scientific reports

Check for updates

OPEN

Application of chemical similarity and bioisosteres to find allosteric inhibitors of type 2 lipid kinase γ

Priyanka Andola, Jishu Pagag, Gatta K. R. S. Naresh & Lalitha Guruprasad✉

Phosphatidylinositol phosphate kinases (PIPKs) exist in three isoforms: I, II, and III. Some of these enzymes are promising drug targets for cancer, metabolic disorders, and neurodegenerative diseases. Type II PIPKs are notable for their dual roles, to perform phosphorylation reactions acting as lipid kinases and to carry out catalytic-independent functions. The dysregulation of Type II PIPKs is linked to several diseases, including psychiatric disorders, cancer, and infections. There is still a need to explore for strong inhibitors of these kinases. This study used a similarity search method to find analogs of the known PI5P4K2C inhibitor, DVF (5-methyl-2-(2-propan-2-ylphenyl)-N-(pyridin-2-ylmethyl)pyrrolo[3,2-d]pyrimidin-4-amine). It utilized open-access platforms like SwissSimilarity, SwissBioisosteres, and the STITCH database for the initial screening of molecules. Drug-likeness assessment of the selected molecules was followed by molecular docking and molecular dynamics simulations to evaluate their binding affinity and stability. Post-simulation analysis revealed four promising hit compounds, each containing a pyrrole-pyrimidine core, which exhibited superior binding free energies and interactions at the allosteric site compared to DVF. These findings highlight potential candidates for further development as PI5P4K2C inhibitors.

Keywords Type II phosphatidylinositol 5-phosphate 4-kinase, Allosteric site, Molecular docking, Molecular dynamics simulations, MM-PBSA

The involvement of phosphatidylinositol phosphate kinases (PIPKs) in multiple diseases, particularly human cancers such as lung cancer and breast cancer, has led to interest in developing small molecule inhibitors that target different classes of these kinases. The three types of PIPKs generating phosphatidylinositol bisphosphates are, Type I phosphatidylinositol-4-phosphate 5-kinases (Type I PI4P5Ks), Type II phosphatidylinositol-5-phosphate 4-kinases (Type II PI5P4Ks) and Type III phosphatidylinositol-3-phosphate 5-kinases (PIKfyve). These enzymes catalyze the hydroxyl phosphorylation reaction of specific PI species to form double-phosphorylated lipids sequentially, regulating both substrate and product levels.

The PI5P4Ks use phosphatidylinositol 5 phosphate (PI5P) as a substrate to produce PI(4,5)P2, which is essential for autophagosome-lysosome fusion[1,2]. The cellular level of PI(5)P is decreased by the Type II PIPK-catalysed reaction that generates PI(4,5)P2 at specific cellular locations[3,4]. The main role of phosphoinositides (PI) is to interact with the specific and functionally important complement peripheral proteins by attracting it towards the membrane. The contribution of PI in the selection of specific peripheral protein for the membrane helps in regulating the activity of integral membrane proteins. The interaction of PI with the molecules that are recruited to the membrane or reside in the membrane for example, ion channels and transporters affect cellular functions. The PI5P4Ks have emerged as promising therapeutic targets for diseases like cancer, immunological disorders, and neurodegenerative conditions, due to their key role in regulating cell signaling pathways.

The Type II PI5P4K family consists of three isoforms, encoded by the PI5P4K2A, PI5P4K2B, and PI5P4K2C genes, which give rise to α, β, and γ forms, respectively. Among these isoforms, PI5P4K2A is the most active, while PI5P4K2C has less catalytic activity[5,6]. PI5P4K2C is found in the Golgi, autophagosomes, and endomembrane compartments[1,7]. The specific function of each isoform is determined by the variable N- and C- terminal regions of the lipid kinase domain and dimerization domain of PI5P4K[8]. PI5P4Ks regulate PI3K/Akt/mTORC pathways, insulin signaling, and are crucial for surviving metabolic stress[9]. Loss of PI5P4Ks leads to increased PI(4,5)P2 levels, which serve as a substrate for PI3K-catalysed production of PI(3,4,5)P3[10]. PI5P4K2C is also known to regulate the Notch pathway[11]. Furthermore, PI5P4Ks play a key role in immune modulation; for example, the lack of PI5P4K2C in mice resulted in increased levels of proinflammatory cytokines and T-helper cells due to mTORC1 signaling hyperactivation. Additionally, several studies have associated PI5P4K2C single nucleotide polymorphisms (SNPs) with autoimmune diseases[12].

School of Chemistry, University of Hyderabad, Hyderabad 500046, India. ✉email: lalitha.guruprasad@uohyd.ac.in

A precise regulation of the various PIPKs is crucial for maintaining the cellular PI pool. The altered expression of the α, β, and γ isoforms of Type II PIPKs has been observed in various cancers, including leukemias, glioblastomas, breast cancer and soft tissue sarcomas[13]. PI5P4Ks are critical for cancer cell survival under oxidative stress, as their depletion leads to decreased glucose metabolism, reduced oxygen consumption, and increased AKT phosphorylation. Many studies suggest that PI5P4Ks are promising therapeutic targets for cancer and other metabolic disorders. Several PI5P4K inhibitors have been developed, including isoform-specific and multi-isoform inhibitors.

The development of selective inhibitors for kinases has traditionally centered on the ATP-binding site. For example, I-OMe Tyrphostin AG-538 is a specific ATP-competitive inhibitor of PI5P4K2A, while SAR088m is a pyrimidine-2,4-diamine inhibitor specific to PI5P4K2B. PI5P4K2C has specific inhibitors, such as NCT 504 and NIH-12848[8,14]. However, this approach often results in cross-reactivity with other kinases, leading to unwanted side effects. In contrast, allosteric inhibition offers a promising alternative, as allosteric sites tend to be more specific to individual isoforms, reducing off-target effects.

Recently, "compound 40", 5-methyl-2-(2-propan-2-ylphenyl)-~{N}-(pyridin-2-ylmethyl)pyrrolo[3,2-d]pyrimidin-4-amine (DVF)[15] was developed from NIH-12848. The pIC_{50} values of DVF against the PI5P4Kα, PI5P4Kβ and PI5P4Kγ + isoforms were reported as < 4.3, < 4.6 and 6.2, respectively as measured by ADP-Glo assay and with a pIC_{50} value of 6.1 against PI5P4Kγ wild-type determined from InCell Pulse in intact cells. The protein PI5P4Kγ + refers to the PI5P4Kγ construct containing mutations corresponding to the residues in PI5P4Kα; three amino acids (QAR) insertion at location139 plus an additional 11 amino acid mutations, S132L, E133P, S134N, E135D, G136S, D141G, G142A, E156T, N198G, E199G, and D200E. The binding constants (K_Ds) were also determined for PI5P4Kγ-wild-type (68 nM) and PI5P4Kβ (> 30,000 nM) using commercial assays. These experimental results indicate the preferential binding of DVF to PI5P4K2C.

Allosteric binding pockets have been well exploited in the recent past for the purpose of drug design. For example, the virtual screening of diverse molecules from commercial libraries, followed by docking calculations, binding energy prediction, and structural clustering identified inhibitors that bind the allosteric pocket in Caspase-6. These computational findings were confirmed experimentally based on fluorescence-based assays[16]. The pharmacophore modeling of allosterically bound methylenetetrahydrofolate dehydrogenase 2 enzyme combined with molecular docking, molecular dynamics (MD) simulations, and free energy calculations was used for inhibitor design[17]. All-atom MD simulations and analysis of protein–ligand interactions were employed to understand the stimulation of ATPase and closure kinetics in the molecular chaperone Hsp90 by allosteric modulators[18]. Asciminib, an FDA-approved drug in October 2021 was designed as an allosteric inhibitor of BCR-ABL1 kinase through fragment-based drug design approach using NMR. The low-affinity fragments were optimised using similarity and pharmacophore searches based on multiple crystallised fragments[19]. Some recent review articles have summarised the merit of structure-based drug design[20–22].

Although potent inhibitors specific to PI5P4K2C have been reported, there is a need to develop more potent inhibitors specific to PI5P4K2C to explore their therapeutic potential as a drug target. Targeting allosteric binding pockets is a worthwhile strategy specifically for a protein target that has several structurally related proteins. Inhibitors that recognise different binding sites to target PI5P4Ks may improve the selectivity and minimize off-target effects, making them more effective therapeutic agents. Molecular similarity is a crucial concept in drug discovery, based on the assumption that structurally similar molecules display similar properties. Similarity assessments between small molecules proven to be effective in drug discovery are widely used in the early stages of drug development to find novel molecules and enhance the potency or pharmacokinetic properties of lead compounds. In the present work, we employed molecular similarity search and bioisosteric replacement strategies to improve the binding and efficacy of the previously identified small molecule DVF as an allosteric inhibitor of PI5P4K2C lipid kinase as summarised in Fig. 1.

Results and discussion

Allosteric site analysis and validation

The structure of PI5P4K2C features an allosteric binding pocket that is bound with DVF (5-methyl-2-(2-propan-2-ylphenyl)-~{N}-(pyridin-2-ylmethyl)pyrrolo[3,2-d]pyrimidin-4-amine) and AMP-PNP located close to the N-terminal β-sheet region deposited in the RCSB PDB under ID 7QPN[15]. Only the B chain in the crystal structure binds to another molecule of DVF. We have considered the consensus binding pocket of DVF in both A and B chains for all the computational studies. The allosteric binding pocket in PI5P4K2C consists of Asp161, Met162, Ser164, Asn165, Leu166, Tyr169, Leu182, Pro183, Phe185, Phe272, Leu273, Leu276, Ile278, Tyr281, Leu330, Ile331, Asp332, Leu334 and Thr335 amino acid residues. Additionally, several non-bonding interactions were observed (Fig. 2) with DVF. Such as, a hydrogen bond is formed between the side chain of Asn165 and the central pyrimidine N4 atom of DVF. The NH group linking the pyrrolopyrimidine and methylpyridine groups in DVF interacts with the carboxylate group of Asp332 via hydrogen bonding. The hydroxyl group of Thr335 forms a hydrogen bond with the pyridine group of DVF, which also engages in a *pi-alkyl* interaction with Pro183. A *pi-alkyl* interaction is also observed between the phenyl ring with an isopropyl group in DVF and Met162. Furthermore, a *pi-sigma* interaction occurs between Ile278 and the pyrrolopyrimidine group, while Asp161 forms a *pi-anion* interaction with the pyrrole group of the inhibitor. The presence of the isopropyl near the pyridine group of DVF contributes to the stability of the ligand conformation. Although the ligand is well-positioned inside the binding pocket, the bicyclic core containing the methyl pyrrole group remains solvent-accessible.

The validation of docking was performed to assess the accuracy of the docking procedure. This process involved redocking the reference molecule (DVF) into the allosteric site, followed by superposition and measuring the root mean square deviation (RMSD). The superposition of the docked conformations of the reference molecule DVF in the PI5P4K2C protein model, using PyRx and AutoDock 4.2 for validation purposes,

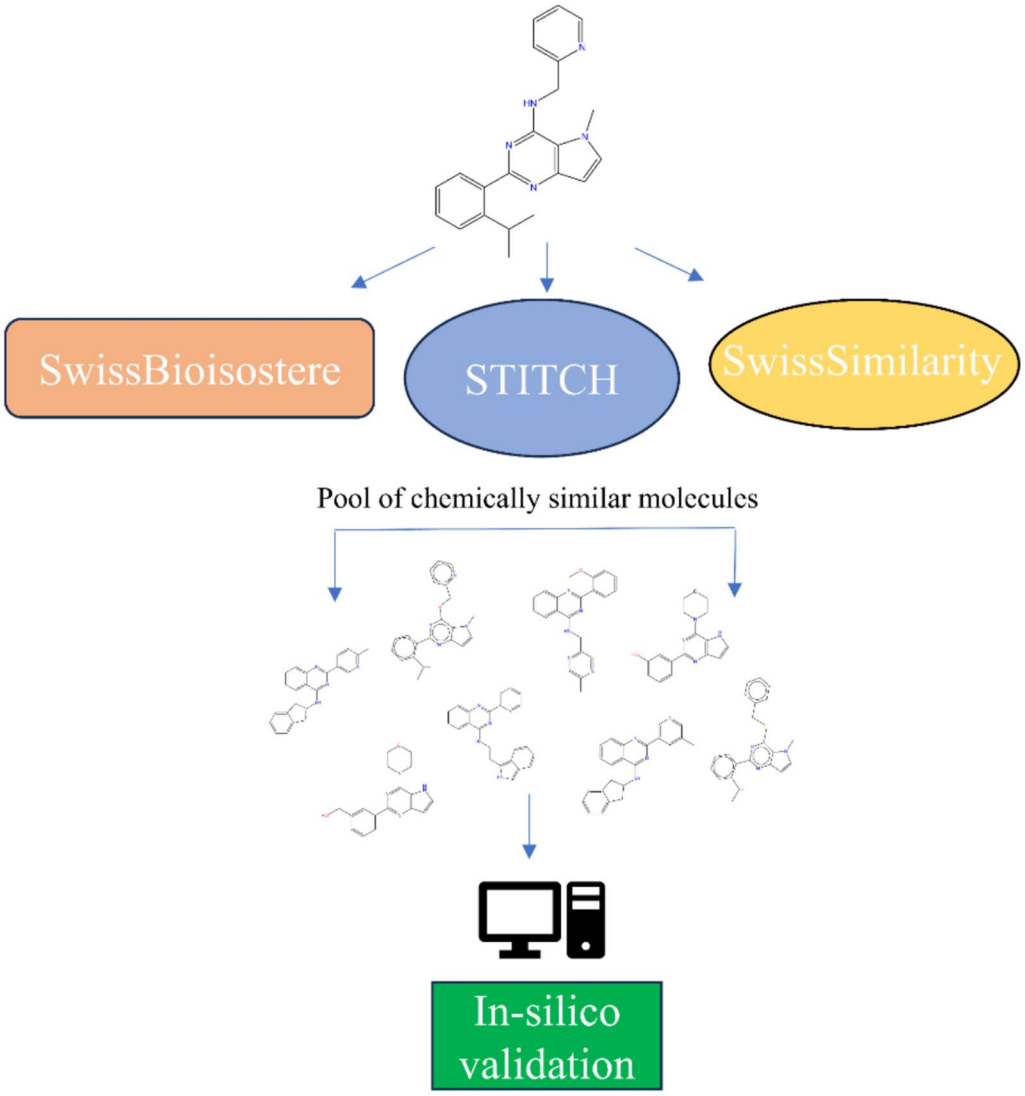


Fig. 1. Summary of workflow for the selection of molecules for in silico investigation as PI5P4K2C inhibitors.

is shown in Fig. 3. Thus, with this approach, we aimed to validate that the docking protocol adopted in this work allowed the re-docked ligand to bind to the allosteric site where it originally bound, with an RMSD of 0.45 Å (AutoDock 4.2) and 1.07 Å (PyRx) between the crystal structure and docked poses.

