

Molecular modeling studies of Pyrazolopyrimidine Derivatives as potent Cyclin Dependent Kinase-2 inhibitors

Bharath Kumar Chagaleti

SRM College of Pharmacy, SRM Institute of Science and Technology

Kathiravan Muthu K (✉ kathirak@srmist.edu.in)


Dr. A. P. J. Abdul Kalam Research Lab, SRM College of Pharmacy, SRM Institute of Science and Technology

Research Article

Keywords: QSAR, Molecular Docking, ADMET, Molecular dynamics simulations, Cyclin-dependent kinase inhibitors

Posted Date: February 5th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-3912493/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

This study addresses the pressing need for innovative cancer treatments in the face of global challenges posed by the widespread occurrence of cancer and increasing treatment resistance. The study looks at cyclin-dependent kinase-2 (CDK2) and uses a methodical computer approach to find possible anticancer compounds with pyrazole and pyrimidine structures. (QSAR) quantitative structure-activity relationship has become crucial in lead optimization over the last three decades. A set of 45 pyrazolopyrimidine derivatives with known IC₅₀ values were used to create and test models using QSARINS software. Model 4, with its high predictive performance ($R^2 = 0.9100$, $R^2_{adj} = 0.8900$, LOF = 0.0394), emerges as the most reliable. The resulting QSAR model proves stable, predictive, and robust, effectively representing the original dataset. Active molecular descriptors are identified for predicting the structure-activity relationship. We used SAR analysis and model equation parameters to create sixty compounds and tested them for their predicted bioactivity using Model 4. These compounds are a series with pyrazolopyrimidine-fused piperidine and hybrid moieties, such as methanethione (20), ethenone (20), and benzamide (20). Among the designed series, 16 compounds exhibited pIC₅₀ values exceeding 7, indicating that they were hit molecules represented as C1-C16. These obtained hit molecules undergo further screening with ADMET, molecular docking, and molecular dynamics simulations. C3 and C7, revealed in docking studies with low-energy conformations and sustained binding during simulations, consistently align their binding modes with the standard drug roscovitine. These compounds emerge as promising leads for targeting CDK2 in the development of groundbreaking cancer therapies.

1. Introduction

Cancer prevention has emerged as a prominent public health challenge in the 21st century, ranking second in global mortality after cardiovascular diseases. The pervasive impact of cancer, characterized by uncontrolled cell growth resulting from genetic mutations, affects one in five individuals worldwide. According to projections from the International Agency for Research on Cancer (IARC), the year 2022 is anticipated to witness approximately 18 million individuals grappling with malignancies, presenting a substantial threat with a mortality rate of 53.3%. Alarming projections based on IARC data indicate a troubling surge in cancer morbidity cases, expecting an increase to 29.5 million by 2040 [1]. According to data provided by the American Association for Cancer Research, approximately 1,958,310 new cases of cancer are projected to be diagnosed in the U.S. alone in the year 2023, with an anticipated mortality toll of 609,820 individuals succumbing to the disease. Despite notable advancements in the field, cancer treatment remains a formidable public health challenge necessitating targeted interventions tailored to diverse cancer types, such as carcinomas, lymphomas, sarcomas, and leukemias, both within the United States and globally. Because of compelling scientific evidence, the implementation of effective primary prevention measures holds promise for mitigating at least 40% of all cancer cases [2, 3]. As we confront this formidable public health challenge, prioritizing proactive preventive interventions becomes imperative in the pursuit of a healthier global population. Additionally, early detection of tumors represents a pivotal strategy for diminishing mortality associated with cancer. Cell cycle dysregulation is a key characteristic feature of cancer and leads to abnormal cell proliferation and a lack of ability for cells to undergo proper differentiation or apoptosis [4]. Cyclin-dependent kinases (CDKs) and their associated cyclin binding partners play crucial roles in tightly regulating the progression of the cell cycle [5]. Abnormalities in the CDK pathway have been identified in a variety of cancers [6]. CDKs have emerged as promising drug targets for cancer therapy due to their involvement in essential cellular processes such as RNA processing, proliferation, and cell survival [7]. CDK2 and CDK4 play crucial roles in facilitating the transition through restriction points and entry into S-phase in both tumorigenic and nontumorigenic cells [8]. CDK2, a pivotal member of the kinase family [9], predominantly forms complexes with cyclins A[10], B, and E, facilitating the transition from G₁ to S and G₂ phases in the cell cycle by sequentially inducing phosphorylation of the retinoblastoma protein (pRB) [11, 12]. In contrast to Cdk4/Cyclin D1 and Cdk2/Cyclin E[13], 16 has additional functions in G₁-S progression distinct from pRb phosphorylation [14, 15].

In light of recent strides in cancer therapeutics, there is a pressing demand for novel pharmaceutical agents characterized by heightened efficacy, minimal toxicity, and the ability to treat resistant variants of cancer beyond the scope of presently accessible drugs. However, there is still a need to address the diverse nature of cancer. Designing safe and selective cancer therapies requires an understanding of cell cycle regulatory mechanisms. Furthermore, the study of cancer helps unravel the intricacies of treatment resistance and recurrence, offering insights into strategies to overcome these challenges. With the global burden of cancer escalating, continuous scientific inquiry is imperative to refine our understanding of the disease, paving the way for novel advancements in prevention, early detection, and therapeutic modalities that could significantly impact patient outcomes and alleviate the societal burden of cancer.

The discovery of a new drug first starts with the identification of lead compounds, which may persist for several years and necessitate substantial financial resources. In the pursuit of designing novel cancer pharmaceutical agents with enhanced therapeutic efficacy and reduced side effects, the integration of computational techniques has emerged as a pivotal methodology [16]. The computational approach to drug discovery (CADD) has provided an enabling environment for expediting the drug discovery process, allowing for the rapid screening and synthesis of extensive compound libraries within a condensed timeframe. Computational investigations are recognized for their ability to save time and financial resources, simultaneously reducing the necessity for unnecessary animal experimentation. The difficulties associated with low effectiveness and high failure rates in drug discovery have been greatly reduced through the development of CADD. Various frequently used CADD methodologies include quantitative structure–activity relationships (QSAR), molecular docking, binding free energy analysis, molecular dynamics (MD) simulation, pharmacokinetics analysis, and other techniques.

In the contemporary landscape of rational drug design, within the realm of computational chemistry, the quantitative structure-activity relationship (QSAR) is a pivotal element in medicinal chemistry research [17]. Over the past three decades, the QSAR has evolved as a crucial tool expediting the lead optimization process [18]. This methodology involves the formulation of mathematical QSAR models to accurately predict the biological activity of novel or untested compounds. This approach serves to streamline ideas in virtual screening and lead optimization. QSAR relies on the concept that alterations in the structural characteristics of compounds lead to corresponding changes in endpoint values. For instance, the biological activity of a compound can be influenced by the removal or substitution of specific functional groups [19]. Currently, computational methods, including the 2D-QSAR algorithm combined with statistical approaches utilizing machine learning algorithms, are pivotal in the identification of novel compounds with enhanced biological activity and safety [20].

This study harnesses the power of quantitative structure–activity relationship (QSAR) analyses using state-of-the-art QSARINS software coupled with molecular docking, absorption, distribution, metabolism, excretion, and toxicity (ADMET) predictions and dynamic studies to comprehensively explore the intricate molecular landscape of pyrazolopyrimidine-based compounds. QSARINS is a computational tool that facilitates the development of a robust model capable of designing new molecules with enhanced anticancer activity [21, 22].

Most of the reported cancer drugs structurally have either pyrazole or pyrimidine moieties [23]. Pyrazolopyrimidines (PPs), which are nitrogen-fused heterocyclic rings, are recognized as bioisosteres of adenine and play crucial roles in various cellular processes [24, 25]. Extensive research has been conducted on the anticancer properties of pyrazolopyrimidine analogs [26], highlighting their ability to inhibit a variety of protein kinase enzymes [27]. Owing to their diverse pharmacological activities and structural versatility, pyrazolopyrimidine derivatives have garnered significant attention in medicinal chemistry, particularly because of their anticancer advantages [28, 29].

Certain pyrazolopyrimidines, exemplified by B-Raf kinase inhibitor I and cyclin-dependent kinase inhibitor II, have been shown to be effective at engaging specific cellular and enzymatic targets associated with dysregulated signaling [30]. Moreover, the pyrazolopyrimidine scaffold interacts with the hinge regions of various kinase enzymes [31]. Therefore, in this study, we employed 45 pyrazolopyrimidine derivatives with documented IC₅₀ values to develop QSAR models through multiple linear regression (MLR) techniques. The application of OECD guidelines for 2D-QSAR modeling ensured the validation of the developed QSAR models. The result of this investigation is a dependable and enduring *in silico* approach for predicting IC₅₀ values (mean ± standard deviation (SD)), offering a positive contribution to the anticancer drug development process by minimizing the financial and temporal resources needed to identify potential anticancer leads. The objective of this study was to create a 2D-QSAR model using known pyrazolopyrimidine derivatives and explore the impact of substitution variations on the CDK inhibitory activity of these compounds. The identification of key descriptor contributions will guide the design of new chemical entities (NCEs), which will undergo activity prediction. Additionally, docking studies will be conducted on the designed NCEs with the crystal structure of CDK2 to gain further insights. Figure 1 shows a schematic representation of the workflow employed in the present study.

This multifaceted approach, which combines QSAR, molecular docking, ADMET prediction, and dynamic studies, provides not only a holistic understanding of the structure–activity relationships of pyrazolopyrimidine derivatives but also aids in the identification and optimization of lead compounds with enhanced pharmacological potential. The outcomes of this research endeavour hold promise for advancing the development of novel cancer therapeutic agents within the realm of pyrazolopyrimidine-based pharmacophores.

2. Methodology

2.1. QSAR

2.1.1 Dataset preparation

In this research study, a set of 45 pyrazolo [3,4-d] pyrimidine scaffolds and their derivatives, accompanied by their respective IC₅₀ data, were obtained from the publication by Jin Zhang et al. in 2020 to construct a quantitative structure–activity relationship (QSAR) model [32]. All the compounds were isolated from the same research group and reported at a single time, and the structures are shown in Table 1. The computational analysis was performed utilizing QSARINS, a software user-friendly platform developed at the University of In subria [33]. The inhibitory effects of these molecules on P21-activated kinase 1 (PAK1) were assessed and are presented as IC₅₀ values (μM). The IC₅₀ values denote the potency of compound inhibition concerning biological or biochemical processes. The activity ranges widely, which further paves the way for improving the quality of compounds with higher activity. Compounds lacking specific numerical data regarding their inhibitory activity against the enzyme under study were excluded from the analysis. To address the skewed nature of the biological data, the IC₅₀ values of all the collected compounds were transformed into pIC₅₀ values using the formula described in Eq. 1.

$$pIC_{50} = 6 - (\text{LOG}_{10}(\text{IC}_{50})) \text{ (Eq. 1)}$$

2.1.2. Molecular sketching and generation of molecular descriptors

Chem Draw 16.0 software was used to construct the chemical structures and generate the SMILES IDs of all the collected compounds. To achieve a well-defined 3D molecular geometry and generate coordinates, the material was processed using Avogadro V1.2.0. The structures were then subjected to minimization through molecular mechanical force fields to obtain a conformer with the lowest global minimum energy and devoid of strain in the molecular structure. Open Babel version 3.1.1 software was used to convert the SMILES ID into the 'smi' file format. The 'smi' file format was compatible with the PaDEL-Descriptor v2.21 software [34, 35]. Molecular descriptors of each compound were calculated by using the PaDEL and Chemdes web servers [36], F which further generated a total of 2432 2-dimensional (2D) molecular descriptors (MDs) and fingerprint molecular descriptors for the 45 pyrazolo-pyrimidine derivatives.

2.1.3. Pretreatment and selection of molecular descriptors

Molecular descriptors were computed using Padel and Chemdes software (Chemopy & RD Kit [37]), encompassing 1D, 2D, and fingerprint descriptors, for a total of 2432 descriptors. Molecular descriptors were selected using a genetic algorithm as part of the variable selection process. A judicious elimination process was implemented, wherein descriptors exhibiting more than 95% constant values and more than 90% correlation among descriptors were excluded. Following this curation, a subset of 300 molecular descriptors was exclusively considered for the subsequent development of the quantitative structure–activity relationship (QSAR) model.

2.1.4 Dataset division and model development

Dataset division

The chosen molecules from the dataset underwent geometry optimization using the MMFF94 force field. The entire dataset was bifurcated into a training set and a test set comprising 75% of the molecules and a test set encompassing the remaining 25%, with further subdivision into a training set of 35 compounds and a test set of 10 compounds for model development and evaluation. The resulting compound dataset, inclusive of compound codes, IUPAC names, IC50 values, and calculated pIC50 values, is presented in Table 2. A meticulous selection of molecules for the training set ensured the representation of structural attributes and biological activity ranges. The training set was utilized for the development of the quantitative structure-activity relationship (QSAR) model, while the test set served for the validation of the established QSAR model [38–40].

