Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# In-Silico Design, QSAR Modeling, Molecular Docking, MD Simulation, Pharmacokinetic and ADMET Studies of 3-Cyanopyridine Derivatives as Potential Anti-Breast Cancer Agents

**Usman Bello**

ub1410@gmail.com

Ahmadu Bello University

**Sagiru Hamza Abdullahi**

Ahmadu Bello University

**Adamu Uzairu**

Ahmadu Bello University

**Muhammad Tukur Ibrahim**

Ahmadu Bello University

**Sani Uba**

Ahmadu Bello University

**Zakari Ya'u Ibrahim**

Ahmadu Bello University

**Renato Araujo da Costa**

Instotuto Federal do Para-Campus Abaetetuba Para Brazil

**Research Article**

**Keywords:** 3-cyanopyridine analogs, breast cancer, MCF-7 cell line, QSAR, molecular docking, molecular dynamics, pharmacokinetics, ADMET, in-silico drug design

**Posted Date:** November 12th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11361799/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# In-Silico Design, QSAR Modeling, Molecular Docking, MD Simulation, Pharmacokinetic and ADMET Studies of 3-Cyanopyridine Derivatives as Potential Anti-Breast Cancer Agents

Usman Bello [a,] *, Sagiru Hamza Abdullahi [a], Adamu Uzairu [a], Muhammad Tukur Ibrahim [a], Sani Uba [a], Zakari Ya'u Ibrahim [a], Renato Araújo da Costa [b]

[a] *Department of Chemistry, Faculty of Physical Sciences, Ahmadu Bello University, P.M.B. 1045, Zaria, Kaduna State, Nigeria*

[b] *Molecular Biology and Microbiology Laboratory, Federal Institute of Pará-Abaetetuba Campus, Avenida, Rio de Janeiro, Brazil*

*Corresponding author: Usman Bello
E-mail: ub1410@gmail.com

## Abstract

**Background:** Breast cancer remains a major global health challenge, creating a need for new compounds with improved antiproliferative activity and drug-like properties. This study investigated 3-cyanopyridine derivatives with reported activity against the MCF-7 breast cancer cell line and explored their potential as PIM-1-targeting anticancer candidates.

**Methods:** A series of 3-cyanopyridine derivatives with reported experimental $IC_{50}$ values was subjected to QSAR modelling using PaDEL descriptors, Kennard–Stone dataset division, genetic function approximation (GFA) and multiple linear regression. Model performance was assessed using $R^2$, adjusted $R^2$, $Q^2_{CV}$, external $R^2_{pred}$, multicollinearity analysis and applicability-domain analysis. Y-scrambling was used to further assess the robustness of the QSAR model against chance correlation by randomly permuting the response variable ($pIC_{50}$) while retaining the original descriptor matrix and repeating model construction for 10 iterations. Five new analogues (A1–A5) were designed from the QSAR model. Their predicted interactions with PIM-1 were evaluated using Glide XP molecular docking, followed by 500-ns molecular dynamics simulations for A1 and A5 and in-silico pharmacokinetic/ADMET prediction.

**Results:** Four QSAR models were developed, and the first model was selected due to its robust internal and external statistical performances as follows: $R^2 = 0.96165$, $R^2_{adj} = 0.952062$, $Q^2_{CV} = 0.934224$ and $R^2_{pred} = 0.749$. The mean $R^2$ of the scrambled models was 0.2336, the mean $Q^2$ was -0.3655, and the calculated $cR^2p$ was 0.8874. The designed analogues had predicted $pIC_{50}$ values of 6.045–7.030, with A1 showing the highest predicted activity (7.030). Applicability-domain assessment of the designed analogues identified A1 and A5 as belonging to the

chemical space represented by the model. Redocking of VRV gave a heavy-atom RMSD of 0.86 Å. A5 showed the most favourable predicted Glide XP docking score (−6.924 kcal $mol^{-1}$), compared with −6.790 kcal $mol^{-1}$ for VRV and −6.811 kcal $mol^{-1}$ for the refence compound AZD1208. The 500-ns simulations indicated stable protein–ligand behaviour for A1 and A5 under the simulated conditions. Predicted pharmacokinetic/ADMET profiles further supported A1 and A5 for further investigation as anti-breast cancer agents.

**Conclusions:** The integrated QSAR, docking, molecular dynamics and ADMET analyses identify A1 and A5 as computationally prioritised 3-cyanopyridine candidates for experimental synthesis, PIM-1 inhibition testing and MCF-7 antiproliferative evaluation.

**Keywords:** 3-cyanopyridine analogs; breast cancer; MCF-7 cell line; QSAR; molecular docking; molecular dynamics; pharmacokinetics; ADMET; in-silico drug design.

**Keywords:** 3-cyanopyridine analogs; breast cancer; MCF-7 cell line; QSAR; molecular docking; molecular dynamics; pharmacokinetics; ADMET; in-silico drug design.

## Introduction

Breast cancer remains a major global health concern, with substantial morbidity and mortality and persistent challenges related to treatment resistance, recurrence and toxicity. Its biological heterogeneity, including aggressive subtypes such as triple-negative breast cancer (TNBC), continues to drive the search for new therapeutic targets and structurally diverse anticancer agents [1,2].

Proviral integration site for Moloney murine leukemia virus 1 (PIM-1) is a constitutively active serine/threonine kinase implicated in cell-cycle progression, survival, metabolism and therapeutic resistance. PIM-1 is reported to be upregulated in breast cancer, including TNBC, and the PIM-1/STAT3 axis has been associated with aggressive disease phenotypes [3-8]. Despite progress in PIM kinase inhibitor discovery, clinically useful inhibitors with an appropriate balance of potency, selectivity, pharmacokinetic properties and safety remain needed [3,4].

3-Cyanopyridines are attractive heteroaromatic scaffolds for medicinal chemistry because their substitution patterns permit systematic structure–activity optimisation. Hussein et al. reported 3-cyanopyridine derivatives with activity against MCF-7 cells and PIM-1, supporting this scaffold for further investigation [9]. However, experimental activity alone does not establish

the structural determinants of potency or whether favourable target interactions are accompanied by dynamic stability and acceptable drug-like properties.

Computer-aided drug design can address these complementary questions by integrating quantitative structure–activity relationship (QSAR) modelling, molecular docking, molecular dynamics (MD) simulation and pharmacokinetic/ADMET prediction [10-16]. QSAR can relate molecular descriptors to biological activity and support the design of new analogues, whereas docking and MD can provide structural and dynamic information about ligand–PIM-1 interactions. Early ADMET prediction can additionally identify potential developability liabilities.

Accordingly, this study applied an integrated in-silico workflow to 3-cyanopyridine derivatives with reported MCF-7 activity. QSAR modelling was used to identify activity-associated descriptors and guide the design of new analogues; docking was then used to examine their predicted interactions with PIM-1, followed by MD simulation of selected complexes and pharmacokinetic/ADMET profiling. The overall aim was to prioritise novel 3-cyanopyridine derivatives for experimental evaluation as potential anti-breast-cancer candidates.

## Methods

### Data set retrieval, geometry optimization, and descriptor calculations

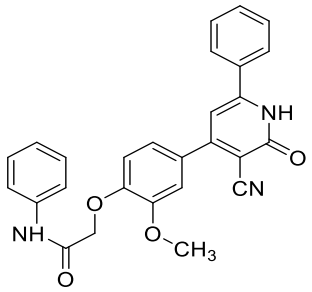
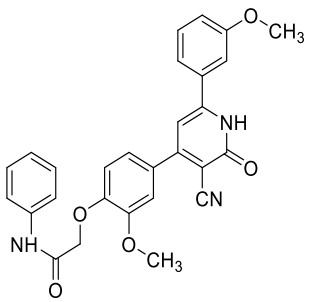
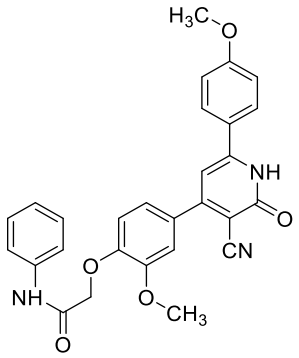
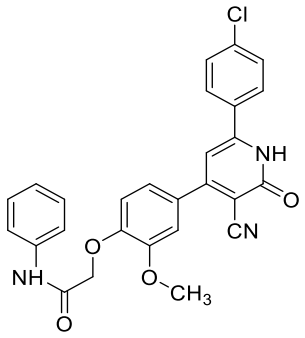
A dataset consisting of 24 structurally diverse 3-cyanopyridine derivatives and their experimentally reported biological activities against the MCF-7 cell line was collected from the literature [9]. The biological activity values were reported as $IC_{50}$ in μM. To obtain a normalised activity scale suitable for QSAR analysis, we converted the values to $pIC_{50}$ using Equation (1):

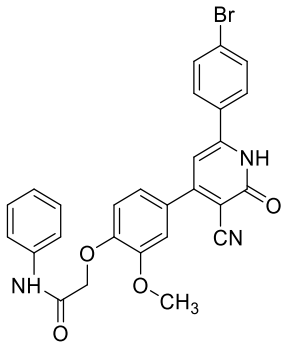
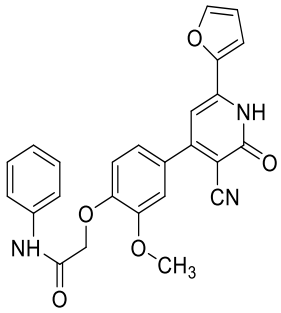
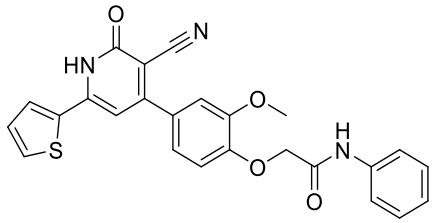
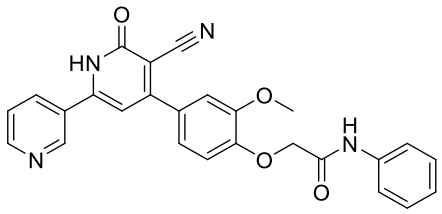
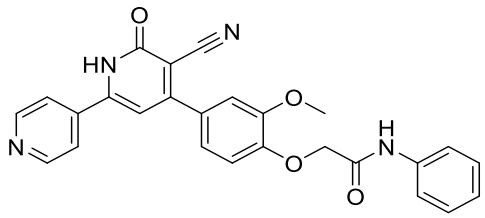
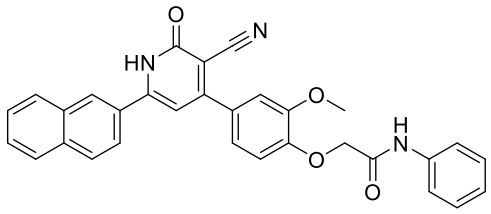
$$pIC_{50} = -\log_{10}(IC_{50} \times 10^{-6}) \quad (1)$$

