligand complex with TOSLab 351. (**c**) protein-ligand complex with TOSLab 111. (**d**) protein-ligand complex with reference ligand

shown deviation but not more than 3 Å and small fluctuation happened in between the time period of 4000 to 5000 ps and TOSLab_111 have a constant RMSD during the whole simulation. These results revealed that constant RMSD has been maintained during the whole simulation time. To confirm the stability of the protein-ligand interaction in the MD simulations we also analyzed the number of hydrogen bonds (H-bond) between TOPK and three selected ligands during 10 ns simulation time. The H-bonds analysis showed that MayBridge_120 has a maximum of six H-bond interactions with the protein and it constantly has four H-bond interactions throughout the whole simulation. The next compound TOSLab_351 shows more H-bond with the protein and constant interactions were shown from zero to four H-bonds in the MD simulation and the final compound TOSLab_111 also shows very constant H-bond interaction between one to four amino acids. This H-bond interaction clearly indicates that protein-ligand complexes are stronger in the MD simulation studies (Fig. 8b). The differences in fluctuation for the apo and holo structures were calculated with the help of RMSF (Supplementary Fig. 4). The stabilizing effects of RMSF values revealed that during the simulation very few residues went away from the 8.0 Å for the total holo enzymes. The RMSF value of the alpha carbon residues were calculated from the whole MD simulation time period (Supplementary Fig. 4). When compared to apo protein the protein-ligand complexes having RMSD in between 1–1.5 Å which reveals the constant RMSD was maintained throughout the MD simulation.

Comparative structural analysis of TOPK-CDK1 and Thr^9 mutated TOPK-CDK1 complexes

In this study we analyzed the functional aspects of TOPK by detecting its relationship to CDK1 complex. N-terminus region of the CDK1 will be phosphorylated by TOPK protein. The downstream targets of TOPK include MAPK's JNK1 and ERK2. ERK2 forming a positive feedback loop with TOPK which will lead to a tumor in colorectal cancer cells [10, 11]. Thus, TOPK has been supposed to play a role with CDK1 in mitotic cell division but the exact mechanism still needs to be elucidated. In this study we are trying to elucidate the protein-protein interaction studies between both the TOPK-CDK1 complexes. The ClusPro results show that the TOPK-CDK1 complex has a weighted score of -906.6 for center and -1198.6 for lowest energy and for the Thr^9 mutated TOPK-CDK1 has a weighted score of -

Table 2 Docking results of three lead molecules with the reference compound using Glide XP in Schrödinger 9.1

Lead molecule[a]	Glide score[b] (kcal/mol)	Glide energy[c] (kcal/mol)	Active residues interaction (D…H-A)[d]	H-bond length (Å)
Maybridge_120	-10.074	-34.103	OH…O (Asp178)	1.647
			NH (Glu180)…O	2.321
			NH (Lys176)…O	1.932
TOSLab_351	-8.946	-37.893	NH (Gly119)…O	2.055
			NH…O (Lys176)	1.973
			NH (Thr181)…O	2.071
			NH…O (Glu180)	2.082
TOSLab_111	-8.390	-32.384	NH (Lys176)…CN	2.128
			NH…O (Glu180)	1.843
*CID_44259	-5.198	-31.623	NH…O (Asp178)	1.935
			NH…O (Glu120)	1.930

[a] Ligand IDs are from the Maybridge and TOSLab databases

[b] Glide score

[c] Glide energy

[d] Number of hydrogen bonds formed (D - donor, H - hydrogen, A - acceptor)

[*] Staurosporine (reference compound)

926.2 and -1045.9 for lowest energy. ClusPro analysis of the best docked complexes showed that energies of each complex are stable and compared with mutated complex, wild type is very stable in nature. It shows that mutation of Thr^{9} to Ala in TOPK protein will cause an increase in energy. The TOPK-CDK1 complexes are found to have energetically stable properties which have -1.709e5 $kcal\,mol^{-1}$ of total energy and Thr^{9} mutated TOPK-CDK1 complexes have less stable -1.694e5 $kcal\,mol^{-1}$ of total energy, as confirmed by MD simulation studies carried out on these complexes. The amino acid interactions between the wild type TOPK and mutant TOPK with CDK1 have common residues which have strong interaction between both the complexes. The Ser32, Glu128, Arg130, Tyr131, and Tyr206 in TOPK and Glu42, Ser46, Thr14, Glu209 and Ser233 in CDK1 are involved in H-bond interactions between both the complexes. The backbone RMSD of the protein complexes