QSAR Model development

Following this dataset division, the QSAR model was constructed employing the multiple linear regression (MLR) method and generally validated according to the chemometric approach [41]. This methodology substantiates the direct correlation between the dependent variable Y and the independent variable X, represented by molecular descriptors [42, 43]. In multiple linear regression (MLR) analysis, the mean of the dependent variable Y is dependent on X (descriptor). Consequently, the MLR equation can incorporate more than one independent variable (descriptor) with a single response variable [44].

$$Y = k_1x_1 + k_2x_2 + k_3x_3 + C \text{ (Eq. 1)}$$

where Y denotes the dependent variable, x denotes the independent variable, 'k's are regression coefficients for each 'x', and 'C' is a regression intercept.

2.1.5 Validation

The developed QSAR model was validated as per OECD guidelines by the following methods [45].

Goodness of fit

A total of six parameters, namely, the coefficient of determination (r^2), adjusted r^2 (r^2_{adj}), predicted r^2 (r^2_{pred}), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE), were used to assess the goodness of fit of the developed QSAR model [46]. The r^2 and r^2_{adj} were obtained after the execution of the MLR technique. Equations 2, 3, 4, and 5 were used to calculate the r^2_{pred} , MSE, RMSE, and MAE, respectively.

$$r^2_{pred} =$$

$$1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (T_i - O_i)^2} \text{ (Eq. 2)}$$

$$\text{Mean squared error} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \text{ (Eq. 3)}$$

where n = the number of datasets, Y_i = the actual value, and \hat{Y}_i = the predicted value.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \text{ (Eq. 4)}$$

where \sum = sum, P_i = predicted value, O_i = observed value, and n = sample size.

$$\text{MAE} = \left(\frac{1}{n}\right) \sum |O_i - P_i| \text{ (Eq. 5)}$$

where n = the total number of observations, \sum = the sum, O_i = the observed value, and P_i = the predicted value.

2.2 Determination of the applicability domain

The method is employed to determine outliers and influential molecules, and it strongly establishes the ability of the model to make predictions in the limited space for which it was developed. Any molecule that lies next to the domain space or range is said to be an outlier. The leveraged approach elaborates on the applicability domain of the QSAR model.

The warning leverage h^* is the limit for checking for the influential molecule. The warning leverage h^* is defined as

$$h^* = 3 \frac{(j+1)}{m}$$

where j is the number of descriptors in the built-up model and m is the number of compounds that make up the training set.

2.3 Quality assurance of the model

The fitting criteria, stability, robustness, reliability, and predictive strength of the obtained model were evaluated via various validation parameters. Generally, an acceptable QSAR model should comply with the minimum recommended statistical values that are tabulated [51–53].

2.4 Physicochemical properties

Drug likeness analysis is an important aspect of pharmaceutical research since it helps medicinal chemists create novel compounds. The failure of numerous drugs in clinical trials can be attributed to adverse ADMET characteristics and deleterious effects on biological systems [54]. The pharmacological impact of a drug is determined by its physical and chemical attributes, including molecular weight, partition coefficient, number of rotatable bonds, topological polar

surface area, hydrogen bond acceptors and donors, among other parameters. Interestingly, the topological polar surface area plays a crucial role in defining drug absorption pathways, including blood–brain barrier penetration, intestinal absorption, bioavailability, and Caco-2 permeability [55–57].

A significant portion of successfully marketed medications have favorable oral bioavailability when they meet Lipinski's rule of five, which consists of a molecular weight of less than 500, five hydrogen bond donors, ten hydrogen bond acceptors, and no more than five rotatable bonds. When assessing critical pharmacokinetic parameters for absorption, distribution, metabolism, and excretion (ADME), this rule is a valuable resource. The application of Lipinski's rule of five significantly enhances the identification of potential therapeutic compounds and aids in drug design by striking a balance between advantageous pharmacokinetic features and the likelihood of successful drug development. In the present investigation, the assessment of ADME properties was conducted utilizing the freely accessible Molinspiration online predictor, a tool designed for the evaluation of small compounds. Additionally, this approach provides information concerning the bioactivity of unknown compounds [58].

2.4.1 Bioactivity

In this study, we employed Molinspiration software to predict the bioactivity scores of a series of compounds (C1 to C16) across various biological modules, including GPCR ligands, ion channel modulators, kinase inhibitors, nuclear receptor ligands, protease inhibitors, and enzyme inhibitors. The Molinspiration bioactivity scores provide a preliminary assessment, and it is essential to validate these predictions through experimental assays.

The bioactivity score serves as a calculated parameter indicating the anticipated biological activity of a given chemical compound. This numerical value provides an estimation of the compound's potential to exhibit specific bioactive effects, such as enzyme inhibition or receptor binding. A higher score is indicative of stronger activity, reflecting the efficacy of the compound in biological interactions. Accordingly, a bioactivity score above 0.00 is indicative of a compound with substantial bioactivity, while scores ranging from -0.50 to 0.00 suggest moderate activity [59]. Conversely, compounds are classified as inactive if their scores fall below or equal to -0.50.

2.4.2 Toxicity profile

In this study, a comprehensive *in silico* toxicity assessment was conducted for sixteen compounds (C1 to C16) to assess their potential hepatotoxicity, carcinogenicity, immunotoxicity, mutagenicity, and cytotoxicity. Toxicity prediction was assessed by using the online free web tool Pro Tox-II, which primarily identified 5 toxicity categories [60, 61]. Further studies, including *in vitro* and *in vivo* experiments, are warranted to confirm the pharmacological activities of these compounds and assess their potential as drug candidates in specific therapeutic areas [62, 63].

2.5 Molecular docking

Molecular docking was used to assess potential interactions between the ligand and the active site of the target. Molecular docking employs a rigid receptor docking approach, rendering CDK2 rigid while allowing flexibility for the ligands [64, 65]. Precise molecular simulations and structural studies were conducted within the MOE using MMFF94x force field settings. X-ray crystal structures of the CDK2 enzyme (PDB ID 3PJ8 cocrystallized with roscovitine) were retrieved from the Protein Data Bank [66]. Energy minimization was completed for the obtained PDB file before beginning the docking process [67]. Each compound underwent five conformations and was given an accurate binding score. The selection of the PDB ID 3PJ8 for molecular docking investigations was based on its relevance to the CDK2 target. This choice is substantiated by the interaction of roscovitine, the cocrystallized ligand, with this protein. Additionally, the PDB entry belongs to the classification of cell cycle transferase, indicating its involvement in cell cycle processes. The chosen protein comprises one chain denoted A with a resolution of 1.96 Å. The removal of water molecules and cocrystals was executed followed by the addition of polar hydrogens and potential correction. Utilizing the site finder tool for active site prediction revealed a total of 21 active sites within the 3PJ8 protein distributed across four distinct pockets, as shown in Fig. 2. Each pocket is distinguished by its size, with pocket 1 highlighted in red, pocket 2 highlighted in yellow, pocket 3 highlighted in a subtle orange, and pocket 4 highlighted in blue. The docking scores, expressed in Kcal/mol, reflect the strength of the interaction between each compound and the target molecule. The number of interacting residues and the specific amino acid interactions are also highlighted, shedding light on the molecular basis of these interactions [68].

2.6 Molecular Dynamics Simulation Studies

The Desmond V 5.9 package from the Schrödinger LLC suite was used to run molecular dynamics simulations to elucidate alterations in the solvent system around the macromolecular complex. Using the OPLS force field, dynamic simulations of the docked complex were performed [69]. To simulate the dynamics, the complex was placed inside an orthorhombic cubic box that included TIP3P water molecules and buffers. The protein atoms were spaced out from the box corners by approximately 10 Å. To neutralize the system, complex counterions, such as Na⁺ and Cl⁻ ions, were transported through a Tobias-Klein barostat to maintain a consistent temperature of 300 K and pressure of 1 atm throughout the simulation [70]. Approximately 100 ns passed at the initiation of the NPT ensemble. The frames that were produced were compiled and examined, and their paths were explored through the use of simulation interaction diagrams [71].

3. Results and discussion

3.1 QSAR Model Information

Numerous QSAR models were generated utilizing QSARINS software; however, only the most robust model 4 is presented herein. The statistical parameters for the developed models (models 1, 2, 3, and 4) were evaluated against predefined threshold values and are shown in Table 3. Notably, Model 4 consistently demonstrates superior performance across various criteria, including higher R², lower S, and higher F values, indicating a better fit to the data. The internal validation criteria revealed robust performance, with all the models surpassing the specified thresholds for Q² LOO, R² – Q² LOO, Q² LMO, and RMSE_{cv}. Model 4 particularly excels in these metrics, suggesting its efficacy in capturing the underlying patterns within the training dataset. The external validation criteria

further affirm the reliability of Model 4, as evidenced by a higher R^2_{ext} , lower RMSE ext, and favorable values for Q^2-F1 , Q^2-F2 , Q^2-F3 , CCC ext, r^2_m , Δr^2_m , K' , and K . These results collectively indicate the robustness and predictive capability of Model 4 in external datasets. The comprehensive evaluation of the statistical parameters underscores the superior performance of Model 4, highlighting its potential as a reliable and accurate predictive model. While some models exhibited higher R^2 and Q^2 LOO values, their external validation was compromised due to the presence of numerous outliers. Typically, a QSAR model needs to undergo cross-validation to assess both its internal performance (robustness) and external performance in terms of predictive capacity. The scatter plot illustrates the experimental vs. calculated inhibitory activities of pyrazolopyrimidine derivatives, with Fig. 3 demonstrating that the predicted values closely aligned with their experimental counterparts. Model 4, developed with the mentioned options, yielded optimal statistical values but presented one outlier (compound 43 from the training set) in the Williams plot at the 2.5 standard deviation unit level, along with high RMSE values and low Q^2 LOO and Q^2 LMO values. Despite these challenges, Model 4 exhibited favorable fitting criteria with commendable validation values, surpassing the internal validation parameters of previous models.

Model 1:

$pKI = 9.7242 - S_{max23} (1.3873) - MATSm2 (1.2379)$

Model 2:

$pKI = 7.4766, SlogP_VSA5 (0.0293), SlogP_VSA5 (0.0293), fr_Ndealkylation2 (0.5116), \text{ and } slogPVSA5 (0.0424)$

Model 3:

$pKI = 10.2052 - SlogP_VSA5 (0.0397) - fr_Ndealkylation2 (0.5704) - slogPVSA5 (0.0409) - diameter (0.1756)$

Model 4:

$pKI = 11.82 - SlogP_VSA5 (0.0690) - fr_Ndealkylation2 (0.6567) + EstateVSA2 (0.0084) + CIC3 (1.5483) - slogPVSA5 (0.0393) - diameter (0.3296).$

The applicability domain of the developed QSAR model was delineated through the utilization of a Williams plot, illustrating the standardized residuals vs. leverage values shown in Fig. 4. These plots can depict whether the molecules are located in the applicability domain of the model. From the plot, the leverage values were confined within the squared region of ± 3 standard residues and warning leverage ($h^* = 0.618$). One outlier was identified, positioned outside the defined applicability domain, exhibiting a standard residue value of 3.412; however, this value fell within the warning leverage threshold ($h^* = 0.618$). These results indicate that there is no correlation and that there is a meaningful relationship between pyrazolopyrimidine derivatives and the corresponding inhibitory agents. K_{xy} and the intercorrelations among descriptors and response vs. Q^2 LMO of Model 4 are plotted in Fig. 5, which shows the LMO parameter values around the model parameters, indicating that the model is robust and stable. A discernibly low R^2_{Yscr} value indicates that the observed favorable outcome was not a result of chance correlation. To substantiate this, Y scrambling or Y randomization analysis was conducted on the training dataset, affirming that the achieved result was not fortuitous. During this procedure, the dependent variable (response value) within the dataset underwent random shuffling, while the independent descriptors remained unchanged; subsequently, multiple linear regression (MLR) analysis was performed on the permuted dataset. Fig. 6 shows the Y-scramble plot of K_{xy} vs. R^2_{Yscr} and Q^2_{Yscr} , which indicates that the correlation coefficients of Model 4 are much greater than those after endpoint scrambling and that a broken relationship can be observed between the structure and the responses.

