Biointerface Research in Applied Chemistry
Open-Access Journal (ISSN: 2069-5837)

Article
Volume 16, Issue 4, 2026, 114
https://doi.org/10.33263/BRIAC164.114

In Silico Optimization of Dihydropteridine Derivatives Targeting PLK1 for Glioblastoma: An Integrated QSAR, Docking, and Molecular Dynamics Study

Youssef Briach [1,*], Mohammed Er-rajy [1], Jamal Elkhabchi [1], Salima Hdiddou [1], Menana Elhallaoui [1]

[1] LIMAS Laboratory, Faculty of Sciences, Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco
* Correspondence: youssef.briach1@usmba.ac.ma;

Received: 21.01.2026; Accepted: 20.05.2026; Published: 15.08.2026

Abstract: A 2D quantitative structure–activity relationship (QSAR) model was developed for a series of 34 dihydropteridine derivatives to predict their antitumor activity against glioblastoma. The multiple linear regression model (MLR), based on four descriptors (ATS7s, AATS8e, AATS3p, and AATS5p), demonstrated satisfactory statistical performance (R^2 = 0.714, R^2_adj = 0.660, Q^2 = 0.513, RMSE = 0.096, F = 13.134, p < 0.0001), with strong external validation (R^2_test = 0645). Model-guided optimization led to the design and screening of new structural analogs, which were subsequently docked against Polo-like kinase 1 (PLK1, PDB ID: 3BD6), a key mitotic regulator overexpressed in glioblastoma. Among the compounds evaluated, X14 exhibited the most favorable docking affinity (-10.5 kcal / mol), compared to -6.6 kcal/mol for Temozolomide (TMZ) and 8.1 kcal/mol for the reference compound N27. Although generally favorable ADMET profiles were observed, hepatotoxicity alerts were predicted for all compounds, which represents an important limitation supporting the prioritization of X14. In general, this study provides. Molecular dynamics simulations over 100 ns supported stable complex formation, with RMSD backbone values stabilizing around 2.5 to 3.0 Å. The ligands remained bound within the binding pocket throughout the simulation, exhibiting RMSD values below 1.5 Å for X14 and temozolomide and below 2.5 Å for N27, while the PLK1–TMZ complex showed higher structural fluctuations. A plausible synthetic route was proposed to assess the experimental feasibility of X14. Overall, these computational findings support the prioritization of X14 for further experimental validation in glioblastoma therapy.

Keywords: QSAR modeling; dihydropteridine derivatives; glioblastoma; molecular docking; PLK1 inhibition; molecular dynamics.

1. Introduction

Glioblastomas are the most common and aggressive primary malignant brain tumors in adults [1]. Recent whole-genome sequencing studies have provided important insights into the genetic susceptibility factors associated with adult glioblastoma [2]. Due to its highly infiltrative nature, complete surgical resection rarely results in a curative outcome. GBM is known to show significant inter- and intratumoral heterogeneity from a biological perspective. This characteristic makes standard treatments less effective, leading to disappointing clinical results [3].

The current standard of care consists of maximal safe surgical resection followed by radiotherapy combined with temozolomide (TMZ) [4]. The incorporation of bevacizumab resulted in tumor stabilization without a substantial improvement in overall survival [5]. Similarly, the combination of everolimus with standard therapy resulted in increased toxicity without significant clinical benefit [6].

Despite the progress made in biomedical research, the outlook for patients with GBM remains bleak. The median survival time for these patients is only 15 months, and the 5-year survival rate is less than 10%, even among those participating in clinical trials [7]. In addition, TMZ is clinically effective but is associated with serious adverse effects, including leukopenia and thrombocytopenia [8]. Furthermore, overexpression of O6-methylguanine-DNA methyltransferase (MGMT) reduces the efficacy of TMZ by repairing alkylated DNA lesions, thereby attenuating its cytotoxic effect [9].

Identifying and developing new molecules that specifically target pathways involved in GBM progression could provide a more effective, less toxic alternative. While cell proliferation is essential for tissue regeneration, it becomes uncontrolled in cancer, promoting tumor growth. Polo-like kinase 1 (PLK1) is a key serine/threonine kinase that regulates mitotic progression, centrosome maturation, spindle assembly, and cytokinesis, and is frequently overexpressed in high-grade gliomas [10].

Structurally, the PLK1 crystal structure (PDB ID: 3BD6) was selected for this study due to its high resolution and the presence of a cocrystallized ATP-competitive inhibitor. This feature enables reliable validation of the docking protocol through redocking procedures and ensures accurate characterization of the ATP-binding pocket. Compared to other available PLK1 crystal structures, including apo forms and lower-resolution complexes, 3BD6 provides a structurally well-resolved ATP-binding pocket with preserved hinge-region interactions, making it particularly suitable for structure-based inhibitor evaluation. The 3BD6 structure presents a crystallographic resolution of 2.0 Å and contains a co-crystallized ATP-competitive inhibitor bound to the catalytic domain. The presence of conserved hydrogen bonds in the hinge region and a well-defined activation loop provides a reliable structural environment for docking validation via redocking. In contrast to apo structures or complexes with an incomplete definition of the binding site, 3BD6 offers a geometrically stable and pharmacologically relevant conformation of the ATP-binding pocket [11].

Recent studies have reported that dihydropteridine derivatives containing an oxadiazole moiety are promising ATP-competitive PLK1 inhibitors [12]. These compounds were designed to interfere with the signaling pathways involved in tumor proliferation and survival.

Furthermore, previous QSAR studies on related scaffolds often relied on limited validation strategies, small datasets, or lacked a mechanistic interpretation of descriptor contributions in relation to PLK1 structural characteristics. Therefore, a rigorously validated and mechanistically interpretable computational framework remains necessary to enhance predictive robustness and rational candidate prioritization [13].

Although several studies have reported the synthesis and preliminary biological evaluation of dihydropteridine derivatives, prior computational investigations have generally lacked rigorous statistical validation (eg, extensive Y-randomization and applicability domain analysis) and rarely integrated QSAR modeling with docking, ADMET prediction, and molecular dynamics simulations within a unified validation framework [14].

In this context, the development of a rigorously validated QSAR model in accordance with OECD principles is essential to ensure statistical reliability, predictive accuracy, and mechanistic interpretability [15].

In particular, compliance with OECD validation criteria - including defined endpoint, unambiguous algorithm, appropriate goodness of fit and robustness measures, applicability domain, and mechanistic interpretation - ensures transparency and reproducibility of predictive modeling [15].

The aim of our research is to use computer-based methods to create structure-activity relationships and to identify molecular architecture characteristics that underpin the efficacy of PLK1 inhibitors. A 2D quantitative structure-activity relationship (QSAR) analysis was performed on a set of 34 distinct molecules.

In the context of glioblastoma research, molecular dynamics (MD) simulations have increasingly been employed to assess the stability of interactions between potential inhibitors and key mitotic regulators, such as PLK1. MD simulations allow dynamic evaluation of ligand–protein stability, conformational flexibility, and interaction persistence within the ATP-binding pocket. Stable root mean square deviation (RMSD) values and favorable binding free energy calculations provide important information on the structural robustness and potential efficacy of candidate inhibitors [16].

Although computational predictions do not directly translate into clinical efficacy, this integrated in silico framework strengthens the prioritization and supports subsequent experimental validation.

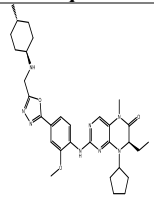
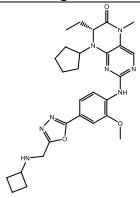
Therefore, the present study aims to develop a rigorously validated 2D-QSAR model for dihydropteridine derivatives targeting PLK1 and to prioritize promising candidates through molecular docking, ADMET evaluation, and molecular dynamics simulations for potential glioblastoma therapy.

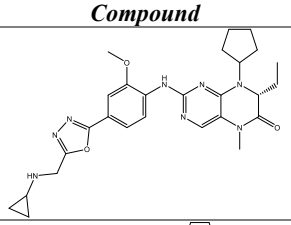
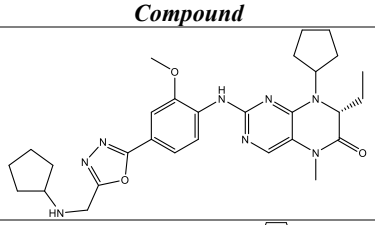
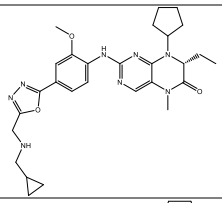
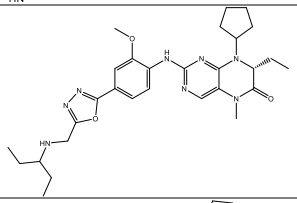
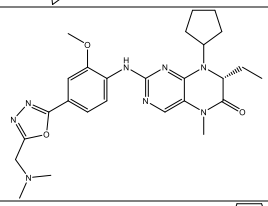
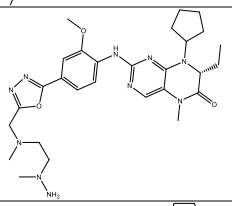
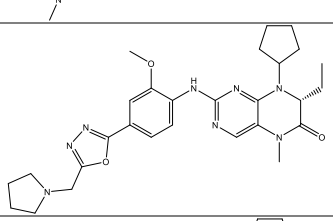
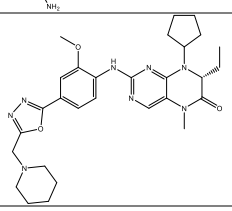
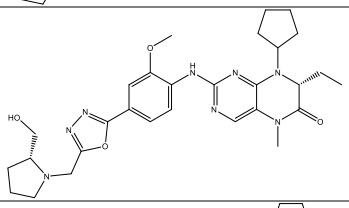
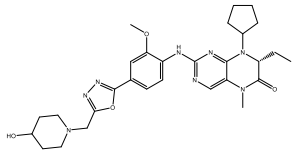
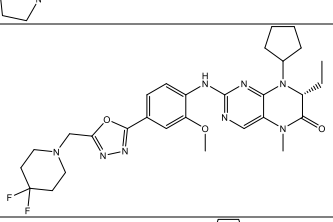
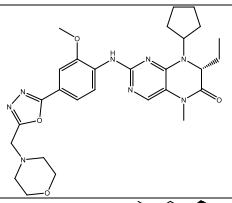
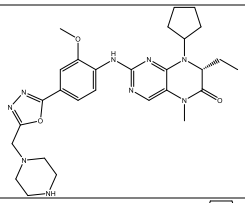
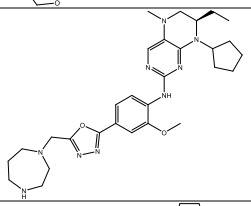
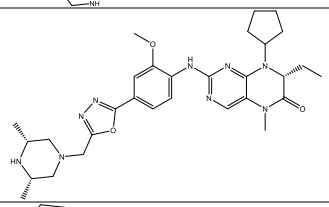
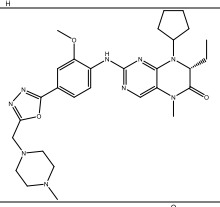
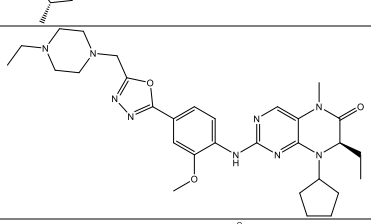
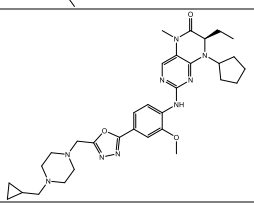
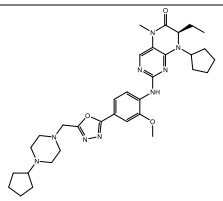
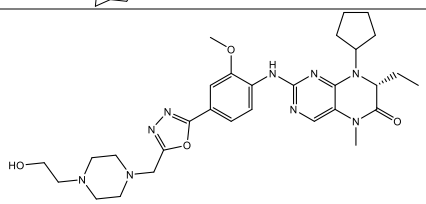
2. Materials and Methods

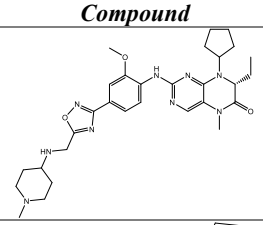
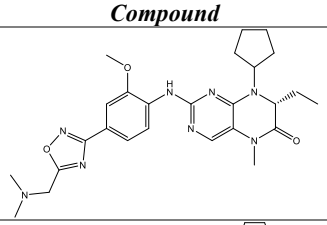
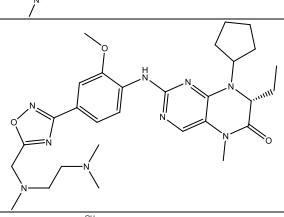
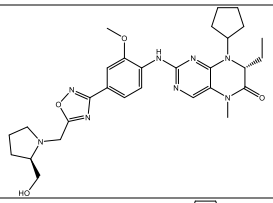
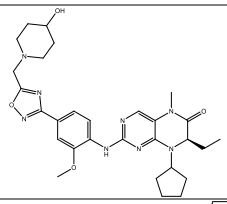
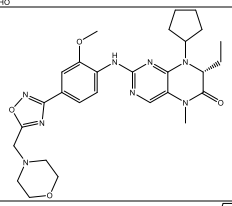
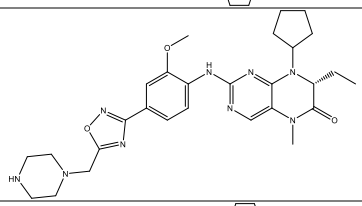
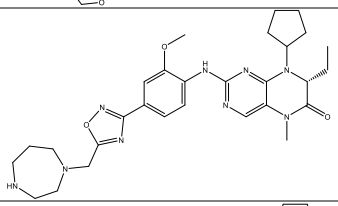
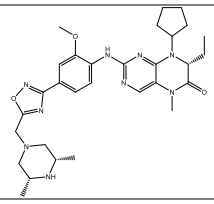
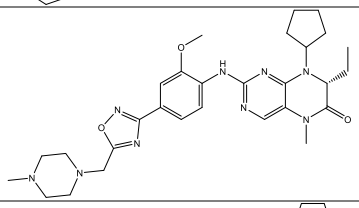
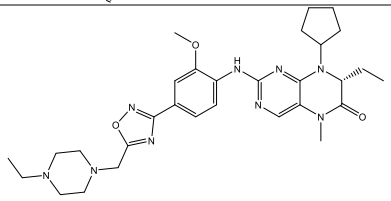
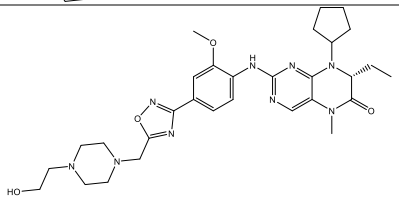
2.1. Experimental dataset.