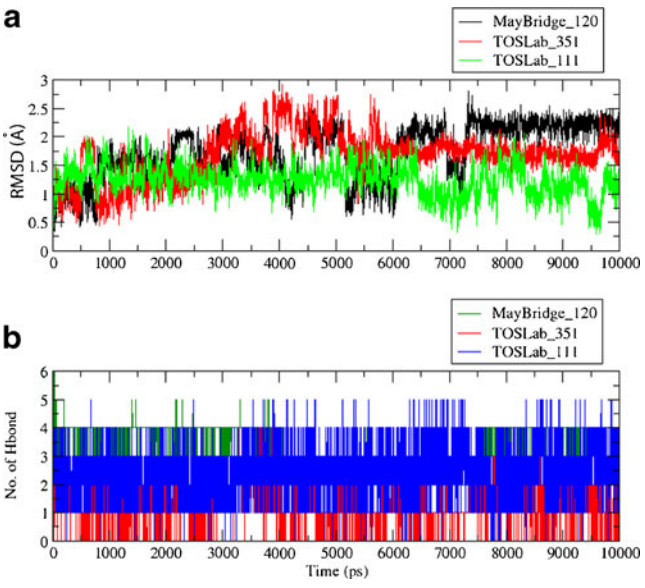

Fig. 8 (**a**) RMSD of small molecules in active site residues during the simulation in TOPK protein. (**b**) Total number of intermolecular H-bond interactions between Maybridge 120, TOSLab 351 and TOSLab 111 in complex with TOPK protein

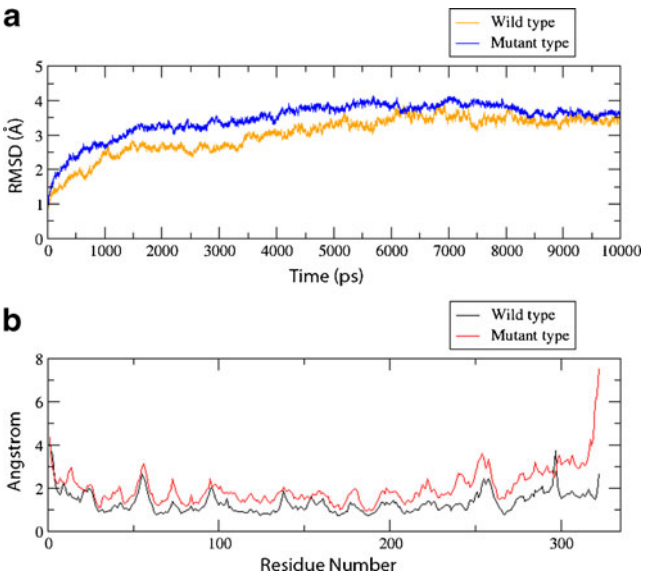

Fig. 9 (**a**) Time-dependent RMSD (Å) from TOPK-CDK1 (brown color) and Thr^{9} mutated TOPK-CDK1 (blue color) in a 10 ns MD simulation (**b**) The RMSFs for each complex (TOPK-CDK1 (black color) and Thr^{9} mutated TOPK-CDK1 (red color) with respect to residues in the wild and mutant TOPK-CDK1 protein

Table 3 ADME properties of the identified lead molecules obtained using QikProp module in Schrödinger 9.1

Lead molecules[a]	Stars[b]	Mol.wt[c]	QplogS[d]	Qplog HERG[e]	QPPcaco2[f]	QPPMdck[g]	Percent human oral absorption[h]	Rule of five[i]	No of metab[j]
Maybridge_120	1	268.32	-3.57	-6.43	104.92	497.34	100	0	2
TOSLab_351	0	364.38	-3.60	-5.04	123.59	228.83	50.92	0	1
TOSLab_111	0	349.19	-3.51	-5.07	131.49	280.66	72.43	0	2

[a] Ligand IDs are from the Maybridge and TOSLab databases.

