ACCSCIENCE PUBLISHING

INNOSC Theranostics and Pharmacological Sciences

ORIGINAL RESEARCH ARTICLE

Rational drug design from phosphatidylinositol 3-kinase-α inhibitors through molecular docking and 3D-QSAR methodologies for cancer immunotherapy

Kevin Tochukwu Dibia[1†*], Sandra Nneka Van-Dibia[2†], and Philomena Kanwulia Igbokwe[1]

[1]Department of Chemical Engineering, Faculty of Engineering, Nnamdi Azikiwe, University, Awka, Anambra, Nigeria

[2]Department of Animal Physiology, College of Animal Science and Livestock Production, Federal University of Agriculture, Abeokuta, Ogun, Nigeria

Abstract

Dysregulation or aberrant activation of the phosphatidylinositol 3-kinase (PI3K) signaling pathway is commonly observed in various cancers and is associated with tumor growth, metastasis, and resistance to therapy. Targeting PI3K-α with appropriate inhibitors can disrupt this pathway, hindering cancer progression, and potentially enhancing the immune system's ability to recognize and eliminate cancer cells. In this study, we aimed to design a novel and potent inhibitor of PI3K-α for cancer immunotherapy using rational drug design techniques, including virtual screening, molecular docking, and 3D-QSAR. We obtained the human PI3K-α protein (6PYS) complexed with (3S)-3-benzyl-3-methyl-5-[5-(2-methylpyrimidin-5-yl)pyrazolo[1,5-a] pyrimidin-3-yl]-1,3-dihydro-2H-indol-2-one (PJ5) from the RCSB Protein Data Bank. Virtual screening of ligands, integrated with predictive computational molecular docking and 3D-field-based-QSAR, was implemented using appropriate Schrödinger Maestro modules. Rational drug design was also carried out, and its clinical relevance was validated across several ADMET descriptors. Docking results suggested that a hybrid of sulfonamide and pyridine-based heterocyclic compounds, functionalized with potent moieties derived from alkaloids, exhibited adequate synergistic biological effects capable of enhancing sufficient biological activity against PI3K-α. A field-based 3D-QSAR model was built on four partial least squares factors, and five statistical metrics were employed to validate the model. The newly designed ligand from this approach, named 6'-amino-5'-(2-fluoro-1,3-oxazol-5-yl)-N-{[3-(hydroxymethyl)oxetan-3-yl]methyl}-3-methyl-[2,3'-bipyridine]-6-sulfonamide or T85, exhibited a predicted bioactivity (pIC_{50}) of 8.25. The predicted ADMET properties of T85 fell reasonably within the range of recommended standards, especially adhering to Lipinski's rule of five and Jorgensen's rule of three. In conclusion, the results of this study offer significant insights into *in silico* drug design using a rational approach, which could expedite the discovery and development of new drug molecules.

Keywords: Cancer; PI3K-α; Molecular docking; 3D-QSAR; ADMET

†These authors contributed equally to the work.

***Corresponding author:**
Kevin Tochukwu Dibia
(tkevin.dibia@gmail.com)

Citation: Dibia KT, Van-Dibia SN, Igbokwe PK. Rational drug design from phosphatidylinositol 3-kinase-α inhibitors through molecular docking and 3D-QSAR methodologies for cancer immunotherapy. *INNOSC Theranostics and Pharmacological Sciences*. 2024;7(2):2340. doi: 10.36922/itps.2340

Received: November 30, 2023

Accepted: February 1, 2024

Published Online: April 15, 2024

Copyright: © 2024 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Human physiology is governed by intricate cellular metabolisms, which exemplify how cell-specific programs efficiently regulate overall health and well-being. The body initiates cascade mechanisms of cellular signaling pathways to control physiological functions such as cell proliferation, growth, survival, differentiation, and metabolism.[1] Cells serve as the building blocks of life; however, genetic mutations or alterations can lead to anomalies, disrupting the mitotic and cytokinetic regulatory mechanisms of the cells. Over time, dysregulation of cellular signaling progresses to carcinogenic effects, affecting various enzyme functions, such as phosphoinositide 3-kinase alpha (PI3K-*α*), an isoform of the phosphatidylinositol-3-kinase (PI3K) family of enzymes.

PI3Ks are lipid kinases responsible for phosphorylating the -OH moiety at the 3' position of the inositol ring, generating phosphatidylinositol-3,4,5-triphosphate (PIP3).[2,3] They play a central role in regulating cell cycle, apoptosis, DNA repair, cellular senescence, angiogenesis, cellular metabolism, and motility. Besides, they serve as intermediate signaling molecules that are well known for their roles in the PI3K/serine-threonine protein kinase (AKT)/mammalian target of rapamycin (mTOR) (PI3K/AKT/mTOR) signaling pathway.[2,4-6] Furthermore, the PI3K/AKT/mTOR pathway is involved in oncogenesis and tumorigenesis,[5,7,8] which is frequently dysregulated in human cancers, such as pancreatic cancer,[2] breast cancer,[5] colorectal, and ovarian cancer.[6] At the molecular level, alterations at numerous nodes are described in different tumor types, including activating mutations and/or the amplification of phosphatidylinositol-4,5-biphosphate 3-kinase catalytic subunit alpha (PIK3CA), mutation or overexpression of upstream receptor tyrosine kinases (RTKs), or inactivating mutations or deletions of phosphatase and tensin homolog (*PTEN*) from chromosome 10.[4,7]

In the recent decade, several studies have described the role of the PI3K pathway,[2-5,9,10] which is a complex cascade of signal transduction that regulates proliferation, growth, differentiation, protein synthesis, glucose metabolism, migration, apoptosis, and other intracellular metabolisms activated in diverse types of cancer.[3,9] In humans, PI3K exists in three classes – Classes I, II, and III,[3,4] and is enzymes with a molecular weight of about 200 – 300 kDa.[4] Class I of the PI3K subfamily is primarily linked to oncogenesis.[4,11-13] The PI3K is a heterodimer comprised both the catalytic and regulatory subunits. The phosphorylation of the regulatory subunit leads to the activation of p110, the catalytic subunit of PI3K. The genes *PIK3CA, PIK3CB, PIK3CG,* and *PIK3CD* encode the four isotypes of p110: alpha, beta, gamma, and delta, respectively.[14] These p110 isotypes activate downstream signaling pathways, including binding corresponding ligands to tyrosine kinase receptors (RKTs).[9] However, the p110-alpha isotype is involved in insulin-like signaling, p110-beta plays a role in platelet-derived aggregation, thrombosis, and insulin signaling, while p110-gamma and p110-delta isotypes are expressed in lymphocyte activation, mast cell degranulation, and chemotaxis.[15] Furthermore, the catalytic p110 subunit forms a complex with one of the three p85-related regulatory subunits: p85-alpha, p85-beta, and p55-gamma,[14] encoded by the genes *PIK3R1, PIK3R2,* and *PIK3R3*, respectively.[14,16,17] Activation of the p85 regulatory subunit occurs through the stimulation of the receptor tyrosine kinase, where it binds to phosphotyrosine residues in the receptor tyrosine kinase. Such a process unleashes the p110 catalytic subunit from inhibition of the p85 regulatory subunit.[3] Progressively, PI3K localizes to the cellular membrane, where the p110 subunit converts phosphatidylinositol 4,5-biphosphate (PIP2) into PIP3, a lipid secondary messenger, which then activates the downstream effector protein kinase B (also known as AKT), promoting its translocation to the inner membrane of cells, where it is phosphorylated and activated by phosphoinositide-dependent protein kinase (PDK)-1, PDK2, and mTor-rictor.[3,5] AKT, a serine/threonine kinase, together with mammalian target of rapamycin complex 1 (mTORC1), further modulates the activities of downstream biomacromolecules, such as B-cell lymphoma-2 (BCL-2), BCL-2 antagonist of cell death (BAD), forkhead box O (FOXO), p53, p27, transcription factor, tuberous sclerosis complex 2 (TSC2), glycogen synthase kinase-3β (GSK3β), insulin-like growth factors, cyclin D1, C-MYC (cellular myelocytomatosis) oncogene, nuclear factor kappa B (NF-κB), caspase-3 and caspase-9, and murine double minute 2 (MDM2) (Figure 1).[3,5] These biomacromolecules regulate protein synthesis, cell survival, cell cycle progression, cellular growth, proliferation, motility transformation, DNA repair, glucose metabolism, and drug resistance.[3,18,19]

PI3K classes II and III also play significant roles in cell-specific functions and metabolism. There are three isoforms of PI3K class II, including PI3KC2-alpha, PI3KC2-beta, and PI3KC2-gamma.[3-5] These isoforms of PI3K class II are monomers with a high molecular weight.[16,20] Moreover, class II monomers have no regulatory subunits[3,5] but possess individual catalytic moieties that interact directly with phosphorylated adapter proteins.[16,20] In contrast to PI3KC2-beta, PI3KC2-alpha appears to be involved in cell migration and neuronal cell survival, as well as clathrin-mediated vesicle trafficking, insulin signaling, neurosecretory granular exocytosis, and smooth muscle contraction.[15] On the other hand, the role of PI3KC2 is

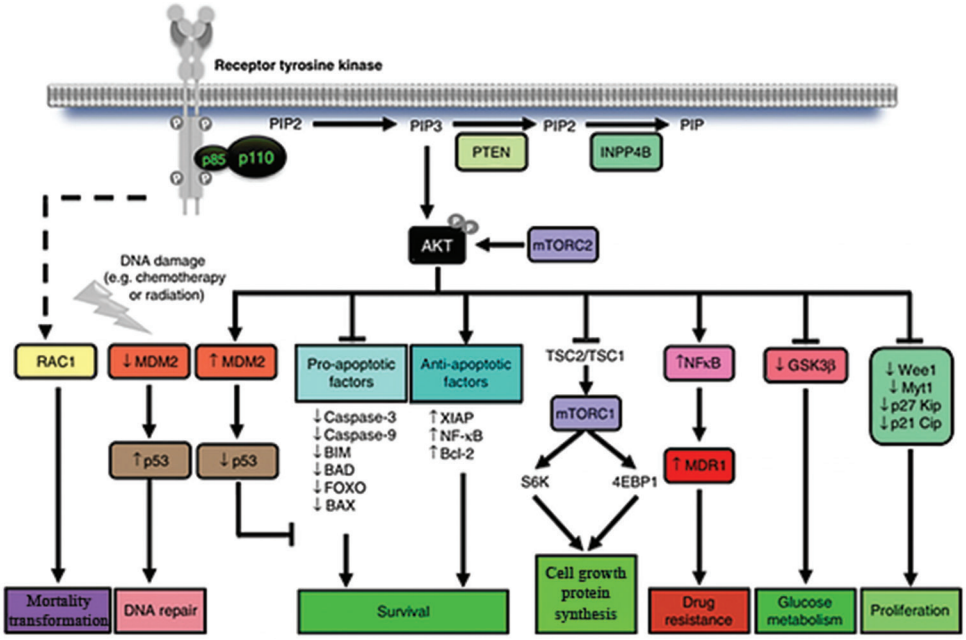


Figure 1. PI3K/AKT/mTOR signaling pathway in human cancer. Exposure of cells to DNA damage reduces intracellular proportions of MDM2, which enables p53 accumulation and stabilization, leading to the initiation of DNA repair pathways.[5]
Abbreviations: BAX: BCL-2-associated X protein; BIM: BCL-2-interacting mediator of cell death; INPP4B: Inositol polyphosphate-4-phosphatase type II B; MDR1: Multidrug resistance protein 1; Myt1: Myt1 kinase; S6K: Ribosomal protein S6 kinase; Wee 1: Wee 1 kinase; XIAP: X-linked inhibitor of apoptosis protein; 4EBP1: Eukaryotic translation initiation factor 4E-binding protein 1.

not established. The PI3KC3 gene produces the class III PI3K, also known as vacuolar protein sorting 34, which is a potential serine/threonine protein tyrosine kinase regulatory subunit. Vacuolar protein sorting 34 is also linked to mTOR signaling and plays a part in the vesicular transport of membrane proteins to the lysosome.[15] Furthermore, the PI3K/AKT/mTOR pathway plays a critical role in the pathogenesis and provides survival advantages in hematologic malignancies such as leukemia, lymphoma, and myeloma.[3]

PI3K-α is a member of the Class 1A PI3K family of enzymes and is activated by various upstream signaling biomacromolecules, such as growth factor receptors, which phosphorylate phosphatidylinositols in the cell membrane. It is linked to several tumorigeneses through amplification, overexpression, and mutation of the PIK3CA gene,[21] following the PI3K/AKT/mTOR pathway. In addition, PI3K-α specifically consists of heterodimers of a p110-α catalytic subunit and a p85 regulatory subunit.[22] The PI3K-α isoform transmits various extracellular stimuli through signaling pathways that regulate numerous cellular processes, including cell proliferation, motility, cell death, and cell invasion, thereby playing a crucial role in the physiology of cells.[23,24]

Recently, it has been demonstrated that selective inactivation of the PI3K-α isoform, which is the most frequently up-regulated isoform of PI3K in human cancer, effectively blocks PI3K/AKT/mTOR signaling in response to diverse growth stimuli.[25] Consequently, PI3K-α has emerged as a key target that is primarily affected by cancer mutations, gene rearrangement, and gene amplification, making it an attractive focus for drug development.[23,26] Ongoing efforts aim to exploit and develop novel selective inhibitors for PI3K-α as a promising drug target for anti-inflammation and anti-cancer therapy. Moreover, several kinase inhibitors for PI3K-α have been discovered and synthesized to regulate the proliferation of cancer cells. These small-molecule drugs exhibit unique potency. Notable examples include buparlisib (BKM120), pilaralisib (XL147), pictilisib (GDC-0941),[27] alpelisib (BYL719), umbralisib,[28] and taselisib (GCD-0032),[4,24] serving as PI3K isoform inhibitors utilized in the treatment of pancreatic cancer,[2] breast cancer,[2,5,29] or ovarian cancer.[6] Despite their development, most inhibitors encountered challenges in clinical trials, displaying poor efficacies due to nanotherapeutic-associated toxicity resulting from on-target and off-target effects.[21,24] To maximize effectiveness and minimize negative side effects, medicinal chemists focus on identifying isoform-selective PI3K-α inhibitors.[21] This discovery of isoform-selective PI3K-α inhibitors remains a priority to engineer therapeutic drugs with enhanced efficacy and reduced side effects[21]

through various scientific approaches. Furthermore, various essential metabolic pathways and key factors serve as a pool of associated therapeutic targets for anticancer therapy development.[30]

The drug discovery process is known for its time-consuming nature and high costs.[28] However, advancements in molecular biology, genomics, and computational technologies have accelerated the understanding of cancer development, leading to the discovery of novel biomarkers, targeted therapies, and immunotherapies. These advancements have significantly improved cancer diagnosis and treatment outcomes. Three-dimensional quantitative structure-activity relationship (3D-QSAR) in combination with molecular dynamics simulations is a hybrid *in silico* approach for the design and synthesis of drugs and is instrumental in identifying new compounds with superior activity.[31,32] The combination of 3D-QSAR and molecular docking approaches continues to demonstrate how *in silico* pharmacological systems can work together to produce novel therapeutics using structural data. The hybrid offers a more comprehensive insight into the structure-activity relationship of a compound (ligand)[33] and its interaction with the target protein. This understanding enables drug designers to modify the compound's structure to enhance its activity, selectivity, and binding affinity, ultimately leading to a rational design of more potent and effective drugs.

