In silico homology modelling and identification of Tousled-like kinase 1 inhibitors for glioblastoma therapy via high throughput virtual screening protein-ligand docking

Kamariah Ibrahim[1]

Abubakar Danjuma[2]

Chyan Leong Ng[3]

Nor Azian Abdul Murad[1]

Roslan Harun[1, 4]

Wan Zurinah Wan Ngah[1, 5]

and *Rahman Jamal[1,5]

[1]UKM Medical Molecular Biology Institute (UMBI), National University of Malaysia, Cheras, Jalan Yaacob Latiff, Bandar Tun Razak, 56000, Cheras, Kuala Lumpur, Malaysia

[2]Kulliyyah of Pharmacy, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang Darul Makmur

[3]Institute of Systems Biology, National University of Malaysia, 436000, Bangi,

[4]KPJ Ampang Puteri Specialist Hospital, Jalan Memanda 9, Taman Dato Ahmad Razali, 68000 Ampang, Selangor, Malaysia

[5]Faculty of Medicine, National University of Malaysia, 56000, Cheras, Kuala Lumpur

***Corresponding author:**

Rahman Jamal

Address: UKM Medical Molecular Biology Institute (UMBI), National University of Malaysia, Jalan Yaacob Latiff, Bandar Tun Razak, 56000, Cheras, Kuala Lumpur

Email: rahmanj@ppukm.ukm.edu.my

Phone: +603-91456321/9239

Fax: +603-91717185

Acknowledgements: This study was funded by the Higher Institution Centre of Excellence (HICoE) research grant (JJ-015-2011), Ministry of Education, Malaysia

Conflicts of interest: All members declare no conflict of interest

Keywords: Tousled-like kinase 1, TLK1, GBM, Homology modelling, *in silico* high throughput virtual screening

Abstract

Background: Glioblastoma multiforme (GBM) is a grade IV brain tumor that arises from star-shaped glial cells supporting neural cells called astrocytes. The survival of GBM patients remains poor despite many specific molecular targets that have been developed and used for therapy. Tousled-like kinase 1 (TLK1), a serine-threonine kinase, was identified to be overexpressed in cancers such as GBM. TLK1 plays an important role in controlling chromosomal aggregation, cell survival and proliferation. *In vitro* studies suggested that TLK1 is a potential target for some cancers; hence, the identification of suitable molecular inhibitors for TLK1 is warranted as a new therapeutic agents in GBM. To date, there is no structure available for TLK1. In this study, we aimed to create a homology model of TLK1 and to identify suitable molecular inhibitors or compounds that are likely to bind and inhibit TLK1 activity via *in silico* high-throughput virtual screening (HTVS) protein-ligand docking.

Methods: 3D homology models of TLK1 were derived from various servers including HOmology ModellER, i-Tasser, Psipred and Swiss Model. All models were evaluated using Swiss Model Q-Mean server. Only one model was selected for further analysis. Further validation was performed using PDBsum, 3d2go, ProSA, Procheck analysis and ERRAT. Energy minimization was performed using YASARA energy minimization server. Subsequently, HTVS was performed using Molegro Virtual Docker 6.0 and candidate ligands from ligand.info database. Ligand-docking procedures were analyzed at the putative catalytic site of TLK1. Drug-like molecules were filtered using FAF-Drugs3, which is an ADME-Tox filtering program. **Results and conclusion:** High quality homology models were obtained from the Aurora B kinase (PDB ID:4B8M) derived from *Xenopus levias* structure that share 33% sequence identity to TLK1. From the HTVS ligand-docking, two compounds were identified to be the potential inhibitors as it did not violate the Lipinski rule of five and the CNS-based filter as a potential drug-like molecule for GBM.

Background

Glioblastoma multiforme (GBM) is the most common primary brain tumor in adults. It is also classified as grade IV glioma which arises from the lineage of star-shaped glial cells known as astrocytes. The survival rate is very poor where only 15% of patients survived more than 24 months due to disease aggressiveness and heterogeneity of the disease (Ohgaki, & Kleihues 2007; Ohgaki et al. 2004). Although several molecular inhibitors have been developed to target aberrantly expressed enzymes and proteins, the results have been very frustrating (Li, & Tu 2015; Piccirillo et al. 2015). Factors contributing to resistant of GBM cells include deregulation of key signalling pathways, namely *PTEN*, *TP53*, *RB* and *PI3K-Akt* (Ohgaki, & Kleihues 2011; Smith et al. 2001), increased in the expression of anti-apoptotic proteins *BCL2* and *survivin* (Guvenc et al. 2013; Ruano et al. 2008), iterative perivascular growth within the highly vascularized brain (Baker et al. 2014), and presence of 30-65% constitutively active *EGFRvIII* mutant in GBM which secretes higher levels of invasion-promoting proteins (Sangar et al. 2014). Studies have revealed that Tousled-like kinase 1 (*TLK1*) is overexpressed in breast cancer (Wolfort et al. 2006), prostate cancer (Ronald et al. 2011), and cholangiocarcinoma (Takayama et al. 2010). In our previous study, we proved that *TLK1* is overexpressed in GBM and silencing of *TLK1* results in a significant decrease in invasion, migration and GBM cells survival (Ibrahim et al. 2013).

Human TLK1 contains 766 amino acids and is one of the members of the Tousled-like kinase family consisting of TLK1 and TLK2 (Pruitt et al. 2012). The gene is mapped on chromosome 2q31.1 and encoded by 25 exons. TLK1 share 85% sequence identity to TLK2, and both share ~50% sequence identity with *Arabidopsis thaliana* where Tousled-like kinase family was initially identified (Takahata, Yu & Stillman 2009). This serine-threonine kinase is an important signalling regulator mainly involved in the cell cycle regulation, cellular mitosis, cell

survival, and proliferation (Sunavala-Dossabhoy, & De Benedetti 2009). In general, the N-terminal domain of Tousled-like kinase is well conserved to include three potential nuclear localization sequences and three putative coiled-coil regions, while the C-terminus region contains the catalytic ATP-binding domain at the region that consists of 456 to 734 amino acid residues. The active binding site is located within the protein kinase domain sequence (Silljé et al. 1999). This ~90 kDa kinase is activated by the *CHK1/ATM DNA* damage pathway (Groth et al. 2003). TLK1 interacts with its substrates, namely Asf1, histone H3 (Carrera et al. 2003), and Rad9 (Sunavala-Dossabhoy, & De Benedetti 2009) to activate DNA damage and DNA repair activity (De Benedetti 2012). It was suggested that when overexpressed, *TLK1* is involved in radioprotection and chemo-resistance of cancer cells (Y. Li et al. 2001; Ronald et al. 2011). Unfortunately, the structure of TLK1 has not been elucidated and this hinders the full understanding of TLK1 biological processes. Nonetheless, the X-ray diffraction data for the kinase domain of human TLK1 family member TLK2 have been recently reported which may shed a light on structural understanding of human Tousled-like kinase (Garrote et al. 2014). No structure is yet available for both TLK1 and TLK2, hence, we perform a homology modelling study of TLK1 structure to understand its function in orchestrating cellular functions particularly in cancer pathways. In this study, we present a structural homology model of the TLK1 catalytic binding domain which may serve as a potential target for molecular inhibitors. We then used the proposed structure to identify potential inhibitors for TLK1 by utilising *in silico* ligand-docking with high throughput virtual screening (HTVS) targeting more than 16,000 candidate compounds.

Materials and methods

Template identification and homology modelling

The amino acid sequence of human TLK1 was retrieved from UniProt with the accession number: Q9UKI8 (http://www.uniprot.org/).The TLK1 FASTA format amino acid sequence was downloaded into the BLASTP and PSI-BLAST search (http://blast.ncbi.nlm.nih.gov/) in order to identify the homologous proteins. An appropriate template for TLK1 was identified based on the e-value and sequence identity ranging from 30% to 33% at the protein kinase domain indicating similarity of structure and function. The template and the target sequences were later aligned using the Clustal Omega program (http://www.ebi.ac.uk/Tools/msa/clustalo/). Subsequently, homology modelling was carried out against the chosen template using HOmology ModellER (Tosatto 2005), I-Tasser (Zhang 2009), and PsiPred (Buchan et al. 2010).