3.2 Molecular descriptor correlation matrix

In Table 4, the molecular descriptor correlation matrix reveals pairwise correlations among the descriptors, including SlogP_VSA5, fr_Ndealkylation2, EstateVSA2, CIC3, slogPVSA5, and diameter. The values displayed represent the correlation coefficients, where positive values indicate a positive correlation, negative values signify a negative correlation, and a value of 1 along the diagonal indicates a perfect correlation with itself. The correlation coefficients between SlogP_VSA5 and fr_Ndealkylation2, EstateVSA2, CIC3, slogPVSA5, and the Diameter morphotype were -0.03837, -0.11680, 0.66568, -0.45560, and -0.46540, respectively. Similarly, fr_Ndealkylation2 was correlated with EstateVSA2, CIC3, and slogPVSA5, with correlation coefficients of 0.24828, -0.07989, -0.11019, and -0.13966, respectively. EstateVSA2 was correlated with CIC3, slogPVSA5, and diameter, with correlation coefficients of -0.17935, -0.10916, and 0.25139, respectively. The correlation coefficient between CIC3 and slogPVSA5 is -0.1248. Additionally, the SlogPVSA5 score and diameter exhibited a positive correlation with a coefficient of 0.56709. These correlation values provide insights into the interdependencies and relationships among the descriptors, which is crucial for understanding the potential impact of each descriptor on the overall model and its predictive performance.

3.3 Molecular descriptor contributions to activity

The molecular descriptors included in Model 4 provide structural information associated with the predicted bioactivity, emphasizing the crucial role of these descriptors in influencing anticancer activity.

The inhibitory activity is influenced by the following molecular descriptors:

i) SlogP_VSA5:

This particular descriptor falls into the category of 2D descriptors known as MOE-type descriptors. The method combines Estate indices and surface area contributions to outline the molecular hydrophobicity and spatial characteristics that are crucial for anticancer activity. This descriptor has a negative impact on the activity according to the final model information.

ii) **fr_Ndealkylation2:**

This constitutes an additional 2D descriptor indicating the count of tert-alicyclic amines. It has proven to be a distinctive indicator closely associated with specific molecular features intricately connected to the observed anticancer properties. This descriptor has a negative impact on the activity according to the final model information.

iii) **EstateVSA2:**

This descriptor is classified as an MOE-type descriptor originating from Estate indices and surface area contributions, representing a 2D characteristic. It is computed using information from Estate indices and surface area contributions, specifically capturing details related to electronic and spatial factors that impact anticancer activity. This descriptor has a positive impact on the activity according to the final model information.

iv) **slogPVSA5:**

SlogPVSA5 is a 2D descriptor that falls under the classification of both a complementary information content index and an MOE-type descriptor. It utilizes S Log P contributions and surface area contributions to provide valuable insights into the molecular lipophilicity and structural features that are relevant to anticancer properties. This descriptor has a negative impact on the activity according to the final model information.

v) **CIC3 (complementary information content index):**

This gene set serves as a topological descriptor associated with the TPSA. Additionally, employing CIC3 is crucial for measuring the distinctiveness and intricacy of molecular information, thereby enhancing our understanding of the molecular underpinnings of anticancer properties. This descriptor has a positive impact on the activity according to the final model information.

vi) **Diameter:**

The diameter, identified as a 2D descriptor, represents the maximum value in the distance matrix, imparting details about the molecular size. Moreover, this approach provides essential information regarding the spatial organization of molecules, providing valuable insights into the structural characteristics influencing anticancer activity. This descriptor has a negative impact on the activity according to the final model information. These findings collectively contribute to a comprehensive understanding of the molecular determinants underlying the observed anticancer effects, paving the way for targeted drug design and development. Understanding these descriptors will enrich the understanding of the molecular underpinnings of anticancer activity. Within the most accurately predicted model (Model 4), a positive correlation with anticancer activity was observed for the molecular descriptors EstateVSA2 and CIC3.

Conversely, the molecular descriptors SlogP_VSA5, fr_Ndealkylation2, Diameter, and slogPVSA5 exhibited negative associations with anticancer activity and with the predicted endpoint vs residual end point, as shown in Fig. 7. The pIC_{50} values of the dataset predicted by Model 4 along with the residuals are denoted in Table 5. Further investigations and analyses of these descriptors can guide the rational design and optimization of novel anticancer agents with enhanced efficacy and specificity.

3.4 SAR studies of dataset compounds

The SAR analysis of the dataset compounds revealed distinct trends in biological activity, shedding light on the impact of different substituents on cancer activity. In the 5th series (1 to 10), the 4-trifluoromethylphenyl, 4-fluorophenyl, and 2,3-dichlorophenyl substituents emerged as promising contributors to enhanced biological activity, with compound 5 h (4-trifluoromethylphenyl) displaying the highest observed activity. Among the 6 series (11 to 27), the 3-chlorophenyl, 2,4-difluorophenyl, and 2,4-dichlorophenyl substituents demonstrated strong bioactivity, with compounds 6 q, 6 h, and 6 j exhibiting the highest potency. In the 7th series (28 to 35), the 4-methyl and 2-chloro substituents emerged as promising contributors to enhanced biological activity, as shown by compounds 7 a and 7 i. Finally, in the 8 series (36 to 45), the 3-chlorophenyl, 4-methylphenyl, and 2,4-difluorophenyl substituents were highlighted as effective contributors to enhanced biological activity, with compound 8 f (3-chlorophenyl) displaying the highest observed activity. Overall, the analysis suggested that electron-donating groups, such as 4-methylphenyl and 4-trifluoromethylphenyl, contribute positively to enhanced biological activity, while certain electron-withdrawing groups, such as 2,6-dimethylphenyl, may have a less favorable impact.

3.4.1 5 SERIES (1 to 10):

Compound 5 h (4-trifluoromethylphenyl substitution) exhibited the highest activity, with an IC_{50} of 6.77 μM , indicating enhanced potency. Similarly, compound 5 d (4-fluorophenyl substitution) displayed an IC_{50} of 7.82 μM , suggesting increased bioactivity. Compound 5 j (2,3-dichlorophenyl substitution) contributed to increased activity, with an IC_{50} of 6.77 μM . Compounds 5 b, 5 f, and 5 i (2-methyl, 2-chloro, and 3-trifluoromethylphenyl substituents, respectively). These compounds exhibited lower activities, with IC_{50} values of 33.21 μM , 6.84 μM , and 50 μM , respectively. The 4-trifluoromethylphenyl substituent in compound 5 h was the most effective, yielding the highest observed activity. This suggests a favorable interaction between this specific substituent and the biological target. Compound 5 d, featuring a 4-fluorophenyl substituent, exhibited moderate activity, indicating the importance of this substituent in enhancing potency. Compound 5 j, with a 2,3-dichlorophenyl substituent, demonstrated increased activity, suggesting the potential impact of this substituent on bioactivity. Compounds 5 b, 5 f, and 5 i exhibited lower activities, suggesting that the 2-methyl, 2-chloro, and 3-trifluoromethylphenyl substituents, respectively, may have a less favorable impact on bioactivity than the other substituents. Among these 5 series, the 4-trifluoromethylphenyl, 4-fluorophenyl, and 2,3-dichlorophenyl substituents emerged as promising contributors to enhanced biological activity.

3.4.2 6 SERIES (11 to 27):

Compound 6q (3-chlorophenyl substitution) displayed the highest activity, with an IC_{50} of 5.17 μ m, indicating enhanced potency. Similarly, compound 6h (2,4-difluorophenyl substitution) exhibited an IC_{50} of 5.24 μ m, suggesting increased bioactivity. Similarly, the 2,4-dichlorophenyl substitution of compound 6j results in heightened activity with an IC_{50} of 5.46 μ m. Compound 6e (2-chlorophenyl substitution) shows notable activity with an IC_{50} of 7.18 μ m. Compounds 6b, 6l, and 6s (with 2,6-dimethylphenyl, naphthyl, and naphthyl substituents, respectively) exhibited comparatively lower activities, with IC_{50} values of 6.27 μ m, 7.36 μ m, and 50 μ m, respectively. Compound 6n (with 2,6-dimethylphenyl substitution) exhibited a moderate IC_{50} of 6.67 μ m. The 3-chlorophenyl substituent in compound 6q appears to be particularly effective, yielding the highest observed activity. This suggests a favorable interaction between this specific substituent and the biological target. Compounds 6h and 6j, with 2,4-difluorophenyl and 2,4-dichlorophenyl substituents, respectively, also demonstrated strong bioactivity, indicating the importance of these substituents in enhancing potency. The 2-chlorophenyl substituent in compound 6e contributes to moderate activity, demonstrating its influence on the effectiveness of the compound. Compounds 6b, 6l, and 6s, featuring 2,6-dimethylphenyl, naphthyl, and naphthyl substituents, respectively, exhibit lower activities. The bulkiness of these substituents may influence the interaction of the compound with the biological target. Test compound 6n, with a 2,6-dimethylphenyl substituent, exhibited a moderate level of activity, suggesting the potential impact of this substituent on bioactivity.

3.4.3 7 SERIES (28 to 35):

Compound 7a (4-methyl substitution) displayed the highest activity, with an IC_{50} of 5.61 μ m, indicating enhanced potency. Compound 7i (2-chloro substitution) exhibited an IC_{50} of 5.24 μ m, suggesting increased bioactivity. Compound 7b (2-methyl substitution) exhibited moderate activity with an IC_{50} of 6.26 μ m. Similarly, compound 7d (4-trimethyl substitution) demonstrated heightened activity with an IC_{50} of 8.82 μ m. Compounds 7c, 7e, 7f, and 7g (2,4,6-trimethyl, 4-fluoro, 4-bromo, and 4-trifluoromethyl) exhibited lower activities, with IC_{50} values of 6.77 μ m, 4.54 μ m, 16.4 μ m, and 5.74 μ m, respectively. The 4-methyl substitution in compound 7a was the most effective, yielding the highest observed activity. This suggests a favorable interaction between this specific substituent and the biological target. Compound 7i, with a 2-chloro substituent, also demonstrated strong bioactivity, indicating the importance of this substituent in enhancing potency. Compound 7b, featuring a 2-methyl substituent, exhibited moderate activity, suggesting that the bioactivity was balanced. The 4-trimethyl substitution in compound 7d contributes to increased activity, albeit to a lesser extent than the 4-methyl and 2-chloro substituents. Compounds 7c, 7e, 7f, and 7g, which have 2,4,6-trimethyl, 4-fluoro, 4-bromo, and 4-trifluoromethyl substituents, respectively, exhibit lower activities. The bulkiness or electronic effects of these substituents may influence the interaction of the compound with the biological target.

3.4.4 8 SERIES (36 to 45):

Compound 8f (3-chlorophenyl substitution) has the highest activity, with an IC_{50} of 1.24 μ m, indicating enhanced potency. Compound 8a (4-methylphenyl substitution) exhibited an IC_{50} of 4.17 μ m, suggesting increased bioactivity. Compound 8d (4-bromophenyl substitution) exhibited moderate activity with an IC_{50} of 4.03 μ m. Similarly, compound 8b (with a 2,6-dimethylphenyl substituent) demonstrated heightened activity with an IC_{50} of 2.16 μ m. Compound 8i (2,4-difluorophenyl substitution) contributes to increased activity, with an IC_{50} of 1.14 μ m. Compounds 8g, 8c, 8e, and 8k (2-chlorophenyl, 4-methoxyphenyl, 4-trifluoromethylphenyl, and 2-naphthyl substituents, respectively) exhibited lower activities, with IC_{50} values of 0.56 μ m, 6.37 μ m, 7.15 μ m, and 11.14 μ m, respectively. The 3-chlorophenyl substituent in compound 8f was the most effective, yielding the highest observed activity. This suggests a favorable interaction between this specific substituent and the biological target. Compound 8a, with a 4-methylphenyl substituent, also demonstrated strong bioactivity, indicating the importance of this substituent in enhancing potency. Compound 8d, featuring a 4-bromophenyl substituent, exhibited moderate activity, suggesting that the bioactivity was balanced. The 2,6-dimethylphenyl substitution in compound 8b contributes to increased activity, albeit to a lesser extent than the 3-chlorophenyl and 4-methylphenyl substituents. Compound 8i, with a 2,4-difluorophenyl substituent, demonstrated increased activity, suggesting the potential impact of this substituent on bioactivity. The SAR analysis highlights the significance of specific substituents in modulating the bioactivity of the compounds.

3.5 SAR and Molecular Descriptor contributions to the designed compounds

Based on the structure-activity relationship (SAR) analysis of the initial dataset and model equation parameters, sixty compounds were designed and assessed for predicted bioactivity using Model 4. These designed series of molecules have pyrazolopyrimidine-fused piperidine moieties with hybrid moieties (methanethione 20, ethanone 20, and benzamide 20). Within the entire designed series, 16 compounds exhibited pIC_{50} values exceeding 7, indicating that they were hit molecules, as depicted in Fig. 8. These hit molecules are denoted as C1-C16. Among the compounds in the methanethione series, 12 compounds predicted by Model 4 are hits and are represented as compounds C1 to C12. Among the ethanone series, model 4 predicted only one compound as a hit and was represented as compound C13, and in the benzamide series, model 4 predicted only 3 compounds as hits and was represented as compounds 14 to 16. Model 4-predicted pIC_{50} values for the designed compounds (hit molecules) are shown in Table 6. Molecular descriptor contributions to the designed compounds are shown in Fig. 9. The distribution of these designed compounds within the chemical space delineated by the model descriptors was systematically examined using principal component analysis (PCA).

