Current Chemistry Letters 13 (2024) 451–466

Contents lists available at GrowingScience

Current Chemistry Letters

homepage: www.GrowingScience.com

Inhibition activity of triazoles as a new family for the inhibition of the Indoleamine 2,3-dioxygenase 1 IDO1 protein using 2D-QSAR approach

Khadija Zaki[a*], Fatimazahra Fakir[a], Abdelouahid Sbai[a], Hamid Maghat[a], Mohammed Bouachrine[a,b] and Tahar Lakhlifi[a]

[a]*Molecular Chemistry and Natural Substances Laboratory, Moulay Ismail University, Faculty of Science, bp 11201 Meknes, Morocco*
[b]*EST Khenifra, Sultan Moulay Sliman University, Beni-Mellal, Morocco*

C H R O N I C L E

Article history:
Received October 4, 2023
Received in revised form
January 10, 2024
Accepted March 21, 2024
Available online
March 21, 2024

Keywords:
IDO1
2D QSAR
PCA
MLR
MNLR

A B S T R A C T

Protein IDO1 (indoleamine 2,3-dioxygenase) occupies a critical position in the regulation of the immune system and is involved in cancer progression and the development of immune diseases. Being a therapeutic target for such critical diseases, we aimed to investigate the IDO1 inhibition activity of thirty-nine triazole derivatives using a quantitative structure-activity relationship. The dataset was under principal component analysis, multiple linear regression, and multiple non-linear regression from which two models were generated. The best 2D-QSAR model was generated using linear regression, demonstrating a determination coefficient of R^2=0.680, a good acceptable internal cross-validated coefficient of R^2_{cv}=0.700, an error of MSE=0.074, and a good predictive potential of R^2_{test}=0.809. The QSAR model was further investigated using the applicability domain, which showed that all molecules were within the applicability domain, hence the absence of an outlier. Overall, the obtained results provide a reliable and highly predictive model for the design and prediction of new IDO1 inhibitors thereby influencing cancer progression and autoimmune disease development.

1. Introduction

Cancer and immune diseases pose significant medical challenges due to their complexity and impact on human health. The discovery of new drugs to treat these conditions is a lengthy and costly process. However, identifying specific therapeutic targets can provide new treatment perspectives. In this context, the protein IDO1 (indoleamine 2,3-dioxygenase) plays a crucial role.[1,2] IDO1 occupies a critical position in the regulation of the immune system and is involved in cancer progression as well as the development of immune diseases.[1,3] Therefore, it represents a promising target for drug development. Nevertheless, the search for molecules capable of selectively inhibiting the activity of IDO1 faces considerable challenges due to the complexity of the molecular interactions involved.

Computational approaches, like QSAR, have emerged as promising techniques to help and direct drug designs.[4] This approach is based on statistical models that establish quantitative relationships between the chemical structure of molecules and their biological activity. By combining experimental data with relevant chemical descriptors, QSAR models can predict the inhibitory activity of molecules against IDO1, thus facilitating a faster and more accurate selection of promising compounds for drug development.[4,5]

* Corresponding author
E-mail address kha.zaki@edu.umi.ac.ma (K. Zaki)

doi: 10.5267/j.ccl.2024.3.004

The QSAR approach offers the advantage of being faster and more cost-effective than traditional drug discovery methods. It enables the optimization of the design of new molecules with improved inhibitory properties, thereby reducing the time and resources required for the drug development process. Therefore, the application of the QSAR approach in drug research targeting IDO1 provides a promising opportunity to accelerate the development of more effective treatments against cancer and immune diseases.

In this regard, the objective of this work is to develop robust and reliable QSAR models, representing the steps of a 2D-QSAR study, to predict and explain the inhibitory activity of the enzyme IDO1 for a series of triazole derivative molecules.

2. Results and Discussion

2.1. Principle Component Analysis (PCA) Results

PCA analysis enables the generation of a correlation matrix (**Table 1**) where each correlation coefficient corresponds to the variables in the same row and column. Every value higher than |0.5| is considered high, meaning the correspondent descriptors give the same information, either positively or negatively depending on the sign of the coefficient. The molecular weight (MW) descriptor was found to be positively correlated with four other descriptors, namely molar volume (MV), molar refractivity (MR), parachor (Pc), and polarizability (ae), meaning they all provided the same information about the variation of the activity in the studied series. In order to avoid any redundancy in the model, we only retained molecular weight (MW) from that group for further analysis. Similarly, the refractivity index (n) and surface tension (γ) are highly correlated. Therefore, we conserved only one descriptor, the refractivity index, for further investigation. As a result, the refractivity index, molecular weight, density, and logP were the descriptors chosen for a regression investigation.

Table 1. Correlation matrix.

Variables	*pIC50*	*MW*	*MR*	*MV*	*Pc*	*N*	*γ*	*D*	*αe*	*logP*
pIC50	1									
MW	0.635	1								
MR	0.710	0.905	1							
MV	0.706	0.786	0.95	1						
Pc	0.718	0.856	0.987	0.986	1					
N	0.017	0.409	0.201	-0.114	0.047	1				
γ	0.047	0.424	0.23	-0.078	0.088	0.971	1			
D	0.108	0.609	0.248	-0.006	0.124	0.814	0.795	1		
αe	0.71	0.905	1	0.950	0.987	0.201	0.229	0.248	1	
logP	0.367	0.502	0.488	0.431	0.457	0.219	0.174	0.270	0.488	1

2.2. Multiple Linear Regression MLR

The linear relationship between the inhibition activity and the chosen variables was explored using XLSTAT2023 software. After several attempts, a model was generated revealing a linear regression between the molecular weight, density, and the coefficient of partition logP. The equation is as follows:

$$pIC_{50} = 5.218 + 6.205\ MW - 1.556\ d - 0.107\ LogP$$

The regression equation between the inhibition activity of IDO1 and the 2D studied descriptors includes variables with positive coefficients and others with negative ones. The molecular weight has a positive coefficient related to it, which means an increase in the molecular weight will positively influence the inhibition activity of IDO1, while a decrease will induce a negative influence. As for the coefficient of partition logP and the density, they both have negative coefficients meaning a reverse correlation with the activity thus, an increase in any/both of those descriptors will lead to a decrease in the inhibition activity and inversely.

Furthermore, the VIF values (Variance Inflation Factor) for the three descriptors fall within the range of 1 to 5, indicating a weak correlation among these descriptors.[6,7]

Table 2. The MLR model parameters and the thresholds between brackets.

R^2	0.68 (> 0.6)
R	0.764 (> 0.5)
R^2 adj	0.58 (> 0.5)
R^2 cv	0.70 (> 0.5)
MCE	0.074 (low value)
RMCE	0.272 (low value)
F	12.647
R^2test	0.809 (> 0.6)

The model explained 68.0% of the variation of the activity, with a 0.70 cross-validation coefficient, an 80.9% predictive ability, a low mean square error of 0.074, and an F value of 12.647. These results follow the acceptance thresholds reported by previous studies.[8–10] **Fig. 1** is a graphical representation of the plot of the pIC50= f(pred(pIC50)). The distribution of the experimental activity values and predicted values is relatively uniform. Overall, these results confirm the statistical significance of the suggested model.

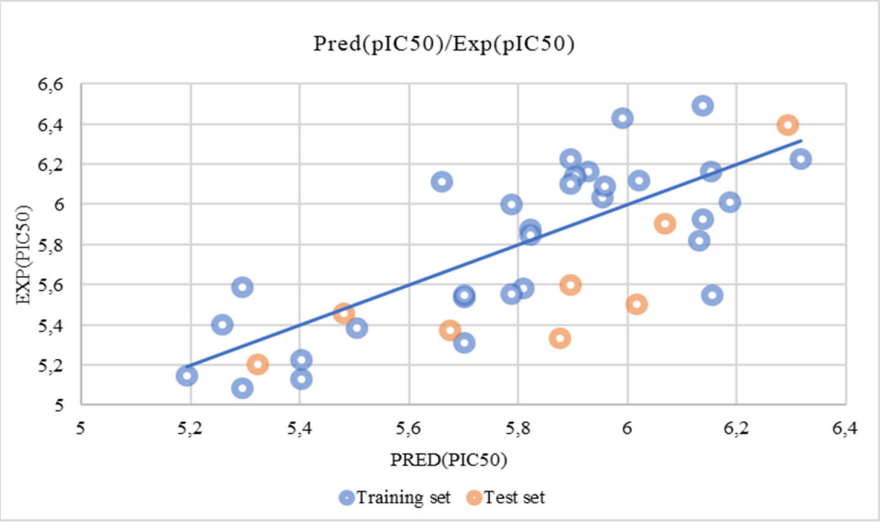


Fig. 1. Graphical presentation of the observed activity vs the predicted activity using the MLR model.