In the present study, the drug discovery process using PI3K-*α* inhibitors involved ligand preparation based on predefined descriptors by assessing the drug-like properties of the ligands and assessing their binding affinity to the target through molecular docking. In addition, the study considered pharmacophore screening with the least energy core to prioritize the binding of ligands to the active site of PI3K-*α*, predicting the most energetically favorable conformations and estimating the binding affinities. Moreover, the study included the prediction of absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiles. These approaches aided in the rational design and optimization of PI3K-*α* inhibitors, contributing to the development of potential therapeutic agents for cancer and other diseases.

Current preclinical and clinical evidence suggests that inhibitors targeting the PI3K/AKT/mTOR pathway are utilized in combination with other anticancer therapies to combat resistance in cancer cells. Multiple ongoing clinical studies are investigating this approach. However, most targeted anticancer therapies, as well as cytotoxic and radiation therapies, are complicated by secondary resistance in cancer cells. Resistance is an intricate phenomenon involving numerous mechanisms, comprising the activation of signaling pathways such as PI3K/AKT/mTOR.[5] Drug discovery and development represent a lengthy and costly process, relying heavily on *in vitro* assays, animal models, and clinical trials for reliable testing of drug molecules. Hence, our objective is to contribute to the discovery of a novel PI3K-*α* inhibitory compound through a rational drug design approach with potential immunomodulatory, immunobiological, and clinical implications. We aim to explore a series of clinically approved selective PI3K-*α* inhibitors with excellent pharmacokinetic properties obtained from the public web-accessible molecular recognition database, BindingDB (http://www.bindingdb.org).[34,35] Rational drug design holds the potential to discover novel drugs or drug combinations and repurpose existing drugs for new indications. This work seeks to introduce innovative methods for discovering potential drug molecules through molecular docking and 3D-QSAR methods, coupled with robust rational drug design techniques focused on selected PI3K-*α* inhibitors.

2. Methods

2.1. Dataset

The dataset, comprised ligands, was retrieved from BindingDB. These cogeneric ligands were represented in three-dimensional (3D) coordinate structures, all sharing common orientation codes. Initially, this exploratory dataset contained 3994 rows of different inhibitory molecules and 56 columns, providing information on the inhibitory molecules targeting the phosphatidylinositol-3 kinase regulatory subunit alpha in *Homo sapiens* (human). Furthermore, the dataset contained some important column attributes, such as the half-maximal inhibition concentration (IC_{50}), BindingDB ligand name, UniProt (SwissProt) primary and secondary IDs of the target chain, molecule ID, and ROMol information. The UniProt serves as a focal point for collecting functional information on proteins with precise, dependable, and extensive annotations.[33] It integrates biologically significant data obtained from selected resources and the manual curation of protein features, such as functional domains and active sites, amino acid variations, ligand binding sites, and post-translational modifications (PTMs). UniProt records provide mechanistic insights into disease-drug relationships.[36] On the other hand, ROMol (the Read-Only molecule) is a representation of a molecule or a chemical structure that is strictly read-only. It is an object within RDKit, an open-source cheminformatics library written in C++ with Python bindings, enabling operations with chemical structures and data.

2.2. Data preprocessing

Data preprocessing is a crucial step in preparing and transforming datasets to improve raw data quality,

considering the inherent complexity and imperfections in data preparation operations.[37,38] It serves as a basis for valid data analysis.[37] Preprocessing includes various techniques such as cleaning, integration, transformation, imputation of missing values, and reduction.[33,38]

In this study, lists of PI3K-α inhibitory molecules were obtained from the binding databank, resulting in a dataset comprising 3994 inhibitory molecules in 3D geometry. Furthermore, the dataset included columns containing IC_{50} values of the molecules, molecule IDs, ligand names, ROMol object information of ligands, etc. The IC_{50} values column contained affinity information, indicating the potency of each molecule against the PI3K-α target. The dataset was in SDF (structure-data file) format, and data preprocessing was performed using the Python programming language.

The IC_{50} values of compounds, expressed in nanomolar (nM) units and ranging from 0.07 – 7200 nM, helped capture a broader chemical space, enhancing the identification of novel ligands. In addition, the IC_{50} column was also used as a reference column, in which duplicate rows sharing the same IC_{50} were dropped. The governing code syntax was specific to maintaining the first entries, as it was assumed that two or more ligands with the same IC_{50} value exhibited similar potency or affinity, pharmacological effects, and functional activities toward the target protein or receptor. Dropping duplicate entries of IC_{50} values offered a normal distribution of values that made the dataset more amenable to statistical analysis. However, docking ligands of similar half-maximal inhibitory concentrations may not provide significant additional insight. Later, the IC_{50} values were converted to pIC_{50} values to enable dataset standardization and consistency.

The IC_{50} values depicted in multiple units can complicate the analysis of results across different concentrations. Hence, it was necessary to convert the IC_{50} values in the dataset to pIC_{50} values. Data presentation in pIC_{50} values, which represent the values as the negative logarithm of the molar concentration of the IC_{50} values, is considered a better approach. This method enhances data clarity, minimizes potential errors in data representation, and improves reproducibility with standardization, linearity, normal distribution, and precision as additional attributes. Relevant columns were selected and preserved for further analysis. The data preprocessing stage functions as a preliminary filtering technique to minimize the compound selection size before executing virtual screening campaigns.

2.3. Protein complex refinement

The PI3K-α protein structure was obtained from the RCSB Protein Data Bank (rcsb.org). The architecture of the human PI3K-α protein, encoded as 6PYS, is a protein complex composed of a ligand and several water molecules. The structure of 6PYS, obtained through X-ray diffraction, exhibits a resolution of 2.19Å, with associated R-values of free, work, and observed, numerically presented as 0.259, 0.2243, and 0.225, respectively. The composition of 6PYS includes a total structural weight of 110.61 kDa, an atom count of 7558, modeled residue counts of 890, deposited residue counts of 945, and one unique protein chain A. Furthermore, no mutations were associated with the 6PYS polymer sequence that was engineered from the reference sequence.

The protein preparation involved isolating the ligand from the 6PYS protein-ligand complex, followed by protein content modification using the protein preparation and refinement wizard embedded in Schrödinger Maestro (Schrödinger Release 2020-3: Maestro, Schrödinger, LLC, United States, 2023). The Maestro software is an intuitive molecular modeling environment for various scientific discoveries based on material science, as well as an integrated predictive computational modeling and machine-learning platform for small-molecule drug development. During refinement, simulation settings were configured for a pH of 7.0, which allowed small molecules (HETs) to detect ligands, metals, and ions. In addition, the refinement process incorporated various measures, such as the assignment of bond orders the employment of a chemical component dictionary (CCD) database to help identify and characterize the ligand present in the protein structure in connection with its binding modes and their potential functional or therapeutic roles; inclusion of missing hydrogens in the protein; the addition of terminal oxygens to the protein; the conversion of selenomethionines to methionines; the filling of missing loops; cap termini; the deletion of water molecules beyond HETs of 0 Å; and the generation of HET state within 7.0 ± 2.0 pH value. The Kabat antibody annotation scheme was employed to facilitate the design and analysis of antibody-based therapeutics by comparing the protein sequences and structures of antibodies. Furthermore, to mimic the natural environment of the protein and prevent unwanted interactions or structural distortions that may arise from exposed termini, the termini of the protein were capped with small fragments of peptides.

Hydrogen bond assignment was carried out in the refinement stage to assign hydrogen bonds to the right geometry. The optimization of the hydrogen bond assignment scheme was carried out using PROPKA, a molecular dynamics program in Maestro that facilitated a quantitative analysis of the protein pKa values of ionizable groups. More specifically, PROPKA was utilized

in identifying the interactions of ionizable residues in the protein together with their structural determinants along with significantly perturbed pKa values that contributed to the stability of the protein. The last step of protein refinement was energy minimization and the deletion of water molecules around the ligand. In this step, the root mean square deviation (RMSD) for heavy atom coverage was set to 0.30 Å. This was carried out while pushing the minimization using the OPLS3 (optimized potentials for liquid simulations-3) force field and concurrently deleting water molecules within 3 Å from the ligand (HETs). The primary objective of protein minimization was to remove steric clashes,[39] reorganize structural features such as bond angles, bond lengths, and torsional angles, and remove strained conformations arising from crystal packing artifacts, experimental errors, or inaccuracies in computational modeling.[40]

2.4. Binding site/pocket identification

With a large amount of protein data available in the UniProt database, only a fraction of it is functionally annotated.[41] Consequently, understanding the biological or biochemical role of protein with an interacting partner in a binding relationship remains a fundamental challenge in the fields of medicinal chemistry, genomic bioinformatics, and pharmacology. This underscores the importance of automated sequencing tools. Protein function is closely correlated with the small molecules that attach to them. These small molecules may function as substrates or products of an enzyme reaction, cofactors that are crucial for catalysis,[41,42] or they may provide crucial structural or regulatory functions.[43] Hence, protein-ligand interaction in the context of binding sites is critical for drug discovery,[44] and understanding responses to drugs.[45]

Binding site identification was conducted following a predictive approach using the SiteMap wizard in Maestro. To conduct binding site prediction, the ligand in the 6PYS protein-ligand complex was masked. To find, visualize, and evaluate protein binding sites, the identification of top-ranked potential receptor binding sites was prioritized. This necessitated the search of at least 15 site points per reported site, from which up to five site-point groupings were to be reported. The strategy for identifying binding pockets involved employing a more restrictive definition of hydrophobicity and having a standard grid and crop site maps at 4 Å from the nearest site point. Furthermore, the binding site attributes such as druggable sites, docking space (site score), and desirable ligand size,[46] as well as other parameters including pocket volume, exposure score, contact score, and hydrophobic interaction score, were computed using the SiteMap algorithm. Furthermore, the respective values of these binding site attributes were evaluated to determine the optimal site for binding. The mathematical correlations depicted in Equations I and II were used by Maestro to compute the druggability and site scores, respectively. These procedures culminated in detecting deep binding sites/pockets for potential receptors, specific to a protein-ligand complex.

$$Druggability\ (D)\ score = 0.094\sqrt{n} + 0.60\delta - 0.34\omega \quad (I)$$

$$Site\ score = 0.0733\sqrt{n} + 0.6688\delta - 0.20\omega \quad (II)$$

Where, n is the number of site points (capped at 100); δ is the enclosure score; and ω is the hydrophilic score.

2.5. Ligand refinement

The idea behind ligand refinement is to transform two-dimensional (2D) or three-dimensional (3D) structures into corresponding low-energy 3D structures in the structure-data file (.sdf) format, with the option to expand each input structure by creating variations on ionization state, tautomers, stereochemistry, and ring confirmations, thereby generating broad chemical and structural diversity from a single input structure. In this study, the LigPrep package in Schrödinger Maestro (Schrödinger Release 2020-3: LigPrep, Schrödinger, LLC, United States, 2023) was employed to prepare the ligands.

The 6PYS is a co-crystallized protein with a ligand, P5J (Figure 2). The ligand P5J was isolated from the protein complex and refined using LigPrep in Maestro. The ligand states were generated at pH 7 ± 2.0 using the Hamette and Taft methodology encoded in the

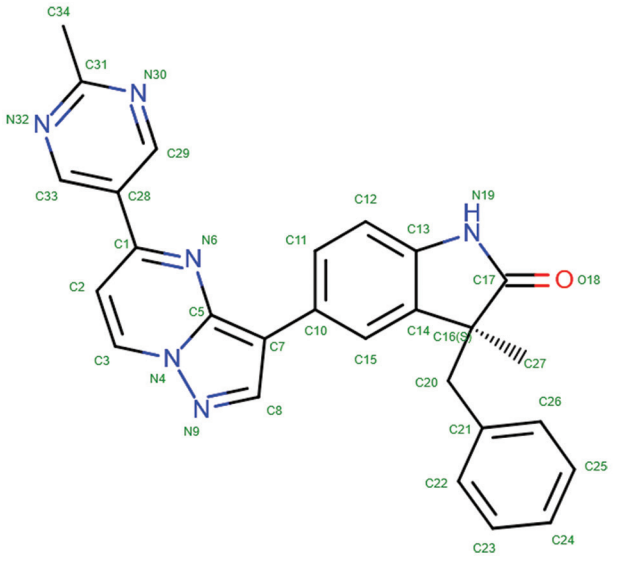


Figure 2. 2D representation of ligand P5J in 6PYS human protein complex. P5J refers to (3S)-3-benzyl-3-methyl-5-[5-(2-methyl pyrimidine-5-yl) pyrazolo[1,5-a]pyrimidin-3-yl]-1,3-dihydro-2H-indol-2-one; chemical structural formula: $C_{27}H_{22}N_6O$; molecular weight: 446.503 g/mol.

Epik wizard of Maestro to improve virtual screening enrichment. Moreover, the ligand was desalted to remove extra molecules, such as water molecules and counter-ions, present in the ligand file from certain structure databases. To define the stereoisomers, the computation was restrained to determine chirality from the 3D structure of P5J to account for keto-enol tautomerization, analogous sulfur and nitrogen tautomerizations, as well as histidine- and DNA-base tautomerizations. However, the generation of tautomeric forms of P5J was avoided to prevent tautomeric duplicates while maintaining accuracy, computational efficiency, quality experimental validation, reliable structure-activity relationship, consistency, and simplifying analysis. Similarly, refined PI3K-α ligand molecules obtained from the binding database (bindingdb.org) were refined for docking using a similar 6PYS protein-ligand complex LigPrep approach.

The LigPrep settings for refining the inhibitory molecules excluded tautomer generation. However, stereoisomer computation was carried out to determine chirality based on the 3D structure with the objective of having the internally produced stereoisomers filtered to remove any structures, fused ring systems, or chirality that were incompatible with that of natural products to generate the desired enantiomers. Ligand alignment was performed on all refined inhibitory molecules, including the P5J cocrystallized ligand. This option aligned structures with similar orientations, facilitating the identification of the ligand pose that maximizes beneficial interactions, such as hydrogen bonding, hydrophilic interactions, and electrostatic interactions, while minimizing detrimental interactions or clashes.

2.6. Ligand virtual screening-molecular docking

A structure-based, *in silico* virtual screening approach using Schrödinger Maestro was applied to predict the interaction, favorable binding orientations, and conformation of the refined ligands within the active site of the target PI3K-α protein. Essentially, for a molecule to tightly bind to a receptor, both geometric (shape) and electrostatic (charge) complementarities must exist.[47,48] These complimentary aspects define the molecular dynamics of the ligand-receptor relationship by incorporating interaction maximization while minimizing the total energy of the complex.

Typically, most *in silico* docking programs are built to predict binding mode and binding affinity between protein and ligand using a hybrid search algorithm and scoring function. While the search algorithm robustly generates multiple poses for a ligand in the binding site of the receptor, the scoring function ranks or orders the conformations to distinguish the experimental binding pose from the rest of the predicted poses. However, in Schrödinger Maestro, the binding affinity was calculated as (Equation III):

$$Binding\ affinity = \vartheta_{L/VdW} + \vartheta_{E} + \vartheta_{Hb} + \vartheta_{SM} + \vartheta_{\pi} + \vartheta_{Halogen} \quad \text{(III)}$$

where $\vartheta_{L/vdW}$ is the energy contribution associated with lipophilic pair concerning total Van der Waals force of interaction, ϑ_{E} is the energy contribution associated with electrostatic interactions, ϑ_{Hb} is the energy contribution associated with hydrogen bond interaction, ϑ_{SM} is the energy contribution associated with site map interactions, ϑ_{π} is the energy contribution associated with pi-cation interaction, and $\vartheta_{Halogen}$ is the energy contribution associated with halogen bond interactions.