Homology models quality estimation

The model quality estimation was performed using the Swiss-Model Qualitative Model Energy Analysis (Q-Mean) Server based on the composite scoring function, which derives a quality estimation on the basis of the geometrical analysis of single models (Benkert, Biasini & Schwede 2011). It also describes the major geometrical aspects of the protein structures. Five different structural descriptors were used. The local geometry was analyzed using the torsion angle potential function over three consecutive amino acids. A secondary structure-specific distance-dependent pairwise residue-level potential was used to assess long-range interactions. A solvation potential describes the burial status of the residues. Two simple terms describing the agreement of predicted and calculated secondary structure and solvent accessibility, were also included. In comparison with other protein structure evaluation servers, the QMEAN

shows a statistically significant improvement over nearly all quality measures describing the ability of the scoring function to identify the native structure and to discriminate good from bad models (Benkert, Tosatto & Schomburg 2008). 3D structure was then visualized using PyMol software (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

Validation of modelled structure

The best homology model created was used for further investigation. We used the latest version of PDBsum (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/) which provides further information on protein function prediction, structural topology, PROCHECK and cleft analysis. We also used ProSA which displays scores and energy plots that highlight potential problems spotted in protein structures (Wiederstein, & Sippl 2007). Prediction of the protein structure function was performed using proteo-genomic analysis software 3d2go (http://www.sbg.bio.ic.ac.uk/phyre/pfd/html/help.html). This allowed full structural scan of the protein structure made against the Structural Classification of Proteins (SCOP) database using a modified version of BLAST (Tung, Huang & Yang 2007). Energy minimization was performed on YASARA server (http://www.yasara.org/minimizationserver.php).

High throughput *in silico* ligand-docking analysis

In silico ligand-docking analysis was performed using Molegro Virtual Docker (MVD version 2013.6.0) to predict protein-ligand interactions. The potential binding sites of selected proteins and candidate small molecules were characterized by the molecular docking algorithm called MolDock which was derived from "*Piecewise Linear Potential* (Sundarapandian et al. 2010). The MolDock score refers to the approximate binding energies between protein and ligand

which is usually expressed in kcal/mol. This software handles all aspects of the docking process from the preparation of the molecules to determine the potential binding site of the target protein, and the predicted binding modes of the ligand. Interestingly, MVD has been shown to provide higher accuracy compared with the other commercially available docking softwares e.g. Glide, Surflex and FlexX (Sivaprakasam, Tosso & Doerksen 2009). Docking requires five steps; importing molecules, importing ligands, molecular preparation, creating template and docking.

Candidate ligands for ligand-docking screening were downloaded from Ligand.Info (http://ligand.info/) which compiles various publicly available databases of small molecules and compounds from ChemBank, KEGG, ChemPDB, Drug-likeness NCI subset and non-annotated NCI subset (von Grotthuss, Pas & Rychlewski 2003). We downloaded a total of 16,358 sdf. format small molecules from KEGG ligands (10,005), ChemBank (2,344) and ChemPDB (4,009) for high throughput screening of potential inhibitors for TLK1. Due to the large number of candidate KEGG ligands, we filtered out some of these compounds based on the relevancy to the present TLK1 3D model using Findsite server (Brylinski, & Skolnick 2008) as a pre-molecular docking step. After filtering these ligands, only 1,386 KEGG ligands were selected for further investigation. Most of the ligands in the database as well as the homology model or molecule did not have correct bond orders and bond angles. Hence, full optimization of molecules and ligand preparation was performed using Molegro Virtual Docking software default setting whereby appropriate missing hydrogen atoms were added, missing bonds were assigned, partial charges were added if necessary and flexible torsions in ligands detected.

Docking study was performed at the catalytic domain of TLK1. Simulation on the modelled protein identified five cavities as potential binding sites. However, only one cavity was used for the ligand-docking study i.e. the cavity with the largest surface area and volume of 214.528 arbitary unit within the catalytic domain sites of TLK1. The predicted sites had a grid resolution of 0.3Å and a binding site of 15Å radius from the template. The Moldock optimizer was used as a search algorithm and the number of runs was set to 10 with a maximum iteration of 1000, scaling factor of 0.50, 0.90 cross over and a population size of 50. The maximum number of poses generated was 5. Potential ligands were selected based on the best MolDock score value that is less than -170.

Visualization of ligand-protein interaction

The three-dimensional and two-dimensional visualisation of ligand-protein interaction were performed using the Maestro software package (Maestro, version 10.4, Schrödinger, LLC, New York, NY, 2015).

***In silico* bioavailability study**

Lead molecules identified from the high throughput ligand-docking screening were subjected to further *in silico* filtering to identify those with the best values in terms of their absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). This was done using the FAF-Drugs3 (November 2014 edition) which is a free ADME-Tox (Miteva et al. 2006) filtering tool. This step will ensure the suitability of lead molecules based on toxicity for future *in vivo* applications. We applied Lipinski's Rule of Five (Lipinski et al. 2001) to remove some reactive groups and compounds. We have also included the Central Nervous System (CNS) drugs physicochemical criteria (Jeffrey, & Summerfield 2010; Pajouhesh, & Lenz 2005), which

includes (1) molecular mass less than 450 Da, (2) partition coefficient (logP) of 0.2 -6.0, (3) hydrogen bond donors not less than three, (4) hydrogen bond acceptors not less than five and (5) topological surface area (tPSA) within 3-118.

Results

Homology modelling of TLK1 serine/threonine kinase

The PSI-BLAST results of TLK1 sequence Q9UKI8 were analysed and we selected the protein hits based on query coverage, similarity and identity. The model structure which was selected showed sequence identity and similarity that ranged from 27% to 37% and a query coverage E-value that ranged from 4e-29 to 9e-15 and covered only the protein kinase domain site (450-756). The homology model was created based on the TLK1 protein kinase catalytic domain sequence. We selected 40 protein sequence templates for homology modelling using various softwares. However, only 18 models were successfully created using HOmology ModellER and i-tasser. We evaluated all the 18 models using Q-Mean Server and identified the Aurora B kinase structure from African clawed frog *Xenopus levias* (PDB ID: 4B8M) as the best template structure for TLK1 producing a Total QMean Score of 0.68 out of 1.0 required for an excellent homology model **(Table 1).** The Aurora B kinase that in complex with inner centromere protein A (VX-680) was determined to 1.85 Å resolution (PDB ID: 4B8M). Pro-Motif analysis showed that the modelled TLK1 structure, with 270 amino acids, contains 4 beta-hairpins, 6-beta bulges, 10 strands, 14 helices, 15 helix-helix interactions, 16 beta-turns and 3 gamma turns **(Figure 1A and 1B).**

The homology model of TLK1 was also assessed using ProSA Z-score. The overall Z-score quality was -4.92 suggesting a good quality model compared with the available structure from

NMR and X-ray **(Figure 2A and 2B).** Ramachandran plot obtained from PROCHECK analysis achieved a good quality model assessment of 90.1% in the favoured region **(Figure 2C).** The plot represents the *psi* and the *phi* angles of the amino acid residues. Details of the analysis plot can be referred to **Table 2.** Analysis from the three dimensional structural superposition (3d-ss) web server (Sumathi et al. 2006) showed the root mean square deviation (RMSD) between template structure and the 3D homology model structure to be 0.543 Å **(Figure 2D).** ERRAT overall quality factor is 53.696% and at least more than 80% of the amino acids have scores more than or equal to 0.2 in the 3D/1D profile. The YASARA public server for energy minimization provided a value of 16140271100.5 kJ/mol to 143790.2 kJ/mol with a score of -1.53 to -0.95.