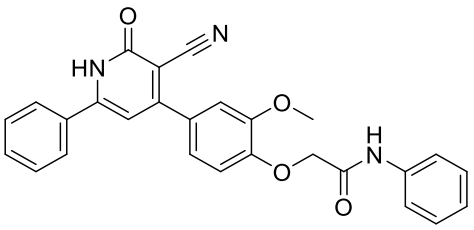
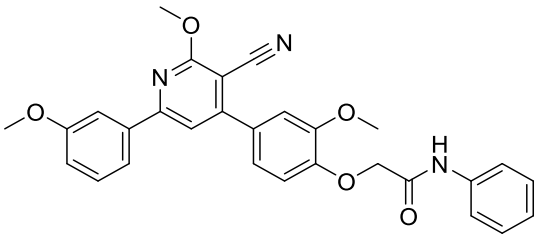
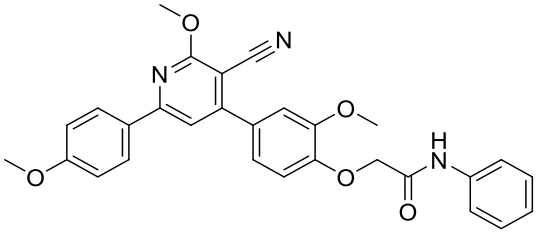
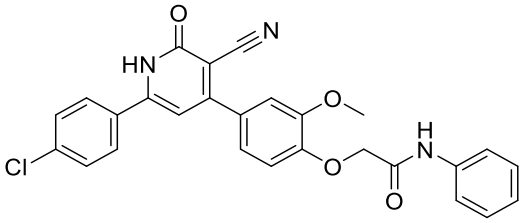
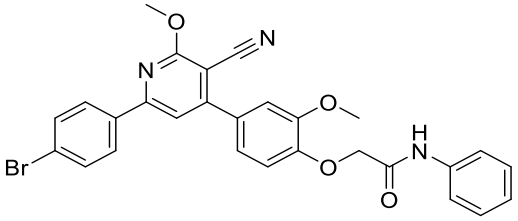
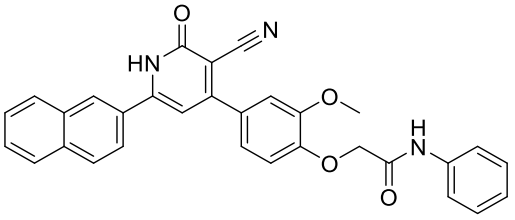
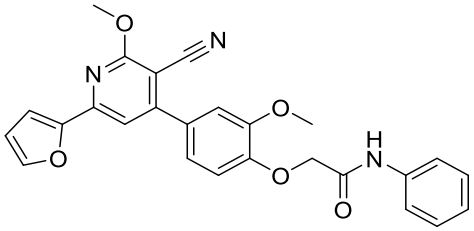
The 2D structures of the derivatives were drawn with ChemDraw 16.0 software and subsequently converted into 3D format by means of Spartan v14.0 software. Using the Spartan interface, geometry optimization of the molecules was carried out through Density Functional Theory (DFT) quantum mechanical calculations employing the B3LYP/631-G* basis set. Geometry optimization was carried out in order to identify the most stable conformer of the molecules, which possesses the lowest energy [17]. The geometrically optimized molecular structures were stored together in one folder in Spatial Document File (SDF) format and afterwards transferred to the PADEL descriptor toolkit, where the key molecular descriptors responsible for the anti-breast cancer activities of the molecules were computed [18]. The 2D

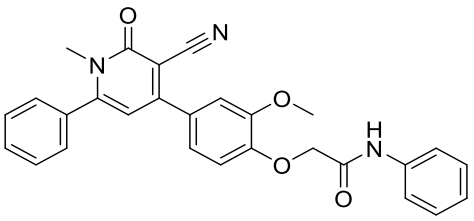
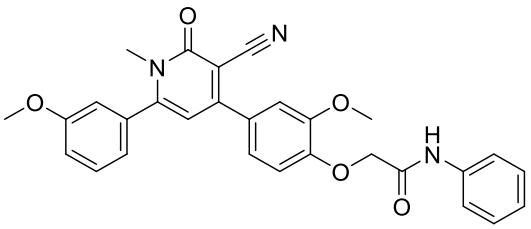
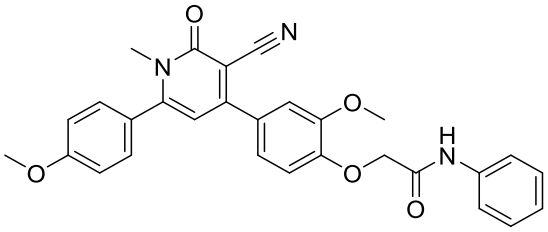
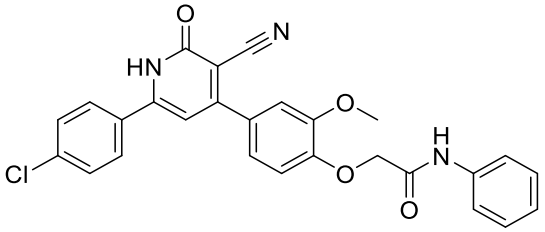
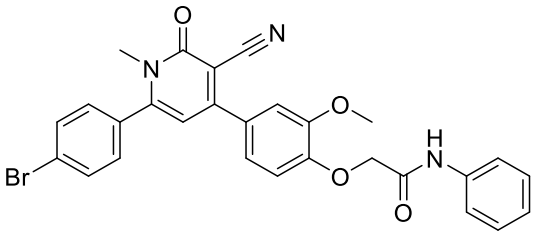
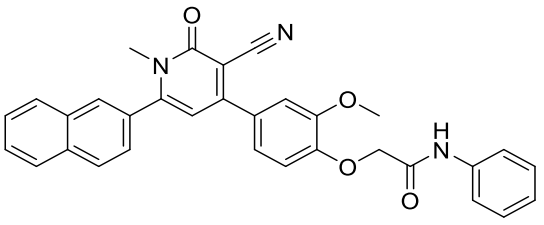
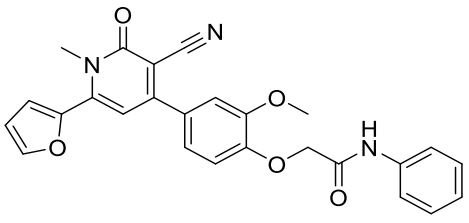
structures, experimental and predicted $pIC_{50}$ values of the 3-cyanopyridine molecules are presented in Table 1.

**Table 1:** 2D structures, experimental and the predicted $pIC_{50}$ of the 3-cyanopyridine molecules.

| S/N | 2D Structure | Exp $pIC_{50}$ | Pred $pIC_{50}$ | Residuals |
|---|---|---|---|---|
| 1 |  | 5.093 | 5.554 | -0.461 |
| 2 |  | 5.699 | 5.622 | 0.077 |
| 3 |  | 5.523 | 5.571 | -0.048 |
| 4 |  | 5.398 | 5.325 | 0.073 |

| | | | | |
|---|---|---|---|---|
| 5 |  | 5.301 | 5.348 | -0.047 |
| 6 |  | 5.222 | 5.222 | 0.000 |
| 7 |  | 5.155 | 5.089 | 0.065 |
| 8 |  | 5.097 | 5.079 | 0.018 |
| 9 |  | 5.046 | 5.259 | -0.213 |
| 10 |  | 5.000 | 4.976 | 0.023 |

| | | | | |
|---|---|---|---|---|
| 11 |  | 4.959 | 5.100 | -0.142 |
| 12 |  | 4.921 | 5.016 | -0.095 |
| 13 |  | 4.886 | 4.965 | -0.079 |
| 14 |  | 4.854 | 4.871 | -0.017 |
| 15 |  | 4.824 | 4.754 | 0.069 |
| 16 |  | 4.796 | 4.976 | -0.181 |
| 17 |  | 4.770 | 4.720 | 0.049 |

| | | | | |
|---|---|---|---|---|
| 18 |  | 4.745 | 4.745 | -0.000 |
| 19 |  | 4.721 | 4.838 | -0.117 |
| 20 |  | 4.699 | 4.789 | -0.089 |
| 21 |  | 4.678 | 4.871 | -0.193 |
| 22 |  | 4.658 | 4.544 | 0.124 |
| 23 |  | 4.638 | 4.651 | -0.012 |
| 24 |  | 4.620 | 4.634 | -0.014 |

**Data pretreatment and splitting**

The computed molecular descriptors of the compounds in the training set were first pretreated manually and afterward with data pretreatment software to cut down redundant and irrelevant descriptors. The pretreated data were then split into training and test sets via the Kennard Stone algorithm in data division software. In this study, the training set contains 16 molecules and the test set 8 molecules.

**QSAR model building and validation**

Training set molecules were used for QSAR model generation and internal validation, whereas test set molecules served for external validation and assessment of predictive strength. Feature selection of relevant descriptors and model generation relied on the genetic function algorithm (GFA) coupled with multi-linear regression (MLR) in Material Studio v8.0. GFA yields a population of models, not a single model [19]. In QSAR modeling, MLR links the independent variables (selected descriptors) to the dependent variables (experimental $pIC_{50}$) [19]. The best QSAR model was chosen using these statistical metrics: training set correlation coefficient ($R^2_{training}$), adjusted correlation coefficient ($R^2_{adj}$), cross-validation coefficient ($Q^2_{cv}$), and external test set correlation coefficient ($R^2_{pred}$). The equations defining the validation parameters are shown in Eqs. (2, 3, 4, and 5), respectively.

$$R^2 = \frac{\sum(Y_{exp} - Y_{pred})^2}{\sum(Y_{exp} - Y_{mtraining})^2} \quad (2)$$

$$R^2_{adj} = \frac{(N-1)R^2 - p}{N - p + 1} \quad (3)$$

$$Q^2_{cv} = 1 - \frac{\sum(Y_{pred} - Y_{exp})^2}{\sum(Y_{exp} - Y_{mtraining})^2} \quad (4)$$

$$R^2_{pred} = 1 - \frac{\sum(Y_{pred} - Y_{exp})^2}{\sum(Y_{exp} - Y_{mtraining})^2} \quad (5)$$

Where p is the independent variable's number, N represents the size of the training set sample, and the experimental, predicted, and mean activities of the training set samples are represented by $Y_{exp}$, $Y_{pred}$, and $Y_{mtraining}$.

**Y-scrambling test**

The Y-scrambling test was carried out to determine how robust the model is and to verify that the model did not arise from chance correlation. When the Y-scrambling test was conducted,

the experimental activities were shuffled while the descriptors remained unchanged, and new QSAR models were built across several trials. It was expected that the newly built QSAR models would show considerably lower $Q^2$ and $R^2$ values than the original model. The statistical parameter used in the Y-scrambling test is the coefficient of determination for the Y-scrambling test, $\boldsymbol{cR_p^2}$ (Eq. 5), and for a model to be considered reliable, its value must exceed 0.5 [20]

$$cR_p^2 = R \times [R^2 - R_r^2]^2 \quad (6)$$

$cR_p^2$ is coefficient of determination for Y-scrambling, R is the coefficient of determination for Y-randomization and $R_r$ is average 'R' of random models.