2.3. Non-Linear Multiple Regression NLMR

The non-linear relationship between the activity and the descriptors was also investigated, using the same software as before XLSTAT2023 and a pre-programmed function within. The equation of the resulting model is as follows:

$$pIC_{50} = -1.412 + 1.141\ MW + 5.888\ d - 0.163\ logP - 4.461\ (MW)^2 - 2.514\ (d)^2 - 4.377\ (logP)^2$$

This model explained 59.8% of the variation, with an R coefficient of correlation of 0.773 and an error of 0.081. It had a cross-validated coefficient of 0.58 and a predictive ability of 0.789. The model had a lower cross-validated coefficient (**R^2 cv** =0.58) than the linear model (**R^2 cv** =0.7).

Table 3. The NLMR models parameters, and thresholds between brackets.

R^2	0.598 (> 0.6)
R	0.773 (> 0.5)
R^2 cv	0.58 (> 0.5)
SCE	1.934
MCE	0.081 (low value)
RMCE	0.284 (low value)
R^2test	0.789 (> 0.6)

The calculated coefficients adhere to all established criteria, confirming that the developed MNLR model can be successfully applied to predict the studied activity. Additionally, according to **Table 4** and **Fig. 2**, a good correlation between the values predicted by the NLMR model and the experimental values has been observed. The nonlinear multiple regression (NLMR) model yielded good results with lower correlation coefficients (R^2), and cross-validated coefficient. However, these results aren't statistically improved than the MLR model. This was confirmed by comparing the predicted activity of the studied dataset using both models (**Table 4**).

In this study, we used Multiple Linear Regression (MLR) and Multiple non-linear regression (MNLR) to develop a model describing the relationship between the structure and the inhibition activity of triazoles.

The results obtained using MLR outperformed those using MNLR, demonstrating that MLR regression is suitable for modeling the inhibition activity of the IDO1 protein. MLR allows for more precise prediction of inhibitory activity as it captures linear combinations of the descriptors.[11]

Analyzing the statistical results of the two models (from Tables 2 and 3), MLR and MNLR, we observe that the MLR model exhibits higher reliability and predictive capability compared to the MNLR model. Calculation errors are lower, and the determination coefficients for the training and prediction sets are higher with the MLR model. Therefore, we conclude

that the linear approach is more suitable for analyzing the quantitative structure-activity relationship of the inhibitory activity of the IDO1 protein with triazole derivatives.

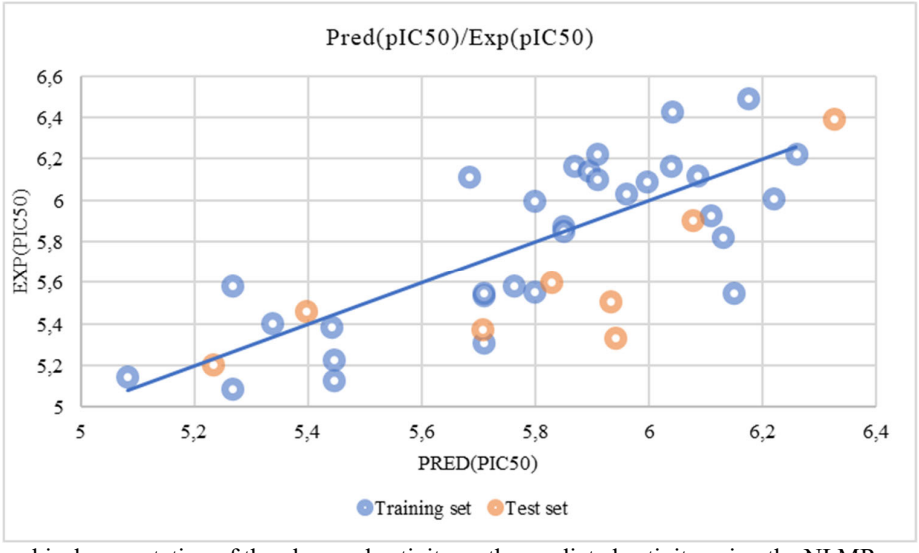


Fig. 2. Graphical presentation of the observed activity vs the predicted activity using the NLMR model.

Table 4. Observed and predicted pIC50 values using MLR and MNLR regressions

		MLR		MNLR	
Molecule	**pIC_{50}**	**pred (pIC50)**	**Residue**	**pred (pIC50)**	**Residue**
Training set					
1	6.167	5.926	0.242	5.867	0.301
3	6.495	6.137	0.358	6.174	0.32
6	5.384	5.504	-0.12	5.44	-0.056
7	5.088	5.294	-0.206	5.266	-0.178
8	5.227	5.401	-0.174	5.444	-0.217
9	5.587	5.294	0.293	5.266	0.32
10	5.129	5.401	-0.272	5.444	-0.315
11	5.401	5.257	0.144	5.337	0.064
12	5.149	5.193	-0.045	5.081	0.068
13	6.229	6.316	-0.087	6.258	-0.029
14	5.55	6.154	-0.604	6.149	-0.599
16	6.013	6.187	-0.174	6.219	-0.206
17	6.036	5.952	0.084	5.959	0.077
19	5.583	5.808	-0.224	5.762	-0.178
20	5.928	6.137	-0.209	6.107	-0.179
21	6.143	5.904	0.239	5.893	0.249
23	5.538	5.699	-0.161	5.708	-0.17
24	5.879	5.821	0.058	5.848	0.031
25	5.312	5.699	-0.387	5.708	-0.396
26	6.114	5.658	0.455	5.683	0.43
27	5.55	5.699	-0.149	5.708	-0.158
29	5.851	5.821	0.03	5.848	0.003
30	5.554	5.787	-0.233	5.797	-0.242
31	6	5.787	0.213	5.797	0.203
33	6.102	5.895	0.208	5.908	0.194
34	6.229	5.895	0.334	5.908	0.321
35	5.821	6.129	-0.308	6.13	-0.309
36	6.432	5.989	0.443	6.041	0.391
37	6.119	6.02	0.099	6.085	0.034
38	6.167	6.151	0.016	6.037	0.131
39	6.092	5.958	0.134	5.996	0.096
Test set					
2	6.398	6.293	0.105	6.325	0.073
5	5.207	5.321	-0.114	5.232	-0.025
4	5.461	5.48	-0.019	5.396	0.065
15	5.907	6.067	-0.161	6.077	-0.17
18	5.506	6.015	-0.509	5.931	-0.426
22	5.372	5.673	-0.302	5.707	-0.336
28	5.602	5.895	-0.293	5.827	-0.225
32	5.335	5.875	-0.54	5.94	-0.605

To determine the outliers and boundaries of our chosen model, (i.e. MLR model), the Williams plot was constructed using matlabR2021a software, where leverage values are calculated and compared to the critical leverage. Lower leverage values than the critical value is considered acceptable, while the contrast indicates an extrapolation of the model and is presented as an outlier. In this study, the training set consists of 31 molecules, and the model is developed using three descriptors resulting in a critical leverage value of h*=0.38. The figure below represents William's plot, all the molecules are within the boundaries of the critical leverage value and the standardized residual, signifying the reliability of the MLR model for the prediction of novel compounds with enhanced activities.

