Indian Journal of Chemistry
Vol. 64, January 2025, pp. 22-37
DOI: 10.56042/ijc.v64i1.13952

Ligand based pharmacophoric discovery of new CLK1 inhibitors

Vidya Jyothi Alli[a,b#], Swapnil Anil Sule[a#], Darna Mounika[a,b], Sai Satya Sri Pulla[a], Pawan Yadav[a] & Surender Singh Jadav*[a,b]

[a] Department of Natural Products and Medicinal Chemistry, CSIR-Indian Institute of Chemical Technology Tarnaka, Uppal Road, Hydrabad 500 037, India

[b] Academy of Scientific and Innovative Research (AcSIR), CSIR-HRDC Campus, Postal Staff College Area Sector 19, Kamla Nehru Nagar, Ghaziabad 201 002, Uttar Pradesh, India

E-mail: surenderjs@iict.res.in

Received 25 September 2024; accepted(revised) 27 December 2024

CLK1 has been recognized as an optimistic target against an array of diseases, including cancer due to their pre-mRNA splicing function. In this study, ligand-based pharmacophore approach has been employed to identify new CLK1 inhibitors from kinase and FDA approved drug libraries. Two ligands, **K6** and **D1** have been extracted, which established consistent interactions with hinge residues while maintaining a web of interactions in DFG region. Additionally, their stable protein dynamics comparable with reference CLK1 inhibitor (**T24**) and Apo ratify them as aspiring CLK1 inhibitors. Further, **D1** has been identified as a CLK1 binder and **K6** as a potential kinase inhibitor by Swiss Target Prediction server. Moreover, their adequate pharmacokinetic and toxicity profiles make them worth future investigations.

Keywords: CLK kinase, Dynamics, *In silico* ADMET study, *In silico* target prediction, Pharmacophore, Rosiglitazone

CDC2-like kinases (CLK1-4) are part of CMGC group, which are also called "LAMMER" kinases due to the conservative 'EHLAMMERILG' motif that define CLKs'substrate specificity[1,2]. CLKs recognize and interact with RNA recognition motifs (RRMs) of SR (Serine/Arginine rich factor) proteins (SRSF1-12) and phosphorylate RS domain (Arginine-Serine repeats) regulating pre-mRNA splicing and pre-mRNA alternative splicing[2]. CLKs also phosphorylate various downstream proteins to promote cytokinesis (Aurora B), modulate c-Myc (USP13) and PTP-1B activity (PTP-1B), dephosphorylate AKT kinase (B56β), and suppress fatty acid metabolism (PGC-1α), and cardiac hypertrophy (NEXN) in addition to activating Wnt and hippo signalling pathways[3]. Dysregulated alternative splicing of tau (exon 10) mediated by CLK1, 2 contribute to neurodegenerative diseases like sporadic Alzheimer's disease (AD) and skipping exon 31 of mutated dystrophin by CLK inhibition is helpful in treating Duchenne muscular dystrophy (DMD)[3,4]. CLK family potentially act as anti-viral target against influenza A virus[3,5], adenovirus[2,6], and HIV-1[2,3,7]. Especially, CLK1 is a promising target against CHIKV[8] and human cytomegalovirus[9]. Furthermore, incorrect alternate splicing of proteins associated with cell cycle, angiogenesis, and apoptosis regulated by abnormal CLK activity leads to a variety of cancers. CLK1, 3 are overexpressed in hypoxic cancers, including prostate cancer, breast cancer, and colorectal cancer[10]. Moreover, in ovarian and pancreatic cancers the overexpression of CLK1 is associated with cell metastasis, increased proliferation, and poor prognosis[11]. Further, in prostate cancer, inhibition of CLK1 with TG003 (**1**) induced caspase 9b expression, which was initially sparse in hypoxic conditions[10]. CLK2 on the other hand, has been linked to breast, larynx, lung cancers, and glioblastomas[11]. CLK3 expression is hypoxia-prompted and is often dysregulated in hepatocellular carcinoma and gastrointestinal cancers[11]. CLK4 is independent of hypoxia but is dysregulated in breast cancer, lung, colorectal, and larynx cancers. Thus, individual inhibition or pan-CLK inhibition is useful in treating a diverse spectrum of cancers[11].

All CLKs possess at least 50% structural and sequential similarities, especially CLK1 and CLK4 (78.4%)[1]. The ATP pocket of CLK1, 2, 4 displayed high structural resemblance with subtle differences in charge distribution, while CLK3 has a wider pocket as the Valine (DFG-1 residue) of other CLKs is replaced

# Authors contributed equally.

with the shorter Alanine residue. This replacement offers selectivity against CLK3[3]. Due to the high therapeutic potential of CLKs, the focus has been shifted towards the discovery and design of CLK inhibitors. Many CLK inhibitors with diverse scaffolds were identified (Fig. 1); some were derived from natural products (**1-6**) and the other synthetic scaffolds were designed based on natural derivatives (**7-26**). These scaffolds possess striking pharmacophoric similarities; hence, in the current study a ligand-based pharmacophoric model was used to identify novel CLK inhibitors by precisely targeting CLK1. Further, their respective pharmacophoric features were analyzed to form a templet for future design.

Materials and Methods

Materials

A GPU-based workstation of 128 GB RAM, with Quadro P5000, P600 graphics and an 88-core processor, containing a 5TB hard disk, CPU E5-2699v4@2.20GHz of intel® xeon®, was used for the present *in silico* study whose protocol is portrayed in Fig. 2.

Data set generation

In the current study, a data set of 69 ligands with pyrrolo-pyridine and imidazo [1,2-a] pyridine scaffolds were taken[12–14]. This data set is further divided into a test set (54 ligands) and a training set (15 ligands) for pharmacophore generation and

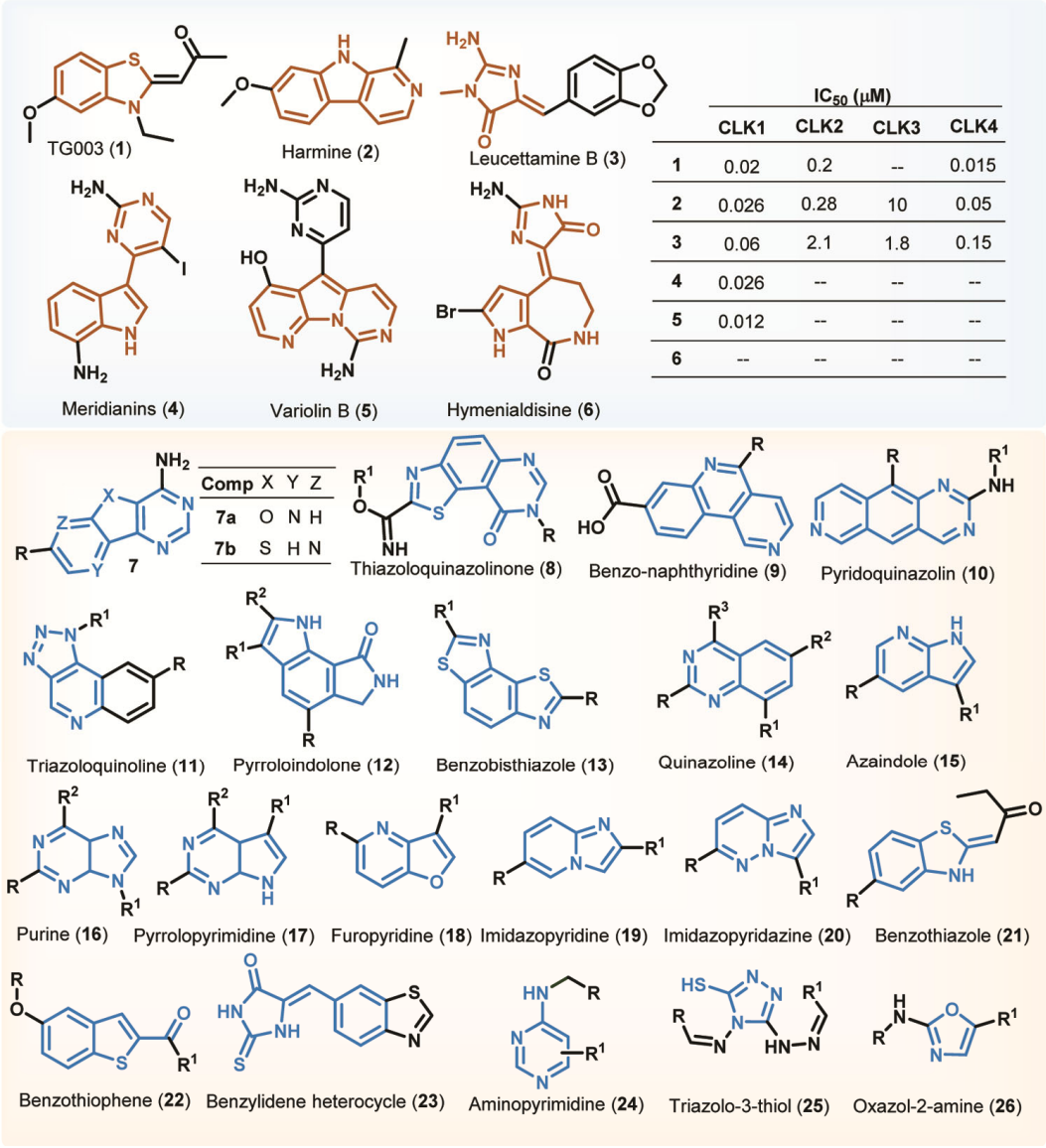


Fig. 1 — Natural (**1-6**) and Synthetic (**7-26**) CLK inhibiting scaffolds

validation and is included in Tables S1 and S2. The ligands of both sets were either active or inactive in relation to their threshold activity of $pIC_{50} > 7$ and $pIC_{50} < 6$, respectively. The pIC_{50} values were derived from their respective nanomolar IC_{50} values (1 nM = 10^9 M) using the following formula [1].

$$pIC_{50}(nM) = 9 - log_{10}(IC_{50}) \quad \ldots (1)$$

The structures of all the ligands were drawn using "2D sketcher" of Schrödinger Suite Release 2024-1 then they were subjected to "LigPrep" using default parameters to generate possible ionized and minimized isomers. The prepared ligands were used for later studies[15].

