RESEARCH PAPER

QSAR and Molecular Docking Studies of Novel 1-Hydroxonapthalene-2-carboxinilide Derivatives as Potential Anticancer Agents for Colorectal Carcinoma

Ms. Kanij Ansari[1], Dr. Sapna Jain Dabade[2*], Mrs. Madhu Singh[3]

[1]Research Scholar, Chemistry under the Faculty of Science, Sage University, Indore
Faculty of Chemistry, Gov. College, Jeerapur (Rajgarh)
Email: kanijfatima123456a@gmail.com

[2*]Associate Professor, Department of Applied Science, Institute of Engineering and Technology, Sage University, Indore (NAAC A+) Email: Sapnadabade@gmail.com

[3]Research Scholar, Chemistry under the Faculty of Science, Sage University, Indore
Assistant Professor, Gov. PG College, Gadarwara
Email: madhusinghp9@gmail.com

ABSTRACT

Identification of novel anticancer agents with high efficacy and good pharmacokinetic properties continues to be an important goal in medicinal chemistry. The present study investigated the anticancer activity of 1-hydroxynaphthalene-2-carboxanilides against Human DNA Topoisomerase IIα (TOP2A) using an integrated computer- based method, which intricate the utility of a quantitative structure–activity relationship (QSAR) model and an in-silico docking study. To ensure a comprehensive evaluation of the structure-activity data, both Partial Least Squares (PLS) regression-based and classification-based QSAR models were developed. The regression model was evaluated by the standard statistical parameters such as coefficient of determination (R^2), adjusted R^2, cross-validated coefficient (Q^2), Mean Absolute Error (MAE), and Rm^2 parameters. The classification model was evaluated by commonly accepted classification parameters such as classification accuracy, sensitivity, specificity, precision, F-value, and Matthews Correlation Coefficient (MCC). Various classification algorithms were studied ranging from linear to nonlinear machine learning. Moreover, the overall discriminatory strength of the model at different threshold levels was measured using Receiver Operating Characteristic (ROC) Curve. In this case, the Matthews correlation coefficient (MCC) of LDA model was 0.763, which shows that the predicted class and observed class have a strong correlation, with no random chance bias. Additionally, the model demonstrates high sensitivity (0.880) and specificity (0.881) in the overall classification, further validating its ability to accurately classify both active inhibitors and inactive compounds. The precision and F-value were both determined to be 0.883 and 0.880, respectively, which indicated good overall precision and low false positive/false negative rates. The MolDock and Re-rank scores of compounds 2 were −117.60 kcal/mol and −135.41 kcal/mol, respectively, which were higher than the score of the reference inhibitor etoposide. Compound 2 formed favourable hydrogen-bonding interactions with Gly1007 and Arg727, as well as extensive steric interactions with Arg672, Arg673, Pro724, Ser717, Ile715, and Glu712, suggesting a good fit within the ATPase binding pocket. By means of a combined QSAR and molecular docking analysis, Compound 2 was found to be the most promising lead candidate among those evaluated 1-hydroxynaphthalene-2-carboxanilide derivatives. The QSAR predictions and docking results provide mechanistic supports for the reported antiproliferative activity in experiments and indicate that the hydroxynaphthalene carboxanilide scaffold is a promising scaffold for designing new anticancer agents targeting the TOP2A.

Keywords: Quantitative Structure Activity Relationship, Molecular Docking, Anticancer, Matthews Correlation Coefficient, Partial Least Square

How to cite this article: Ansari K, Dabade SJ, Singh M., QSAR and Molecular Docking Studies of Novel 1-Hydroxonapthalene-2-carboxinilide Derivatives as Potential Anticancer Agents for Colorectal Carcinoma. Int J Drug Deliv Technol. 2026;16(76s): 220-234; DOI: 10.25258/ijddt.16.76s.23

Source of support: Nil

Conflict of interest: None

1. INTRODUCTION

Cancer remains a leading cause of morbidity and mortality globally, resulting in almost 20 million new cases annually and approximately 9.7 million deaths every year. Although there has been much progress in diagnosis and treatment, resistance to existing chemotherapeutic agents, lack of selectivity and systemic toxicity remain a problem. Thus, the development of novel anticancer compounds of better efficacy and safety is a big challenge for medicinal chemistry. However, computational technologies have become useful solutions to speed up lead identification and optimization, thereby decreasing the throughput time and outlay in novel drug delivery systems[1-3].

Colorectal carcinoma is prevalent forms of cancer in the world and the HCT116 human colorectal carcinoma cell line is commonly used to assess the antiproliferative activity of potential anticancer agents. 1-Hydroxynaphthalene-2-carboxanilides are interesting bioactive molecules with various pharmacological properties such as anticancer activity. previously reported a series of substituted 1-hydroxynaphthalene-2-carboxanilide derivatives with high antiproliferative activity against HCT116 cells that act in a p53-independent manner. Their conformational activity and their statistical correlation with predicted Biotargets, however, an obscure. QSAR analysis is a well-entrenched algorithmic method of correlate the molecular features with pharmacological activity which can be employed to aid lead optimization and virtual screening[4,5]. While regression-based QSAR models are used to predict continuous biological responses, classification-based models are used to classify compounds into actives and inactive, which gives complementary information for compound prioritisation, especially in the case where the range of activity in the datasets is limited[6,7] .

While QSAR is able to identify molecular features that relate to biological activity, it is not able to give information on ligand–target interactions. In addition to the advantages of QSAR, molecular docking is performed to find out the way a ligand docks, its binding strength and important intermolecular forces within the site of action receptor protein[1,8]. Human DNA Topoisomerase IIα (TOP2A) is a validated anticancer target that is critical for DNA replication and chromosome segregation, and is over-expressed in fast-growing cancers such as colorectal carcinoma. The therapeutic action of several clinically important anticancer drugs relies on the inhibition of TOP2A, including etoposide, doxorubicin and mitoxantrone[9,10].

In the current research, a combined computational approach of regression-based QSAR, classification-based QSAR and molecular docking was used to examine the structure–activity relationships of 1-hydroxynaphthalene-2-carboxanilide derivatives. The experimental compounds with antiproliferative activity against HCT116 were used to develop the QSAR model, and there were 75 such compounds. On the basis of our previously reported ADMET analysis 30 compounds with suitable pharmacokinetic properties were identified for molecular docking against Human DNA Topoisomerase IIα (PDB ID: 4FM9) with Molegro Virtual Docker (MVD). Docking was done to get an experimentally observed antiproliferative activity in a molecular context, but not necessarily to directly inhibit TOP2A. The integrated strategy is targeted to the recognition of the molecular properties that are amenable for anticancer activity and to rational design of novel TOP2A targeting anticancer agents.

2. METHODOLOGY

2.1 QSAR Analysis

Dataset Collection

A dataset consisting of 75 1-hydroxynaphthalene-2-carboxanilide derivatives with experimentally determined inhibitory activity against the HCT116 human colorectal carcinoma cell line was collected from published literature[11]. The biological activity was reported as IC50 values and subsequently converted into logarithmic units (pIC50 or log-transformed activity) to obtain a more suitable response variable for QSAR modeling. Variance-stabilizing transformation minimizes lopsidedness in the activity distribution, reduces the influence of extreme values, and facilitates the development of statistically robust predictive models. The 2-D chemical structures of all compounds were manually rendered via ChemDraw software and modified into SMILES (Simplified Molecular Input Line Entry System) notation[12,13]. The SMARTS strings and its corresponding biological activity values were organized in Microsoft Excel format, which served as the input dataset for subsequent descriptor calculation and model development.