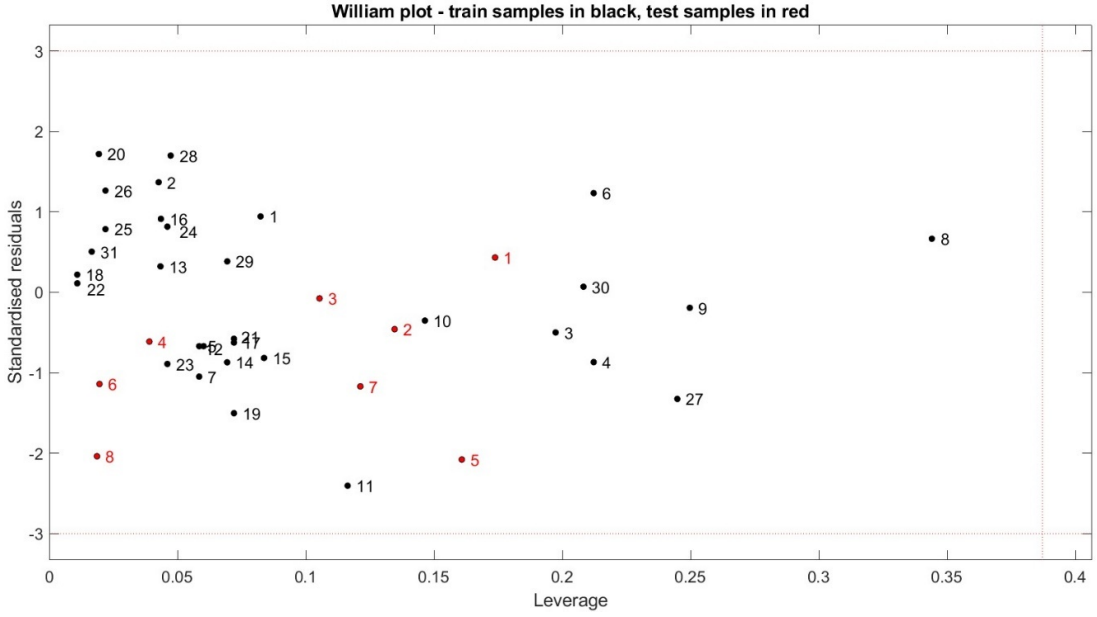


Fig. 3. Applicability domain of the MLR model.

3. Conclusions

Our work has been dedicated to modeling the quantitative structure-activity relationship of triazole inhibitors, aiming to construct reliable, robust, and accurate QSAR models capable of effectively predicting this activity. For this purpose, we conducted our study using a library of thirty-nine molecules derived from triazoles.

Our investigation successfully identified a substantial correlation between biological activity and three specific descriptors (MW, d, and logP). These findings hold promise for gaining insights into the intricate relationship between chemical structure and biological activity. Employing both Multiple Linear Regression (MLR) and Nonlinear Regression (MNLR), we developed two significant models to predict the inhibition activity of the IDO1 protein.

The first model assumes a linear relationship between the molecular descriptors (independent variables) and biological activity (dependent variable). On the other hand, the second model incorporates more complex patterns by introducing non-linearity in the relationship between molecular descriptors and activity. Despite the complexity introduced by MNLR, the MLR model demonstrated superior performance in capturing the variation of activity in the data, indicating a linear trend in the studied triazoles' activity.

All statistical parameters and validation criteria for the MLR model fall within acceptable thresholds. The model exhibited a high coefficient of determination, a strong cross-validated coefficient, a low mean square error, and demonstrated high predictive power. William's plot visualization confirmed that all molecules lie within the applicability domain, with no outliers detected. In summary, our MLR model proved to be well-fitted, robust, and highly predictive. Its successful performance makes it a valuable tool for predicting the activities of newly designed molecules in future studies.

4. Experimental

4.1. Materials and Methods

4.1.1. Data collection

In a QSAR study, the quality and size of the database are essential to obtain reliable predictions. Therefore, in this study, we collected 39 triazole derivatives from the literature.[12–15] We investigated the biological activity IC50, after a

transformation to pIC50, [16] as the dependent variable for the development of a QSAR model. **Table 5** groups the studied molecules with their activities.