In this study, a series of 34 dihydropteridone derivatives, considered potential therapeutic candidates for their inhibitory activity against glioblastoma (GBM) [12], was investigated. The main objective was to establish a quantitative structure-activity relationship (QSAR) to examine the relationship between chemical structure and inhibitory efficacy in GBM. The molecular structures of the derivatives studied are presented in Table 1, along with their values of experimental biological activity (IC_{50}). To standardize and optimize the use of the data, the minimum inhibitory concentration values in molar units (M) (IC_{50}) were converted to their logarithmic form (pIC_{50}), which are presented in the table below.

Table 1. Chemical structures and their biological activity for 34 derivatives.

No.	*Compound*	*pIC_{50}*	*No.*	*Compound*	*pIC_{50}*
1		6.52	*2*		6.28

No.	Compound	pIC_{50}	No.	Compound	pIC_{50}
3		6.72	4		6.40
5		6.41	6		6.16
7		6.64	8		6.19
9		6.22	10		6.21
11		6.64	12		6.26
13		5.97	14		6.38
15		6.48	16		6.38
17		6.30	18		6.10
19		6.20	20		6.28
21		6.47	22		6.36

No.	*Compound*	*pIC50*	*No.*	*Compound*	*pIC50*
23		*6.21*	*24*		*5.99*
25		*6.28*	*26*		*6.30*
27		*6.74*	*28*		*6.12*
29		*6.35*	*30*		*6.49*
31		*6.43*	*32*		*6.04*
33		*6.26*	*34*		*6.59*

pIC50 value was determined against the glioblastoma cell line (GBM).

2.2. Calculation of the descriptors.

In this section, 144 molecular descriptors were calculated using PaDEL software for each compound in the series studied [17]. The complete list of descriptors calculated for all molecules is provided in the supplementary material (Table S1). These descriptors cover a wide range of geometric, physicochemical, and electronic properties and are divided into several distinct classes (Table 2).

Table 2. Description of molecular descriptors calculated by PaDEL.

No.	Descriptors	Types	Main role/interpretation
1	Gibbs Free Energy	Quantum / Thermodynamic	Indicates molecular stability and spontaneity of formation.
2	Mol Refractivity	Constitutional / Physicochemical	Molar refractivity: an indicator of molecular polarizability
3	LogP	Physicochemical	Hydrophobicity index; octanol-water partition coefficient
4	LogS	Physicochemical	Aqueous solubility estimate
5	Num Rotatable Bonds	Topological	Flexibility of the molecule
6	Polar Surface Area	Topological / Geometrical	Sum of surfaces of polar atoms; related to membrane permeability

No.	Descriptors	Types	Main role/interpretation
7	ALogP	Physicochemical	LogP estimation using the Ghose-Crippen method
8	ATS3i	Autocorrelation (charge)	3D distribution of ionization potential
9	AATS7e	Average autocorrelation (electronegativity)	Mean electronegativity at lag 7
10	GATS4v	Geometrical / Autocorrelation	Geometrical autocorrelation with atomic volume at lag 4
11	BCUTp-1h	BCUT/Polarizability	The highest polarizability-weighted BCUT descriptor
12	MATS	Autocorrelation	Autocorrelation based on atom properties
13	VR_DzZ SM_Dzi VE_Dzi SM_DzZ	WHIM/Edge/Vertex indices	3D geometrical or edge/vertex descriptor
14	SpMax_Bhm SpMin_Bhm	Topological Indices	Eigenvalues from Burden matrix

2.3. Principal component analysis (PCA).

The construction of the QSAR model required a statistical pre-processing phase. This phase had two objectives: reducing data dimensionality and avoiding overfitting. The aim was to eliminate irrelevant or highly redundant descriptors that could compromise the model's stability and interpretability.

To this end, the XLStat software, version 2016, was used to apply PCA [18]. This multivariate method transforms an initial set of correlated variables into a reduced number of uncorrelated principal components while retaining the essential information contained in the data. The PCA was then used to evaluate the correlation structure between descriptors and select a relevant subset for modeling.

The selected descriptors include physicochemical descriptors (such as polarity, lipophilicity, refractivity, etc.) as well as topological and autocorrelative descriptors (of types AATS, ATS, and ATSC). These descriptors reflect atomic relationships weighted by various properties across the molecule. These variables were shown to be the most informative in predicting biological activity (pIC_{50}) while ensuring the optimal representativeness of structural diversity.

The correlation matrix derived from the PCA analysis played a central role in this process, allowing the elimination of variables that were highly interdependent and those whose contributions to total variance did not justify their retention in the final model.

2.4. Data set partitioning.

The data set (n = 34) was randomly divided into a training set (n = 26) and an external test set (n = 8), corresponding to 76% and 24% of the total dataset, respectively. Random selection was performed using XLSTAT to ensure reproducibility of the partitioning process. The distribution of pIC_{50} values was examined after partitioning, and the activity values of the test set were confirmed to fall within the range of the training set, thus preventing extrapolation beyond the domain of applicability of the model [14,15].

2.5. Model construction.

The construction of the QSAR model was performed using the multiple linear regression (MLR) method. This approach is frequently used in medicinal chemistry modeling studies because of its simplicity, statistical robustness, and ease of interpreting the coefficients

obtained. The method is based on the assumption that the biological response (pIC_{50}) is a linear combination of selected molecular descriptors.

RLM allows identifying significant descriptors with a statistically validated influence on biological activity, while minimizing the discrepancy between experimental values and those predicted by the model. Furthermore, this approach is often used as a preliminary step for more complex modeling techniques as a variable selection tool [19].

The final model was generated using XlSTAT software, which offers a powerful interface for multiple regression and a suite of statistical evaluation tools, including analysis of variance (ANOVA) and Fisher's test. It also calculates the coefficients of determination, R^2, and Q^2.

2.5. Validation of the model.

2.5.1. Internal model validation.

The robustness and reliability of a QSAR model can be assessed through internal validation, which is an essential step. Performance within the training set forms the basis for statistical analysis, and indicators such as the adjusted coefficient of determination (R^2_adj), cross-validation (Q^2), root mean square error (RMSE), and randomization tests (Y-scrambling) are used. Cross-validation using leave-one-out (LOO) or leave-many-out (LMO) can be used with these methods to estimate the model's internal predictive power while highlighting the risk of overfitting. In general, a model is considered robust if $Q^2 > 0.5$ and $R^2 > 0.6$. Therefore, internal validation is a necessary but not sufficient condition and must be supplemented by rigorous external validation [18,19].

2.5.2. External validation.

Golbraikh and Tropsha (2002) proposed a set of mathematical criteria to rigorously assess the external predictive quality of a QSAR model [14]. These criteria confirm that the relationship between the experimental and predicted values is significant, symmetric, linear, and reliable.

Table 3. Summary of the Golbraikh and Tropsha validation parameters for QSAR models.

Parameter	Mathematical relation	Recommended threshold	Interpretation
Predictive squared correlation coefficient	R^2_pred	> 0.6	Indicates good predictive capacity on the external test set
Correlation deviation (direct)	$(r^2 - r_0^2)/r^2$	< 0.1	Measures how much the regression without an intercept deviates from the actual correlation
Correlation deviation (inverse)	$(r^2 - r'_0{}^2)/r^2$	< 0.1	The same as above, but for inverse regression
Slope of direct regression	K	$0.85 < k < 1.15$	The slope of Y vs. without intercept should be close to 1
Slope of inverse regression	k'	$0.85 < k' < 1.15$	Slope of $\hat{Y}$ vs Y without intercept; should be close to 1
Concordance Correlation Coefficient	CCC	> 0.85	Evaluates both the accuracy and precision of the model predictions
Average r^2_m	r^2_m (mean)	> 0.5	Indicates balanced predictive power and correlation symmetry
Root Mean Squared Error	RMSE_test	Low value desirable	Measures the magnitude of the average prediction error magnitude
Mean Absolute Error	MAE_test	< 0.15	Measures the average of absolute prediction errors
Y-randomization check	R^2_rand, Q^2_rand	Near 0	Ensures that the model is not a result of chance correlation

The criteria apply exclusively to molecules in the test set (external validation) and are presented in Table 3.

These criteria confirm that the model predictions are correlated with experimental values and free of regression bias and imbalance between direct and inverse linear trends. The incorporation of slopes k and k′ is particularly noteworthy, as it prevents the model from systematically underestimating or overestimating predictive values. As already shown, a model with a high R^2_(pred) can still be inaccurate if the other conditions are not met [20]. Therefore, evaluating a QSAR model using the criteria of Golbraikh and Tropsha is an essential step in determining its relevance in a regulatory or application context.

2.5.3. Model robustness and Y-randomization test to evaluate the robustness of the

developed a QSAR model and excluded the possibility of chance correlations; a Y-randomization test was performed. In this procedure, the response variable (pIC50) was randomly shuffled while keeping the descriptor matrix unchanged, and new models were generated using the same regression procedure. This process was repeated multiple times, and the resulting R^2 and Q^2 values were compared with those of the original model. Significantly lower R^2 and negative Q^2 values for the randomized models indicate that the proposed QSAR model is not the result of chance correlation [14].

2.5.4. Applicability domain.

Assessing the applicability domain (AD) is essential to ensure the reliability of QSAR predictions. According to the recommendations of the OECD, the chemical space in which the model can make valid predictions [15]. William's diagram is the benchmark approach among the available methods. The William diagram represents standardized residues based on leverage values (hi), thus identifying influential or poorly predicted molecules. AD is considered to include a compound if the standardized residue is greater than ± 3 or the leverage is greater than the critical threshold:

$$h^* = \frac{(3(p+1))}{n} \qquad (1)$$

The number of descriptors is represented by p, and the number of compounds in the training set is represented by n. This methodology facilitates the identification of both outliers (exhibiting elevated residuals) and structurally atypical points (demonstrating elevated leverage). Some authors restrict this analysis to the test set, but a more rigorous approach extends it to all predicted molecules, including those generated in silico [21].

2.6. Molecular docking modeling.

Molecular docking is a fundamental tool in modern drug discovery [22]. It offers a computational approach to examine the interactions between small molecules and their macromolecular targets. In this study, molecular docking was performed online using the PCB-Dock platform, which automatically integrates ligand and protein preparation, active site prediction, and binding conformation optimization (http://www.caddlab.scbdd.com/pcb-dock). The crystallographic structure of PLK1 (PDB ID: 3BD6) was retrieved at 2.10 Å resolution. Before docking, all water molecules were removed using the Discovery Studio 2021 software.

Polar hydrogens were added and partial charges assigned using the Gasteiger–Hückel method [11].

To validate the docking protocol, a redocking procedure was performed using the native cocrystallized ligand (B6H) extracted from the PLK1 crystal structure (PDB ID: 3BD6). The ligand was removed from the binding pocket and subsequently redocked into the ATP-binding site using the same docking parameters as those applied for all the compounds studied. The docking grid box was centered on the original ligand binding site with coordinates center_x = 30.426, center_y = 0.38, and center_z = 52.846, and grid dimensions of 22 × 22 × 22 Å, ensuring full coverage of the active site region. The docking calculations generated multiple binding poses, and the best-ranked pose was selected for further analysis. The predicted pose was then superimposed onto the crystallographic conformation, and the root-mean-square deviation (RMSD) between the heavy atoms was calculated to assess the precision of the docking protocol [23].