Molecule selection

All the compounds retrieved from similarity criteria by SwissSimilarity, STITCH, and SwissBioisostere were filtered based on their drug-likeness to eliminate the unfavoured molecules from the present study.

Initially, SwissSimilarity generated a total of 800 molecules, including the two classes: ChEMBL (actives only) and ZINC (lead-like) compounds. These molecules were then evaluated based on a similarity score, with a threshold set at 0.75. From this analysis, 143 molecules from the ChEMBL actives class and 115 molecules from the ZINC lead-like class were found to have a similarity score above the threshold, indicating a stronger potential for relatedness. Additionally, the query returned nine molecules in the STITCH server, that possessed Tanimoto scores between 0.53 to 0.57. A higher Tanimoto score (closer to 1) indicates greater similarity between the molecules. Molecules have been classified as similar molecules (Tanimoto scores are > 0.7) and as medium similar molecules (Tanimoto scores between 0.5 and 0.7)[23]. The molecules generated by the STITCH database had a Tanimoto score between (0.53 to 0.57) with the reference molecule, DVF, indicating that the molecules share only medium similarity with the query structure. Since we are targeting the allosteric site, we believe that the newly identified molecules may not bind the closely related off-targets strongly.

The identified molecules were CHEMBL2418346, CHEMBL2418347, AGN-PC-069BQU, AGN-PC-069BM8, AGN-PC-069BM9, AGN-PC-BMA, AGN-PC-ALR, AGN-PC-oMU1T5, AGN-PC-oN9PGP. Chemical-protein interactions (Fig. 4) were observed between CHEMBL2418346 and the mechanistic target of rapamycin (MTOR), similar to the interactions seen for CHEMBL2418347 and AGN-PC-069BQU. Additionally, a chemical-protein interaction was observed between AGN-PC-069BM8 and both phosphatidylinositol-4,5-bisphosphate 3-kinase,

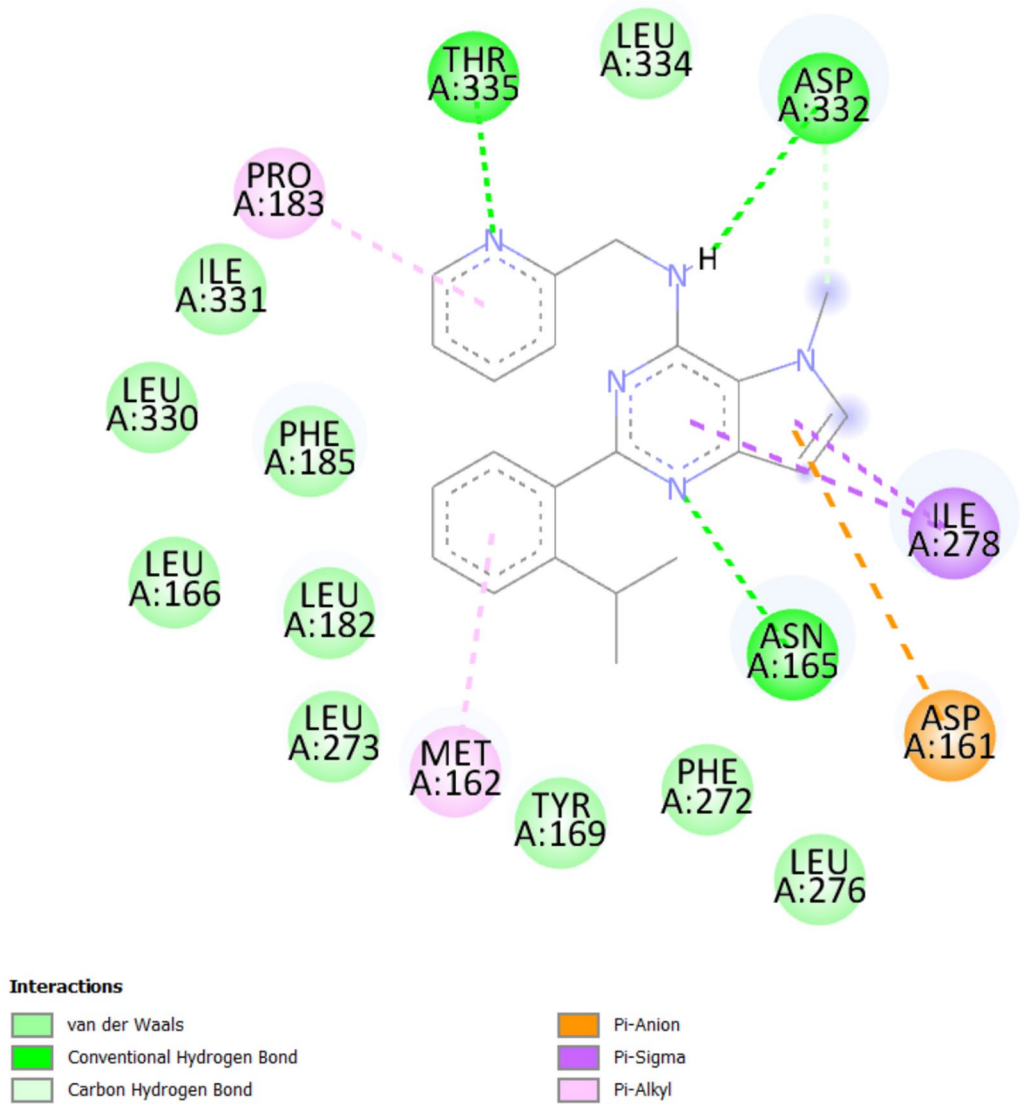


Fig. 2. The non-bonding interactions of the reference molecule, DVF at the allosteric site of PI5P4K2C.

catalytic subunit alpha (PIK3CA) and MTOR. Moreover, AGN-PC-069BM9 interacted with PIK3CA, and AGN-PC-oN9PGP with neuropeptide Y receptor Y5 (NPY5R). The network distance is representative of the strength and confidence such as the evidence suggesting a functional link between the predicted functional partners in a chemical-protein interaction. The green line indicates a chemical protein interaction with stronger associations represented by thicker lines. The protein–protein interactions are indicated in grey color.

The SwissBioisostere database was utilized to identify similar structures or fragments that returned 33 entries as bioisosteric replacements for the pyrrole-pyrimidine group in the reference molecule DVF. The database provided several fragments along with information on their impact on activity, as well as changes in properties like ΔLogP, ΔTPSA, etc., based on data from literature (PubMed) or bioassays (Assay ID). The results (Fig. S1) were analyzed, and two bioisosteric replacements containing oxygen and sulfur heteroatoms as a replacement for nitrogen heteroatoms[24] were selected. These selections were made based on their demonstrated improvement in activity, while maintaining minimal changes in LogP, TPSA, and molecular weight (MW).

ADME study

The molecules obtained from different sources were assessed for their drug-likeness properties using a webserver called SwissADME. Among the molecules above the similarity threshold of 0.75, around 133 molecules from ChEMBL and 84 molecules from ZINC qualified for the preliminary drug-likeness properties and were taken further for molecular docking studies. Nine molecules obtained from the STITCH server and two molecules selected from SwissBioisostere also maintained favorable drug-likeness properties and were used for docking studies. The details about the pharmacokinetic properties of the molecules selected for molecular docking studies are provided in Table S1.

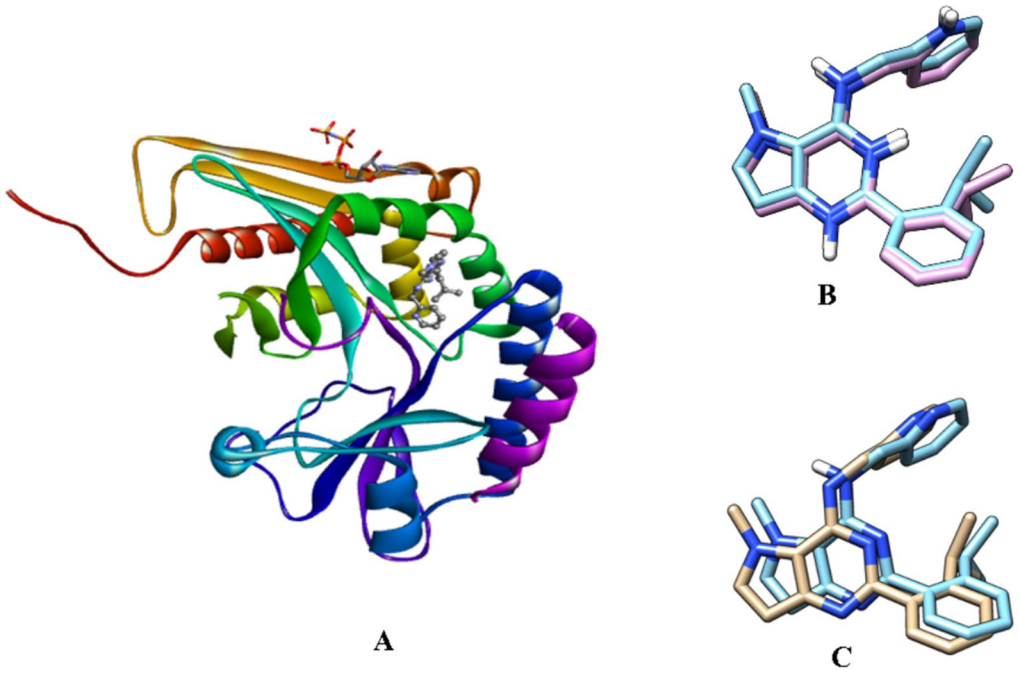


Fig. 3. (**A**) The PI5P4K2C bound to inhibitor molecule DVF deposited as PDB ID: 7QPN and docked conformer of reference molecule DVF in (**B**) PyRx (Co-crystallised DVF (cyan) and docked conformer (pink)). (**C**) AutoDock 4.2 (Co-crystallised DVF (cyan) and docked conformer (brown)).

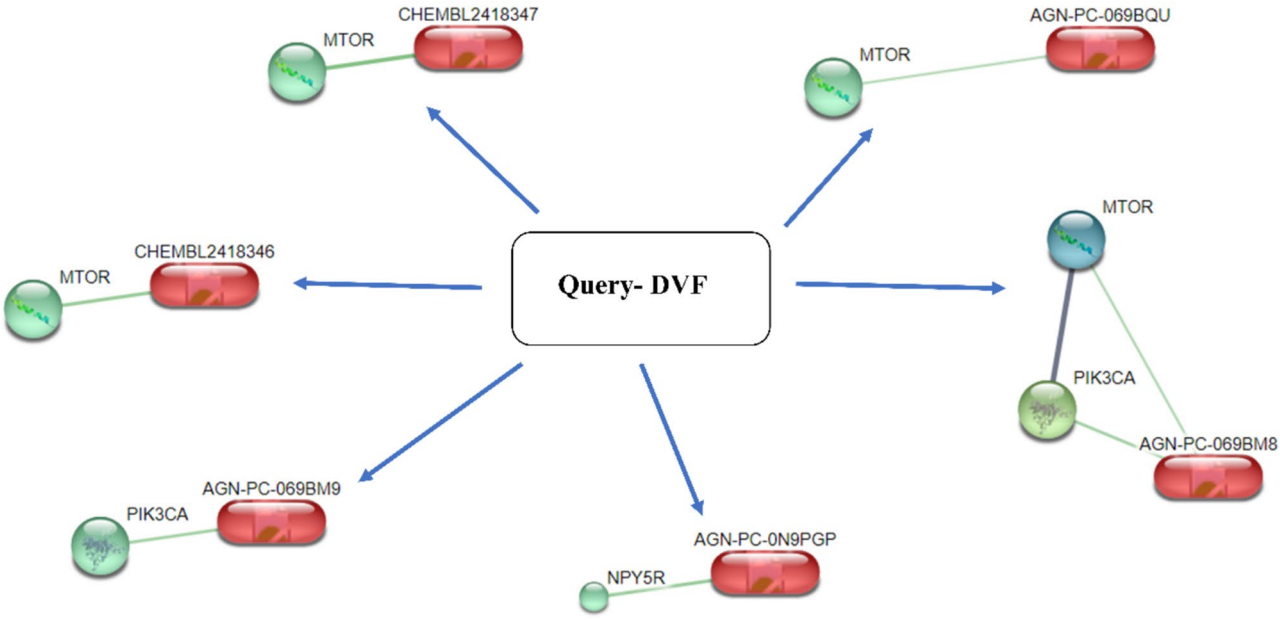


Fig. 4. The chemical-protein interactions existing between the chemical structures and protein targets.

The computational prediction of absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties is important for the subsequent steps of in-silico drug design. The predicted ADME properties of chemical compounds derived from the SwissADME webserver were analysed and the following results were observed:

- The TPSA values of the selected molecules ranged between 33.95 to 114.51 $Å^2$ within the optimal range of TPSA = 0 to 140 $Å^2$ suggestive of good oral bioavailability.