Proteogenomic analysis

Functional analysis of the TLK1 modelled structure performed using 3d2go web server identified the following activities with the highest confidence value of 1.0: phosphotransferase activity alcohol group as the acceptor, protein amino acid phosphorylation, protein serine/threonine kinase activity and nucleotide binding. Nucleus and protein binding functions were predicted with a confidence value of 0.89. Functional prediction in cell cycle, mitosis, phosphoinositide-mediated signalling (confidence value of 0.86), centrosome, spindle organization, regulation of protein stability, ubiquitin protein ligase binding (confidence value of 0.85) were all in concordance with experimental data (Kelly, & Davey 2013; Pilyugin et al. 2009). These findings were predicted to be similar with the function of human Aurora kinase2 (PDB ID: 2J4Z). Interestingly, with a confidence value of 0.79, the modelled TLK1 structure is also predicted to be involved in insulin receptor signalling pathway and actin cytoskeleton organization which is similar to the human PDK1 (PDB ID:1UU3). This indicates that TLK1

could be involved in the regulation of actin filament organization particularly in controlling cancer cell motility.

High throughput virtual ligand-docking screening

The cut-off point of the MolDock docking scoring was set at less than -170 to select ligands that predicted to have high binding affinity to TLK1. We identified 192 lead molecules, and ATP was the top scoring molecule in the docking procedure with a MolDock score of -193.654. The amino acid residues that found to involve in the protein-ligand interactions were GLY463, ARG464, GLY465, GLY466, PHE467, SER468, GLU469, VAL470 and LYS485. The compounds that utilized in the screening were initially not known until we have completed the identification procedure. The results showed that ATP docked accurately within the cavity, suggesting the robustness of the *in silico* experiment.

In silico pharmacokinetic analysis

The 192 compounds with the best MolDock scores were submitted to the Free ADME-Tox filtering tool 3 (November 2014 edition) for pharmacokinetic analysis. Analysis were subjected to the Lipinski's Rule of Five (RO5) (Lipinski et al. 2001) and filters for CNS drugs (Jeffrey, & Summerfield 2010; Pajouhesh, & Lenz 2005) to ensure the efficacy and safety of the candidate compounds. The final filtering process revealed that only two compounds passed this assessment without violating the general Lipinski's RO5 and the CNS rule. These compounds were identified as ID352 and ID1652 from the ChemBank database **(Table 3).** Their chemical structures, IUPAC names, the radar plot of physicochemical analysis, oral absorption estimation data and the Pfizer 3/75 Rule Positioning plot, which estimated drug-like molecules that are likely to cause toxicity and experimental promiscuity, are presented in **Figure 3A-H.**

ID1652 is known as beraprost which is a prostacyclin analogue used in the treatment of arterial hypertension (Galiè et al. 2002). It has a better docking score, with no violation of Lipinki's rule of five and a low promiscuous toxicity as compared to ID352 or bepridil which is a calcium channel blocker for anti-angina (Rae et al. 1985). Beraprost also has a better hydrogen bonding score from the ligand-docking simulation. Results from receptor-ligand interactions **(Figure 4)** revealed a common cavity for ATP, ID352 and ID1652 binding. The residues that are involved in the interactions include GLY465, GLY466, PHE467, SER468, VAL470, and LYS485. These suggested that both of the two compounds bind to catalytic site of TLK1 ATP binding pocket.

Discussion

GBM remains as the solid tumour with the poorest survival in adults since the past few decades. The search for the right molecular target is still ongoing and one of the many approaches is by using computer-aided drug discovery tools. Our recent *in vitro* study identified TLK1 as a potential target for glioblastoma multiforme. We found *TLK1* to be overexpressed and the knockdown of *TLK1* reduced cellular proliferation and invasion (Ibrahim et al. 2013). An auto-phosphorylated chemical inhibition screen on recombinant TLK1B, which is a known splice variant, has been performed by Ronald et al, using more than 6,000 compounds. This study identified four inhibitors belonging to the class of phenothiazine antipsychotics that are structurally and chemically similar. The same study also showed that thioridazine was able to sensitize prostate cancer cells when used with doxorubixin (Ronald et al. 2013). Although chemical library screening for drug discovery seems promising, it is very expensive and time consuming. A study using the ChemBL database and Kinase SaRfari application identified 74 "hits" compounds that can potentially bind to TLK1 (Bento et al. 2014). However, no details

were reported on the specific biding sites and the specific TLK1 structure that were used for the screen. In this study we used a computational approach to identify suitable TLK1 inhibitors based on a homology model that has been created.

The 3D structure of TLK1 is currently not available for drug design strategy, hence we used 18 PDB templates that shared 30% to 33% sequence identity, to create homology models of TLK1. As a result, Aurora B kinase (PDB ID: 4B8M) was identified as the most suitable homology template by the HOmology modellER server. This model allows us to perform ligand-docking analysis to identify potential inhibitors for TLK1.

One of the major challenges for optimal therapeutic intervention for glioblastoma and other types of brain tumor is to achieve maximal penetration across the blood brain barrier (BBB). The BBB is a structure composed of endothelial cells which is associated with perivascular neurons, pericytes and astrocytic end-feet processes. The endothelial cells connected by tight junctions form an almost impenetrable barrier to all compounds except highly lipidized small molecules of less than 400 Da (Nathanson, & Mischel 2011). Although many studies have identified drug-like molecules from high throughput virtual screening, most only follow the Lipinski's rule of five and have neglected the probability calculations for the molecules to cross the BBB. This eventually led to dismal results in *in vivo* studies (Gidda et al. 1995; Pardridge 1998). We used the recent version of the free ADME-TOX software and utilized the CNS filter to identify drug-like molecules that are able to cross the BBB. With this approach we identified bepridil and beraprost as the two compounds which may bind specifically at the catalytic site of TLK1 receptor protein and also fulfilled the CNS drugs selection criteria (Jeffrey, & Summerfield 2010; Pardridge 1998). We observed that more than 80% of the interactions

involved between ligands and receptor are hydrophobic. We have also identified other lead compounds for TLK1 such as the imidazole-pyrrole polyamide derivatives with better binding affinity (with Moldock Score of -208.44 to -209.34) compared to bepridil and beraprost. Unfortunately, these compounds violated the Lipinski's Rule of Five and have molecular masses of more than 450 Da which are not suitable to cross the blood brain barrier.

Beraprost, an analogue to prostacyclin or PGI_2, is commonly used for arterial pulmonary hypertension and has multiple physiological effects such as endothelial vasodilation, inhibition of platelet aggregation, leukocyte adhesion, and vascular smooth muscle cell proliferation (Wang et al. 2011). Activation of the PGI_2 signalling pathway by beraprost sodium suppressed lung cancer metastases by preventing maturation of angiogenesis (Yoshinori Minami et al. 2012). It was also reported to enhance permeability and retention (EPR) of solid tumors by decreasing tumor blood flow by 70%, hence inhibiting tumor growth. Morever, it did not affect normal cells and systemic blood flow (Tanaka et al. 2003). Since this compound mimics structurally related lipid soluble hormone PGI_2, it was predicted that the efficacy of the compound will be high as it will be able to cross the BBB (Moga 2013).

Bepridil is a known sodium-calcium channel blocker that is use for anti-arrythmias. An earlier study reported that bepridil caused tumor growth inhibition in neuroblastoma and astrocytoma cells by causing a prolonged increase in free intracellular calcium concentration when cells were co-treated with anti-estrogens (Yong, & Wurster 1996). Bepridil has been experimentally found to bind to the N-domain pocket of cardiac troponin C but with negative cooperativity (Varguhese, & Li 2011). Even though, theoretically, bepridil can cross blood brain barrier effectively (Muehlbacher et al. 2012), our findings showed that it may have non-specific

binding properties towards TLK1. Hence, it will be an added value if some chemical modification can be made to increase its selectivity towards TLK1. It is worth to note that S-bepridil was found to have a higher binding affinity towards the p53 binding domain in MDM2 (Warner et al. 2012). In order to enhance binding affinity between TLK1 receptor and these two identified ligands, as well as preventing cross binding towards other types of receptors, modification of current ligand structure by QSAR fragment based on pharmacophore analysis is warranted for future study.