2.7. Molecular dynamics simulation.

The preparation of the molecular system for molecular dynamics simulations was carried out in several steps [16]. First, during the import and processing phase, bond orders were assigned, and zero-order bonds were created for metal atoms, while disulfide bridges were automatically generated [24]. Water molecules located at a distance greater than 0.00 Å from heteroatomic groups were removed, and the appropriate ionization states were determined at pH 7.0 using Epik [25]. In the refinement stage, optimization was performed by adjusting the orientations of water molecules and minimizing hydrogen-containing species using the pKa prediction model (PRPKA) at pH 7.0. A minimization step was then applied, including the elimination of water molecules that form fewer than three hydrogen bonds with non-water atoms and the convergence of heavy atoms to an RMSD threshold of 0.3. Hydrogen atoms were added using the OPLS3e force field. The system was embedded in an orthorhombic simulation box with dimensions a, b, and c fixed at 10 Å to reduce total volume. The ions were recalculated and added to obtain a final NaCl concentration of 0.15 M [26]. The molecular dynamics simulation was configured for a total duration of 100 ns, with trajectory frames saved every 4.8 ps and energy values recorded every 1.2 ps, producing approximately 250 frames. The simulation was carried out in an NPT ensemble at 300 K and a pressure of 1.01325 bar. Before the production phase, a relaxation protocol was applied to stabilize the system and ensure proper equilibration [16].

3. Results and Discussion

3 .1. Principal component analysis (PCA).

To assess the redundancy or independence of the various descriptors, a correlation matrix was created. The complete results are available in the Supplementary Material (Table S2). This matrix shows the correlation coefficients between each pair of descriptors. Descriptors with a very high correlation ($R > 0.95$), suggesting that similar information about the property under study is transmitted by them, were excluded. Similarly, a low correlation ($R < 0.3$) was excluded from the analysis due to its low informative contribution.

The multicollinearity statistics of the selected descriptors are summarized in Table 4.

Table 4. Multicollinearity statistics.

	ATS7s	AATS8e	AATS3p	AATS5p
Tolerance	0.257	0.386	0.193	0.458
VIF	3.894	2.588	5.177	2.182

Multicollinearity (Table 4) among the descriptors included in the multiple linear regression model (MLR) was evaluated using the variance inflation factor (VIF). The values obtained for the four selected descriptors, namely ATS7s, AATS8e, AATS3p, and AATS5p, are 3.894, 2.257, 5.177, and 2.182, respectively. These results indicate low to moderate collinearity among the independent variables, which helps ensure good model stability. Furthermore, the tolerance values associated with these descriptors, with the exception of AATS3p, are overall above the critical threshold of 0.2, confirming the absence of significant collinearity within the model (Figure 1).

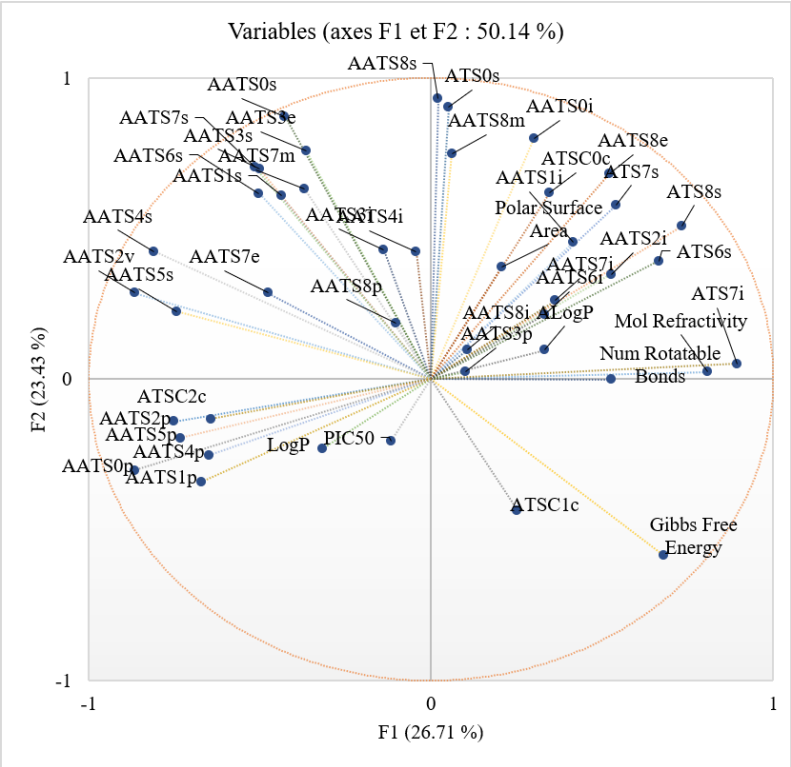


Figure 1. Correlation circle of model descriptors in QSAR analysis.

The principal PCA performed on the molecular descriptors used in the QSAR model is illustrated in Figure 2. This projection shows that the first two main components (F1 and F2) collectively explain 50.14% of the total variation in the data. The figure also reveals significant covariation among variables, including descriptors such as AATS3p, AATS5p, AATS7s, and AATS8e. These descriptors show a notable degree of correlation with the principal components. Their similarity is evident. They are also proximate and oriented similarly. This suggests a significant contribution to explaining biological activity. In contrast, descriptors such as Gibbs Free Energy and ATSC1c are oriented in the opposite direction and show a negative correlation with pIC_{50}. This suggests an inverse influence on the activity studied. Descriptors near the origin may be considered less informative and prioritized. This analysis validates the importance of the chosen descriptors and suggests ways to improve the QSAR model.

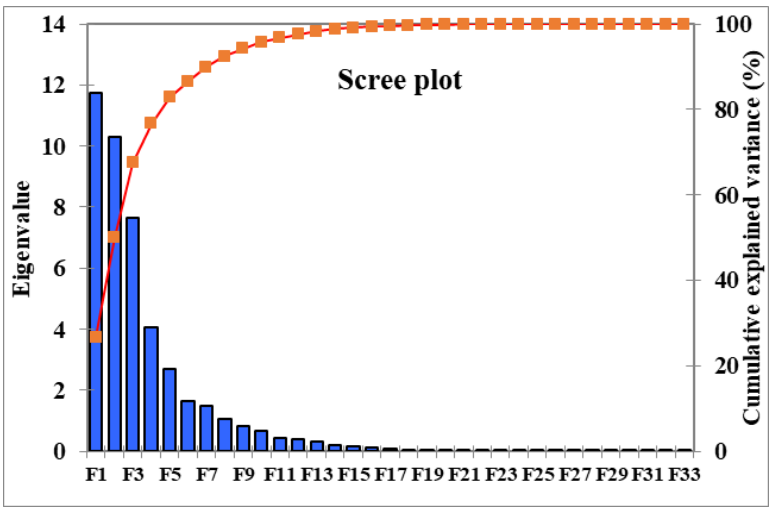


Figure 2. The percentage of variance of PCA.

The scree plot indicates that the first six principal components have eigenvalues greater than 1, collectively accounting for nearly 80% of the total variance. According to Kaiser's criterion [27], only components with eigenvalues greater than 1 should be retained for analysis. Thus, the retention of these components is justified. After the sixth element, the eigenvalues drop rapidly, suggesting a minimal impact on total information. The visible inflection in the curve also confirms the relevance of this threshold according to the elbow method.

3.2. Multiple linear regression (MLR).

Descriptor reduction and variable selection. A total of 144 molecular descriptors were initially calculated using the PaDEL-Descriptor. Descriptors with constant or near-constant values were first eliminated. Principal Component Analysis (PCA) was then performed as an exploratory tool to evaluate the intercorrelations of the descriptors and identify redundancy within the descriptor pool. Based on PCA loadings and correlation analysis, redundant descriptors that contributed similarly to the same principal components were removed, reducing the set from 144 to 43 representative original variables (Supplementary Table S2). Importantly, the principal components were not used as regression variables. The subsequent Multiple Linear Regression (MLR) model was constructed using the original descriptors. The selection was carried out using an iterative stepwise procedure, and the candidate models were evaluated based on statistical significance ($p < 0.05$), adjusted R^2, predictive performance, and multicollinearity diagnostics (VIF < 5). The final model included four descriptors, ensuring an adequate compound-to-descriptor ratio and statistical robustness.

The model obtained is as follows:

$$pIC_{50}=-10.600 - 3.723E\text{-}03*ATS7s + 4.252*AATS8e - 12.720*AATS3p + 2.682*AATS5p \quad (2)$$

By examining the descriptors included in the QSAR model, we can pinpoint the structural characteristics that contribute to biological activity. The distribution of atomic properties along the molecular graph is assessed using the ATS (Autocorrelation of Topological Structure) and AATS (Average-Centered Autocorrelation) descriptors. These descriptors provide a comprehensive numerical representation of structural and physicochemical properties relevant to bioactivity [28]. A high ATS7s value indicates a marked distribution of this 7-bond property. In our study, we observed a negative correlation between

ATS7s and activity, suggesting that a high density of this long-range property could disrupt the optimal alignment of the ligand with the biological target.

A positive correlation between AATS8e and biological activity suggests that long-range electrostatic interactions enhance molecular recognition.

AATS3p shows a negative correlation with activity, suggesting that local polarization can lead to unfavorable electronic contributions. However, AATS5p shows a positive correlation, indicating that moderate polarization at intermediate distances promotes stabilizing interactions between the ligand and the target. In summary, these descriptors reflect the critical influence of electronic properties (electronegativity and polarizability) on modulating biological activity and confirm the importance of long-range interactions in inhibitory mechanisms.

The statistical evaluation of the regression model showed good performance across both the training and validation sets, as shown in Table 5.

Table 5. Adjustment coefficients (pIC_{50}).

Statistics	Training sample
Number of compounds (n)	26
R^2	0.714
Adjusted R^2	0.660
MSE	0.009
RMSE	0.096
MAPE	1.132
DW	2.019
Cp	5.000
AIC	-117.245
SBC	-110.954
PC	0.422

The R^2 determination coefficient was 0.714, indicating that 71.4% of the variance in the dependent variable was explained by the model. This value is supported by an adjusted R^2 of 0.660, which confirms the relevance of the explanatory variables without overestimation due to an excessive number of descriptors.

Prediction errors are low, as evidenced by the root mean square error (MSE = 0.009), the root mean square error (RMSE = 0.096), and the mean absolute percentage error (MAPE = 1.132%), reflecting high model accuracy.

The Durbin–Watson statistic (DW = 2.019) is close to the ideal value of 2, indicating the absence of significant autocorrelation in the residuals. The model selection criteria also support the regression model's robustness, with a Mallows Cp value of 5.000, suggesting an appropriate balance between model complexity and predictive performance.

Furthermore, the Akaike Information Criterion (AIC = 117.245) and the Schwarz Bayesian Criterion (SBC = 110.954) have low values, confirming that the model achieves a good compromise between goodness of fit and parsimony. The Amemiya prediction criterion (PC = 0.422) also indicates an acceptable predictive stability of the regression model.

In general, these statistical indicators confirm that the MLR model is statistically reliable and provides a satisfactory description of the relationship between the selected descriptors and biological activity.

When evaluating the standardized coefficients of the multiple linear regression model, the relative importance of the different descriptors in modeling biological activity (pIC_{50}) is revealed. The descriptor AATS3p stands out with a strongly negative coefficient (-1.5561), indicating a significant unfavorable contribution to the increase in pIC_{50}. On the contrary,

AATS8e shows a significant positive influence (+0.6697), while AATS5p also exerts a positive but more moderate effect (+0.5044). The ATS7s descriptor shows an intermediate negative influence (-0.5360). These results indicate that electronic and topological descriptors are crucial for predicting activity, which justifies their priority in the development of the QSAR model (Figure 3).

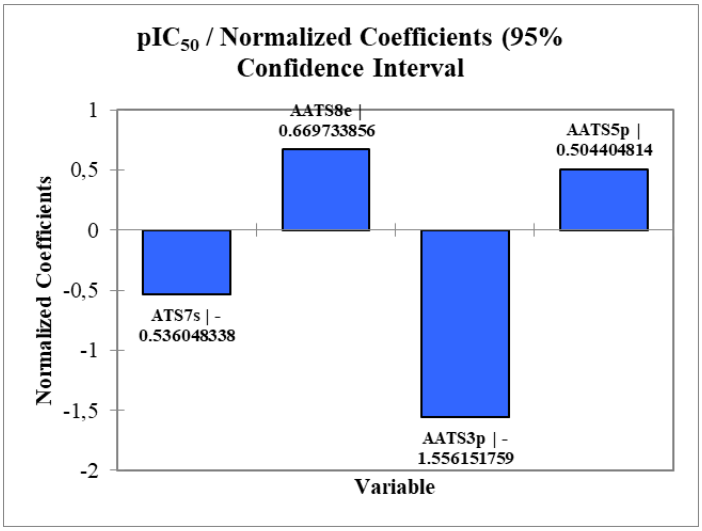


Figure 3. Box plot of standardized coefficients as a function of variables.