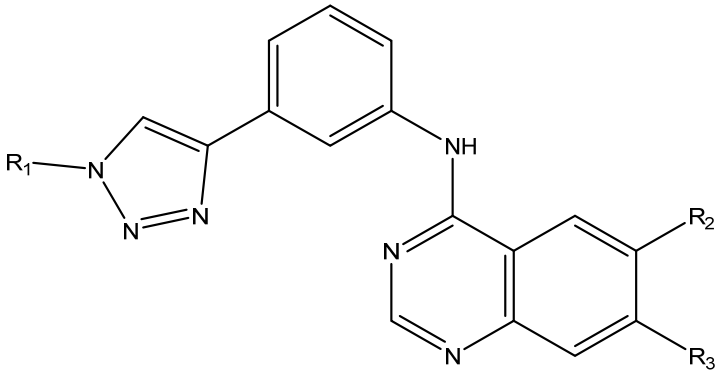
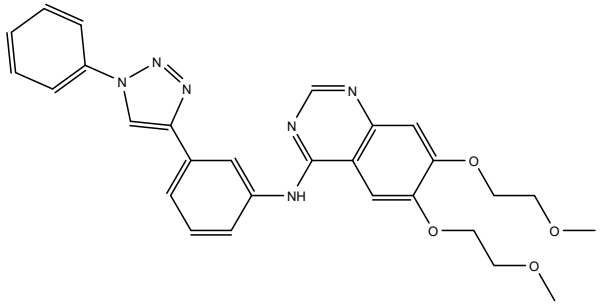
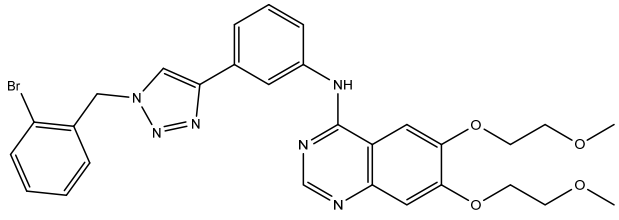
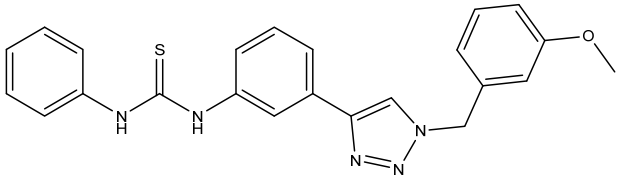
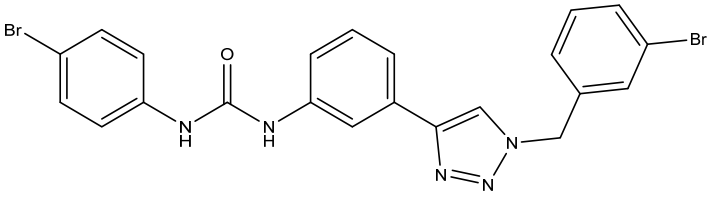
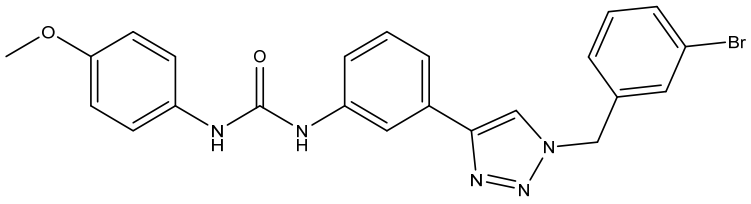
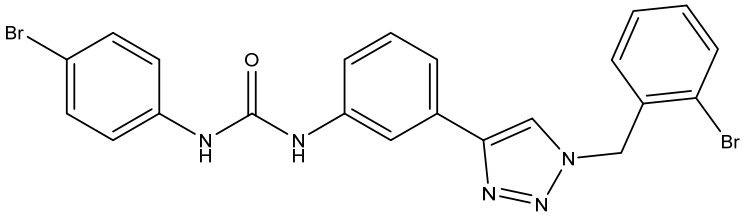
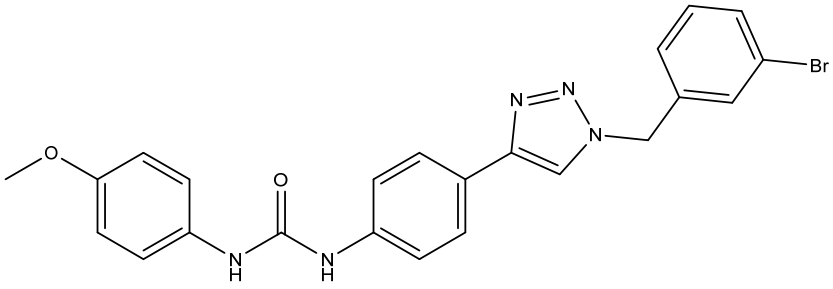
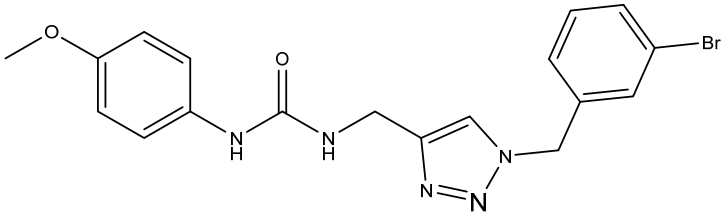
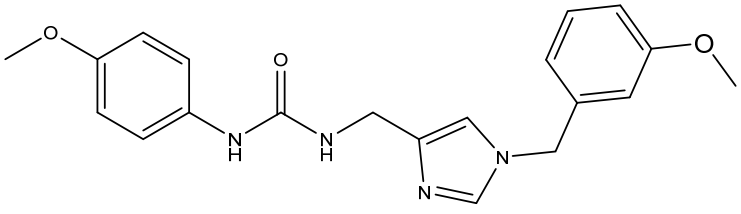
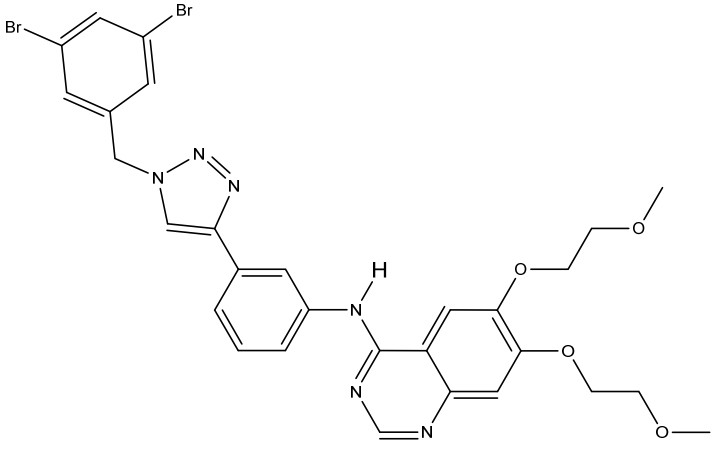
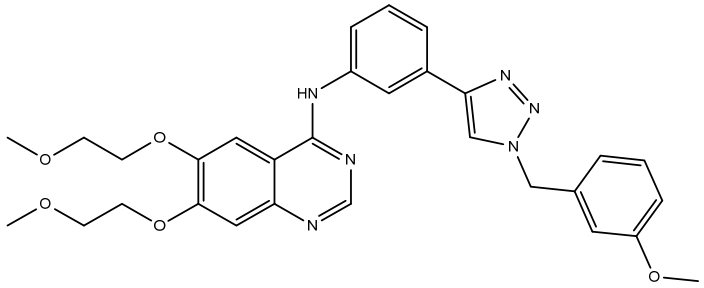
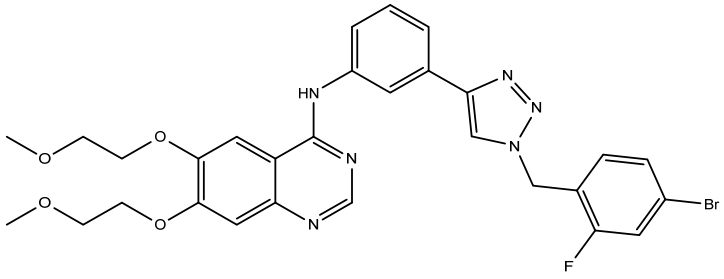
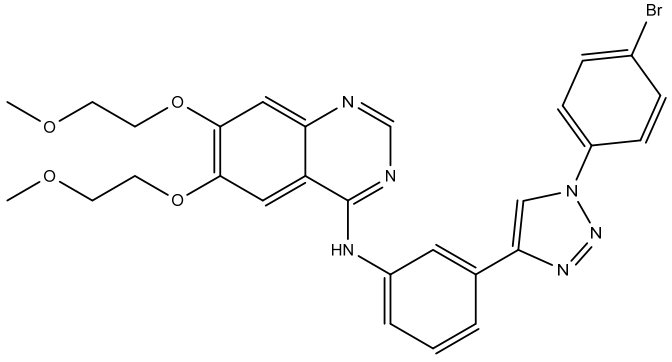
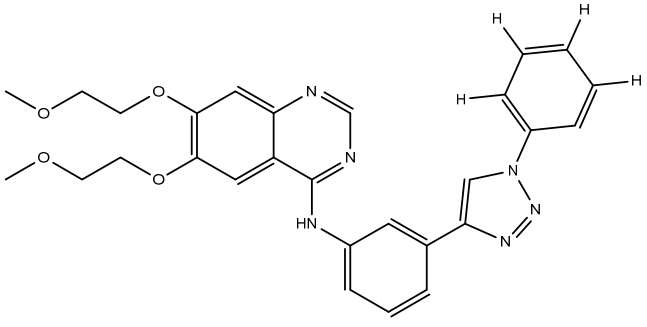
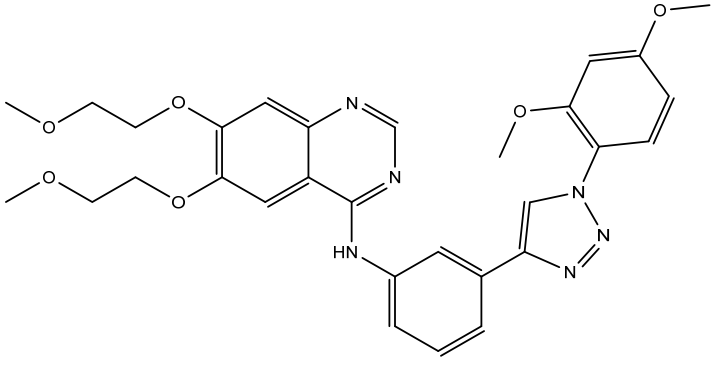
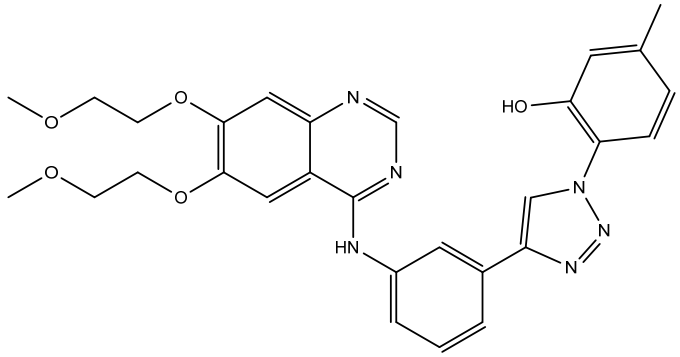
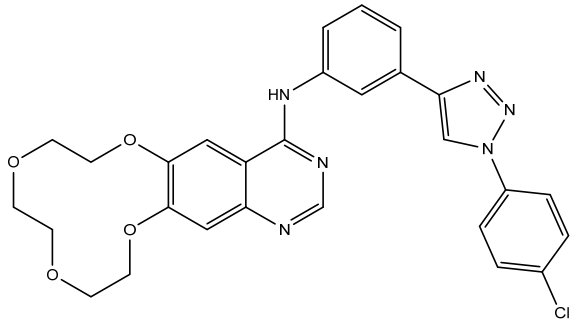
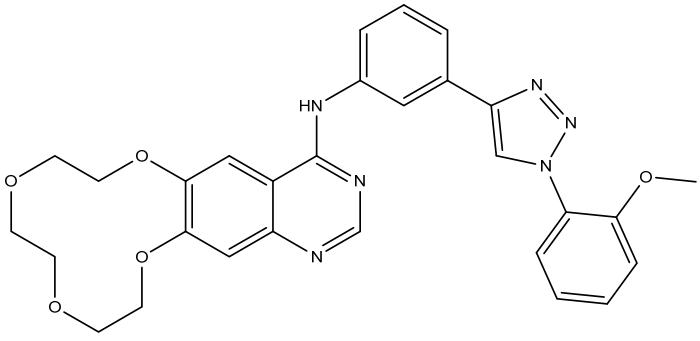
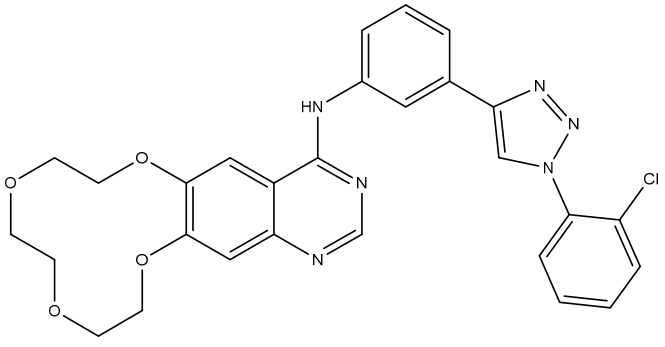
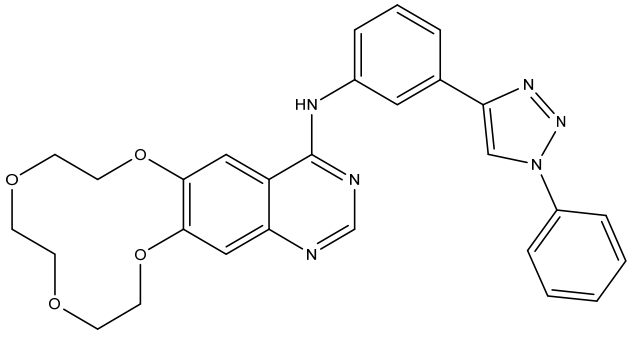