Pharmacophore hypothesis generation and validation

Ligand-based pharmacophore was developed using "PHASE" module of Schrödinger Suite by taking 'multiple ligand alignment' protocol to identify common pharmacophoric features, including hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), aromatic ring (R), positive charged group (P), and negative charged group (N). Five models with five pharmacophoric sites were generated and subjected to validation (Fig. S1)[16]. The enrichment study was done by employing the ligand dataset (both training and test sets) and decoy set extracted from Directory of Useful Decoys (DUD.E, https://dude.docking.org) server and actives were set to own at least three out of five pharmacophoric features. The output was then analyzed in "enrichment viewer" panel and the best hypothesis (**ADRRR_1**) with excellent % screen plot and enrichment parameters was designated for further studies[17]. The enrichment parameters include, the enrichment factor (EF), Phase Hypo Score, Receiver operator characteristic area under the curve (ROC), Boltzmann-enhanced discrimination ROC (BEDROC160.9), and Area under the accumulation curve (AUAC). Table 1 discloses enrichment parameters of each hypothesis and their respective % screen plots are enclosed in Fig. S2.

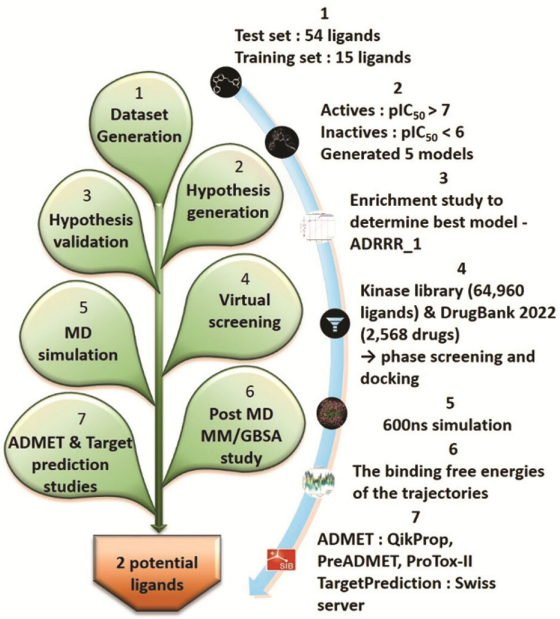


Fig. 2 — Overview of the study procedure

The EF discloses the fraction of known actives recovered for a fraction of database and it is calculated using the below formula [2][18]. Phase Hypo Score indicates the rank of hypothesis with respective to its performance. The BEDROC160.9 predicts and classifies the actives before decoys and a value of > 0.5 is considered good. ROC expresses the degree of discrimination of actives and inactives, calculated by eq. [3] and ranges from 0 (bad performance) to 1 (good performance)[19]. AUAC is obtained by calculating area under the curve using eq. [4]; it is a probability of ranking an active defined by rank-order before any random ligand from a uniform distribution. Further, AUAC lies between 0 (bad value) to 1 (good value) similar to ROC[19].

$$EF = \frac{H_a \times D}{H_t \times A} \quad \ldots (2)$$

Table 1 — Parameters of obtained pharmacophore models

S. No	Hypothesis	[a]EF1%	Phase Hypo Score	[b]ROC	[c]BEDROC160.9	[d]AUAC
1	DDRRR_1	15.49	0.89	0.82	0.94	0.80
2	***ADRRR_1***		***1.23***	***0.94***	***0.99***	***0.91***
3	ADRRR_2		1.01	0.90	0.97	0.88
4	ADRRR_3		1.22	0.89	0.97	0.87
5	AADRR_1		1.00	0.90	0.98	0.87

[a]Enrichment factor; [b]Receiver operator characteristic area under the curve; [c]Boltzmann-enhanced discrimination receiver operator characteristic area under the curve; and [d]Area under the accumulation curve.

$$ROC = \int_0^1 F_a(x) f_i(x) dx \quad (3)$$

$$AUAC = \int_0^1 F_a(x) dx \quad (4)$$

Here, H_a is number of active hits, H_t is number of hits, D is total number of ligands in the database, and A is the total number of actives in the database. $F_a(x)$ is the probability of an active ranking before the relative rank x.$f_i(x)$ is the of inactives from order list.

Virtual screening

The best hypothesis **ADRRR_1** was employed for virtual screening of kinase library of Enamine database (64,960 compounds, https://enamine.net/compound-libraries/targeted-libraries/kinase-library), and FDA approved library from DrugBank 2022 (2,568 drugs, https://go.drugbank.com) by matching four out of five pharmacophoric features. The fitness score (*FS*) of all ligands was generated that typically ranges from –1 to 3 (higher value indicate better matching) and it is a linear sum of site score (*SS*), vector alignment score (V_aS), and volume score (*VS*) as shown in eq. [5][20].

$$FS = SS + V_aS + VS \quad \ldots (5)$$

Simultaneously, CLK1 (PDB ID: **5X8I**, http://www.pdb.org) was preprocessed, curated, optimized, and minimized using "Protein Preparation" wizard from Schrödinger Suite using default parameters[21]. The prepared complex was then used to generate a cubic receptor grid defined with co-crystal ligand with internal and external dimensions of 25 Å. The generated grid was employed to dock ligands from each library with $FS \geq 1.0$ using "GLIDE" software in flexible Extra Precision (XP) scoring function[22].

Prime MM/GBSA study

The binding free energies (delta G_{bind}or ΔG_{bind}) of the ligands were computed by Molecular Mechanic-Generalized Born Surface Area (MM/GBSA) study, keeping default parameters, including the VSGB solvation method in the "PRIME" module[23]. The binding free energies (ΔG_{bind}) are obtained from formula [6]. The two ligands (one from kinase library and one from DrugBank) with best docking scores and binding patterns were handpicked to generate possible poses using induced-fit docking (IFD) by standard protocol[24]. Then the best pose of each ligand with good interacting network was picked for Molecular Dynamic (MD) studies.

$$\Delta G_{bind} = G_C - (G_P + G_L) \quad \ldots (6)$$

Here, G_C, G_P, and G_L are the free energies of the complex, protein, and ligand, respectively.

Molecular Dynamic Simulation and post dynamic MM/GBSA study

Using "Desmond 2020", MD simulations were carried out for each protein-ligand complex to better predict their respective protein stability and interaction pattern[25]. The simulation was carried out using SPC cubic water solvent model with buffer dimensions of 10 Å and the system was neutralized using appropriate number of sodium (Na^+) and chloride (Cl^-) ions. Then the system was minimized for 5ns (1 ns = 10^9 s) using OPLS_3e force field using default method of unrestrained minimization (combination of steepest descent and conjugate gradient methods)[26,27]. Moreover, the MD simulation was performed by taking 5ns of relaxation time, temperature of 300 K was maintained using Nose-Hoover method, 1.01325 Bar (1 Bar = 10^5 Pa) pressure was retained using NPT ensemble and other default parameters (Time step = 2.0 fs and cut-off radius of van der Waals = 9.0 Å). Initially, 100ns simulation was done to check the stability and equilibrate the system, then 600ns simulation was performed while recording at every 600ps. All the trajectories were studied and Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), and Radius of Gyration (RoG) graphs were plotted and analyzed for structural convergence of Cα-backbone atoms to equilibrium using "Simulation Event Analysis" and "Simulation Interaction Diagram" panels. Furthermore, the binding free energies of the trajectories were computed by MM/GBSA study using "thermal_mmgbsa.py" script[23].

Principal Component Analysis (PCA) and Free energy landscape (FEL)

The maestro trajectory files of Apo and complexes were converted into DCD files by VMD 1.9.4a53 and were subjected to Principal Component Analysis (PCA) by Bio3D to examine their variance[28,29]. Further, residue cross-correlation was also determined using this tool. Then the Free energy landscape of the trajectory was plotted using the free-energy-landscape 1.0.3 package provided by Python Package Index (PyPI) that calculates free energy from collective variables, in our case, RMSD and Radius of Gyration[30].

ADMET study and Target Prediction

The ADMET properties of selected two ligands were predicted using the "QikProp" module of Schrödinger suit with default parameters and also using PreADMETserver (https://preadmet.webservice.bmdrc.org)[31,32]. Further, through ProTox-II online free server (https://tox-new.charite.de/protox_II), the toxicity parameters were estimated[33]. Moreover, Swiss Target Prediction server (http://www.swisstargetprediction.ch) was used to speculate possible targets for each ligands[34].