Data Curation

Data curation is a crucial step in QSAR model development, while the quality of the outcome model is largely influenced by the quality of the data fed into it. Before descriptor calculation and model building, the dataset was carefully examined for possible inconsistencies and error[14]. The following curation steps were performed:

- Verification of chemical structures to ensure correctness of the molecular representation.
- Examination of duplicated compounds and duplicated response values.
- Identification of structural outliers that significantly differ from the remaining compounds.

- Detection of response outliers having unusual biological activity values.
- Analysis of possible activity cliffs, where structurally analogous compounds exhibit substantially different biological activities.

Since the dataset was collected from a peer-reviewed publication and consisted of chemically related analogues, no significant inconsistencies were observed during the curation process. Consequently, all compounds were retained for the modeling study.

Molecular Descriptor and Fingerprint Calculation

To quantitatively represent the chemical structures, a comprehensive set of chemical feature and bitstring representation was computed using PaDEL-Descriptor software[15]. PaDEL-Descriptor generates a wide variety of descriptor classes, including: constitutional descriptors, topological descriptors, electronic descriptors, geometrical descriptors, BCUT descriptors, autocorrelation descriptors, Burden descriptors, atom-type descriptors, functional group descriptors, and hybrid descriptors. In addition to molecular descriptors, several fingerprint representations were also generated, including widely used structural fingerprints that encode the presence or absence of predefined molecular fragments. The use of multiple descriptors and fingerprint representations enabled the exploration of different structural information and increased the possibility of identifying informative molecular features responsible for anticancer activity.

Development of QSAR Models

Both regression-based and classification-based QSAR models were developed for ensure a comprehensive evaluation of the structure-activity data. MetaQSAR software was employed for performing all the task relevant to QSAR modeling throughout the study[16].

Regression Modeling

For regression analysis, the experimentally determined log-transformed IC50 values were directly used as the dependent variable. Several linear and nonlinear machine learning algorithms were investigated during model development, including algorithms capable of capturing both simple linear relationships and more complex nonlinear relationships between molecular descriptors and biological activity. Among all evaluated algorithms, Partial Least Squares (PLS) regression produced the most reliable and interpretable predictive model and was thus selected as the final regression model. PLS is particularly suitable for QSAR studies including a relatively small number of compounds and a comparatively large number of correlated molecular descriptors. The algorithm reduces descriptor dimensionality while simultaneously maximizing the covariance between descriptor variables and biological activity.

Classification Modeling

Considering the relatively narrow activity range observed in the dataset, classification modeling was also performed. The compounds were categorized into two activity classes using a threshold value of 5.2 log units: Active compounds (activity $\geq$ 5.2) and Inactive compounds (activity $<$ 5.2). This binary classification approach transforms the prediction problem into identifying whether a compound possesses relatively higher or lower biological activity instead of predicting an exact activity value. Several classification algorithms were investigated, including both linear and nonlinear machine learning methods. Among them, Linear Discriminant Analysis (LDA) demonstrated the best balance between predictive performance and model simplicity and emerged as the final classification model.

Model Validation Strategy

The dataset consisted of only 75 compounds, which is relatively small for machine learning applications. Conventionally, QSAR studies employ an external training-test split for model validation. However, partitioning such a limited dataset into independent training and test subsets substantially reduces the number of compounds available for model training and often produces unstable performance estimates. Therefore, cross-validation was adopted as the primary validation strategy. Cross-validation repeatedly partitions the dataset into different subsets, allowing every compound to serve as both training and validation data during different iterations. This strategy enables a more robust assessment of model performance for small datasets while maximizing the utilization of all available experimental data. Regression models were evaluated using standard statistical parameters such as coefficient of determination (R^2), adjusted R^2, Q^2, MAE, and Rm^2 parameters. Classification models were evaluated using commonly accepted classification metrics including classification accuracy, sensitivity, specificity, precision, F-value, and MCC.

2.2 Molecular Docking

Although QSAR analysis provides valuable information regarding the relationship between molecular structure and biological activity, it cannot adequately explain the molecular interactions leading to the anticancer effects observed. Molecular docking is hence an essential tool in CADD to predict the binding orientation, affinity and interaction pattern of small molecules to the active site of biologically relevant proteins. Molecular docking has therefore become an indispensable tool in CADD for predicting the binding orientation, affinity, and interaction pattern of small molecules within the active site of biologically relevant proteins. By identifying H bonds, hydrophobic bonding, electrostatic contacts, and steric complementarity between ligands and receptors, docking studies provide mechanistic insight into ligand–protein

recognition and facilitate the rational optimisation of lead compounds[1,8,20-23].

Human DNA Topoisomerase IIα (TOP2A) is one of the most well studied enzymes amongst the molecular targets associated with cancer progression, due to the critical role it plays in DNA replication, chromosome segregation, transcription and cell-cycle progression. In fast-growing tumour cells, such as colorectal carcinoma, TOP2A is often over-expressed and its high expression has been associated with unfavourable clinical outcomes to and greater tumour aggressiveness. Therefore, inhibitors of TOP2A have been established as a cancer chemotherapy strategy. The clinical usage of anticancer drugs like etoposide, doxorubicin, and mitoxantrone that act as inhibitors of TOP2A demonstrates its significance as a therapeutic target in anticancer drug discovery[9,10,17].

For the present study, the crystal structure of Human DNA Topoisomerase IIα (PDB ID: 4FM9) was accessed from the Protein Data Bank and molecular docking was executed. To evaluate the binding avidity and interaction pattern of the reported 1-hydroxynaphthalene-2-carboxanilide derivatives with the catalytic binding pocket of TOP2A, docking simulations were performed using Molegro Virtual Docker (MVD Version 6.0). The docking protocol was tested by MolDock Score, Re-ranking Score, H-Bond Score, hydrogen-bond interactions and steric interaction analysis, and the reference inhibitor etoposide was used for comparative evaluation. The investigation of the docking was carried out to offer a structural and mechanistic idea about the antiproliferative activity observed against the HCT116 human colorectal carcinoma cell line and to determine the amino acid residues involved in the ligand recognition. It should be noted that the docking results are only for proposing a plausible molecular mechanism and do not necessarily mean to provide direct experimental inhibition of TOP2A.

ADMET-Based Selection of Compounds for Molecular Docking

Previously, in our published study, we investigated the Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties of the investigated 1-hydroxynaphthalene-2-carboxanilide derivatives. A total of 30 compounds with desirable ADMET profile, drug-like properties and low toxicity prediction were selected for molecular docking studies[18]. All other compounds were removed due to lacking one or more of the predefined ADMET criteria. No redoing ADMET analysis was done in the present work, the results reported in the previous work were used to prioritize the compounds for structure based molecular docking. This approach allowed docking investigations to be directed towards compounds with well-balanced antiproliferative activity and predicted favourable pharmacokinetic properties, enhancing the chances of identifying viable lead compounds.

Protein Selection and Preparation

Human DNA Topoisomerase IIα (TOP2A) 3D crystal structure was obtained from the PDB. These crystal structure used (4FM9) was chosen as it is the ATPase domain of human DNA Topoisomerase II alpha (Top2) from Homo sapiens determined by X-ray diffraction at 2.90 Å resolution. This high-resolution structure is suitable for molecular docking studies for investigation among the interactions between the ligands and proteins.

Protein preparation was done with the MVD version 6.0. All structural water molecules, buffer molecules and non-essential heteroatoms were omitted while maintaining the integrity of the active site. The structure of the protein was prepared in MVD with the default protein preparation protocol by adding the polar hydrogens, assigning bond orders. Using the cavity detection algorithm, the binding cavity corresponding to the co-crystallized ligand was automatically recognized and was selected for the docking simulation.