- The lipophilicity measured as consensus Log Po/w varied between 0.90 and 4.62 within the optimal range of Log Po/w = -4.0 to 5.6 indicating that molecules can permeate cell membranes easily and distribute effectively.
- The water solubility determined by Log S value ranged between -2.42 to -6.06 concerning the optimal value of Log S < 6 for the selected molecules suggestive of moderate solubility.
- The skin permeation factor also termed as Log Kp was recorded between -4.22 cm/s to -7.69 cm/s with respect to optimal value of Log Kp = -6.1 to -0.19 cm/s indicative of moderate to low skin permeability.
- The synthetic accessibility suggesting ease of synthesis varied from 2.10 to 3.86 with respect to synthetic accessibility scale = 1 to 10.

The selected compounds from the applied workflow (Fig. 5) were considered promising candidates for further investigation based on their structural and chemical properties.

Molecular docking analysis

After the assessment of the drug-likeness properties, the ligand structures obtained from various sources like ChEMBL (actives), ZINC (lead-like), STITCH, and SwissBioisosteres were prepared and utilized for docking purposes in PyRx and AutoDock software. The docking was first conducted in PyRx as a screening tool to identify molecules with greater binding affinity at the allosteric site. These molecules were docked again in AutoDock 4.2 tools to come up with the best hit molecules. In the docking process, a binding affinity value is assigned to each molecule to rank the molecules based on their binding potential. The lower the binding affinity score, the better the binding potential of the ligand.

The allosteric binding site comprises residues, Asp161, Met162, Ser164, Asn165, Leu166, Tyr169, Leu182, Pro183, Phe185, Phe272, Leu273, Leu276, Ile278, Tyr281, Leu330, Ile331, Asp332, Ile333, Leu334 and Thr335. The docking results were exported to Discovery Studio Visualizer to examine the binding interactions of the ligands with the receptor. Analysis of two-dimensional (2D) ligand interactions was done to eliminate the molecules from the study that did not form any conventional hydrogen bonds or other non-bonding interactions with the residues present in the PI5P4K2C allosteric site. These 2D protein–ligand interactions are shown in Table 1. A summary of the interactions is provided in Table 1 along with docking scores obtained from the docking studies. The reference molecule DVF achieved a binding score of -9.26 kcal/mol in AutoDock 4.2 and -10.4 kcal/mol in PyRx. The docking scores of the hit molecules ranged between -7.06 and -10.0 kcal/mol in PyRx and -7.31 kcal/mol to -11.1 kcal/mol in AutoDock 4.2. Analysis of the docked poses revealed that hydrophobic amino acid residues, Met162, Leu166, Tyr169, Leu182, Pro183, Phe185, and Ile278 interacted with the ligands via *Pi-alkyl* interactions, *Pi-Pi* stacked interactions, *Pi-Pi* T-shaped interactions, *Pi-sigma* interactions. The hydrophilic residues Asp161, Asn165, Asp332, and Thr335, formed hydrogen bonds with the ligands. Leu330 and Ile331 were also occasionally involved through hydrogen bonding and van der Waals (vdW) interactions with the ligands.

Thus, from this overall analysis, it was observed that the selected molecules were involved in the conventional non-bonding interactions with the allosteric binding pocket of PI5P4K2C. The protein–ligand interactions formed an essential criterion for the examination of hit molecules for further studies. In addition, the similarity scores between the reference and hit molecules (I-XI) selected from molecular docking were computed using

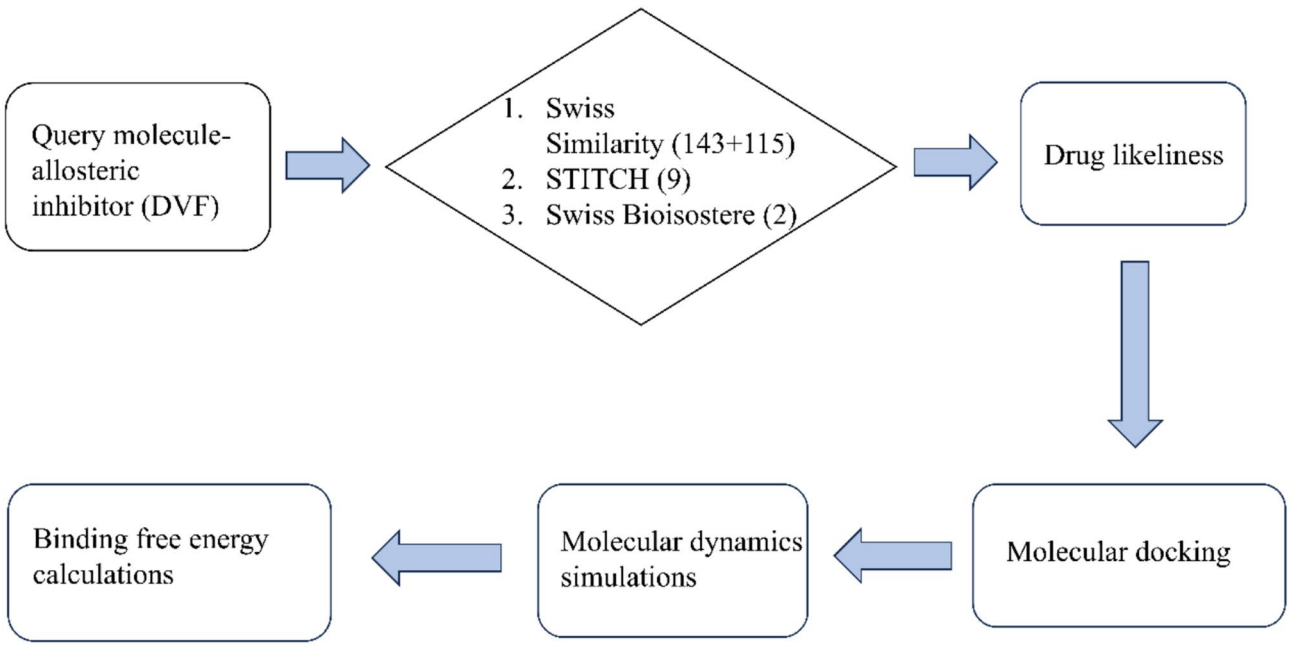
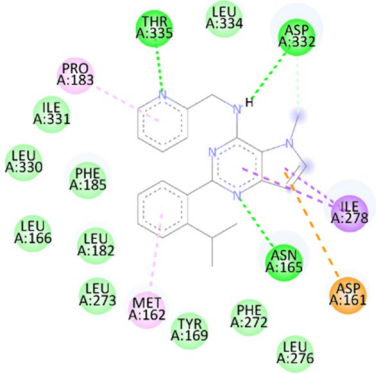
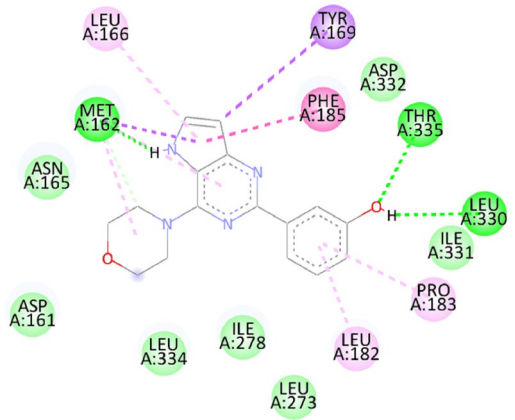
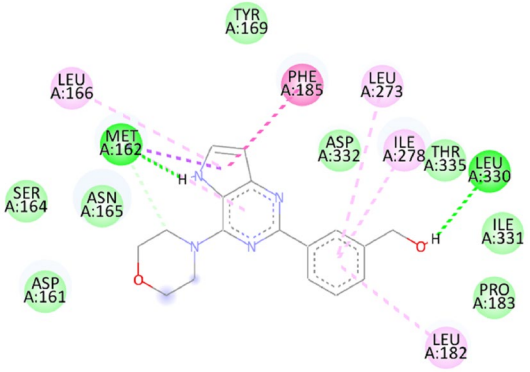
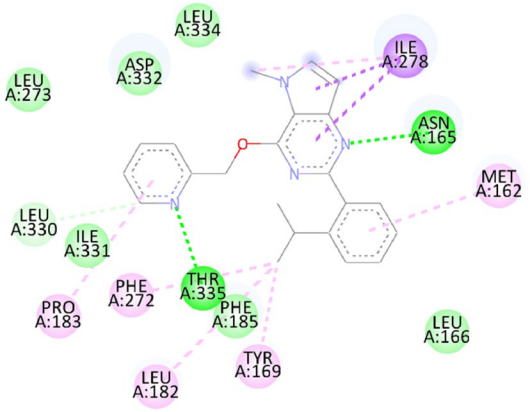
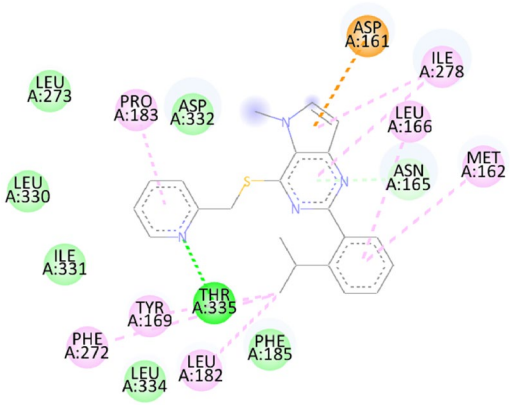
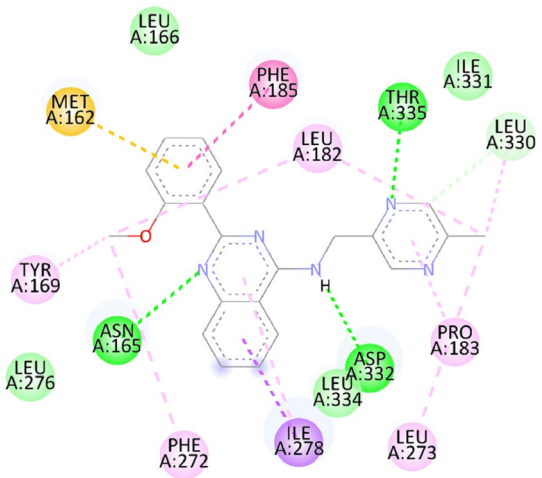
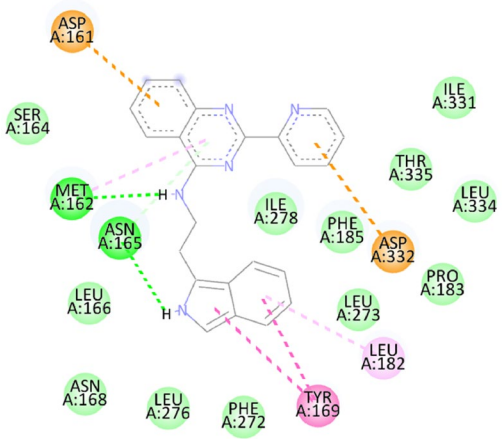


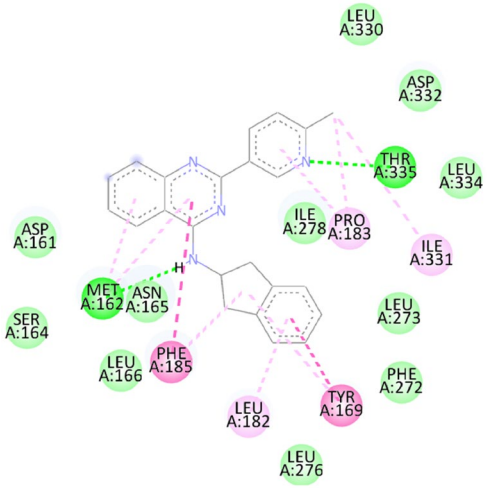
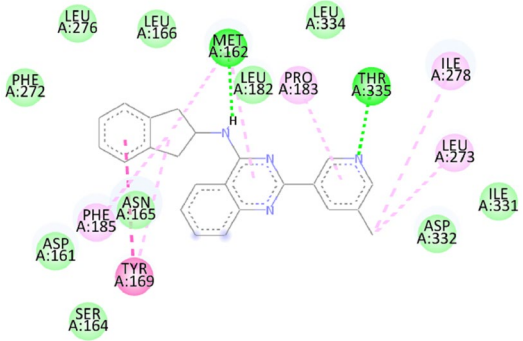
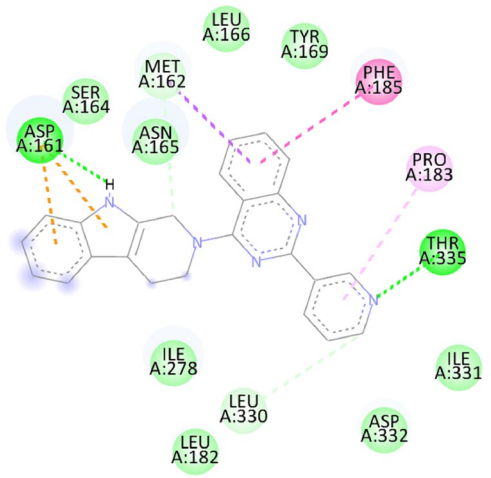
Fig. 5. The workflow of methods employed to identify potential hit molecules using allosteric inhibitor DVF as the query.