**Multicollinearity and Descriptor Independence**

Variance inflation factor (VIF) analysis was conducted to investigate possible multicollinearity among the descriptors included in the final model. VIF values were evaluated to examine the extent of correlation between the selected descriptors. A model is considered acceptable if the VIF values obtained from its descriptors ranged from 1 to 5; values > 10 indicate serious multicollinearity; whereas a value of more than 10 implies that the model needs to be reinvestigated [21]. The VIF values of the selected descriptors were calculated using Equation 7.

$$VIF = \frac{1}{1-R^2} \quad (7)$$

Where: $R^2$ is the correlation coefficient for each descriptor pair.

The correlation between the selected descriptors was further evaluated using Pearson correlation analysis. Descriptor independence was considered acceptable when the correlation coefficients remained below the acceptable ($R^2 \geq 0.6$) [21]. Analysis of variance (ANOVA) was used to evaluate the statistical significance of the descriptor–activity relationships at the [95%] confidence level.

The contribution of each descriptor in the reported model was evaluated by calculating its mean effect value (ME) using the equation 8 [22].

$$ME_j = \frac{\beta_j \sum_{j=1}^{i=n} d_{ij}}{\sum_j^m \beta_j \sum_i^n d_{ij}} \quad (8)$$

Where $MF_j$ is the mean effect of a descriptor j in a model, βj depicts the coefficient of the descriptor J in the model and $d_{ij}$ is the value of the descriptor in the data matrix per sample in the training set, m is the descriptor's numbers that turn up in the model and n is the number of

samples in the training set. The magnitude and sign of the descriptors indicate the various directions of either increase or decrease in the biological activity of a molecule [22].

**Applicability Domain Analysis**

The applicability domain of the developed QSAR model was evaluated using the leverage approach and the Williams plot. The leverage and its cut-off values were calculated using 9 and 10.

$$h_i = x_i(X^T X)^{-1} X_i^T \quad (9)$$

$$h^* = \frac{3(Q+1)}{N} \quad (10)$$

Where x represents the descriptor vector of a referred molecule, X refers to the descriptor matrix originating from the modeling set descriptor values, Q represents the number of independent variables in the developed model, and N is the number of training set compounds. Compounds with standardized residuals beyond ±3 were considered potential response outliers, whereas compounds with leverage values greater than h* were considered outside the leverage-based applicability domain [23]. The applicability domain analysis was used to determine whether the predictions for the designed compounds were made within the chemical space represented by the training dataset.

**Molecular Design of New Compounds**

Based on the results of the QSAR analysis, an in-silico design approach was used to generate more effective 3-cyanopyridine analogues. The developed QSAR equation was employed to predict the biological activities of the newly designed molecules. The predicted activities were then compared with those of the template compound 4.

**Molecular Docking Studies**

**Software and Protein Selection**

Molecular docking studies were carried out using Schrödinger Release 2024-1 (Maestro version 13.9, LLC, New York, NY). The X-ray crystal structure of human PIM-1 kinase in complex with a staurosporine analog (PDB ID: 2OBJ, resolution 2.0 Å) was retrieved from the RCSB Protein Data Bank (www.rcsb.org). Structure 2OBJ was selected because it provides a well-defined ATP-binding pocket with conserved key residues LYS67, GLU89, ASP186 and hinge residue GLU121, and is widely used for PIM-1 docking studies. The co-crystallized ligand in 2OBJ and clinical stage AZD1208 were used as reference inhibitors [24-26].

**Protein Preparation**

The protein structure was prepared using the Protein Preparation Wizard module in Maestro. Bond orders were assigned using the Chemical Component Dictionary database, zero-order bonds to metals were created, disulfide bonds were assigned, and missing hydrogens were added. Missing side chains and loops were filled using Prime module. Water molecules beyond 5.0 Å from the co-crystallized ligand were deleted, and waters beyond 3.0 Å not participating in water bridges were also removed. The ionization states were generated at pH 7.0 ± 2.0 using PROPKA, and H-bond assignment was optimized. Finally, restrained energy minimization was performed using the OPLS4 force field, converging heavy atoms to RMSD 0.30 Å [24,27,28]. The prepared protein structure is shown in Figure 1.

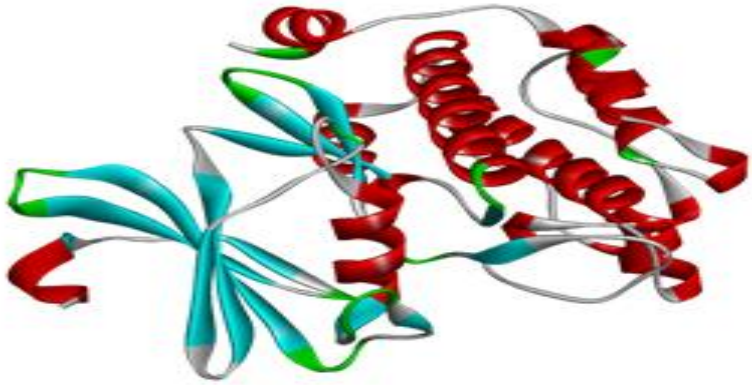


**Figure 1.** Three-dimensional structure of the prepared PIM-1 receptor (PDB 2OBJ).

**Ligand Preparation**

The 3D structures in SDF format of the original dataset, the designed analogues and the reference compounds were imported into Maestro Workspace Navigator and converted to low-energy 3D conformers using LigPrep tool. The OPLS4 force field was employed for energy minimization. Possible ionization and tautomeric states were generated using Epik at pH 7.0 ± 2.0 with desalting option [24,28].

**Receptor Grid Generation**

Receptor grid was generated using Glide's Receptor Grid Generation tool. The grid box was centered at the coordinates X = –18.07 Å, Y = 35.6 Å, Z = –0.98 Å, around the centroid of the co-crystallized ligand, with van der Waals radius scaling factor 1.0 and partial charge cutoff 0.25 [24,28,29].

**Molecular Docking and Validation**

Molecular docking was performed using Glide module in Maestro. Compounds were docked flexibly within the generated grid using the Extra Precision (XP) scoring function with post-docking minimization. Glide docking score was used to rank docked poses, and XP descriptor data were used to analyze interactions [24,28-30].

The docking protocol was validated by re-docking the native co-crystallized ligand into its binding pocket. The root mean square deviation (RMSD) between the native and re-docked pose was calculated by structure superposition. An RMSD ≤ 2.0 Å was considered as a valid protocol, confirming the ability of the algorithm to reproduce the experimental binding mode [26,31].

**Interaction Analysis**

Protein-ligand interactions were visualized using Ligand Interaction Diagram in Maestro and Biovia Discovery Studio Visualizer 2025 Client. Key interactions including hydrogen bonds, hydrophobic contacts, π-π stacking and salt bridges with catalytic residues were analyzed.

**Molecular Dynamics Simulation**

For the selected PIM-1–A1 and PIM-1–A5 complexes, molecular dynamics simulations were run in explicit solvent over a period of 500 ns, employing the General Amber Force Field (GAFF) [32,33]. Simulations were carried out with the Amber18 software package [34]. TIP3P water molecules were used to solvate each complex within a cubic periodic box, and sodium ions were added for neutralisation [35]. To relieve steric clashes, energy minimisation was carried out in several successive stages. The first minimisation step, targeting water molecules and counterions, comprised 2000 steepest-descent steps followed by 3000 conjugate-gradient steps. An analogous procedure was used to relax the hydrogen atoms and water molecules. Afterwards, the whole system was subjected to a final minimisation round made up of 5000 steepest-descent steps followed by 5000 conjugate-gradient steps [35].

Over 200 ps, the system was heated progressively from 0 to 300 K, after which 300 ps of density equilibration was performed under constant-volume conditions, with positional restraints imposed on the protein atoms. The next equilibration stage consisted of a 10 ns unrestrained MD simulation at constant pressure, permitting the protein–ligand complexes to relax fully before the production runs. A Langevin thermostat, set to a collision frequency of 2 $ps^{-1}$, was used to control the temperature. Non-bonded interactions were cut off at 10 Å, and long-range electrostatic interactions were handled with the Particle Mesh Ewald (PME) approach. The SHAKE algorithm was applied to constrain bond lengths involving hydrogen atoms [35].

Production MD simulations ran for 500 ns at 300 K. From the resulting trajectories, analyses were performed for backbone and ligand RMSD, per-residue RMSF, radius of gyration (Rg), solvent-accessible surface area (SASA), and protein–ligand hydrogen-bond occupancy [35].

## Drug-Likeness and ADMET Prediction

The drug-likeness and pharmacokinetic properties of the designed compounds were predicted using SwissADME (https://www.swissadme.ch/) and the pkCSM web server (https://biosig.lab.uq.edu.au/pkcsm/). The assessed endpoints included molecular weight, hydrogen-bond donors and acceptors, topological polar surface area, lipophilicity, gastrointestinal absorption, blood–brain barrier and CNS permeability, P450 inhibition, total clearance, drug-likeness rules, synthetic accessibility, and selected toxicity endpoints. Because these values are computational predictions, they were interpreted as preliminary developability indicators rather than experimentally established pharmacokinetic or safety properties.

## Results and Discussion

### Result of QSAR studies

Four (4) QSAR models were built with Material Studio v8.0 in order to provide an efficient prediction of the anti-breast cancer activities of the molecules under study. The mathematical expressions corresponding to the generated models are presented below:

***Model 1***

$pIC_{50}$ = ─1.319372518 * AATS0s ─ 4.426202062 * GATS5i ─ 0.327544961 * CrippenLogP + 17.686394605

***Model 2***

$pIC_{50}$ = 0.034948948 * AATS4v ─ 2.530853463 * AATS4s ─ 0.363051604 * CrippenLogP + 8.395491313

***Model 3***

$pIC_{50}$ = 0.375922666 * SpMAD_Dzs ─ 15.431711505 * BCUTc-1h ─ 0.284011840 * CrippenLogP + 5.684523366

***Model 4***

pIC50 = ─1.184272836 * AATS0s + 4.466455887 * MATS5i ─ 0.317205915 * CrippenLogP + 12.680179326

**Results of internal and external validations of the developed models**

Tables 2 and 3 display the findings from both internal and external validations of the QSAR models that were developed, and these models were determined to satisfy the minimum criteria for an acceptable QSAR model [22]. The model that exhibited the strongest internal validation metrics, model 1, was selected to undergo external validation procedures involving the test set molecules. Using this model, the activities of the test set molecules were predicted, and Eq. 5 was applied to calculate the predicted correlation coefficient ($R^2_{pred}$) so as to evaluate the predictive capability of the selected model. Table 2 provides the experimental and predicted activities of the test set molecules, as well as the step-by-step computations of the external prediction correlation coefficient ($R^2_{pred}$).