Ligand Preparation

The structure of 30 ADMET screened 1-hydroxynaphthalene-2-carboxanilide derivatives chosen for docking were converted to 3D molecular conformation. Hydrogen atoms were added, bond order was checked and optimized geometries were computed with molecular mechanics force field in the software Molegro Virtual Docker. The optimized structure was then imported in docking workspace as Mol2 format.

Molecular Docking Procedure

Molecular docking was carried out via the MVD Version 6.0, where utilizes the MolDock Optimizer, a guided differential evolution algorithm for conformational search and forecast of bound conformation. Ligand–protein interaction energies were estimated using the MolDock Score as the main scoring function to perform docking calculations**.**

The cavity for the docking search space was determined according to the bound structure of the PDB structure of the complexed ligand in the ATP-binding domain of Type IIA topoisomerase IIα (PDB ID: 4FM9). All parameters were remaining at their default settings except the docking simulation, which was performed using 10 independent runs, 50 individuals or population, 1500 maximum iterations and a grid resolution of 0.30Å. Each compound was subjected to multiple ligand conformations and the pose with the lowest MolDock score, along with biologically relevant interaction with the active-site residues was chosen for further analysis.

Validation of the Docking Protocol

Validity of the docking methodology was analysed by using the validated MolDock Optimizer (MVO)

implemented in the production version of MVD, and by choosing the crystallographic binding cavity of the structure 4FM9 which corresponds to the co-crystallised ligand. Each ligand was docked multiple times to guarantee reproducibility of the docked conformation.

Investigation of Molecular Docking

The visualization tools of Molegro Virtual Docker (MVD) were used to analysis of the docked ligand–protein complexes. The binding affinity was calculated using the MolDock Score, Re-Rank Score and H-Bond Score, with lesser negative scores representing better predicted ligand-protein interactions. The number of hydrogen bonds, the number of hydrogen-bonding amino acid residues, steric interactions and the orientation of the ligands in the ATPase binding pocket of Human DNA Topoisomerase IIα (PDB ID: 4FM9) were also investigated. The docking poses were compared with the binding conformation of the reference inhibitor etoposide to the active Center to identify catalytic site residues for interaction. In cases of compounds with multiple docking conformations, the pose with the lowest overall docking energy and the most biologically relevant interaction pattern was chosen. Evaluated the docking poses via the MolDock Score, Re-rank Score, H-Bond Score and the preservation of interactions with catalytically important amino acid residues. Specific attention was focused on the interaction with the residues Arg672, Arg673, Arg727, Gly1007, Pro724, Met699 and Glu712 that are involved in ligand binding to the ATPase domain. A clinically approved TOP2A inhibitor, etoposide, was chosen as a reference ligand for comparison of docking scores and docking interactions.

The unification of regression and classification-based QSAR and molecular docking provides a comprehensive computational framework for understanding the structure–activity relationships of 1-hydroxynaphthalene-2-carboxanilide derivatives. The molecular descriptors that govern anticancer activity are derived in QSAR while docking analysis provides structural information on the interaction of ligands with proteins, which is valuable for rational design and optimization of novel TOP2A-targeted anticancer agents.

3. RESULTS AND DISCUSSION

Dataset Characteristics

The final curated dataset contained 75 structurally related 1-hydroxynaphthalene-2-carboxanilide derivatives with experimentally determined activity against the HCT116 cancer cell line. One important characteristic of the dataset was the relatively small activity range of only 1.88 log units. Such a narrow response range presents a significant challenge for regression-based QSAR modeling because statistical measures based on explained variance become highly sensitive to even small prediction errors[19]. Although the dataset contained chemically meaningful structural diversity through substitution patterns around the common scaffold, the limited variation in biological activity restricted the maximum achievable regression performance.

Regression QSAR Model (Partial Least Squares)

Multiple descriptor sets, fingerprint representations, and machine learning algorithms were investigated during model development. Among all evaluated approaches, Partial Least Squares (PLS) generated the most satisfactory regression model. The final optimized PLS model was constructed using the curated training set of 75 compounds with an optimum selection of 4 latent components. The quantitative relationship between the 8 selected molecular descriptors and the biological activity (pIC50) against the HCT116 cell line is defined by the following linear equation:

$$\textit{ACTIVITY} = 6.343 - 0.529(\textit{PubchemFP25}) + 0.964(\textit{SM1\_Dze}) - 0.285(\textit{PubchemFP758}) - 0.65(\textit{IC4}) + 4.281(\textit{VE1\_D}) + 0.525(\textit{PubchemFP643}) + 0.183(\textit{APC2D4\_N\_Cl}) - 0.261(\textit{APC2D4\_F\_Cl})$$

The internal validation and fitness parameters for this final regression model are tabulated in Table 1.

Table 1: Model Validation and Fitness parameters

Validation Metric	Value
Number of Training Datapoints (N)	75
Number of Selected Features	8
Optimum Number of Components	4
Coefficient of Determination (R^2)	0.656
Adjusted Coefficient of Determination (R^2 Adjusted)	0.614
Cross-Validated Coefficient (Q^2 LOO)	0.563

Scaled Average Rm^2 (LOO)	0.442
Scaled Delta Rm^2 (LOO)	0.185
Mean Absolute Error (MAE-LOO)	0.234

The predictive capability of the optimized PLS regression model is visually evaluated by comparing the experimentally observed biological activities against the model-predicted values in the Response Scatter Plot (Figure 1).

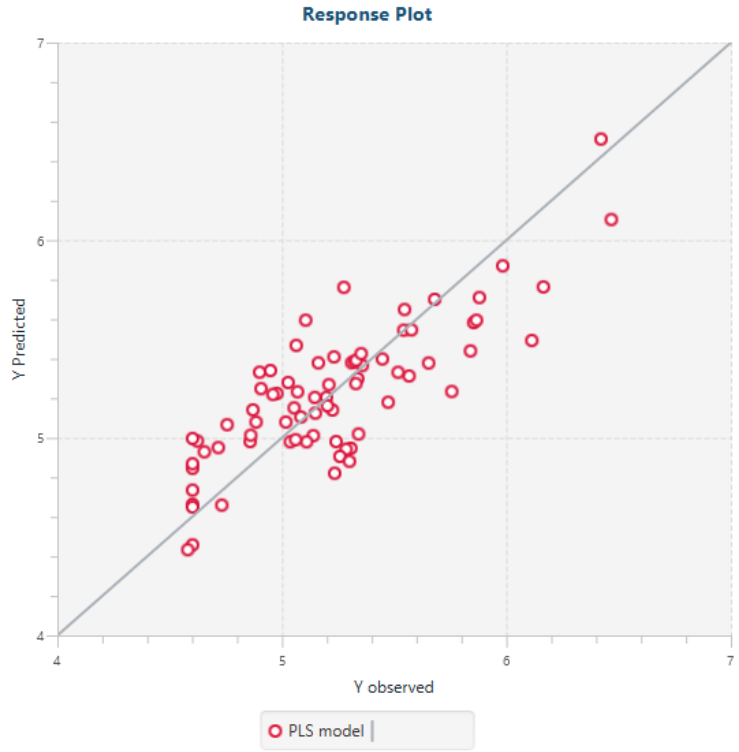


Figure 1: Response Plot showing Experimentally Observed vs. PLS-Predicted Biological Activity

As shown in Figure 1, the datapoints cluster closely around the diagonal line of ideal fit (y = x) across the entire modeled spectrum. This close proximity graphically substantiates the low Leave-One-Out Mean Absolute Error (MAE-LOO) of 0.234 log units, illustrating that the model's actual prediction errors are remarkably small on a practical scale. However, the plot also clearly exposes the statistical challenge discussed previously: the total distribution of the response variable is tightly confined between approximately 4.5 and 6.5 log units, presenting a narrow total activity range of only 1.88 log units. Because the variance of the underlying dataset is so small, minor vertical deviations from the line of ideal fit represent a large fraction of the total vertical spread. This narrow data distribution is the direct mathematical reason the model yields a moderate R^2 of 0.656 and a Q^2-LOO of 0.563, leading to an automated software classification of 'BAD' prediction quality based strictly on standard threshold criteria. Visually, it is evident that the model is neither underfitted nor wildly inaccurate; rather, it is highly precise within this localized therapeutic window. The response plot firmly confirms that error-based metrics (MAE) offer a far more scientifically honest appraisal of this model's utility than traditional variance-based correlation coefficients.