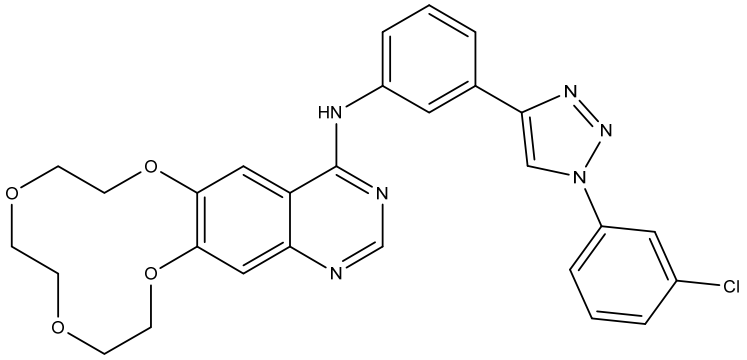
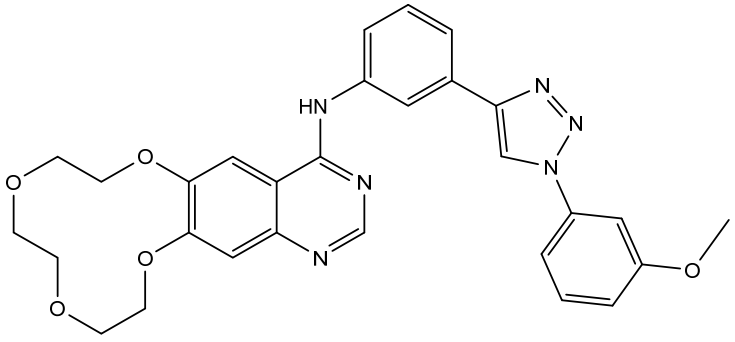
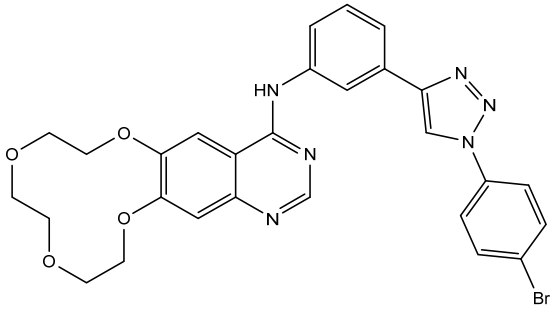
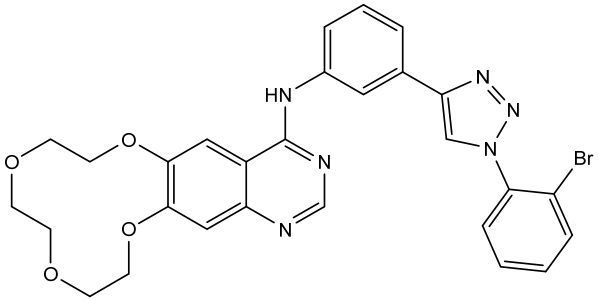
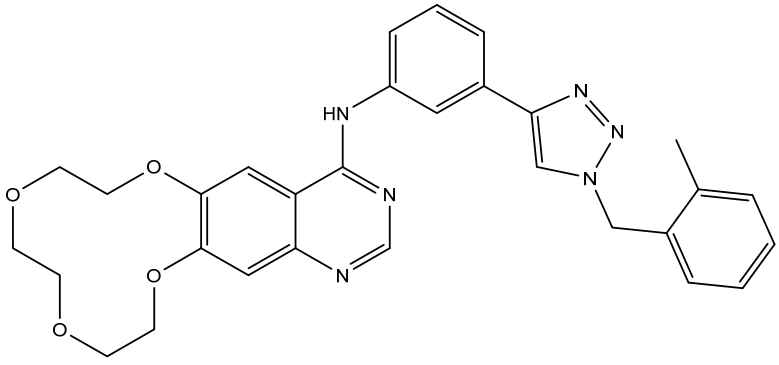
Table 5. Dataset molecules along with their pIC50 values.

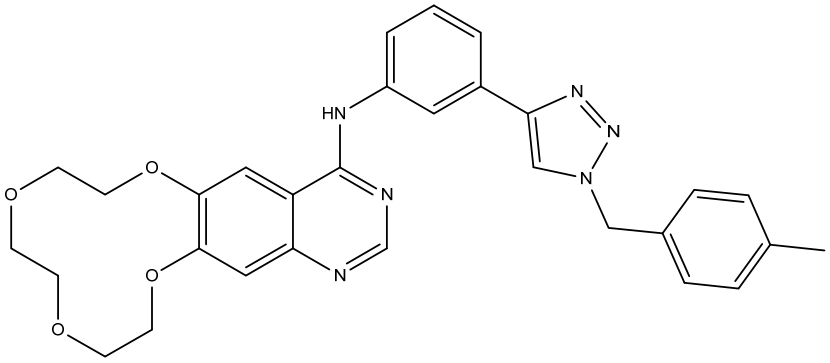
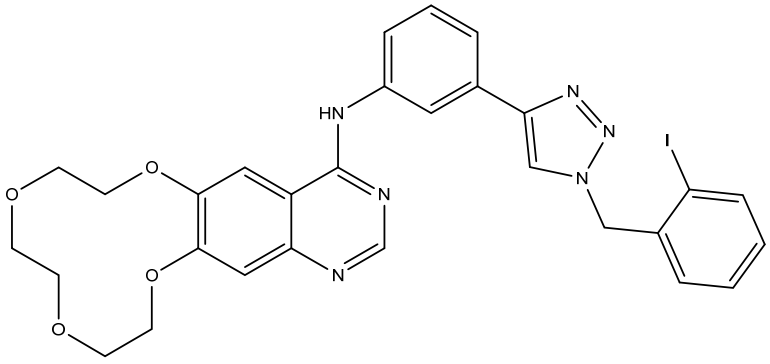
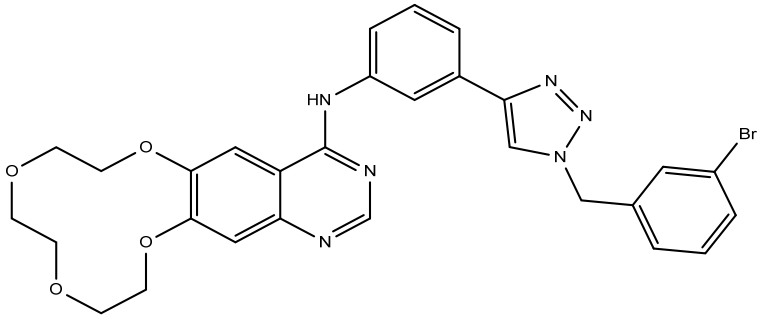
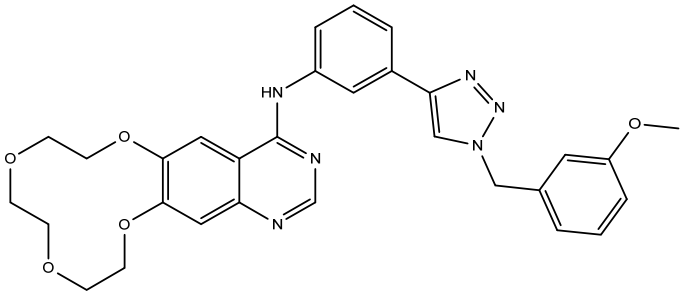
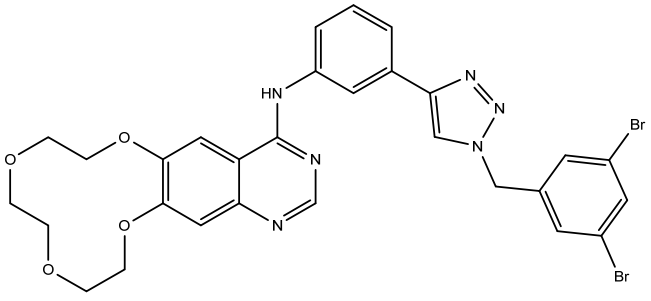
Compound	Structure	pIC_{50}
1		6.167
3		6.495
6		5.384
7		5.088