**Table 2:** Minimum acceptable values of the developed QSAR models

| **Symbol** | **Definition** | **Threshold value** | **Model 1** | **Model 2** | **Model 3** | **Model 4** |
|---|---|---|---|---|---|---|
| $R^2$ | Correlation coefficient of the training set | ≥0.6 | 0.961650 | 0.947702 | 0.946816 | 0.946093 |
| $R^2_{adj}$ | Adjusted $R^2$ | ≥0.6 | 0.952062 | 0.934627 | 0.933520 | 0.932616 |
| $Q^2_{cv}$ | Cross validation coefficient | ≥0.5 | 0.934224 | 0.885844 | 0.900184 | 0.909052 |
| $R^2$ - $Q^2_{cv}$ | Difference between $R^2$ and $Q^2_{cv}$ | ≤0.3 | 0.027426 | 0.061858 | 0.046632 | 0.037041 |
| P(95%) | Confidence interval at 95% confidence level | <0.05 | 0.052477 | 0.061281 | 0.061797 | 0.062216 |

**Table 3:** Results of the external validation of the selected model

| **ID** | **$Y_{exp}$** | **$Y_{pred}$** | **$(Y_{exp} - Y_{pred})$** | **$(Y_{exp} - Y_{mtrng})$** | **$(Y_{exp} - Y_{pred})^2$** | **$(Y_{exp} - Y_{mtrng})^2$** |
|---|---|---|---|---|---|---|
| 1 | 6.000000 | 5.554004 | 0.445996 | 1.064591427 | 0.198912 | 1.133355 |
| 9 | 5.045757 | 5.258563 | -0.21281 | 0.110348918 | 0.045286 | 0.012177 |
| 13 | 4.886057 | 4.964564 | -0.07851 | -0.049351925 | 0.006163 | 0.002436 |
| 16 | 4.79588 | 4.976389 | -0.18051 | -0.139528556 | 0.032584 | 0.019468 |
| 19 | 4.721246 | 4.837922 | -0.11668 | -0.214162174 | 0.013613 | 0.045865 |
| 20 | 4.69897 | 4.788805 | -0.08983 | -0.236438569 | 0.00807 | 0.055903 |
| 21 | 4.677781 | 4.871 | -0.19322 | -0.257627868 | 0.037334 | 0.066372 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 22 | 4.657577 | 4.53394 | 0.123637 | -0.277831254 | 0.015286 | 0.07719 |
| | Sum: 39.48327 | | | Sum | 0.357249 | 1.412767 |
| | $Y_{mtrng}$: 4.935409 | | | | | |

$$R^2_{pred} = 1 - \frac{\sum(Y_{exp} - Y_{pred})^2}{\sum(Y_{exp} - Y_{mtrng})^2} = 1 - \frac{0.357249}{1.412767} = 0.747$$

Figures 2 and 3 show the plots of experimental against predicted activities for the training and test sets, respectively.

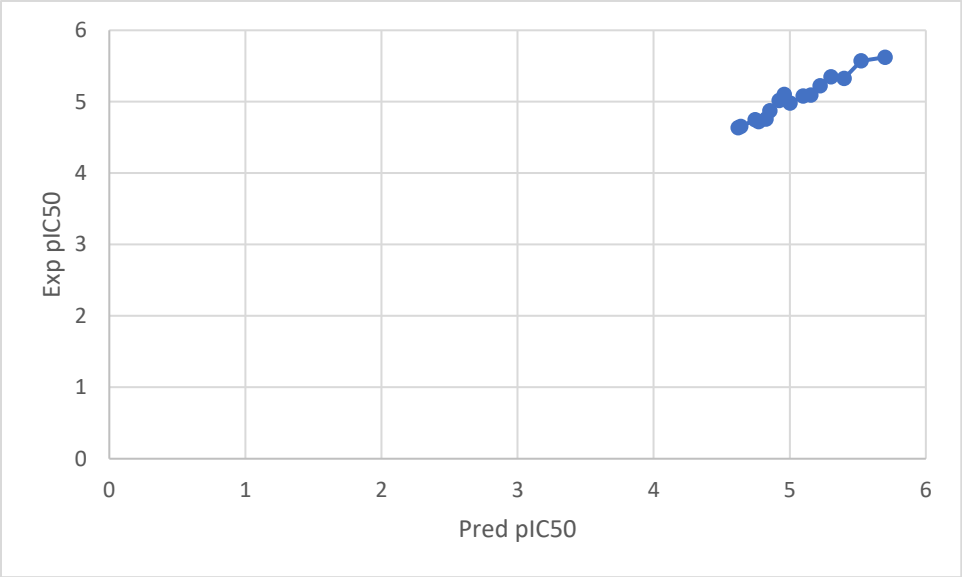



**Figure 2.** A plot of experimental against predicted activities ($pIC_{50}$) for the training set

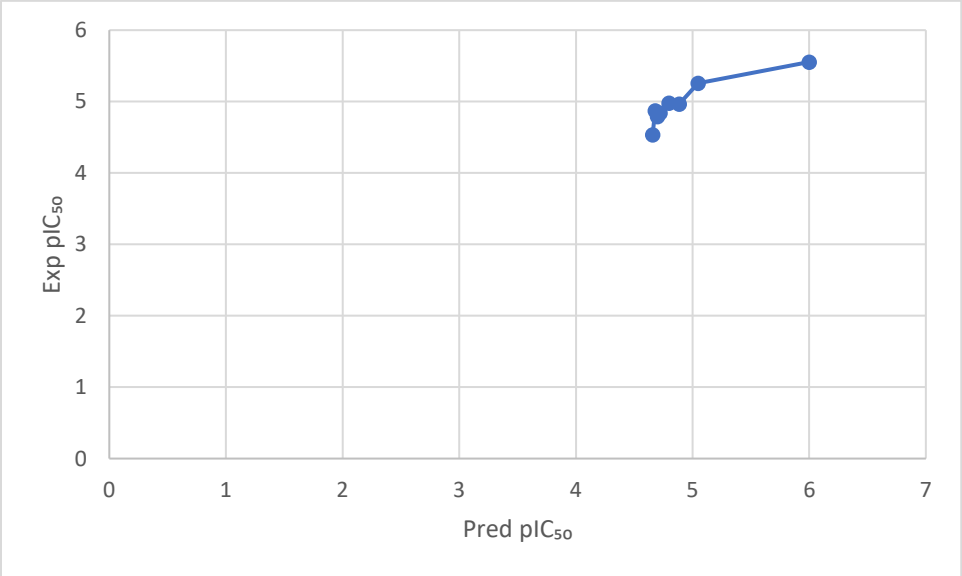



**Figure 3.** A plot of experimental against predicted activities ($pIC_{50}$) for the test set

**Results of the Y-Scrambling Test**

The robustness of the QSAR model against chance correlation was assessed by randomly permuting the response variable ($pIC_{50}$) while retaining the original descriptor matrix and repeating model construction for 10 iterations. The mean $R^2$ of the scrambled models was 0.2336, and the mean $Q^2$ was -0.3655, both of which are significantly lower than $R^2 = 0.96165$ and $Q^2 = 0.96165$ respectively for the original model. The calculated $cR^2p$ of the scrambled models was 0.8874, which is higher than the value of 0.5 required for passing this test. The developed model was considered resistant to chance correlation since it satisfied all the required validation criterion as illustrated in Table 4 presents the Y-scrambling results.

**Table 4. Results of the Y-scrambling test for 10 iterations**

| **MODEL TYPE** | **R** | **$R^2$** | **$Q^2$LOO** |
|---|---|---|---|
| Original | 0.980637 | 0.96165 | 0.934224 |
| Random 1 | 0.210515 | 0.044317 | -0.63143 |
| Random 2 | 0.498564 | 0.248566 | -0.52029 |
| Random 3 | 0.31375 | 0.098439 | -0.4303 |
| Random 4 | 0.286332 | 0.081986 | -0.65967 |
| Random 5 | 0.263943 | 0.069666 | -0.58113 |
| Random 6 | 0.183203 | 0.033563 | -0.56344 |
| Random 7 | 0.461994 | 0.213438 | -0.46981 |
| Random 8 | 0.574428 | 0.329968 | -0.17946 |
| Random 9 | 0.519193 | 0.269561 | -0.65701 |
| Random 10 | 0.467076 | 0.21816 | -0.2624 |

| **Random Models Parameters** | |
|---|---|
| Average R | 0.432694 |
| Average $R^2$ | 0.233574 |
| Average $Q^2$ | -0.36552 |
| $cRp^2$ | 0.887377 |

**Descriptor Interpretation**

The Variance inflation factor (VIF) values of all the three descriptors were found in the range of 1-5 (Table 5), which proved the absence of collinearity. Pairwise Pearson's correlation coefficients were found to be $|r|<0.8$. Moreover, ANOVA p-values for all three descriptors were found to be $<0.05$ (Table 5), which proved their significance for the model.

**Mean effect**

The impact and contribution of every descriptor in a QSAR model were assessed by calculating its mean effect value (MF) [22]. The mean effect values for the chosen descriptors are presented in Table 5. The size and direction of a descriptor are linked to the biological activity of a compound [17]. A descriptor with a negative sign indicates that the biological activity of a

compound declines as its value rises, whereas a positive sign implies that biological activity increases as its value rises.

In this work, the most significant molecular descriptor is AATS0s, an averaged 2D autocorrelation of lag 0, weighted by I-state (intrinsic state). Its mean effect value is 0.48274, indicating that a rise in its value influences the anti-cancer activities of the compounds in a positive way. The least significant descriptor is CrippenLogP, or the Logarithm of octanol/water partition coefficient predicted by the Wildman-Crippen method; its mean effect value is 0.142362, which likewise indicates that a rise in its value influences the anti-cancer activities of the compounds positively.

Another significant descriptor present in the model is GATS5i, a 2D Geary autocorrelation of lag 5, weighted by first ionization potential. Its positive mean effect value of 0.3749 indicated that the growing biological activity of the compounds is linked to the increase in the value of this descriptor. A chart displaying the mean effect values of the descriptors included in the model is presented in Figure 4.