To rigorously test the statistical stability of the PLS regression model and rule out any chance correlations, a Y-randomization test was carried out by randomly shuffling the biological activity vector 50 times while keeping the descriptor matrix intact. The results are illustrated in Figure 2.

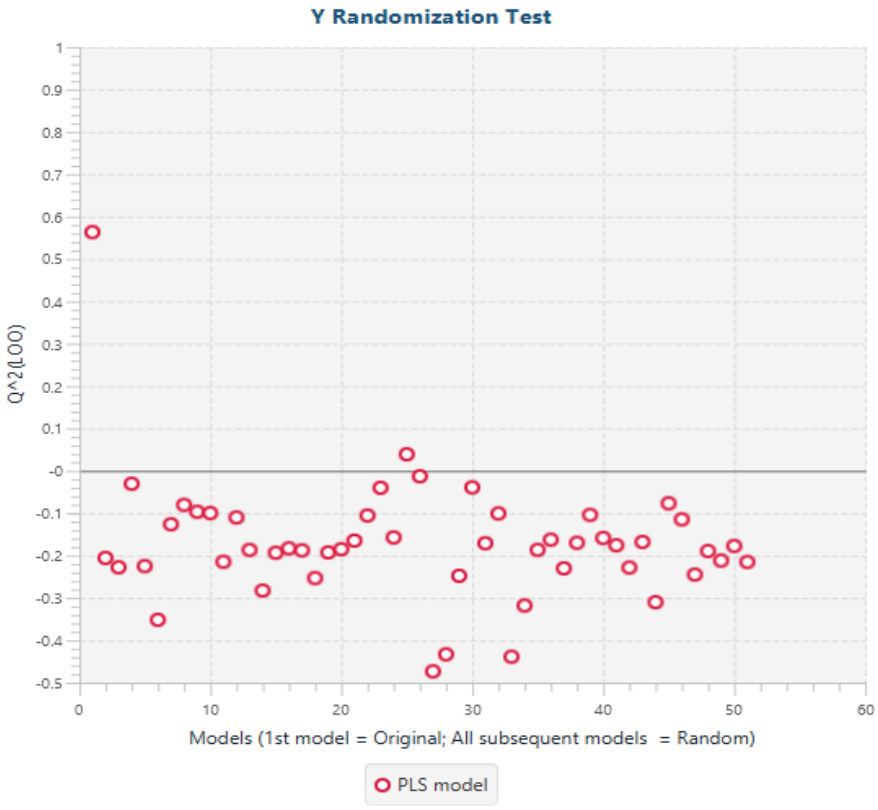


Figure 2: Y-Randomization Test Scatter Plot for the PLS Regression Model showing Q2(LOO) values

As illustrated in Figure 2, the original PLS model exhibits a robust cross-validated Q^2(LOO) of 0.563 (plotted clearly on the far left). In contrast, all 50 randomized models generated from the shuffled datasets produced substantially lower and predominantly negative Q2 values, falling within a range between approximately 0.05 and -0.47. The wide, distinct gap between the true model and the random permutations firmly demonstrates that the PLS model is highly stable, free from random chance correlations, and possesses genuine structural predictability.

Classification QSAR Model (Linear Discriminant Analysis)

To overcome the limitations imposed by the narrow activity range, binary classification modeling was performed using an activity threshold of 5.2 log units. Among the evaluated machine learning algorithms, Linear Discriminant Analysis (LDA) demonstrated superior classification performance. The optimized LDA model was developed through a curated training set of 75 compounds utilizing 8 selected molecular features. Internal cross-validation yielded robust predictive metrics, which are summarized in Table 2.

Table 2: Cross-validation Metrics of LDA model

Metric (Weighted)	Value
Number of Features	8
Matthews Correlation Coefficient (MCC)	0.763
Sensitivity (TPR)	0.880
Specificity (TNR)	0.881
Precision	0.883
F-value	0.880

The final LDA model attained a high Matthews Correlation Coefficient (MCC) of 0.763, reflecting a strong correlation between the predicted and observed classes free from random chance bias. The overall classification accuracy is further supported by a balanced sensitivity of 0.880 and a specificity of 0.881, confirming those the model equally identifies both active inhibitors and inactive compounds with high fidelity. The precision and F-value were both calculated at 0.883 and 0.880, respectively, demonstrating excellent overall reproducibility and minimal false-positive/false-negative rates. The high validation statistics achieved indicate that the 8 selected molecular descriptors successfully captured the critical structural differences associated with higher and lower biological activity.

To evaluate the statistical significance of the LDA classification model, a Y-randomization test was executed across 50 permutations, shuffling the active/inactive class labels. The scatter plot mapping the Matthews Correlation Coefficient (MCC) across these random runs is shown in Figure 3.

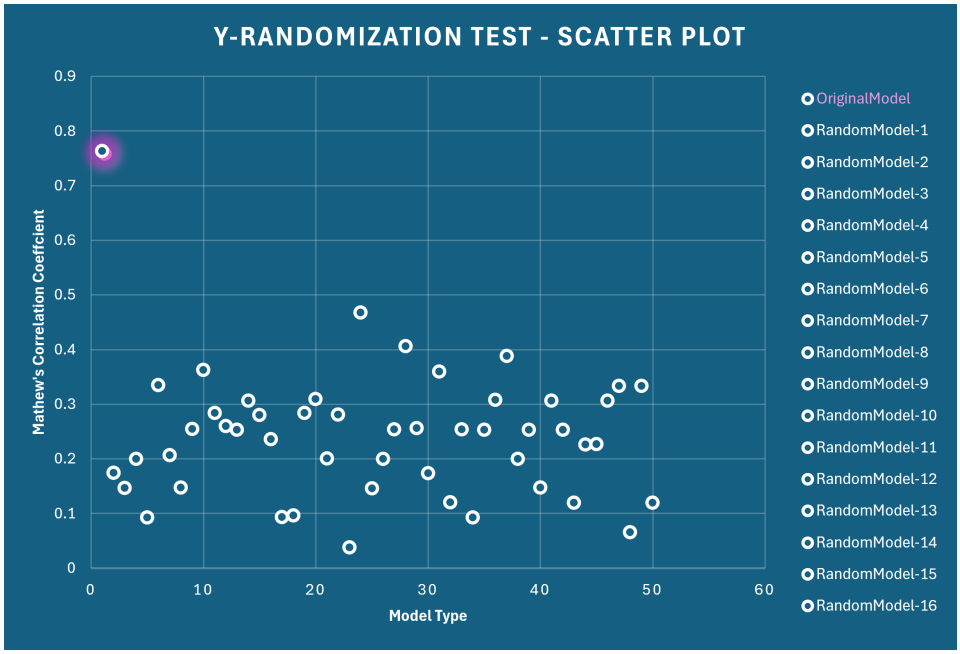


Figure 3: Y-Randomization Test Scatter Plot for the LDA Classification Model showing MCC values

As shown in Figure 3, the original LDA model exhibits a distinctively superior MCC of 0.763. In stark contrast, all 50 randomized models generated from the shuffled datasets produced drastically lower MCC values, remaining well below 0.50 (ranging between approximately 0.04 and 0.47). Because the performance of the true model sits entirely outside the distribution of the random permutations, this confirms that the original classification model's predictive power is highly robust, statistically significant, and firmly rooted in true structure-activity relationships rather than mere chance or overfitting.