In this study, a receptor grid was initially generated around the region occupied by P5J in the 6PYS protein, aiming to map the properties of the binding site onto a grid. Next, the refined ligands were docked into the minimized 6PYS protein through Schrödinger's Virtual Screening Workflow panel, with customized settings that automated a virtual screening-molecular docking workflow. The virtual screening workflow was automated to screen the ligands through successive stages, starting from high-throughput virtual screening mode (HTVS) to standard precision (SP) mode, and lastly, through extra-precision (XP) mode. The submission ratio of screened ligands at each stage was set to 70%, 60%, and 8%, respectively, as depicted in Figure 3. A trade-off between the speed and accuracy of the virtual screening served as the basis for selecting this ratio. By filtering out a large fraction of compounds in the early stages, the workflow could save time and resources, considering that docking and post-processing are computationally intensive and time-consuming compared to ligand preparation. However, by retaining a sufficient

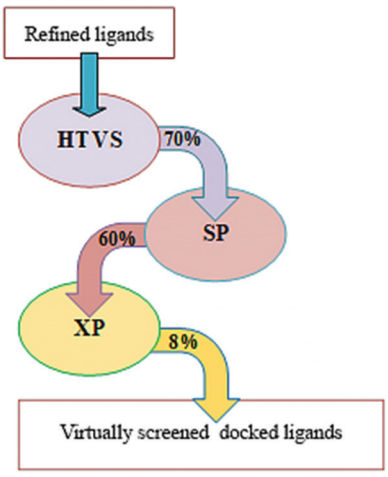


Figure 3. Schematics of submission ratio for virtual screening workflow for docking in Glide.
Abbreviations: HTVS: High-throughput virtual screening; SP: Standard precision; XP: Extra-precision.

fraction of compounds in the later stages, the workflow could ensure that the final hits are diverse, relevant, and reliable, as docking and post-processing are more accurate and informative than ligand preparation. Therefore, this ratio was chosen as a default value that balances the speed and accuracy of the virtual screening workflow. In addition, the virtual screening workflow aided in the computation of performance scores, including Glide, docking, interaction, and penalty scores, as well as similarity scores, to validate the refined ligands. Furthermore, during virtual screening, interaction scores for residues within 12 Å of the grid center were considered.

2.7. Implementation of 3D-QSAR

The computational modeling technique employed in this study was field-based 3D-QSAR to analyze and predict the relationship between the 3D structure of the refined ligands based on their alignment, similarity to a known pharmacophore, and biological activity. The pIC_{50} values served as a measure of the potency or biological activity of the ligands.

The pIC_{50} values of the refined ligands ranged from 4.866 to 9.398. In Schrödinger Maestro, structural alignment was deployed to identify similar ligand structures, focusing on identifying the core for each structure to align the molecules effectively. The field-based model was built on the Gaussian field domain that utilized a training set of 70% of the total input of refined ligands and a random seed set to 0, with a maximum of four partial least square factors. Both the steric and the electrostatic force fields were set truncated at 30.0 kcal/mol, and the cross-validation was performed by leaving out just one ligand.

Statistical analysis methods, such as comparative molecular field analysis (CoMFA) or comparative molecular similarity indices analysis (CoMSIA), as well as partial least squares (PLS), were applied to correlate the calculated descriptors with the activity values of the ligands in the training set. The generated correlation was utilized to predict the activity of new compounds based on their 3D structures. In addition, five Gaussian field fractions, including steric, electrostatic, hydrophobic, hydrogen bond acceptor, and hydrogen bond donor, were evaluated to provide insight into the field interactions of the ligands within the binding pocket of the receptor. An optimal number of PLS factors that can balance the trade-off between data fitting and model prediction was chosen.

2.8. Rational design of a new ligand

The rational design of a new ligand in this study involved a robust approach to iteratively modifying the skeletal structure of a lead compound, considered the reference ligand. The reference ligand was obtained after virtual screening-molecular docking, and the interactive pose prediction (IPP) panel offered in Schrödinger Maestro was employed for a new ligand design. The IPP operated under the maximum common substructure (MCS) constrained docking type, which simultaneously docked the compounds into the binding site of proteins using a grid-based approach. The GlideScore value and predicted biological activity from the 3D-QSAR model were the metrics employed for performance verification between the newly designed compound and the reference lead compound. Furthermore, the interaction pattern of the designed ligand within the protein was also studied. In addition, the ADMET-related indices of the new ligand were assessed using the QikProp program in Schrödinger Maestro at normal mode to evaluate its pharmacokinetics, efficacy, and safety profiles. All ADMET-related indices for the new compounds were evaluated using a total of 50 descriptors with a #star parameter as an indicator of several property descriptors computed by QikProp that violate a given optimum range of values for 95% of known drugs.

3. Results and discussion

3.1. Data preprocessing

Several ligands obtained from the database were dropped during the data preprocessing step due to missing column information, inconsistencies, and ambiguities in the data structure. This step resulted in reducing the initial dataset size from 3994 rows of ligands to 2972 rows and 48 columns, ensuring a clean dataset for analysis. In addition, the computed pIC_{50} values (activity) for each ligand ranged from 4.54 to 10.15 throughout the dataset. Using pIC_{50} as a measure of activity guarantees that the potency of different compounds can be precisely compared, facilitating the evaluation of their efficacy and the selection of the most promising candidates for further research or drug development.

3.2. Protein complex refinement

In mechanistic studies involving drug design, molecular docking, and prediction of protein functions, refining and minimizing protein structures play a significant role in improving their utility in pharmaceutical applications.[49] In Figure 4A, the 6PYS protein structure representing the human PI3K-α protein complex possesses inherent local and global errors, including irregular contacts or hydrogen bonds, chain breaks and atomic clashes, and unusual bond angles and lengths.[50] However, refining a protein obtained from a database before docking improves the accuracy and reliability of docking results.

Figure 4B illustrates the schematics of the refined 6PYS protein complex with the necessary side-chain

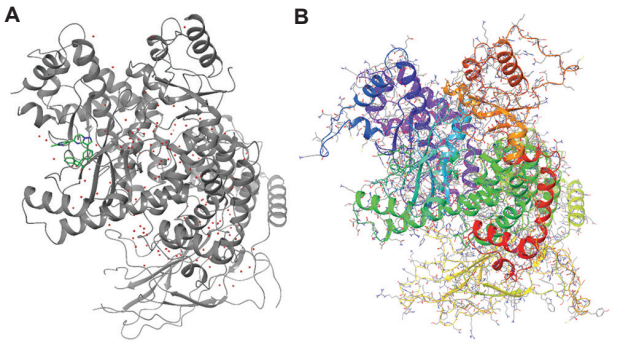


Figure 4. The human PI3K-α protein complex (Protein Data Bank ID.: 6PYS) with its cocrystallized ligand, (3S)-3-benzyl-3-methyl-5-[5-(2-methylpyrimidin-5-yl)pyrazolo[1,5-a]pyrimidin-3-yl]-1,3-dihydro-2H-indol-2-one (P5J). (A) Cartoon representation of the raw form; and (B) minimized-refined form. Red dots in (A) represent water molecules.

conformations, addressed chain breaks, added missing atoms or residues, assigned bond orders, converted selenomethionines, and deleted far-water molecules. In addition, refinement of the cocrystallize ligand enables the optimization of its positional and thermal parameters, resulting in a more precise representation of its interactions within the active site of the receptor molecule. This optimization accounted for the spatial arrangement of atoms, bond lengths, bond angles, and torsional angles of the ligand. Moreover, ligand refinement allows for the assessment of ligand-receptor interactions, such as hydrogen bonding, Van der Waals contacts, and electrostatic interactions. Precise refinement of these interactions provides insights into the binding affinity, specificity, and structural basis of ligand recognition. Furthermore, ligand refinement evaluates the underlying principles of binding affinity, specificity, and structural basis of the ligand concerning optimized ligand refinement parameters, including the assignment of proper bond orders, generation of accessible tautomers and ionization states, and prior virtual screening.[50]

3.3. Protein-ligand binding pocket identification

The binding pockets/sites of several therapeutic targets are significantly impacted by protein dynamics.[51] Moreover, the structural information of the protein-ligand complex can accelerate the optimization process of potential lead compounds and help solve problems related to compound selectivity, pharmacokinetics, and patentability.[52] To ensure specificity and further relay information between active and allosteric sites, protein-ligand binding is enhanced by physical interactions between the binding site residues of the protein and the ligand.[53]

Identifying binding sites that predict the concavity where the core scaffold can bind with the protein is essential for rational drug design. Binding pocket identification for

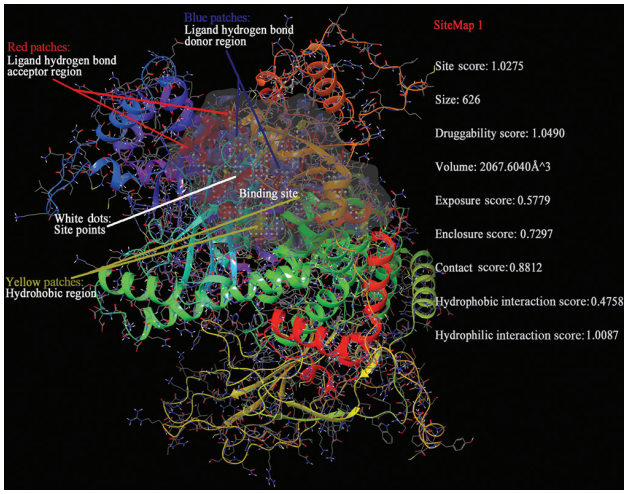


Figure 5. A visual representation of potential 6PYS (human PI3K-α protein complex) binding site (SiteMap 1).

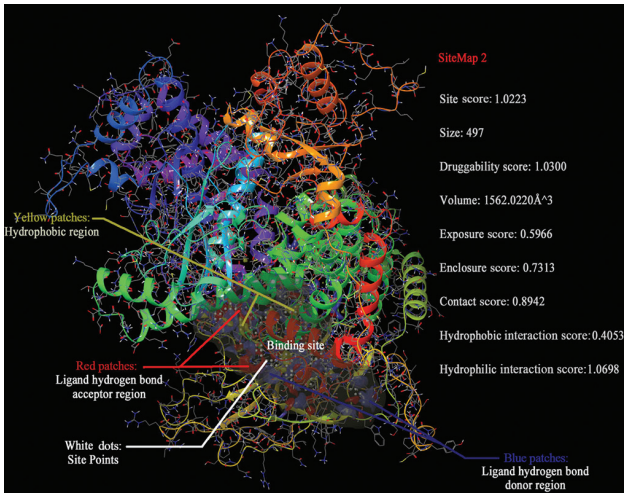


Figure 6. A visual representation of potential 6PYS (human PI3K-α protein complex) binding site (SiteMap 2).

6PYS protein and P5J ligand is depicted in Figures 5-9. In each figure, the binding site constituted some ligand hydrogen bond donor and acceptor regions, along with sites where hydrophobic interactions could occur. These specific properties of the protein significantly influence ligand binding and interaction.

The binding site score is a metric used to assess the likelihood or potency of an interaction between a small-molecule ligand and a protein at a specific binding site. It quantifies the propensity of the ligand to bind to the local site of the protein, facilitating firm ligand binding and the formation of a stable complex.[54] The score is usually presented as a numerical value and can be used to rank or prioritize ligands based on their potential binding affinity to the protein. It is computed based

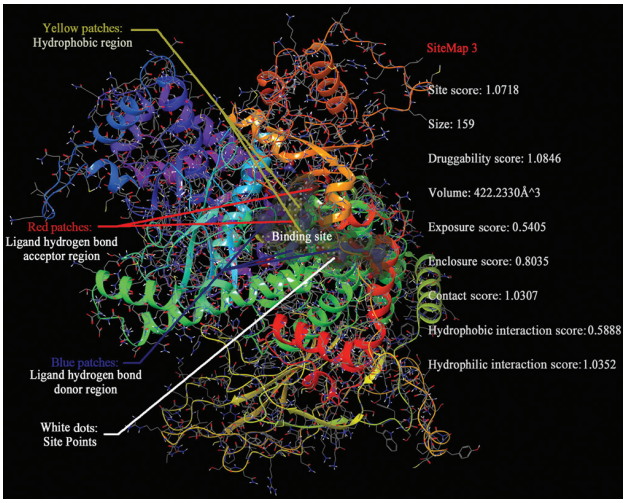


Figure 7. A visual representation of potential 6PYS (human PI3K-α protein complex) binding site (SiteMap 3).

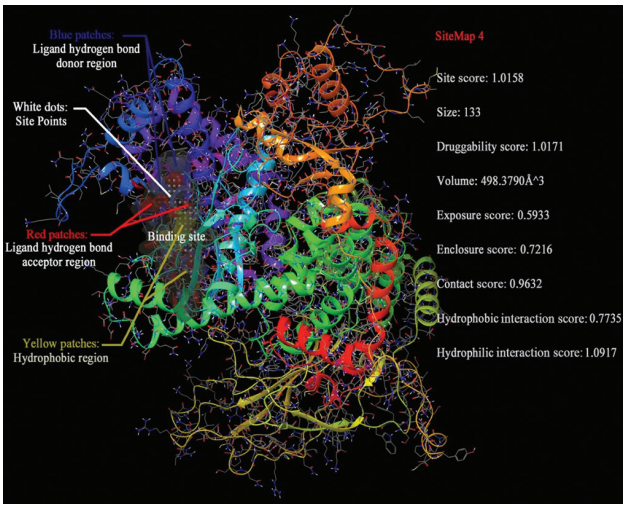


Figure 8. A visual representation of potential 6PYS (human PI3K-α protein complex) binding site (SiteMap 4).

on factors such as the exposure of the site to solvent, the degree of enclosure by the protein, and the degree of hydrophilicity/hydrophobicity. A higher binding site score indicates a stronger predicted interaction, suggesting a higher likelihood of optimized binding and potential biological activity. Conversely, a lower score indicates a weaker predicted interaction. A site score above 1 indicates a highly druggable protein pocket, while a score of 0.8 distinguishes between drug-binding and non-drug-binding sites.[54] The average site score value for sub-micromolar binding sites is 1.01.[55] In Figures 5-9, the computed site scores were mostly above 0.95. However, despite the nearly identical site score values, the site score value of SiteMap 3 (Figure 7) ranked highest at 1.0718.

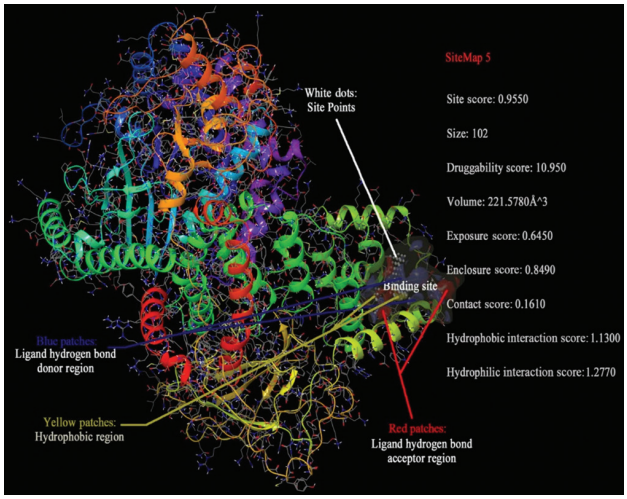


Figure 9. A visual representation of potential 6PYS (human PI3K-α protein complex) binding site (SiteMap 5).