Results and Discussion

Pharmacophore hypothesis generation and validation

Five ligand-based pharmacophores were generated and validated based on the pIC_{50} values of ligand datasets (Fig. S1). All the hypothesises not only obtained same EF% of 15.49 (indicating excellent early recognition of actives) but the other parameters were also excellent with nearer values to 1 (Table 1). Additionally, the % screen curves of all models lie above the reference line, indicating accurate discrimination of actives from inactives (Fig. S2). Among all the models, **ADRRR_1** displayed highest ROC, BEDROC, and AUAC values of 0.94, 0.99, and 0.91 thus nominated and Fig. 3a & b shows the corresponding distance and angle measurements between each site.

Virtual screening approach

By employing **ADRRR_1**, kinase library (64,960 compounds) and DrugBank 2022 (2,568 drugs) were screened by matching 4 out of 5 pharmacophoric sites. Instead of having only excellent pharmacophore mapping, ligands should also possess good receptor affinity, therefore, from each library, ligands with FS > 1.0 were subjected to ligand docking using flexible docking protocol against CLK1 to identify better ligands from wider chemical space and top 10 hits from each library were tabulated in Table 2 and 3. Two ligands, one from kinase library (**K6**), and another from DrugBank (Rosiglitazone, **D1**) were

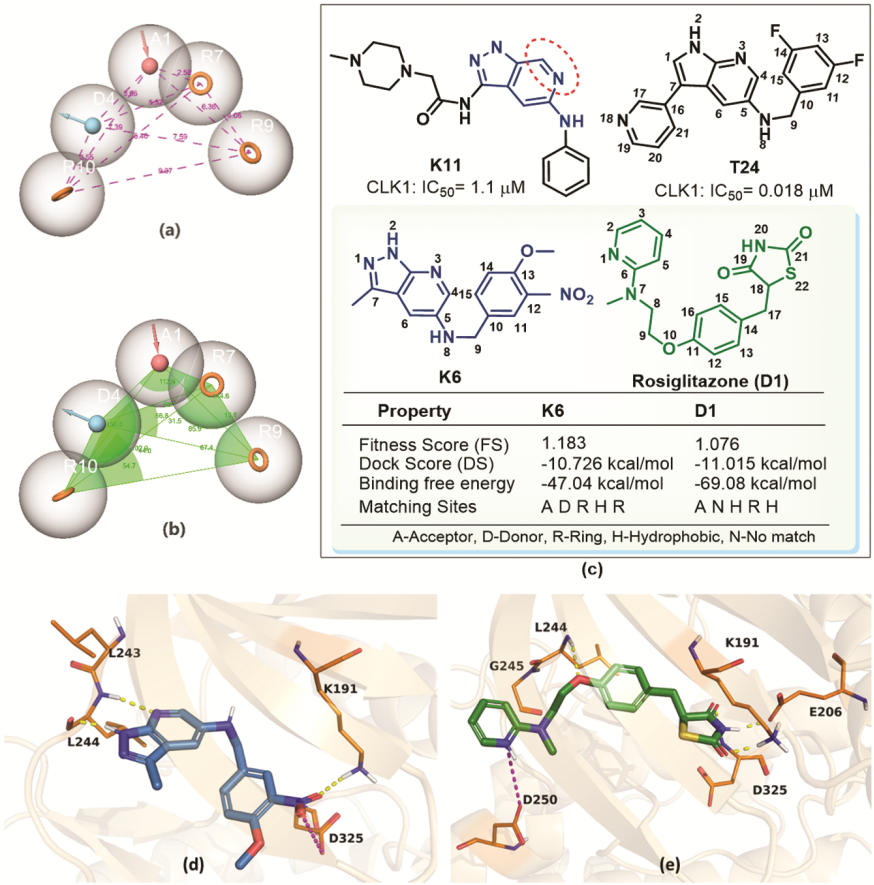
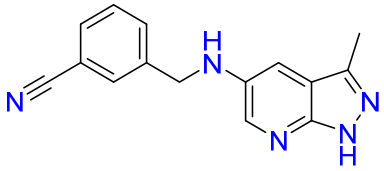
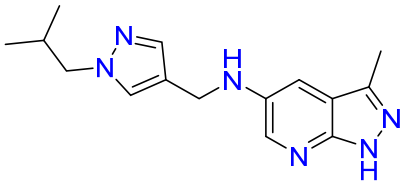
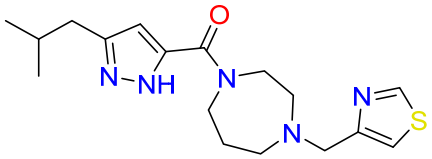
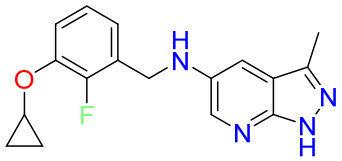
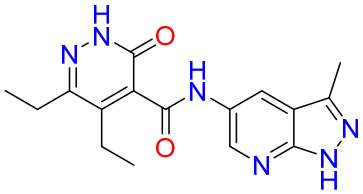
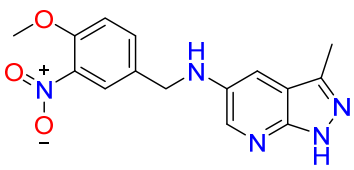
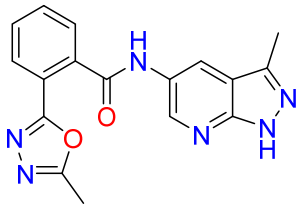
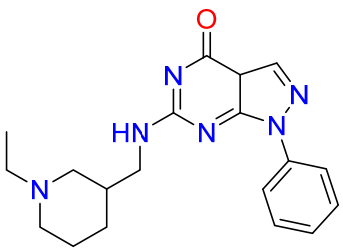


Fig. 3 — a & b) Distance and angle measurements of ADRRR_1; c) The obtained hits (**K6** and **D1**) of virtual screening (Pharmacophoric screening and docking), similar CLK1 inhibitor (**K11**), and the reference ligand (**T24**) and the summary of selected criteria; d & e) Interaction pattern of **K6** (Blue colour) and **D1** (Green colour) with CLK1, respectively

Table 2 — Top 10 hits obtained from Kinase library

S. No	Ligand structure and name	Docking Score (kcal/mol)	Fitness score
K1		–11.095	1.260
K2		–11.093	1.157
K3		–10.837	1.140
K4		–10.808	1.194
K5		–10.736	1.162
K6		–10.726	1.183
K7		–10.472	1.087
K8		–10.264	1.137

(*Contd.*)

Table 2 — Top 10 hits obtained from Kinase library (*Contd.*)

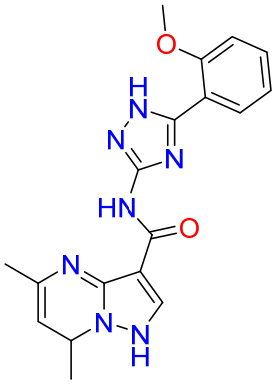
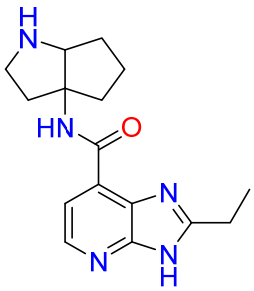
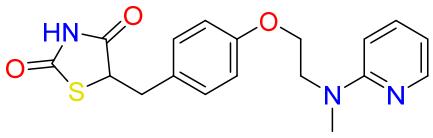
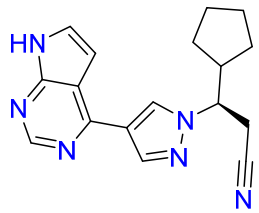
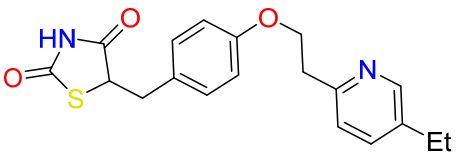
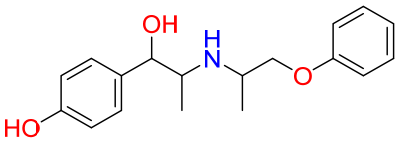
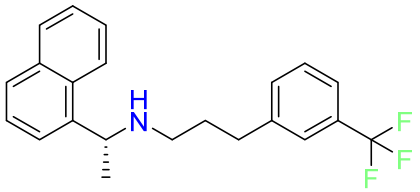
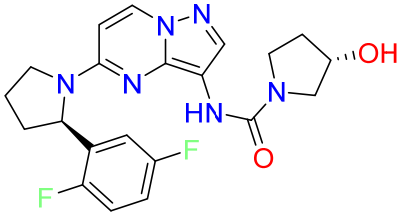
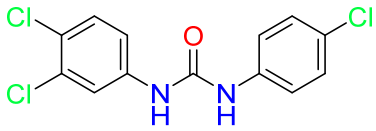
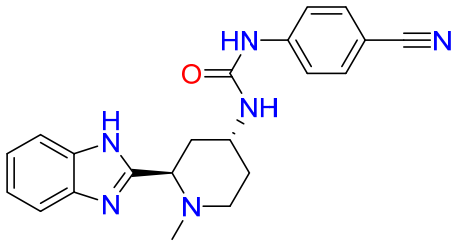
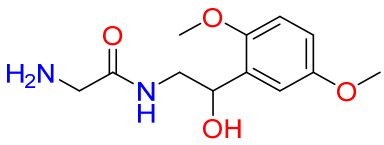
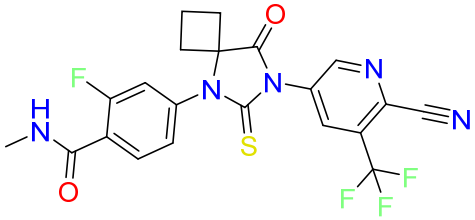
S. No	Ligand structure and name	Docking Score (kcal/mol)	Fitness score
K9		–10.195	1.001
K10		–10.144	1.063