Furthermore, the overall discriminatory power of the model across various threshold settings was evaluated using a Receiver Operating Characteristic (ROC) curve, shown in Figure 4.

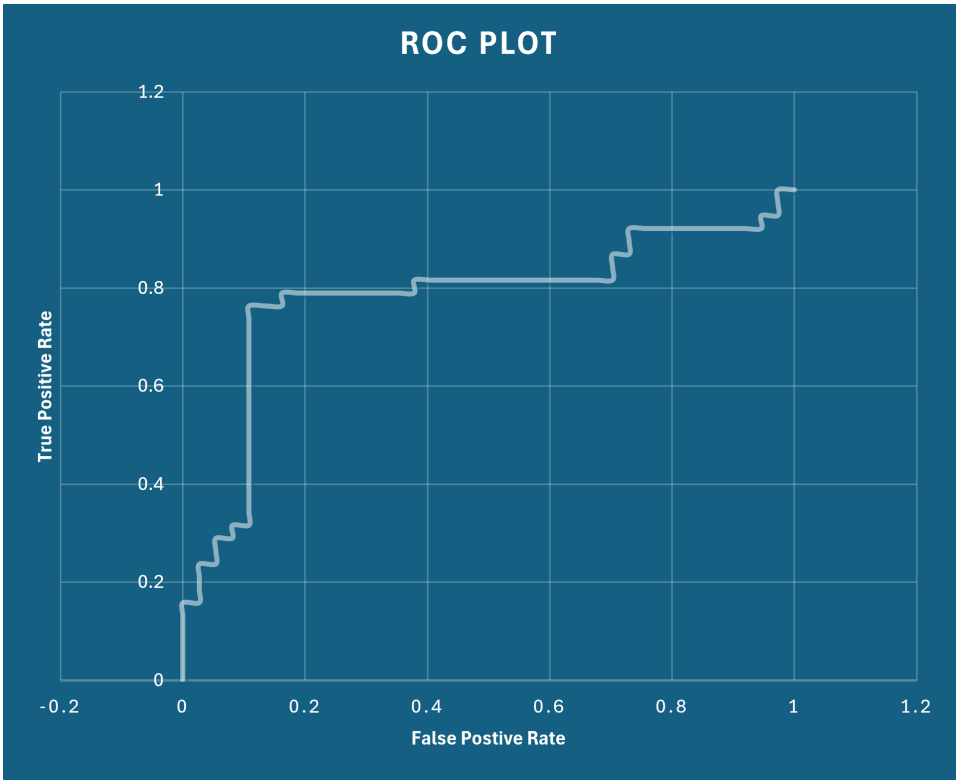


Figure 4: ROC Plot for the LDA Classification Model

As shown in Figure 4, the curve climbs steeply near the y-axis, reaching a True Positive Rate (Sensitivity) of nearly 0.80 while maintaining an exceptionally low False Positive Rate of approximately 0.11. This sharp initial ascent reflects the model's strong capacity to correctly classify active 1-hydroxynaphthalene-2-carboxanilide derivatives while minimizing the risk of misclassifying inactive analogues as active. The substantial area captured under the curve (AUC-ROC) confirms that the model possesses excellent overall discriminatory power rather than reliance on a single arbitrary threshold setting.

Comparison of Regression and Classification Models

The comparative analysis demonstrates the importance of selecting an appropriate modeling strategy based on dataset characteristics. Regression models generally require a sufficiently wide activity distribution to accurately learn continuous quantitative relationships between molecular structure and biological activity. In this study, the narrow activity range limited the maximum achievable R^2 values despite relatively low prediction errors (MAE = 0.234). Conversely, classification modeling simplified the prediction problem by categorizing compounds into biologically meaningful active and inactive classes. This transformation reduced the influence of small experimental variations and resulted in substantially improved predictive performance (MCC = 0.763, Accuracy ~ 88%). The findings indicate that classification modeling represents a more suitable approach for datasets exhibiting limited activity variation. Nevertheless, the regression model still provides useful quantitative information, particularly when interpreted using error-based evaluation metrics rather than relying solely on variance-based statistics. Overall, both modeling approaches complement each other. The regression model provides approximate quantitative activity predictions, whereas the classification model offers reliable discrimination between active and inactive compounds. Together, these models provide valuable computational tools for prioritizing new 1-hydroxynaphthalene-2-carboxanilide derivatives for future synthesis and biological evaluation against HCT116 colorectal cancer cells.

Evaluation of Molecular Docking Results

The possible molecular basis of the reported antiproliferative activity was studied by molecular docking in Molegro Virtual Docker (MVD) of the compounds against Human DNA Topoisomerase IIα (TOP2A; PDB ID: 4FM9). The MolDock Score, Re-rank Score, H-Bond Score, hydrogen-bond interactions, and steric interactions within the ATPase binding pocket were used to evaluate docking poses. The clinically proven TOP2A inhibitor etoposide was chosen as a reference ligand. The docking results of the investigated compounds with the reference compound, etoposide, are compiled in Table 3 along with significant observations are depicted in Figure 5 and Figure 6.

Table 3: Molecular Docking output of 1-hydroxynaphthalene-2-carboxanilides derivatives in the Topoisomerase II (4FM9) Binding Pocket Obtained Using MVD

Compound No.	**Rank score** kcal/mol	**Mole Dock Score** kcal/mol	**No of H bonds**	**Amino Acid comprehend in H Bonding**	**Amino Acid include in Steric Interaction**	**Hydrogen bond score**
Comp 3	-135.41	-117.60	4	Gly1007, Arg727	Arg672, Arg673, Gly1007, Pro724, Arg727, Ser717, Ile715, Glu712	-7.17
Comp 4 (Pose 1)	-132.02	-114.17	3	Arg672, Arg673, Glu712	Arg672, Arg673, Met699	-5.19
Comp 4 (Pose 2)	-131.02	-112.92	7	Ser591, Leu685, Leu592, Ser547, Lys550	Pro593, Gln544, Asp543, Gln542, Lys701, Asp683, Leu705	-6.67
Control Drug Etoposide	-129.68	-107.88	6	Met699, Arg672, Arg673, Glu837. Lys728	Arg672, Arg673, Pro724, Gly1007, Lys728	-3.96