This study has identified potential inhibitors that binds at the catalytic site of TLK1. However, identification of inhibitors that can bind to the non-catalytic component of a particular kinase would also be useful as they would also play significant roles in the regulation of cellular functions (Romano, & Kolch 2011). Further studies of TLK-ligand complex structure will allow identification of allosteric inhibition sites to provide much specific TLK1 regulatory inhibitory effects.

Conclusion

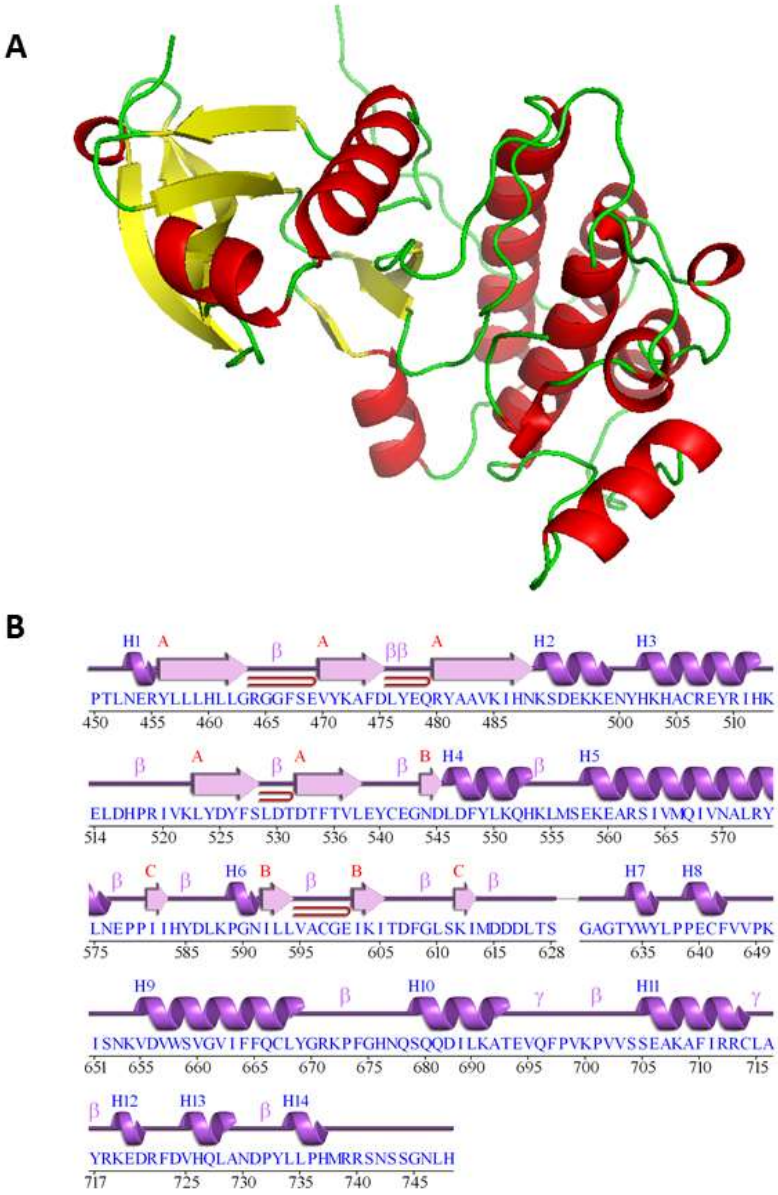
We have successfully created a 3D structure for the catalytic domain for TLK1 which was predicted to be a potential molecular target for GBM. We have performed vigorous analysis to determine the suitability and stability of the modelled structure through various quality control platforms. We identified beraprost and bepridil as the two candidate compounds that will bind to TLK1. These two drugs are commonly used for cardiovascular diseases. Further *in vitro* and *in vivo* studies need to be performed to validate the therapeutic value of these compounds for GBM.

Figures and Tables

Table 1: Top 20 models generated from two homology modelling servers; Homology modeller (HOMER) and i-Tasser. TLK1homer4B8M was selected as our homology model for subsequent analysis.

Model name	C_beta interaction energy	All-atom pairwise energy	Solvation energy	Torsion angle energy	Secondary structure agreement	Solvent accessibility agreement	Total QMEAN-score
TLK1homer4B8M	-61	-5727.18	-16.37	-16.13	89.30%	79.30%	0.68
TLK1homer4FR4	-74.72	6094.74	-11.26	-17.32	85.70%	77.60%	0.648
TLK1homer4DFX	-54.53	-6539.34	-22.18	-20.69	85.90%	77.10%	0.634
TLK1homer3SOA	-45.31	-5046.76	-4.80	-0.19	81.20%	78.80%	0.625
TLK1homer4M7N	-89.83	-5128.66	-15.31	-0.89	80.80%	79.10%	0.617
TLK1homer4FGB	-79.38	-6276.79	-16.15	-14.23	79.30%	77.50%	0.61
TLK1homer4L44m	-62.07	4967.85	-8.67	-4.86	86.60%	75.60%	0.604
TLK1homer3Q5I	-67.53	5211.28	-3.89	-6.1	77.20%	77.20%	0.596
TLK1homer4KIKB	-46.22	-3504.47	-4.66	-7.84	74.80%	77.20%	0.585
TLK1homer3TAC	-67.15	-5787.83	-12.33	5.77	81.70%	76.90%	0.582
TLK1homer1KOB	-67.51	-5454.16	-17.93	10.78	80.30%	77.30%	0.566
TLK1homer2Y94	-88.61	-5646.09	-14.67	13.09	81.40%	76.30%	0.558
TLK1homer2YCF	-71.96	-5698.7	-1.22	0.66	76.90%	74.40%	0.551
TLK1homer4EQC	-44.07	-5230.6	-15.55	0.2	79.40%	72.60%	0.511
TLK1homer3ZDU	-20.33	-3543.81	6.09	0.85	79.50%	70.90%	0.509
TLK1homer2ETR	-57.27	-5573.2	-12.92	-18.61	76.30%	68.60%	0.471
TLK1homer4FIE	-48.17	-4413.68	-7.96	-12.1	79.60%	78.60%	0.471
TLK1homer3I6U	-75.46	-6091.17	-6.14	4.72	68.40%	70.50%	0.443
tlk1model2itasser	-211.76	-9701.51	-35.09	43.13	77.80%	65.40%	0.371
tlk1model1itasser	-114.9	-6928.17	-22.73	29.02	71.00%	61.60%	0.294

383



384

385 **Figure 1: (A)** Secondary structure of TLK1 homology model generated from Homology
386 Modeller server. Visualization was performed using The PyMOL Molecular Graphics System,
387 Version 1.5.0.4 Schrödinger, LLC.; b-sheets, alpha-helices and loops are in yellow, red and
388 green respectively. **(B)** Depiction of the amino acid residues that used in secondary structure
389 analysed from Pro-Motif analysis using PDBsum server.