Table 3 — Top 10 hits obtained from Drug Bank

S. No	Drug structure and name	Docking Score (kcal/mol)	Fitness score	Biological Activity
D1	 Rosiglitazone	–11.015	1.076	PPARγ agonist; Antidiabetic
D2	 Ruxolitinib	–10.511	1.032	JAK 1, 2 inhibitor; Anticancer
D3	 Pioglitazone	–9.845	1.072	PPARγ agonist; Antidiabetic
D4	 Isoxsuprine	–9.41	1.05	β-adrenergic agonist; Vasodilator

(*Contd.*)

Table 3 — Top 10 hits obtained from Drug Bank (*Contd.*)

S. No	Drug structure and name	Docking Score (kcal/mol)	Fitness score	Biological Activity
D5	 Cinacalcet	–8.941	1.042	Calcium-sensing Receptor Agonist
D6	 Larotrectinib	–8.215	1.004	Tropomyosin receptor kinase (Trk) inhibitor; Anticancer
D7	 Triclocarban	–8.192	1.004	Soluble epoxide hydrolase (sEH); Antiseptic
D8	 Glasdegib	–8.044	1.092	Hedgehog signalling inhibitor (SMO receptor antagonist); Anticancer
D9	 Midodrine	–7.774	1.039	α-adrenergic agonist; Vasoconstrictor
D10	 Apalutamide	–7.769	1.054	Androgen receptor (AR) antagonist; Anticancer

selected whose FS, docking scores, binding free energy, and matching pharmacophoric sites were included in Fig. 3c. The IFD protocol was then employed to generate possible poses and the interaction pattern of each ligand-protein complex was analysed and summarized in Table 4. Simultaneously, the Scifinder (https://scifinder-n.cas.org) and PubChem (https://pubchem.ncbi.nlm.nih.gov) search on each ligand was done for relevant literature.

Table 4 — Summary of interactions of selected ligands

S.No	Ligand	Interacting Residue	Type of interaction	Bond distance (Å)
1	K6	[B]L244	[D] H-bond	2.33
			[A] H-bond	1.96
		[S]K191	[D] H-bond	2.28
		D325	Salt bridge	4.08
2	D1	L244	[D] H-bond	2.79
		[S]D250	Salt bridge	2.87
		[S]K191	[D] H-bond	2.04
		[S]E206	[A] H-bond	2.04
		[B]D225	[D] H-bond	2.21

[B] Backbone interaction; [S] Sidechain interaction; [A] Acceptor; [D] Donor

K6 was initially selected because of its nitro group making excellent interacting network with the back pocket near DFG motif; similar to the nitro group containing **DB18** and its analogues characterized by favourable binding and high activity (CLK1: IC_{50} = 11 nM)[35]. No relevant literature was found about **K6** but an almost similar scaffold (**K11**) inhibiting CLK1 with an IC_{50} of 1.1 μM(1 μM = 10^6 M) was found that confirms the ability of **K6** to inhibit CLK1[36]. The main scaffold (pyrazolopyridine) of **K6** showed hydrogen bond (H-bond) interactions with hinge backbone L244 and L243. While the nitro group associate with K191 *via* H-bond and D325 *via* salt bridge. These interactions at DFG pocket might enhance its inhibitory activity and are presented in Fig. 3d.

Rosiglitazone (**D1**) is a Peroxisome Proliferator Activated Receptor gamma (PPARγ) agonist that enhance insulin sensitivity, regulate fatty acid metabolism, and reduce gluconeogenesis; thus, used to treat type 2 diabetes (T2D). PPARγ-Rosiglitazone complex also promote neuroprotection *via* increasedB-cell lymphoma 2 (BCL-2) protein levels[37]. Superposition of PPARγ and CLK3 kinase identified remarkable similarities in the binding pocket topology and thus, PPARγ can be considered a potential distant off-target of CLK3 and *vice versa*[38]. In the current study, Rosiglitazone (**D1**), a PPARγ agonist was procured as a CLK1 binder with a dock score of –11.015 kcal/mol and the thiazolidinedione (glitazone) moiety was associated with the DFG pocket of CLK1 as seen in Fig. 3e. The *'10O'* acted as a hydrogen bond acceptor towards the backbone hinge L244 residue and the *'1N'* formed a salt bridge with D250. The '*19CO'* and *'21CO'* carbonyl oxygens accepted backbone hydrogens from D325 and K191, respectively; while the *'20N'* acted as a hydrogen bond donor to E206; thus, showing DFG-in conformation (Fig. 3e). Interestingly, another PPARγ agonist, Pioglitazone, was also a top hit with a dock score of –9.845 kcal/mol with a similar binding pattern as **D1**. It was not further studied, but the identified hit confirms the off-target correlation between CLKs and PPARγ.

Table 5 — Summary of Protein-Ligand contacts during MD simulation

Ligand	Interacting residue	Type of interaction	Threshold
T24	L244	H-bond	98% (*2N*)
	E242	H-bond	99% (*3N*)
	K191	Water bridge	21% (*18N*)
K6	L244	H-bond	75% (*1N*)
	E242	H-bond	75% (*2N*)
		Water bridge	25% (*3N*)
	K191	Salt bridge	27% (nitro)
		π-cation	16% (phenyl)
D1	L244	2H-bonds	32% ($1NH^+$), 89% (*10O*)
	D325	H-bond	79% (*21CO*)
		Water bridge	65% (*20NH*)
	K191	Water bridge	27% (*20NH*)
	E206	2Water bridges	72% (*20NH*), 70% (*21CO*)
	G327	Water bridge	37% (*21CO*)
	F326	Water bridge	44% (*21CO*)

Molecular Dynamics

The selected poses of each ligand were put through a 600ns MD simulation and the interactions were analysed. Both ligands (**K6** and **D1**) exhibited H-bonding with backbone L244 hinge residue and were seen associating with DFG pocket. Table 5 summarizes the interaction profiles of **K6** and **D1**.

Interaction profile

The ligand **K6** showed remarkable structural similarity with the test and training set ligands, especially with reference ligand **T24**. The ligand **T24** has a pyridine on position-7 that made water bridge with K191 (21%), but **K6** with a –Me group lost this interaction. On the other hand, nitro group on the tail of **K6** made interactions like salt bridge (27%), H-bond, π-cation (16%), and water bridges with K191 that account to its high labyrinth of interactions; The lack of this group incapacitates these interactions in **T24**; thus, significantly reducing the number of interacting residues (Fig. 4a-f). Moreover, the presence of an excess '*1N'* in **K6** enabled a complete novel binding pattern. Unlike **T24**, *'1N'* and *'2N'* of **K6** formed H-bond interactions with L244 (75% of

the simulation time) and E242 (75% of the time), respectively, including an additional water bridge between E242 and *'2N'* for 25% of the time. Nevertheless, the lack of *'1N'* in **T24** enabled *'2N'* and *'3N*' interactions with L244 (98%) and E242 (99%), respectively for a relatively prolonged period. **T24** maintained small but constant interactions with 20 amino acids while **K6** made fragmented interactions with a wider network of around 30 residues (Fig. 4a-f).

Rosiglitazone (**D1**) established an incredible binding pattern, especially with the DFG motif. Unlike classic kinase inhibitors, it lacks heterocyclic hinge binding scaffolds (like pyrimidines) but the aliphatic oxygen accepted H-bond from L244 89% of the trajectory time and the pyrimidine moiety also barely donated H-bond to L244 for 32% of time. Both these atoms compensated for the lack of hinge binding scaffolds and together contributed to prolonged contact throughout the simulation time (Fig. 4g-i). Interestingly, the glitazone moiety showed excellent interactions with the DFG motif. The glitazone *'20NH'* and *'21CO'* each made three water bridges with K191 (27%), D325 (65%), and E206 (72%); and F326 (44%), E206 (70%), and G327 (37%), respectively. Additionally, *'21CO'* also accepted a H-bond from D325 with a threshold of 79% (Fig. 4h). The alliance of H-bond (79%) and the water bridge (65%) make D325 the highest interacting residue, followed by E206 making intense water bridges (70%, 72%) (Fig. 4g-i).

Protein Dynamics

The RMSD

The Apo protein retained the highest stability with an average Cα-RMSD around 0.24 nm and the presence of the ligand seems dismissive, as depicted in Fig. 5a. Furthermore, the Apo protein very slowly progressed towards stabilization and achieved it from 400 ns. Further, all the complexes followed the same trend. Moreover, the **T24** and **K6** protein complexes displayed huge spikes at various intervals. On the other hand, **D1** complex afforded minimal deviations. The Cα-RMSD of overall simulations remained around 0.30 nm indicating fairly stable protein complexes.