In Table 6, the experimental and predicted pIC_{50} values are summarized, along with the corresponding residuals, for the molecular series.

Table 6. The values of chemical descriptors and the activities observed and predicted using the MLR models for the test set.

Observation	ATS7s	AATS8e	AATS3p	AATS5p	pIC_{50}	Predicted (pIC_{50})	Residual	Standardized residual
N_1	536.083	7.544	1.296	1.226	6.376	6.286	0.090	0.934
N_2	563.861	7.555	1.297	1.262	6.301	6.313	-0.012	-0.127
N_3	531.25	7.589	1.321	1.215	6.102	6.148	-0.046	-0.476
N_4	565.083	7.537	1.27	1.212	6.468	6.441	0.027	0.276
N_5	547.833	7.601	1.296	1.194	6.356	6.399	-0.043	-0.444
N_6	560.333	7.567	1.292	1.216	6.207	6.318	-0.111	-1.148
N_7	481	7.549	1.354	1.308	5.991	5.995	-0.004	-0.037
N_8	541.333	7.564	1.295	1.223	6.301	6.356	-0.055	-0.573
N_9	552.722	7.601	1.289	1.246	6.744	6.609	0.135	1.400
N_{10}	513.667	7.563	1.313	1.251	6.119	6.301	-0.182	-1.891
N_{11}	515.5	7.558	1.299	1.255	6.346	6.462	-0.116	-1.202
N_{12}	535.5	7.561	1.299	1.224	6.494	6.317	0.177	1.839
N_{13}	535.583	7.576	1.306	1.189	6.259	6.197	0.062	0.639
N_{16}	514.25	7.546	1.311	1.253	6.376	6.257	0.119	1.231
N_{17}	516.083	7.54	1.297	1.257	6.481	6.414	0.067	0.697
N_{18}	536.083	7.544	1.296	1.226	6.376	6.286	0.090	0.934
N_{21}	563.861	7.555	1.297	1.262	6.301	6.313	-0.012	-0.127
N_{22}	531.25	7.589	1.321	1.215	6.102	6.148	-0.046	-0.476
N_{23}	536.167	7.56	1.304	1.191	6.2	6.158	0.042	0.435
N_{24}	559.222	7.563	1.291	1.235	6.275	6.368	-0.093	-0.970
N_{26}	565.083	7.537	1.27	1.212	6.468	6.441	0.027	0.276
N_{27}	547.833	7.601	1.296	1.194	6.356	6.399	-0.043	-0.444
N_{28}	560.333	7.567	1.292	1.216	6.207	6.318	-0.111	-1.148
N_{29}	481	7.549	1.354	1.308	5.991	5.995	-0.004	-0.037
N_{30}	564.5	7.652	1.334	1.234	6.275	6.178	0.097	1.012
N_{33}	541.333	7.564	1.295	1.223	6.301	6.356	-0.055	-0.573
N_{14*}	514.25	7.546	1.311	1.253	6.376	6.257	0.119	1.231
N_{15}*	516.083	7.54	1.297	1.257	6.481	6.414	0.067	0.697
N_{19}*	536.167	7.56	1.304	1.191	6.2	6.158	0.042	0.435
N_{20}*	559.222	7.563	1.291	1.235	6.275	6.368	-0.093	-0.970
N_{25}*	564.5	7.652	1.334	1.234	6.275	6.178	0.097	1.012

Observation	ATS7s	AATS8e	AATS3p	AATS5p	pIC_{50}	Predicted (pIC_{50})	Residual	Standardized residual
N_{31}*	563.278	7.571	1.299	1.26	6.431	6.353	0.078	0.814
N_{32}*	530.667	7.605	1.323	1.213	6.04	6.187	-0.147	-1.529
N_{34}*	547.25	7.617	1.299	1.192	6.585	6.425	0.160	1.656

*: test set

In general, the predicted values agree well with the experimental data, demonstrating the reliability of the QSAR model. Most standardized residuals fall within the range of -2 to +2, indicating that most predictions are within an acceptable error range. Extreme residuals were not observed, indicating the robustness of the model and the absence of significant outliers. These results confirm the model's predictive performance in estimating the biological activity (pIC_{50}) of the compounds studied (Figure 4).

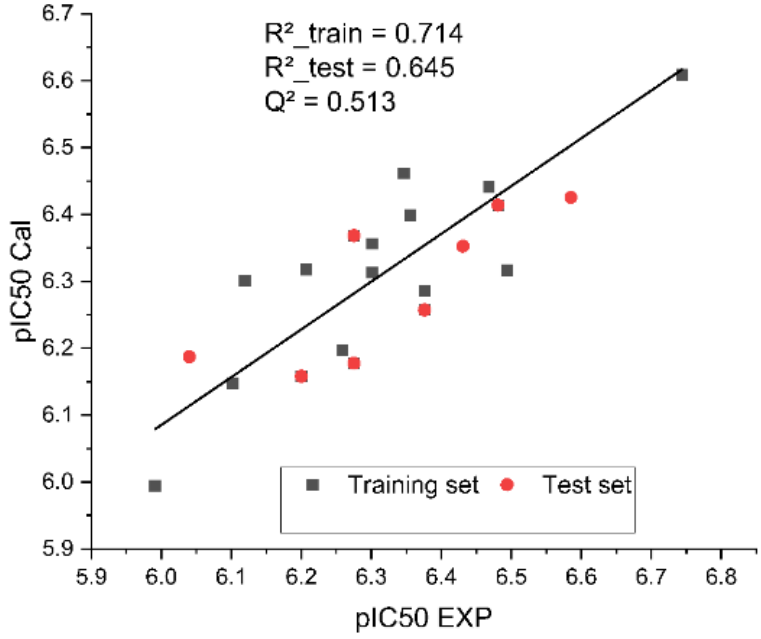


Figure 4. Correlations of observed and predicted activity values calculated using MLR models.

3.3. Internal and external validation of the QSAR model.

The performance of the QSAR model was evaluated in a way that followed the recommendations of the OECD (2007) and the principles created by Golbraikh and Tropsha [14]. This ensured that the results were reliable, robust, and could be reproduced.

3.3.1. Internal validation.

The regression statistics from the training set show that the model fits the data well. The determination coefficient (R^2 = 0.714) shows that the four selected molecular descriptors (ATS7s, AATS8e, AATS3p, and AATS5p) explain 71.4% of the variance observed in biological activity (pIC_{50}), demonstrating their structural relevance. The adjusted value of 0.660 confirms the model's stability and performance without overfitting, given the number of variables.

The low mean square error (MSE = 0.009), combined with a mean absolute percentage error (MAPE) of 1.132, highlights the precision of internal predictions. Furthermore, the absence of autocorrelation in residuals (DW = 2.019) validates the assumptions of the linear model. Information criteria (AIC = −117.245; BIC = -110.954) and Mallows' Cp (5.000) suggest an optimal balance between complexity and performance. Finally, leave-one-out cross-validation (Q^2 = 0.513) confirms good generalizability.

3.3.2. External validation.

External validation was performed on an independent test set to evaluate the predictive transferability of the developed QSAR model. The predictive squared correlation coefficient was R^2_test = 0.645, exceeding the recommended threshold (> 0.6).

The Golbraikh–Tropsha criteria were numerically satisfied:

- $\frac{(r^2 - r_0^2)}{r^2} = 0.04 (< 0.1)$
- $\frac{(r^2 - r'_0{}^2)}{r^2} = 0.05 (< 0.1)$
- k = 0.97 and k′ = 1.02 (0.85 < k, k′ < 1.15)

The root mean squared error for the test set was RMSE_test ≈ 0.107, indicating acceptable predictive accuracy.

The concordance correlation coefficient (CCC_test = 0.84) further confirmed the agreement between the experimental and predicted values.

These numerical results demonstrate that the model has an acceptable external predictive capability, as summarized in Table 7.

Table 7. Statistical parameters for the internal and external validation of the QSAR model.

Statistics	Training set	Test set
Number of compounds (n)	26	8
R^2	0.714	
Adjusted R^2	0.660	
Q^2 (LOO)	0.513	
Press	0.332	0.092
R^2pred		0.645
r^2m		0.595
Δr^2m		0.03
CCC		0.84
K		1.02
k′		0.97

3.3.3. Model robustness and Y-randomization analysis.

To further evaluate the robustness of the QSAR model and to address potential concerns related to the compound-to-descriptor ratio ($\frac{n}{p}$ = 8.5), additional validation procedures were performed.

A Y-randomization (Y-scrambling) test was conducted using 1000 permutations, in which the response variable (pIC_{50}) was randomly shuffled while keeping the descriptor matrix unchanged. The resulting randomized models produced significantly lower statistical parameters compared to the original model. The mean values of R^2_random and Q^2_random were 0.164 and -0.275, respectively, while the original model showed R^2 = 0.714.

None of the randomized models reached the predictive performance of the actual QSAR model, clearly demonstrating that the observed correlation is not due to chance but reflects a genuine structure–activity relationship.

Additionally, rm^2 metrics were evaluated to assess predictive symmetry. The mean r^2m value (0.595) exceeded the recommended threshold (> 0.5), while Δr^2m (0.03) was well below the critical value of 0.2, indicating balanced predictive performance.

Overall, these robustness analyses demonstrate that the developed QSAR model is statistically stable, non-random, and reliable for predicting the biological activity of structurally related compounds, as summarized in Table 8.

Table 8. Results of the Y-randomization test performed to evaluate the robustness of the developed QSAR model.

Validation test	Result
Number of random runs	50
Mean R^2_random	0.164
Mean Q^2_random	−0.275
R^2_model	0.714
Q^2_model	0.513

3.3.4. Domain of applicability (DA).

William's diagram (Figure 5) was used to evaluate the applicability domain, with leverage values (hi) and standardized residuals cross-referenced. The critical threshold (h*) was calculated using the formula $h^* = \frac{(3(p+1))}{n}$, where p represents the number of descriptors (four), and n represents the number of training molecules (26). This calculation yielded h* = 0.577 [29].

The results show that the applicability domain includes all molecules in the training set ($h_i < h^*$), which reinforces the statistical reliability of the model. In terms of the test set, four compounds exceed this threshold.

This indicates that these compounds fall outside the descriptor space defined by the training set and, therefore, represent structural extrapolations rather than interpolations within the calibrated chemical domain.

However, previous studies have shown that moderate leverage deviations do not necessarily compromise prediction reliability, provided that standardized residuals remain within acceptable limits (±3) [30,31].

However, despite their standardized residuals remaining within ±3, predictions for these high-leverage compounds should be interpreted with caution, as extrapolation beyond the applicability domain may reduce predictive reliability.

Thus, while some points are technically out of range, their predictions may still be useful, particularly in an exploratory or prospective context.

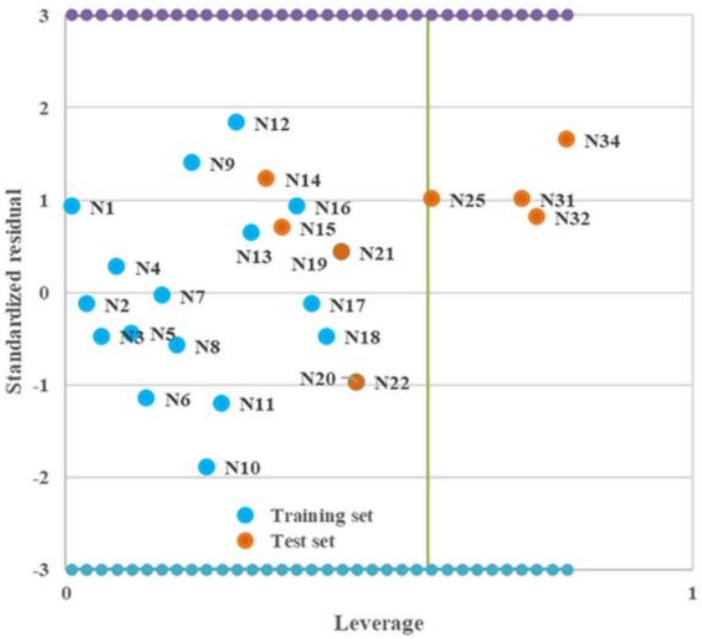


Figure 5. The William diagram illustrates the leverage values and standardized residuals of the compounds in the training and validation sets.

Consequently, the predictive reliability of these compounds is inherently limited by their structural distance from the training chemical space, and conclusions derived from these predictions should be considered with caution.

The internal and external validation results confirm that the proposed QSAR model is statistically robust, predictive, and scientifically relevant. Therefore, it is a reliable tool to predict the biological activity of analogous compounds.

3.4. ADMET.

An important step in selecting a drug candidate is evaluating ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties. This technology allows us to identify molecules that behave well in the body, even before experimental testing. In this study, we used the Swiss ADME online platform [31].