Sl. no	IUPAC nomenclature of ligand	2D plot of protein–ligand interactions	Docking Score AutoDock/ PyRx (kcal/ mol)	Amino acid residues in non-bonding interactions
1	DVF 2-(2-isopropylphenyl)-5-methyl-N-(pyridin-2-ylmethyl)-5H-pyrrolo[3,2-d] pyrimidin-4-amine		-9.26/-10.4	Hydrogen bond- Asn165, Asp332, Thr335 Pi Anion- Asp161 Pi Sigma- Ile278 Pi Alkyl- Met162, Pro183
2	3-(4-morpholino-5H-pyrrolo[3,2-d]pyrimidin-2-yl)phenol		-7.38/-8.0	Hydrogen bond- Met162, Leu330, Thr335 Pi Sigma-Tyr169 Pi-Pi T shaped- Phe185 Alkyl and Pi Alkyl- Leu166, Leu182, Pro183
3	(3-(4-morpholino-5H-pyrrolo[3,2-d]pyrimidin-2-yl)phenyl)methanol		-7.31/-7.8	Hydrogen bond- Met162, Leu330 Pi-Pi T shaped- Phe185 Pi Alkyl- Leu166, Leu182, Leu273, Ile278
4	2-(2-isopropylphenyl)-5-methyl-4-(pyridin-2-ylmethoxy)-5H-pyrrolo[3,2-d]pyrimidine		-9.07/-9.9	vdW and Carbon Hydrogen bond- Leu330 Hydrogen bond- Asn165, Thr335 Alkyl Pi Alkyl- Met162, Tyr169, Leu182, Pro183, Phe272 Pi Sigma- Ile278

Continued

Sl. no	IUPAC nomenclature of ligand	2D plot of protein–ligand interactions	Docking Score AutoDock/ PyRx (kcal/ mol)	Amino acid residues in non-bonding interactions
5	2-(2-isopropylphenyl)-5-methyl-4-((pyridin-2-ylmethyl)thio)-5H-pyrrolo[3,2-d]pyrimidine		**-9.18/-8.7**	Pi Donor hydrogen bond- Asn165 Hydrogen bond- Thr335 Pi Anion- Asp161 Alkyl and Pi Alkyl- Met162, Leu166, Tyr169, Leu182, Pro183, Phe272, Ile278
6	2-(2-methoxyphenyl)-N-((5-methylpyrazin-2-yl) methyl)quinazolin-4-amine		**-8.48/-9.6**	Carbon Hydrogen bond- Leu330 Hydrogen bond- Asn165, Asp332, Thr335 Pi Sigma- Ile278 Pi-Pi Stacked- Phe185 Pi Sulfur- Met162 Alkyl and Pi Alkyl- Tyr169, Leu182, Pro183, Phe272, Leu273
7	N-(2-(2H-isoindol-1-yl)ethyl)-2-(pyridin-2-yl) quinazolin-4-amine		**-9.35/-10.0**	Hydrogen bond- Met162, Asn165 Pi Anion- Asp161, Asp332 Pi-Pi Stacked- Tyr169 Pi Alkyl-Leu182

Continued

Sl. no	IUPAC nomenclature of ligand	2D plot of protein–ligand interactions	Docking Score AutoDock/ PyRx (kcal/ mol)	Amino acid residues in non-bonding interactions
8	N-(2,3-dihydro-1H-inden-2-yl)-2-(6-methylpyridin-3-yl)quinazolin-4-amin		-9.75/-10.9	Hydrogen bond- Met162, Thr335 Pi-Pi Stacked and Pi Pi T Shaped- Tyr169, Phe185 Alkyl and Pi Alkyl- Leu182, Pro183, Ile331
9	N-(2,3-dihydro-1H-inden-2-yl)-2-(5-methylpyridin-3-yl)quinazolin-4-amine		-10.0/-11.1	Hydrogen bond- Met162, Thr335 Pi Pi Stacked- Tyr169 Alkyl and Pi Alkyl- Pro183, Phe185, Leu273, Ile278
10	2-(2-(pyridin-3-yl)quinazolin-4-yl)-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indole		-9.11/-10.2	Hydrogen bond- Asp161, Thr335 vdW and Carbon Hydrogen bond- Asn165, Leu330 Pi Sigma- Met162 Pi Pi T shaped- Phe185 Pi Alkyl- Pro183 Pi Anion- Asp161

Continued

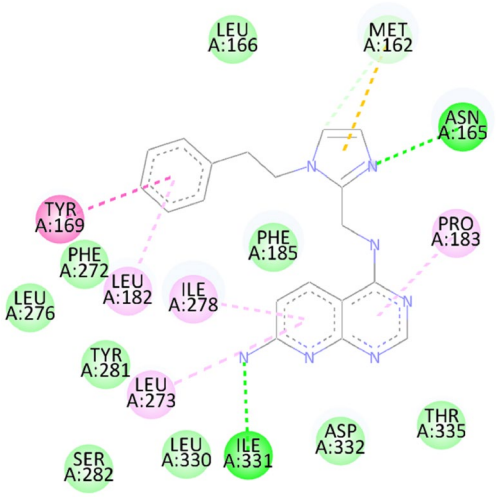
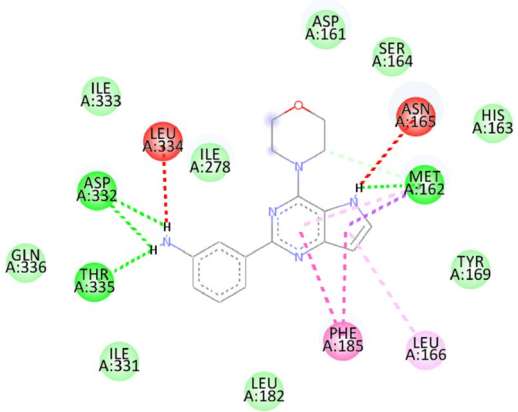
Sl. no	IUPAC nomenclature of ligand	2D plot of protein–ligand interactions	Docking Score AutoDock/ PyRx (kcal/ mol)	Amino acid residues in non-bonding interactions
11	N4-((1-phenethyl-1H-imidazol-2-yl)methyl) pyrido[2,3-d]pyrimidine-4,7-diamine		-8.86/-9.3	vdW and Carbon Hydrogen bond- Met162 Hydrogen bond- Asn165, Ile331 Pi Pi Stacked- Tyr169 Hydrogen bond- Asn165, Ile331 Pi Alkyl- Leu182, Pro183, Leu273, Ile278
12	3-(4-morpholino-5H-pyrrolo[3,2-d]pyrimidin-2-yl)aniline		-7.06/-8.0	Hydrogen bond- Met162, Asp332, Thr335 Pi Sigma- Met162 Pi-Pi T shaped- Phe185 Pi Alkyl- Leu166

Table 1. The docking scores of reference and ligand molecules at the allosteric site of PI5P4K2C along with the non-bonding interactions.

ChemMine Tools available at https://chemminetools.ucr.edu/similarity/ and provided in Table S2. During the analysis of compound pair similarity, the ChemMine interface calculates the atom pair and maximum common substructure (MCS) similarities with Tanimoto coefficients. The MCS is a graph-based similarity concept defined as the largest substructure shared between the two compounds. It is a pair-wise concept and the results (size of MCS relative to source structure) can be applied to compute similarity coefficients. MCS provides an accurate and sensitive similarity measure even for compounds with large size differences. The MCS Tanimoto and MCS size for the hit molecules varied from 0.25–0.63 and 11–21, respectively. A lower MCS Tanimoto score signifies that the molecules are structurally different with a small portion of common substructure. A higher MCS Tanimoto score suggests that the maximum common substructure between the molecules is large and they share similar bonding patterns or chemical motifs. The AP Tanimoto scores varied between 0.25–0.86 for the selected hit molecules obtained from molecular docking. A low score near 0.2 is suggestive of weak similarity in terms of commonality or overlap whereas a high score near 0.8 is representative of strong similarity or substantial overlap between the elements in the compared structures.

Molecular dynamics simulations

The MD simulations studies were used to identify stable protein–ligand complexes. Thirteen molecular systems were analyzed to investigate the protein dynamics of PI5P4K2C, both in its apo form and when complexed with the reference molecule DVF and hit compounds (Hit I to Hit XI), using GROMACS 5.1.4 software. Of these systems, four molecular complexes (Hit I to Hit IV) demonstrated greater stability during 250 ns MD simulations. The extent of deviations in the Cα atoms throughout the MD run was assessed using parameters like RMSD and root mean square fluctuations (RMSF), which were plotted as a function of time.

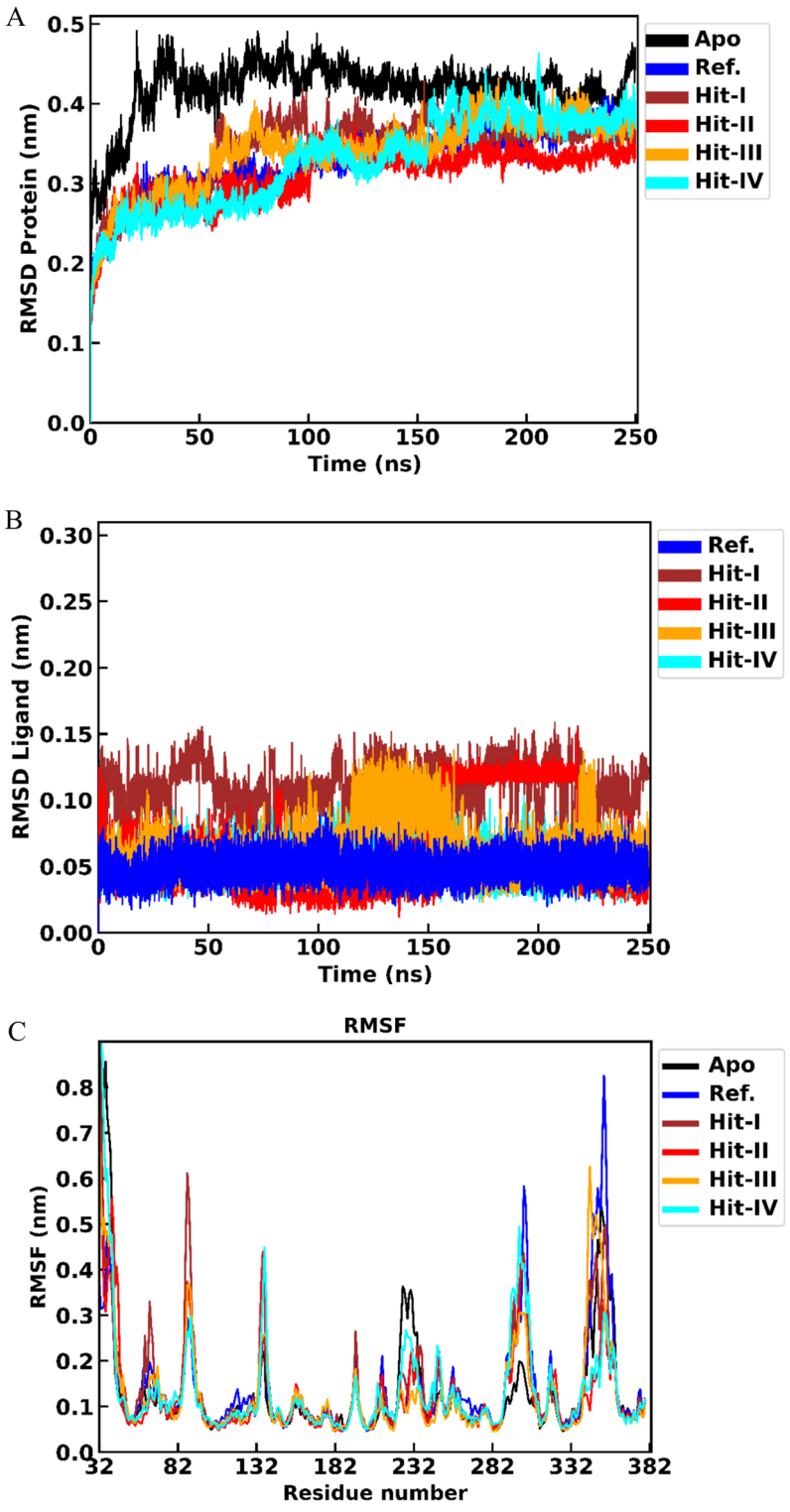


Fig. 6. The RMSD plots for the apo and complexed states of PI5P4K2C bound to reference and hit molecules as (**A**) RMSD protein (**B**) RMSD ligand (**C**) RMSF.

From the protein RMSD plots (Fig. 6A) it was observed that the RMSD of the apo protein reached up to 0.47 nm, while the protein RMSD was lower when complexed with DVF and the hit molecules, indicating the enhanced stability of the protein–ligand complexes. Among all molecules, Hit II displayed the lowest protein RMSD. The PI5P4K2C when complexed with the reference and hit molecules, the RMSD increased to 0.3 nm till 50 ns and further slowly increased to 0.37 nm till 150 ns. The clustering of all molecular complexes was observed at RMSD below 0.4 nm, however, in the Hit IV bound molecular complex, a noticeable change was

observed after 102 ns (0.33 nm) and 197 ns (0.38 nm) after which the complex remained stable. The RMSD of ligands (Fig. 6B) remained below 0.15 nm with reference molecule and Hit IV showing a minimum deviation of 0.05 nm. Thus, from the RMSD plots it was observed that the molecular systems attained stability when bound to hit and reference molecules when compared to the apo state alone. The identification of flexible residues in the molecular system was understood from the RMSF plot (Fig. 6C) which showed larger fluctuations (> 0.2 nm) of residues towards the N-terminal end of the protein. In all the complexes, residues from Glu295-Phe305 and Lys341-Thr360 were found to be flexible long loops. Incidentally, these regions were not defined in the three-dimensional crystal structure (PDB ID:7QPN). In addition, the regions, Ser225-Lys231 and Asn247-Asn249 located around the hinge region were also found to be flexible. In the N-terminal domain residual fluctuations were seen for Asn86-Arg91 that is present as a connecting loop between the two β strands. Similarly, flexibility was seen for the glycine-rich loop situated at the top of the hinge binding pocket and comprising residues Gly135-Asp138. It was studied that these fluctuations occurred outside the allosteric binding site and hence did not lead to conformational changes at the binding site. Further significant fluctuations in these regions reveal the structural alterations leading to protein conformational changes required for the enzyme activity. Low RMSF values (< 0.2 nm) observed for the allosteric binding site residues are indicative of the stability gained upon ligand binding. It was studied that all the interacting amino acid residues present in the allosteric site exhibited a low RMSF value of less than 0.2 nm, suggesting that the hit-bound molecular systems showed good conformational stability.