The Radius of Gyration (RoG)

Protein and its complexes in Fig. 5b displayed varying radius of gyration during the course of 600 ns. The protein portrayed the lowest RoG owing to its high compactness. Moreover, during the initial 300 ns, the protein RoG remained stable < 2.15 nm, and then a momentary loss of compactness was observed but later reached stability with a RoG > 2.15 nm. Furthermore, the stability shrinkage in presence of the

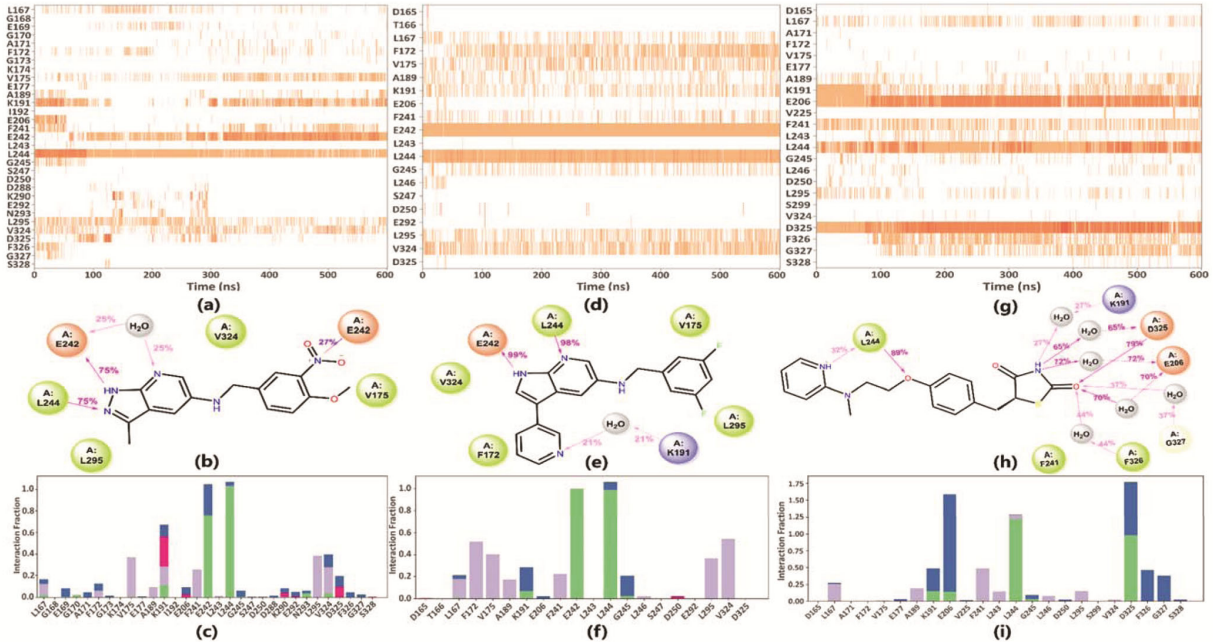


Fig. 4 — a, d, & g) Heat map summarising protein-ligand contacts of each residue during the course of trajectory; b, e, & h) the 2d-representations; and c, f, & i) Histograms summarising interaction fraction of each residue of protein-ligand contacts of **K6**, **T24**, and Rosiglitazone (**D1**) with CLK1

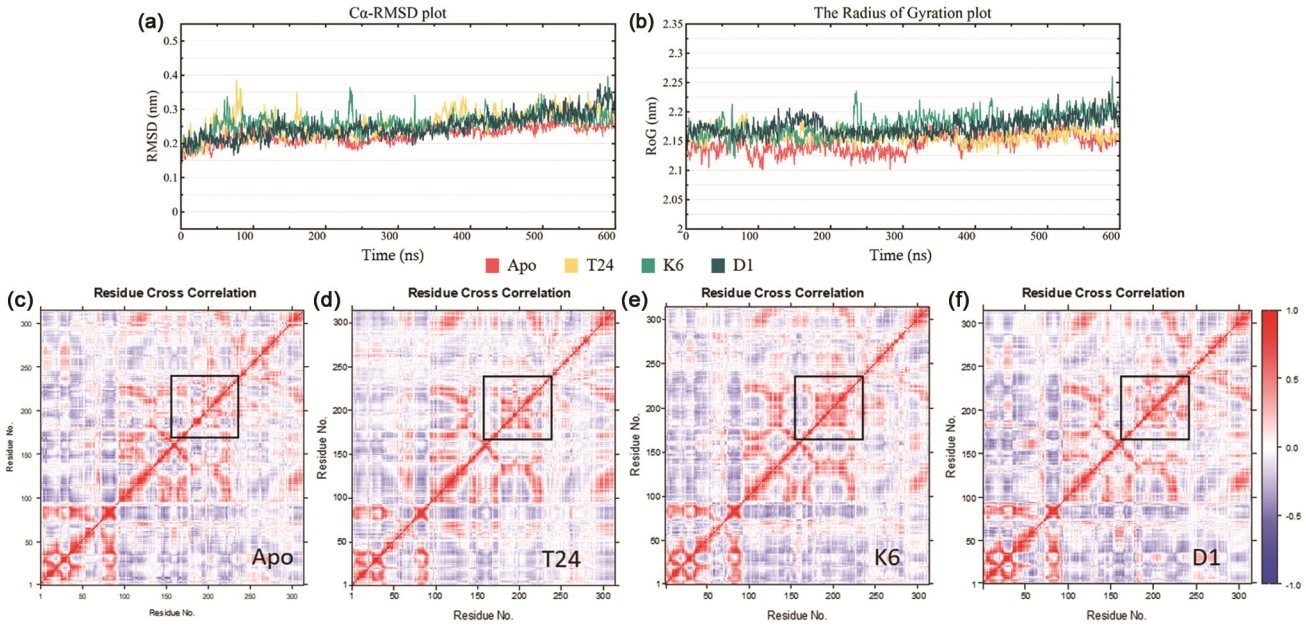


Fig. 5 — a) Cα-RMSD plot showing stable complexes; b) Radius of Gyration plot depicting compromised stability when bound; c-f) DCC plots, highlighted region shows difference between the residue cross-correlation of Activation loop between Apo and bound protein

ligand was also reflected in RoG plot and similar to RMSD, the RoG plot showed a linear progression. Unlike RMSD, the compactness of **T24** complex coincides with Apo, while the **K6** and **D1** exhibited high but hardly stable RoGs.

Dynamic cross-correlation (DCC) plot and RMSF plot

The residue cross-correlation depicted in Fig. 5c-f, showed good correlation between Apo and complexes and the Fig. 6a clearly depicts the Cα-RMSF of all the residues of the protein and its complexes. The RMSF plot illustrates four major peaks denoted by integer 1-4 each representing their corresponding regions portrayed in Fig. 6b. The region-1 (loop of β-hairpin insertion, P307-K310) revealed a high Cα-RMSF values > 0.35 nm, especially K308 displayed excessive fluctuations in Apo protein and all of its complexes (Cα-RMSF ≈ 0.4nm). Additionally, the Apo protein showed spiked RMSF at the region-2 (activation loop, E334-S341) and region-4 (strand connecting helixes αH and αI) with Cα-RMSF of 0.53 nm and 0.35 nm, respectively (Fig. 6a and c). Complexes only retained fluctuations of region-1 but region-2 remained stable (Fig. 6c-f). This phenomenon is also reflected in DCC plot (Fig. 5c-f). In **K6** complex, the region-3 (strand connecting MAPK-like insertion and αH-helix, D420-G426) owned Cα-RMSF ranging from 0.27 to 0.48 nm with highly fluctuating H422 (Cα-RMSF = 0.48 nm) (Fig. 6a & e). The RMSF plots of **T24** and **D1** complexes overlap with Apo but **D1**-complex showed no conformational changes at region-4 (Fig. 6f).

Secondary structure

The secondary structure dynamics of the protein and protein complexes are almost identical as depicted in Fig. S3. The β-hairpin region (D300-I317) varied tremendously in protein and all of its complexes especially the occasional disappearance in **D1**-complex. The MAPK-like insertion also showed highly unstable secondary structures *i.e.*, the both the β-sheets (Y411-H413 and D415-L417) rarely appeared in protein and **T24**-complex. However, after 200 ns they disappeared completely in **K6**-complex and fade in and out in **D1**-complex frequently and the stable α-helix (S424-A432) vanished completely in **K6**-complex after 400 ns.