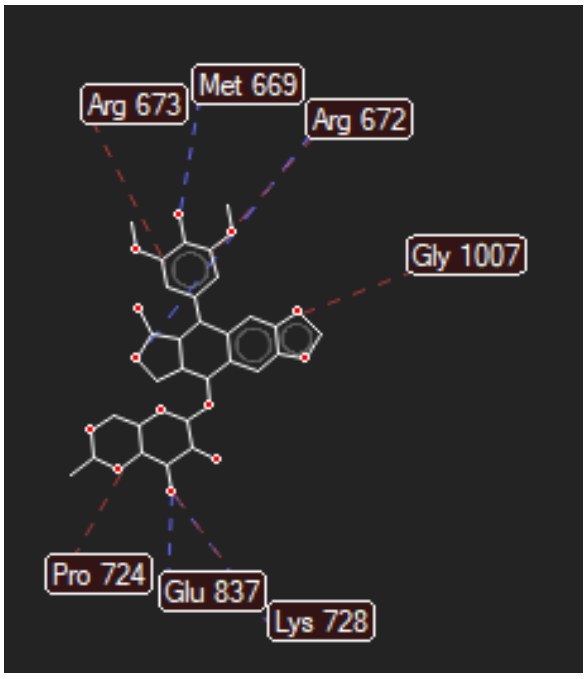
As illustrated in the table, among the compounds investigated, the compound 3 was having the greatest predicted binding affinity for a MolDock Score of −117.60 kcal/mol compared to the corresponding Re-rank Score of etoposides (−129.68 kcal/mol) which was higher than the Re-rank Score of compounds 3 (−135.41 kcal/mol). Also, the compound showed the maximum H-Bond Score (−7.17 kcal/mol) which revealed that it has the highest hydrogen bond stabilization ability. Hydrogen-bond interactions were seen with Gly1007 and Arg727, with additional steric interactions with Arg672, Arg673, Pro724, Ser717, Ile715 and Glu712 which further stabilized the ligand in the active-site cavity. The interplay of these interactions with residues that bind to etoposide indicate that Compound 2 has a biologically relevant binding orientation, and is likely to be a good binder of the catalytic region of TOP2A. There were two energetically favourable docking conformations for Compound 4. Pose 1 gave a MolDock Score of −114.17 kcal/mol and Re-rank Score of −132.02 kcal/mol with H-bond with Arg672, Arg673, Glu712 and hydrophobic interaction with Met699. These residues are also found in the interactions with the reference inhibitor, thus the pose of 1 is very similar to the known one of TOP2A inhibitors.

Pose 2 on the other hand had a slightly lower overall binding affinity (MolDock Score = −112.92 kcal/mol; Re-rank Score = −131.02 kcal/mol) yet more extensive hydrogen-bonding network (7 hydrogen bonds) with Ser591, Leu592, Leu685, Ser547 and Lys550, which made its H-Bond Score (−6.67 kcal/mol) more favourable than Pose 1. This indicates an alternative stabilized orientation in the binding cavity with Pose 2. At the same time, Pose 1 showed the better overall docking energy and captured the same interactions with catalytically important residues as observed in etoposide, thus making Pose 1 the preferred and biologically more relevant binding pose of Compound 4.

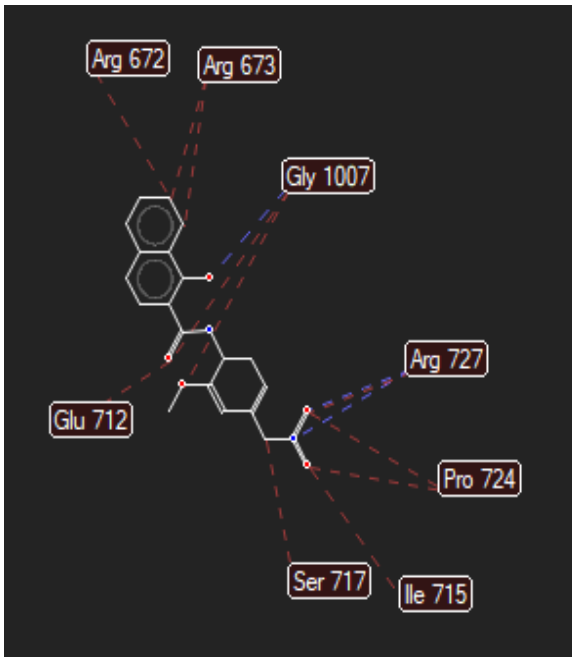
In general, the docking results suggest that the synthesized 1-hydroxynaphthalene-2-carboxanilide derivatives have more binding affinities towards TOP2A than the reference inhibitor etoposide. Among them, Compound 2 was the most promising candidate that showed high docking scores, along with interactions via important key active site residue.

The docking results agree with the QSAR results, which corroborate the significance of the hydroxynaphthalene carboxanilide scaffold for the anticancer activity and indicate that TOP2A could be a reasonable molecular target for the antiproliferative activity of the hydroxynaphthalene carboxanilide derivatives reported experimentally against the HCT116 cell line. However, biochemical enzyme inhibition assays and molecular dynamics simulations are needed at prove that suggested mode of action.

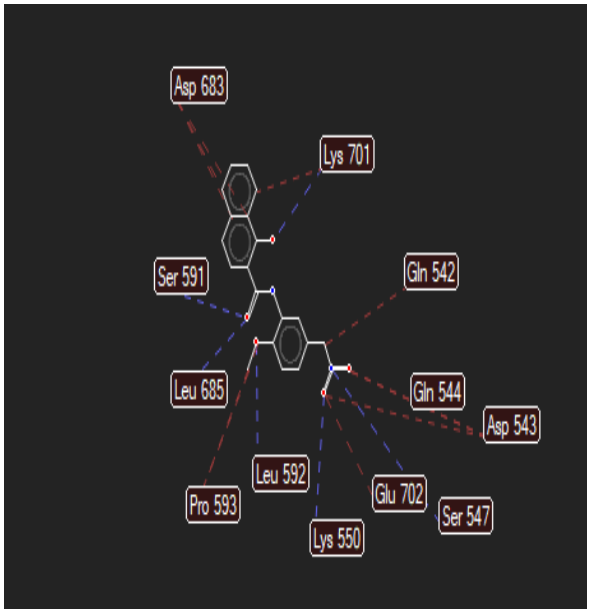
Control Drug (Etoposide)



Compound 3



Compound 4 (Pose1)



Compound 4 (Pose2)

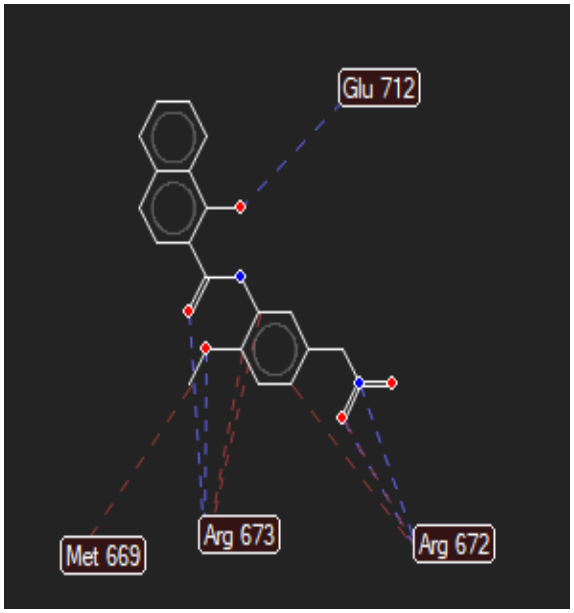
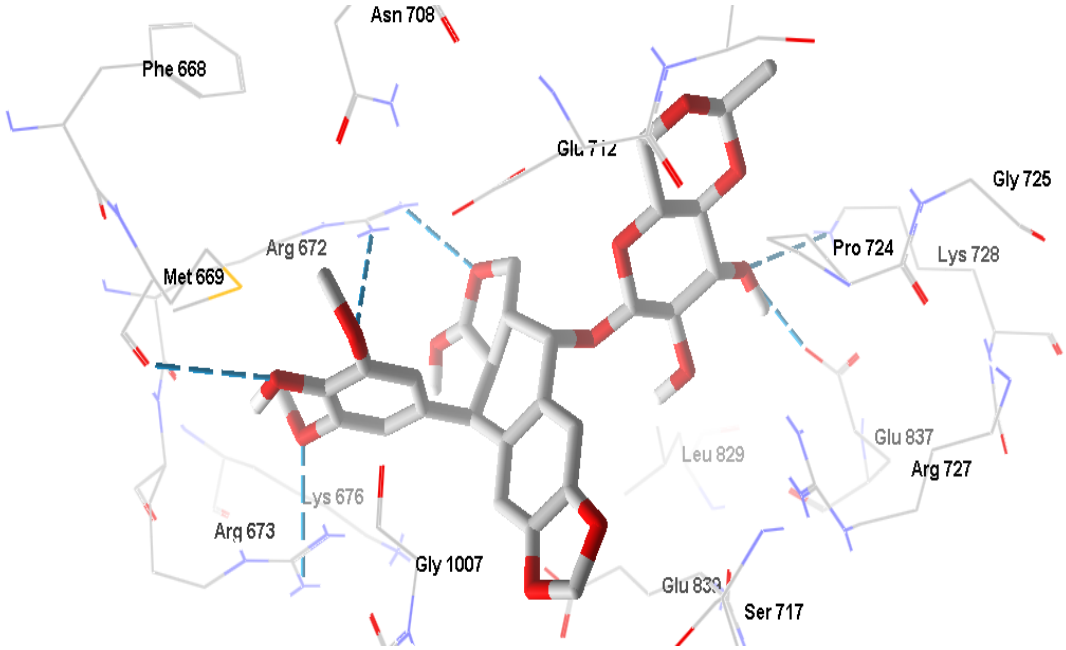