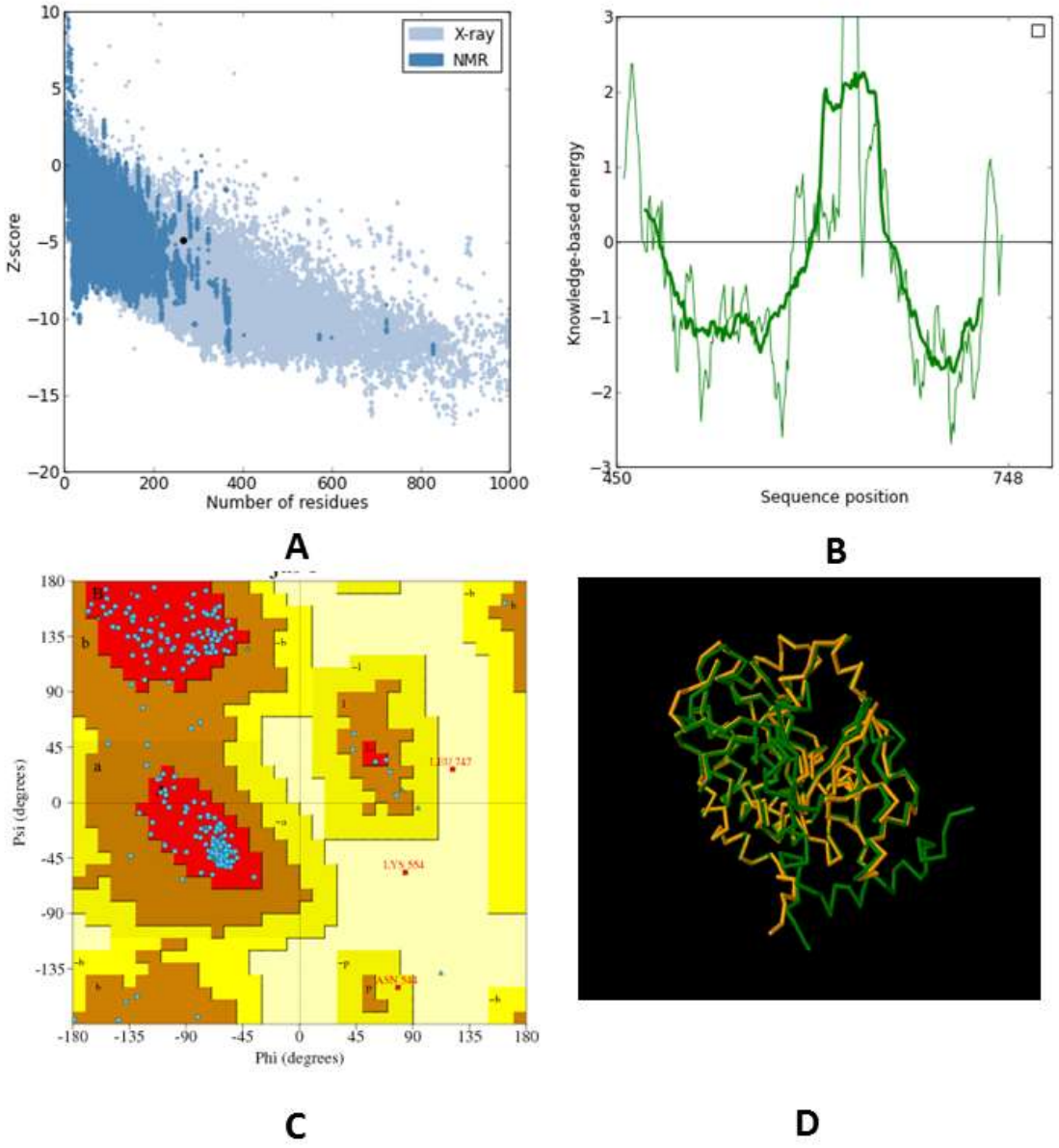


Figure 2: (A) ProSA shows the overall quality model of TLK1 with score of -4.92 **(B)** ProSA comparison results of energy-plots for TLK1 model structure with the PDB ID: 4B8M. **(C)** Ramachandran plot analysis using PROCHECK shows 90.1% of amino acids are generously in the allowed region. **(D)** 3D structural superposition of Aurora B kinase (PDB ID:4B8M) (green) and homology model of TLK1 (yellow).

Table 2: Ramachandran plot statistics of TLK1 homology model structure obtained from PROCHECK analysis.

Parameter	Value in percentage
Most favoured region	90.1
Additional allowed region	8.7
Generously allowed region	0.4
Disallowed region	0.8
Amino acid residues accepted in the analysis	242 out of 270
G-factor average score	0.22
Main chain bond angles	0.41
Main chain bond lengths	0.62

Table 3: Lead molecules with their docking scores and amino acids interaction identified. In bold, are common residues that involved in the ATP, Bepridil and Beraprost binding.

Lead molecules ID	Chemical name	MolDock Score	Rerank Score	H-Bond score	Amino acids involved in interaction
352	Bepridil	-170.518	-109.678	-2.5	**GLY465**, **GLY466, PHE467**, **SER468, VAL470, LYS485**, **HIS487, GLU496, TYR501, HIS502, HIS504, ALA505**, **TYR509**, GLU508, HIS512, LEU523, THR536, LEU538, THR606, ASP607, PHE608,
1652	Befaprost/ Beraprost	-181.124	37.981	-5.35	**GLY465**, **GLY466, PHE467**, **SER468, VAL470, LYS485**, **HIS487, GLU496, TYR501, HIS502, HIS504, ALA505,** CYS506, **TYR509,** SER528, THR533
*367 (Control)	ATP	-186.431	-49.1549	-9.386	GLY463, ARG464, **GLY465**, **GLY466, PHE467**, **SER468**, GLU469, **VAL470, LYS485**

412 **Table 4:** Physiochemical properties of ligands from the docking study that passes ADME-TOX

413 Lipinski rule of five and CNS filtering.

Parameters	ID352	ID1652
MW	366.54	402.52
logP	5.31	4.1
logSw	-4.94	-4.33
tPSA	16.91	89.52
Rotatable bonds	10	10
Rigid Bonds	17	16
Flexibility	0.37	0.38
HB Donors	0	3
HB Acceptors	3	5
HBD_HBA	3	8
Number of system ring	3	1
Max Size System Ring	6	12
Charge	1	1
Total charge	1	-1
Heavy atoms	27	29
C atoms	24	24
Heteroatoms	3	5
Ratio H/C	0.12	0.21
Lipinski violation	1	0
Solubility mg/ml	2613.49	5304.9
Solubility forecast index	Reduced solubility	Reduced solubility
Phospholipidosis	Non-inducer	Non-inducer
Stereocenters	1	6
iPPI	No	No
Status	Accepted	Accepted