3.6 Physicochemical properties

The physicochemical properties of the designed compounds were assessed through Molinspiration software, and the detailed results are presented in Table S1. In this study, we performed a comprehensive analysis of a series of compounds (C1 to C16) using Lipinski's Rule of Five, which is widely employed in drug discovery to assess the drug likeness and pharmacokinetic properties of small molecules, as they exhibit characteristics such as a log P value of approximately 5, a molecular weight close to 500 Daltons, approximately 10 hydrogen bond acceptors, and approximately 5 hydrogen bond donors. Deviations from these parameters might necessitate a reevaluation of a molecule's bioavailability potential. The computed log P values for all the compounds fell within the range of 0. These values suggest moderate hydrophobicity for the investigated compounds, which is generally favorable for membrane permeability, a crucial aspect of oral bioavailability. Additionally, the molecular weights of the compounds ranged from 314.27 to 429.51. All the compounds

fall within the commonly accepted range for drug-like molecules, below the 500 Dalton threshold, suggesting that they are of an appropriate size for effective absorption and distribution. Furthermore, the TPSA values varied from 69.73 to 124.18. Compounds with TPSA values less than 140 Å are generally considered to have good oral bioavailability, indicating that these compounds may be absorbed efficiently in the gastrointestinal tract.

The compounds exhibited 4 to 7 hydrogen bond acceptors and 2 to 5 hydrogen bond donors. These values are within the acceptable limits defined by Lipinski's rule, indicating favorable interactions with biological targets. The number of rotatable bonds ranged from 3 to 5, suggesting moderate flexibility in the molecular structure. This flexibility can be advantageous for binding interactions while maintaining the overall stability of the compound. Notably, the absence of violations in Lipinski's rule for all the compounds underscores their drug-like properties, rendering them promising candidates for further exploration in pharmaceutical research. These findings provide valuable insights for the systematic design and refinement of potential drug candidates, emphasizing the importance of their physicochemical attributes and adherence to established drug likeness criteria in drug development efforts.

3.6.1 BIOACTIVITY

The bioactivity scores provide insights into the potential pharmacological activities of these compounds, and the detailed results can be found in Table S2. The GPCR ligand scores ranged from 0.15 to 0.34. These scores indicate the likelihood of the compounds interacting with G protein-coupled receptors (GPCRs). Higher scores suggest a stronger potential for GPCR modulation. The ion channel module scores ranged from -0.52 to 0.08. These scores represent the compounds' potential as ion channel modulators. Positive scores suggest activation, while negative scores suggest inhibition. The range observed here indicates diverse activities across ion channels. The kinase inhibition scores ranged from 0.17 to 0.62. Higher scores in this module suggest the potential of the compounds to act as kinase inhibitors, indicating possible roles in signal transduction pathways and cell regulation. Compounds C13 and C14 had elevated scores in the kinase inhibitor and protease inhibitor modules, suggesting potential roles in cell signaling and proteolytic regulation. The nuclear receptor ligand scores ranged from -0.82 to -0.49. These scores represent the affinity of the compounds for nuclear receptors. Negative scores suggest potential antagonistic effects, while positive scores indicate agonistic activity. Protease inhibition ranged from -0.87 to -0.24. These scores suggest that these compounds may act as protease inhibitors, with lower scores indicating stronger inhibitory effects. The enzyme inhibitor scores ranged from 0.01 to 0.13. Positive scores suggest potential enzyme inhibitory activity, with a higher score indicating a stronger inhibitory effect.

3.6.2 Toxicity profile

Table S3 presents the results of toxicity predictions using Pro Tox-II software, categorizing the compounds into five primary toxicity categories. The toxicity data indicate diverse toxicological profiles across the compounds. Compounds C1, C3, C6, C11, and C13 exhibited active hepatotoxicity, suggesting potential harm to the liver. Moreover, C1, C3, C6, C11, and C13 displayed active carcinogenicity, indicating a potential risk of causing cancer. Immunotoxicity assessment revealed activity in C1, C3, C6, C11, and C13, signifying an impact on the immune system. C1, C3, C6, C11, and C13 exhibited mutagenicity, suggesting the potential for inducing genetic mutations. Cytotoxicity analysis identified C1, C3, C6, and C13 as compounds with active cytotoxic effects. LD50 values, representing acute toxicity, varied among compounds, with C3 and C11 exhibiting moderate toxicity (300 mg/kg) and C16 exhibiting the highest LD50 value of 4000 mg/kg. Overall, these findings underscore the importance of a comprehensive *in silico* toxicity assessment for predicting potential hazards associated with chemical compounds, aiding in early-stage screening and decision-making in drug development and safety evaluation. The assignment of toxicity classes further facilitates the categorization and prioritization of compounds based on their potential adverse effects.

3.7 Molecular docking

In the virtual screening, a total of 16 hits were identified from the QSAR studies utilizing our developed equation model-4. These 16 hits were subsequently subjected to molecular docking simulations to predict their binding affinity with the CDK2 target molecule. Docking analysis of various compounds (C1-C16 and standard) with a target molecule revealed diverse interaction profiles involving specific amino acid residues. The docking scores of the obtained hits (C1-C16) are listed in Table 7 and were compared with those of the standard roscovitine. Among the obtained hit series, C3, C6, C7, C15 and C1 were more potent at binding to the target compound and exhibited good binding affinities greater than -7.0 K/Cal.

The binding energy of the standard drug roscovitine is -7.1 kcal/mol, representing the highest docking score and the formation of hydrogen bonds with Asp86 and Gly11. These findings reaffirm the strong binding affinity of the compound and establish it as a benchmark for the investigated compounds. Compound C3 demonstrated the highest docking score of -7.2 kcal/mol and engaged with three key residues (Asp 145, Glu 13, and Gln 131) through hydrogen bonding. This compound's pronounced affinity is substantiated by the involvement of multiple amino acids, suggesting a robust binding mode. Compounds C6, C7, and C15 also exhibited the best docking scores of -7.0, -7.1, and -7.0 Kcal/mol, respectively. These compounds formed hydrogen bonds with Asp145, emphasizing the significance of this interaction for their binding affinity. Several compounds demonstrated consistent interactions with Asp145, forming hydrogen bonds and suggesting a common binding pattern at this residue. Additionally, C15 engaged in hydrophobic interactions with Gly 11, Thr 165, and Val 164, further enhancing its binding profile. In contrast, compounds C2 and C16 displayed comparatively lower docking scores of -5.5 and -5.1 kcal/mol, respectively. These lower scores indicate weaker binding interactions, supported by the involvement of fewer residues in hydrogen bonding. Fig. 10 shows the 2D and 3D interactions of compounds C3 and C7 at the active site of 3PJ8 along with the standard. Due to their comparable interactions and superior binding scores, these compounds underwent additional molecular dynamics studies to assess their stability in relation to the protein. While other compounds displayed binding affinities that were nearly as significant as the standard, the 2D interactions of all the hit compounds can be observed in Fig. 11.

3.8 Molecular dynamics results

3.8.1 Protein–ligand RMSD

The protein–ligand RMSD is a critical metric in molecular dynamics (MD) simulations, indicating the stability and conformational changes of the ligand within the protein binding site over time. A lower RMSD value suggests greater stability and a more favorable binding interaction between the protein and

ligand.

The protein–ligand RMSD of Compound 3 was consistently within the range of 2-2.4 Å over the 40-nanosecond simulation period, as shown in Fig. 12A. This narrow RMSD range indicates that the binding mode was stable throughout the simulation, emphasizing the robust interaction between Compound 3 and the protein target. In contrast, compound 7 displayed higher RMSD values or greater deviations than compound 3 during the simulation, suggesting a less stable binding interaction, as depicted in Fig. 12B. The deviation in the RMSD of Compound 7 may indicate conformational changes, fluctuations, or weaker binding affinities, which could impact its efficacy as a potential drug candidate. Notably, the standard roscovitine used for comparison exhibited a consistent deviation in the RMSD throughout the simulation time, as shown in Fig. 12C. This deviation might be indicative of a less stable binding interaction between the standard compound and the protein, raising concerns about its effectiveness as a potential therapeutic agent. The compatibility of Compound 3 with a narrow RMSD range suggested that it maintained a stable conformation within the protein binding site. This stability is crucial for the sustained interaction required for a drug to exert its therapeutic effects, making Compound 3 a favorable candidate for further development.

3.8.2 RMSF

Root-mean-square fluctuation (RMSF) analysis was carried out to obtain information on the flexibility of the protein–ligand complexes. Ligand RMSF provides insights into the flexibility and dynamic behavior of ligand atoms during molecular dynamics simulations. Lower RMSF values suggest less fluctuation, indicating a more stable binding conformation. The ligand RMSF for Compound 3 was observed to be 0.8 Å over the 40-nanosecond simulation. This low RMSF value indicates that Compound 3 exhibits minimal fluctuation, suggesting a stable and well-defined binding conformation within the protein binding site. In contrast, Compound 7 displayed a higher ligand RMSF value of 1 Å over the simulation time. The increased fluctuation in Compound 7 may suggest more dynamic behavior, potentially leading to a less stable binding interaction than that of Compound 3. Interestingly, the standard compound showed even higher ligand RMSF values, ranging from 2-4 Å. These significant fluctuations indicate a less stable binding conformation for the standard compound, reinforcing concerns about its efficacy and reliability as a potential therapeutic agent. The low ligand RMSF value for Compound 3 aligns with the stable RMSD results, collectively suggesting that the binding conformation was well maintained and stable throughout the simulation. The stability of the ligand conformation is crucial for drug efficacy, as it ensures consistent interaction with the target protein throughout treatment. Both the ligand RMSD and RMSF analyses collectively support the superior stability of Compound 3 over Compound 7 and the standard.

The identification and characterization of protein–ligand interactions are fundamental aspects of drug discovery. The protein–ligand interactions of compounds 3 and 7 with the CDK2 target, emphasizing key residues such as GLU 81, LEU 83, GLU 12, and ASP 145. Our findings reveal distinct protein–ligand interaction patterns, with compound 3 exhibiting superior characteristics. Specifically, Compound 3 exhibited remarkable protein–ligand interactions, with two strong hydrogen bonds consistently forming with GLU (81; 97%) and LEU 83 (approximately 80% of the simulation time), indicating its potential as a promising drug candidate, as depicted in Fig. 13A. In contrast, compound 7 exhibited interactions with GLU 81 and LEU 83, with frequencies of 85% and 80%, respectively, as depicted in Fig. 13B. However, the overall interaction profile of compound 7 was less favorable than that of compound 3. Additionally, compound 7 displayed significantly weak H-bond interactions with GLU12 and ASP145, indicating potential limitations in its binding affinity. The standard does not show any H bonding, as depicted in Fig. 13C. The superior performance of compound 3 can be attributed to its sustained and strong hydrogen bonding with the key residues GLU 81 and LEU 83. These interactions, persisting at 97% and 80% of the simulation time, respectively, signified a stable and reliable binding mode.

3.8.4 Protein Secondary structure Elements:

Stable secondary structure elements are essential for maintaining the overall structural integrity of a protein, potentially influencing its binding affinity and interaction with ligands.

The protein secondary structures, namely, α -helices (25.72%) and β -helices (14.24%), and a total of 39.46% of the proteins participated in the simulation; these numbers were less than those of compound 7 and greater than those of compound 3. Protein secondary structure elements, such as α -helices and β -sheets, play crucial roles in determining the overall structural stability and function of proteins. Analyzing these structural elements during molecular dynamics simulations provides insights into the dynamic behavior of the protein and its interactions with ligands. Protein secondary structure analysis revealed that α -helices accounted for 25.72% and β -sheets for 14.24% of the simulation time, resulting in a total of 39.46% participation in these secondary structure elements. The participation of α -helix and β -sheet elements indicates the presence of stable and well-defined structural motifs in the protein during the simulation. Comparatively, the percentage participation of α -helix and β -sheet elements was greater for Compound 3 than for Compound 7. This greater participation suggested that the protein with Compound 3 maintained a more ordered and stable secondary structure throughout the simulation, potentially contributing to a more favorable binding interaction.