**Table 5:** Pearson's correlation matrix, P-value, VIF and ME of the developed model

| | *AATS0s* | *GATS5i* | *CrippenLogP* | *P-value* | *VIF* | *ME* |
|---|---|---|---|---|---|---|
| AATS0s | 1 | | | 1.4945E-07 | 2.331279 | 0.482736 |
| GATS5i | -0.37732 | 1 | | 8.31162E-05 | 1.415352 | 0.374902 |
| CrippenLogP | -0.75483 | 0.539772 | 1 | 3.61718E-08 | 2.8214 | 0.142362 |

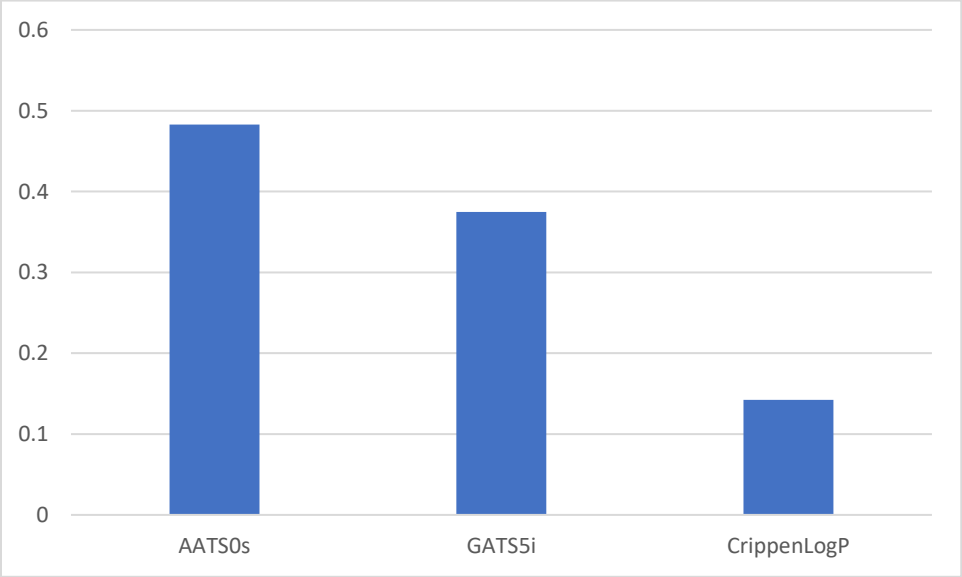



**Figure 4. Mean effect values of the descriptors used in the selected model**

**Applicability Domain**

The warning leverage (h = 0.75) was calculated, and it was found that the leverages of all compounds from both sets were below this threshold and the standardised residuals were within ±3, which confirmed that all compounds were inside the applicability domain of the model (Figure 5).

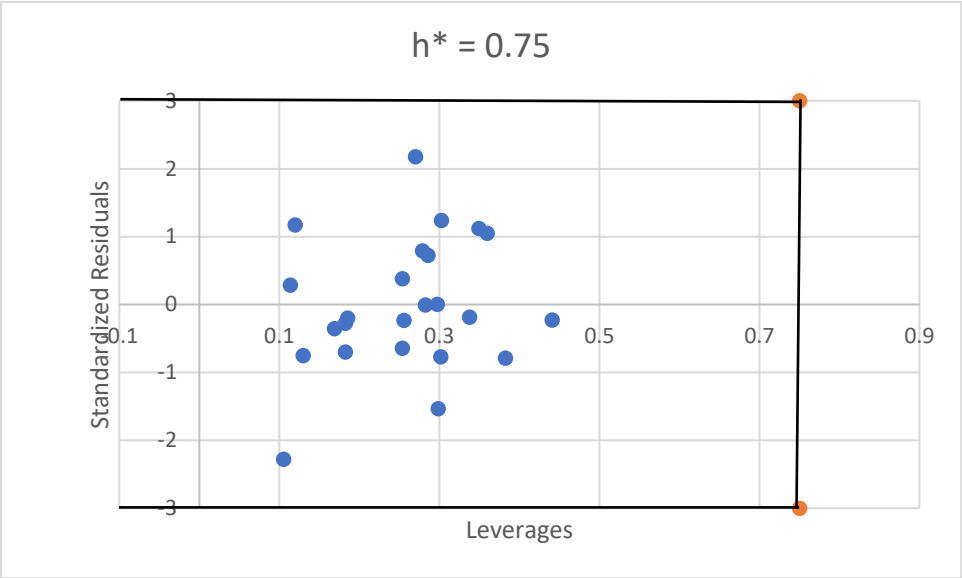



**Figure 5.** William's plot of the developed QSAR model (h* = 0.75).

**Design of Novel Compounds**

The training set compound 4 (Figure 6) was selected as a design template based on its relatively high experimental activity ($pIC_{50}$ = 5.398) and low residual (0.073) together with a low leverage value (0.120). The structural modifications were introduced at positions A and B of the scaffold, mainly by the incorporation of electron-donating groups (dimethylamino, amino, hydroxy, methoxy), and an imidazoline-containing side chain.

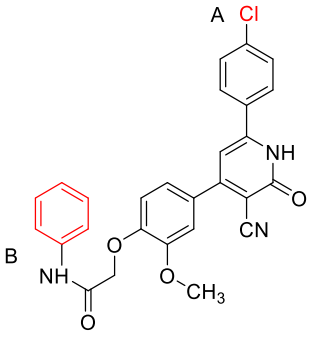
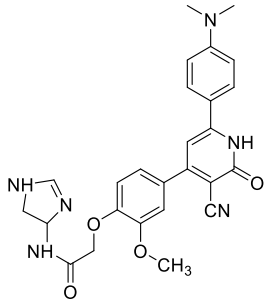
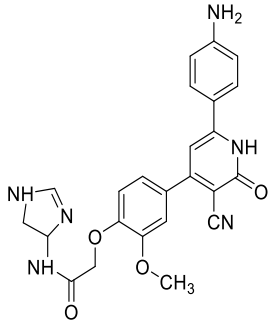
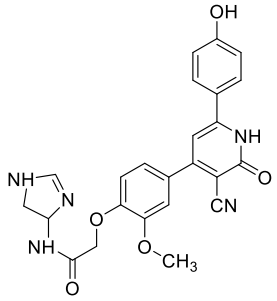


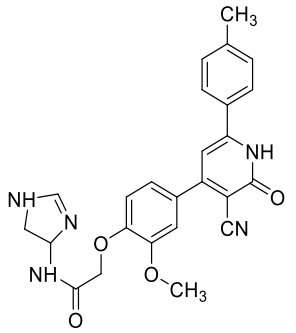
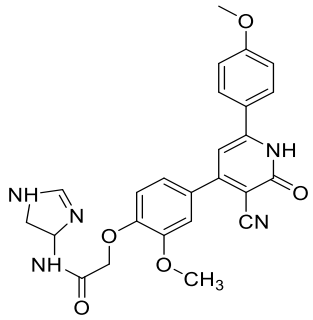

**Figure 6.** Chemical structure of the design template, compound 4.

Five novel analogues (A1–A5) were designed. Their predicted $pIC_{50}$ values ranged from 6.045 to 7.030, compared with an experimental $pIC_{50}$ of 5.398 for the template compound 4.

A1 showed the highest predicted $pIC_{50}$ (7.030), followed by A4 (6.954), A2 (6.704), A5 (6.625) and A3 (6.045). The structural modifications were selected to explore the effect of different substituents on the descriptor space represented by the final QSAR model. The structures, molecular descriptors, and predicted $pIC_{50}$ values of the designed compounds are presented in Table 6.

**Table 6.** Structures, molecular descriptors and predicted $pIC_{50}$ values of the designed compounds

| ID | Structure | AATS0s | GATS5i | CrippenLogP | Pred $pIC_{50}$ |
|---|---|---|---|---|---|
| A1 |  | 4.3993 | 0.9870 | 1.4749 | 7.030 |
| A2 |  | 4.8544 | 0.9608 | 0.9911 | 6.704 |
| A3 |  | 5.2729 | 0.9759 | 1.1145 | 6.045 |

| | | | | | |
|---|---|---|---|---|---|
| A4 |  | 4.5857 | 0.9515 | 1.4369 | 6.954 |
| A5 |  | 4.7134 | 0.9893 | 1.4175 | 6.625 |

Applicability-domain assessment of the designed analogues revealed that designed compounds A1 and A5 are within the chemical space predicted by the model, whereas the other analogues (A2, A3 and A4) are outside the applicability domain due to their leverages being higher than the warning leverage (h* = 0.75). The calculated leverages of the designed analogues are presented in Table 7.

**Table 7.** Calculated leverages of the designed analogues

| **Compound** | **Leverage ($h_i$)** | **Status** |
|---|---|---|
| A1 | 0.678637 | < h*; inside AD |
| A2 | 0.967995 | > h*; outside AD |
| A3 | 0.872192 | > h*; outside AD |
| A4 | 0.999342 | > h*; outside AD |
| A5 | 0.481834 | < h*; inside AD |

**Molecular Docking Results**

Redocking of the co-crystallised ligand VRV reproduced the experimental binding mode with a heavy-atom RMSD of 0.86 Å, well within the accepted limit of ≤2.0 Å, confirming the reliability of the Glide XP protocol. The docking scores for the original series (compounds 1 – 24), the designed compounds (A1-A5), the co-crystallized ligand VRV and the reference compound AZD1208 are listed in Table 8. Four of the designed compounds returned more favourable scores than the template compound 4.

**Table 8:** Docking scores of the original data set (1-24), designed compounds (A1-A5) and the reference compounds.

| Entry Name | Docking score (kcal/mol) |
|---|---|
| A5 | -6.924 |
| AZD1208 | -6.811 |
| VRV | -6.79 |
| A2 | -6.584 |
| A3 | -6.546 |
| A1 | -6.521 |
| 4 | -6.502 |
| 5 | -6.406 |
| 8 | -6.215 |
| 14 | -6.196 |
| 24 | -6.175 |
| 7 | -6.175 |
| A4 | -6.047 |
| 6 | -6.034 |
| 16 | -5.984 |
| 19 | -5.737 |
| 12 | -5.625 |
| 9 | -5.561 |
| 11 | -5.542 |
| 10 | -5.528 |
| 21 | -5.396 |
| 18 | -5.234 |
| 22 | -5.221 |
| 1 | -5.194 |
| 2 | -5.15 |
| 23 | -5.142 |
| 15 | -5.131 |
| 13 | -4.855 |
| 17 | -4.647 |
| 20 | -4.544 |
| 3 | -4.413 |

**Binding Interaction Analysis of the Most Promising Compounds**

**Compound A5**

Compound A5 exhibited the best XP docking score among the designed A-series (–6.924 kcal $mol^{-1}$). Its pose was stabilised by conventional hydrogen bonds with LYS67 and ASP131, a water-mediated interaction involving HOH A:464, and extensive hydrophobic contacts with ALA65, VAL52, LEU44, LEU177, ILE104 and ILE185. The docking score of A5 was slightly more favourable than those of the co-crystallised ligand VRV (–6.79 kcal $mol^{-1}$) and the

reference inhibitor AZD1208 (–6.811 kcal $mol^{-1}$), indicating that A5 can establish a competitive interaction profile inside the PIM-1 ATP-binding site (figure 7).