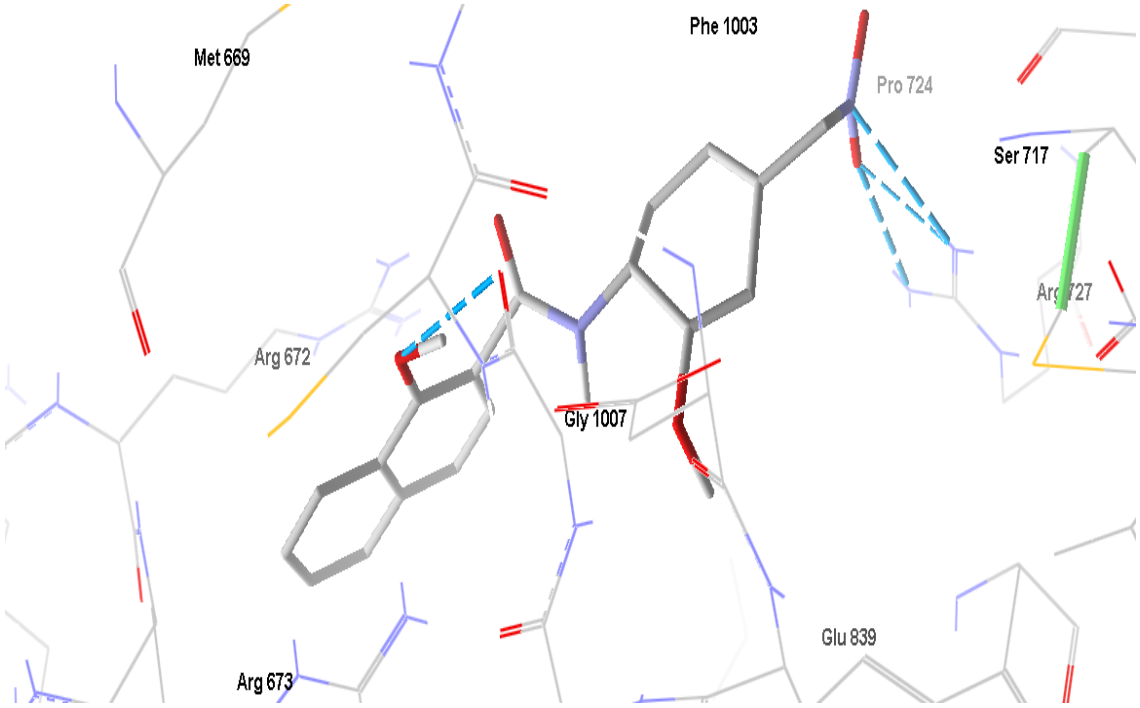
Figure 5: Protein–ligand interaction diagrams of **Control Drug (Etoposide)**

Compound 3, Compound 4 (Pose 1), Compound 4 (Pose 2), and the reference inhibitor etoposide docked into the ATPase binding site of Human DNA Topoisomerase IIα (PDB ID: 4FM9). H-bond interactions are denoted by blue dashed lines, although steric hindrance is shown by Red dashed lines

Control Drug (Etoposide)



Compound 3



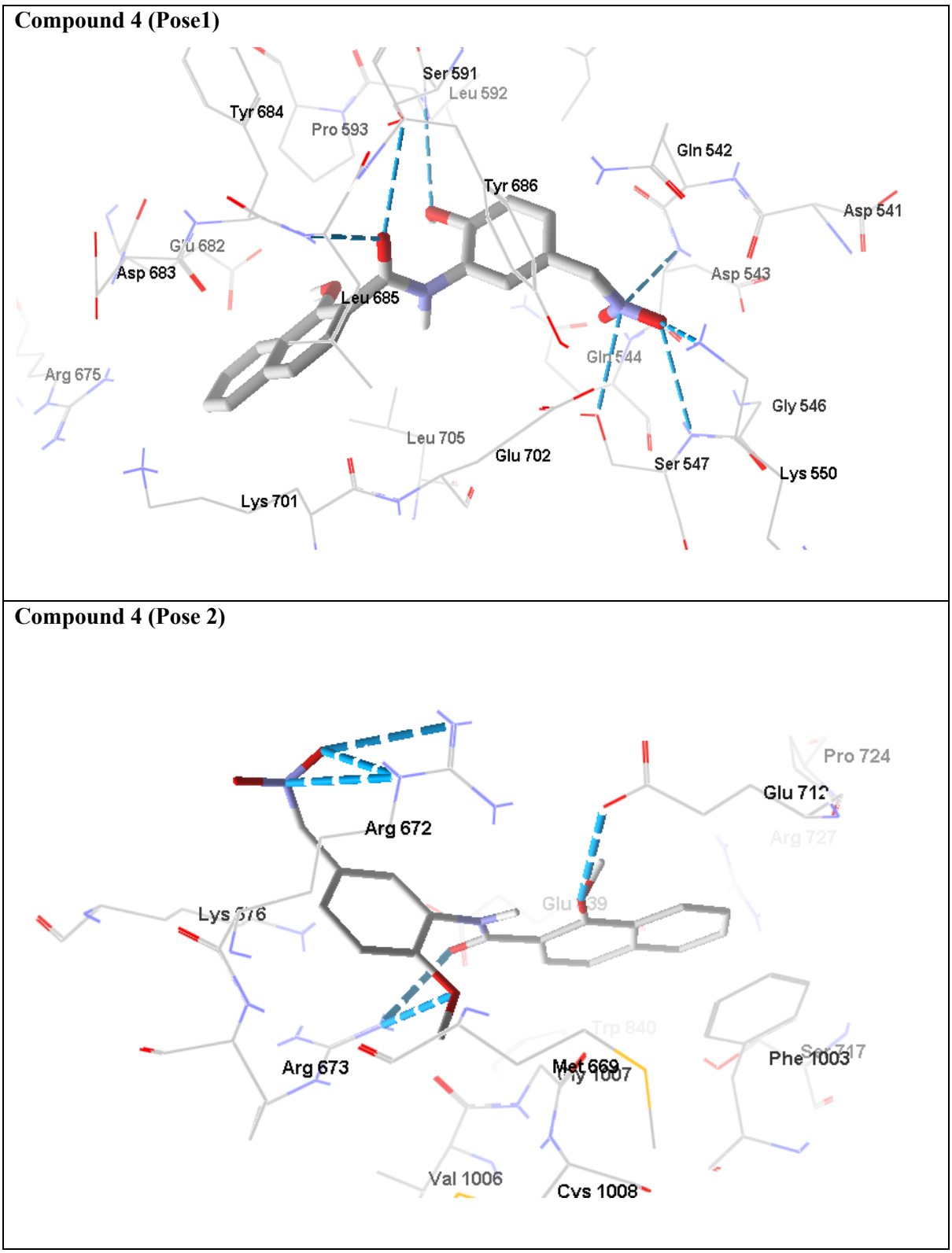


Figure 6: 3D protein–ligand H- bond interaction (Blue dashed lines) diagrams of Compound 3, Compound 4 (Pose 1), Compound 4 (Pose 2), and the reference inhibitor etoposide docked into the ATPase binding site of Human DNA Topoisomerase IIα (PDB ID: 4FM9).