414

415

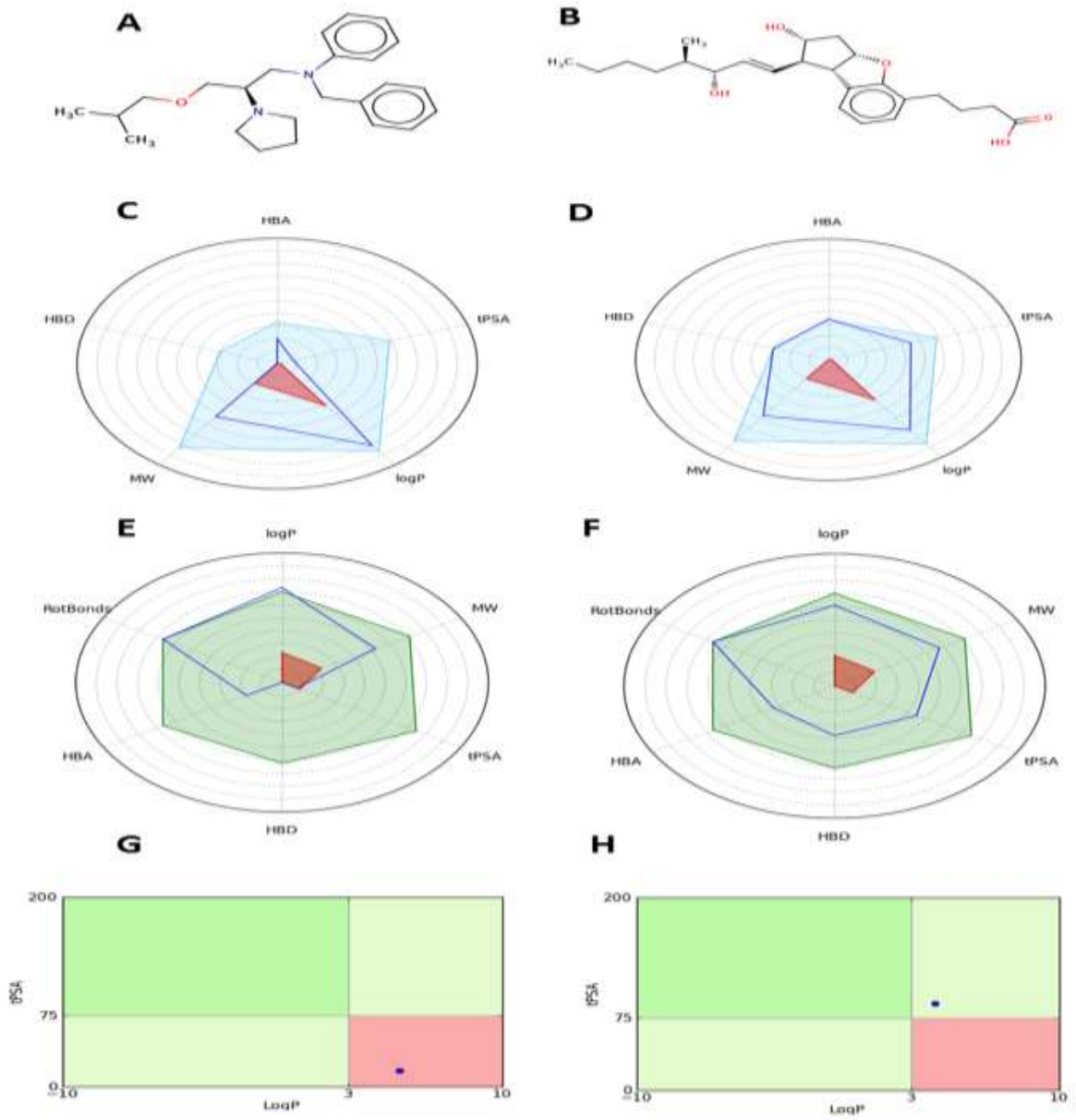
416

417

418

419

420



421

422 **Figure 3: (A)** and **(B)** Structure of identified compounds ID352; *N-benzyl-N-(3-isobutoxy-2-*
423 *pyrrolidin-1-yl-propyl)aniline* and ID1652; *2,3,3a,8b-tetrahydro-2-hydroxy-1-(3-hydroxy-4-*
424 *methyl-1-octen-6-ynyl)-1H-cyclopenta(b)benzofuran-5-butanoic acid* respectively. **(C)** and
425 **(D)** Physico-chemical profile of compounds ID352 and ID1652, respectively. A radar plot
426 representing the computed compound profile blue line that should cover within the CNS filter
427 area in red and must be within the blue field. (**E)** and (**F)** Oral absorption estimation of ID352
428 and ID1652, whereby the compound values should fall within RO5 and Veber rule area; light

green and red area. **(G)** and **(H)** Shows oral bioavailability profile (compound blue dot should fall within the optimal dark green and light green area and red ones being extreme zones generally indicating low oral bioavailability). ID352 were predicted to cause toxicity compared to ID1652 whereby dot plot falls within the green area which is less likely to cause toxicity.

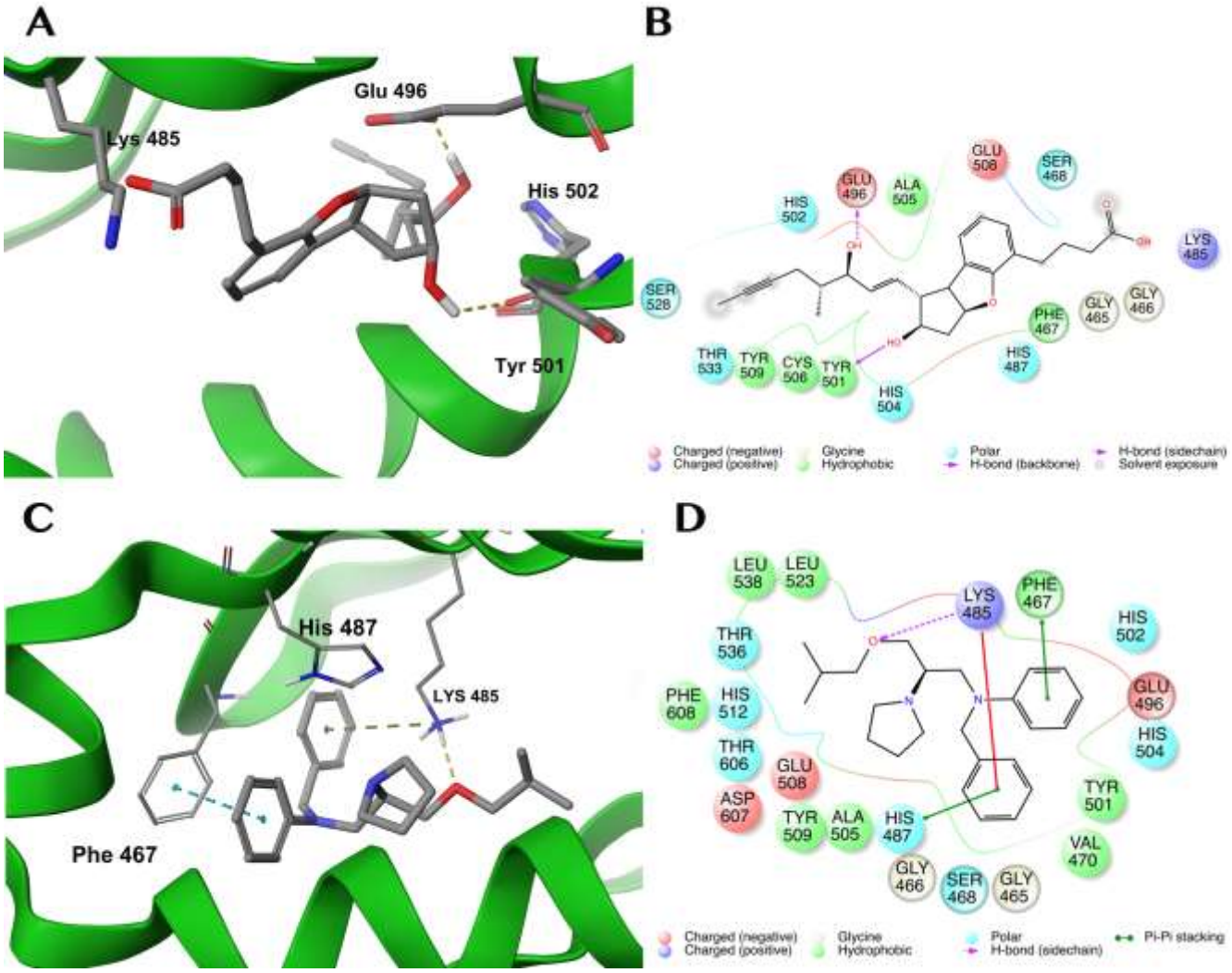


Figure 4: (A) 3D binding mode of the ligands ID1652 in the ATP binding site of the homology modeled TLK1 protein. The docking pose between ligand 1652 and the ATP binding site of TLK1 protein shows two backbone hydrogen bonds between the ligand and TYR501 and GLU496. **(B)** 2D ligand interaction diagram showing presence of hydrophobic interactions between the ligand and PHE467, ALA505, TRY501, CYS506 and TYR509. **(C)** 3D docking pose between ligand 352 and the ATP binding site of TLK1 showing an aromatic-aromatic and amino-aromatic interactions between the ligands and PHE467 and HIS487 respectively. There is also a hydrogen bond between the ligand and the LYS485. **(D)** 2D ligand interaction diagram showing hydrophobic interactions between the ligand and LEU523, LEU538, VAL470,

PHE467, TYR501, TYR509, ALA505 and PHE608. The fact that ligand ID1652 has more activity than ligand 353 demonstrated the importance of hydrogen bonding rather than the aromatic-aroamtic and amino-aromatic interactions. Visualization of ligand-protein interaction. The three-dimensional and two-dimensional visualisation of ligand-protein interaction were performed using Maestro software package (Maestro, version 10.4, Schrödinger, LLC, New York, NY, 2015).