These results suggested that while the site score may aid in deciding the target site, it may not provide sufficient information about the drug-binding state of the site,[55] as it relies on computational estimations and is primarily useful for predicting ligand-protein interactions.

The size of the binding site corresponds to the size of the interacting ligand, determining the maximum ligand size that can fit and interact with the protein. However, ligand binding pockets varied widely in size, as shown in Figures 5-9. Comparatively, a binding site size of 102 (Figure 9**,** SiteMap 5) would typically hold a smaller ligand than a binding site of size 162 (Figure 8**,** SiteMap 4). Moreover, the binding of ligands of different sizes, shapes, and compositions implies a different geometry of the receptor binding sites that allow the movement of backbone and side-chain atoms, resulting in differences in the details of residue interactions.

Druggability refers to the likelihood of a drug-like compound or molecule modulating or exhibiting protein interaction[54] at a therapeutically useful level of affinity. Here, the druggability of the protein was evaluated using SiteMap. Interestingly, Figures 5-9 revealed a wide range of druggability scores (0.9504 – 1.0846) for the refined protein. Among the identified druggable pockets, SiteMap 3 (Figure 7) ranked the highest, with a score of 1.0846 as compared to others. It is noteworthy that a druggability score over 0.5 and closer to 1.0 indicates good druggability. The druggability score is computed based on the enclosure and hydrophilicity (hydrophilic interaction score) of the tested pocket (Equation I).

In Figures 5-9, the enclosure score ranged from 0.6314 – 0.8035. There is a positive correlation between the

enclosure score and the druggability score, indicating that a rise in the druggability score corresponds to a higher degree of cavity enclosure, represented by the enclosure score. A higher enclosure score suggests an envelope with a tighter barrier that would prevent ligands from releasing once it is enclosed. Hence, an enclosure score of 0.8035 depicted a high-affinity enclosure (potentially stimulus-sensitive) for the 6PYS human PI3K-*α* protein to envelope a ligand and act as a diffusion barrier, providing an excellent slow release.[56]

Conversely, the pocket hydrophilicity (hydrophilic interaction score) of the protein depicted an inverse correlation with the druggability score, as illustrated in Figures 5-9. The hydrophilicity of the binding pockets primarily favors polar ligands,[57] which have a strong affinity for water and are soluble in water. Although the results shown in Figures 5-9 are approximately identical, SiteMap 5 obtained the best hydrophilic interaction score of 1.1301, indicating that the binding site provided a better hydrophilic environment with a higher affinity for interacting with polar ligands and water compared to other binding sites.

In addition, Figures 5-9 present the computed results of the volume of potential binding sites from the respective SiteMaps. The volume of the binding pocket plays a significant role in the process of rational drug discovery. The size and shape of the drug molecule that can successfully bind to the protein depend on the volume of the binding pocket. While a smaller binding pocket might be better suited for tiny molecules, a larger binding pocket might be able to accommodate larger medication molecules. In Figures 5-9, binding SiteMap 1 suggested a larger binding volume of 2067.6040 $Å^3$ when compared to the other binding SiteMaps. In any case, a well-fitted ligand into the receptor pocket can trigger signaling transduction, while a mismatch in volume fitting may decrease responses.

Figures 5-9 display the result of the ligand-receptor complex contact score. The contact score is a property that evaluates the strength of the average site-point interactions, such as hydrogen bonding, electrostatic interactions, hydrophobic interactions, and non-bonded Van der Waals interactions with the receptor.[55] These interactions contribute to the stability and strength of the ligand-receptor complex and are essential for effective binding. The contact score provides a numerical value representing the quality of the ligand-receptor interaction, calibrated such that the average score for the sub-micromolar sites is 1.0, facilitating comparison between sites, where a higher value indicates better interaction. In Figures 5-9, the contact score ranged from 0.8492 – 1.0307. SiteMap 5 possessed the highest contact score value (1.0307), indicating a strong and favorable interaction of the ligand with the target receptor, suggesting a higher potential for binding and therapeutic activity of optimal efficacy.

Figures 5-9 depict the computed results for the binding site exposure score, which explains the degree of accessibility or availability of the binding site to interact with a ligand, as well as its openness to solvents. However, in Figures 5-9, the exposure scores of the SiteMaps range from 0.5405 – 0.6446. A lower exposure score indicates a deeper or well-encapsulated site favorable for excellent ligand-protein binding. More so, for a favorable binding, the Van der Waals contact with the receptor should lie at least 4 Å from the nearest protein atom. Hence, the numeric value of the exposure score of SiteMap 5 (Figure 9) suggested that better molecular mechanotransduction could occur, due to a deeper and better-encapsulated site location in the receptor for ligand binding.[55]

The dynamics around the hydrophobicity of a binding site is another important property that ensures the presence of hydrophobic characters necessary for high binding affinity. Hydrophobicity is a characteristic that measures the extent to which a molecule is able to repel water. Proteins tend to bury hydrophobic residues within their core during the folding process to stabilize the protein structure and prevent aggregation. However, the hydrophobic interaction score measures how effectively hydrophobic molecules interact with each other in a solvent environment. Figures 5-9 depict the results of the respective SiteMaps' hydrophobic interaction scores. The hydrophobic interaction score for SiteMap 4 (Figure 8) was 0.7735, ranking top among all the SiteMaps assessed. This result depicted a region with a higher propensity for stronger hydrophobic binding sites to bind to ligands with higher affinity. In general, entropic phenomena cause hydrophobic molecules to interact with each other in water.[57]

3.4. Molecular docking analyses

In this study, we report a comprehensive workflow involving three stages of virtual screening processes: HTVS, SP, and XP modes that simultaneously incorporated molecular docking at each stage to discover potential top-performing PI3K-*α* inhibitors for the suppression of human PI3K-*α*-related problems.

Table 1 reveals four potential candidates identified as top-ranked inhibitory molecules from the virtual screening and docking of several ligands on the human 6PYS protein complex. The molecular docking results of these compounds revealed that the structural motifs of the top-ranked potential inhibitory molecules that could bind to the 6PYS human PI3K-*α* protein consisted of cyclic sulfonamide derivatives attached to pyridine cores,

Table 1. Molecular docking result from the three-stage virtual screening workflow

Reordered ligand candidacy number	Candidate 1	Candidate 2	Candidate 3	Candidate 4
Ligand digital nomenclature	00310001.cdx	00254001.cdx	00137001.cdx	00199001.cdx
Ligand nomenclature	3-(6-amino-5-(2-methyl oxazol-5-yl) pyridin-3-yl)-N-((3-hydroxy oxetan-3-yl) methyl)-4-methylbenzenesulfonamide	3-(6-Amino-5-(isoxazol-5-yl) pyridin-3-yl)-N-(2-hydroxy-2-methylpropyl) -4-methylbenzenesulfonamide	3-[6-Amino-5-(2-methyl-oxazol-5-yl)-pyridin-3-yl]-N-(3-hydroxy-3-methyl-butyl)-4-methylbenzenesulfonamide	3-(6-Amino-5-(3-methyl-1,2,4-oxadiazol-5-yl) pyridin-3-yl)-N-((4-(hydroxymethyl) tetrahydro-2H-pyran-4-yl) methyl)-4-methylbenzenesulfonamide
Activity (pIC_{50})	6.5086	6.4685	6.8539	6.2518
glide gscore	−12.8675	−12.6943	−12.7015	−12.6736
glide evdw	−38.4417	−41.9224	−43.5924	−45.2563
glide energy	−49.8733	−49.7553	−54.6716	−53.5503
glide einternal	5.0994	6.2456	5.8060	4.5817
glide emodel	−80.7610	−77.5800	−84.7264	−81.8777
XP HBond	−2.1467	−1.9817	−1.9926	−1.7768
XP PhobEn	−2.0000	−1.9500	−1.9500	−2.0750
XP LowMW	−0.0651	−0.1584	−0.0649	0.0000
XP RotPenal	0.1931	0.2181	0.2253	0.1894
XP LipophilicEvdW	−5.8921	−6.2412	−5.9585	−6.1663
XP Electro	−0.8574	−0.5875	−0.8309	−0.6220
XP Sitemap	−0.1494	−0.1468	−0.1799	−0.3769
XP ExposPenal	0.0000	0.1032	0.0000	0.1040
XP ClBr (halogen contribution)	0.0000	0.0000	0.0000	0.0000
XP PiCat	0.0000	0.0000	0.0000	0.0000
glide eff state penalty	0.0206	0.0008	0.0206	0.0008
Binding affinity	−9.0456	−8.9569	−8.9619	−8.9420
Computed RMSD against reference ligand structure (Å)	2.98	2.98	2.97	2.94

along with several other substituent moieties. While these moieties vary, they may be derived from alkaloids.

Surprisingly, sulfonamide derivatives represent a class of intriguing compounds with a diverse range of pharmacological activities,[58-60] including anti-cancer, anti-bacterial,[59,61-66] anti-fungal,[59,67] anti-oxidant,[59,68,69] anti-inflammatory,[59,62,70,71] and anti-diabetic[60,72] activities. However, the results presented in Table 1 suggest that a hybrid of sulfonamide and pyridine-based heterocyclic compounds functionalized with other potent moieties exhibits adequate synergistic biological effects capable of enhancing sufficient biological activity against PI3K-*α*. The binding affinity values for the respective candidates demonstrated a reasonable reflection of a spontaneous and favorable free energy change toward the bound complex. In addition, in Table 1, the RMSD values between the docked compounds and the cocrystallized ligand (P5J) were within the threshold of 2 – 3Å, suggesting a successful docking pose prediction with appreciable biological relevance, attributed to the chemostructural similarities of the candidates. Furthermore, in Table 1, the values of other crucial parameters for the respective candidates offer significant insights into key mechanistic interactions between a ligand and the target, PI3K-*α*.

3.5. Field-based 3D-QSAR analysis

To establish the structure-activity relationships of the selected human PI3K-*α* inhibitory compounds, a robust field-based 3D-QSAR model was employed via an *in silico* technique. The field-based 3D-QSAR model was a hybrid model composed of a force field model and a Gaussian

model. While the force field model is similar to the comparative molecular field analysis (CoMFA) models, the Gaussian model is similar to the comparative molecular structure analysis (CoMSIA). However, the hybrid model is built on the OPLS-3e force field, which requires a set of aligned compounds in a 3D space with known activities for predictive operation.

Table 2 summarizes the performance of the robust field-based 3D-QSAR model, organized into four different partial least squares (PLS) regression sub-models. Each row illustrates the application of special multivariate statistical analysis routines – PLS factors – used in building the model. The PLS factors are linear combinations of the original predator variables that are used to fit a linear regression model. The number of PLS factors or latent variables was employed to extract the most relevant information from the molecular descriptors and the biological activity data of the inhibitors. Based on the pIC_{50} as the activity property of the inhibitory compounds, the model performance was validated across several statistical indices, such as standard deviation (SD), coefficient of determination (R^2), R^2 cross-validation for the training set, R^2 scramble (R^2 *Scr*), model stability, *F*-value, *p*-value, root mean square error (RMSE), cross-validated coefficient of determination (Q^2) for the test set, and the Pearson-*r* correlation with the test set of the model. Hence, the PLS factors are compared with the best performer selected to depict the field-based 3D-QSAR model.

The SD is a descriptive statistic used in the majority of clinical and experimental studies, and it illustrates how the mean represents sample data while also evaluating the variation in a dataset that follows a normal distribution.[73-75] A low SD indicates more stable and consistent sample data whose values are close to the mean, while a high standard deviation indicates that the values are spread out (scattered) over a wide range, away from the mean. Thus, as shown in Table 2, the order of significance of the SD for the four PLS factors is illustrated by the following trend:

SD: PLS factor 4 ⟹ *PLS factor* 3 ⟹ *PLS factor* 2 ⟹ *PLS factor* 1

The coefficient of determination (R^2) is the measure of the proportion of variance in the observed activity that is explained by the model.[76] Statistically, the R^2 value ranges from 0 to 1, with 1 indicating the excellent explanatory power of a model. Furthermore, a high R^2 value indicates a good fit of the model data. Hence, in Table 2, the hierarchy of importance in the value of R^2 across each PLS factor is depicted in the following trend:

R^2: *PLS factor* 4 ⟹ *PLS factor* 3 ⟹ *PLS factor* 2 ⟹ *PLS factor* 1

Emphatically, it is important to note that a high value of R^2 does not necessarily translate to a better model when interpreting the R^2 statistic of a QSAR model for novel drug discoveries, since a high R^2 may be achieved by overfitting the model to the training data. Furthermore, it does not imply a causality nor provide a mechanistic interpretation of the QSAR model,[77,78] which is often misconceived among many researchers.

The coefficient of determination obtained by cross-validation (R^2 cross-validation) is a statistical method used to assess the predictive performance of the model on new data. In addition, the R^2 cross-validation values were computed from the predictions obtained by a leave-one-out (LOO) approach as a measure of the predictability of the CoMFA mode. According to the CoMFA approach, changes in the biological activities or binding affinities of sample compounds correlate with the variations in the steric and electrostatic fields of the molecules since drug-receptor interactions are typically non-covalent. In a typical CoMFA procedure, the steric and electrostatic fields around each molecule under study are sampled using probe atoms, typically sp3 carbon atoms with +1 charge, on a rectangular grid that includes the structurally aligned molecules.[79] The R^2 cross-validation value greater than 0.3 is considered significant. However, in Table 2, only PLS factor 1 deviated from significance in that its R^2 cross-validation value was 0.2157. Hence, the following trend depicted the order of significance among the four PLS factors:

R^2CV: *PLS factor* 4 ⟹ *PLS factor* 3 ⟹ *PLS factor* 2 ⟹ *PLS factor* 1

Table 2. Summary of Field-based 3D-QSAR model performance

# PLS factors	Standard deviation	R^2	R^2 cross-validation	R^2 scramble	Stability	*F*-value	*P*-value	RMSE	Q^2	Pearson-*r*
1	0.7356	0.2815	0.2157	0.0499	0.994	37.2	2.28E-08	0.67	0.3583	0.6035
2	0.5484	0.6049	0.4424	0.1749	0.963	72.0	1.11E-19	0.50	0.6486	0.8062
3	0.4370	0.7517	0.4643	0.2736	0.875	93.9	4.94E-28	0.45	0.7185	0.8488
4	0.3622	0.8313	0.4502	0.3674	0.721	113.3	1.11E-34	0.40	0.7776	0.8825

Abbreviations: PLS: Partial least squares; RMSE: Root mean square error.