4. CONCLUSION

Our present study successfully combined regression approach and classification approach based QSAR modeling with molecular docking approach for the investigation of anticancer potential of 1-hydroxynaphthalene-2-carboxanilide derivatives against Human DNA Topoisomerase IIα (TOP2A). The regression model gave quantitative insights into the structure–activity relationship while the classification model showed excellent performance in predicting the investigated dataset with an MCC of 0.763, a classification accuracy of 88.1%, and ROC-AUC of 0.80. The results show that classification modeling is applicable to datasets with a relatively low range of biological activities, whereas regression modeling can be used for datasets with quantitative activities for which error-based evaluation methods are useful.

Based on our previously reported ADMET study, 30 compounds that showed good ADMET applied to prioritised for molecular docking studies. Compound 2 become identified as one of the most promising hits with the lowest MolDock Score (−117.60 kcal/mol) and Re-rank Score (−135.41 kcal/mol) with good hydrogen-bonding and steric interactions in the ATPase binding pocket in the docked complex. Mechanistic support for the experimentally reported antiproliferative activity of this scaffold was obtained from the agreement obtained between the QSAR predictions and the docking results.

Overall, this integrated computational workflow demonstrates that 1-hydroxynaphthalene-2-carboxanilide analogues signify promising lead structures for the development of TOP2A-targeted anticancer agents.

REFERENCES

1. Ferreira LG, dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. Molecules. 2015;20(7):13384-13421. DOI: 10.3390/molecules200713384.

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209-249. DOI: 10.3322/caac.21660.

3. Ansari K, Jain DS, Madhu S. Advances in CADD driven oncology: From lead identification to clinical success. Research Journal of Biotechnology. 2026;21(8):279-289. DOI: 10.25303/218rjbt2790289.

4. Organisation for Economic Co-operation and Development (OECD). Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. OECD Series on Testing and Assessment No. 69. Paris: OECD Publishing; 2007. 154 p. DOI: 10.1787/9789264085442-en.

5. Gramatica P. Principles of QSAR models validation: Internal and external. QSAR Comb Sci. 2007;26(5):694-701. DOI: 10.1002/qsar.200610151.

6. Gramatica P. On the development and validation of QSAR models. Methods Mol Biol. 2013; 930:499-526. DOI: 10.1007/978-1-62703-059-5_21.

7. Roy K, Kar S, Das RN. Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. London: Academic Press; 2015. 479 p. DOI: 10.1016/C2014-0-00286-9.

8. Thomsen R, Christensen MH. MolDock: A new technique for high-accuracy molecular docking. J Med Chem. 2006;49(11):3315-3321. DOI: 10.1021/jm051197e.

9. Nitiss JL. DNA topoisomerase II and its growing repertoire of biological functions. Nat Rev Cancer. 2009;9(5):327-337. DOI: 10.1038/nrc2608.

10. Pommier Y, Sun Y, Huang SY, Nitiss JL. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. Nat Rev Mol Cell Biol. 2016;17(11):703-721. DOI: 10.1038/nrm.2016.111.

11. Spaczyńska E, Mrozek-Wilczkiewicz A, Malarz K, Kos J, Gonec T, Oravec M, Gawecki R, Bak A, Dohanosova J, Kapustikova I, Liptaj T, Jampilek J, Musiol R. Design and synthesis of anticancer 1-hydroxynaphthalene-2-carboxanilides with a p53 independent mechanism of action. Sci Rep. 2019;9(1):6387. DOI: 10.1038/s41598-019-42595-y.

12. ChemDraw Ultra 8.0.3. Cambridge Soft Corporation, Cambridge, MA, USA.

13. Chem3D Ultra 8.0.3. Cambridge Soft Corporation, Cambridge, MA, USA.

14. Fourches D, Muratov E, Tropsha A. Trust, but verify II: A practical guide to chemogenomics data curation. J Chem Inf Model. 2016;56(7):1243-1252. DOI: 10.1021/acs.jcim.6b00129.

15. Yap CW. PaDEL-descriptor: An open-source software to calculate molecular descriptors and fingerprints. J Comput Chem. 2011;32(7):1466-1474. DOI: 10.1002/jcc.21707.

16. Ambure P, Serrano-Candelas E, Bhat-Ambure J, Gozalbes R. MetaQSAR: A comprehensive tool for automated QSAR modeling. In: Cheminformatic Modeling and Data Gap Filling for a Green and Sustainable Environment. Amsterdam: Elsevier; 2026. p. 993-1019. DOI: 10.1016/B978-0-443-36474-7.00003-X.

17. Okoro CO, Fatoki TH. A mini review of novel topoisomerase II inhibitors as future anticancer agents. Int J Mol Sci. 2023;24(3):2532. DOI: 10.3390/ijms24032532.

18. Ansari MK, Dabade SJ, Singh M. In silico pharmacokinetics and toxicological studies of 1-hydroxynaphthalene-2-carboxanilide derivatives as potential anticancer drug candidates. The Bioscan. 2026;21(3):139-153. DOI: 10.63001/tbs.2026.v21. i03.

19. Roy K, Das RN, Ambure P, Aher RB. Be aware of error measures: Further studies on validation of predictive QSAR models. Chemom Intell Lab Syst. 2016; 152:18-33. DOI: 10.1016/j.chemolab.2016.01.008.

20. Dabade SJ, Mandloi D, Bajaj A. Molecular docking and QSAR studies of coumarin derivatives as NMT inhibitors: Simple structural features as potential modulators of antifungal activity. Lett Drug Des Discov. 2020;17(10):1293-1308. DOI: 10.2174/1570180817999200617105711.

21. Dabade SJ, Mandloi D, Bajaj AV, Dhingra N. GA-SMLR-based QSAR modeling and molecular docking studies of bisamidine derivatives as NMT inhibitors. Int J Quant Struct Prop Relat. 2020;5(4):1-31. DOI: 10.4018/IJQSPR.2020100101.

22. Dabade SJ, Mandloi D, Bajaj AV, Thakur A. In silico evaluation of inhibitory potential of novel triazole derivatives against therapeutic target myristoyl-CoA: Protein N-myristoyltransferase (NMT) of Candida albicans. Netw Model Anal Health Inform Bioinform. 2020;9(1):60. DOI: 10.1007/s13721-020-00265-4.

23. Madhu S, Jain DS, Ansari K. Novel triazole analogues as SGL1 inhibitors of Cryptococcus neoformans: An integrative QSAR-molecular docking study. Res J Chem Environ. 2026;30(5):119-134. DOI: 10.25303/305rjce1190134.