The interactions between the hit molecules and residues within the allosteric binding site were examined using the structures from most populated cluster obtained from MD simulations trajectories (Table 2). Key residues, such as Met162, Asn165, Leu330 and Thr335 were involved in hydrogen bonding across nearly all protein–ligand complexes. This observation is further supported by the number of frames (Table S3) and distance plots (Fig. S2), which indicate the presence of these intermolecular hydrogen bonds across 25,000 frames per complex. The Fig. S3 illustrates the number of hydrogen bonds as a function of time throughout the MD simulations. In addition, a comparison of the average structure and the structure from cluster-1 is shown in Table S4.

Mechanical stiffness and normal mode analysis

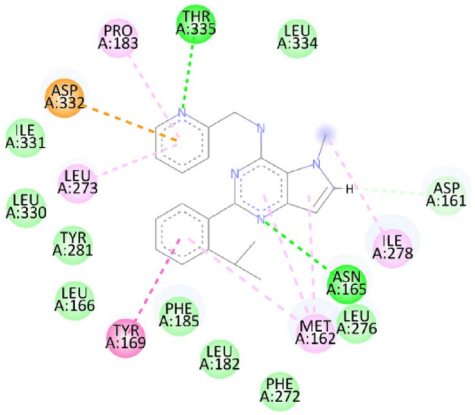
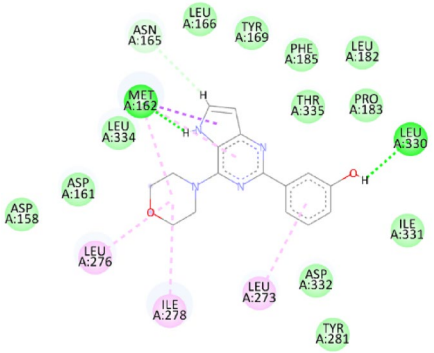
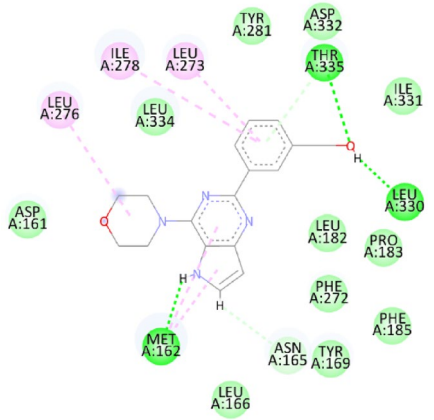
The mechanical stiffness plots of all the molecular systems were generated using the Anisotropic Network Model. The study was carried out to gain information about the relative mobilities of regions under tension and resistance to deformation as they might be relevant for biological function. The mean value of the effective spring constant was computed for all the molecular systems.

In the mechanical resistance map (Fig. S4) the blue-colored regions indicate mechanically stiff regions while red-colored regions denote easily deformable parts of the protein. Interestingly, in the apo protein, only the disordered N-terminal amino acid region Val34-Gln36 was mechanically weak with a spring constant value close to 6.3 k (a.u), while the rest of the apo PI5P4K2C displayed a relative stable structure with > 7 k (a.u). The Hit II and Hit IV showed a mechanically weak region with a spring constant value close to 6.4 and 6.5 k (a.u) at Gly303-Phe305. While the PI5P4K2C complex with Hit III displayed mechanically weak region around His345-Thr360 with a spring constant value close to 6.2 k (a.u). In general, the regions Asn86-Arg91, glycine-rich loop Gly135-Asp138, Ser225-Lys231, Asn247-Asn249, Glu295-Phe305 and Lys341-Thr360 with an effective spring constant value lower than 8 k (a.u) reflected the lower mechanical stiffness. These regions are more disposed to deformation than others. While the stable regions in the protein with well-defined secondary structures displayed an effective spring constant value around 11 k (a.u) in the mean plots. Thus, through this study elastic nature of the PI5P4K2C protein was studied.

NMA is a simple and fast method to determine the vibrational modes and flexibility in the protein. It provides a time-independent approach to gain an understanding of the mechanism of slow and large amplitude motions. The protein dynamics analysis using NMA was carried out and ten normal modes were calculated for each molecular system. Among these first three modes were chosen and studied for comparison with the apo structure of PI5P4K2C. The study revealed the flexible regions corresponding to the N-terminal loop region in the apo PI5P4K2C and in complex with Hit II and Hit III. In all the structures, residues Asn86-Arg91 that form a connecting loop between the two β strands in the N-terminal domain showed highest mobility. The complex with Hit IV showed the least mobility among all the molecular systems studied. The complex with Hit II displayed a dynamical structure in highly mobile loops situated near the hinge binding pocket *i.e.* Glu295-Phe305 and the helix-loop region (Lys341-Thr360) that connects to the N-terminal domain in PI5P4K2C. These large motions from the NMA plots are indicated in Fig. S4.

Principal component analysis and free energy landscapes

The various structural and conformational changes involved during protein–ligand binding can be studied using the essential dynamics, employing PCA. In this method, eigen vectors and eigen values are calculated to describe the motion of atoms as well as the atomic contribution for that movement during MD trajectories of protein. The PCA calculations were performed for the 250 ns MD simulations trajectories of each molecular system. The resultant 25,000 frames were utilized to know the Cα atom's motion as a function of time. The PC1 and PC2 projected into two-dimensional space as they captured most of the variance and conformational ensembles of PI5P4K2C were analysed. The representation of scatter plots for each molecular system is provided in Fig. 7. From the plots, it is observed that conformational changes took place in the protein when it is bound to the ligand as seen from scatter plots (Fig. 7 B-F) when compared to apo-protein (A) indicating that ligand binding is responsible for different conformational changes in protein dynamics. In addition, the conformational distribution of PI5P4K2C in protein–ligand complexes when bound to reference molecule DVF (B), Hit I (C), Hit II (D), and Hit III (E) were almost similar indicative of the similarity between reference and hit molecules. The lesser variance observed in PI5P4K2C- Hit I, PI5P4K2C-Hit II and PI5P4K2C-Hit III complexes is suggestive

Sl. No	Ligands	Cluster1	Non-bonding interactions in the complex
1	DVF 2-(2-isopropylphenyl)-5-methyl-N-(pyridin-2-ylmethyl)-5H-pyrrolo[3,2-d]pyrimidin-4-amine		vdW and carbon hydrogen bond-Asp161 Hydrogen bond- Asn165, Thr335 Pi-Pi T shaped- Tyr169 Alkyl and Pi alkyl-Met162, Pro183, Leu273, Ile278
2	Hit I 3-(4-morpholino-5H-pyrrolo[3,2-d]pyrimidin-2-yl)phenol		vdW and carbon hydrogen bond- Asn165 Hydrogen bond- Met162, Leu330 Alkyl and Pi-alkyl-Leu273, Leu276, Ile278
3	Hit II (3-(4-morpholino-5H-pyrrolo[3,2-d]pyrimidin-2-yl)phenyl) methanol		vdW and carbon hydrogen bond-Asn165 Hydrogen bond- Met162, Thr335 Alkyl and Pi-alkyl-Leu273, Leu276, Ile278

Continued

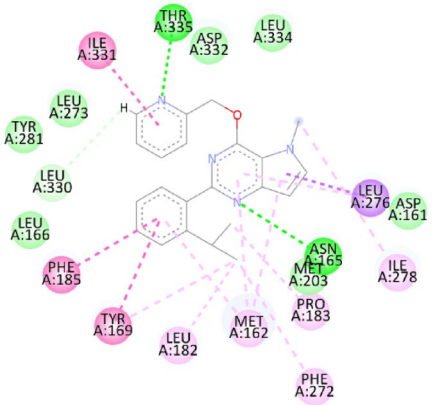
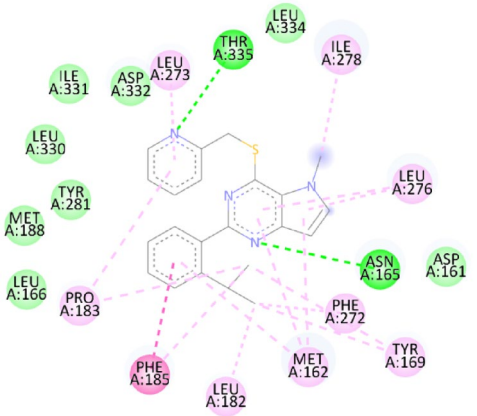
Sl. No	Ligands	Cluster1	Non-bonding interactions in the complex
4	Hit III 2-(2-isopropylphenyl)-5-methyl-4-(pyridin-2-ylmethoxy)-5H-pyrrolo[3,2-d]pyrimidine		vdW and carbon hydrogen bond-Leu330 Hydrogen bond- Asn165, Thr335 Pi-sigma-Leu276 Pi-Pi T-shaped and Amide-Pi stacked- Tyr169, Phe185, Ile331 Alkyl and Pi-alkyl-Met162, Leu182, Pro183, Phe272, Ile278
5	Hit IV 2-(2-isopropylphenyl)-5-methyl-4-((pyridin-2-ylmethyl)thio)-5H-pyrrolo[3,2-d]pyrimidine		Hydrogen bond- Asn165, Thr335 Pi-Pi stacked- Phe185 Alkyl and Pi-alkyl-Met162, Tyr169, Leu182, Pro183, Phe272, Leu273, Leu276, Ile278

Table 2. Intermolecular interactions of reference molecule and identified hit molecules (I—IV) bound to PI5P4K2C from the structure of most populated clusters from MD simulations trajectories.

of the stability gained upon binding to the hit molecules with similar structural characteristics. PI5P4K2C-Hit II molecular system, in particular, showed less overall motion suggestive of the stability acquired whereas Cα deviations were greater for apo PI5P4K2C and PI5P4K2C-Hit IV molecular systems that displayed high-frequency motion among all the systems studied. Thus, large protein conformational changes were observed in the apo PI5P4K2C and PI5P4K2C-Hit IV bound molecular systems.

The 2D free energy landscape (FEL) plots of PI5P4K2C in apo form and when bound to reference and hit molecules (Hit I to Hit IV) were constructed to understand the free energy profiles (Fig. S5) due to changes in the conformational dynamics during 250 ns MD simulations run. The plots were generated using the eigen vectors obtained from PCA. The 2D contour maps indicate the regions with high or low energy structures. The red colour patterns determine the maximum energy structures and the regions with minimum energy structures are indicated with a deep blue colour. The apo PI5P4K2C covered larger ranges of PC1 and PC2 (18.1 kJ/mol), indicating a more rugged free energy surface than the complexed protein. The apo protein clearly showed one free energy well in four basins. When complexed with reference, Hit II and Hit III molecules there are two energy wells in four basins. While the Hit I and Hit II displayed one conspicuous free energy well in more than 3 basins. The apo PI5P4K2C showed Gibbs free energy in the range of 0 to 18.1 kJ/mol while the PI5P4K2C bound to reference molecule DVF and other hit molecules exhibited Gibbs free energy in the range of 0 to 17 kJ/mol. This observation suggested that the apo protein underwent significant structural alterations compared to its initial state while the molecular complexes of PI5P4K2C bound to reference and hit molecules were more stable.

CarcinoPred-EL study

The toxicity of reference and selected hit molecules as potential carcinogens were studied from CarcinoPred-EL server. It was found that all the four hit molecules were predicted to be non-carcinogens by Ensemble RF, Ensemble SVM, and Ensemble XGBoost models. The details involved in the prediction of class for the studied molecules are represented in Table 3. This study suggests that the hit molecules may serve as clinically safer compounds than the reference molecule DVF.

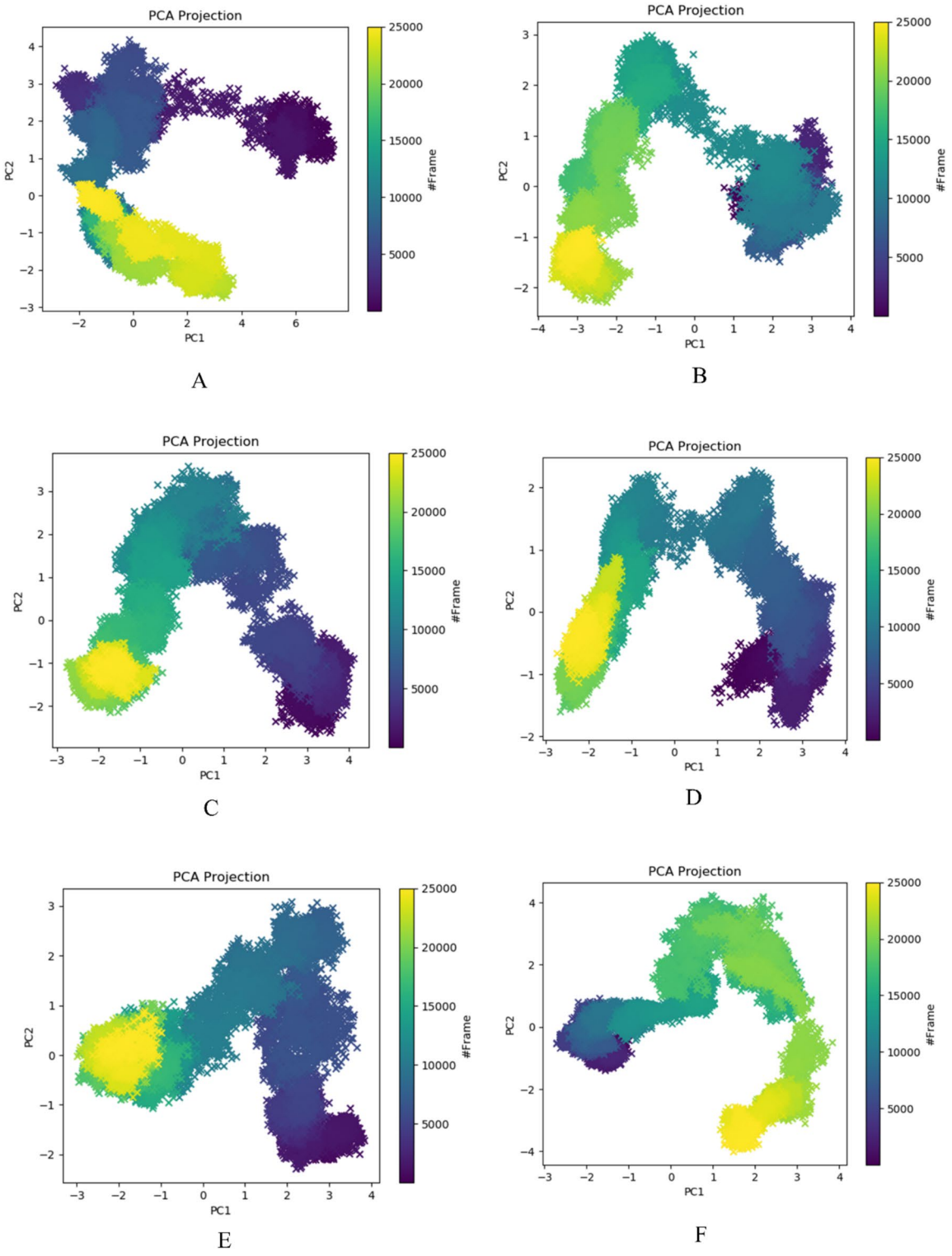


Fig. 7. The PCA scatter plots for PI5P4K2C in (A) apo and complexes bound to (B) DVF (C) Hit I (D) Hit II (E) Hit III (F) Hit IV.