The coefficient of determination obtained by scrambling, R^2 scramble (R^2Scr), is one of the statistical methods to test the significance of a 3D-QSAR model. It is the average value of R^2 from a series of models built using scrambled activities and measures the degree to which the molecular fields can fit meaningless data. Table 2 shows the values of R^2 scramble for the respective PLS factors, reveals that PLS factor 1 scored lowest with a value of 0.0499, while PLS factor 4 scored a value of 0.374. Moreover, a high R^2 scramble indicates that the model is not meaningful and may be overfitting the data. Hence, the model containing a PLS factor emerged as the most significant model with meaningful data fitting. Concerning R^2Scr values for various PLS factors in Table 2, the following trend summarizes the order of meaningfulness of models as obtained by random shuffles of the values of the bioactivity response variable:

R^2Scr: PLS factor 4 ⟹ *PLS factor* 3 ⟹ *PLS factor* 2 ⟹ *PLS factor* 1

Table 2 also shows the value of stability of respective PLS factors. Stability accounts for how stable the PLS factors are when different subsets of data are used to fit the model. This statistic ranges from 0 to 1, where 1 means that the factors are identical for all subsets (stable), and 0 means that they are completely different. However, it is inferred from Table 2 that PLS factor 1 is more stable at model predictions to changes in the training set composition than other PLS factors, as highlighted in the following trend:

Stability: PLS factor 4 ⟹ *PLS factor* 3 ⟹ *PLS factor* 2 ⟹ *PLS factor* 1

F-value and *p*-value are also statistical indices used to validate the 3D-QSAR model. *F*-value statistic tests whether adding a new PLS factor to the model significantly improves its fit or not, while *p*-value is a probability that measures how likely it is to obtain *F*-value as large or larger than the observed one by chance alone, assuming that adding a new PLS factor does not improve the fit of the model. *F*-value is computed by comparing the sum of square errors (SSE) of two nested modes: one with k PLS factors and one with k+1 PLS factors. A higher *F*-value means that adding a new PLS factor reduces the SSE significantly and improves the fit of the model. Conversely, a lower *p*-value means that adding a new PLS factor is more significant and not due to chance. Table 2 suggested that PLS factor 4 had a higher *F*-value (113.3) when compared with the other PLS factors, which means that its addition to the model would significantly improve its fit and reduce the SSE. The hierarchy of *F*-value significance is depicted below:

F-value PLS factor 4 ⟹ *PLS factor* 3 ⟹ *PLS factor* 2 ⟹ *PLS factor* 1

PLS factor 4 had the lowest *P*-value (1.11E-34) among all four PLS factors, and adding it to the model would improve the model significantly, but not due to chance. Hence, the following trend is the case for the order of *P*-value significance among the PLS Factors:

P-value PLS factor 4 ⟹ *PLS factor* 3 ⟹ *PLS factor* 2 ⟹ *PLS factor* 1

The remaining aspects of Table 2 were statistics typically associated with the test set of the input data used to build the 3D-QSAR model. The RMSE is the statistic that describes how close the predicted values of the dependent variable are to its actual values. A lower RMSE means that the predictions are more accurate and have fewer errors. However, Table 2 revealed the lowest RMSE value for PLS factor 4, while the RMSE value of 0.67, which represented PLS factor 1, suggested a slightly weaker model prediction with somewhat more errors than other PLS factors. The trend for the order of significance is presented as follows:

RMSE: PLS factor 4 ⟹ *PLS factor 3* ⟹ *PLS factor 2* ⟹ *PLS factor 1*

The predictive squared correlation coefficient of the test set, Q^2, is a statistical property that reinforces the validity of a QSAR model based on quantifying the predictive ability of the model in the aspects concerning reliability, accuracy, and applicability domain of the model. The Q^2 is obtained from methods based on simple reuse, such as leave-one-out and leave-many-out cross-validation.[80] This parameter is important and has become well-known because it takes values in a normalized range (i.e., ≤1), thereby permitting a trivial understanding of its values and easy comparison of different QSAR models and the different performance of fitting and predictive capabilities of a model. However, a higher Q^2 means that the predictions are reliable and have less uncertainty. Hence, in Table 2, it is revealed that the computed value for Q^2 for PLS factor 4 exceeded the Q^2 values of other PLS factors. This implied a more confident final result valid both for internal validation, such as cross-validation or bootstrap, as well as external validation[80] and the following trend depicts the reliability order of Q^2 on model:

Q^2: PLS factor 4 ⟹ *PLS factor* 3 ⟹ *PLS factor* 2 ⟹ *PLS factor* 1

The Pearson correlation coefficient, or Pearson-*r* statistic, assesses the strength of the correlation between two continuous variables, ranging from 0 to 1. A value of 1 indicates a stronger correlation, while a value of 0 represents a weaker correlation. Table 2 illustrates the Pearson correlation coefficient, which estimates the degree of correlation between the respective PLS factors and the predicted activities of the model using the test set. It is therein in Table 2 that PLS factor 4, with a Pearson-*r*

value of 0.8825, held a stronger correlation with the model than other PLS factors, even as PLS factor 3 ranked next. However, the rank which the degree of correlation is depicted in the following trend:

Pearson–r: *PLS factor* 4 ⟹ *PLS factor* 3 ⟹ *PLS factor* 2 ⟹ *PLS factor* 1

The table can be used to compare different models with different numbers of factors and select the best one based on the given statistical criteria. However, a good QSAR model should not only have a high R^2 value but also have high predictive power, robustness, interpretability, applicability, etc. It is logical to choose a model that has high R^2, Q^2], and Pearson-*r*, and a low RMSE and *p*-value, while avoiding overfitting (indicated by a large gap between R^2 and R^2 CV or a high R^2 *Scr*) and instability (indicated by a low stability). It is inferred from Table 2 that PLS factors 4 and 3 were the best performers among other PLS factors. To compare, PLS factor 4 performed better than PLS factor 3 in terms of fit, significance, accuracy, reliability, and correlation, but worse in terms of variation, prediction, generalization, and stability. Based on these critical criteria, PLS factor 3 was considered the optimal PLS factor, which offered the safest approach that accounted for model performance and quality.

Table 3 displays the percentage contribution of different Gaussian descriptors in the QSAR model for each number of PLS factors used in the model. The Gaussian descriptors encode the mean and covariance information of local features in a graphical object.[81] In Table 3, PLS factor 3 comprised 27.3% steric, 10.5% electrostatic, 19.4% hydrophobic, 25.8% hydrogen bond acceptor, and 17% hydrogen bond donor descriptors. These values indicate the contribution of each Gaussian descriptor to PLS factor 3 in analyzing the relationship between the predictor variables and the response variable in the regression model. The result suggests that most of the binding energy predominantly emanated from steric (27%) and hydrogen bond acceptor (25.8%) interactions. Steric and hydrogen bond acceptor interactions influence protein-ligand binding affinity and specificity. Steric interactions depict the shape complementarity and spatial fit of the ligand, as well as influence the entropy of the binding process of the protein-ligand complex.[82] On the other hand, hydrogen bond acceptor interactions illustrate how ligands accept electrons from a protein donor, stabilize the complex, and enhance ligand recognition within the protein,[83] in which the hydrogen bond paring influences shifts in pKa values of interacting groups and also affects charge distribution and electrostatic potential.[84]

In conjunction with Table 3, the visualization of the interactive contributions of the different Gaussian descriptors in the context of PLS factor 3 under study is graphically presented in Figure 10.

3.6. Rational design of a new ligand

3.6.1. Skeletal modifications

Medicinal chemistry continues to be impacted by innovative *in silico* methods, especially at the drug discovery stage using QSAR models. In this study, we attempted to design a new compound with enhanced binding affinity and specificity by enacting the desired chemical transformations on a reference ligand in a concise and chemospecific fashion. The concept of "molecular editing" involves building onto, modifying, or pruning molecules atom by atom, utilizing transformations that are adequately mild and selective for application in the later stages of drug synthesis and sequencing.[85]

The skeletal editing of the top-ranked hit compound (Candidate 1, in Table 1), which was obtained from the robust virtual screening-molecular docking workflow, involved some heterocyclic modifications that were in conformation with the congeneric series of molecules used in docking. The molecular modification between the reference compound and the resulting T85 is illustrated in Figure 11. The skeletal editing of the reference hit compound was associated with molecular modifications in the heterocycles of its oxazole and the oxetane-pyridine sub-cores, with fluorination and bipyridine formation dispatched, respectively. However, the fluorination involved the replacement of the 2-methyl group on the oxazole ring by a fluorine atom, and the pyridine ring was fused with another pyridine ring at the 3-position. The resulting putative ligand, referred to as T85, holds promise as the next synthetic hit compound.

Table 3. Molecular field fraction analysis

# PLS factors	Gaussian steric (%)	Gaussian electrostatic (%)	Gaussian hydrophobic (%)	Gaussian hydrogen bond acceptor (%)	Gaussian hydrogen bond donor (%)
1	0.428119	0.056828	0.172533	0.217674	0.124847
2	0.315070	0.093522	0.197463	0.237846	0.156099
3	0.273183	0.105146	0.194088	0.257736	0.169848
4	0.312171	0.093195	0.190332	0.260085	0.144218

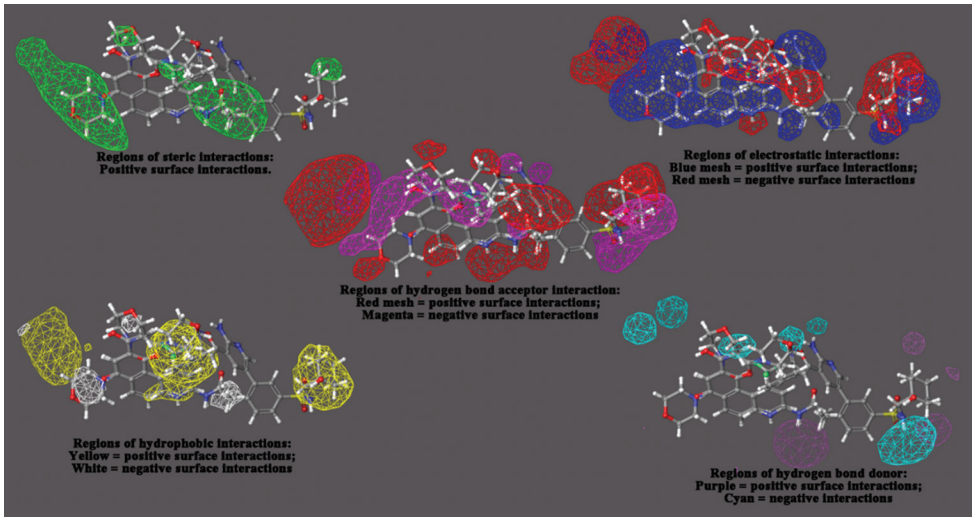


Figure 10. Field-based QSAR visualization of Gaussian descriptors for PLS factor 3.

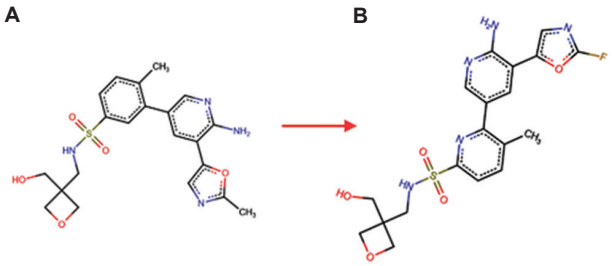


Figure 11. T85 (B) from the structural modification of Candidate 1 (reference hit compound [A]). The nomenclature of (A) and its SMILES notation is 3-[6-amino-5-(2-methyl-1,3-oxazol-5-yl)pyridin-3-yl]-N-{[3-(hydroxymethyl)oxetan-3-yl]methyl}-4-methylbenzene-1-sulfonamide and Cc1ncc(o1)-c1cc(cnc1N)-c1cc(ccc1C)S(=O)(=O)NCC1(CO)COC1, respectively, while the nomenclature of T85 (B) and its SMILES notation is: 6'-amino-5'-(2-fluoro-1,3-oxazol-5-yl)-N-{[3-(hydroxymethyl)oxetan-3-yl]methyl}-3-methyl-[2,3'-bipyridine]-6-sulfonamide and Cc1ccc(nc1-c1cnc(N)c(c1)-c1cnc(F)o1)S(=O)(=O)NCC1(CO)COC1, respectively.

The concept of incorporating a fluorine atom in the compound was to furnish T85 with unconventional and distinctive properties. Although fluorine addition is not a panacea, its presence in the drug molecule is intended to improve the drug's biological activity, affecting both the pharmacokinetics and dynamic properties,[86] ultimately aiding in saving lives. For example, the production of fluorine-associated novel therapeutic drugs approved by the FDA in 2021 was utilized for controlling the COVID-19 pandemic and treating various diseases.[87,88]

The low metabolic stability of drugs is one of the critical problems in drug development. However, this could easily be circumvented by blocking the metabolically labile sites with fluorine substituents. Moreover, the carbon-fluorine (C-F) bond is one of the strongest single bonds due to the high electronegativity of fluorine, giving the bond a significant dipole moment or polarity. In addition, the C-F bond is relatively short, enacted by its partial ionic character, which also affects the strengths of other bonds, rendering other parts of the drug harder to degrade.[89,90] Hence, the drug persists longer, exerting a prolonged effect on the targeted disease or condition. We anticipate this phenomenon to hold true for T85.

Improving the lipophilicity of the intended drug was another design consideration in line with drug development protocols. For effective passive transportation across the cell membrane, the drug must readily traverse the lipid membrane without any hindrance.[87]Therefore, for a better drug compound, moderate lipophilicity is necessary. Fluorination is often a better option due to its high lipophilicity, which increases drug absorption.[87,91] In light of the reported findings, the deliberate inclusion of fluorine during the design of T85 can productively influence its intrinsic potency and membrane permeability, thereby fostering effective metabolic pathways and pharmacokinetic properties against human PI3K-α. Moreover, the addition of fluorine to a molecule to inhibit a protein kinase involves modulating the binding affinity and selectivity of the drug to its target receptor by altering the electronic and steric properties of the molecule, influencing the conformation and solubility of the drug.

The asymmetrical combination of pyridines – bipyridine – in the design of T85 is believed to confer various advantages, such as enhanced binding affinity, selectivity, stability, and drug solubility. This modification can also alter the electronic and steric properties of the molecule, facilitating interactions with different molecules through non-covalent interactions and potentially resulting in the formation of supramolecular structures with interesting

properties. Furthermore, bipyridine is capable of inducing chirality through ring functionalization or restricted rotation (atropisomerism), thus increasing its relevance in asymmetry-based applications. However, the inclusion of fluorine in the design of T85 is still a limited study, and its full potential has yet to be clinically verified.

Table 4 summarizes the performance comparison between the reference hit compound and T85 obtained from the IPP under the MCS-constrained docking type and 3D-QSAR model prediction. To compare docking scores, a docking score of -9.88 was better than -9.51. This meant that T85 was predicted to have a stronger binding affinity or interaction with the target protein compared to Candidate 1. In addition, Table 4 indicates that Candidate 1 possessed a lower biological activity (6.51) than T85 (8.25) in inhibiting the target protein.

3.6.2. T85-6PYS interactions

In this work, we studied the interactive behavior of T85 within the receptor grid of the target protein in a 2D space. Figure 12 illustrates the amino acid residues of the 6PYS protein that enveloped T85 in the target binding site. The interaction diagram revealed both hydrogen bonds and non-bonded contacts, presenting a list of amino acid residues along with their corresponding positions in the protein sequence. Together, these elements formed the characteristic binding pocket of the 6PYS protein in complex with the T85 compound. The diagram revealed hydrogen bond interactions between T85 and the 6PYS protein complex, particularly in the regions where LYS 802 and VAL 851 amino acid residues were located. Furthermore, π-π bonded interactions occurred between the pyridine ring of T85 and the amino acid residue TYR 836, as well as between the oxazole ring of T85 and the amino acid residue TRP 780. The π-π bonded interactions are non-covalent interactions among the residues of proteins and nucleic acids, between ligands and proteins, which have extraordinary significance in interpreting the dynamics of intricate biological systems and the biological activity necessary for drug discovery.[92] These interactions contribute to the stability and selectivity of the protein-ligand complexes by providing additional attractive forces and complementarity between the binding partners. These suggest that T85 exhibits an adequate number of interactions, indicating optimal complementarity between the ligand and the receptor.