To estimate the «druglikeness» of predicted compounds based on five recognized filters: Lipinski, Ghose, Veber, Egan, and Muegge (Table9).

Table 9. Evaluation of drug similarity of compounds according to Lipinski, Ghose, Veber, Egan, and Muegge rules.

Comp	Lipinski	Ghose	Veber	Egan	Muegge
N27	No	No	Yes	No	Yes
X1	No	No	No	No	No
X2	No	No	No	No	No
X3	No	No	Yes	No	No
X4	No	No	Yes	No	No
X5	No	No	No	No	No
X6	No	No	No	No	No
X7	No	No	Yes	Yes	Yes
X8	No	No	No	No	No
X9	No	No	Yes	Yes	No
X10	No	No	Yes	No	Yes
X11	No	No	Yes	Yes	No
X12	No	No	No	No	No
X13	No	No	No	No	No
X14	No	No	Yes	Yes	No
TMZ	Yes	No	Yes	Yes	No

The results indicate that temozolomide, the reference drug, satisfies the Lipinski, Veber, and Egan criteria, affirming its favorable pharmacokinetic profile. However, several predictive molecules did not meet some of these rules, notably those of Lipinski. However, the X14 molecule stood out by meeting the Veber and Egan criteria, indicating good molecular flexibility and adaptive polarity, two favorable factors for efficient oral absorption.

These results suggest that X14 may have a favorable pharmacokinetic profile. They also suggest that it has potential biological activity. This indicates that X14 could be considered a promising candidate for further development, as summarized in Table 10**.**

Table 10. ADME and basic toxicity parameters of the studied ligands.

Compound	% absorption	LogBBB	AMES toxicity	Max. tolerated dose (human)	hERG I inhibitor	hERG II inhibitor
N27	96.726	-1.676	No	-0.359	No	Yes
X1	93.236	-1.743	No	-0.311	No	Yes
X2	89.141	-1.715	No	-0.153	No	Yes
X3	99.362	-1.994	No	-0.258	No	Yes
X4	95.248	-1.980	No	-0.111	No	Yes
X5	87.761	-1.959	No	0.166	No	Yes
X6	98.393	-2.142	No	0.379	No	Yes
X7	98.117	-1.789	No	-0.462	No	Yes
X8	90.621	-3.285	No	-0.122	No	Yes

Compound	% absorption	LogBBB	AMES toxicity	Max. tolerated dose (human)	hERG I inhibitor	hERG II inhibitor
X9	99.599	-1.969	No	-0.412	No	Yes
X10	94.781	-1.988	No	-0.065	No	Yes
X11	100.000	-1.805	No	-0.457	No	Yes
X12	86.154	-1.856	No	-0.309	No	Yes
X13	90.903	-1.975	No	0.255	No	Yes
X14	**90.156**	**-1.649**	**No**	**-0.101**	**No**	**Yes**
TMZ	76.019	-1.347	Yes	0.661	No	No

Overall, the toxicity and ecotoxicological profiles of the studied ligands were systematically evaluated to assess their potential environmental and biological safety, as summarized in Table 11.

Table 11. Predicted toxicity and ecotoxicological parameters of the studied ligands.

Compound	Oral rat acute toxicity (LD_{50}, mol/kg)	Chronic toxicity (LOAEL, log g/kg/day	Hepatotoxicity	Skin sensitisation	*T. pyriformis* toxicity (log µg/L)	Minnow toxicity (log mM)	Bioaccumulation logL/Kg)	Blood–brain permeability (LogPS)
N27	2.941	1.533	Yes	No	0.285	1.362	1.295	-3.276
X1	2.791	1.530	Yes	No	0.285	1.858	1.138	-3.344
X2	3.093	1.624	Yes	No	0.285	1.783	1.364	-3.386
X3	2.911	1.384	Yes	No	0.285	1.301	1.140	-3.316
X4	3.089	0.857	Yes	No	0.285	2.417	0.555	-3.283
X5	3.057	1.077	Yes	No	0.285	2.879	0.632	-3.398
X6	2.657	1.683	Yes	No	0.285	2.241	1.313	-3.809
X7	3.349	-0.216	Yes	No	0.288	2.170	0.183	-3.171
X8	3.305	0.689	Yes	No	0.286	2.632	0.183	-3.285
X9	3.308	-0.197	Yes	No	0.287	1.904	0.146	-3.188
X10	3.045	0.641	Yes	No	0.287	2.440	0.487	-3.231
X11	3.278	0.409	Yes	No	0.288	1.869	0.174	-3.178
X12	2.814	1.581	Yes	No	0.285	1.411	0.450	-3.249
X13	3.153	0.606	Yes	No	0.285	2.118	0.055	-3.290
X14	**2.980**	**0.189**	**Yes**	**No**	**0.290**	**1.570**	**-0.089**	**-3.178**
TMZ	2.708	0.552	Yes	No	0.285	2.983	0.338	-3.261

LD_{50}: median lethal dose; LOAEL: lowest observed adverse effect level; BBB: blood–brain barrier.

The results obtained (Tables 10 and 11) justify the inclusion of additional ADMET parameters to better assess the therapeutic potential of the studied molecules, thereby helping determine their effectiveness and potential benefits. Analyzing the pharmacokinetic and toxicological properties reveals the exceptional characteristics of certain compounds, particularly X14.

It is important to note that the hepatotoxicity model predicted potential hepatotoxicity for all the compounds evaluated, including the reference drug temozolomide (TMZ). This uniform prediction suggests that the in silico model may provide conservative alerts for this chemical class rather than a clear discriminatory toxicity profiling. Therefore, these predictions of hepatotoxicity should be interpreted as preliminary safety indicators that require experimental validation rather than definitive toxicological conclusions.

Although X14 meets only two of the «druglikeness» criteria, it has a generally favorable ADMET profile. It is characterized by high absorption (90.15%), absence of mutagenic toxicity (AMES test: negative outcome [NO]), and lack of hERG I channel inhibition, reducing the risk of adverse cardiac effects. Furthermore, its acute and chronic oral toxicity is low (LD50 = 2.98 mol/kg; LOAEL = 0.189 mol/kg) with predicted hepatotoxicity but without skin sensitization effects.

It is important to note that the hepatotoxicity model predicted potential hepatotoxicity for all the compounds evaluated, including the reference drug temozolomide (TMZ). This

uniform prediction suggests that the in silico model may provide conservative alerts for this chemical class rather than clear discriminatory toxicity profiling. Therefore, these predictions of hepatotoxicity should be interpreted as preliminary safety indicators that require experimental validation rather than definitive toxicological conclusions.

The X7 molecule shows high absorption (98.11%) and a satisfactory pharmacokinetic profile. However, it exhibits slight chronic toxicity (negative LOAEL) and inhibition of the hERG II channel, which can pose a risk. On the contrary, temozolomide, an approved drug, has certain limitations: lower absorption (76.01%) and proven mutagenic toxicity (AMES test: yes).

In general, these results highlight the promising therapeutic potential of the 14 molecules. It combines satisfactory biological tolerance with favorable pharmacokinetic characteristics. However, its predicted low BBB permeability (logBBB < –1) may limit its direct application in glioblastoma therapy and warrants further optimization or delivery-based enhancement strategies.

Since blood–brain barrier permeability is a critical requirement for glioblastoma therapy, compounds exhibiting low predicted logBBB values represent pharmacokinetic limitations that may compromise in vivo efficacy despite favorable docking performance. Therefore, structural optimization aimed at reducing the polar surface area and fine-tuning lipophilicity should be considered in future development.

However, this limitation, common in the development of drugs targeting the central nervous system, can be overcome through vectorization strategies or targeted chemical modifications, as reported in previous studies [31,32].

3.5. Docking results.

3.5.1. Complex temozolomide-3BD6.

The 3BD6 temozolomide complex is stabilized by several interactions, mainly hydrogen bonds and a pi-cation interaction. Cysteine 119 (CYS119) and arginine 122 (ARG122) form hydrogen bonds with the ligand at distances of 2.10 Å and 2.78 Å, respectively, indicating strong interactions that enhance the anchoring of the ligand.

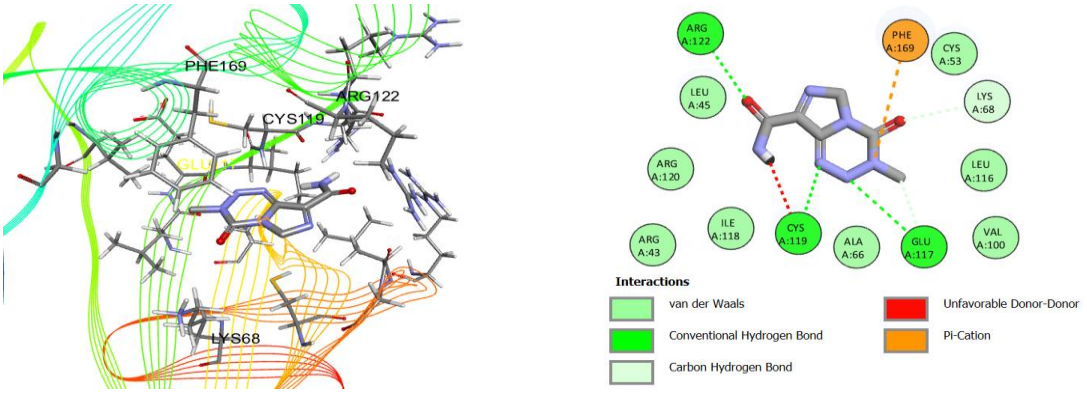


Figure 6. Interaction profile of the temozolomide–PLK1 (3BD6) complex.

Furthermore, glutamate 117 (GLU117) and lysine 68 (LYS68) form multiple hydrogen bonds at 2.76, 3.24, 2.82, and 3.56 Å, contributing to stabilization of the complex through electrostatic interactions between opposite charges of the functional groups. Furthermore, phenylalanine 169 (PHE169) participates in a pi-cation interaction with the ligand at a distance of 4.95 Å, where the electron cloud of its aromatic ring interacts with a positive charge of the

ligand, affecting its position and stability at the active site (Figure 6). The total energy affinity of -6.6 kcal/mol reflects moderately strong binding, suggesting efficient molecular recognition between temozolomide and the 3BD6 protein.

3.5.2. Complex N27-3BD6.

The complex N27-3BD6.PDB establishes several key interactions with target protein residues, explaining its affinity of -8.1 kcal/mol. Hydrogen bond interaction with PHE50 (2.88) improves the stability of the complex. The presence of pi-sigma interactions with VAL70 (3.64 Å) and pi-sulfur with MET84 (5.45 Å) suggests an important role of π forces in stabilization. Alkyl interactions with LYS68, LYS83, VAL114, and LEU116 (distances between 4.55 and 5.36 Å) and Pi-Alkyl with several residues, including HIS91, ALA51, LYS68, VAL70, and VAL114 (distances between 4.51 and 5.43 Å), also contribute to ligand anchoring through hydrophobic interactions. Finally, an essential role in molecular recognition is played by the pi-cation interaction with PHE169 (4.95 Å). Together, these interactions explain the complex's high affinity and its potential to stabilize the protein's active site (Figure 7).

Figure 7. Visualization of the interactions of the complex N27-PLK1 (3BD6).

3.5.3. Complex X9-3BD6.

The X9 forms with protein (ID: 3BD6) several stabilizing interactions with the surrounding residues, which promote its affinity (-8.9 Kcal/mol) with the target.

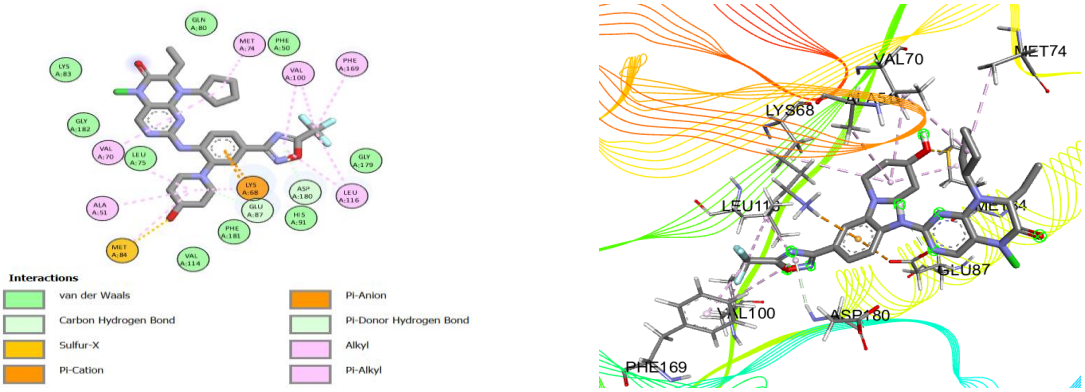


Figure 8. Visualization of the interactions of the complex X9-PLK1 (3BD6).