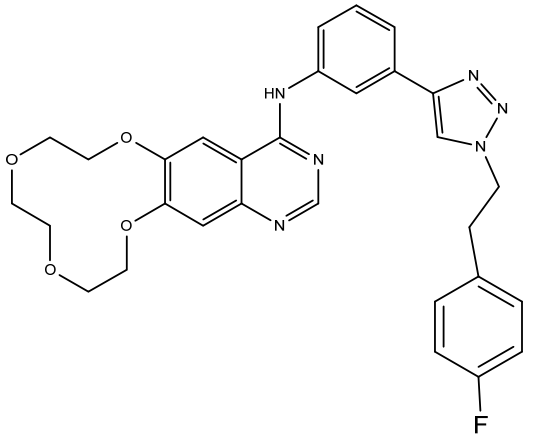
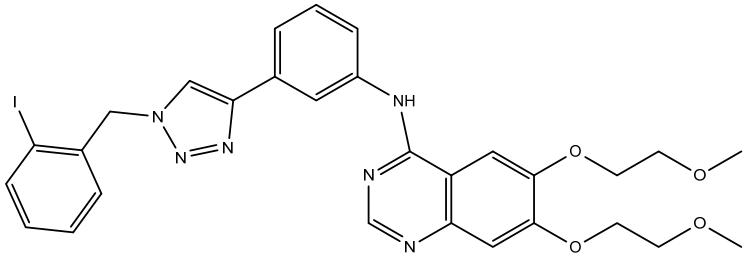
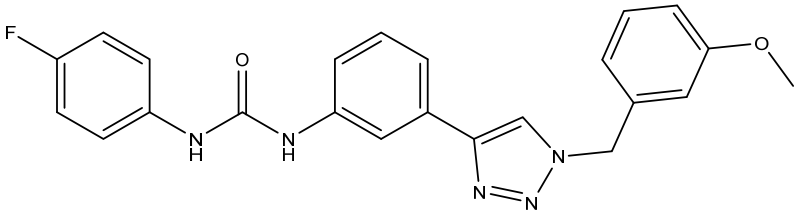
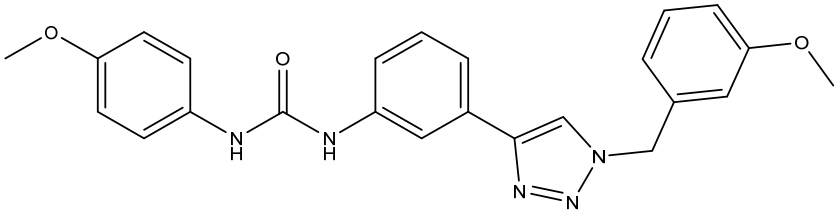
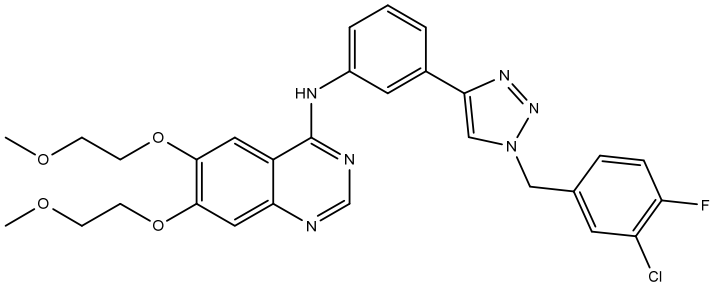
8		5.227
9		5.587
10		5.129
11		5.401
12		5.149
13		6.229

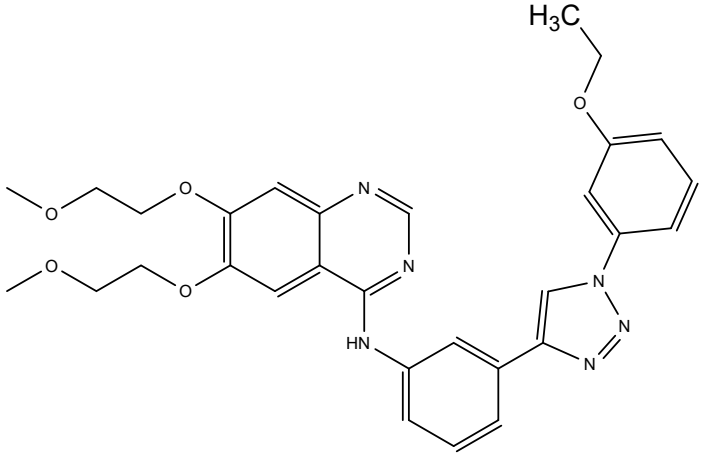
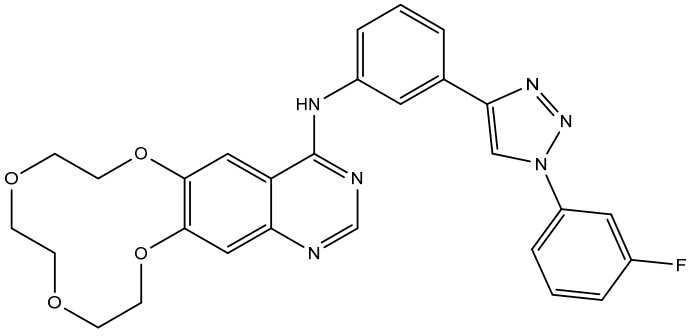
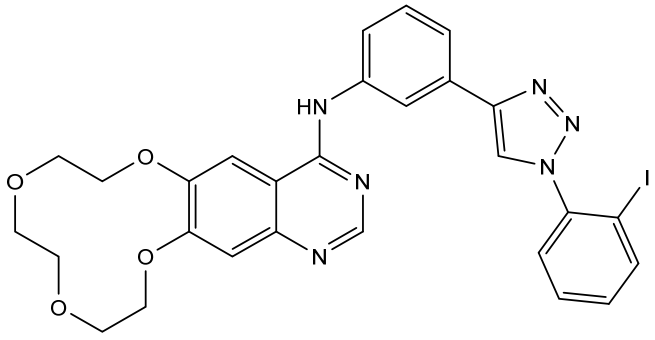
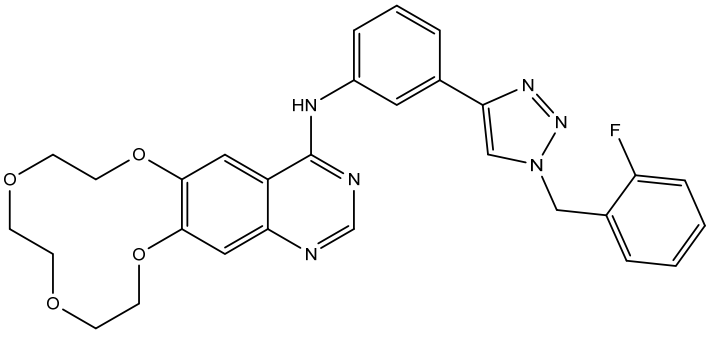
14		5.550
16		6.013
17		6.036
19		5.583
20		5.928