Figure 13 describes the non-covalent interaction of the hydrophobic type established in the enclosure of T85 by several hydrophobic amino acid residues. Interestingly, the result revealed that 11 hydrophobic amino acid residues enveloped T85, with more isoleucine (ILE) amino acid residues involved in the ligand-receptor hydrophobic interaction network at different positions in the protein sequence. However, the presence of multiple hydrophobic

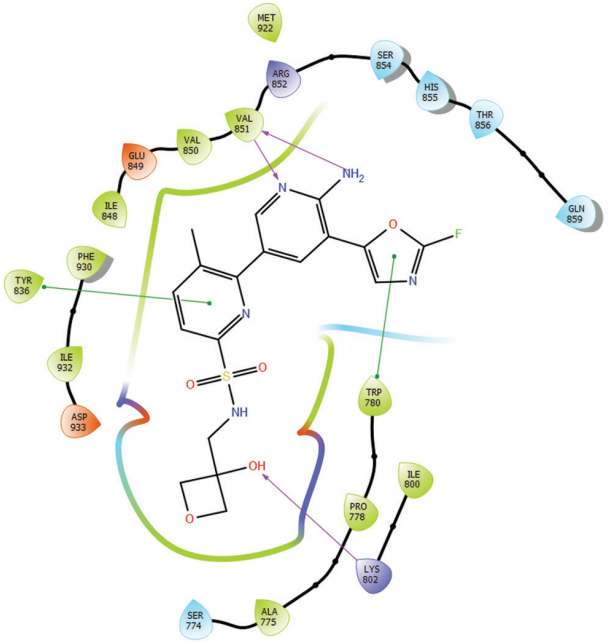


Figure 12. Two-dimensional schematics of T85 in the binding cavity 6PYS, surrounded by amino acids in their respective positions in the protein sequence.
Abbreviations: ALA: Alanine: ARG: Aginine; ASP: Aspartic acid; GLN: Glutamine; GLU: Glutamic acid; HIS: Histidine; ILE: Isoleucine; LYS: Lysine; MET: Methionine; PHE: Phenylalanine; PRO: Proline; SER: Serine; THR: Threonine; TRP: Tryptophan; TYR: Tyrosine; VAL: Valine.

Table 4. Performance comparison data between hit reference ligand and T85

Compound	Nomenclature	MCS- constrained docking score	Activity (pIC_{50})
Candidate 1 (hit reference compound)	3-[6-amino-5-(2-methyl-1,3-oxazol-5-yl) pyridin-3-yl]-N-{[3-(hydroxymethyl) oxetan-3-yl] methyl}-4-methylbenzene-1-sulfonamide	–9.51	6.51
T85	6'-amino-5'-(2-fluoro-1,3-oxazol-5-yl)-N-{[3-(hydroxymethyl) oxetan-3-yl] methyl}-3-methyl-[2,3'-bipyridine]-6-sulfonamide	–9.88	8.25

Abbreviation: MCS: Maximum common substructure.

amino acid residues around a ligand is crucial to driving coherent functional control of biomolecules.[93] Hence, from our simulation result, the formation of a hydrophobic core by the numerous hydrophobic amino acid residues around T85 demonstrated a strong interaction that indicated adequate ligand-receptor stability, binding specificity, and selectivity.

In addition, Figure 13 reveals the synergy between hydrophobic interactions and hydrogen bonds within the binding pocket of the 6PYS PI3K-α protein. The diagram depicted a hydrophobically packed correlated hydrogen bond, which is regarded as a signature type of interaction that explained the coexistence of polar (hydrogen bond) and non-polar (hydrophobic) interactions at certain locations of the binding cavity of the simulated T85-6PYS complex. The schematic revealed that hydrophobic interactions emanated from several unique hydrophobic amino acid residues of 6PYS, including VAL 851, TRP 780, VAL 850, TYR 836, PHE 930, ILE 800, ILE 848, LYS 802, and ILE 932.

The complexity of the interactions also involved the presence of hydrogen bonds within the hydrophobic vicinity of VAL 851. At a distance of 1.88 Å, a favorable intermolecular hydrogen bond interaction was formed between the polar hydrogen atom of T85 and an oxygen atom of the carbonyl functional group of the VAL 851 amino acid side chain (Figure 13). In addition, in the binding cavity of the protein, the nitrogen atom in the pyridine ring of T85 established another hydrogen bond with a hydrogen atom of the VAL 851 side chain with an interactive distance of 2.12 Å. These hydrogen bonds illustrated a strategic molecular recognition of T85 in its tautomeric state, signaling substantial therapeutic effects. The entrapment of T85 within the binding cavity of the 6PYS protein further illustrated that for T85 to be released, it would need to simultaneously break both hydrogen bonds with VAL 851 and all hydrophobic interactions between the hydrophobic amino acid residues of the receptor, implying a strong binding of the ligand to the receptor.

All hydrophobic amino acid residues are depicted as CPK representation, while the portions of the T85 that formed hydrophobic interaction are revealed in ball-and-tube representation.

Figure 14 describes the details of non-covalent interactions, especially hydrogen bonds and π-π (pi-pi) stacking interactions within the receptor-binding domain of the target 6PYS. Non-covalent interactions, such as hydrogen bonds, are ubiquitous in nature and serve a variety of significant roles in protein folding, protein-ligand interactions, catalysis, and maintaining specific

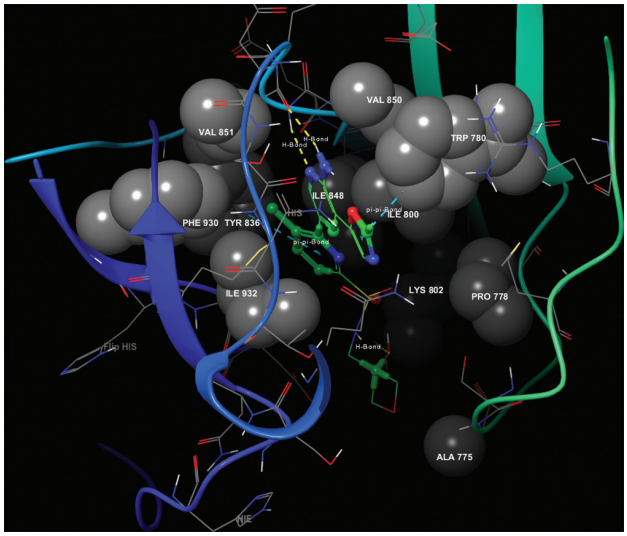


Figure 13. Hydrophobic encapsulation of T85 among several 6PYS protein residues.

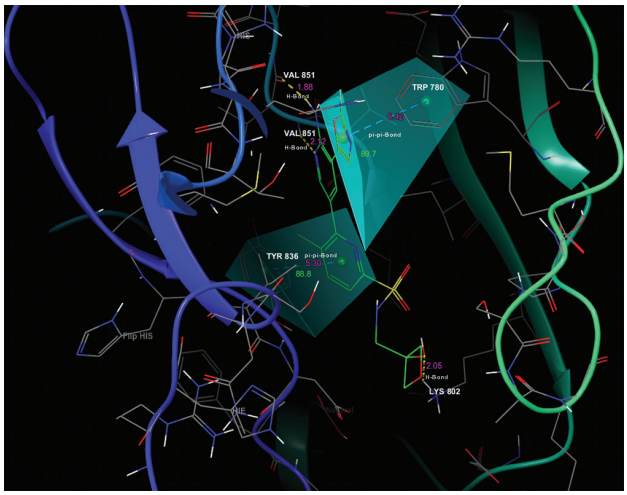


Figure 14. Hydrogen bond and π-π stacking interactions.

molecular conformations. Therefore, they are particularly important in biological systems.[94] Hence, numerous biological, chemical, and physical systems and processes are impacted by hydrogen bond interactions, which are essential in controlling the structure, characteristics, and activities of biomolecules.[95] On the other hand, π-π stacking interactions are a form of attractive and non-destructive non-covalent interactions between aromatic rings, which encompass π bonds, and whose contributions in a supramolecular assembly are vital for understanding the intrinsic nature of receptor-ligand binding domain.[96,97]

The hydrogen bond contributions between T85 and 6PYS are shown in Figure 14. The hydrogen bond interactions developed in the binding pocket of the receptor were

characterized by bond distances between certain atoms of T85 and side chains of the target, including VAL 850, VAL 851, and LYS 802. The architecture of the hydrogen bond interactions includes the H-N•••H between T85 and VAL 851 residue at a bond distance of 2.12 Å; the C-O•••H between VAL 851 residue and T85 at a bond distance of 1.88 Å; and the N-H•••O between amino acid residue LYS 802 and T85 at a bond distance of 2.05 Å. The architecture suggests that the hydrogen bonds formed within the computed bond distances indicate strong hydrogen bonds necessary for the effective binding of T85.

Figure 14 also presents another type of non-covalent interaction in our simulation, revealed as π-π stacking interactions. The distance and angle of the intermolecular contribution of π-π stacking interaction between the planes of the fluorine-associated oxazole aromatic ring of T85 and the indole ring of TRP 780 side-chain was computed at 5.42 Å and 89.7°. More so, another π-π stacking interaction existed at a different space within the binding site. This interaction manifested between the hydroxyphenyl ring along the side-chain of TYR 836 and the pyridine ring of T85 at a distance of 5.30 Å and an angle of 88.8°. Both π-π stacking interactions were coordinated in T-shaped configurations within the binding site of the 6PYS protein complex. This type of configuration involves quadrupole interactions among delocalized electrons in π-orbitals, leading to enhanced intermolecular electrostatic interactions.[97,98] Therefore, having multiple π-π stacking interactions between aromatic compounds between T85 and 6PYS suggested increased binding affinity and specificity of the ligand-protein interaction, which provided additional stabilizing forces and shape complementarity.

3.6.3. ADMET Investigations

To investigate the metabolic stability and safety profiles of T85, we assessed its clinical relevance using the QikProp program in Schrödinger Maestro. For this, we utilized all generated descriptors by QikProp, relevant for ADMET predictions to evaluate the pharmacokinetic profiles of T85. Table 5 displays a host of descriptors[99] computed by QikProp with their optimal tolerance ranges following

Table 5. Evaluation of the pharmacokinetic attributes of T85 ligand from Schrödinger's ADMET descriptors[99]

Descriptors	Description	Range or recommended values	T85
#stars	The quantity of attribute or descriptor values for known medications that are outside the 95% range of comparable values. A molecule with more stars than a few indicates that it is less drug-like than the other molecule. The quantity, #rotor, donorHB, accptHB, glob, QPpolrz, PlogPC16, QPlogPoct, QPlogPw, QPlogPo/w, logS, QPLogKhsa, QPlogBB, #metabol, and the following attributes and descriptors are taken into account while determining the #stars.	0 – 5	0
#amine	Quantity of amine groups that are not conjugated.	0 – 1	0
#amidine	Quantity of amidine and guanidine groups.	0	0
#acid	Count of groups of carboxylic acid.	0 – 1	0
#amide	Count of non-conjugated amide groups.	0 – 1	0
#rotor	Number of rotatable bonds that are non-trivial (not CX3) and non-hindered (not amide, tiny ring, or alkene).	0 – 15	7
#rtvFG	Reactive functional group count. These groups have the potential to cause toxicity, decomposition, or reactivity issues *in vivo,* as well as false-positive results in HTS assays.	0 – 2	0
CNS	CNS is measured on a scale ranging from -2 (indicating no activity) to+2 (indicating full activation).	−2 (inactive), +2 (active)	-2
Mol_MW (in Daltons, Da)	The molecule's molecular weight.	130.0 – 725.0	435.429
dipole†	The molecule's calculated dipole moment.	1.0 – 12.5	10.969
SASA	SASA (total solvent accessible surface area), measured in square angstroms with a 1.4 Å radius probe.	300.0 – 1000.0	648.546
FOSA	The SASA's hydrophobic component (saturated carbon and attached hydrogen).	0.0 – 750.0	208.350
FISA	The SASA's hydrophilic component (SASA on N, O, H on heteroatoms, and carbonyl C).	7.0 – 330.0	205.152

(Cont'd..)

Table 5. ***(Continued)***

Descriptors	Description	Range or recommended values	T85
PISA	The SASA's π (carbon and attached hydrogen) component.	0.0 – 450.0	182.467
WPSA	Weakly polar SASA component (halogens, P, and S).	0.0 – 175.0	52.577
Volume	The volume of solvent accessible in cubic angstroms, measured with a 1.4 Å radius probe.	500.0 – 2000.0	1179.394
donorHB	The approximate quantity of hydrogen bonding that the solute would provide to the molecules of water in an aqueous solution. The values represent averages calculated across multiple combinations.	0.0 – 6.0	3
accptHB	The approximate quantity of hydrogen bonding that a solute in an aqueous solution would accept from water molecules.	2.0 – 20.0	10.75
dip^2/V†	The molecule volume is split by the square of the dipole moment. This is the crucial term in the Kirkwood-Onsager equation, which calculates the free energy of solvation of a dipole with volume V.	0.00 – 0.13	0.102009
AC×DN^0.5/SA	Cohesive interaction index in solids. This term embodies the correlation (accptHB($\sqrt{donerHB}$))/(SA)	0.0 – 0.05	0.028710
glob	Globularity descriptor, $(4\pi r^2)$/(SASA); r=Radius of a sphere; volume equal to the molecular volume. Globularity is equal to 1.0 for a spherical molecule.	0.75 – 0.95	0.832396
QPpolrz	Polarizability prediction in cubic angstroms.	13.0 – 70.0	38.293
QPlogPC16	Predicted hexadecane/gas partition coefficient.	4.0 – 18.0	12.254
QPlogPoct‡	Predicted octanol/gas partition coefficient.	8.0 – 35.0	24.625
QPlogPw	Predicted water/gas partition coefficient.	−2.0 – 6.5	17.373
QPlogPo/w	Predicted octanol/water partition coefficient.	8.0 – 35.0	1.122
QPlogS	Aqueous solubility prediction, log S; S in mol dm^{-3} is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid.	−6.5 – 0.5	-3.633
CIQPlogS	Conformation-independent predicted aqueous solubility, log S. S in mol dm^{-3}=concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid.	−6.5 – 0.5	-4.59
QPlogHERG	Predicted IC_{50} value for blockage of HERG K^+channels.	Concern ≤- 5	-5.114
QPPCaco	Anticipated apparent permeability of Caco-2 cells in nm/sec. A model for the gut-blood barrier is provided by Caco-2 cells. The QikProp predictions pertain to passive transportation.	<25 poor, >500 great	112.317
QPlogBB	Estimated brain/blood partition coefficient. Note: QikProp predictions are for orally delivered drugs.	−3.0 – 1.2	-1.703
QPPMDCK	Estimated apparent permeability in nm/sec for MDCK cells. It is thought that MDCK cells are an excellent representation of the blood-brain barrier. For non-active transport, QikProp predictions are made.	<25 poor, >500 great	90.367
QPlogKp	Predicted skin (dermal) permeability, log Kp	-8.0 –-1.0	-3.986
IP (eV)†	PM3 calculated ionization potential (negative of HOMO energy).	7.9 – 10.5	8.667
EA (eV)†	PM3 calculated electron affinity (negative of LUMO energy).	−0.9 – 1.7	1.343
#metab‡	The quantity of possible metabolic reactions.	1 – 8	3
QPlogKhsa	Prediction of binding to human serum albumin.	−1.5 – 1.5	-0.44
HumanOralAborption	Human oral absorption predicted at the qualitative level: 1, 2, or 3 for low, medium, or high. The evaluation employs a set of knowledge-based guidelines that include determining appropriate values for logP, solubility, cell permeability, metabolite count, number of rotatable bonds, and percent human oral absorption.		3

(Cont'd..)