A carbon-hydrogen interaction is observed with GLU87 (3.71 Å), as well as electrostatic interactions, in particular Pi-cation / Pi donor with LYS68 (2.84 Å) and Pi donor with ASP180 (2.81 Å). The latter is associated with an affinity of -8.9 kcal/mol, indicating a strong interaction. GLU87 is also involved in a Pi anion interaction (3.37 Å). A specific sulfur-X interaction is observed with MET84 (3.29 Å). Hydrophobic interactions are also well

represented, with alkyl-alkyl contacts involving ALA51, LYS68, VAL70, MET74, MET84, VAL100, and LEU116 at distances ranging from 4.31 Å to 5.45 Å, strengthening the stability of the complex. Finally, pi-alkyl interactions are established with PHE169, VAL100, and LEU116 (4.20 to 5.39 Å), which contribute to the overall affinity of the ligand for its target. Taken together, these interactions suggest a stable complex involving electrostatic, hydrophobic, and specific forces (Figure 8).

3.5.4. Complex X7-3BD6.

The interaction between the X7 ligand and the 3BD6 protein, with an affinity of -9.1 kcal/mol, is mediated by a carbon-hydrogen interaction with ASP180 (2.64 Å and 3.79 Å), Pi cation/ Pi donor interactions with LYS68 (2.81 Å) and Pi anion interactions with GLU87 (3.38 Å), and alkyl-alkyl interactions with ALA51 (5, 28 Å and 5.33Å), LYS68 (5.23Å and 4.87Å), VAL70 (5.18Å and 5.22Å) and MET74 (4.29Å and 4.44Å), and finally a Pi-Alkyl interaction with PHE169 at 4.29Å, demonstrating a combination of hydrophobic, electrostatic and Van der Waals stabilization forces that are favorable for binding (Figure 9).

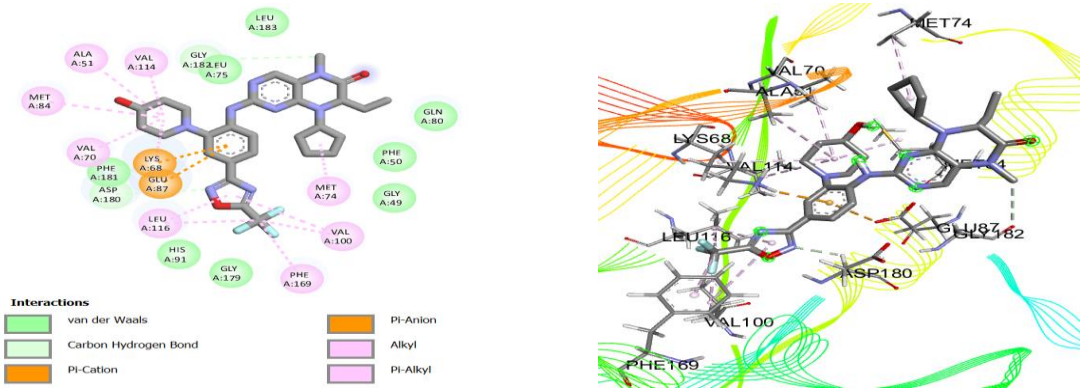


Figure 9. Visualization of the interactions of the X7–PLK1 (3BD6) complex.

3.5.5. Complex X11-3BD6.

Ligand X11 interacts with glycine (GLY179), aspartic acid (ASP180), glutamine (GLN80), and lysine (LYS68) via carbon-hydrogen (2.93-3.75 Å) with an affinity of -9.1 kcal/mol, with glutamic acid (GLU87) via pi-anion (3.59-3.62 Å), with lysine (LYS83) and valine (VAL100) via alkyl-alkyl (4. (LYS68), (VAL100) and (LEU116) through pi-alkyl interactions (4.02-4.92Å), and finally with (PHE169) through a pi-pi stack interaction (5.53Å), demonstrating its stability and affinity with the target (Figure 10).

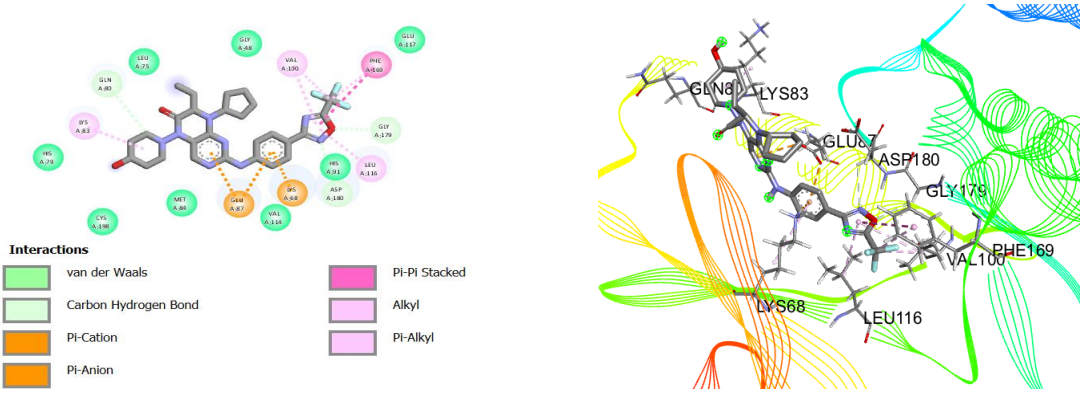


Figure 10. Visualization of the interactions of the X11–PLK1 (3BD6) complex.

3.5.6. Complex X14-3BD6.

The X14 complex establishes several key interactions with the target protein, which is explained by its high binding affinity, estimated at -10.5 kcal/mol. Conventional hydrogen bonds with CYS119 (2.47 Å and 2.20 Å) contribute significantly to the stability of the complex by forming intense polar contacts. The presence of π-type interactions enhances ligand stabilization at the active site: a π-cation interaction and a π-donor hydrogen bond with LYS68 (2.64 Å), a π-anion interaction with GLU87 (3.54 Å), a π-sulfur interaction with CYS119 (4.50 Å), and a π-π stack with PHE169 (4.61 Å and 4.54 Å). This highlights the role of aromatic forces in molecular recognition. Display Omitted Hydrophobic interactions also play a central role in ligand anchoring, including alkyl-alkyl interactions with ALA51, ALA66, LEU45, CYS119, MET84, and VAL114 (at distances between 4.20 Å and 5.06 Å), and π-alkyl interactions with LYS68 (5.47 Å and 4.72 Å), VAL100 (4.36 Å and 4.27 Å), LEU116, and ALA66. These essential hydrogen bonds, aromatic contacts, and hydrophobic interactions are responsible for the high affinity of the X14 ligand and its efficient stabilization in the protein (Figure 11).

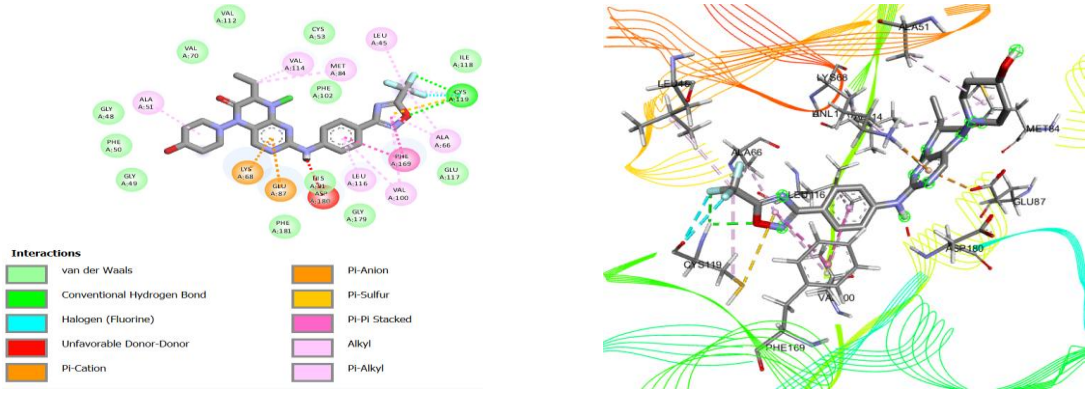


Figure 11. Visualization of the interactions of the X14–PLK1 (3BD6) complex.

3.6. Validation of the docking protocol.

To validate the reliability of the docking protocol, a redocking procedure was performed using the native co-crystallized ligand retrieved from the PLK1 crystal structure (PDB ID: 3BD6). The ligand was first removed from the binding pocket, and the protein was prepared following the same protocol applied to all tested compounds. The extracted ligand was subsequently redocked into the ATP-binding site using AutoDock Vina with identical docking parameters [33].

The grid box was centered on the original ligand binding site with the following coordinates: center_x = 30.426, center_y = 0.38, and center_z = 52.846, and grid dimensions of 22 × 22 × 22 Å, ensuring full coverage of the active site region.

The predicted docking pose was then superimposed onto the original crystallographic conformation, and the RMSD between heavy atoms was calculated. An RMSD value of 0.82 Å was obtained, which is well below the commonly accepted threshold of 2.0 Å, confirming the accuracy and reliability of the docking protocol used in this study.

Furthermore, the superimposition of the crystallographic and redocked conformations revealed very good spatial overlap, supporting the robustness of the docking procedure, as shown in Figure 12.

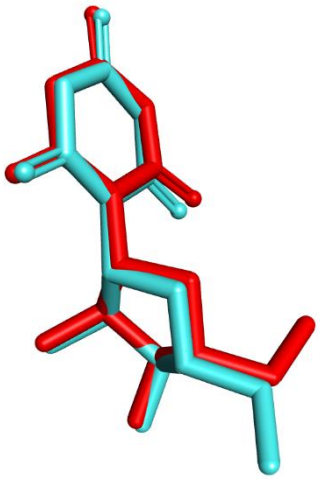


Figure 12. Validation of the docking protocol by redocking of the native ligand (B6H) into the PLK1 binding site (PDB ID: 3BD6). The crystallographic ligand conformation (cyan) is superimposed with the redocked pose (red), showing a very good overlap with an RMSD of 0.82 Å.

The detailed protein–ligand interaction profiles are provided in Supplementary Table S3.

3.7. Molecular dynamics.

3.7.1. Root mean square deviation (RMSD).

RMSD is considered a key indicator of the conformational stability of ligand-protein complexes (Figure 13). The temozolomide complex exhibits an exceptionally low and stable RMSD of approximately 0.75 Å, indicating a rigid binding of the ligand within the pocket. The excellent structural stability of the complex during simulation is reflected in this low variability. The predicted X14 molecule also shows a low RMSD of 0.5-1.2 Å, indicating exceptional conformational stability and a robust, durable interaction with the target protein.

Although X14 exhibits slightly lower RMSD values compared to temozolomide, these differences remain modest and fall within the typical fluctuation range observed in molecular dynamics simulations. Therefore, the results suggest comparable conformational stability rather than definitive superiority over the reference drug. This stability is reinforced by a network of specific interactions within the active site.

On the contrary, the ligand N27 exhibits slightly higher RMSD values, ranging from 1.5 to 2.0 Å. Although these values are still within an acceptable range, they indicate that the molecule is more mobile within the cavity. This suggests that the molecule has greater conformational flexibility or is able to adapt more dynamically.

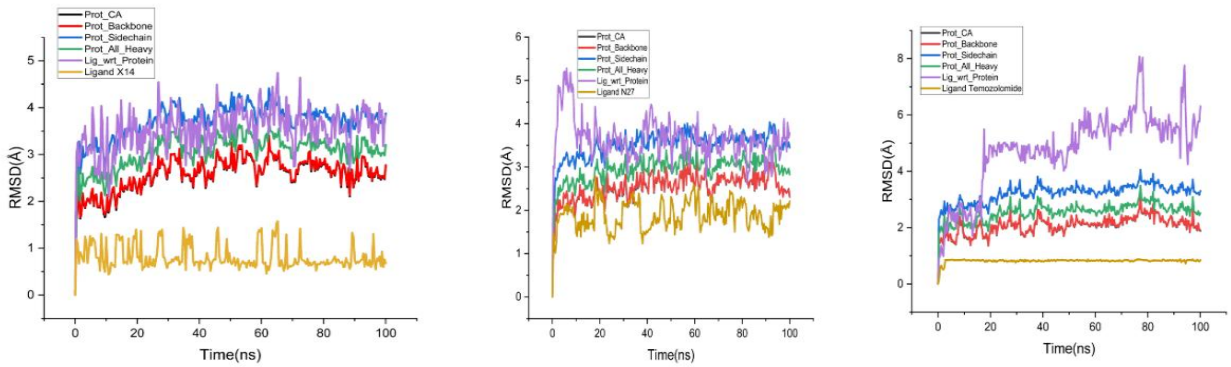


Figure 13. RMSD values of the complexes during 100 ns of MD simulations.