References

Baker, G. J., Yadav, V. N., Motsch, S., Koschmann, C., Calinescu, A.-A., Mineharu, Y., Camelo-Piragua, S. I. et al. 2014. Mechanisms of glioma formation: iterative perivascular glioma growth and invasion leads to tumor progression, VEGF-independent vascularization, and resistance to antiangiogenic therapy. *Neoplasia (New York, N.Y.)*, *16*(7), 543–561. doi:10.1016/j.neo.2014.06.003

Benkert, P., Biasini, M. & Schwede, T. 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics (Oxford, England)*, *27*(3), 343–350. doi:10.1093/bioinformatics/btq662

Benkert, P., Tosatto, S. C. E. & Schomburg, D. 2008. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*, *71*(1), 261–277.

Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., Krüger, F. A. et al. 2014. The ChEMBL bioactivity database: an update. *Nucleic acids research*, *42*(Database issue), D1083–90.

Brylinski, M. & Skolnick, J. 2008. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(1), 129–134. doi:10.1073/pnas.0707684105

Buchan, D. W. A., Ward, S. M., Lobley, A. E., Nugent, T. C. O., Bryson, K. & Jones, D. T. 2010. Protein annotation and modelling servers at University College London *38*(May), 563–568. doi:10.1093/nar/gkq427

Carrera, P., Moshkin, Y. M., Gronke, S., Sillje, H. H. W., Nigg, E. A., Jackle, H. & Karch, F. 2003. Tousled-like kinase functions with the chromatin assembly pathway regulating nuclear divisions. *Genes & development*, *17*(20), 2578–2590. doi:10.1101/gad.276703

De Benedetti, A. 2012. The Tousled-Like Kinases as Guardians of Genome Integrity. *ISRN molecular biology*, *2012*, 627596.

Galiè, N., Humbert, M., Vachiéry, J.-L., Vizza, C. D., Kneussl, M., Manes, A., Sitbon, O. et al. 2002. Effects of beraprost sodium, an oral prostacyclin analogue, in patients with pulmonary arterial hypertension: a randomized, double-blind, placebo-controlled trial. *Journal of the American College of Cardiology*, *39*(9), 1496–1502.

Garrote, A. M., Redondo, P., Montoya, G. & Muñoz, I. G. 2014. Purification, crystallization and preliminary X-ray diffraction analysis of the kinase domain of human tousled-like kinase 2. *Acta crystallographica. Section F, Structural biology communications*, *70*(Pt 3), 354–357.

Gidda, J. S., Evans, D. C., Cohen, M. L., Wong, D. T., Robertson, D. W. & Parli, C. J. 1995. Antagonism of serotonin3 (5-HT3) receptors within the blood-brain barrier prevents cisplatin-induced emesis in dogs. *The Journal of pharmacology and experimental therapeutics*, *273*(2), 695–701.

Groth, A., Lukas, J., Nigg, E. a., Silljé, H. H. W., Wernstedt, C., Bartek, J. & Hansen, K. 2003. Human Tousled like kinases are targeted by an ATM- and Chk1-dependent DNA damage checkpoint. *EMBO Journal*, *22*(7), 1676–1687. doi:10.1093/emboj/cdg151

Guvenc, H., Pavlyukov, M. S., Joshi, K., Kurt, H., Banasavadi-Siddegowda, Y. K., Mao, P., Hong, C. et al. 2013. Impairment of glioma stem cell survival and growth by a novel inhibitor for Survivin-Ran protein complex. *Clinical cancer research : an official journal of the American Association for Cancer Research*, *19*(3), 631–642. doi:10.1158/1078-0432.CCR-12-0647

Ibrahim, K., Mat, F. C., Harun, R., Ngah, W. Z. W., Mokhtar, N. M. & Jamal, R. 2013. Silencing of Tousled-like Kinase 1 (TLK1) Reduces Survival, Migration and Invasion of Glioblastoma multiforme cells. *Asia Pacific Journal of Molecular Medicine*, *41*(1st National Conference for Cancer Research 5th Regional Conference on Molecular Medicine RCMM).

Jeffrey, P. & Summerfield, S. 2010. Assessment of the blood-brain barrier in CNS drug discovery. *Neurobiology of disease*, *37*(1), 33–37. doi:10.1016/j.nbd.2009.07.033

Kelly, R. & Davey, S. K. 2013. Tousled-like kinase-dependent phosphorylation of Rad9 plays a role in cell cycle progression and G2/M checkpoint exit. *PloS one*, *8*(12), e85859. doi:10.1371/journal.pone.0085859

Li, Q. & Tu, Y. 2015. Genetic Characteristics of Glioblastoma: Clinical Implications of Heterogeneity. *Cancer Translational Medicine*, *1*(5), 176. doi:10.4103/2395-3977.168573

Li, Y., DeFatta, R., Anthony, C., Sunavala, G. & De Benedetti, A. 2001. A translationally regulated Tousled kinase phosphorylates histone H3 and confers radioresistance when overexpressed. *Oncogene*, *20*(6), 726–738. doi:10.1038/sj.onc.1204147

Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, *46*(1-3), 3–26.

Miteva, M. A., Violas, S., Montes, M., Gomez, D., Tuffery, P. & Villoutreix, B. O. 2006. FAF-Drugs: free ADME/tox filtering of compound collections. *Nucleic acids research*, *34*(Web Server issue), W738–44. doi:10.1093/nar/gkl065

Moga, T. 2013. The 2-Series Eicosanoids in Cancer: Future Targets for Glioma Therapy? Journal of Cancer Therapy. *Journal of Cancer Therapy*, 338–352SRC – GoogleScholar.

Muehlbacher, M., Tripal, P., Roas, F. & Kornhuber, J. 2012. Identification of Drugs Inducing Phospholipidosis by Novel in vitro Data. *Chemmedchem*, *7*(11), 1925–1934. doi:10.1002/cmdc.201200306

Nathanson, D. & Mischel, P. S. 2011. Charting the course across the blood-brain barrier. *The Journal of clinical investigation*, *121*(1), 31–33. doi:10.1172/JCI45758

Ohgaki, H., Dessen, P., Jourde, B., Horstmann, S., Nishikawa, T., Di Patre, P.-L., Burkhard, C. et al. 2004. Genetic pathways to glioblastoma: a population-based study. *Cancer research*, *64*(19), 6892–6899. doi:10.1158/0008-5472.CAN-04-1337

Ohgaki, H. & Kleihues, P. 2007. Genetic pathways to primary and secondary glioblastoma. *The American journal of pathology*, *170*(5), 1445–1453. doi:10.2353/ajpath.2007.070011

Ohgaki, H. & Kleihues, P. 2011. Genetic profile of astrocytic and oligodendroglial gliomas. (1). *Brain tumor pathology*, *28*(3), 177–183. doi:10.1007/s10014-011-0029-1

Pajouhesh, H. & Lenz, G. R. 2005. Medicinal chemical properties of successful central nervous system drugs. *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics*, *2*(4), 541–553.

Pardridge, W. M. 1998. CNS drug design based on principles of blood-brain barrier transport. *Journal of neurochemistry*, *70*(5), 1781–1792.