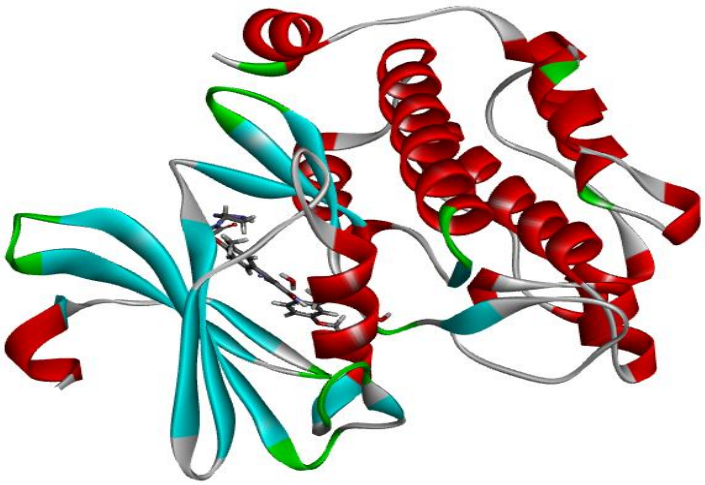


**Figure 7a.** Three-dimensional interaction diagrams of designed compound A5 with the PIM-1 active site.

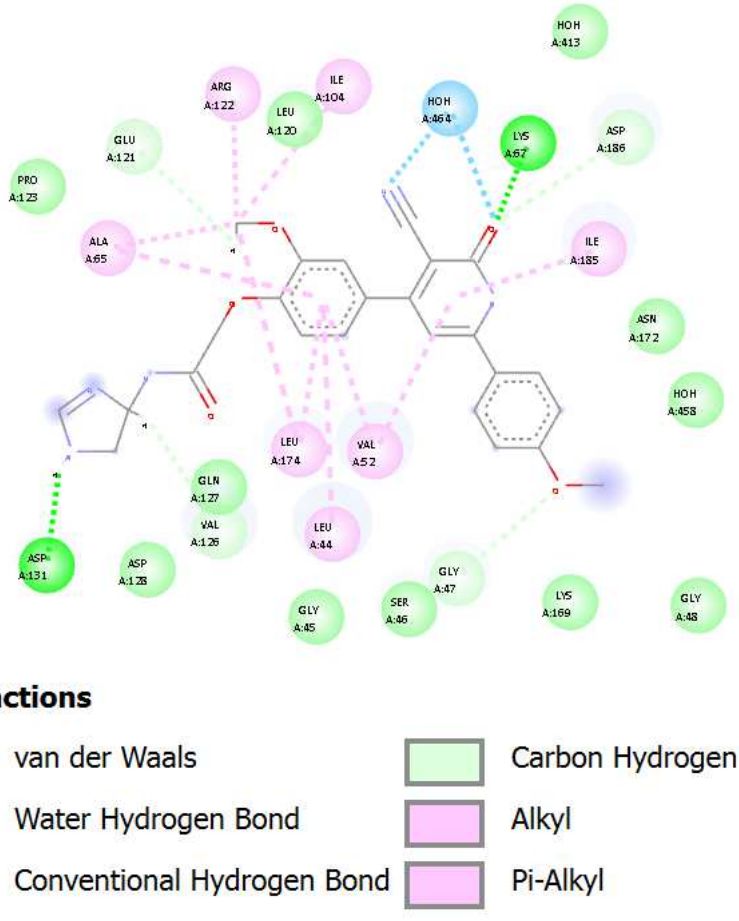


**Figure 7b.** Two-dimensional interaction diagrams of designed compound A5 with the PIM-1 active site.

**Compound A1**

Compound A1 with the docking score of –6.521 kcal $mol^{-1}$ adopted a binding orientation that also featured hydrogen bonds to LYS67 and ASP131, a water-mediated contact with HOH A:464, and hydrophobic interactions with ALA65, VAL52, LEU120, LEU174, ILE104 and ILE185, together with a π–anion interaction involving ASP186. The overall interaction pattern was similar to that of A5, although the hydrophobic packing appeared slightly less optimal, consistent with its comparatively lower docking score (Figure 8).

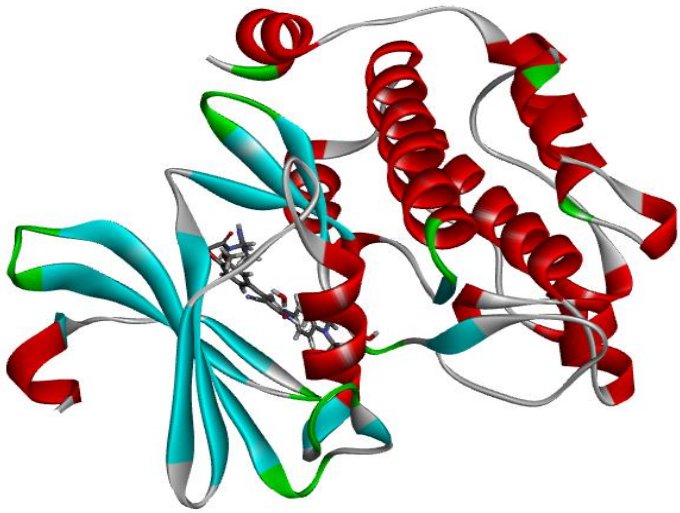


**Figure 8a.** Three-dimensional interaction diagrams of designed compound A1 with the PIM-1 active site.

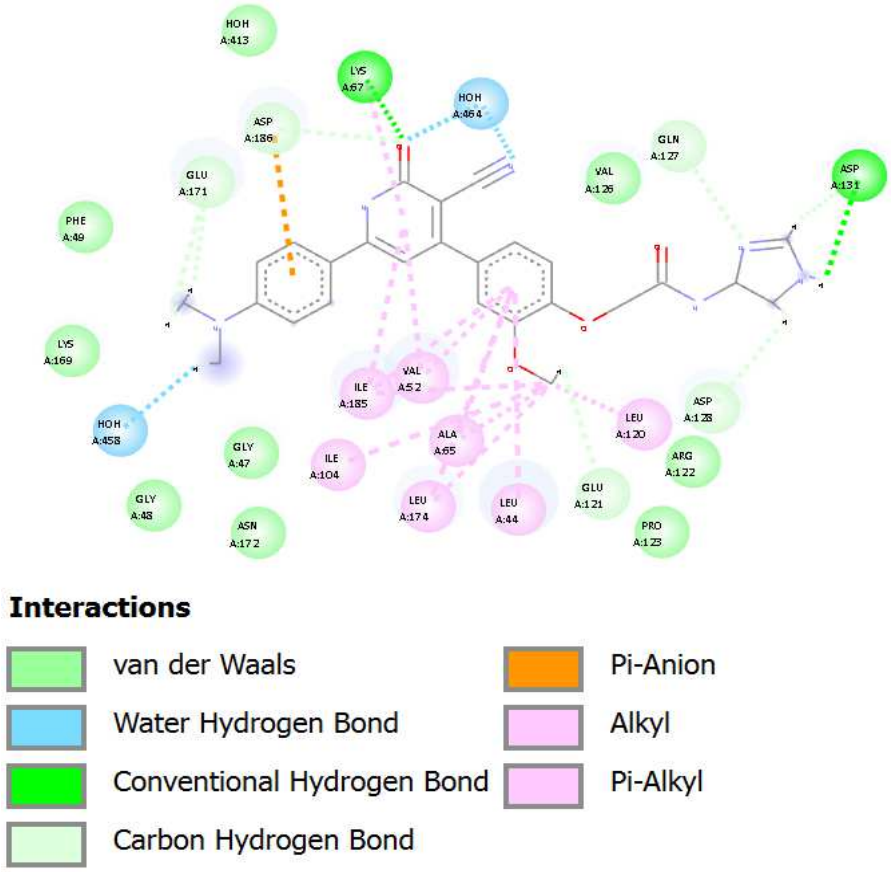



**Figure 8b.** Two-dimensional interaction diagrams of designed compound A1 with the PIM-1 active site.

### Reference Compound AZD1208

AZD1208 formed extensive hydrophobic interactions with the same set of non-polar residues identified for A5 and A1, in addition, it engaged in hydrogen-bonding interactions with ASP122/GLY47 and π–anion contact with ASP186. Its predicted docking score of –6.811 kcal $mol^{-1}$ was comparable to the value for A5, indicating that PIM-1 kinase can accommodate a range of polar interactions while maintaining high predicted affinity (Figure 9).

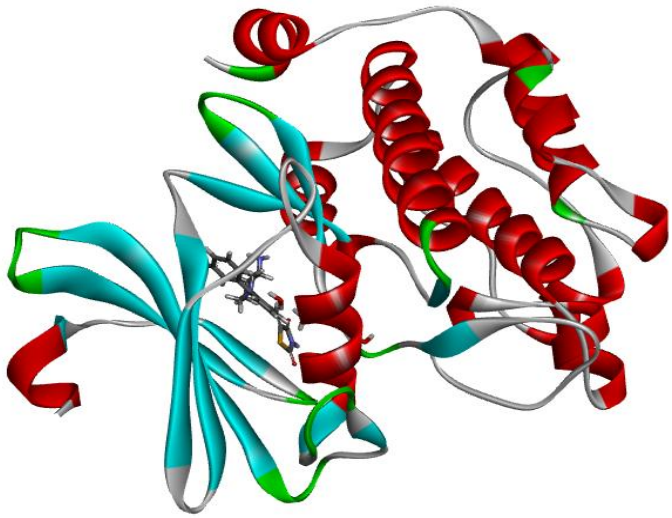


**Figure 9a.** Three-dimensional interaction diagrams of the reference inhibitor AZD1208 with the PIM-1 active site.

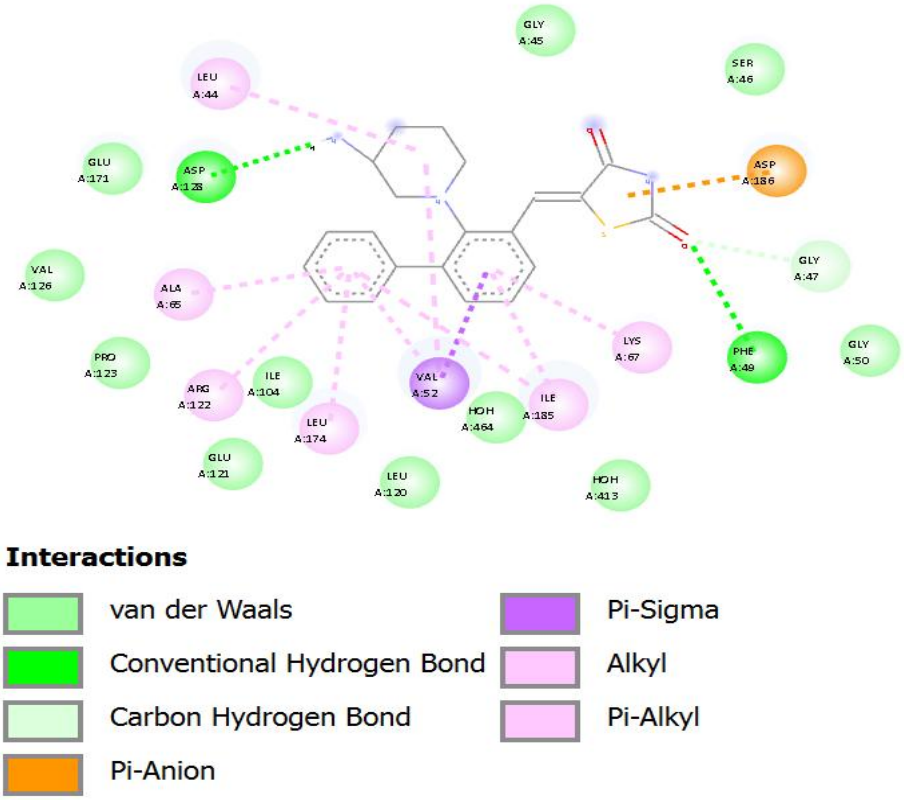



**Figure 9b.** Three-dimensional interaction diagrams of the reference inhibitor AZD1208 with the PIM-1 active site.

**Molecular Dynamics Simulation of A1 and A5 in Complex with PIM-1 Kinase**

To assess the binding stability of the designed 3-cyanopyridine analogs, we performed 500 ns molecular dynamics simulations for PIM-1 kinase in complex with compound A1 and compound A5. Long timescale simulations are recommended for kinases because of the slow conformational transitions of their loop regions that dictate ligand dissociation rates [36].