[b] Number of property or descriptor values that fall outside the 95 % range of similar values for known drugs (0–5 acceptable)

[c] Molecular weight of the molecule (130.0–725.0 acceptable)

[d] Predicted aqueous solubility, log S. S in moldm–3 is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid (–6.5–0.5)

[e] Predicted IC50 value for blockage of HERG K+ channels (concern below –5)

[f] Predicted apparent Caco-2 cell permeability in nm/sec. Caco- 2 cells are a model for the gut- blood barrier. QikProp predictions are for non-active transport (<25 poor, >500 great)

[g] Predicted apparent MDCK cell permeability in nm/sec. MDCK cells are considered to be a good mimic for the bloodbrain barrier. QikProp predictions are for non-active transport (<25 poor, >500 great)

[h] Predicted human oral absorption on 0 to 100 % scale. The prediction is based on a quantitative multiple linear regression model. This property usually correlates well with human-oral-absorption, as both measures the same property (>80 % is high <25 % is poor)

[i] Number of violations of Lipinski's rule of five. The rules are: mol_MW<500, QPlogPo/w<5, donorHB≤5, accptHB≤10. Compounds that satisfy these rules are considered druglike (maximum is 4)

[j] Number of likely metabolic reactions (1–8)

obtained after the simulation studies revealed very small deviations among wild type TOPK protein and mutant TOPK with CDK1 complexes (Fig. 9a). The maximum RMSD value observed in the case of Thr^9 mutated TOPK-CDK1 was about 3.6 Å, whereas it was 3.3 Å for TOPK-CDK1. The change in the amino acid in the position of ninth residue leads to the increase of RMSD which results in less stable interaction between TOPK-CDK1 complexes. After various time periods of MD simulations, no changes in the structures were observed. The H-bond interaction between the Thr^9 mutated TOPK-CDK1 was not stable in the whole course of MD simulations compared with TOPK-CDK1 complexes. To observe the differences of structural behaviors and flexibilities in the wild and mutant TOPK protein with CDK1 complexes, we analyzed and compared the results with root mean square fluctuation (RMSF) of each residue. The wild type TOPK-CDK1 was not detected with any noticeable aspect of conformational changes whereas Thr^9 mutated TOPK-CDK1 has displayed conspicuous structural changes (Fig. 9b) especially from the residues of 250 to 316 amino acids present in the helix regions. The RMSF analysis showed that the helix region of TOPK-CDK1 has lower flexibility than Thr^9 mutated TOPK-CDK1 complexes. These interactions affected the rigidity of the helix region in the mutated system.

Pharmacokinetic prediction of lead molecules

The drug-likeness of the three lead molecules were evaluated by calculating their physicochemical properties that includes stars (number of property or descriptor values that fall outside the 95 % range of similar values for known drugs) (<5), molecular weight (<500 Daltons) and rule of five. All these properties were in acceptable range. Water solubility (QPlogS) calculations have been done to estimate the absorption and distribution of drugs within the body. The values of QPlogS ranged between ~-3.5 and ~-3.6. Drug metabolism and its access to biological membranes were calculated using cell permeability method (QPPCaco2) and the IC_{50} values for blockage of HERG K+channels were predicted using QPlogHERG. The QPPCaco2 ranged from ~104 to ~131 and QPlogHERG ~-5 to ~-6. While the cell permeability (QPPMDCK) of the lead molecules ranges from ~228 to ~497. The number of metabolic reactions was found to be from ~1 to ~2. The percent of human oral adsorption ranges from ~50 to ~100 (Table 3). Overall these three lead molecules are within the acceptable range for human usage. PASS prediction of Maybridge_120, TOSLab_351 and TOSLab_111 shows activity against protein kinase. The most probable activities for a given compound are characterized by Pa values close to 1, and Pi values close to 0. Their Pa values are 0.569, 0.558, 0.415 and Pi values are 0.001, 0.002, and 0.044 respectively.

Conclusions

We have modeled the TOPK protein systematically and evaluated its structure by using various structure analysis tools such as PROCHECK, ERRAT and ProSA. The

stability and energy minimization process has been done with MD simulation studies. The surface energy was calculated by electrostatic analysis. Active site was predicted with the help of SiteMap, and the result has been supported by the available experimental data. These results support the reliability of the modeled structure. Docking study was performed with reference ligand and the result indicated that position and orientation of the molecules bind in the active site region of the modeled protein. Based on HTVS, three potential lead molecules were identified (Maybridge_120, TOSLab_351 and TOSLab_111). MD simulation of the protein-ligand complexes supported the docking results and it has also been proved that the inhibitors maintained its stability inside the binding site of TOPK throughout the simulation time. They have shown stable interactions with the same residues that are predicted in the rigid docking studies during the whole simulation time. In relation with the protein-protein interaction studies the TOPK-CDK1 and Thr^{9} mutated TOPK-CDK1 analyses of both in docking and dynamics interaction correlates with experimental identification of the energy of both the complexes. These results provide a new interaction mechanism of wild and mutant type TOPK with CDK1 complexes. The ADME properties of the best three compounds are in acceptable range. Interestingly, some of the compounds showed activity against TOPK protein by PASS biological activity prediction. The results from this study confirmed that these compounds may be potential lead molecules against TOPK.