3.8.5 Molecular surface area

The molecular surface area is a critical parameter influencing a compound's interaction with its environment. Compound 3 has a molecular surface area of 342 Å², closely resembling that of compound 7. However, when compared to the standard, compound 3 consistently exhibited higher values throughout the simulation time. The surface area of compound 3 fluctuates dynamically, reflecting its adaptability in response to varying environmental conditions. In contrast, compound 7 has a comparable molecular surface area but consistently lags behind compound 3 and fails to reach the same elevated values. The surface area of compound 7 also fluctuated, albeit to a lesser extent than that of compound 3. The consistently greater molecular surface area of compound 3 compared to that of both compound 7 and the standard indicates its superior adaptability and potential for increased interactions with its surroundings. The fluctuations in surface area throughout the simulation time suggest dynamic adjustments in the molecular conformation, enabling compound 3 to explore a broader range of interactions. The comparable but consistently lower surface area of compound 7 implies a limitation in its adaptability and a reduced ability to explore a diverse set of interactions. This reduced adaptability might hinder its efficacy in binding to different sites or conformations within the target

protein. The similarity in surface area between compounds 3 and 7 suggested that they share certain structural characteristics, yet the consistently greater values for compound 3 indicate that it possesses unique features enabling more extensive interactions.

3.8.6 Solvent accessible surface area (SASA)

The solvent accessible surface area (SASA) is a crucial metric for understanding the exposure of a molecule to its surrounding solvent environment. The SASA values and fluctuations were analyzed to assess the adaptability of the compounds and interactions with the solvent. Compound 3 exhibited a fluctuating SASA between 60 and 150 Å² over the first 20 ns of simulation. Compound 7 shows a slightly wider range, with SASA values ranging from 80–160 Å². However, the standard compound demonstrated a consistently higher SASA, reaching approximately 250 Å² at 60 ns, with fluctuations observed throughout the simulation time. The observed SASA ranges provide insights into the adaptability and exposure of compounds to the solvent environment. The fluctuating SASA of compound 3 suggested dynamic conformational changes, allowing the molecule to adapt to varying solvent conditions.

3.8.7 Polar Surface Area

The polar surface area is a critical parameter for understanding the interaction of a ligand with its biological environment. The focus is on assessing stability, as stability is often correlated with the ligand's potential to exert biological activity. Compound 3 exhibited an excellent polar surface area of 120 Å², with minimal fluctuations throughout the simulation. In contrast, both compound 7 and the standard compound exhibited fluctuations in polar surface area during the simulation. The overall comparison of ligand properties highlights the stability of compound 3, suggesting its potential efficacy in rendering biological activity. The stable and excellent polar surface area exhibited by compound 3 is indicative of its ability to interact with polar regions in the biological environment. This stability may contribute to the ligand's reliable and sustained interactions with the target protein, supporting its potential for biological activity. The lack of significant fluctuations suggested that compound 3 maintains a structurally stable conformation, which is crucial for its efficacy as a drug candidate. In contrast, the observed fluctuations in polar surface area for compound 7 and the standard compound may indicate variations in their interactions with the biological environment. These fluctuations could lead to structural instability, impacting the ability of ligands to consistently engage with the target protein. Such instability may raise concerns about the overall efficacy of compound 7 and the standard compound in exerting their intended biological effects. A comparison of the ligand properties, with a focus on polar surface area stability, underscores the favorable attributes of compound 3. Due to its stable and excellent polar surface area coupled with minimal fluctuations, compound 3 is a promising candidate over compound 7 and the standard.

4. Conclusion

The present study focused on performing QSAR analysis of 45 pyrazolopyrimidine derivatives utilizing QSARINS software. The four developed QSAR models satisfy the OECD criteria for significant and accurate prediction and have a good model fit. Model 4 was chosen as the optimal model based on its statistical parameters. The QSAR equation revealed essential insights, highlighting the positive correlation of EstateVSA2 and CIC3 with anticancer activity, while SlogP_VSA5, fr_Ndealkylation2, Diameter, and slogPVSA5 exhibited negative correlations. The quality of the model was rigorously assessed through various statistical methods, including fitting, internal, and external validation criteria, ensuring its reliability. Furthermore, the hit compounds identified in this study adhere to the Lipinski rule, reinforcing their potential for further development as drug candidates. Molecular docking revealed that compounds C3 (methoxy) and C7 (trifluoromethyl) exhibited low-energy conformations (-7.2 Kcal and -7.1 Kcal) compared to the conformations of the standard drug roscovitine with the 3PJ8 target. The present study provides the basic requirements for designing and optimizing novel pyrazolopyrimidines with anticancer activity. The stabilities of lead compounds C3 and C7 were determined using molecular dynamics simulations, and upon analysis, compound C3 emerged as the best lead compound for targeting CDK2.

Declarations

Supplementary Information

The online version contains supplementary material available at

Acknowledgments

We thank the Management of SRM College of Pharmacy and SRM Institute of Science and Technology for their constant support and encouragement. We thank the QSARINS software license provider Dr Gramatica P. and also D. E. Shaw research group and Schrödinger, Inc., for providing an academically free version of the Desmond suite.

Author contributions

BK: data collection, conceptualization, data curation, investigation, methodology, validation, writing-original draft. & editing. MK: Resources, software, supervision, writing-review & editing.

Funding

This research project did not receive any funding.

Data availability

The datasets generated during the current study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare no conflicts of interest.

Institutional review board statement

Not applicable.

References

1. World Health Organization (2022) Cancer Statistics. <https://www.who.int/news-room/fact-sheets/detail/cancer>
2. Angre T, Kumar A, Singh AK, et al (2022) Role of Collagen Regulators in Cancer Treatment: A Comprehensive Review. *Anticancer Agents Med Chem* 22:2956–2984. <https://doi.org/10.2174/1871520622666220501162351>
3. Sung H, Ferlay J, Siegel RL, et al (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 71:209–249. <https://doi.org/10.3322/CAAC.21660>
4. Malumbres M, Barbacid M (2009) Cell cycle, CDKs and cancer: A changing paradigm. *Nat. Rev. Cancer* 9
5. Deshpande A, Sicinski P, Hinds PW (2005) Cyclins and cdks in development and cancer: a perspective. *Oncogene* 24:2909–2915. <https://doi.org/10.1038/sj.onc.1208618>
6. Vijayaraghavan S, Molder S, Keyomarsi K, Layman RM (2018) Inhibiting CDK in Cancer Therapy: Current Evidence and Future Directions. *Target Oncol* 13:21–38. <https://doi.org/10.1007/S11523-017-0541-2/TABLES/4>
7. Singh SK, Dessalew N, Bharatam P V. (2006) 3D-QSAR CoMFA study on indenopyrazole derivatives as cyclin dependent kinase 4 (CDK4) and cyclin dependent kinase 2 (CDK2) inhibitors. *Eur J Med Chem* 41:1310–1319. <https://doi.org/10.1016/J.EJMECH.2006.06.010>
8. Sabnis RW (2020) Novel CDK2 Inhibitors for Treating Cancer. *ACS Med. Chem. Lett.* 11:2346–2347
9. Chagaleti BK, Reddy MBR, Saravanan V, et al (2023) An overview of mechanism and chemical inhibitors of shikimate kinase. *J. Biomol. Struct. Dyn.*
10. Echalié A, Bettayeb K, Ferandin Y, et al (2008) Meriolins (3-(pyrimidin-4-yl)-7-azaindoles): Synthesis, kinase inhibitory activity, cellular effects, and structure of a CDK2/Cyclin A/meriolin complex. *J Med Chem* 51:737–751. https://doi.org/10.1021/JM700940H/SUPPL_FILE/JM700940H-FILE003.PDF
11. Chohan TA, Qian HY, Pan YL, Chen JZ (2015) Molecular simulation studies on the binding selectivity of 2-anilino-4-(thiazol-5-yl)-pyrimidines in complexes with CDK2 and CDK7. *Mol Biosyst* 12:145–161. <https://doi.org/10.1039/C5MB00630A>
12. Morgan DO (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol* 13:261–291. <https://doi.org/10.1146/ANNUREV.CELLBIO.13.1.261>
13. Fagundes R, Teixeira LK (2021) Cyclin E/CDK2: DNA Replication, Replication Stress and Genomic Instability. *Front. Cell Dev. Biol.* 9
14. Ohtsubo M, Theodoras AM, Schumacher J, et al (1995) Human Cyclin E, a Nuclear Protein Essential for the G 1 -to-S Phase Transition . *Mol Cell Biol* 15: <https://doi.org/10.1128/mcb.15.5.2612>
15. Xu X, Nakano T, Wick S, et al (1999) Mechanism of Cdk2/Cyclin E Inhibition by p27 and p27 Phosphorylation. *Biochemistry* 38:8713–8722. <https://doi.org/10.1021/BI9903446>
16. Hope I, Endicott JA, Watt JE (2023) Emerging approaches to CDK inhibitor development, a structural perspective. *RSC Chem Biol* 4:146–164. <https://doi.org/10.1039/D2CB00201A>
17. Girgis AS, Panda SS, Aziz MN, et al (2015) Rational design, synthesis, and 2D-QSAR study of anti-oncological alkaloids against hepatoma and cervical carcinoma. *RSC Adv* 5:28554–28569. <https://doi.org/10.1039/C4RA16663A>
18. Gramatica P, Papa E, Sangion A (2018) QSAR modeling of cumulative environmental end-points for the prioritization of hazardous chemicals. *Environ Sci Process Impacts* 20:38–47. <https://doi.org/10.1039/C7EM00519A>
19. Radha S Lead optimization of 4-(thio)-chromenone 6-O-sulfamate analogs using QSAR, molecular docking and DFT-a combined approach as steroidal sulfatase inhibitors. <https://doi.org/10.1080/10799893.2020.1794004>
20. Lan P, Chen WN, Xiao GK, et al (2010) 3D-QSAR and docking studies on pyrazolo[4,3-h]quinazoline-3-carboxamides as cyclin-dependent kinase 2 (CDK2) inhibitors. *Bioorg Med Chem Lett* 20:6764–6772. <https://doi.org/10.1016/J.BMCL.2010.08.131>
21. B S, M. K K (2020) Insights into structures of imidazo oxazines as potent polyketide synthase XIII inhibitors using molecular modeling techniques. *J Recept Signal Transduct Res* 40:313–323. <https://doi.org/10.1080/10799893.2020.1742740>
22. Patan A, Aanandhi M V, Gopinath P (2023) Molecular dynamics simulation approach of hybrid chalcone–thiazole complex derivatives for DNA gyrase B inhibition: lead generation. *RSC Adv* 13:24291–24308. <https://doi.org/10.1039/D3RA00732D>
23. Kumar A, Singh AK, Singh H, et al (2023) Nitrogen Containing Heterocycles as Anticancer Agents: A Medicinal Chemistry Perspective. *Pharm* 2023, Vol 16, Page 299 16:299. <https://doi.org/10.3390/PH16020299>
24. Li Y, Gao W, Li F, et al (2013) An in silico exploration of the interaction mechanism of pyrazolo[1,5-a]pyrimidine type CDK2 inhibitors. *Mol Biosyst* 9:2266–2281. <https://doi.org/10.1039/C3MB70186G>
25. Hossan A, Alrefaei AF, Katouah HA, et al (2023) Synthesis, anticancer activity, and molecular docking of new pyrazolo[1,5-a]pyrimidine derivatives. *J Saudi Chem Soc* 27:101599. <https://doi.org/10.1016/J.JSCS.2023.101599>
26. Chohan TA, Chen JJ, Qian HY, et al (2016) Molecular modeling studies to characterize N-phenylpyrimidin-2-amine selectivity for CDK2 and CDK4 through 3D-QSAR and molecular dynamics simulations. *Mol Biosyst* 12:1250–1268. <https://doi.org/10.1039/C5MB00860C>