Bioactivity prediction study

The selected hit molecules were evaluated for the potential to inhibit Type II PI5P4K2C using the predict bioactivity tool in Molinspiration Chemoinformatics [Molinspiration Cheminformatics free web services, Slovensky Grob, Slovakia]. The predicted bioactivity scores as kinase inhibitor for reference and hit molecules (I-IV) were found to be 0.24, 0.33, 0.71, 0.47, and -0.07 respectively. The bioactivity prediction analysis indicated that the hit molecules can exhibit promising activity towards PI5P4K2C, as measured by bioactivity scores for

Ligand	Ensemble	CDK	CDKExt	CDKGraph	KR	KRC	MACCS	Pubchem	Average	Class
DVF	RF	0.45	0.46	0.37	0.44	0.43	0.57	0.37	0.44	Non-Carcinogen
	SVM	0.60	0.43	0.49	0.46	0.46	0.60	0.52	0.51	Carcinogen
	XGBoost	0.07	0.19	0.47	0.45	0.85	0.63	0.76	0.49	Non-Carcinogen
Hit I	RF	0.42	0.34	0.34	0.46	0.62	0.34	0.42	0.42	Non-Carcinogen
	SVM	0.39	0.40	0.39	0.38	0.49	0.33	0.43	0.40	Non-Carcinogen
	XGBoost	0.07	0.19	0.47	0.45	0.78	0.63	0.76	0.48	Non-Carcinogen
Hit II	RF	0.39	0.36	0.35	0.39	0.54	0.30	0.32	0.38	Non-Carcinogen
	SVM	0.40	0.38	0.41	0.32	0.44	0.29	0.39	0.37	Non-Carcinogen
	XGBoost	0.07	0.19	0.47	0.45	0.78	0.63	0.76	0.48	Non-Carcinogen
Hit III	RF	0.40	0.39	0.48	0.36	0.53	0.35	0.40	0.42	Non-Carcinogen
	SVM	0.31	0.41	0.53	0.36	0.35	0.35	0.44	0.39	Non-Carcinogen
	XGBoost	0.07	0.19	0.47	0.45	0.70	0.63	0.76	0.47	Non-Carcinogen
Hit IV	RF	0.45	0.39	0.40	0.29	0.50	0.44	0.29	0.39	Non-Carcinogen
	SVM	0.36	0.38	0.52	0.24	0.28	0.38	0.37	0.36	Non-Carcinogen
	XGBoost	0.07	0.19	0.47	0.45	0.70	0.63	0.76	0.47	Non-Carcinogen

Table 3. Prediction of chemical carcinogenicity for reference and identified hit molecules.

Sl. No	Ligand	vdW Energy	Electrostatic Energy	Polar Solvation Energy	SASA Energy	$\Delta G_{(MM\text{-}PBSA)}$
	Ref	-101.044 ± 2.127	-23.765 ± 0.519	65.324 ± 1.241	-8.954 ± 0.190	-68.439 ± 1.799
	Hit I	-102.695 ± 1.475	-35.003 ± 0.595	83.532 ± 1.062	-10.711 ± 0.149	-64.925 ± 1.192
	Hit II	-148.159 ± 1.479	-40.713 ± 0.422	99.208 ± 0.862	-13.890 ± 0.139	-103.639 ± 1.258
	Hit III	-215.035 ± 0.551	-41.866 ± 0.172	111.946 ± 0.365	-18.734 ± 0.048	-163.677 ± 0.488
	Hit IV	-200.877 ± 0.941	-35.015 ± 0.269	103.247 ± 0.520	-17.908 ± 0.084	-150.650 ± 0.849

Table 4. Binding free energies (kJ/mol) from MM-PBSA study of reference molecule DVF, and identified hit molecules bound to PI5P4K2C.

kinase inhibitors. The hit molecules specifically Hit I to Hit III, showed higher scores than the reference molecule, with scores ranging from 0.33 to 0.71 as kinase inhibitors. In contrast, Hit IV exhibited the least potential and showed a negative value of predicted kinase inhibitory activity.

Binding free energy calculations

The binding energies of the molecular systems were calculated using the *g_mmpbsa* tool, based on the MD trajectories. The Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) ΔG analysis (Table 4) revealed that Hit II (-103.64 kJ/mol), Hit III (-163.67 kJ/mol), and Hit IV (-150.65 kJ/mol) exhibited significantly lower ΔG values compared to the reference molecule DVF (-68.44 kJ/mol), indicating stronger binding affinity and greater stability in these complexes. While Hit I displayed a ΔG value of -64.92 kJ/mol, comparable to that of DVF, suggesting a similar level of binding efficiency.

Further breakdown of the energy components; vdW, electrostatic, polar, and apolar energies showed that the hit molecule-bound systems had higher contributions from these interactions than the reference molecule. This suggests that the enhanced stability of Hit II, Hit III, and Hit IV is likely due to stronger non-covalent interactions, which contribute to their more favorable binding energies. The contribution of individual residues to the binding free energy was also analyzed, as illustrated in Fig. 8.

From the residue contribution plot, it was observed that amino acid residues such as Met162, Tyr169, Pro183, Phe185, Leu273, Leu276, Ile278, and Leu334 in the allosteric site contributed more towards the binding free energy values. The findings of the study aligned with results obtained from docking studies since these residues were seen to be involved in several non-bonding interactions with ligands. Thus, from this study, it is inferred that Hit II, Hit III, and Hit IV molecules possess lower ΔG values in comparison to the reference molecule.

Conclusions

PI5P4K lipid kinases are critical regulators of various cellular processes, including lipid transport, integral membrane protein signaling, and the recruitment of proteins to different intracellular locations. Previous studies have identified PI5P4Ks as promising drug targets for the treatment of neurodegenerative and immunological disorders, as well as cancer. However, the development of potent inhibitors against these lipid kinases has been limited, highlighting the need for further research. While the 3D crystal structures of PI5P4K2C complexed with inhibitors have been solved, these inhibitors bind at the ATP-binding site, presenting a challenge for achieving selectivity. Thus, targeting allosteric sites on lipid kinases offers a more selective therapeutic approach.

One such allosteric inhibitor, DVF, has been identified to bind to PI5P4K2C, as in PDB ID 7QPN. This study aimed to identify structural analogs from different sources that bear similarity to the DVF molecule and

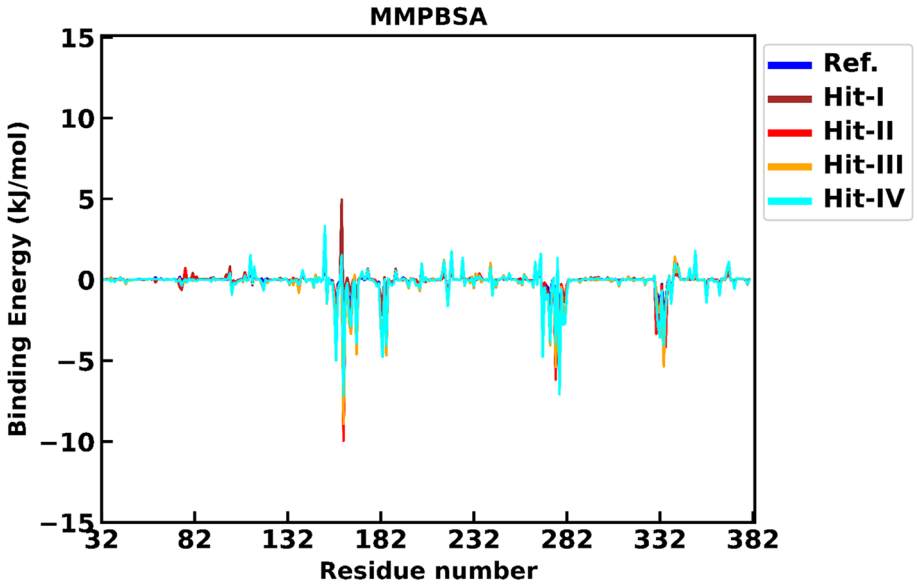


Fig. 8. The residue contribution in binding free energy values calculated for PI5P4K2C complexes bound to reference molecule DVF and identified hit molecules (Hit I-Hit IV).

investigate them as potential allosteric inhibitors by in-silico methods. Hence, structurally similar compounds were explored to identify potent molecules with enhanced binding affinity and improved pharmacokinetic properties than the reference molecule DVF. To achieve this, chemical and structurally similar molecules (~269) were collected from sources such as SwissSimilarity, STITCH, and Swiss Bioisostere. Initial screening of the generated output molecules for drug-likeness was followed by molecular docking and visualization of protein–ligand interactions to evaluate binding affinity.

From the docking studies, 12 hit molecules were identified, which were further evaluated for their stability through MD simulations when complexed with PI5P4K2C. The dynamic behaviour and structural stability of PI5P4K2C bound to hit molecules were monitored. Post-simulation analysis revealed that molecular complexes involving Hit I to Hit IV molecules demonstrated strong binding at the allosteric site of PI5P4K2C and remained stable throughout the 250 ns MD simulations. In addition, NMA, mechanical stiffness, and PCA were carried out to study the structural conformations and variations of Type II PI5P4K2C bound to these hit molecules. Evaluation of the molecules (Hit I to Hit IV) for carcinogenicity and predicted bioactivity suggested Hit I, Hit II, and Hit III molecules to be better structural analogs for DVF as they exhibited greater predicted kinase inhibitory potential than DVF and were non-carcinogenic. Moreover, Hit II, Hit III, and Hit IV identified as (3-(4-morpholino-5H-pyrrolo[3,2-d]pyrimidin-2-yl)phenyl)methanol, 2-(2-isopropylphenyl)-5-methyl-4-(pyridin-2-ylmethoxy)-5H-pyrrolo[3,2-d]pyrimidine, and 2-(2-isopropylphenyl)-5-methyl-4-((pyridin-2-ylmethyl)thio)-5H-pyrrolo[3,2-d]pyrimidine, respectively exhibited greater binding free energies than the reference molecule, DVF. Thus, these findings highlight Hit I, Hit II, and Hit III with their favorable non-bonding interactions and greater binding free energies compared to DVF as potential candidates for further development as PI5P4K2C allosteric inhibitors. These hit molecules with pyrrole-pyrimidine as the core moiety similar to DVF have the potential to bind to Type II PI5P4K2C protein with good affinity and can be used to design novel anticancer agents. These analyses provide valuable insights for the future design of potent drug candidates, particularly for targeting cancer. Further experimental studies may be needed to gain a clear understanding of the allosteric inhibitory potential of Hit I, Hit II, and Hit III molecules against PI5P4K2C.

Materials and methods

Protein preparation

The crystal structure of the γ isoform, PI5P4K2C, bound to the allosteric inhibitor DVF, with PDB ID 7QPN, was selected as the target protein for this study. The missing residues in the crystal structure of the protein (PDB ID: 7QPN) were built into the model structures predicted by the I-TASSER server[25] and used for this study. The best model was selected based on its C-score, TM-score, and cluster density. The output models, along with their prediction details, are provided in Fig. S6. Throughout the study, the co-crystallized inhibitor DVF was used as the reference molecule in the protein–ligand complex. The binding site residues were identified using the "define and edit binding site" feature in the Discovery Studio visualizer.

Construction of ligand library

Drugs with similar chemical structures have been found to exhibit similar pharmacological actions[26]. The concept of similarity has been recently used in drug repurposing studies for inhibition of SARS-CoV2 replication[27]. We obtained chemical molecules from freely available web servers, including SwissSimilarity[28], STITCH database[29], and SwissBioisostere[30], using the similarity approach.

SwissSimilarity

SwissSimilarity is a web tool that rapidly screens small molecules for potential ligands. It uses various approaches, including 2D molecular fingerprints (like FP2) and 3D similarity methods (such as Align-IT and Shape-IT), to make efficient predictions. SwissSimilarity has been used to build libraries for identifying histamine H3 receptor ligands[31] and analogs of PKMYT1 inhibitors[32].

In this study, we chose the ChEMBL database[33], which contains 402,239 active molecules, to represent bioactive compounds. We also used the ZINC database[34], which includes 2,817,546 purchasable lead-like compounds, as a compound library. The combination of FP2 and electroshape methods was used for searching similar molecules. Molecules with a similarity score of > 0.75 were selected from both the ChEMBL (actives only) and ZINC (lead-like) compound libraries for further study on drug-likeness.

STITCH database

The chemical structures available in the STITCH database, accessible at (http://stitch.embl.de), were retrieved by providing the query molecule DVF as a smiles string. The database offers several query options, including full-text searches for common names of proteins or chemicals. Additionally, similar chemicals stored in the database can be searched by providing the SMILES strings of the query compound. The chemical structures that matched the input structure displayed with Tanimoto scores were generated from the STITCH database.