Table 5. ***(Continued)***

Descriptors	Description	Range or recommended values	T85
PercentHumanOralAbsorption	On a 0 – 100% scale, the predicted oral absorption by humans. An analysis of multiple linear regression data provides the basis for the forecast. HumanOralAbsorption and this attribute measure the same thing, hence their correlation is typically good.	>80% is high, <25% is poor	70.217
SAFluorine	The solvent-accessible surface area of fluorine atoms.	0.0 – 100.0	51.967
SAamideO	The solvent-accessible surface area of amide oxygen atoms.	0.0 – 35.0	0
PSA	Van der Waals surface area of polar nitrogen and oxygen atoms and carbonyl carbon atoms.	7.0 – 200.0	141.097
#NandO	Number of nitrogen and oxygen atoms.	2 – 15	10
RuleOfFive	Number of violations of Lipinski's rule of five. The rules are: mol_MW<500, QPlogPo/w < 5, donorHB ≤ 5, accptHB ≤ 10. Compounds that satisfy these rules are considered drug like. (The "five" refers to the limits, which are multiples of 5.).	Maximum is 4	0
RuleOfThree	Number of violations of Jorgensen's rule of three. The three rules are QPlogS > -5.7, QP PCaco>22 nm/s, # Primary Metabolites<7. Compounds with fewer (and preferably no) violations of these rules are more likely to be orally available.	Maximum is 3	0
#ringatoms	Number of atoms in a ring.		21
#in34	Number of atoms in 3- or 4-membered rings.		4
#in56	Number of atoms in 5- or 6-membered rings.		17
#noncon	Number of ring atoms not able to form conjugated aromatic systems (e.g., sp^3 C).		3
#nonHatm	Number of heavy atoms (non-hydrogen atoms).		30
Jm	Predicted maximum transdermal transport rate, Kp×MW×S ($\mu g\ cm^{-2}\ hr^{-1}$). The parameters (K_p and S) are obtained from the aqueous solubility and skin permeability, QPlog K_p and QPlogS.		0.01

95% of known drugs gleaned from a retrospective evaluation of drugs by the World Drug Index (WDI) to serve as a baseline for molecules. The drug-likeness and pharmacokinetic attributes of a compound are important considerations in drug development to guarantee that the molecule exhibits favorable characteristics for absorption, distribution, metabolism, excretion, and toxicity properties while possessing appropriate physicochemical properties for potential therapeutic use. Interestingly, T85 complied with the ADMET descriptors put forward by QikProp, and results were within optimal tolerance bounds, especially Lipinski's rule of five, Jorgensen's rule of three, the blood/brain barrier (BBB) penetration, access to the central nervous system, dermal penetration parameter ($QP\log K_p$), blockage of human ether-a-go-go-related gene potassium ion (hERG K^+), and so on, even the Verber rule.[100] The result culminated in the suitability of T85's development as a promising inhibitor for PI3K-*α*.

4. Conclusion

The enzyme PI3K-*α* plays a crucial role in regulating the growth, division, and survival of cells by phosphorylating specific lipids in the cell membrane, thereby activating downstream signaling pathways. It is connected to the frequently dysregulated PI3K/AKT/mTOR pathway in cancer. Increased signaling through this route can result from abnormal activation or mutations in PI3K-*α*, which supports cell survival, proliferation, and resistance to cell death. Such an imbalance may contribute to the development and spread of cancer. Mutations in the *PIK3CA* gene, which codes for the PI3K-*α* catalytic subunit, are frequently observed in cancer types such as lung, ovarian, colorectal, and breast cancers, contributing to increased PI3K activity and abnormal signaling that promotes cancer. To effectively target PI3K-*α* and other members of the PI3K/AKT/mTOR pathway, it is critical to rationally design potent multitarget inhibitors to address the heterogeneity and complexity of cancer cells, as well as the emergence of drug resistance and toxicity. In this study, we obtained the human PI3K-*α* protein (6PYS) co-crystallized with a ligand (PJ5) and selected PI3K-*α* inhibitors from protein and binding databases, respectively. The dataset comprised a congeneric series of 3D structures of selected PI3K-*α* inhibitors, which were

refined to improve the quality, accuracy, consistency, and completeness of the data while reducing noise, errors, outliers, and inconsistencies. A virtual screening workflow coupled with molecular docking evaluated the binding potentials of the refined ligands within the binding pocket 6PYS. Furthermore, the field-based QSAR technique elucidated the potential interactions between the ligands and probe atoms having specific attributes such as hydrophobic, electrostatic, hydrogen bond donor, or acceptor at each point in a 3D grid, highlighting key features influencing ligand activity, potency, and selectivity within the target binding site. In addition, the field-based QSAR model captured the shape, size, and flexibility of the ligands, along with their complementarity with the target. Multivariate PLS regression was employed to evaluate the relationship between the ligand molecular fields (independent variables) and the biological activity (dependent variable). Four PLS factors revealed important information from the data, indicating the influence of variance from the data on the 3D-QSAR model. A rational drug design approach was employed to design a molecule that could bind to the target and modulate its functions. The clinical relevance and ADMET properties of the newly designed molecule were predicted and evaluated according to recognized standards, including Lipinski's rule of five and Jorgensen's rule of three. Overall, the study proposes the use of advanced and reliable computational techniques to design novel and potent inhibitors of PI3K-α, a potential target for cancer immunotherapy, along with the repurposing of existing drugs for new indications.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Kevin Tochukwu Dibia, Sandra Nneka Van-Dibia
Data curation: Kevin Tochukwu Dibia
Formal Analysis: Kevin Tochukwu Dibia
Investigation: Kevin Tochukwu Dibia, Sandra Nneka Van-Dibia, Philomena Kanwulia Igbokwe
Methodology: Kevin Tochukwu Dibia
Writing—Original Draft: Kevin Tochukwu Dibia
Writing—Review & Editing: Kevin Tochukwu Dibia, Sandra Nneka Van-Dibia

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Python codes for data preprocessing are available on request.

References

1. Toppmeyer DL, Press MF. Testing considerations for phosphatidylinositol-3-kinase catalytic subunit Alpha as an emerging biomarker in advanced breast cancer. *Cancer Med*. 2020;9:6463-6472.

   doi: 10.1002/cam4.3278

2. Mansour MA, Lasheen DS, Gaber HM, Abouzid KAM. Elaborating piperazinyl-furopyrimidine based scaffolds as phosphoinositol-3-kinase enzyme alpha (PI3Kα) inhibitors to combat pancreatic cancer. *RSC Adv*. 2020;10(53):32103-32112.

   doi: 10.1039/D0RA06428A

3. Castillo JJ, Furman M, Winer ES. CAL-101: A phosphatidylinositol-3-kinase p110-delta inhibitor for the treatment of lymphoid malignancies. *Expert Opin Investig Drugs*. 2012;21:15-22.

   doi: 10.1517/13543784.2012.640318

4. Akinleye A, Avvaru P, Furqan M, Song Y, Liu D. Phosphatidylinositol 3-kinase (PI3K) inhibitors as cancer therapeutics. *J Hematol Oncol*. 2013;6:88.

   doi: 10.1186/1756-8722-6-88

5. Burris HA. Overcoming acquired resistance to anticancer therapy: Focus on the PI3K/AKT/mTOR pathway. *Cancer Chemother Pharmacol*. 2013;71:829-842.

   doi: 10.1007/s00280-012-2043-3

6. Popova NV, Jücker M. The role of mTOR signaling as a therapeutic target in cancer. *Int J Mol Sci*. 2021;22(4):1743.

   doi: 10.3390/ijms22041743

7. Liu P, Cheng H, Roberts TM, Zhao JJ. Targeting the phosphoinositide 3-kinase pathway in cancer. *Nat Rev Drug Discov*. 2009;8(8):627-644.

   doi: 10.1038/nrd2926

8. Saal LH, Johansson P, Holm K, *et al*. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proc Natl Acad Sci U S A*. 2007;104(18):7564-7569.

   doi: 10.1073/pnas.0702507104

9. Millis SZ, Ikeda S, Reddy S, Gatalica Z, Kurzrock R.

Landscape of phosphatidylinositol-3-kinase pathway alterations across 19 784 diverse solid tumors. *JAMA Oncol*. 2016;2(12):1565-1573.

doi: 10.1001/jamaoncol.2016.0891

10. Vivanco I, Sawyers C. The phosphatidylinositol 3-Kinase-AKT pathway in human cancer. *Nat Rev Cancer*. 2002;2:489-501.

doi: 10.1038/nrc839

11. Markman B, Dienstmann R, Tabernero J. Targeting the PI3K/Akt/mTOR pathway--beyond rapalogs. *Oncotarget*. 2010;1:530-543.

doi: 10.18632/oncotarget.188

12. Courtney KD, Corcoran RB, Engelman JA. The PI3K pathway as drug target in human cancer. *J Clin Oncol*. 2010;28:1075-1083.

doi: 10.1200/JCO.2009.25.3641

13. Vanhaesebroeck B, Leevers SJ, Panayotou G, Waterfield MD. Phosphoinositide 3-kinases: A conserved family of signal transducers. *Trends Biochem Sci*. 1997;22:267-272.

doi: 10.1016/S0968-0004(97)01061-X

14. Reif K, Okkenhaug K, Sasaki T, Penninger JM, Vanhaesebroeck B, Cyster JG. Cutting edge: Differential roles for phosphoinositide 3-Kinases, p110gamma and p110delta, in lymphocyte chemotaxis and homing. *J Immunol*. 2004;173:2236-2240.

doi: 10.4049/jimmunol.173.4.2236

15. Williams R, Berndt A, Miller S, Hon WC, Zhang X. Form and flexibility in phosphoinositide 3-kinases. *Biochem Soc Trans*. 2009;37(4):615-626.

doi: 10.1042/BST0370615

16. Engelman JA, Luo J, Cantley LC. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nat Rev Genet*. 2006;7:606-619.

doi: 10.1038/nrg1879

17. Katso R, Okkenhaug K, Ahmadi K, White S, Timms J, Waterfield MD. Cellular function of phosphoinositide 3-Kinases: Implications for development, homeostasis, and cancer. *Annu Rev Cell Dev Biol*. 2001;17:615-675.

doi: 10.1146/annurev.cellbio.17.1.615

18. Raffoul JJ, Kucuk O, Sarkar FH, Hillman GG. Dietary agents in cancer chemoprevention and treatment. *J Oncol*. 2012;2012:749310.

doi: 10.1155/2012/749310

19. Kasala ER, Bodduluru LN, Barua CC, Sriram CS, Gogoi R. Benzo(a)pyrene induced lung cancer: Role of dietary phytochemicals in chemoprevention. *Pharmacol Rep*. 2015;67:996-1009.

doi: 10.1016/j.pharep.2015.03.004

20. Backer JM. The regulation and function of class III PI3Ks: Novel roles for Vps34. *Biochem J*. 2008;410:1-17.

doi: 10.1042/BJ20071427

21. Yadav RR, Guru SK, Joshi P, *et al*. 6-Aryl substituted 4-(4-cyanomethyl) phenylamino quinazolines as a new class of isoform-selective PI3K-alpha inhibitors. *Eur J Med Chem*. 2016;122:731-743.

doi: 10.1016/j.ejmech.2016.07.006

22. Hauptman N, Jevšinek D, Glavač D. Noncoding RNA alterations in cancer molecular pathways. In: *Cancer and Noncoding RNAs*. Cambridge: Academic Press; 2018. p. 247-268.

doi: 10.1016/b978-0-12-811022-5.00014-0

23. Furet P, Guagnano V, Fairhurst RA, *et al*. Discovery of NVP-BYL719 a potent and selective phosphatidylinositol-3 kinase alpha inhibitor selected for clinical evaluation. *Bioorg Med Chem Lett*. 2013;23(13):3741-3748.

doi: 10.1016/j.bmcl.2013.05.007

24. Nunnery SE, Mayer IA. Management of toxicity to isoform α-specific PI3K inhibitors. *Ann Oncol* . 2019;30:x21-x26.

doi: 10.1093/annonc/mdz440

25. Utermark T, Rao T, Cheng H, *et al.* The p110α and p110β isoforms of PI3K play divergent roles in mammary gland development and tumorigenesis. *Genes Dev*. 2012;26(14):1573-1586.

doi: 10.1101/gad.191973.112

26. Sabbah DA, Simms NA, Brattain MG, Vennerstrom JL, Zhong H. Biological evaluation and docking studies of recently identified inhibitors of phosphoinositide-3-kinases. *Bioorg Med Chem Lett*. 2012;22(2):876-880.

doi: 10.1016/j.bmcl.2011.12.044

27. Bhaskar BV, Rammohan A, Babu TM, *et al*. Molecular insight into isoform specific inhibition of PI3K-α and PKC-η with dietary agents through an ensemble pharmacophore and docking studies. *Sci Rep*. 2021;11(1):12150.

doi: 10.1038/s41598-021-90287-3

28. Ali AM, Makki AA, Ibraheem W, *et al*. Design of novel phosphatidylinositol 3-kinase inhibitors for non-hodgkin's lymphoma: Molecular docking, molecular dynamics, and density functional theory studies on gold nanoparticles. *Molecules*. 2023;28(5):2289.

doi: 10.3390/molecules28052289

29. Dirican E, Akkiprik M. Phosphatidylinositol 3-kinase regulatory subunit 1 and phosphatase and tensin homolog as therapeutic targets in breast cancer. *Tumour Biol*. 2017;39:101042831769552.

doi: 10.1177/1010428317695529

30. Saldanha SN, Tollefsbol TO. The role of nutraceuticals in chemoprevention and chemotherapy and their clinical

outcomes. *J Oncol*. 2012;2012:192464.

doi: 10.1155/2012/192464

31. Mirzaei S, Ghodsi R, Hadizadeh F, Sahebkar A. 3D-QSAR-Based pharmacophore modeling, virtual screening, and molecular docking studies for identification of tubulin inhibitors with potential anticancer activity. *Biomed Res Int*. 2021;2021:6480804.

doi: 10.1155/2021/6480804

32. Wasukan N, Kuno M, Maniratanachote R. Molecular docking as a promising predictive model for silver nanoparticle-mediated inhibition of cytochrome P450 enzymes. *J Chem Inf Model*. 2019;59:5126-5134.

doi: 10.1021/acs.jcim.9b00572

33. Dibia KT, Igbokwe PK, Ezemagu GI, Asadu CO. Exploration of the quantitative structure-activity relationships for predicting cyclooxygenase-2 inhibition bioactivity by machine learning approaches. *Results Chem*. 2022;4:100272.

doi: 10.1016/j.rechem.2021.100272

34. Chen X, Liu M, Gilson MK. BindingDB: A web-accessible molecular recognition database. *Comb Chem High Throughput Screen*. 2001;4(8):719-725.

doi: 10.2174/1386207013330670

35. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res*. 2007;35:D198-D201.

doi: 10.1093/nar/gkl999

36. Bateman A, Martin MJ, Orchard S, *et al*. UniProt: The universal protein knowledgebase in 2023. *Nucleic Acids Res*. 2023;51(D1):D523-D531.

doi: 10.1093/nar/gkac1052

37. Fan C, Chen M, Wang X, Wang J, Huang B. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Front Energy Res*. 2021;9:652801.

doi: 10.3389/fenrg.2021.652801

38. Alasadi SA, Bhaya WS. Review of data preprocessing techniques in data mining. *J Eng Appl Sci*. 2017;12(16):4102-4107.