27. Sherbiny FF, Bayoumi AH, El-Morsy AM, et al (2021) Design, Synthesis, biological Evaluation, and molecular docking studies of novel Pyrazolo[3,4-d]Pyrimidine derivative scaffolds as potent EGFR inhibitors and cell apoptosis inducers. *Bioorg Chem* 116:105325. <https://doi.org/10.1016/J.BIOORG.2021.105325>
28. Cherukupalli S, Chandrasekaran B, Kryštof V, et al (2018) Synthesis, anticancer evaluation, and molecular docking studies of some novel 4,6-disubstituted pyrazolo[3,4-d]pyrimidines as cyclin-dependent kinase 2 (CDK2) inhibitors. *Bioorg Chem* 79:46–59. <https://doi.org/10.1016/J.BIOORG.2018.02.030>
29. Nassar IF, Abdel Aal MT, El-Sayed WA, et al (2022) Discovery of pyrazolo[3,4-d]pyrimidine and pyrazolo[4,3-e][1,2,4]triazolo[1,5-c]pyrimidine derivatives as novel CDK2 inhibitors: synthesis, biological and molecular modeling investigations. *RSC Adv* 12:14865–14882. <https://doi.org/10.1039/D2RA01968J>
30. Kim DC, Lee YR, Yang BS, et al (2003) Synthesis and biological evaluations of pyrazolo[3,4-d]pyrimidines as cyclin-dependent kinase 2 inhibitors. *Eur J Med Chem* 38:525–532. [https://doi.org/10.1016/S0223-5234\(03\)00065-5](https://doi.org/10.1016/S0223-5234(03)00065-5)
31. Verkhivker GM (2017) Network-based modeling and percolation analysis of conformational dynamics and activation in the CDK2 and CDK4 proteins: dynamic and energetic polarization of the kinase lobes may determine divergence of the regulatory mechanisms. *Mol Biosyst* 13:2235–2253. <https://doi.org/10.1039/C7MB00355B>
32. Zhang J, Zou L, Tang P, et al (2020) Design, synthesis and biological evaluation of 1H-pyrazolo [3,4-d]pyrimidine derivatives as PAK1 inhibitors that trigger apoptosis, ER stress and anti-migration effect in MDA-MB-231 cells. *Eur J Med Chem* 194. <https://doi.org/10.1016/J.EJMECH.2020.112220>
33. Gramatica P, Cassani S, Chirico N (2014) QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. *J Comput Chem* 35:1036–1044. <https://doi.org/10.1002/JCC.23576>
34. Kumar S, Manoharan A, Jayalakshmi J, et al (2023) Exploiting butyrylcholinesterase inhibitors through a combined 3-D pharmacophore modeling, QSAR, molecular docking, and molecular dynamics investigation. *RSC Adv* 13:9513–9529. <https://doi.org/10.1039/D3RA00526G>
35. chemistry CY-J of computational, 2011 undefined (2011) PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Wiley Online Libr* 32:1466–1474. <https://doi.org/10.1002/jcc.21707>
36. Cao DS, Xu QS, Hu QN, Liang YZ (2013) ChemoPy: freely available python package for computational biology and chemoinformatics. *Bioinformatics* 29:1092–1094. <https://doi.org/10.1093/BIOINFORMATICS/BTT105>
37. Bento AP, Hersey A, Félix E, et al (2020) An open source chemical structure curation pipeline using RDKit. *J Cheminform* 12:1–16. <https://doi.org/10.1186/S13321-020-00456-1/FIGURES/6>
38. Jillella GK, Roy K (2022) QSAR modeling of organic dyes for their acute toxicity in *Daphnia magna* using 2D-descriptors. *SAR QSAR Environ Res* 33:111–139. <https://doi.org/10.1080/1062936X.2022.2033318>
39. Khan PM, Roy K (2021) QSPR modeling for investigation of different properties of aminoglycoside-derived polymers using 2D descriptors. *SAR QSAR Environ Res* 32:595–614. <https://doi.org/10.1080/1062936X.2021.1939150>
40. Kumar V, Ojha PK, Saha A, Roy K (2020) Exploring 2D-QSAR for prediction of beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease. *SAR QSAR Environ Res* 31:87–133. <https://doi.org/10.1080/1062936X.2019.1695226>
41. Gramatica P (1AD) Principles of QSAR Modeling: Comments and Suggestions From Personal Experience. <https://services.igi-global.com/resolvedoi/resolve.aspx?doi=104018/IJQSPR20200701.0a1> 5:61–97. <https://doi.org/10.4018/IJQSPR.20200701.0A1>
42. Moulishankar A, Sundarrajan T (2023) QSAR modeling, molecular docking, dynamic simulation and ADMET study of novel tetrahydronaphthalene derivatives as potent antitubercular agents. *Beni-Suef Univ J Basic Appl Sci* 12:1–20. <https://doi.org/10.1186/S43088-023-00451-Z/FIGURES/14>
43. Moulishankar A, Thirugnanasambandam S (2023) Quantitative structure-activity relationship (QSAR) modeling study of some novel thiazolidine 4-one derivatives as potent anti-tubercular agents. *J Recept Signal Transduct.* <https://doi.org/10.1080/10799893.2023.2281671>
44. Gramatica P, Sangion A (2016) A Historical Excursus on the Statistical Validation Parameters for QSAR Models: A Clarification Concerning Metrics and Terminology. *J Chem Inf Model* 56:1127–1131. https://doi.org/10.1021/ACS.JCIM.6B00088/ASSET/IMAGES/MEDIUM/CI-2016-000886_0001.GIF
45. Gramatica P, Chirico N, Papa E, et al (2013) QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *Artic J Comput Chem* 34:2121–2132. <https://doi.org/10.1002/jcc.23361>
46. Roy K, Kar S, Das RN (2015) Statistical Methods in QSAR/QSPR. 37–59. https://doi.org/10.1007/978-3-319-17281-1_2
47. Cañizares-Carmenate Y, Campos Delgado LE, Torrens F, Castillo-Garit JA (2020) Thorough evaluation of OECD principles in modeling of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine derivatives using QSARINS. *SAR QSAR Environ Res* 31:741–759. <https://doi.org/10.1080/1062936X.2020.1810116>
48. Rücker C, Rücker G, Meringer M (2007) Y-randomization and its variants in QSPR/QSAR. *J Chem Inf Model* 47:2345–2357. <https://doi.org/10.1021/CI700157B/ASSET/IMAGES/MEDIUM/CI700157BN00001.GIF>
49. Adeniji SE, Uba S, Uzairu A, Arthur DE (2019) A Derived QSAR Model for Predicting Some Compounds as Potent Antagonist against *Mycobacterium tuberculosis*: A Theoretical Approach . *Adv Prev Med* 2019:1–18. <https://doi.org/10.1155/2019/5173786>
50. Kalil AC, Patterson TF, Mehta AK, et al (2021) Baricitinib plus Remdesivir for Hospitalized Adults with Covid-19. *N Engl J Med* 384. <https://doi.org/10.1056/nejmoa2031994>
51. Nath A, Ojha PK, Roy K (2023) QSAR assessment of aquatic toxicity potential of diverse agrochemicals. *SAR QSAR Environ Res* 34:923–942. <https://doi.org/10.1080/1062936X.2023.2278074>
52. Chatterjee M, Roy K (2022) Application of cross-validation strategies to avoid overestimation of performance of 2D-QSAR models for the prediction of aquatic toxicity of chemical mixtures. *SAR QSAR Environ Res* 33:463–484. <https://doi.org/10.1080/1062936X.2022.2081255>
53. Paul R, Roy J, Roy K (2023) Prediction of soil ecotoxicity against *Folsomia candida* using acute and chronic endpoints. *SAR QSAR Environ Res* 34:321–340. <https://doi.org/10.1080/1062936X.2023.2211350>

54. Molinspiration Cheminformatics free web services, Slovensky Grob S <https://www.molinspiration.com>. <https://www.molinspiration.com>
55. Jena A, Prakashraj C, Chagaleti BK, et al (2023) In silico Design and Synthesis of Some New Imidazole Derivatives for Tuberculosis. *Indian J Heterocycl Chem* 33:. <https://doi.org/10.59467/ijhc.2023.33.43>
56. Pranathi L, Bharath Kumar C, Kumar AA, Devi KV Design, Synthesis, and Docking Studies of Some New 5-((1,1'-Biphenyl]-4-yloxy) methyl)-2-(N-methylene amino-1,3,4-thiadiazoles as Anti-inflammatory Agents
57. Shanthakumar B, Saravanan V, Chagaleti BK, Kathiravan MK (2023) Design synthesis and biological evaluation of thiophene 2-pentafluoro benzamide derivatives as antitubercular agent. *J Med Pharm Allied Sci* 12:. <https://doi.org/10.55522/jmpas.V12I1.4402>
58. G.Slovensky (1986) G.Slovensky, Cheminformatics free web services, 1986, <https://www.molinspiration.com>.
59. Chagaleti BK, Saravanan V, Vellapandian C, Kathiravan MK (2023) Exploring cyclin-dependent kinase inhibitors: a comprehensive study in search of CDK-6 inhibitors using a pharmacophore modeling and dynamics approach. *RSC Adv* 13:33770–33785. <https://doi.org/10.1039/D3RA05672D>
60. Banerjee P, Eckert AO, Schrey AK, Preissner R (2018) ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res* 46:W257–W263. <https://doi.org/10.1093/NAR/GKY318>
61. Banerjee P, Dehnbostel FO, Preissner R (2018) Prediction is a balancing act: Importance of sampling methods to balance sensitivity and specificity of predictive models based on imbalanced chemical datasets. *Front Chem* 6:387941. <https://doi.org/10.3389/FCHEM.2018.00362/BIBTEX>
62. Banerjee P, Eckert AO, Schrey AK, Preissner R (2018) ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res* 46:W257. <https://doi.org/10.1093/NAR/GKY318>
63. Sathish S, Devaraju P, Julius A, et al (2023) Identification of selective inhibitors for Janus kinase 1: an integrated drug repurposing strategy for breast cancer. *Chem Pap* 78:245–262. <https://doi.org/10.1007/S11696-023-03070-1/TABLES/7>
64. Moulishankar A, Lakshmanan K (2020) Data on molecular docking of naturally occurring flavonoids with biologically important targets. *Data Br* 29:105243. <https://doi.org/10.1016/J.DIB.2020.105243>
65. K G, Lakshmanan K, Moulishankar A, et al (2020) Screening of Kabasura Kudineer Chooranam Against COVID-19 Through Targeting of Main Protease and RNA-Dependent RNA Polymerase of SARS-CoV-2 by Molecular Docking Studies. *SSRN Electron J*. <https://doi.org/10.2139/SSRN.3625653>
66. RCSB PDB: Homepage. <https://www.rcsb.org/>. Accessed 28 Jan 2024
67. Chohan TA, Qian HY, Pan YL, Chen JZ (2015) Molecular simulation studies on the binding selectivity of 2-anilino-4-(thiazol-5-yl)-pyrimidines in complexes with CDK2 and CDK7. *Mol Biosyst* 12:145–161. <https://doi.org/10.1039/c5mb00630a>
68. Shivakumar D, Williams J, Wu Y, et al (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the opl force field. *J Chem Theory Comput* 6:1509–1519. https://doi.org/10.1021/CT900587B/SUPPL_FILE/CT900587B_SI_003.ZIP
69. Venkatesan Saravanan, Kumar Chagaleti B, Devi Packiapalavesam S, Muthukumaradoss Kathiravan (2024) Ligand based pharmacophore modeling and integrated computational approaches in the quest for small molecule inhibitors against hCA IX. *RSC Adv* 14:3346–3358. <https://doi.org/10.1039/D3RA08618F>
70. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105:6474–6487. https://doi.org/10.1021/JP003919D/SUPPL_FILE/JP003919D_S.PDF
71. Shinoda W, Mikami M (2003) Rigid-body dynamics in the isothermal-isobaric ensemble: A test on the accuracy and computational efficiency. *J Comput Chem* 24:920–930. <https://doi.org/10.1002/JCC.10249>