3.7.2. Root mean square fluctuation (RMSF).

The evaluation of protein residue flexibility and ligand conformational stability during molecular dynamics simulations was performed through RMSF analysis. The residue RMSF curves indicate that all three complexes exhibit acceptable overall structural stability. However, some notable differences have been observed (Figure 14). The X14 protein complex exhibits moderate fluctuations, ranging from 1.0 to 2.5 Å. Some localized peaks reach approximately 3.5 Å, remaining lower than the maximum fluctuations observed with temozolomide (up to 4.5 Å) and comparable to those of N27. It is important to note that there is still significant flexibility in the key parts of the active site. This suggests that the interaction between the ligand and the protein is well-maintained.

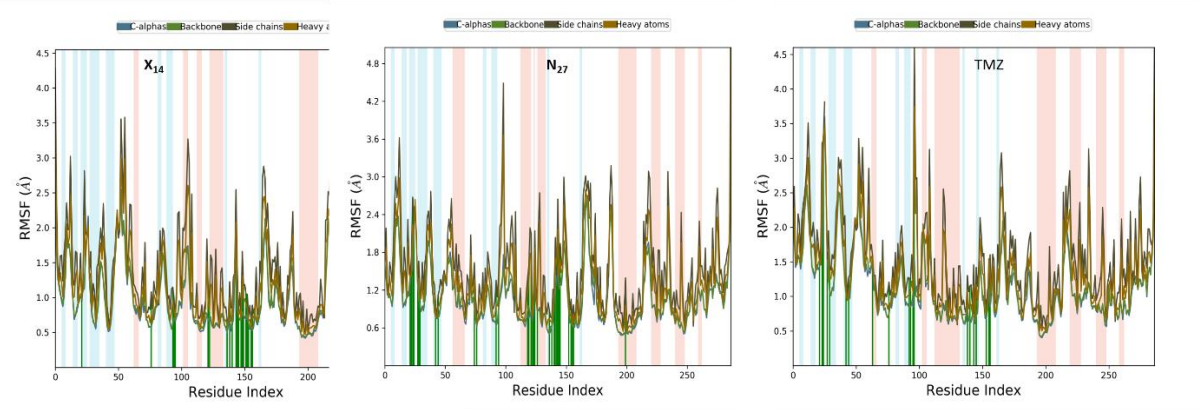


Figure 14. RMSF values of the complexes during 100 ns of MD simulations.

3.7.3. PL contact histogram.

Non-covalent ligand-protein interactions play a key role in binding affinity and specificity. Temozolomide establishes five hydrogen bonds. These are between CYS119, GLU117, and ARG122. There is also one π-cation interaction. This is between PHE169. These interactions indicate a moderate but targeted interaction within the active site.

The formation of a complex and diverse network of interactions by the X14 molecule includes strong hydrogen bonds (CYS119), π-anion (GLU87), π-cation (LYS68), π-Sulfur, and stacked π-π interactions (PHE169). The robust and specific binding of X14 is particularly notable due to its high binding affinity (-10.5 kcal/mol) and interactional richness.

The N27 ligand establishes a variety of interactions, such as H bonds (PHE50), π-sigma (VAL70), and π-Sulfur (MET84), as well as numerous hydrophobic interactions (ALKYL, π-ALKYL). Although these interactions are less intense and concentrated than those of X14, they remain functional and compatible with stable binding (Figure 15).

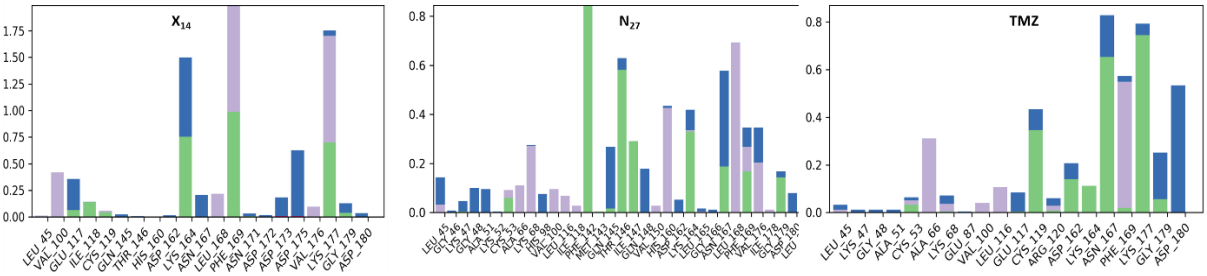


Figure 15. Protein–ligand contact histogram.

3.7.4. Properties of ligands.

The structural compactness of the ligand or complex is indicated by its radius of gyration (Rg). Temozolomide maintains a constant Rg of around 2.65 Å. This confirms a compact and rigid structure. This structure is conducive to stable binding. Rg fluctuates between 6.0 and 6.3 Å during dynamic simulation, with an average of 6.15 Å. Despite this range, Rg remains relatively stable over time, suggesting a constant conformation and an absence of major unfolding. This indicates a stable interaction within the complex.

For N27, Rg is measured throughout the complex (protein + ligand) and oscillates between 6.6 and 6.75 Å. This indicates a good overall compactness with no significant unfolding of the protein structure. The degree of exposure of the ligand to the aqueous environment is measured by the solvent-accessible surface area (SASA). During simulation, temozolomide shows a reduction in SASA from 120 Å to approximately 60 Å. This decrease suggests that the drug is gradually becoming more deeply embedded in the binding pocket, a process conducive to its therapeutic effectiveness.

Ligand X14 exhibits a relatively stable SASA (80-100 Å), reflecting consistent interaction with the receptor without significant burial, thereby maintaining a balanced exposure.

N27 SASA is also stable (~90-120 Å²), but slightly higher. This indicates that the molecule is partially exposed to the solvent and, therefore, less integrated into the cavity. This may affect the binding intensity (Figure 16).

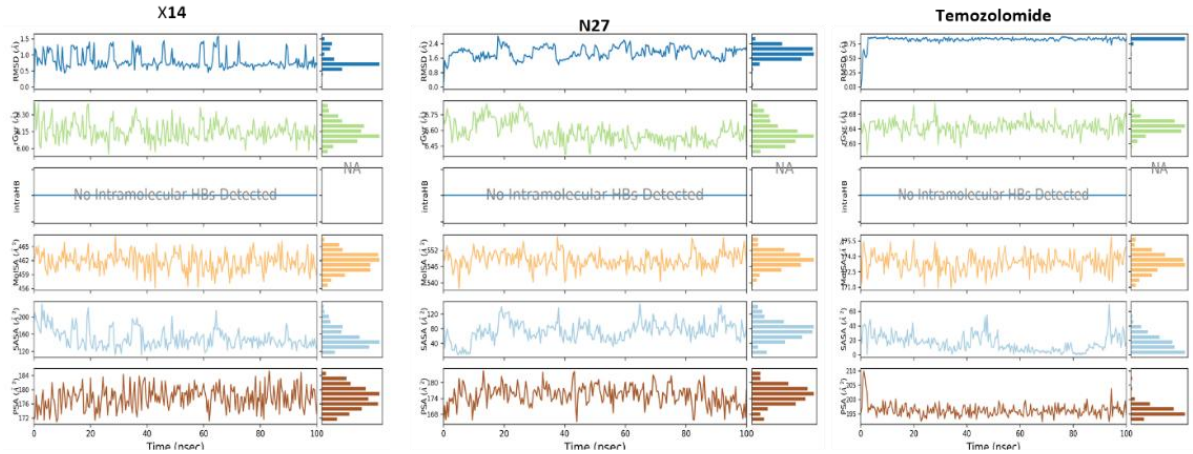


Figure 16. Comprehensive analysis of ligand properties influencing protein binding.

3.7.5. Secondary structure analysis.

The stability of the secondary structure during 100-ns molecular dynamics simulations was evaluated by analyzing the evolution of secondary structure elements (SSE).

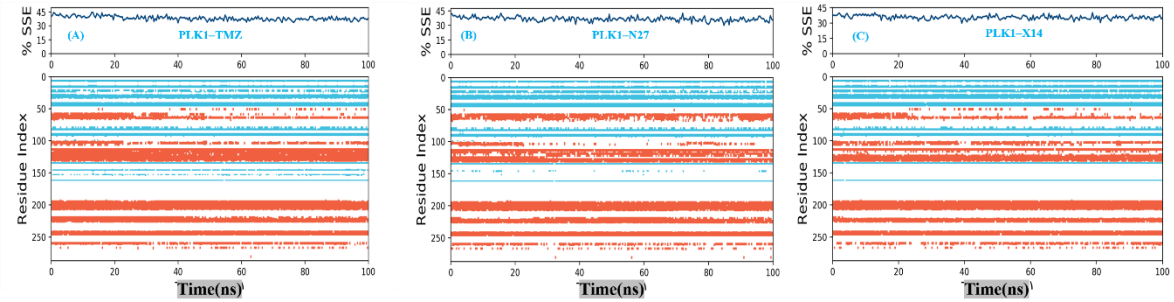


Figure 17. Evolution of the secondary structure of PLK1 in complex with TMZ, N27, and X14.

As shown in Figure 17, the percentage of SSE remains relatively stable for the three complexes (PLK1–TMZ, PLK1–N27, and PLK1–X14), fluctuating around 33–40% throughout the simulation.

The PLK1–TMZ complex exhibits a stable secondary structure profile, indicating that the reference ligand does not induce structural perturbations in the protein. A similar behavior is observed for the PLK1–N27 and PLK1–X14 complexes, where the α-helices and β-strands remain well preserved during the trajectory.

In general, the comparable SSE patterns observed across the three systems suggest that ligand binding does not significantly affect PLK1's global structural stability during simulation.

A comparative analysis of the molecular dynamics trajectories indicates that the three complexes maintain a stable structural profile throughout the simulation. The PLK1–TMZ complex exhibits a stable conformational behavior consistent with the reference drug. Similar stability is observed for the PLK1–N27 and PLK1–X14 systems, with only minor fluctuations occurring in flexible loop regions. Overall, the comparable dynamic behavior observed for the three complexes suggests that the predicted ligands maintain stable interactions within the PLK1 binding pocket during the 100 ns simulation.

3.7.6. Ligand RMSF analysis.

The conformational flexibility of the ligands during the molecular dynamics simulations was further evaluated by ligand RMSF analysis.

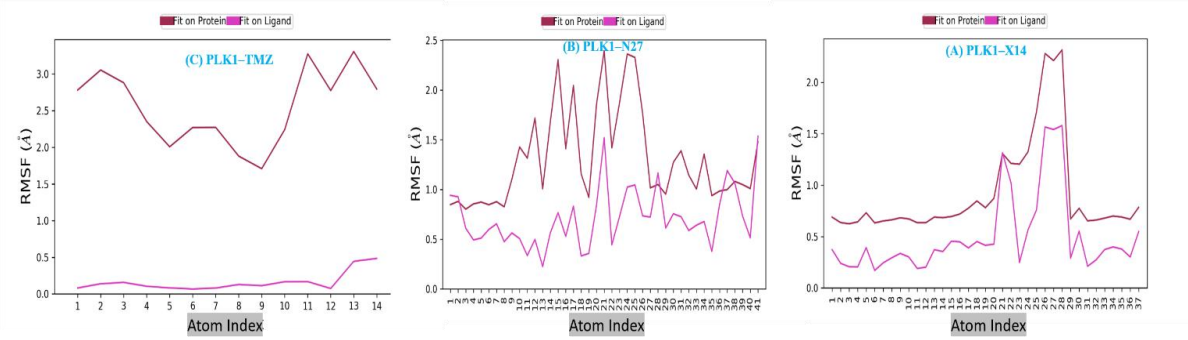


Figure 18. RMSF analysis of PLK1 complexes during MD simulations.

As shown in Figure 18, most of the ligand atoms exhibit relatively low fluctuation values, indicating stable conformations within the PLK1 binding pocket. The reference drug, temozolomide (TMZ), shows greater fluctuations in some atoms, suggesting greater flexibility in certain regions of the molecule. In comparison, the predicted compounds N27 and X14 show moderate fluctuations, mainly located in peripheral atoms that are more exposed to solvent. Overall, the RMSF profiles indicate that all three ligands maintain stable binding conformations within the active site during the 100 ns simulation.

Although binding free energy calculations such as MM-PBSA/MM-GBSA can provide additional quantitative insight into ligand affinity, the stability of the protein–ligand complexes was comprehensively evaluated through multiple complementary MD descriptors, including RMSD, RMSF, secondary structure evolution, protein–ligand interaction profiles, and ligand flexibility analysis. These combined analyses consistently indicate stable binding behavior of the investigated ligands within the active site.

Overall, molecular dynamics simulations consistently indicate that the PLK1–TMZ, PLK1–N27, and PLK1–X14 complexes remain structurally stable throughout the 100 ns

trajectory. The combined analyses of RMSD, RMSF, protein–ligand interaction profiles, ligand flexibility, and secondary structure evolution collectively support the stability of the investigated ligands within the PLK1 binding pocket.

4. Conclusions

In this study, an integrated in silico approach combining QSAR modeling, molecular docking, molecular dynamics simulations, and ADMET profiling was employed to evaluate the anticancer potential of two predictive derivatives, X14 and N27, in comparison with temozolomide, the standard chemotherapeutic agent used in glioblastoma treatment. This multilevel computational strategy provided structural and pharmacokinetic insights into the behavior of these candidates.