References

1. Abe Y, Matsumoto S, Kito K, Ueda N (2000) Cloning and expression of a novel MAPKK-like protein kinase, lymphokine-activated killer T-cell-originated protein kinase, specifically expressed in the testis and activated lymphoid cells. J Biol Chem 275:21525–21531
2. Gaudet S, Branton D, Lue RA (2000) Characterization of PDZ-binding kinase, a mitotic kinase. Proc Natl Acad Sci USA 97:5167–5172
3. Simons-Evelyn M, Bailey-Dell K, Toretsky JA, Ross DD, Fenton R, Kalvakolanu D, Rapoport AP (2001) PBK/TOPK is a novel mitotic kinase which is upregulated in Burkitt's lymphoma and other highly proliferative malignant cells. Blood Cells Mol Dis 27:825–829
4. Côté S, Simard C, Lemieux R (2002) Regulation of growth-related genes by interleukin-6 in murine myeloma cells. Cytokine 20:113–120
5. Yuryev A, Wennogle LP (2003) Novel raf kinase protein-protein interactions found by an exhaustive yeast two-hybrid analysis. Genomics 81:112–125
6. Dougherty JD, Garcia ADR, Nakano I, Livingstone M, Norris B, Polakiewicz R, Wexler EM, Sofroniew MV, Kornblum HI, Geschwind DH (2005) PBK/TOPK, a proliferating neural progenitor-specific mitogen-activated protein kinase kinase. J Neurosci 25:10773–10785
7. Matsumoto S, Abe Y, Fujibuchi T, Takeuchi T, Kito K, Ueda N, Shigemoto K, Gyo K (2004) Characterization of a MAPKK-like protein kinase TOPK. Biochem Biophys Res Commun 325:997–1004
8. Nandi AK, Rapoport AP (2006) Expression of PDZ-binding kinase (PBK) is regulated by cell cycle-specific transcription factors E2F and CREB/ATF. Leuk Res 30:437–447
9. Park JH, Lin ML, Nishidate T, Nakamura Y, Katagiri T (2006) PDZ-binding kinase/T-LAK cell-originated protein kinase, a putative cancer/testis antigen with an oncogenic activity in breast cancer. Cancer Res 66:9186–9195
10. Zhu F, Zykova TA, Kang BS, Wang Z, Ebeling MC, Abe Y, Ma WY, Bode AM, Dong Z (2007) Bidirectional signals transduced by TOPK-ERK interaction increase tumorigenesis of HCT116 colorectal cancer cells. Gastroenterology 133:219–231
11. Oh SM, Zhu F, Cho YY, Lee KW, Kang BS, Kim HG, Zykova T, Bode AM, Dong Z (2007) T-lymphokine-activated killer cell-originated protein kinase functions as a positive regulator of c-Jun-NH2-kinase 1 signaling and H-Ras-induced cell transformation. Cancer Res 67:5186–5194
12. Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28:45–48
13. Wang Z, Liu J, Sudom A, Ayres M, Li S, Wesche H, Powers JP, Walker NPC (2006) Crystal structures of IRAK-4 kinase in complex with inhibitors: a serine/threonine kinase with tyrosine as a gatekeeper. Structure 14:1835–1844
14. Brennan DF, Dar AC, Hertz NT, Chao WCH, Burlingame AL, Shokat KM, Barford D (2011) A Raf-induced allosteric transition of KSR stimulates phosphorylation of MEK. Nature 472:366–369
15. Chrencik JE, Patny A, Leung IK, Korniski B, Emmons TL, Hall T, Weinberg RA, Gormley JA, Williams JM, Day JE, Hirsch JL, Kiefer JR, Leone JW, Fischer HD, Sommers CD, Huang HC, Jacobsen EJ, Tenbrink RE, Tomasselli AG, Benson TE (2010) Structural and thermodynamic characterization of the TYK2 and JAK3 kinase domains in complex with CP-690550 and CMP-6. J Mol Biol 400:413–433
16. Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798–804
17. Šali A, Potterton L, Yuan F, van Vlijmen H, Karplus M (1995) Evaluation of comparative protein modeling by MODELLER. Proteins 23:318–326
18. Delano WL (2002) The PyMOL molecular graphics system. DeLano Scientific, Palo Alto, CA
19. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519
20. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:407–410
21. Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. Proteins 17:355–362
22. Wallner B, Elofsson A (2003) Can correct protein models be identified? Protein Sci 12:1073–1086
23. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118:11225–11236
24. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105:6474–6487
25. Jorgensen WL, Madura JD (1985) Temperature and size dependence for Monte Carlo simulations of TIP4P water. Mol Phys 56:1381–1392
26. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N center-dot. log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092
27. Tuckerman M, Berne BJ, Martyna GJ (1992) Reversible multiple time scale molecular dynamics. J Chem Phys 97:1990–2001