Tables

Tables 1 to 7 are available in the Supplementary Files section

Figures

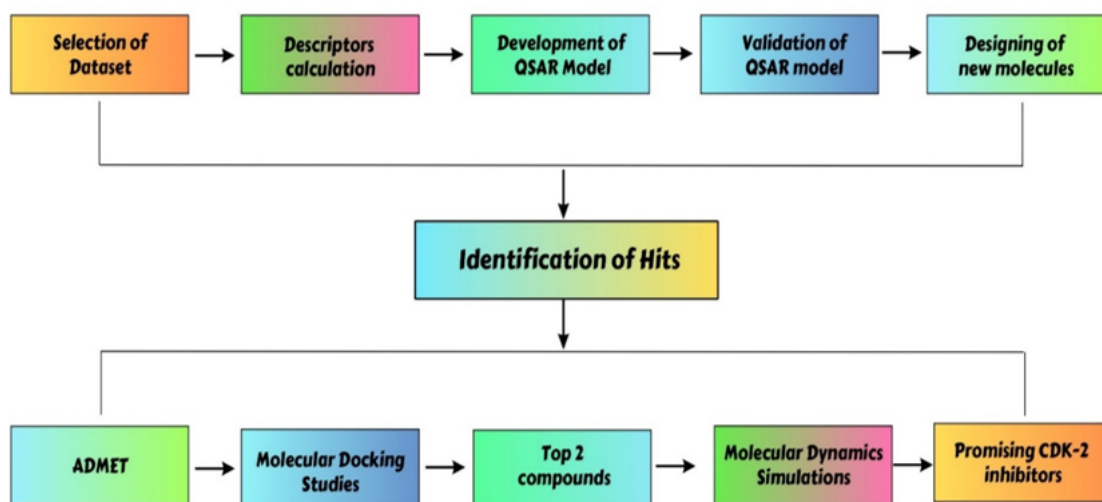


Figure 1

Schematic representation of the present work flow

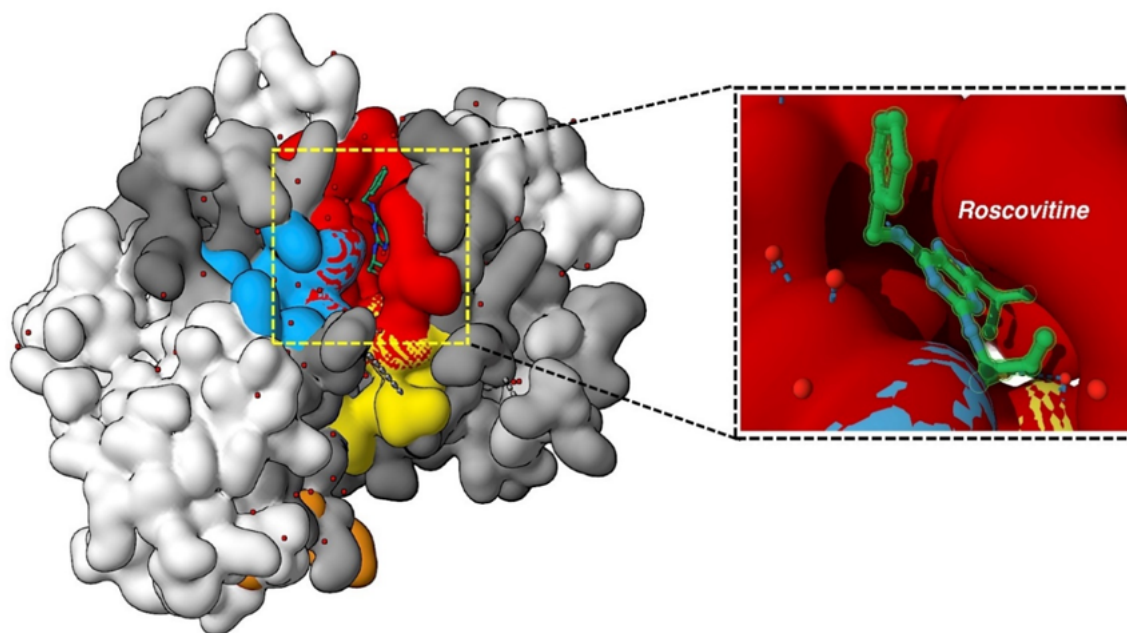


Figure 2

Visual representation of four pockets: Pocket 1 (red), Pocket 2 (yellow), Pocket 3 (subtle orange), and Pocket 4 (blue), detailing Roscovitine's spatial arrangement in Pocket 1.

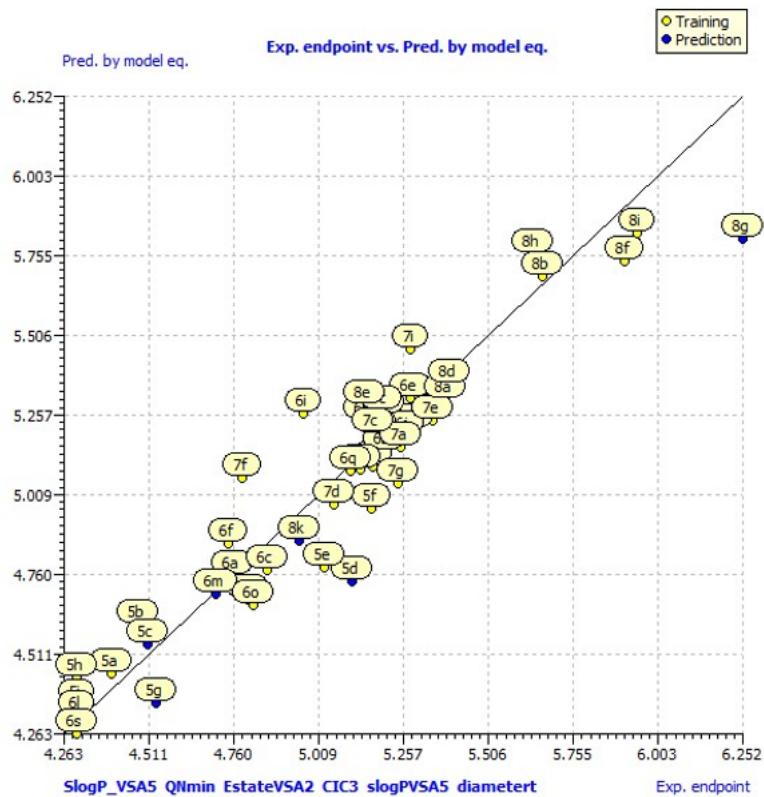


Figure 3
Experimental endpoint vs. prediction by LOO.

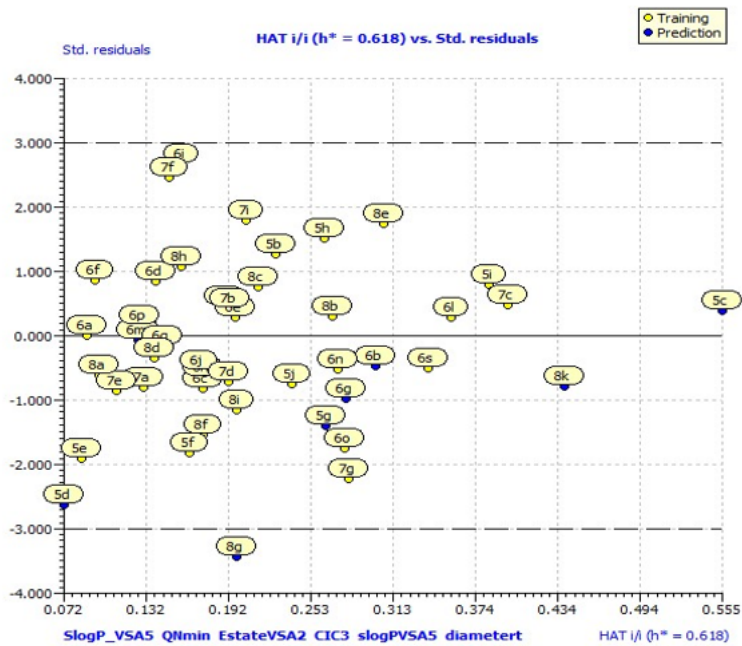


Figure 4
William's plot for the best model

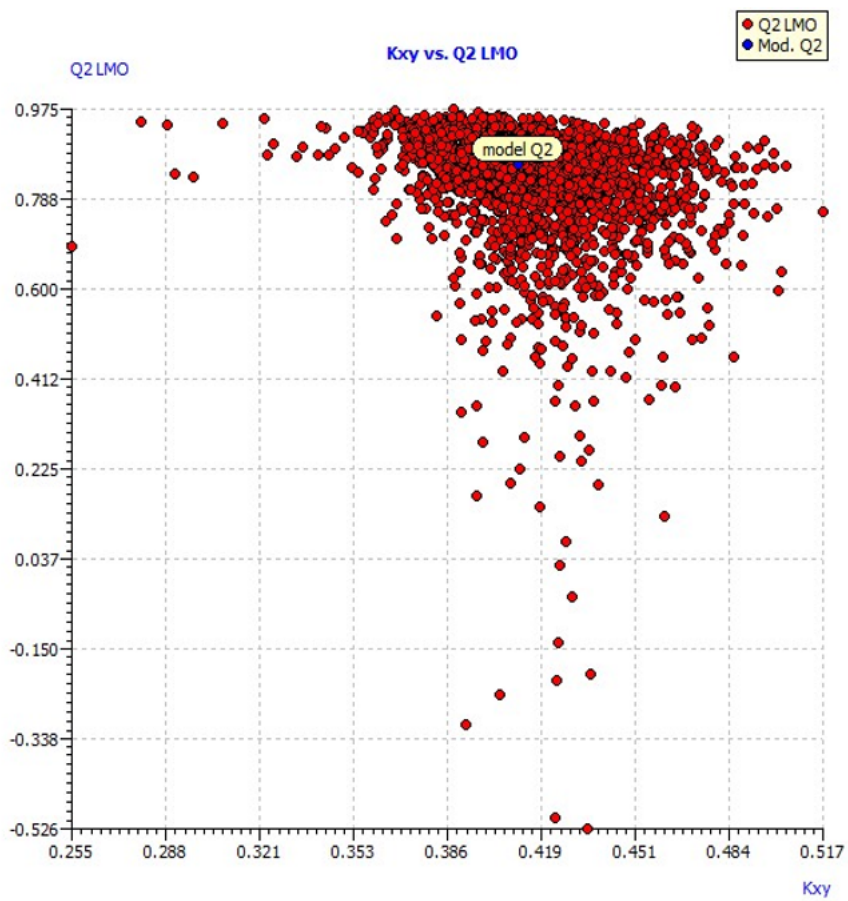


Figure 5

Q²LMO plot explaining the intercorrelation among the descriptors (Kxy vs. Q²LMO)

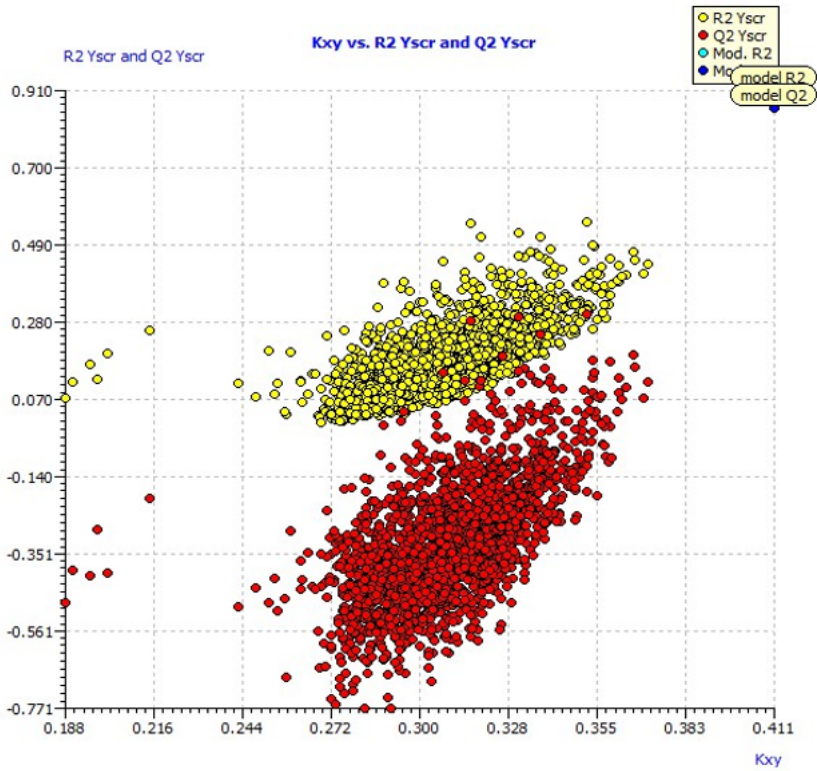


Figure 6

The k_{xy} vs. R^2 and Q^2_{LOO} (Y-scramble plot explaining the external validation parameters).