SwissBioisostere

The lead optimization step in drug discovery is both time-consuming and challenging. This is because even slight modifications to active compounds against a target of interest can disrupt the balance between potency and key parameters like toxicity, bioavailability and metabolic stability. However, understanding bioisosteres can help identify analogs for synthesis, allow fine-tuning the biological and biophysical properties of the active compound, as seen with the reference molecule DVF in our study.

The SwissBioisostere database, accessible at http://www.swissbioisostere.ch, offers a wealth of information on molecular replacements for a query molecule, including their performance in biochemical assays. This enables researchers in drug discovery to identify bioisosteric modifications of lead molecules and view details of possible molecular replacements. The SwissBioisostere homepage features two chemical sketchers, allowing users to submit queries and perform replacements for a particular substructure. The results can be sorted by differences in LogP, TPSA, class counts, MW, and chemical similarity, all presented in a tabular format. A scoring scheme ranks the replacements based on observed bioactivity differences, target classes, number of targets, and Murcko scaffold families[35]. The results are based on values provided by the ChEMBL database, corresponding to the replaced fragment of a particular compound. The filtered results can be exported as a CSV file. Consequently, the reference molecule was used as a query in SwissBioisostere to search for potential replacements in the structure. By incorporating the best replacements into the original structure, we developed the reference molecule into bioisosteres.

The number of molecules with similar chemical structures to reference molecule DVF was compiled and downloaded in SDF format from SwissSimilarity, STITCH database, and SwissBioisostere. Before proceeding to molecular docking studies, these molecules were examined for their drug likeliness as a pre-filtering step.

ADME study

For a molecule to act as a drug, it must have the right physicochemical and pharmacokinetic properties. Studying ADMET properties of chemical structures early in the drug discovery pipeline can reduce failures in later clinical trials. Parameters like polarity, lipophilicity, size, water solubility, saturation, and gastrointestinal absorption are useful in determining a chemical compound's drug-likeness. Moreover, evaluating molecules based on relevant pharmaceutical physicochemical properties—such as MW, number of rotatable bonds, number of hydrogen bond acceptors, number of hydrogen bond donors, and Lipinski's rule of five helps ensure the study of clinically safer compounds.

In the present study, we estimated lipophilicity, polar surface area, solubility, skin permeation, and synthetic accessibility scale (1–10) measured by descriptors like Log Po/w (-4.0 to 5.6), TPSA (0 to 140 $Å^2$), Log S (< 6), and Log Kp (-6.1 to -0.19 cm/s) respectively from the SwissADME web server[36,37]. After the preliminary screening of the molecules for drug-likeness, we selected those with favorable pharmacokinetic properties for molecular docking studies.

Molecular docking

The type II lipid kinase PI5P4K2C bound to the reference molecule DVF was chosen as the target structure for virtual screening using PyRx[38]. The receptor and reference molecule DVF were imported into PyRx using the 'Import Molecule' tool. The 'make macromolecule' and 'make ligand' options were used to prepare the macromolecule and ligand structures for docking and to save them in pdbqt format. The ligands were energy-minimized using PyRx 0.8 (force field: Universal force field; optimization algorithm: conjugate gradients; the number of steps: 200; the number of steps for update: 1; stop if energy difference < 0.1). The binding site of the allosteric inhibitor, consisting of 5 Å residues around DVF, was used for docking the ligands in PyRx. The predefined allosteric binding site coordinates were adjusted to align the grid box, ensuring it covered all the amino acids at the protein's binding site. The output results, including docking scores and RMSD values for each ligand conformer, were analyzed to select the best molecules based on higher docking scores and lower RMSD values. The best molecules selected from PyRx screening were then docked using AutoDock 4.2 tools[39] at the receptor's allosteric site. The receptor and ligand molecules were loaded and prepared before docking. In the protein preparation step, polar hydrogens were added to the receptor, torsions were set, and the structure was saved in pdbqt format.

Similarly, ligands were loaded, and torsions were set to save them in pdbqt format. A grid box was defined around the inhibitor molecule DVF, centered at X = 63.470 Å, Y = 53.110 Å, and Z = 47.470 Å, which included the binding site residues. The grid box had a spacing of 0.375 Å. For each ligand, 10 docking poses were generated, with a maximum of 27,000 generations per GA run. The crossover rate was set to 0.8, and the gene mutation rate was set to 0.02. The conformers obtained for each ligand were visualized in the Discovery studio visualizer to analyze the non-bonding interactions at the allosteric site. To validate the docking protocol, the co-crystallized inhibitor DVF was re-docked at its allosteric site in the protein. The re-docked complex was superposed with the initial complex in UCSF Chimera, and the RMSD between the two complexes was calculated. This analysis assessed the docking protocols' efficiency and accuracy. The molecules with high binding affinity and a greater number of interactions in the allosteric site were selected for further MD studies.

Molecular dynamics simulations

The MD simulations study provides information about the conformational stability of protein–ligand complexes. To investigate this, MD simulations were conducted for the Type 2 PI5P4K2C model structure in both its apo form and complexed state, where the latter was complexed with the reference molecule and shortlisted compounds from molecular docking. The simulations ran for 250 ns using GROMACS 5.1.4 software[40]. This study aimed to analyze the stability of the docked protein-hit complexes. The force fields for hit molecules were generated in Antechamber[41] using the ACPYPE[42], while protein force fields were generated by AMBER ff99SB[43]. The complexes were solvated in a cubic box with a simple point charge (SPC)[44] water molecules and neutralized by adding Na^+ and Cl^- ions at a concentration of 0.15 M NaCl. To relieve the system of short-range bad contacts, energy minimization was performed using the steepest descent method for 50,000 steps. The system was then equilibrated in an NVT step until it reached 300 K, followed by equilibration in an NPT step to attain a suitable density (1 atm, 300 K). Both equilibration steps were performed for a 100 ps timescale. The V-rescale thermostat[45] method was used to maintain the temperature, while the Parinello-Rahman method[46] controlled the pressure. The Particle Mesh Ewald[47] method with a PME order of 6 and a relative tolerance of 10^{-6} was used to handle long-range electrostatics. Short-range interactions were evaluated using a neighbor list of 10 Å.

The Lennard–Jones and real space electrostatic interactions were cut off at 9 Å, and hydrogen bonds were constrained using the LINCS algorithm[48]. The final models of all complexes were produced by averaging snapshots from the MD simulation trajectories. To assess conformational changes in the apo and ligand bound states of the complexes, RMSD and RMSF for Cα atoms were calculated. The convergence of the MD simulations was also evaluated by analyzing RMSD. Protein conformational clustering is an effective method that can be carried out on the simulation data to detect the structural changes in the system ensemble[49]. The RMSD clustering analysis was performed to evaluate atomic heterogeneity in the protein conformers from simulation. The RMSD conformational clustering was executed by the Gromos algorithm[50] and the gmx cluster module in GROMACS was employed to group similar conformations based on structural similarity. In this work we adopted a Cα RMSD cut-off of 0.15 nm to obtain clusters. The clustering analysis enabled identification of dominant conformational states and dynamics of all the molecular systems. The CPPTRAJ module in Amber was used to monitor the hydrogen bonds formed during the trajectory. The hydrogen bond donors and hydrogen bond acceptors are specified with the "hbond" command as masks in CPPTRAJ. The standard convention is followed to define a hydrogen bond donor consisting of a heavy atom and hydrogen while hydrogen bond acceptor contains a single atom. The presence of a hydrogen bond is determined using the distance and angle criterion to give distance, angle of each hydrogen bond and average occupancy as the output[51]. The best molecules identified after post-MD analyses and ADMET prediction were replicated for reproducibility of MD results (Fig. S7).

Mechanical stiffness and normal mode analysis

The MechStiff application of ProDy[52], a Python package to study protein dynamics by structure-based analysis was employed for evaluating the mechanical stiffness in PI5P4K2C protein. The MechStiff tool measures the mechanical resistance in response to an external force applied at specific positions on the 3D structure. The input used in MechStiff can be a known modeled structure as an elastic network for visualization and analysis of anisotropic fluctuations and conformational dynamics. The rigidity in the protein structure is calculated for each residue, averaged over all pairs is represented in the form of a 2D map. The force constant measured as effective stiffness or resistance in response to all possible pulling directions is also determined from the calculation. The information obtained from these plots can help to identify the anisotropic response of protein to external forces. The mechanical stiffness plots were obtained for all the molecular systems using an Anisotropic network model (ANM)[53]. The application of normal modes in the prediction of mobile regions inside the protein has been widely used in research. In particular, the lowest modes are focused upon as they hold information about most movable parts of the protein or slow large amplitude motions. The large motions describe the conformational changes required for protein functioning and biological processes such as recognition of ligands or binding. The information concerning structural variations can be utilized to gain an understanding of biological activity[54,55]. Therefore, the normal mode wizard available in ProDy was used to build an elastic network-based model and for performing ANM calculations to identify flexible regions and large motions in the PI5P4K2C.

Principal component analysis and free energy landscapes

The global or correlated motions of the protein[56] can be determined from PCA. The structural and conformational changes detected from the PCA scatter plots of protein are useful to monitor the motions in the apo structure and protein–ligand complexes during MD simulations. PCA helps in the dimensionality reduction of data while keeping the significant information. Through this analysis additional details like the conformational space accessible to protein can be known, which may not be captured through the RMSD plots. The PCA was crucial to understanding the conformation alterations of KRAS4B[57] due to partner binding and G12C mutation. The

PCA analysis suggested that phosphorylation alters structural fluctuation in the switch domains of KRAS[58]. The projection of 3D coordinates over the first two principal components was done to obtain 2D maps. The 2D scatter plots are representative of protein dynamics during the simulated time through which information about conformational convergence or macrostates explored by protein can be known. Therefore, the 3N × 3N covariance matrix where N signifies the number of atoms was generated from the 3D coordinates. Eigen values were obtained by the diagonalization of the matrix and principal components were the corresponding eigenvectors. MODE-TASK[59] tools were utilized to carry out PCA for studying the overall motion of PI5P4K2C during the 250 ns MD simulations trajectories in all the molecular systems.

FEL analysis was performed for all the molecular complexes from 250 ns MD simulation trajectories to understand the essential movement of PI5P4K2C protein in apo state and in hit molecules bound complexes. The GROMACS 'gmx covar', 'gmx anaeig' and 'gmx sham' modules were employed to construct FEL from the principal components (PC1 and PC2). The distribution of stable conformations and structural states were analysed from these landscapes. The following equation was used to calculate FEL:

$\Delta G(PC1,PC2) = -k_B T \ln P(PC1,PC2)$.

Here PC1 and PC2 represent the reaction coordinates, Boltzmann constant is represented by k_B, T stands for temperature of the system, and P(PC1,PC2) represents the probability distribution.

Carcinogenicity prediction study

Evaluation of the proposed hit molecules for toxicity remains the most important step in the drug development process. Cancer-causing chemicals known as carcinogens may result in serious health effects. Therefore, to prevent drug-induced cancer carcinogenicity tests must be performed for the new compounds before approving their curability. An online carcinogenicity prediction server, CarcinoPred-EL[60] (Carcinogenicity Prediction using Ensemble learning methods) was used to test the studied molecules for carcinogenicity. The server that employs three different classification models such as Ensemble SVM, Ensemble RF, and Ensemble XGBoost was used for the estimation of chemical compounds to be carcinogenic or non-carcinogenic.

Bioactivity prediction

The inhibitory potential of the hit molecules that formed stable protein–ligand complexes was examined through the prediction of bioactivities using an online server called Molinspiration Chemoinformatics available at https://www.molinspiration.com [Molinspiration Cheminformatics free web services, Slovensky Grob, Slovakia]. The molecules were drawn in the molecule sketcher. The predict bioactivity tool allowed prediction of the activity of ligands to different receptors such as ion channel modulator, protease inhibitor, kinase inhibitor, etc. with respective Molinspiration bioactivity scores. The predicted bioactivity scores of reference and hit molecules were recorded to rank them in the order of kinase inhibitory potential.

Binding free energy calculations

The molecular processes including molecular associations, chemical reactions, and protein folding are all driven by free energy. Moreover, the binding affinity of ligands in the protein–ligand complex can be assessed from the binding free energy calculations. The MM-PBSA[61] is a widely used method for measuring the binding free energy of protein–ligand complexes and plays an important role in drug discovery. The g_mmpbsa tool computes the free energy of the system in three consecutive steps. The first step involves the calculation of potential energy in a vacuum followed by the computation of polar and non-polar solvation energies of the system in the next steps.

The expression of $\Delta G_{binding}$ in a protein–ligand complex is computed by the following equation[62,63]

$$\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand});$$

$G_{complex}$ represents the total free energy of the protein–ligand complex,

$G_{protein}$ and G_{ligand} are the total free energies of protein and ligand in the solvent in isolated forms.

Thus g_mmpbsa tool was utilized to compute binding free energies for protein–ligand complexes comprising the reference molecule DVF and hit molecules from MD studies. The output trajectories obtained from MD simulations (250 ns) in GROMACS were employed for evaluating the $\Delta G_{binding}$ values for all molecular systems. The various contributions from electrostatic, vdW, polar solvation energy, and SASA energies (kJ/mol)) were calculated during this step. The residue-wise contribution to the binding free energies was also recorded.

Data availability

The datasets used and/or analysed during the current study are available from the supplementary data.

Received: 16 January 2025; Accepted: 16 June 2025
Published online: 30 September 2025

References

1. Lundquist, M. R. et al. Phosphatidylinositol-5-phosphate 4-kinases regulate cellular lipid metabolism by facilitating autophagy. *Mol. Cell* **70**(3), 531–544. https://doi.org/10.1016/j.molcel.2018.03.037 (2018).
2. De Leo, M. G. et al. Autophagosome–lysosome fusion triggers a lysosomal response mediated by TLR9 and controlled by OCRL. *Nat. Cell Biol.* **18**(8), 839–850. https://doi.org/10.1038/ncb3386 (2016).
3. Jones, D. R. et al. Nuclear PtdIns5P as a transducer of stress signaling: an in vivo role for PIP4Kbeta. *Mol. Cell* **23**(5), 685–695. https://doi.org/10.1016/j.molcel.2006.07.014 (2006).
4. Wilcox, A. & Hinchliffe, K. A. Regulation of extranuclear PtdIns5P production by phosphatidylinositol phosphate 4-kinase 2α. *FEBS Lett.* **582**(9), 1391–1394. https://doi.org/10.1016/j.febslet.2008.03.022 (2008).