The results indicate that X14 exhibits binding characteristics comparable to temozolomide, with stable RMSD values during MD simulations and a consistent interaction network within the active site. Although docking scores and interaction patterns suggest favorable affinity, observed differences relative to temozolomide remain moderate and should be interpreted with caution. N27 also demonstrated stable binding behavior, albeit with slightly lower predicted interaction strength.

ADMET analysis revealed generally favorable pharmacokinetic profiles, including high intestinal absorption and absence of predicted mutagenicity. However, universal hepatotoxicity alerts across the data set and variability in blood–brain barrier (BBB) permeability represent important pharmacological considerations. In particular, the predicted limited permeability of the BBB of X14 (logBBB < –1) constitutes a significant limitation for glioblastoma therapy, where penetration into the central nervous system is essential. Therefore, structural optimization or alternative delivery strategies would be required to enhance cerebral bioavailability before considering translational development.

Importantly, the present findings are based exclusively on computational methods. Therefore, experimental validation is essential to confirm the predicted biological activity, safety profile, and pharmacokinetic behavior of these derivatives. Future studies must include in vitro cytotoxicity assays, PLK1 inhibition experiments, and in vivo pharmacokinetic and biodistribution evaluations.

Overall, this study provides a computational framework to guide the rational optimization of PLK1-targeting compounds, while clearly acknowledging the need for experimental confirmation before any therapeutic application can be considered.

Authors' Contributions

Conceptualization, Y.B. and M.E.; methodology, Y.B.; software, Y.B.; validation, Y.B., M.E.-R., and J.E.; formal analysis, Y.B.; investigation, Y.B.; resources, M.E.; data curation, Y.B.; writing—original draft preparation, Y.B.; writing—review and editing, Y.B., M.E.-R., J.E., and S.H.; visualization, Y.B.; supervision, M.E.; project administration, Y.B.; funding acquisition, M.E. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data supporting the findings of this study are available upon reasonable request from the corresponding author.

Funding

This research received no external funding.

Acknowledgments

The authors acknowledge the LIMAS Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco, for providing the computational facilities and academic environment that supported this research. The authors also gratefully acknowledge the Moroccan Association of Theoretical Chemists (MATC) for technical assistance and access to computational programs that facilitated this study. The individuals and organizations acknowledged had no role in the study design, data collection, analysis, interpretation, manuscript preparation, or the decision to publish.

Conflicts of Interest

The authors declare no conflict of interest.

Role of Funders

The funders had no role in the design of the study, the collection, analysis, or interpretation of data, the writing of the manuscript, or the decision to publish the results.

Supplementary Materials

Table S1. List of calculated molecular descriptors for each compound in the studied series.
Table S2. Correlation matrix used for principal component analysis (PCA).
Table S3. Detailed protein–ligand interaction profiles of the studied compounds obtained from molecular docking analysis.
The supplementary file is available for download here.

References

1. Stupp, R.; Hegi, M.E.; Mason, W.P.; van den Bent, M.J.; Taphoorn, M.J.B.; Janzer, R.C.; Ludwin, S.K.; Allgeier, A.; Fisher, B.; Belanger, K.; Hau, P.; Brandes, A.A.; Gijtenbeek, J.; Marosi, C.; Vecht, C.J.; Mokhtari, K.; Wesseling, P.; Villa, S.; Eisenhauer, E.; Gorlia, T.; Weller, M.; Lacombe, D.; Cairncross, J.G.; Mirimanoff, R.-O. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol.* **2009**, *10*, 459-466, https://doi.org/10.1016/S1470-2045(09)70025-7.
2. van Opijnen, M.P.; van Valkengoed, D.R.; de Ligt, J.; de Vos, F.Y.F.; Broekman, M.L.D.; Cuppen, E.; Koster, R. Whole genome sequencing-based analysis of genetic predisposition to adult glioblastoma. *npj Genom. Med.* **2025**, *10*, 70, https://doi.org/10.1038/s41525-025-00526-z.

3. Le Rhun, E.; Preusser, M.; Roth, P.; Reardon, D.A.; van den Bent, M.; Wen, P.; Reifenberger, G.; Weller, M. Molecular targeted therapy of glioblastoma. *Cancer Treat. Rev.* **2019**, *80*, 101896, https://doi.org/10.1016/j.ctrv.2019.101896.
4. Silginer, M.; Nagy, S.; Happold, C.; Schneider, H.; Weller, M.; Roth, P. Autocrine activation of the IFN signaling pathway may promote immune escape in glioblastoma. *Neuro-Oncology* **2017**, *19*, 1338-1349, https://doi.org/10.1093/neuonc/nox051.
5. Gilbert Mark, R.; Dignam James, J.; Armstrong Terri, S.; Wefel Jeffrey, S.; Blumenthal Deborah, T.; Vogelbaum Michael, A.; Colman, H.; Chakravarti, A.; Pugh, S.; Won, M.; Jeraj, R.; Brown Paul, D.; Jaeckle Kurt, A.; Schiff, D.; Stieber Volker, W.; Brachman David, G.; Werner-Wasik, M.; Tremont-Lukats Ivo, W.; Sulman Erik, P.; Aldape Kenneth, D.; Curran Walter, J.; Mehta Minesh, P. A Randomized Trial of Bevacizumab for Newly Diagnosed Glioblastoma. *N. Engl. J. Med.* **2014**, *370*, 699-708, https://doi.org/10.1056/NEJMoa1308573.
6. Chinnaiyan, P.; Won, M.; Wen, P.Y.; Rojiani, A.M.; Werner-Wasik, M.; Shih, H.A.; Ashby, L.S.; Michael Yu, H.-H.; Stieber, V.W.; Malone, S.C.; Fiveash, J.B.; Mohile, N.A.; Ahluwalia, M.S.; Wendland, M.M.; Stella, P.J.; Kee, A.Y.; Mehta, M.P. A randomized phase II study of everolimus in combination with chemoradiation in newly diagnosed glioblastoma: results of NRG Oncology RTOG 0913. *Neuro-Oncology* **2018**, *20*, 666-673, https://doi.org/10.1093/neuonc/nox209.
7. Wick, W.; Osswald, M.; Wick, A.; Winkler, F. Treatment of glioblastoma in adults. *Ther. Adv. Neurol. Disord.* **2018**, *11*, 1756286418790452, https://doi.org/10.1177/1756286418790452.
8. Gállego, J.; Honnorat, J.; Chinot, O.; Catry-Thomas, I.; Taillandier, L.; Guillamo, J.S.; Campello, C.; Monjour, A.; Mokhtari, K.; Tanguy, M.L.; Delattre, J.Y. Chimiothérapie par temozolomide dans le traitement des glioblastomes des sujets âgés présentant un état fonctionnel altéré. *Rev. Neurol.* **2012**, *168*, S31-S32, https://doi.org/10.1016/S0035-3787(12)70042-5.
9. Hegi Monika, E.; Diserens, A.-C.; Gorlia, T.; Hamou, M.-F.; de Tribolet, N.; Weller, M.; Kros Johan, M.; Hainfellner Johannes, A.; Mason, W.; Mariani, L.; Bromberg Jacoline, E.C.; Hau, P.; Mirimanoff René, O.; Cairncross, J.G.; Janzer Robert, C.; Stupp, R. MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. *N. Engl. J. Med.* **2005**, *352*, 997-1003,https://doi.org/10.1056/NEJMoa043331.
10. Zhu, Y.; Carabenciov, D.D.; Johnson, D.R.; Trejo-Lopez, J.A.; Nguyen, A.T.; Raghunathan, A.; Lanzino, G.; Ida, C.M.; Zepeda-Mendoza, C.J.; Dasari, S.; Russler-Germain, E.; Dahiya, S.; Quezado, M.; Aldape, K.; Giannini, C. Molecular profile of adult primary leptomeningeal gliomatosis aligns with glioblastoma, IDH-wildtype. *Brain Pathol.* **2025**, *35*, e13326, https://doi.org/10.1111/bpa.13326.
11. Kothe, M.; Kohls, D.; Low, S.; Coli, R.; Cheng, A.C.; Jacques, S.L.; Johnson, T.L.; Lewis, C.; Loh, C.; Nonomiya, J.; Sheils, A.L.; Verdries, K.A.; Wynn, T.A.; Kuhn, C.; Ding, Y.-H. Structure of the Catalytic Domain of Human Polo-like Kinase 1. *Biochemistry* **2007**, *46*, 5960-5971, https://doi.org/10.1021/bi602474j.
12. Pan, M.; Cheng, L.; Wang, Y.; Lyu, C.; Hou, C.; Zhang, Q. Exploration of 2D and 3D-QSAR analysis and docking studies for novel dihydropteridone derivatives as promising therapeutic agents targeting glioblastoma. *Front. Pharmacol.* **2023**, *14*, 1249041, https://doi.org/10.3389/fphar.2023.1249041.
13. Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R.; Consonni, V.; Kuz'min, V.E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57*, 4977-5010, https://doi.org/10.1021/jm4004285.
14. Tropsha, A.; Gramatica, P.; Gombar, V.K. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* **2003**, *22*, 69-77, https://doi.org/10.1002/qsar.200390007.
15. OECD. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, OECD Series on Testing and Assessment, No. 69; OECD Publishing: Paris, **2014**; https://doi.org/10.1787/9789264085442-en.
16. Hollingsworth, S.A.; Dror, R.O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129-1143, https://doi.org/10.1016/j.neuron.2018.08.011.
17. Yap, C.W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466–1474, https://doi.org/10.1002/jcc.21707.
18. XLSTAT. Available online: https://www.xlstat.com/ (accessed on Day Month **Year**) (accessed on 07 June 2026).

19. P, U.P.; Suresh, M.; Tolasa, F.T.; Bonyah, E. QSPR/QSAR study of antiviral drugs modeled as multigraphs by using TI's and MLR method to treat COVID-19 disease. *Sci. Rep.* **2024**, *14*, 13150, https://doi.org/10.1038/s41598-024-63007-w.
20. Sastry, M.; Lowrie, J.F.; Dixon, S.L.; Sherman, W. Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments. *J. Chem. Inf. Model.* **2010**, *50*, 771-784, https://doi.org/10.1021/ci100062n.
21. Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **2007**, *26*, 694-701, https://doi.org/10.1002/qsar.200610151.
22. Ferreira, L.G.; Dos Santos, R.N.; Oliva, G.; Andricopulo, A.D. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* **2015**, *20*, 13384-13421, https://doi.org/10.3390/molecules200713384.
23. Xuan-Yu, M.; Hong-Xing, Z.; Mihaly, M.; Meng, C. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput.-Aided Drug Des.* **2011**, *7*, 146-157, https://doi.org/10.2174/157340911795677602.
24. Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. -Aided Mol. Des.* **2013**, *27*, 221-234, https://doi.org/10.1007/s10822-013-9644-8.
25. Shelley, J.C.; Cholleti, A.; Frye, L.L.; Greenwood, J.R.; Timlin, M.R.; Uchimaya, M. Epik: a software program for pKaprediction and protonation state generation for drug-like molecules. *J. Comput. -Aided Mol. Des.* **2007**, *21*, 681-691, https://doi.org/10.1007/s10822-007-9133-z.
26. Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J.M.; Lu, C.; Dahlgren, M.K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R.A.; Harder, E.D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1863-1874, https://doi.org/10.1021/acs.jctc.8b01026.
27. Kaiser, H.F. The Application of Electronic Computers to Factor Analysis. *Educ. Psychol. Meas.* **1960**, *20*, 141–151, https://doi.org/10.1177/001316446002000116.
28. Gonçalves, R.B.; Ferraz, W.R.; Calil, R.L.; Scotti, M.T.; Trossini, G.H.G. Convergent QSAR Models for the Prediction of Cruzain Inhibitors. *ACS Omega* **2023**, *8*, 38961-38982, https://doi.org/10.1021/acsomega.3c03376.
29. Roy, K.; Kar, S.; Ambure, P. On a simple approach for determining applicability domain of QSAR models. *Chemom. Intell. Lab. Syst.* **2015**, *145*, 22-29, https://doi.org/10.1016/j.chemolab.2015.04.013.
30. Pardridge, W.M. Drug Transport across the Blood–Brain Barrier. *J. Cereb. Blood Flow Metab.* **2012**, *32*, 1959-1972, https://doi.org/10.1038/jcbfm.2012.126.
31. Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717, https://doi.org/10.1038/srep42717.
32. Saraiva, C.; Praça, C.; Ferreira, R.; Santos, T.; Ferreira, L.; Bernardino, L. Nanoparticle-mediated brain drug delivery: Overcoming blood–brain barrier to treat neurodegenerative diseases. *J. Controlled Release* **2016**, *235*, 34-47, https://doi.org/10.1016/j.jconrel.2016.05.044.
33. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461, https://doi.org/10.1002/jcc.21334.

Publisher's Note & Disclaimer