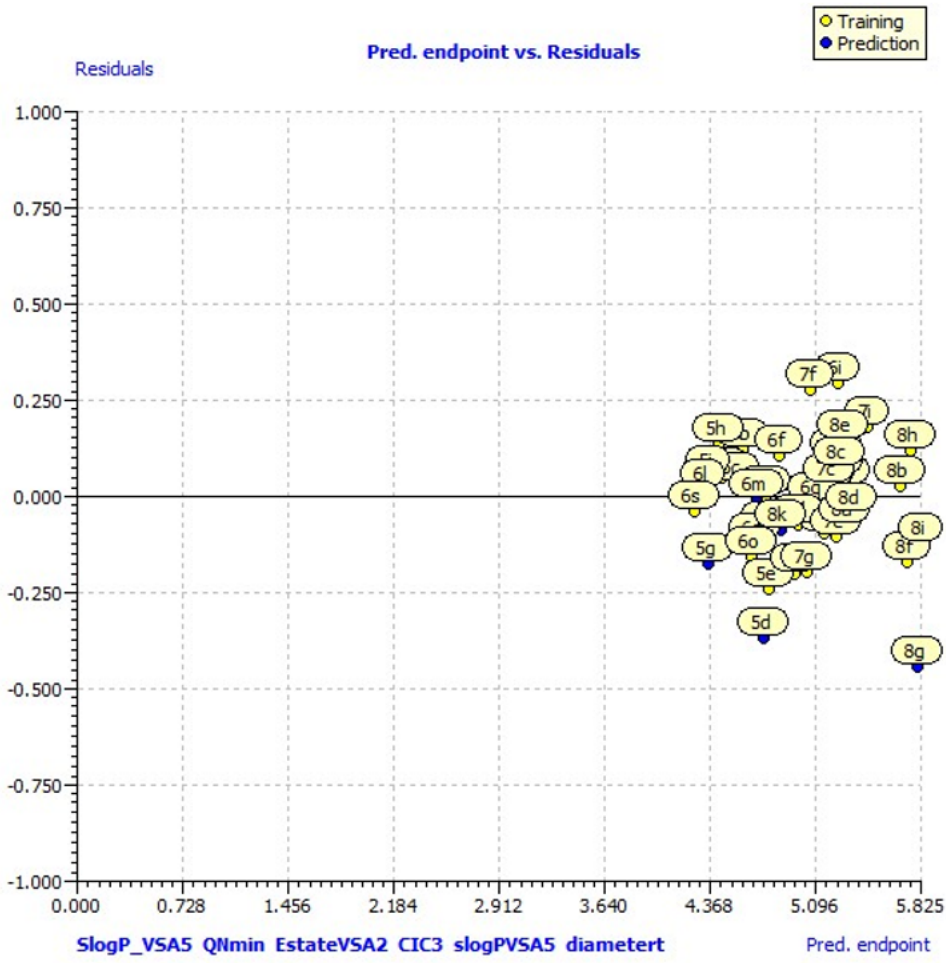


Figure 7

Predicted endpoint vs Residuals end point

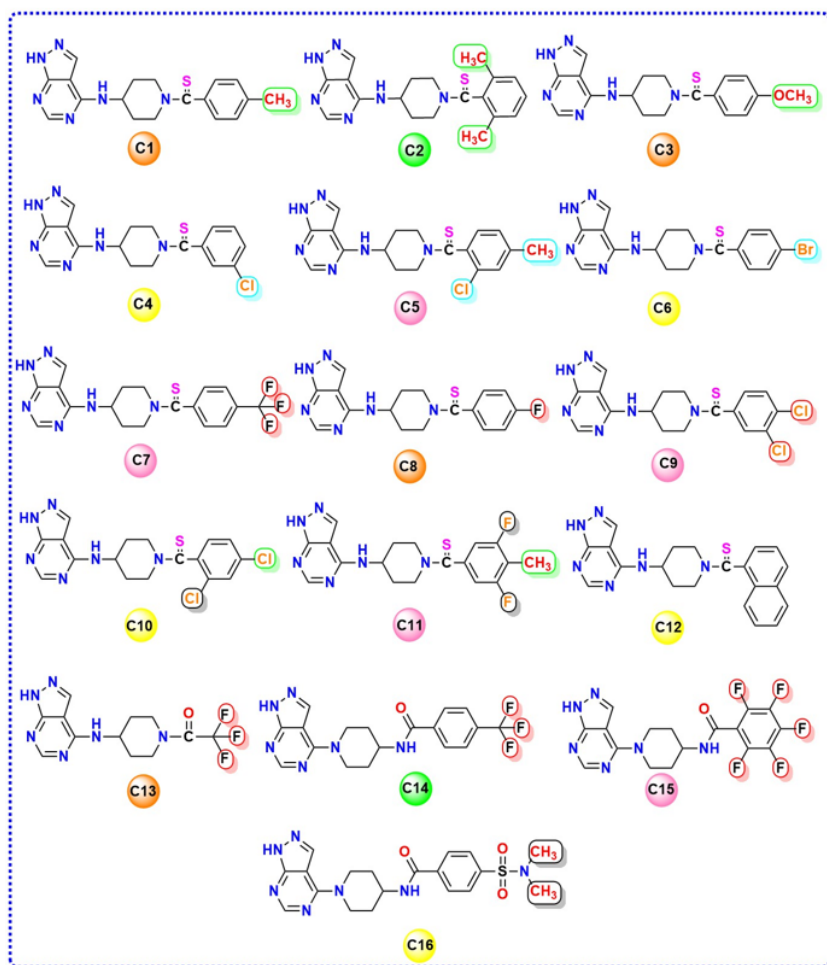


Figure 8

Structures of hit molecules predicted by Model 4 from the designed compounds.

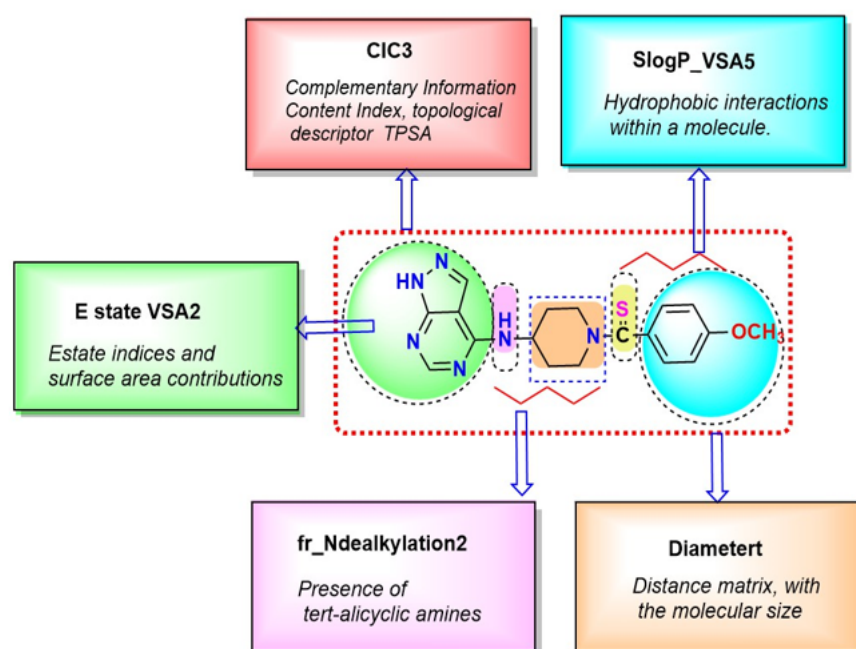


Figure 9

Molecular Descriptors contribution for the designed compounds.

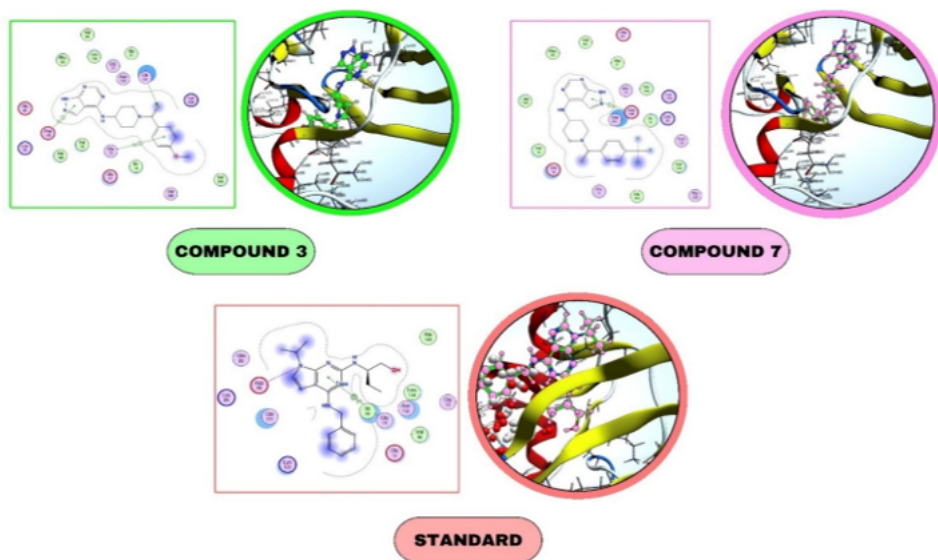


Figure 10

2D & 3D interactions of the Compounds C3, C7 and standard.

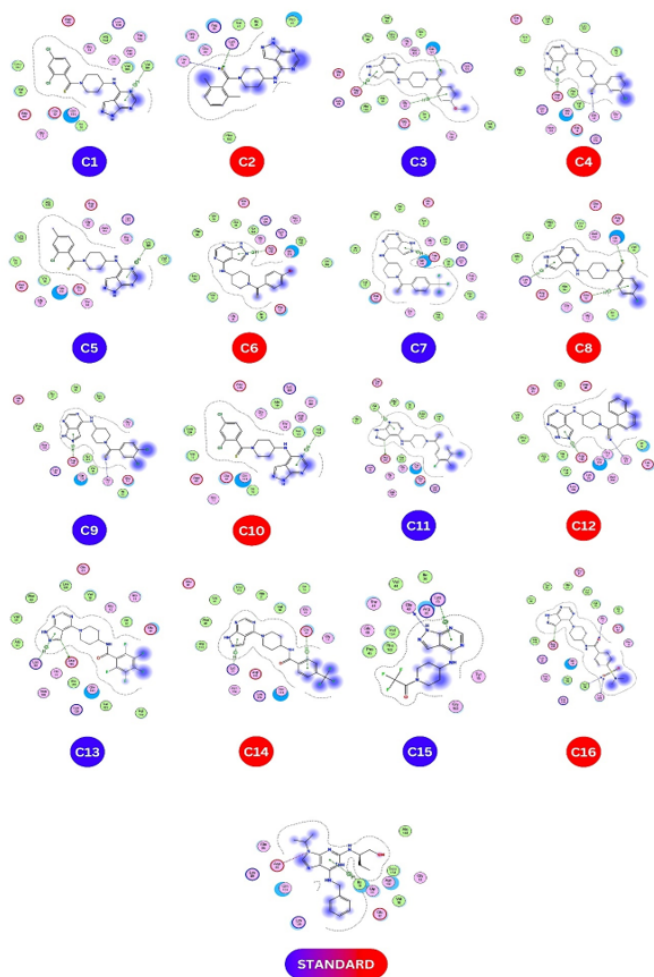


Figure 11

2D interactions of all hit molecules C1-C16.

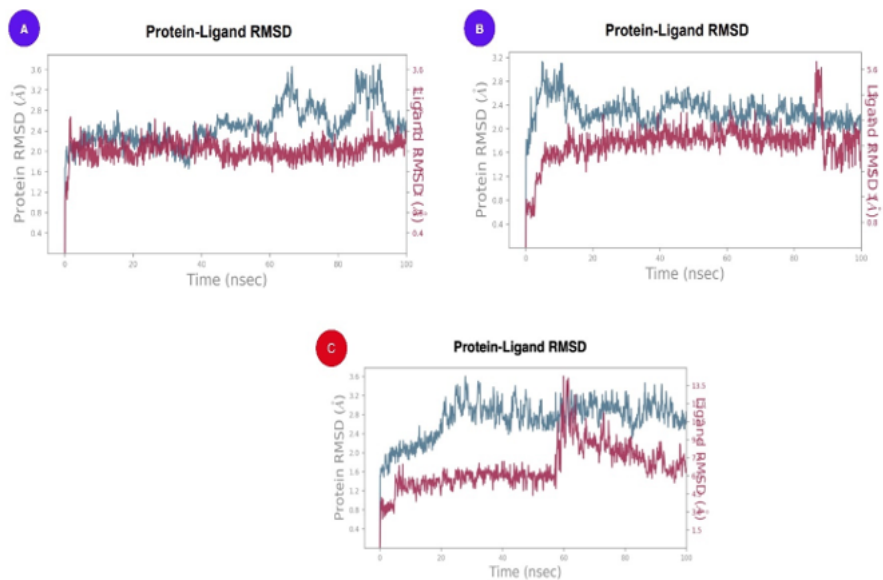


Figure 12

Protein–ligand RMSD of C3(A), C7(B) and Standard

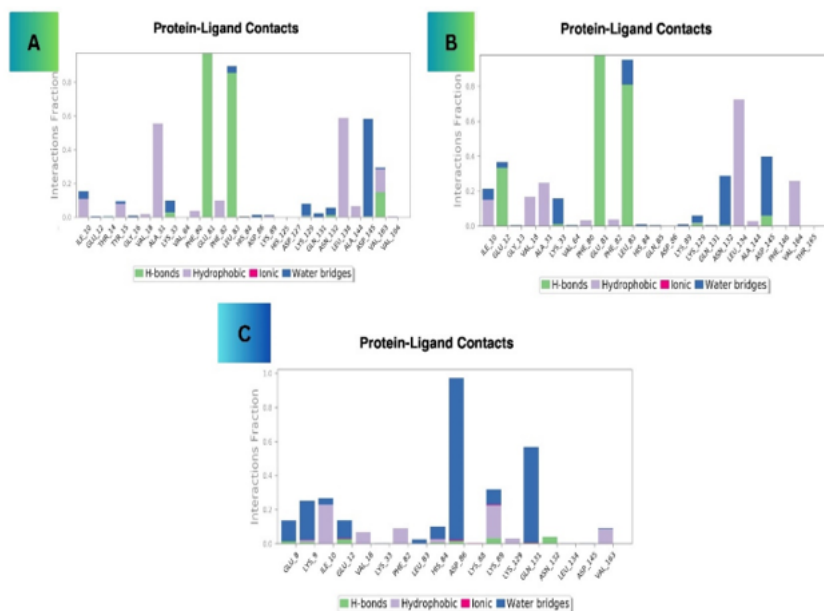


Figure 13

H-Bond plot of C3(A), C7(B) and Standard

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplimentaryJCAM.docx](#)
- [Tables.docx](#)