5. Bulley, S. J., Clarke, J. H., Droubi, A., Giudici, M. L. & Irvine, R. F. Exploring phosphatidylinositol 5-phosphate 4-kinase function. *Adv. Biol. Reg.* **1**(57), 193–202. https://doi.org/10.1016/j.jbior.2014.09.007 (2015).
6. Clarke, J. H. & Irvine, R. F. The activity, evolution and association of phosphatidylinositol 5-phosphate 4-kinases. *Adv. Biol. Regul.* **52**(1), 40–45. https://doi.org/10.1016/j.advenzreg.2011.09.002 (2012).
7. Bultsma, Y., Keune, W. J. & Divecha, N. PIP4Kβ interacts with and modulates nuclear localization of the high-activity PtdIns5 P-4-kinase isoform PIP4Kα. *Biochem. J.* **430**(2), 223–235. https://doi.org/10.1042/BJ20100341 (2010).
8. Al-Ramahi, I. et al. Inhibition of PIP4Kγ ameliorates the pathological effects of mutant huntingtin protein. *Elife* **6**, e29123. https://doi.org/10.7554/eLife.29123 (2017).
9. Wang, D. G. et al. PIP4Ks suppress insulin signaling through a catalytic-independent mechanism. *Cell Rep.* **27**(7), 1991–2001. https://doi.org/10.1016/j.celrep.2019.04.070 (2019).
10. Sharma, S., Mathre, S., Ramya, V., Shinde, D. & Raghu, P. Phosphatidylinositol 5 phosphate 4-kinase regulates plasma-membrane PIP3 turnover and insulin signaling. *Cell Rep.* **27**(7), 1979–1990. https://doi.org/10.1016/j.celrep.2019.04.084 (2019).
11. Zheng, L. & Conner, S. D. PI5P4Kγ functions in DTX1-mediated Notch signaling. *Proc. Natl. Acad. Sci.* **115**(9), E1983–E1990. https://doi.org/10.1073/pnas.1712142115 (2018).
12. Raychaudhuri, S. et al. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat. Genet.* **40**(10), 1216–1223. https://doi.org/10.1038/ng.233 (2008).
13. Arora, G. K., Palamiuc, L. & Emerling, B. M. Expanding role of PI5P4Ks in cancer: A promising druggable target. *FEBS Lett.* **596**(1), 3–16. https://doi.org/10.1002/1873-3468.14237 (2022).
14. Clarke, J. H. et al. The function of phosphatidylinositol 5-phosphate 4-kinase γ (PI5P4Kγ) explored using a specific inhibitor that targets the PI5P-binding site. *Biochem. J.* **466**(2), 359–367. https://doi.org/10.1042/BJ20141333 (2015).
15. Boffey, H. K. et al. Development of selective phosphatidylinositol 5-phosphate 4-kinase γ inhibitors with a non-ATP-competitive, allosteric binding mode. *J. Med. Chem.* **65**(4), 3359–3370. https://doi.org/10.1021/acs.jmedchem.1c01819 (2022).
16. Tubeleviciute-Aydin, A. et al. Identification of allosteric inhibitors against active caspase-6. *Sci. Rep.* **9**(1), 5504. https://doi.org/10.1038/s41598-019-41930-7 (2019).
17. Rana, N. et al. Targeting allosteric binding site in methylenetetrahydrofolate dehydrogenase 2 (MTHFD2) to identify natural product inhibitors via structure-based computational approach. *Sci. Rep.* **13**(1), 18090. https://doi.org/10.1038/s41598-023-45175-3 (2023).
18. Vettoretti, G. et al. Molecular dynamics simulations reveal the mechanisms of allosteric activation of Hsp90 by designed ligands. *Sci. Rep.* **6**(1), 23830. https://doi.org/10.1038/srep23830 (2016).
19. Schoepfer J, Jahnke W, Berellini G, Buonamici S, Cotesta S, Cowan-Jacob SW, Dodd S, Drueckes P, Fabbro D, Gabriel T, Groell JM. Discovery of asciminib (ABL001), an allosteric inhibitor of the tyrosine kinase activity of BCR-ABL1. https://doi.org/10.1021/acs.jmedchem.8b01040
20. Guruprasad L, Andola P, Banerjee A, Laxman D, Naresh GK. Structure-based methods in drug design. InCheminformatics, QSAR and Machine Learning Applications for Novel Drug Development 2023 Jan 1 (pp. 205–237). Academic Press. https://doi.org/10.1016/B978-0-443-18638-7.00003-7
21. Li, R., He, X., Wu, C., Li, M. & Zhang, J. Advances in structure-based allosteric drug design. *Curr. Opin. Struct. Biol.* **1**(90), 102974. https://doi.org/10.1016/j.sbi.2024.102974 (2025).
22. Zhu, R., Wu, C., Zha, J., Lu, S. & Zhang, J. Decoding allosteric landscapes: computational methodologies for enzyme modulation and drug discovery. *RSC Chem. Biol.* https://doi.org/10.1039/D4CB00282B (2025).
23. He, J. et al. Transformer-based molecular optimization beyond matched molecular pairs. *J. Cheminformatics.* **14**(1), 18. https://doi.org/10.1186/s13321-022-00599-3 (2022).
24. Oguro, Y. et al. Design, synthesis, and evaluation of 5-methyl-4-phenoxy-5H-pyrrolo [3, 2-d] pyrimidine derivatives: novel VEGFR2 kinase inhibitors binding to inactive kinase conformation. *Bioorg. Med. Chem.* **18**(20), 7260–7273. https://doi.org/10.1016/j.bmc.2010.08.017 (2010).
25. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinf.* **9**, 1–8. https://doi.org/10.1186/1471-2105-9-40 (2008).
26. Martin, Y. C., Kofron, J. L. & Traphagen, L. M. Do structurally similar molecules have similar biological activity?. *J. Med. Chem.* **45**(19), 4350–4358. https://doi.org/10.1021/jm020155c (2002).
27. MacMahon, M. et al. An in silico drug repurposing pipeline to identify drugs with the potential to inhibit SARS-CoV-2 replication. *Inf. Med. Unlocked.* **1**(43), 101387. https://doi.org/10.1016/j.imu.2023.101387 (2023).
28. Zoete V, Daina A, Bovigny C, Michielin O. SwissSimilarity: a web tool for low to ultra high throughput ligand-based virtual screening. https://doi.org/10.1021/acs.jcim.6b00174
29. Kuhn, M., von Mering, C., Campillos, M., Jensen, L. J. & Bork, P. STITCH: Interaction networks of chemicals and proteins. *Nucl. Acids Res.* **36**(1), 684–688. https://doi.org/10.1093/nar/gkm795 (2007).
30. Wirth, M., Zoete, V., Michielin, O. & Sauer, W. H. SwissBioisostere: a database of molecular replacements for ligand design. *Nucleic Acids Res.* **41**(D1), D1137–D1143. https://doi.org/10.1093/nar/gks1059 (2013).
31. Ghamari, N. et al. Histamine H3 receptor ligands by hybrid virtual screening, docking, molecular dynamics simulations, and investigation of their biological effects. *Chem. Biol. Drug Des.* **93**(5), 832–843. https://doi.org/10.1111/cbdd.13471 (2019).
32. Najjar, A. et al. Computer-aided design, synthesis and biological characterization of novel inhibitors for PKMYT1. *Eur. J. Med. Chem.* **1**(161), 479–492. https://doi.org/10.1016/j.ejmech.2018.10.050 (2019).
33. Zdrazil, B. et al. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* **52**(D1), D1180–D1192. https://doi.org/10.1093/nar/gkad1004 (2024).
34. Irwin, J. J. et al. ZINC20—a free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* **60**(12), 6065–6073. https://doi.org/10.1021/acs.jcim.0c00675 (2020).
35. Bemis, G. W. & Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **39**(15), 2887–2893. https://doi.org/10.1021/jm9602928 (1996).
36. Daina, A., Michielin, O. & Zoete, V. iLOGP: a simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach. *J. Chem. Inf. Model.* **54**(12), 3284–3301. https://doi.org/10.1021/ci500467k (2014).
37. Daina, A., Michielin, O. & Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **7**(1), 42717. https://doi.org/10.1038/srep42717 (2017).
38. Dallakyan, S. & Olson, A. J. Small-molecule library screening by docking with PyRx. *Chem. Biol. Methods Protocols.* https://doi.org/10.1007/978-1-4939-2269-7_19 (2015).
39. Morris, G. M. et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**(16), 2785–2791. https://doi.org/10.1002/jcc.21256 (2009).
40. Hess, B., Kutzner, C., Van Der Spoel, D. & Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**(3), 435–447. https://doi.org/10.1021/ct700301q (2008).
41. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**(2), 247–260. https://doi.org/10.1016/j.jmgm.2005.12.005 (2006).
42. Sousa da Silva, A. W. & Vranken, W. F. ACPYPE-Antechamber python parser interface. *BMC Res. Notes.* **5**, 1–8. https://doi.org/10.1186/1756-0500-5-367 (2012).
43. Hornak, V. et al. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Prot. Struct. Funct. Bioinf.* **65**(3), 712–725. https://doi.org/10.1002/prot.21123 (2006).

44. Berendsen, H. J., Postma, J. P., van Gunsteren, W. F., & Hermans, J. (1981). Interaction models for water in relation to protein hydration. In Intermolecular forces (pp. 331–342). Springer, Dordrecht. https://doi.org/10.1007/978-94-015-7658-1_21
45. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**(1), 014101. https://doi.org/10.1063/1.2408420 (2007).
46. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**(12), 7182–7190. https://doi.org/10.1063/1.328693 (1981).
47. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**(12), 10089–10092. https://doi.org/10.1063/1.464397 (1993).
48. Hess, B., Bekker, H., Berendsen, H. J. & Fraaije, J. G. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**(12), 1463–1472 (1997).
49. Purohit, R. Role of ELA region in auto-activation of mutant KIT receptor: a molecular dynamics simulation insight. *J. Biomol. Struct. Dyn.* **32**(7), 1033–1046. https://doi.org/10.1080/07391102.2013.803264 (2014).
50. Daura, X. et al. Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed.* **38**(1–2), 236–240. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2%3c236::AID-ANIE236%3e3.0.CO;2-M (1999).
51. Roe, D. R. & Cheatham, T. E. III. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**(7), 3084–3095. https://doi.org/10.1021/ct400341p (2013).
52. Bakan, A., Meireles, L. M. & Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27**(11), 1575–1577. https://doi.org/10.1093/bioinformatics/btr168 (2011).
53. Eyal, E., Yang, L. W. & Bahar, I. Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* **22**(21), 2619–2627. https://doi.org/10.1093/bioinformatics/btl448 (2006).
54. Bakan, A. & Bahar, I. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proc. Natl. Acad. Sci.* **106**(34), 14349–14354. https://doi.org/10.1073/pnas.0904214106 (2009).
55. Yang, L., Song, G., Carriquiry, A. & Jernigan, R. L. Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure.* **16**(2), 321–330. https://doi.org/10.1016/j.str.2007.12.011 (2008).
56. Maisuradze, G. G., Liwo, A. & Scheraga, H. A. Principal component analysis for protein folding dynamics. *J. Mol. Biol.* **385**(1), 312–329. https://doi.org/10.1016/j.jmb.2008.10.018 (2009).
57. Chen, J., Wang, J., Yang, W., Zhao, L. & Hu, G. Conformations of KRAS4B affected by its partner binding and G12C mutation: insights from GaMD trajectory-image transformation-based deep learning. *J. Chem. Inf. Model.* **64**(17), 6880–6898. https://doi.org/10.1021/acs.jcim.4c01174 (2024).
58. Chen, J. et al. Molecular mechanism of phosphorylation-mediated impacts on the conformation dynamics of GTP-Bound KRAS probed by GaMD trajectory-based deep learning. *Molecules* **29**(10), 2317. https://doi.org/10.3390/molecules29102317 (2024).
59. Ross, C. et al. MODE-TASK: large-scale protein motion tools. *Bioinformatics* **34**(21), 3759–3763. https://doi.org/10.1093/bioinformatics/bty427 (2018).
60. Zhang, L. et al. CarcinoPred-EL: Novel models for predicting the carcinogenicity of chemicals using molecular fingerprints and ensemble learning methods. *Sci. Rep.* **7**(1), 2118. https://doi.org/10.1038/s41598-017-02365-0 (2017).
61. Kumari, R., Kumar, R. & Lynn, A. A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* **54**(7), 1951–1962. https://doi.org/10.1021/ci500020m (2014).
62. Kollman, P. A. et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**(12), 889–897. https://doi.org/10.1021/ar000033j (2000).
63. Gilson, M. K. & Honig, B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins: Struct., Funct., Bioinf.* **4**(1), 7–18. https://doi.org/10.1002/prot.340040104 (1988).

Acknowledgements

P.A. thanks UGC, New Delhi for research fellowship, J.P. thanks University of Hyderabad for UGC Non-NET fellowship and IoE PhD fellowship, G.K.R.S.N. thanks UGC Non-NET fellowship. The authors thanks CMSD, University of Hyderabad for computational facilities. The authors thank DST-PURSE and UGC UPE2 and IOE for funding and CMSD for computational facilities.

Author contributions

P.A. carried out the research work, methodologies and wrote the manuscript. J.P. assisted in carrying out some methodologies. G.K.R.S.N. provided skill support and discussed the methods. L.G. conceived the idea, supervised and wrote the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-07480-x.

Correspondence and requests for materials should be addressed to L.G.