PCA analysis

The Principal Component Analysis (PCA) was done to analyse the protein's dynamic nature and the collective motions of Cα-atoms of MD trajectories for the first 20 modes. The PCA presents the diagonalized covariance matrix with eigenvectors and eigenvalues and the obtained principal components (PCs) depicts variation in protein conformations. From Fig. 7,it can be observed that PC1-3 are responsible for majority of variance. The first three PCs account for 44.4% for Apo, 30.5% for **T24**-complex, 38.6% for **K6**-

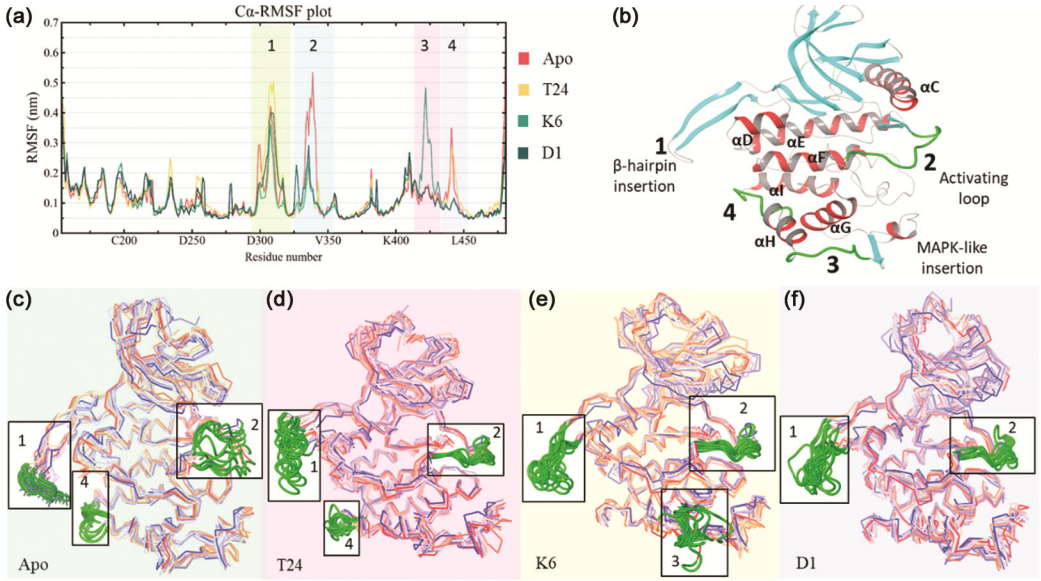


Fig. 6 — a) Cα-RMSF plot; b) Protein highlighting the RMSF peaks in green along with their designated integers; c-f) Superimposition of C traces of eleven frames (every 100th frame of 1003 frames) of Apo, **T24**, **K6**, and **D1** complexes, respectively. The frames are coloured with passage of time from blue to red and the green regions (black box) are peak fluctuating regions showed by RMSF plot

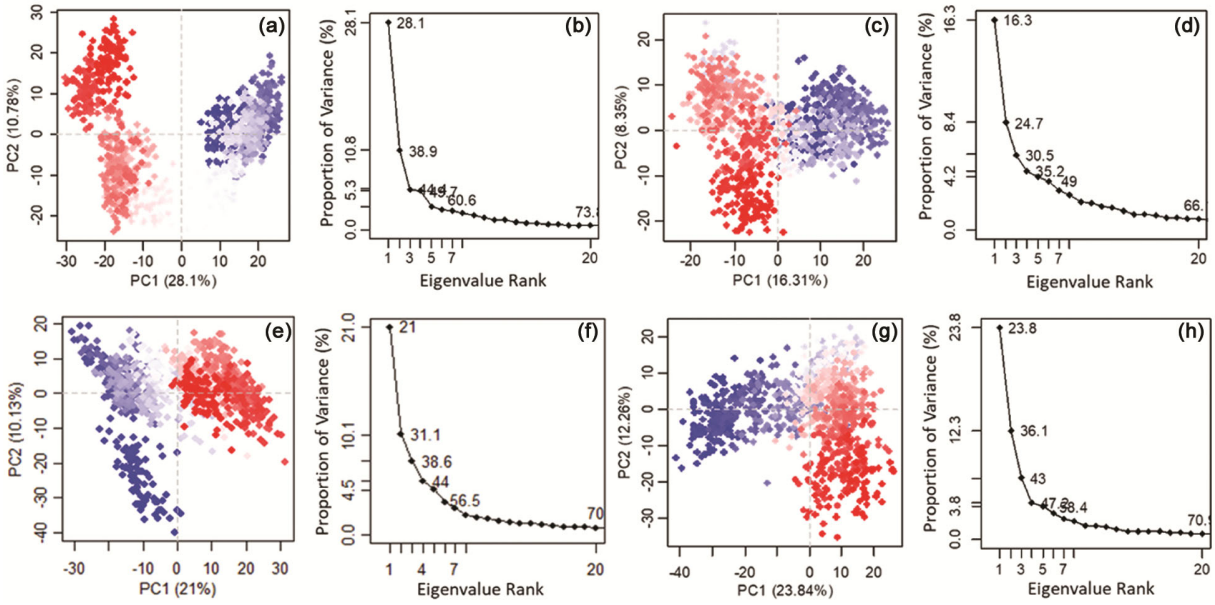


Fig. 7 — Plots of Principal components of the trajectories & proportion of variance of, a & b) Apo; c & d) **T24**; e & f) **K6**; and g & h) **D1** depicting the distribution of clusters of conformations from blue to red with the flow of time

complex, and 43% for **D1**-complex. PC1 is an essential factor that showed 28.1% variance for Apo, 16.31% for **T24**-complex, 21% for **K6**, and 23% for **D1**. This suggests that the binding of ligand slightly reduced the protein conformations. In the plot PC1 *vs* PC2 (Fig. 7), Apo showedstable conformations as the distribution of red (conformations at the end of simulation) and blue (conformations at the begining of simulation) clusters is assimilated at two distinct sides with white (intermediate states i.e, periodiotic oscillations between conformations) clusters in between. The residue-wise loading of the first two principle components (Fig. S4) coincide with RMSF plots and shows the contribution of each residue for each PC.

Post MD-MM/GBSA study

The MM/GBSA of the trajectories in Fig. 8a and 9 was obtained through thermal_mmgbsa.py script. The reference ligand **T24** expressed a stable binding free

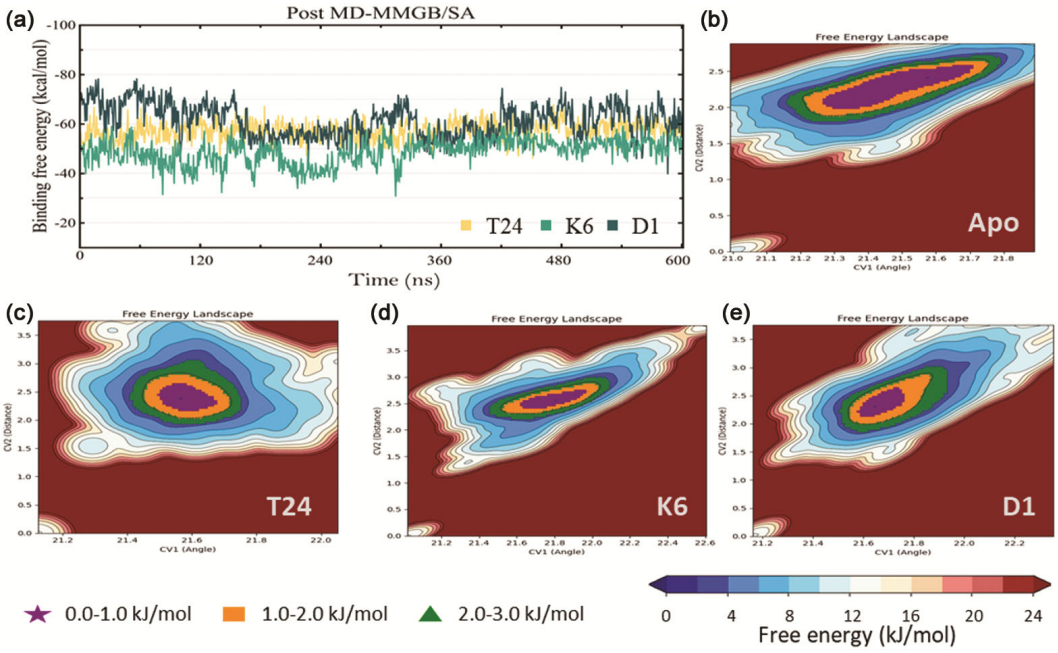


Fig. 8 — a) Post MD-MM/GBSA plot of **T24**, **K6**, and **D1**showing a stable **T24**-complex, gradual stabilisation of **K6**-complex, and destabilised **D1**-complex that later achieved stability; b-e) 2d diagrams of Free Energy Landscapes depicting the existence of only one energy minima