**Conformational Stability and RMSD Analysis**

The root-mean-square deviation (RMSD) of the protein backbone was monitored to assess conformational stability during the simulations (Figure 10a). Both protein–ligand systems showed RMSD values of approximately 1.0–1.6 Å and reached an apparent equilibration phase within the first ~50 ns, after which no major long-term drift was observed. At approximately 200 ns, the A5 complex showed RMSD values of ~1.25–1.35 Å compared with ~1.40–1.50 Å for the A1 complex, while an increase in RMSD was observed for A1 around 350 ns. These observations indicate somewhat more stable backbone behaviour for the A5 complex under the simulated conditions.

**Structural Compactness – Rg and SASA Analysis**

The radius of gyration (Rg) was calculated to monitor changes in protein structural compactness (Figure 10b). The A5 complex showed Rg values of approximately 19.35–19.45 Å, compared with 19.25–19.35 Å for A1, indicating only modest differences in global compactness. The solvent-accessible surface area (SASA) was also monitored (Figure 10c). A5 showed slightly higher SASA values (~13,200–13,400 Å²) than A1 (~12,700–13,000 Å²). These differences were relatively small and are therefore interpreted as minor conformational adjustments rather than evidence of substantial unfolding or destabilisation.

**Residual Flexibility and RMSF Analysis**

Root-mean-square fluctuations (RMSF) were calculated to determine the flexibility of individual amino-acid residues in the protein (Figure 10d). RMSF analysis aids in identifying the regions of the protein that determine the binding specificity and flexibility of the ligand. In general, the RMSF values of the PIM-1 residues were distributed around the same trend for the two protein-ligand complexes. The catalytic and hinge residues characteristic of ATP binding sites exhibited low RMSF values (<1.0 Å), indicating that they adopted stable conformations during the MD run [36]. Most flexible residues were observed in the loop regions spanning ∼50, 70, 230 and the C-terminus, typical for protein kinases. Notably, no significant variations

in RMSF were observed for the active site residues for either complex. This result suggests that there is no induced flexibility in the hinge region upon binding of A1 and A5 [37].

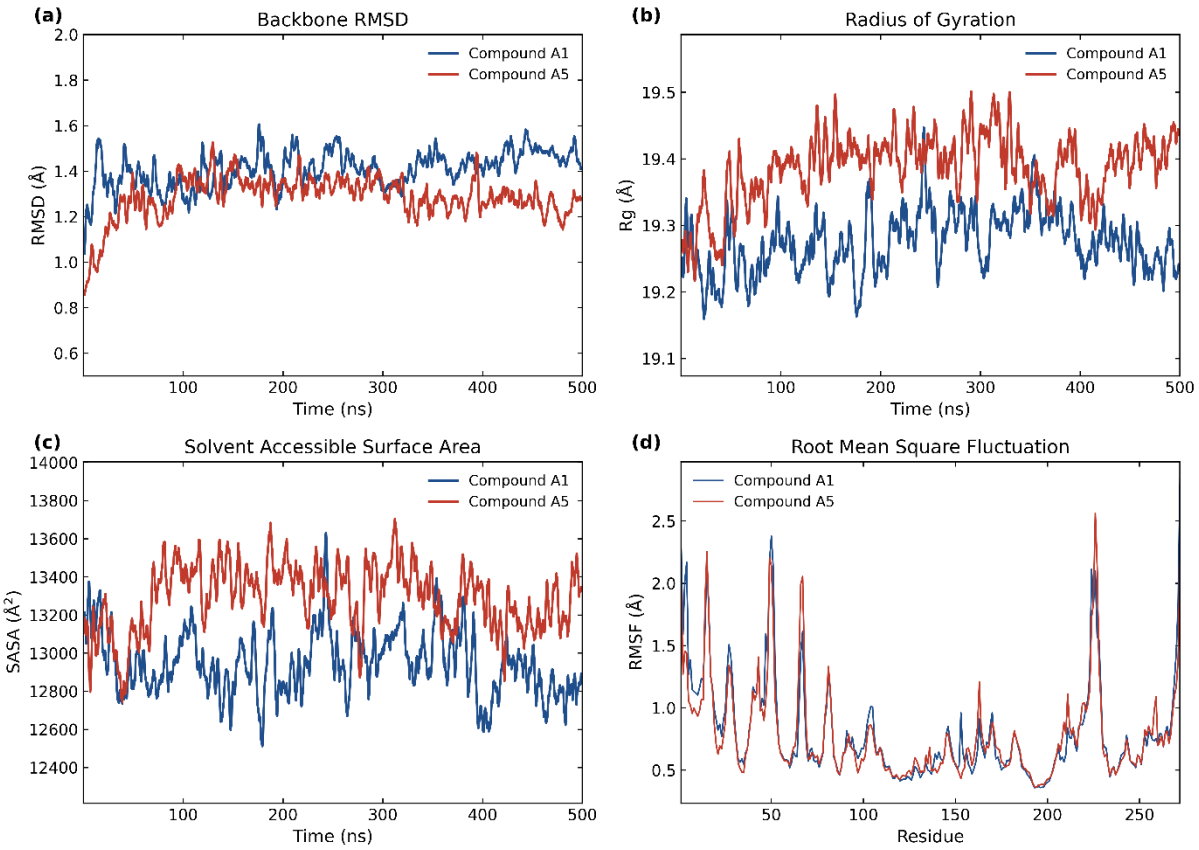



**Figure 10.** MD simulation analysis of PIM-1 in complex with Compound A1 and A5: (a) Backbone RMSD, (b) Radius of Gyration, (c) SASA, (d) RMSF per residue over 500 ns.

**Protein-Ligand Interaction and Hydrogen Bond Analysis**

The number and persistence of protein–ligand hydrogen bonds were monitored during the MD simulations (Figure 11). The trajectories showed an average of approximately 8–20 hydrogen bonds per complex, with transient peaks of up to 25–30 according to the analysis used. A1 showed more frequent high-count peaks during approximately 300–350 ns and 450–500 ns, whereas A5 displayed a comparatively more consistent interaction pattern.

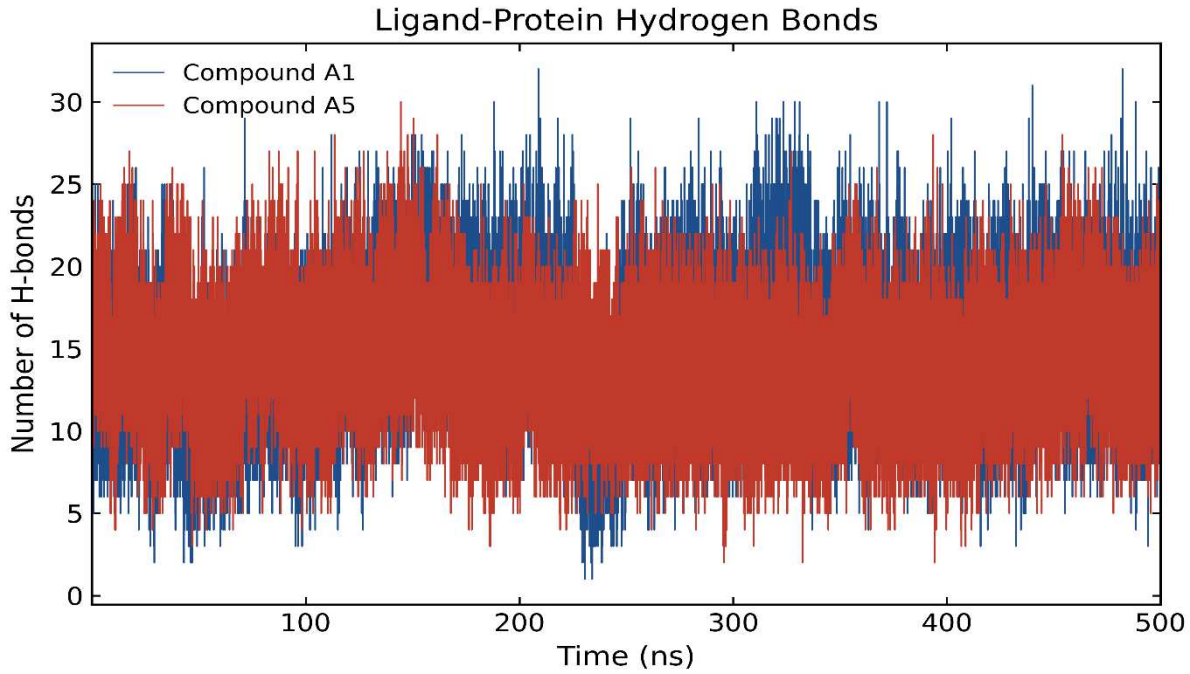



**Figure 11.** Number of hydrogen bonds formed between PIM-1 and Compound A1/A5 over 500 ns.

Taken together, the RMSD, Rg, SASA, RMSF and protein–ligand interaction analyses indicate that both A1 and A5 maintained stable complexes with PIM-1 over the simulated period. A5 showed somewhat more stable overall conformational behaviour than A1, consistent with its more favourable docking score. These findings support A5 as the leading computational docking/MD candidate, while experimental PIM-1 inhibition and binding studies are required to establish its actual affinity and mechanism.

**Pharmacokinetics and ADMET Properties**

The predicted pharmacokinetic and ADMET properties of the five designed analogues are presented in Tables 9 and 10. All compounds satisfied Lipinski's rule of five with zero predicted violations and had a predicted bioavailability score of 0.55. Predicted human intestinal absorption ranged from 72.7 to 80.1%. Synthetic-accessibility scores were below 5 for all analogues. Predicted BBB penetration was low (logBB < −0.9) and predicted CNS permeability was low (logPS < −2.7), which may be desirable for a peripherally acting oncology candidate. However, all five compounds were predicted to inhibit CYP2C9 and CYP3A4, representing an important potential metabolic-drug-interaction liability that requires experimental investigation. A2 additionally produced a toxicity alert and was therefore deprioritised. A1, A3, A4 and A5 showed no predicted toxicity alert under the selected pkCSM endpoints. These

results indicate favourable predicted drug-likeness for several candidates but should not be interpreted as evidence of experimental safety or oral bioavailability.