28. Oh SM, Zhu F, Cho YY, Lee KW, Kang BS, Kim H-G, Zykova T, Bode AM, Dong Z (2007) T-lymphokine activated killer cell originated protein kinase functions as a positive regulator of c-Jun-NH2-kinase 1 signaling and H-Ras-induced cell transformation. Cancer Res 67:5186–5194
29. Jørgensen AM, Topiol S (2008) Driving forces for ligand migration in the leucine transporter. Chem Biol Drug Des 72:265–272
30. Singh K, Kirubakaran P, Nagarajan S, Sakkiah S, Muthusamy K, Velmurgan D, Jeyakanthan J (2012) Homology modeling, molecular dynamics, e-pharmacophore mapping and docking study of Chikungunya virus nsP2 protease. J Mol Model 18:39–51
31. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 98:10037–10041
32. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res 32:665–667
33. Sitkoff D, Sharp KA, Honig B (1994) Accurate calculation of hydration free energies using macroscopic solvent models. J Phys Chem 98:1978–1988
34. Glasser L, Herraez A, Hanson RM (2009) Interactive 3D phase diagrams using Jmol. J Chem Educ 86:566–null
35. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein - ligand complexes. J Med Chem 49:6177–6196
36. Kozakov D, Hall DR, Beglov D, Brenke R, Comeau SR, Shen Y, Li K, Zheng J, Vakili P, Paschalidis IC, Vajda S (2010) Achieving reliability and high accuracy in automated protein docking: Cluspro, PIPER, SDU, and stability analysis in CAPRI rounds 13–19. Proteins 78(15):3124–3130
37. Kozakov D, Brenke R, Comeau SR, Vajda S (2006) PIPER: an FFT-based protein docking program with pairwise potentials. Proteins 65(2):392–406
38. Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 20(1):45–50
39. Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: a fully automated algorithm for protein-protein docking. Nucleic Acids Res 32(suppl 2):W96–W99
40. Schrödinger LLC (2011) QikProp, version 3.4. Schrödinger, LLC, New York, NY
41. Poroikov VV, Filimonov DA, Ihlenfeldt W-D, Gloriozova TA, Lagunin AA, Borodina YV, Stepanchikova AV, Nicklaus MC (2002) PASS biological activity spectrum predictions in the enhanced open NCI database browser. J Chem Inf Comput Sci 43(1):228–236
42. Larsson P, Wallner B, Lindahl E, Elofsson A (2008) Using multiple templates to improve quality of homology models in automated homology modeling. Protein Sci 17:990–1002
43. Venclovas C (2003) Comparative modeling in CASP5: progress is evident, but alignment errors remain a significant hindrance. Proteins 53:380–388
44. Mobarec JC, Sanchez R, Filizola M (2009) Modern homology modeling of G-protein coupled receptors: which structural template to use? J Med Chem 52:5207–5216
45. Kirubakaran P, Muthusamy K, Dhanachandra Singh K, Nagamani S (2012) Homology modeling, molecular dynamics, and molecular docking studies of Trichomonas vaginalis carbamate kinase. Med Chem Res 21:2105–2116
46. Curman D, Cinel B, Williams DE, Rundle N, Block WD, Goodarzi AA, Hutchins JR, Clarke PR, Zhou B-B, Lees-Miller SP, Andersen RJ, Roberge M (2001) Inhibition of the G2 DNA damage checkpoint and of protein kinases Chk1 and Chk2 by the marine sponge alkaloid debromohymenialdisine. J Biol Chem 276(21):17914–17919