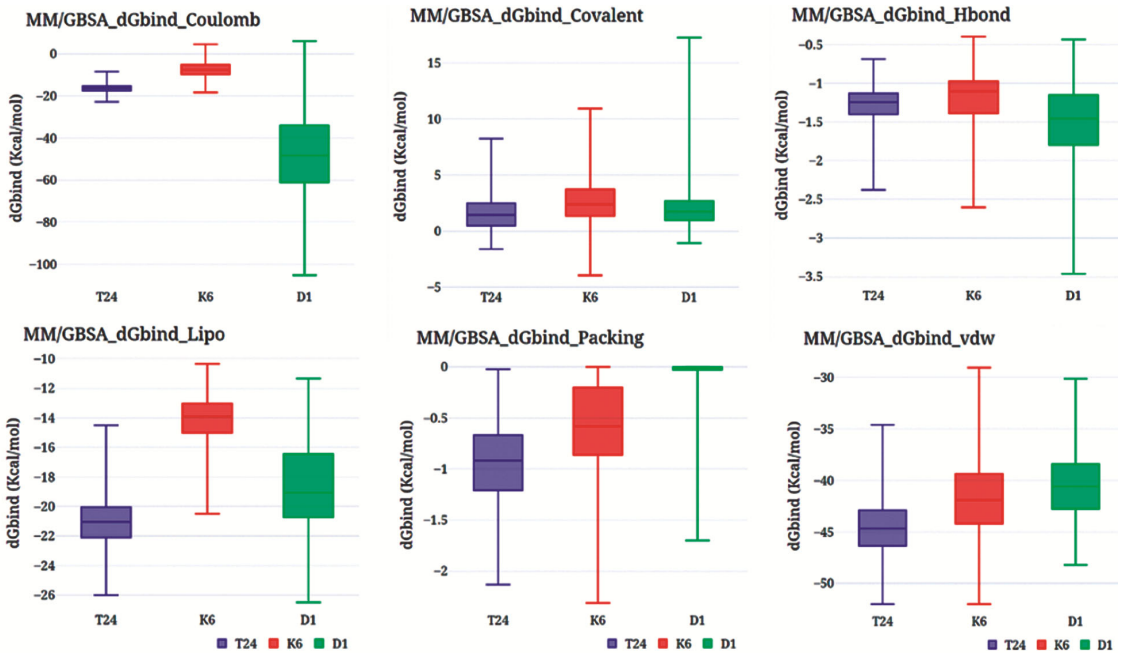


Fig. 9 — Ranges of interaction energies of ligands observed in post MM/GBSA

energy through out the simulation around –60 kcal/mol. However, the ligands **K6** and **D1** exhibited fluctuating energies. The MM/GBSA plot of **K6** clearly displays intial instability (MM/GBSA < –50 kcal/mol) followed by gradual stabulization (> –50 kcal/mol). However, during the entire tragectory, **K6** never reached the standard –60 kcal/mol, indicating a non favourable complex. On the other hand, the **D1** showed excellent stability, far exceeding the standard which may be contributed by coulombic interactions (Fig. 9). Further, at two instances, the **D1** plot coincided with the standard **T24** plot with binding

free energy < –60 kcal/mol. This loss of stability is also reflected in RoG plot of ligand (Fig. 10b). The pre dynamic MM/GBSA also revealed **D1** as a better CLK1 inhibitor compared to **K6**. Eventhough, **K6** afforded more precise pharmacophoric features, it possessed limited affinity as reflected by its MD parameters and MM/GBSA studies (Pre dynamic and post dynamic).

Free energy landscape

Free energy landscapes (FELs) demonstrated in Fig. 8b-e shows the journey of a protein achieving global minima. Furthermore, the FEL of Apo and complexes showed only one basin which indicates that the protein structure has reached lowest energy levels during the duration of the simulation and suggests the existence of only one energy minima. The ligand bound proteins all have a deep but narrow energy wells representing highly stabile conformation compared to the Apo protein with wide basin.

Ligand Dynamics

The RMSD of the selected ligands **D1** and **K6** remained below 0.25 nm but showed high deviations (especially **K6**). The **K6** initially displayed huge fluctuations ranging from 0.05 – 0.25 nm, which may be due to its five rotatable bonds (Fig. S5). Another reason may be the interaction of *'2N'* and *'3N'* with L244, which later shifted to *'1N'* and *'2N'*. On the contrary, **K6** displayed a very stable RoG around 0.45 nm, MolSA (Molecular Surface Area) around 2.95 nm^2, and PSA (Polar Surface Area around 1.95 nm^2). However, the SASA (Solvent Accessible Surface Area) highly fluctuated till the initial 300 ns then stabilised around 0.75 nm^2 (Fig. 10). The unstable SASA may be a result of **K6** coming in and out of the solvent accessible area of the binding pocket, further, the SASA plot also coincides with the RMSD plot representing this motion. Coming to **D1**, the RMSD is very stable at 0.02 nm even though it had seven rotatable bonds (Fig. S5). This may be due to strong and constant interactions with the protein that locked the ligand. The RoG fluctuations may be due to fluctuating pyridine. All the Surface Area properties showed constant values as the ligand is locked inside the pocket (Fig. 10).

ADMET Study

Table S3 and S4 summarised the pharmacokinetic properties of the selected ligands. Both the compounds were inactive against CNS (central nervous system) but both the ligands are permeable through BBB (blood brain barrier) reflected by their logBB and MDCK cell permeability values, especially **D1**. The Caco2 and % Human oral absorption values of the compounds showcase their sound intestinal absorption. These ligands also possessed sound plasma protein binding and skin permeability. Furthermore, **K1** farely inhibited hERG while **D1** displayed threatening hERG inhibition and the literature also associated **D1** with CHF (congestive heart failure)[37]. The authenticity of the qikprop parameters are further confirmed by PreADMET server that produced similar results. Moreover, this server also identified both the ligands as inhibitors of CYP_2C9 and Pgp, and weak substrates of CYP_3A4.

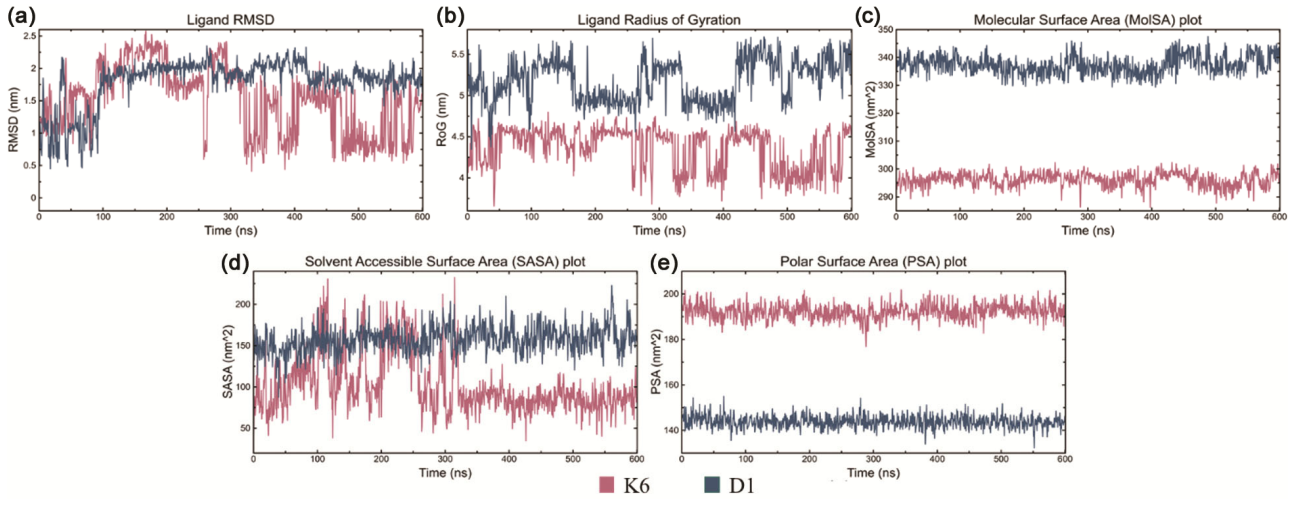


Fig. 10 — Overlap of **K6** and **D1** ligand properties of 600 ns MD trajectories, a)RMSD; b) Radius of Gyration; c) Molecular Surface Area (MolSA); d) Solvent Accessible Surface Area (SASA); and e) Polar Surface Area (PSA)

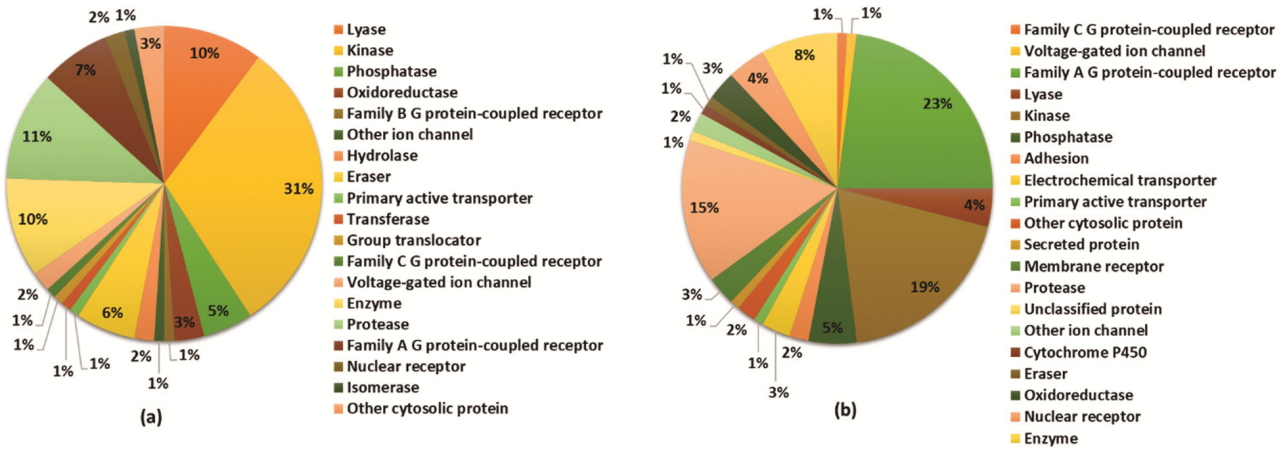


Fig. 11 — Targets predicted by Swiss, (a) Targets for **K6** and (b) Targets for **D1**

K1 is exceptionally druglike as it obeyed all rules of drug likeness except one violation in MDDR (MDL Drug Data Report) like rule due to presence of only 5 rotatable bonds. On the other hand, **D1** made two violations in Lead like rule due to its Mol.Wt and logP values and one violation in WDI (World Drug Index) like rule due to Kier-α3 index. Coming to the toxicity, both the drugs showed an LD_{50} (median lethal dose) of category 4. Furthermore, both ligands are immunotoxic; **K6** is also a mutagen and out of four selected strains of *Salmonella typhimurium* three showed positive results, while **D1** is active against MMP (Mitochondrial Membrane Potential), it retained only positive result for only one of the four *Salmonella* strains. **K6** displayed negative carcinogenicity in both mouse and rat but **D1** is positive in mouse, but the animal studies, including mouse and human trials deemed it non carcinogenic[37].