Piccirillo, S. G. M., Spiteri, I., Sottoriva, A., Touloumis, A., Ber, S., Price, S. J., Heywood, R. et al. 2015. Contributions to Drug Resistance in Glioblastoma Derived from Malignant Cells in the Sub- Ependymal Zone *2004*(10), 194–203. doi:10.1158/0008-5472.CAN-13-3131

Pilyugin, M., Demmers, J., Verrijzer, C. P., Karch, F. & Moshkin, Y. M. 2009. Phosphorylation-mediated control of histone chaperone ASF1 levels by Tousled-like kinases. *PloS one*, *4*(12), e8328. doi:10.1371/journal.pone.0008328

Pruitt, K. D., Tatusova, T., Brown, G. R. & Maglott, D. R. 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic acids research*, *40*(Database issue), D130–5. doi:10.1093/nar/gkr1079

Rae, A. P., Beattie, J. M., Lawrie, T. D. & Hutton, I. 1985. Comparative clinical efficacy of bepridil, propranolol and placebo in patients with chronic stable angina. *British journal of clinical pharmacology*, *19*(3), 343–352.

Romano, J. D. & Kolch, W. 2011. The secret life of kinases functions beyond catalysis. *Cell Communication Signalling*

Ronald, S., Awate, S., Rath, A., Carroll, J., Galiano, F., Kleiner-hancock, H., Mathis, J. M. et al. 2013. Phenothiazine Inhibitors of TLKs Affect Double-Strand Break Repair and DNA Damage Response Recovery and Potentiate Tumor Killing with Radiomimetic Therapy. doi:10.1177/1947601913479020

Ronald, S., Sunavala-Dossabhoy, G., Adams, L., Williams, B. & De Benedetti, A. 2011. The expression of Tousled kinases in CaP cell lines and its relation to radiation response and DSB repair. *The Prostate*, *71*(13), 1367–1373. doi:10.1002/pros.21358

Ruano, Y., Mollejo, M., Camacho, F. I., de Lope, A., Fiaño, C., Ribalta, T., Martínez, P. et al. 2008. Identification of survival-related genes of the phosphatidylinositol 3'-kinase signaling pathway in glioblastoma multiforme. *Cancer*, *112*(7), 1575–1584. doi:10.1002/cncr.23338

Sangar, V., Funk, C. C., Kusebauch, U., Campbell, D. S., Moritz, R. L. & Price, N. D. 2014. Quantitative proteomic analysis reveals effects of epidermal growth factor receptor (EGFR) on invasion-promoting proteins secreted by glioblastoma cells. *Molecular & cellular proteomics : MCP*, *13*(10), 2618–2631. doi:10.1074/mcp.M114.040428

Silljé, H. H., Takahashi, K., Tanaka, K., Van Houwe, G. & Nigg, E. A. 1999. Mammalian homologues of the plant Tousled gene code for cell-cycle-regulated kinases with maximal activities linked to ongoing DNA replication. *The EMBO journal*, *18*(20), 5691–5702. doi:10.1093/emboj/18.20.5691

Sivaprakasam, P., Tosso, P. N. & Doerksen, R. J. 2009. Structure-activity relationship and comparative docking studies for cycloguanil analogs as PfDHFR-TS inhibitors. *Journal of chemical information and modeling*, *49*(7), 1787–1796. doi:10.1021/ci9000663

Smith, J. S., Tachibana, I., Passe, S. M., Huntley, B. K., Borell, T. J., Iturria, N., O'Fallon, J.

R. et al. 2001. PTEN mutation, EGFR amplification, and outcome in patients with anaplastic astrocytoma and glioblastoma multiforme. *Journal of the National Cancer Institute*, *93*(16), 1246–1256.

Sumathi, K., Ananthalakshmi, P., Roshan, Mnam. & Sekar, K. 2006. 3dSS: 3D structural superposition. *Nucleic Acids Research*, *34*, W128–W132.

Sunavala-Dossabhoy, G. & De Benedetti, A. 2009. Tousled homolog, TLK1, binds and phosphorylates Rad9; TLK1 acts as a molecular chaperone in DNA repair. *DNA repair*, *8*(1), 87–102. doi:10.1016/j.dnarep.2008.09.005

Sundarapandian, T., Shalini, J., Sugunadevi, S. & Woo, L. K. 2010. Journal of Molecular Graphics and Modelling Docking-enabled pharmacophore model for histone deacetylase 8 inhibitors and its application in anti-cancer drug discovery. *Journal of Molecular Graphics and Modelling*, *29*(3), 382–395. doi:10.1016/j.jmgm.2010.07.007

Takahata, S., Yu, Y. & Stillman, D. J. 2009. The E2F functional analogue SBF recruits the Rpd3(L) HDAC, via Whi5 and Stb1, and the FACT chromatin reorganizer, to yeast G1 cyclin promoters. *The EMBO journal*, *28*(21), 3378–3389. doi:10.1038/emboj.2009.270

Takayama, Y., Kokuryo, T., Yokoyama, Y., Ito, S., Nagino, M., Hamaguchi, M. & Senga, T. 2010. Silencing of Tousled-like kinase 1 sensitizes cholangiocarcinoma cells to cisplatin-induced apoptosis. *Cancer letters*, *296*(1), 27–34. doi:10.1016/j.canlet.2010.03.011

Tanaka, S., Akaike, T., Wu, J., Fang, J., Sawa, T., Ogawa, M., Beppu, T. et al. 2003. Modulation of tumor-selective vascular blood flow and extravasation by the stable prostaglandin 12 analogue beraprost sodium. *Journal of drug targeting*, *11*(1), 45–52.

Tosatto, S. C. E. 2005. The victor/FRST function for model quality estimation. *Journal of computational biology : a journal of computational molecular cell biology*, *12*(10), 1316–1327. doi:10.1089/cmb.2005.12.1316

Tung, C.-H., Huang, J.-W. & Yang, J.-M. 2007. Kappa-alpha plot derived structural alphabet and BLOSUM-like substitution matrix for rapid search of protein structure database. *Genome biology*, *8*(3), R31. doi:10.1186/gb-2007-8-3-r31

Varguhese, J. F. & Li, Y. 2011. Molecular dynamics and docking studies on cardiac troponin C. *Journal of biomolecular structure & dynamics*, *29*(1), 123–135.

von Grotthuss, M., Pas, J. & Rychlewski, L. 2003. Ligand-Info, searching for similar small compounds using index profiles. *Bioinformatics (Oxford, England)*, *19*(8), 1041–1042.

Wang, J., Zhang, J., Sun, J., Han, J., Xi, Y., Wu, G., Duan, K. X. et al. 2011. Prostacyclin administration as a beneficial supplement to the conventional cancer chemotherapy. *Medical hypotheses*, *76*(5), 695–696.

Warner, W. A., Sanchez, R., Dawoodian, A., Li, E. & Momand, J. 2012. Identification of FDA-approved drugs that computationally bind to MDM2. *Chemical biology & drug design*, *80*(4), 631–637.

Wiederstein, M. & Sippl, M. J. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, *35*(Web Server issue), W407–10. doi:10.1093/nar/gkm290

Wolfort, R., de Benedetti, A., Nuthalapaty, S., Yu, H., Chu, Q. D. & Li, B. D. 2006. Up-

regulation of TLK1B by eIF4E overexpression predicts cancer recurrence in irradiated patients with breast cancer. *Surgery*, *140*(2), 161–169.

Yong, S. L. & Wurster, R. D. 1996. Bepridil enhances in vitro antitumor activity of antiestrogens in human brain tumor cells. *Cancer Lett*, *110 SRC* - , 243–248.

Yoshinori Minami et al. 2012. Activating the prostaglandin I2-IP signaling suppresses metastasis in lung cancer. *Cancer Research*, *4379*(In Proceedings of the 103rd Annual Meeting of the American Association for Cancer Research).

Zhang, Y. 2009. I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins: Structure, Function, and Bioinformatics*, *77*, 100–113.