21		6.143
23		5.538
24		5.879
25		5.312
26		6.114

27		5.550
29		5.851
30		5.554
31		6
33		6.102

34		6.229
35		5.821
36		6.432
37		6.119
38		6.167

39		6.092
2		6.398
5		5.207
4		5.461
15		5.907

18		5.506
22		5.372
28		5.602
32		5.335

4.1.2. Descriptors calculation

The chemical structure of the molecules was transformed into various descriptors. For this purpose, Chemsketch software (http://www.acdlabs.com/) was used to calculate molecular weight (MW), molar refractivity (MR) in cm^3, molar volume (MV) in cm^3, diffraction index (n), density (d) in g/cm^3, parachor (Pr) in cm^3, γ surface tension in dynes/cm, and ae polarizability in cm^3. Additionally, Chemdraw software (http://www.perkinelmer.co.uk/category/chemdraw) was employed to determine the Octanol/Water partition coefficient (ClogP) values (**Table 6**).

Table 6. Calculated descriptors.

Molecule	pIC50	MW	MR	MV	Pc	N	γ	D	αe	logP
1	6.167	526.58	147.8	414.6	1094.6	1.631	48.5	1.26	58.62	5.56
3	6.495	605.48	155.42	427.1	1138.1	1.647	50.3	1.41	61.61	5.98
6	5.384	415.51	123.05	332.4	876.8	1.662	48.3	1.24	48.78	3.37
7	5.088	527.21	125.87	324.5	890.5	1.703	56.7	1.62	49.9	6.26
8	5.227	478.34	124.13	333.6	897.2	1.666	52.3	1.43	49.21	5.2
9	5.587	527.21	125.87	324.5	890.5	1.703	56.7	1.62	49.9	6.26
10	5.129	478.34	124.13	333.6	897.2	1.666	52.3	1.43	49.21	5.2
11	5.401	415.28	105.19	293.2	777.8	1.636	49.5	1.41	41.7	3.19
12	5.149	367.4	101.89	290.9	765.6	1.617	47.9	1.26	40.39	3.19
13	6.229	684.37	162.98	439.6	1181.7	1.663	52.1	1.55	64.61	6.84
14	5.55	556.61	153.68	436.2	1144.9	1.622	47.4	1.27	60.92	5.03
16	6.013	623.47	155.29	430	1138.3	1.641	49.1	1.44	61.56	6.12
17	6.036	591.45	150.81	411.1	1099.5	1.654	51.1	1.43	59.78	6.6
19	5.583	512.55	143.26	398.6	1056	1.637	49.2	1.28	56.79	5.56
20	5.928	572.61	154.88	441.8	1156.5	1.618	46.9	1.29	61.4	5.82
21	6.143	542.58	148.53	411	1092.7	1.642	49.9	1.32	58.88	5.82
23	5.538	544.98	145.68	381.1	1045.1	1.69	56.5	1.42	57.75	6.42
24	5.879	540.56	146.89	393.4	1066.5	1.669	53.9	1.37	58.23	5.75
25	5.312	544.98	145.68	381.1	1045.1	1.69	56.5	1.42	57.75	6.42
26	6.114	510.54	141.08	371.8	1016.2	1.683	55.7	1.37	55.92	5.53
27	5.55	544.98	145.68	381.1	1045.1	1.69	56.5	1.42	57.75	6.42
29	5.851	540.56	146.89	393.4	1066.5	1.669	53.9	1.37	58.23	5.75
30	5.554	589.43	148.63	384.3	1059.8	1.7	57.7	1.53	58.92	6.57
31	6	589.43	148.63	384.3	1059.8	1.7	57.7	1.53	58.92	6.57
33	6.102	538.59	150.11	403	1085.9	1.667	52.6	1.33	59.51	5.53
34	6.229	538.59	150.11	403	1085.9	1.667	52.6	1.33	59.51	5.53
35	5.821	650.46	158.34	405.5	1117.6	1.709	57.7	1.6	62.77	5.9
36	6.432	603.46	153.24	400	1098.4	1.691	56.6	1.5	60.75	5.94
37	6.119	554.59	151.5	409.5	1105.1	1.661	53	1.35	60.06	5
38	6.167	682.36	160.8	412.9	1141.9	1.706	58.5	1.65	63.74	6.81
39	6.092	556.58	150.17	406.7	1093.6	1.66	52.2	1.36	59.53	5.55
2	6.398	652.48	160.52	432.2	1157.3	1.665	51.4	1.5	63.63	5.94
5	5.207	417.43	116.45	323.9	853.8	1.638	48.2	1.28	46.16	4.6
4	5.461	429.47	122.39	342.7	903.9	1.632	48.3	1.25	48.52	4.25
15	5.907	579.02	152.34	426.7	1123.6	1.632	48	1.35	60.39	5.97
18	5.506	556.61	153.68	436.2	1144.9	1.622	47.4	1.27	60.92	6.32
22	5.372	528.53	140.95	374.7	1016.4	1.675	54.1	1.41	55.87	5.85
28	5.602	636.44	153.73	389.4	1079	1.719	58.9	1.63	60.94	6.83
32	5.335	542.56	145.56	390.7	1055	1.667	53.1	1.38	57.7	5.22

Highlighted compounds are test-set compounds

4.2. Statistical Methods

4.2.1. Principle Component Analysis

The principal component analysis (PCA) method is essential to reduce the number of descriptors before applying any regression analysis to eliminate any redundancy. In our study, we performed PCA to detect correlations among the selected descriptors. The results obtained through PCA were fundamental for further analysis.

4.2.2. Multiple Linear Regression

Multiple linear regression is a method used to determine the relationship between a dependent variable Y (activity) and multiple independent descriptors (X1, X2, X3,...Xn) by obtaining a significant correlation. When dealing with a set of diverse descriptors, if the correlation coefficient is statistically significant ($|R| > 0.5$), it indicates a strong correlation between the descriptors.[17]

4.2.3. Multiple Nonlinear Regression

Nonlinear multiple regression is a powerful method for evaluating the complex relationship between structure and activity, taking into account the nonlinear interactions among molecular descriptors. Unlike multiple linear regression,

which assumes a linear relationship between descriptors and activity, nonlinear multiple regression allows the capturing of more complex and nonlinear relationships. This is particularly important when descriptors vary significantly, and their relationship with activity cannot be simply modeled by a linear function.

4.3 Validation techniques and outliers detection

4.3.1. QSAR Validation

The validation of a QSAR model is the most critical step in QSAR approaches, for it determines the utility and validity of the generated model. Since the dataset is split into two parts, a training set, and a test set, there are two validation steps to be executed. An internal validation and an external validation. The first is reserved for the training set, while the second is for the test set.[18,19]

Models can internally be validated using multiple methods, among them we can cite leave-one-out cross-validation LOO [20,21], leave-many-out cross-validation LMO [20], and Y-randomization.[22] In our study, we have chosen to work with leave-one-out cross-validation LOO. This latter consists of deleting one molecule of the training set from the data, and then developing the model, which is therefore used to predict the activity of the deleted molecule. This operation is repeated until all the molecules of the training set are deleted from the data once. The average of the coefficient of determination is calculated at the end. A value higher than 0.5 is considered acceptable.[23]

The external validation is dedicated to assessing the predictive ability of the generated model. The model is used to predict the activity of an external set of molecules. Then the correlation between the observed (experimental activity) and the predicted activity is determined, a value higher than 0.6 is considered acceptable.[24]

4.3.2. Outliers Detection

The available experimental data used in QSAR studies is limited to a specific group of molecules, making each derived model exclusive to that particular group. In other words, the model is valid only for molecules that are similar to the data used for its development. To identify the application boundaries of each model and determine the reliability of predicted biological activity, an applicability domain is established.[25]

In this study, the applicability domain was defined using the well-known William's plot [25], which uses Hotelling's test and associated leverage statistics. The plot consists of an X-axis representing leverage values and a Y-axis representing standardized residuals. It also includes two reference axes: a vertical axis corresponding to the warning leverage (h*), typically set at h* = 3(p+1)/n, where p is the number of descriptors in the model and n is the number of molecules in the training set. The horizontal axis represents three standard deviation units (±3σ). The leverage of a molecule is calculated as follows: $h_i = x_i^T (X^T X)^{-1} x_i$, where xi is the descriptor row of the query compound, and X is the descriptor matrix of the remaining compounds used to develop the model.[26] It is important to note that any leverage value exceeding the warning leverage is considered an extrapolation of the model, indicating that the corresponding compound is an outlier.