Target Prediction

To identify specific targets for the ligands the Swiss server was used, which provided 100 targets for each ligand and a lion's share of 31% was owned by kinases for **K6** and 23% was owned by G protein coupled receptors of family A for **D1** (Fig. 11). The identified kinase targets for **K6** include CDCs, Tyrosine kinases and not CLK kinases but DYRK1A was indeed a target. Intrestingly for **D1**, CLK and DYRK family of kinases were identified as targets, including CLK1 and CLK2 but not CLK3 whose pocket was congruent to PPARγ (Table S5).

Conclusions

The current study to identify potential CLK1 inhibitors *via* ligand-based pharmacophore approach afforded **K6** and **D1**. Both the ligands potentially targeted CLK1 with seldomly unstable complexes with an average Cα-RMSD around 0.247 nm and stable conformations depicted by PCA plots. The ligand **K6** maintained consistent interactions with the hinge residues while procuring a web of interactions near DFG motif owning to its nitro group and the Swiss server recognised **K6** as a promising kinase inhibitor but not as a CLK1 targeting agent. However, slight structural modifications might afford potent CLK1 inhibition. Furthermore, **D1** established DFG-in mode of binding while associating with hinge L244. The glitazone moiety established numerous water bridges and suprisingly the target prediction confirms **D1** as a CLK1 binder. Both the ligands also expressed sound PK profile making them excellent candidates worth studying.

Acknowledgements

VJA expressess gratitude to CSIR-Indian Institute of Chemical Technology (CSIR-IICT), Hyderabad for the Fellowship (Project: Design and Synthesis of Novel Kinase inhibitors; MLP-0102). DM thanks DST-INSPIRE for the Follewship (DST/INSPIRE Felloship/2021/IF210336). PY also appreciates Anti Viral Mission (HCP-041). SSJ thanks CSIR for providing infrastructure facility. IICT/Pubs./2023/368.

Supplementary Information

Supplementary information is available in the website http://nopr.niscpr.res.in/handle/123456789/58776.

References

1 Bullock A N, Das S, Debreczeni J E, Rellos P, Fedorov O, Niesen F H, Guo K, Papagrigoriou E, Amos A L, Cho S, Turk B E, Ghosh G & Knapp S, *Structure,*17 (2009) 352.

2 El Hady A K, El-Gamil D S, Abadi A H, Abdel-Halim M & Engel M, *Med Res Rev,*43 (2023) 343.

3 Song M, Pang L, Zhang M, Qu Y, Laster K V & Dong Z, *Sig Trans Tar Ther*,8 (2023) 148.
4 Lindberg M F & Meijer L, *Int J Mol Sci,*22 (2021).
5 Li C, Lu L J, Lian W-W, Pang X-C, Jia H, Liu A L & Du G H, *Acta Pharm Sin,*39 (2018) 1913.
6 Duncan P I, Stojdl D F, Marius R M & Bell J C, *Mol Cell Biol,*17 (1997) 5996.
7 Wong R, Balachandran A, Mao A Y Q, Dobson W, Gray-Owen S & Cochrane A, *Retrovirology*,8 (2011) 47.
8 Karlas A, Berre S, Couderc T, Varjak M, Barun P, Meyer M, Gangneux N, Karo-Astover L, Weege F, Raftery M, Schonrich G, Klemm U, Wurzlbauer A, Bracher F, Merits A, Meyer T F & Lecuit M, *Nat Comm*,7 (2016) 11320.
9 Beelontally R, Welkie G S, Lau B, Goodmaker C J, Ho C M K, Swanson C M, Deng X, Wang J, Gray N S, Davison A J & Strang B L, *Antiviral Res.*138, 61–67 (2017).
10 Bowler E, Porazinski S, Uzor S, Thibault P, Durand M, Lapointe E, Rouschop K M A, Hancock J, Wilson I & Ladomery M, *BMC Cancer*18, 355 (2018).
11 Blackie A C & Foley D J, *Bioorg Med Chem*,70 (2022) 116914.
12 Zhang Y, Xia A, Zhang S, Lin G, Liu J, Chen P, Mu B, Jiao Y, Xu W, Chen M & Li L, *Bioorg Med Chem Lett*, 41 (2021) 127881.
13 Yang T, Yang Y, Chen Y, Tang M, Shi M, Tian Y, Yuan X, Yang Z & Chen L, *Eur J Med Chem*, 250 (2023) 115168.
14 Zhou Q, Phoa A F, Abbassi R H, Houqe M, Reekie T A, Font J S, Ryan R M, Stringer B W, Day B W, Johns T G, Munoz L & Kassiou M, *J Med Chem*, 60 (2017) 2052.
15 *Schrödinger Release 2024-1: LigPrep, Schrödinger, LLC, New York, NY, 2024-1.*
16 Dixon S L, Smondyrev A M, Knoll E H, Rao S N, Shaw D E & Friesner R A, *J Comp Aided Mol Des*, 20 (2006) 647.
17 Mysinger M M, Carchia M, Irwin J J & Shoichet B K, *J Med Chem*, 55 (2012) 6582.
18 Yu Y, Dong H, Peng Y & Welsh W J, *Molecules*, 26 (2021) 5270.
19 Truchon J-F & Bayly C I, *J Chem Inf Mod*, 47 (2007) 488.
20 What are the ranges for fitness, vector, volume, and site scores in Phase? schrodinger Article ID: 1646. (https://support.schrodinger.com/s/article/1646#:~:text=In%20Table%206.1%2C%20you%20will,the%20default%20weight%20of%201.0.).
21 Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R & Sherman W, *J Comp Aided Mol Des*, 27 (2013) 221.
22 Friesner R A, Murphy R B, Repasky M P, Frye L L, Greenwood J R, Halgren T A, Sanschagrin P C & Mainz D T, *J Med Chem*, 49 (2006) 6177.
23 Jacobson M P, Pincus D L, Rapp C S, Day T J F, Honig B, Shaw D E & Friesner R A, *Prot Struc Fun Bioinfo*, 55 (2004) 351.
24 Sherman W, Beard H S & Farid R, *Chem Biol Drug Des*, 67 (2006) 83.
25 Bowers K J, Chow D E, Xu H, Dror R O, Eastwood M P & Gregersen B A, Proceedings of the 2006 ACM/IEEE Conference on Supercomputing,(IEEE Eplore, Tampa, FL, USA) 2006.(https://doi.org/10.1109/SC.2006.54).
26 Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang J Y, Wang L, Lupyan D, Dahlgren M K, Knight J L, Kaus J W, Cerutti D S, Krilov G, Jorgensen W L, Abel R & Friesner R A, *J Chem Theory Comp*, 12 (2016) 281.
27 Roos K, Wu C, Damm W, Reboul M, Stevenson J M, Lu C, Dahlgren M K, Mondal S, Chen W, Wang L, Abel R, Friesner R A & Harder E D, *J Chem Theory Comp*, 15 (2019) 1863.
28 Grant B J, Skjærven L & Yao X-Q, *Protein Sci*, 30 (2021) 20.
29 Humphrey W, Dalke A & Schulten K, *J Mol Graph*, 14 (1996) 33.
30 Python Package Index - PyPI. free-energy-landscape 1.0.3.
31 *Schrödinger Release 2024-1: QikProp*, (Schrödinger, LLC, New York, NY) 2024.
32 Lee S K, Lee I H, Kim H J, Chang G S, Chung J E & No K T, *Euro QSAR*, 9 (2004) 5.
33 Banerjee P, Eckert A O, Schrey A K & Preissner R, *Nucleic Acids Res*, 46 (2018) W257.
34 Daina A, Michielin O & Zoete V, *Nucleic Acids Res*, 47 (2019) W357.
35 Brahmaiah D, Bhavani A K D, Aparna P, Kumar N S, Solhi H, Guevel R L, Baratte B, Ruchaud S, Bach S, Jadav S S, Reddy C R, Roisnel T, Mosset P, Levoin N & Gree R, *Bioorg Med Chem*, 31 (2021) 115962.
36 Sklepari M, Lougiakis N, Papastathopoulos A, Pouli N, Marakos P, Myrianthopoulos V, Robert T, Bach S, Mikros E & Ruchaud S, *Chem Pharm Bull (Tokyo)*, 65 (2017) 66.
37 National Center for Biotechnology Information. PubChem Compound Summary for CID 77999, Rosiglitazone. https://pubchem.ncbi.nlm.nih.gov/compound/Rosiglitazone. (2023).
38 Schumann M & Armen R S, *PLoS One*, 8 (2013) e83533.