**Table 9.** Predicted pharmacokinetic properties of designed 3-cyanopyridine analogs

| Compound | Lipinski Violations | Bioavailability Score | HIA (%) | logBB | logPS | Synthetic Accessibility |
|---|---|---|---|---|---|---|
| A1 | 0 | 0.55 | 80.13 | -0.95 | -2.71 | 4.15 |
| A2 | 0 | 0.55 | 72.70 | -1.01 | -2.93 | 4.10 |
| A3 | 0 | 0.55 | 73.83 | -1.13 | -2.99 | 4.07 |
| A4 | 0 | 0.55 | 77.26 | -0.97 | -2.87 | 4.32 |
| A5 | 0 | 0.55 | 78.32 | -1.13 | -2.99 | 4.19 |

**Table 10.** Predicted ADMET and toxicity properties of designed compounds

| Compound | CYP2C9 Inhibitor | CYP3A4 Inhibitor | Total Clearance (log ml/min/kg) | Toxicity Alerts |
|---|---|---|---|---|
| A1 | Yes | Yes | 0.565 | None |
| A2 | Yes | Yes | 0.403 | Potential Alert |
| A3 | Yes | Yes | 0.45 | None |
| A4 | Yes | Yes | 0.651 | None |
| A5 | Yes | Yes | 0.545 | None |

## Conclusion

In this study, a QSAR model was developed for 24 3-cyanopyridine derivatives with reported antiproliferative activity against the MCF-7 breast cancer cell line. The final model incorporated AATS0s, GATS5i and CrippenLogP and showed strong internal and external predictive statistics ($R^2$ = 0.96165, $Q^2$ = 0.934224 and $R^2pred$ = 0.749). Five new analogues (A1–A5) were designed using the model, with predicted $pIC_{50}$ values ranging from 6.045 to 7.030. Molecular docking against PIM-1 identified A5 as the compound with the most favourable predicted docking score (−6.924 kcal $mol^{-1}$), while A1 also showed a favourable combination of predicted cellular activity and docking score. 500-ns molecular dynamics simulations indicated stable conformational behaviour for both A1 and A5, with A5 showing somewhat greater overall stability under the simulated conditions. Predicted pharmacokinetic and ADMET profiles further supported A1 and A5 for experimental follow-up, although the predicted CYP2C9/CYP3A4 inhibition and the limitations inherent to in-silico toxicity prediction warrant further investigation. The Y-scrambling test produced mean $R^2$ and $Q^2$

values of 0.2336 and −0.3655 for the scrambled models—far below the original model's $R^2$ = $Q^2$ = 0.96165—along with a $cR^2p$ of 0.8874 (> 0.5). These outcomes demonstrate that the QSAR model is resistant to chance correlation and satisfies all required validation criteria. Overall, A1 and A5 are proposed as computationally prioritized candidates for experimental synthesis, PIM-1 inhibition testing and MCF-7 antiproliferative evaluation.

### Author Contributions

UB: conceptualization, methodology, data curation, formal analysis, and writing—original draft.

SHA: methodology, computational analysis, visualization, and writing—review and editing.

MTI, ZYI, and RAdC: validation, investigation, and review of this manuscript.

AU and SU: supervision, project administration, and critical revision of this manuscript. All authors have read and agreed to the published version of this manuscript.

### Funding

This research received no external grant from any funding agency in the public, commercial, or not-for-profit sectors.

### Data Availability Statement

All data generated or analysed during this study are available within the paper. The datasets and online databases used in the study are publicly available from their cited references.

## Declarations

### Ethics approval and consent to participate

Not applicable. This is a purely computational study, and no human participants, human data, or live vertebrate animals were involved. This study did not involve human participants.

### Consent for publication

Not applicable. This study did not involve any individual participant data requiring consent for publication

### Competing interests

The authors declare no competing interests.

### Acknowledgments

The authors extend their appreciation to the physical chemistry research group under the supervision of Prof. Adamu Uzairu for technical assistance, computational facilities, supervision and motivation.

## References

1. Sung H, et al. Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. CA Cancer J Clin. 2024;76:e70090.
2. World Health Organization. Breast cancer [Internet]. 2024 [cited 2026 Sep 23]. Available from: https://www.who.int/news-room/fact-sheets/detail/breast-cancer.
3. Choudhury R, et al. PIM1 kinase and its diverse substrate in solid tumors. Cell Commun Signal. 2024;22:529.
4. Chen S, et al. Targeting PIM kinases in cancer therapy: an update on pharmacological small-molecule inhibitors. Eur J Med Chem. 2024;264:116016.
5. Rout AK, et al. A review on structure-function mechanism and signaling pathway of serine/threonine protein PIM kinases as a therapeutic target. Int J Biol Macromol. 2024;270:132030.
6. Chen J, Tang G. PIM-1 kinase: a potential biomarker of triple-negative breast cancer. Onco Targets Ther. 2019;12:6267-6273.
7. Mahata S, et al. PIM1/STAT3 axis: a potential co-targeted therapeutic approach in triple-negative breast cancer. Med Oncol. 2022;39:74.
8. Horiuchi D, et al. PIM kinase inhibition presents a novel targeted therapy against triple-negative breast tumors with elevated MYC expression. Nat Med. 2016;22:1321-1329.
9. Hussein BRM, et al. Design, synthesis, and anti-breast cancer activity evaluation of novel 3-cyanopyridine derivatives as PIM-1 inhibitors. Mol Divers. 2025;29:2565-2584.
10. Tropsha A, et al. Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR. Nat Rev Drug Discov. 2024;23:141-155.
11. Kataria A, et al. Systematic computational strategies for identifying protein targets and lead discovery. RSC Med Chem. 2024;15:2254-2269.
12. Naithani U, Guleria V. Integrative computational approaches for discovery and evaluation of lead compound for drug design. Front Drug Discov. 2024;4:1362456.
13. Wu Z, et al. Current perspectives and trend of computer-aided drug design: a review and bibliometric analysis. Int J Surg. 2024;110:3848-3878.
14. Niazi SK, Mariam Z. Recent advances in machine-learning-based chemoinformatics: a comprehensive review. Int J Mol Sci. 2023;24:11488.
15. Wei H, McCammon JA. Structure and dynamics in drug discovery. npj Drug Discov. 2024;1:1. doi:10.1038/s44386-024-00001-2.
16. AlRawashdeh S, Barakat KH. Applications of molecular dynamics simulations in drug discovery. Methods Mol Biol. 2024;2714:127-141.
17. Abdullahi SH, Uzairu A, Ibrahim MT, Umar BA. Chemo-informatics activity prediction, ligand based drug design, Molecular docking and pharmacokinetics studies of some series of 4, 6-diaryl-2-pyrimidinamine derivatives as anti-cancer agents. Bull Natl Res Cent. 2021;45:167. doi:10.1186/s42269-021-00631-w.
18. Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem. 2011;32(7):1466-1474.

19. Abdullahi SH, Uzairu A, Shallangwa GA, Uba S, Umar AB. In-silico design of some pyrazolopyridine analogs as breast cancer inhibitors: QSAR modeling, molecular docking, MD simulation, and pharmacological features profiling. J Indian Chem Soc. 2023;100:101053.
20. Adeniji SE, Arthur DE, Abdullahi M, Haruna A. Quantitative structure-activity relationship model, molecular docking simulation and computational design of some novel compounds against DNA gyrase receptor. Chem Afr. 2020;3:391-408. doi:10.1007/s42250-020-00132-9.
21. Abdullahi SH, Uzairu A, Shallangwa GA, Uba S, Umar AB. In-silico activity prediction, structure-based drug design, molecular docking and pharmacokinetic studies of selected quinazoline derivatives for their antiproliferative activity against triple negative breast cancer (MDA-MB231) cell line. Bull Natl Res Cent. 2022;46:2. doi:10.1186/s42269-021-00690-z.
22. Gramatica P, Giani E, Papa E. Statistical external validation and consensus modeling: a QSPR case study for Koc prediction. J Mol Graph Model. 2007;25(6):755-766. doi:10.1016/j.jmgm.2006.06.005.
23. Fatima A, Choudhary MI, Siddiqui S, Zafar H, Hu K, Wahab A-t. Insights into the molecular interactions between urease subunit gamma from MRSA and drugs: an integrative approach by STD-NMR and molecular docking studies. RSC Adv. 2024;14(42):30859-30872. doi:10.1039/D4RA01732C.
24. Jacobs MD, Black J, Futer O, Swenson L, Hare B, Fleming M, et al. Pim-1 ligand-bound structures reveal the mechanism of serine/threonine kinase inhibition by LY294002. J Biol Chem. 2005;280(14):13728-13734. doi:10.1074/jbc.M413155200.
25. Bullock AN, et al. Structure and substrate specificity of the Pim-1 kinase. J Biol Chem. 2005;280(50):41675-41682.
26. Sastry GM, et al. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013;27(3):221-234.
27. El-Zahabi MA, et al. Design, synthesis, molecular docking, and molecular dynamic studies of novel quinazoline derivatives as phosphodiesterase 7 inhibitors. Front Pharmacol. 2023;14:1151491.
28. Friesner RA, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. 2004;47(7):1739-1749.
29. Friesner RA, et al. Extra Precision Glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem. 2006;49(21):6177-6196.
30. Alharbi MA, Almagati AA, Gineinah MM, El-Malah AA, Hawsawi MB, Mantargi MJS, et al. New quinazoline derivatives with anti-inflammatory activity and computationally predicted PDE7A binding: synthesis, molecular modeling, and in silico pharmacokinetic evaluation. Front Pharmacol. 2026;17:1859798. doi:10.3389/fphar.2026.1859798.
31. Anza M, Endale M, Cardona L, Cortes D, Eswaramoorthy R, Zueco J, et al. Antimicrobial activity, in silico molecular docking, ADMET and DFT analysis of

secondary metabolites from roots of three Ethiopian medicinal plants. Adv Appl Bioinform Chem. 2021;14:117-132. doi:10.2147/AABC.S323657.
32. Halimi M, Bararpour P. Natural inhibitors of SARS-CoV-2 main protease: structure based pharmacophore modeling, molecular docking and molecular dynamic simulation studies. J Mol Model. 2022;28(9):279. doi:10.1007/s00894-022-05286-6.
33. El-Mernissi R, Alaqarbeh M, Khaldan A, et al. 3D-QSAR, molecular docking, ADMET, simulation dynamic, and retrosynthesis studies on new styrylquinolines derivatives against breast cancer. Open Chem. 2024;22:20240041. doi:10.1515/chem-2024-0041.
34. Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE 3rd, Cruzeiro VWD, et al. AMBER 2018. San Francisco (CA): University of California; 2018.
35. Abdullahi SH, Uzairu A, Shallangwa GA, Uba S, Umar AB. Structure based design of some novel 3-methylquinoxaline derivatives through molecular docking and pharmacokinetics studies as novel VEGFR-2 inhibitors. Chem Afr. 2022;5(6):1967-1978. doi:10.1007/s42250-022-00485-3.
36. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. Neuron. 2018;99:1129-1143.
37. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci. 2010;35:539-546.