References

1. Tang, K., Wu, Y.-H., Song, Y., Yu, B. (**2021)** Indoleamine 2,3-Dioxygenase 1 (IDO1) Inhibitors in Clinical Trials for Cancer Immunotherapy. *J Hematol Oncol. 14* (1), 68.
2. Pallotta, M. T., Rossini, S., Suvieri, C., Coletti, A., Orabona, C., Macchiarulo, A., Volpi, C., Grohmann, U. **(2022)** Indoleamine 2,3-dioxygenase 1 (IDO1): An Up-to-date Overview of an Eclectic Immunoregulatory Enzyme. *The FEBS Journal*. *289* (20), 6099–6118.
3. Qian, S., Zhang, M., Chen, Q., He, Y., Wang, W., Wang, Z. **(2016)** IDO as a Drug Target for Cancer Immunotherapy: Recent Developments in IDO Inhibitors Discovery. *RSC Adv. 6* (9), 7575–7581.
4. Bastikar, V., Bastikar, A., Gupta, P. **(2022)** Quantitative Structure–Activity Relationship-Based Computational Approaches. In *Computational Approaches for Novel Therapeutic and Diagnostic Designing to Mitigate SARS-CoV-2 Infection*, Elsevier. 191–205.
5. Myrko, I., Chaban, T., Demchuk, Y., Drapak, Y., Chaban, I., Drapak, I., Pankiv, M., Matiychuk, V. **(2024)** Current Trends of Chemoinformatics and Computer Chemistry in Drug Design: A Review. *10.5267/j.ccl. 13* (1), 151–162.
6. Zrinej, J., ElMchichi, L., Lakhlifi, T., Bouachrine, M. **(2022)** Curcumin and Derivatives against Human Colon Cancer (HCT-116). Computationnel. *RHAZES: Green and Applied Chemistry*. 83-102.
7. Ouabane, M., Hajji, H., Belhassan, A., Koubi, Y., Elbouhi, M., Badaoui, H., Sekkat, C., Lakhlifi, T. **(2022)** 2D-QSPR of the Retention/Release Property for Odorant Molecules in Pectin Gels of Different Concentration. *RHAZES: Green and Applied Chemistry*. 15-35.
8. Tropsha, A., Gramatica, P., Gombar, V. K. (**2003)** The Importance of Being Earnest: Validation Is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci. 22* (1), 69–77.
9. Adeniji, S. E., Uba, S., Uzairu, A. (**2018)** Theoretical Modeling and Molecular Docking Simulation for Investigating and Evaluating Some Active Compounds as Potent Anti-Tubercular Agents against MTB CYP121 Receptor. *Future Journal of Pharmaceutical Sciences*, *4* (2), 284–295.

10. Adeniji, S. E., Uba, S., Uzairu, A., Arthur, D. E. (**2019)** A Derived QSAR Model for Predicting Some Compounds as Potent Antagonist against *Mycobacterium Tuberculosis* : A Theoretical Approach. *Advances in Preventive Medicine*, *2019*, 1–18.
11. Saxena, A. K., Prathipati, P. (**2003)** Comparison of MLR, PLS and GA-MLR in QSAR Analysis. *SAR and QSAR in Environmental Research*, *14* (5–6), 433–445.
12. Mao, L., Wang, Y., Zhao, J., Xu, G., Yao, X., Li, Y.-M. **(2020)** Discovery of Icotinib-1,2,3-Triazole Derivatives as IDO1 Inhibitors. *Front. Pharmacol.11*, 579024.
13. Hou, X., Gong, X., Mao, L., Sun, G., Yang, J. **(2022)** Design, Synthesis and Biological Evaluation of Erlotinib-Based IDO1 Inhibitors. *Front. Pharmacol. 13*, 940704.
14. Xu, G.-Q., Gong, X.-Q., Zhu, Y.-Y., Yao, X.-J., Peng, L.-Z., Sun, G., Yang, J.-X., Mao, L.-F. **(2022)** Novel 1,2,3-Triazole Erlotinib Derivatives as Potent IDO1 Inhibitors: Design, Drug-Target Interactions Prediction, Synthesis, Biological Evaluation, Molecular Docking and ADME Properties Studies. *Front. Pharmacol. 13*, 854965.
15. Hou, X., Gong, X., Mao, L., Zhao, J., Yang, J. **(2022)** Discovery of Novel 1,2,3-Triazole Derivatives as IDO1 Inhibitors. *Pharmaceuticals*. *15* (11), 1316.
16. Zaki, K., Sbai, A., Bouachrine, M., Lakhlifi, T. **(2022)** Statistical QSAR Investigations Using QSAR Techniques to Study Aminopyrimidine-Based CXCR4 Antagonists. *RHAZES: Green and Applied Chemistry.*
17. Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T. D., McDowell, R. M., Gramatica, P. **(2003)** Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environ Health Perspect*. *111* (10), 1361–1375.
18. Walker, J. D., Jaworska, J., Comber, M. H. I., Schultz, T. W., Dearden, J. C. **(2003)** GUIDELINES FOR DEVELOPING AND USING QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIPS. *Environ Toxicol Chem*. *22* (8), 1653.
19. Gramatica, P. **(2007)**Principles of QSAR Models Validation: Internal and External. *QSAR Comb. Sci. 26* (5), 694–701.
20. Baumann, K. **(2003)** Cross-Validation as the Objective Function for Variable-Selection Techniques. *TrAC Trends in Analytical Chemistry*. *22* (6), 395–406.
21. Yadav, M., Narasimhan, B., Kapoor, A. **(2024)** Development of 2-Dimensional and 3-Dimensional QSAR Models of Indazole Derivatives as TTK Inhibitors Having Anticancer Potential. *10.5267/j.ccl*. *13* (1), 225–240.
22. Rücker, C., Rücker, G., Meringer, M. **(2007)** Y-Randomization and Its Variants in QSPR/QSAR. *J. Chem. Inf. Model. 47* (6), 2345–2357.
23. Golbraikh, A., Tropsha, A. **(2002)** Beware of Q2! *Journal of Molecular Graphics and Modelling. 20* (4), 269–276.
24. Elbouhi, M., Badaoui, H., Ouabane, M., Alaoui, M. A., Koubi, Y., Mokhlis, Y., ElKamel, K., Lakhlifi, T. **(2022)** Anti-Tumor Activity of Novel Benzimidazole-Chalcone Hybrids as Non-Intercalative Topoisomerase II Catalytic Inhibitors: 2D-QSAR Study. *RHAZES: Green and Applied Chemistry.* 62-75.
25. Netzeva, T. I., Worth, A. P., Aldenberg, T., Benigni, R., Cronin, M. T. D., Gramatica, P., Jaworska, J. S., Kahn, S., Klopman, G., Marchant, C. A., Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G. Y., Perkins, R., Roberts, D. W., Schultz, T. W., Stanton, D. T., Van De Sandt, J. J. M., Tong, W., Veith, G., Yang, C. **(2005)** Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships: The Report and Recommendations of ECVAM Workshop 52 ·. *Altern Lab Anim*. *33* (2), 155–173.
26. Garg, R., Smith, C. J. **(2014)** Predicting the Bioconcentration Factor of Highly Hydrophobic Organic Chemicals. *Food and Chemical Toxicology*. *69*, 252–259.