doi: 10.3923/jeasci.2017.4102.4107

39. Kanwar G, Kumar A, Mahajan A. Open source software tools for computer aided drug design. *Int J Res Pharm Sci*. 2018;9(1):86-95.

doi: 10.26452/ijrps.v9i1.1191

40. Lindahl ER. Molecular dynamics simulations. *Methods Mol Biol*. 2008;443:3-23.

doi: 10.1007/978-1-59745-177-2_1

41. McGreig JE, Uri H, Antczak M, Sternberg MJE, Michaelis M, Wass MN. 3DLigandSite: Structure-based prediction of protein-ligand binding sites. *Nucleic Acids Res*. 2022;50(W1):W13-W20.

doi: 10.1093/nar/gkac250

42. Mukhopadhyay A, Borkakoti N, Pravda L. Finding enzyme cofactors in Protein Data Bank. *Bioinformatics*. 2019;35(18):3510-3511.

doi: 10.1093/bioinformatics/btz115

43. Torrance JW, MacArthur MW, Thornton JM. Evolution of binding sites for zinc and calcium ions playing structural roles. *Proteins*. 2008;71(2):813-830.

doi: 10.1002/prot.21741

44. Zhao J, Cao Y, Zhang L. Exploring the computational methods for protein-ligand binding site prediction. *Comput Struct Biotechnol J*. 2020;18:417-426.

doi: 10.1016/j.csbj.2020.02.008

45. Roche DB, Brackenridge DA, McGuffin LJ. Proteins and their interacting partners: An introduction to protein-ligand binding site prediction methods. *Int J Mol Sci*. 2015;16:29829-29842.

doi: 10.3390/ijms161226202

46. Kudo G, Hirao T, Yoshino R, Shigeta Y, Hirokawa T. Pocket to concavity: A tool for the refinement of protein-ligand binding site shape from Alpha spheres. *Bioinformatics*. 2023;39(4):btad212.

doi: 10.1093/bioinformatics/btad212

47. Heifetz A, Katchalski-Katzir E, Eisenstein M. Electrostatics in protein-protein docking. *Protein Sci*. 2009;11(3):571-587.

doi: 10.1110/ps.26002

48. Palma PN, Krippahl L, Wampler JE, Moura JJ. Bigger: A new (soft) docking algorithm for predicting protein interactions. *Proteins*. 2000;39(4):372-384.

doi: 10.1002/(SICI)1097-0134(20000601)39:4<372:AID-PROT100>3.0.CO;2-Q

49. Adiyaman R, McGuffin LJ. Methods for the refinement of protein structure 3D models. *Int J Mol Sci*. 2019;20:2301.

doi: 10.3390/ijms20092301

50. Madhavi G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des*. 2013;27(3):221-234.

doi: 10.1007/s10822-013-9644-8

51. Lerm JA, Meyners C, Christmann A, *et al*. Binding pocket stabilization by high-throughput screening of yeast display libraries. *Front Mol Biosci*. 2022;9:1023131.

doi: 10.3389/fmolb.2022.1023131

52. Müller I. Guidelines for the successful generation of protein-

ligand complex crystals. *Acta Crystallogr D Struct Biol*. 2017;73:79-92.

doi: 10.1107/S2059798316020271

53. Guterres H, Park SJ, Jiang W, Im W. Ligand-binding-site refinement to generate reliable holo protein structure conformations from apo structures. *J Chem Inf Model*. 2020;61(1):535-546.

doi: 10.1021/acs.jcim.0c01354

54. Shin WH, Kumazawa K, Imai K, Hirokawa T, Kihara D. Current challenges and opportunities in designing protein-protein interaction targeted drugs. *Adv Appl Bioinform Chem*. 2020;13:11-25.

doi: 10.2147/AABC.S235542

55. Halgren TA. Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model*. 2009;49(2):377-389.

doi: 10.1021/ci800324m

56. De Wolf FA, Brett GM. Ligand-binding proteins: Their potential for application in systems for controlled delivery and uptake of ligands. *Pharmacol Rev*. 2000;52(2):207-236.

57. Sriramulu DK, Lee SG. Combinatorial effect of ligand and ligand-binding site hydrophobicities on binding affinity. *J Chem Inf Model*. 2020;60(3):1678-1684.

doi: 10.1021/acs.jcim.9b01143

58. Shin YS, Lee JY, Noh S, *et al*. Discovery of cyclic sulfonamide derivatives as potent inhibitors of SARS-CoV-2. *Bioorg Med Chem Lett*. 2021;31:127667.

doi: 10.1016/j.bmcl.2020.127667

59. Wan Y, Fang G, Chen H, Deng X, Tang Z. Sulfonamide derivatives as potential anti-cancer agents and their SARs elucidation. *Eur J Med Chem*. 2021;226:113837.

doi: 10.1016/j.ejmech.2021.113837

60. Hamed FM, Hassan BA, Abdulridha MM. The antitumor activity of sulfonamides derivatives: Review. *Int J Pharm Res*. 2020;12:2512.

doi: 10.31838/ijpr/2020.SP1.390

61. Ghorab MM, Alsaid MS, El-Gaby MSA, Safwat NA, Elaasser MM, Soliman AM. Biological evaluation of some new N-(2,6-dimethoxypyrimidinyl) thioureido benzenesulfonamide derivatives as potential antimicrobial and anticancer agents. *Eur J Med Chem*. 2016;124:299-310.

doi: 10.1016/j.ejmech.2016.08.060

62. Bano S, Javed K, Ahmad S, Rathish IG, Singh S, Alam MS. Synthesis and biological evaluation of some new 2-pyrazolines bearing benzene sulfonamide moiety as potential anti-inflammatory and anti-cancer agents. *Eur J Med Chem*. 2011;46(12):5763-5768.

doi: 10.1016/j.ejmech.2011.08.015

63. Alaoui S, Dufies M, Driowya M, *et al*. Synthesis and anti-cancer activities of new sulfonamides 4-substituted-triazolyl nucleosides. *Bioorg Med Chem Lett*. 2017;27(9):1989-1992.

doi: 10.1016/j.bmcl.2017.03.018

64. Stokes SS, Albert R, Buurman ET, *et al*. Inhibitors of the acetyltransferase domain of N-acetylglucosamine-1-phosphate-uridylyltransferase/glucosamine-1-phosphate-acetyltransferase (GlmU). Part 2: Optimization of physical properties leading to antibacterial aryl sulfonamides. *Bioorg Med Chem Lett*. 2012;22(23):7019-7023.

doi: 10.1016/j.bmcl.2012.10.003

65. Konda S, Raparthi S, Bhaskar K, *et al*. Synthesis and antimicrobial activity of novel benzoxazine sulfonamide derivatives. *Bioorg Med Chem Lett*. 2015;25(7):1643-1646.

doi: 10.1016/j.bmcl.2015.01.026

66. Durgun M, Turkmen H, Zengin G, Zengin H, Koyunsever M, Koyuncu I. Synthesis, characterization, *in vitro* cytotoxicity and antimicrobial investigation and evaluation of physicochemical properties of novel 4-(2-methylacetamide)benzenesulfonamide derivatives. *Bioorg Chem*. 2017;70:163-172.

doi: 10.1016/j.bioorg.2016.12.007

67. Lal J, Gupta SK, Thavaselvam D, Agarwal DD. Biological activity, design, synthesis and structure activity relationship of some novel derivatives of curcumin containing sulfonamides. *Eur J Med Chem*. 2013;64:579-588.

doi: 10.1016/j.ejmech.2013.03.012

68. Ning X, Guo Y, Ma X, *et al*. Design, synthesis and pharmacological evaluation of (E)-3,4-dihydroxy styryl sulfonamides derivatives as multifunctional neuroprotective agents against oxidative and inflammatory injury. *Bioorg Med Chem*. 2013;21(17):5589-5597.

doi: 10.1016/j.bmc.2013.05.043

69. Abbas A, Murtaza S, Tahir MN, *et al*. Synthesis, antioxidant, enzyme inhibition and DNA binding studies of novel N-benzylated derivatives of sulfonamide. *J Mol Struct*. 2016;1117:269-275.

doi: 10.1016/j.molstruc.2016.03.066

70. Chandna N, Kumar S, Kaushik P, *et al*. Synthesis of novel celecoxib analogues by bioisosteric replacement of sulfonamide as potent anti-inflammatory agents and cyclooxygenase inhibitors. *Bioorg Med Chem*. 2013;21(15):4581-4590.

doi: 10.1016/j.bmc.2013.05.029

71. Lu XY, Wang ZC, Ren SZ, Shen FQ, Man RJ, Zhu HL. Coumarin sulfonamides derivatives as potent and selective COX-2 inhibitors with efficacy in suppressing cancer proliferation and metastasis. *Bioorg Med Chem Lett*. 2016;26(15):3491-3498.

doi: 10.1016/j.bmcl.2016.06.037

72. Naim MJ, Alam O, Alam MJ, *et al*. Design, synthesis and

molecular docking of thiazolidinedione based benzene sulphonamide derivatives containing pyrazole core as potential anti-diabetic agents. *Bioorg Chem*. 2018;76:98-112.

doi: 10.1016/j.bioorg.2017.11.010

73. Mechelke M, Habeck M. Robust probabilistic superposition and comparison of protein structures. *BMC Bioinformatics*. 2010;11:363.

doi: 10.1186/1471-2105-11-363

74. Darling HS. Do you have a standard way of interpreting the standard deviation? A narrative review. *Cancer Res Stat Treat*. 2022;5(4):728-733.

doi: 10.4103/crst.crst_284_22

75. Lee DK, In J, Lee S. Standard deviation and standard error of the mean. *Korean J Anesthesiol*. 2015;68(3):220-223.

doi: 10.4097/kjae.2015.68.3.220

76. Zhang D. A coefficient of determination for generalized linear models. *Am Stat*. 2017;71(4):310-316.

doi: 10.1080/00031305.2016.1256839

77. Prinz F, Schlange T, Asadullah K. Believe it or not: How much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011;10:712.

doi: 10.1038/nrd3439-c1

78. Jawarkar RD, Bakal RL, Khatale PN, *et al*. QSAR, pharmacophore modeling and molecular docking studies to identify structural alerts for some nitrogen heterocycles as dual inhibitor of telomerase reverse transcriptase and human telomeric G-quadruplex DNA. *Futur J Pharm Sci*. 2021;7(1):231.

doi: 10.1186/s43094-021-00380-7

79. Tropsha A, Cho SJ. *Perspectives in Drug Discovery and Design*. Vol. 12. United States: Springer; 1998. p. 57-69.

doi: 10.1023/a:1017017601586

80. Consonni V, Ballabio D, Todeschini R. Comments on the definition of the Q2 Parameter for QSAR Validation. *J Chem Inf Model*. 2009;49(7):1669-1678.

doi: 10.1021/ci900115y

81. Nguyen XS, Mouaddib AI, Nguyen TP. Hierarchical gaussian descriptor based on local pooling for action recognition. *Mach Vis Appl*. 2019;30(2):321-343.

doi: 10.1007/s00138-018-0989-9

82. Raschka S, Wolf AJ, Bemister-Buffington J, Kuhn LA. Protein-ligand interfaces are polarized: Discovery of a strong trend for intermolecular hydrogen bonds to favor donors on the protein side with implications for predicting and designing ligand complexes. *J Comput Aided Mol Des*. 2018;32(4):511-528.

doi: 10.1007/s10822-018-0105-2

83. Chen D, Oezguen N, Urvil P, Ferguson C, Dann SM, Savidge TC. Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Sci Adv*. 2016;2(3):e1501240.

doi: 10.1126/sciadv.1501240

84. Bulusu G, Desiraju GR. Strong and weak hydrogen bonds in protein-ligand recognition. *J Indian Inst Sci*. 2020;100:31-41.

doi: 10.1007/s41745-019-00141-9

85. Jurczyk J, Woo J, Kim SF, Dherange BD, Sarpong R, Levin MD. Single-atom logic for heterocycle editing. *Nat Synth*. 2022;1:352-364.

doi: 10.1038/s44160-022-00052-1

86. Barnes-Seeman D, Beck J, Springer C. Fluorinated compounds in medicinal chemistry: Recent applications, synthetic advances and matched-pair analyses. *Curr Top Med Chem*. 2014;14(7):855-864.

doi: 10.2174/1568026614666140202204242

87. Chandra G, Singh DV, Mahato GK, Patel S. Fluorine-a small magic bullet atom in the drug development: Perspective to FDA approved and COVID-19 recommended drugs. *Chem Zvesti*. 2023;77, 4085-4106.

doi: 10.1007/s11696-023-02804-5

88. *U.S. Food and Drugs Administrations: Novel Drug Approvals for 2021*; 2021. Available from: https://www.fda.gov/drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products/novel-drug-approvals-2021 [Last accessed on 2023 Sep 11].

89. Pal S, Chandra G, Patel S, Singh S. Fluorinated nucleosides: Synthesis, modulation in conformation and therapeutic application. *Chem Rec*. 2022;22:e202100335.

doi: 10.1002/tcr.202100335

90. Shet H, Sahu R, Sanghvi YS, Kapdi AR. Strategies for the synthesis of fluorinated nucleosides, nucleotides and oligonucleotides. *Chem Rec*. 2022;22:e202200066.

doi: 10.1002/tcr.202200066

91. Grygorenko OO, Melnykov KP, Holovach S, Demchuk O. Fluorinated cycloalkyl building blocks for drug discovery. *ChemMedChem*. 2022;17:e202200365.

doi: 10.1002/cmdc.202200365

92. Jena S, Dutta J, Tulsiyan KD, Sahu AK, Choudhury SS, Biswal HS. Noncovalent interactions in proteins and nucleic acids: Beyond hydrogen bonding and π-stacking. *Chem Soc Rev*. 2022;51:4261-4286.

doi: 10.1039/d2cs00133k

93. Hirota S, Lin YW. Design of artificial metalloproteins/metalloenzymes by tuning noncovalent interactions. *J Biol Inorg Chem*. 2018;23:7-25.

doi: 10.1007/s00775-017-1506-8

94. Yunta MJR. It is important to compute intramolecular

hydrogen bonding in drug design? *Am J Model Optim*. 2017;5(1):24-57.

doi: 10.12691/ajmo-5-1-3

95. Gilli G, Gilli P. *The nature of the hydrogen bond: outline of a comprehensive hydrogen bond theory*. In: International Union of Crystallography Monographs on Crystallography. Oxford: Oxford Academic; 2009.

doi: 10.1093/acprof:oso/9780199558964.001.0001

96. Deng JH, Luo J, Mao YL, *et al*. Π-π stacking interactions: Non-negligible forces for stabilizing porous supramolecular frameworks. *Sci Adv*. 2020;6(2):eaax9976.

doi: 10.1126/sciadv.aax9976

97. Chen T, Li M, Liu J. Π-π stacking interaction: A nondestructive and facile means in material engineering for bioapplications. *Cryst Growth Des*. 2018;18(5):2765-2783.

doi: 10.1021/acs.cgd.7b01503

98. Martinez CR, Iverson BL. Rethinking the term "pi-stacking". *Chem Sci*. 2012;3:2191-2201.

doi: 10.1039/c2sc20045g

99. Press S. *QikProp 4.4 User Manual*. New York, NY: LLC, Schrödinger Press; 2015.

100. Zhu Y, Alqahtani S, Hu X. Aromatic rings as molecular determinants for the molecular recognition of protein kinase inhibitors. *Molecules*. 2021;26(6):1776.

doi: 10.3390/molecules26061776