Moroccan Journal of Chemistry
ISSN :2351-812X

Mor. J. Chem., 2023, Volume 11, Issue 4, Page 1137-1182
https://revues.imist.ma/index.php/morjchem
https://doi.org/10.48317/IMIST.PRSM/morjchem-v11i04.39694

QSAR model to develop newer generation GSK-3β inhibitors targeting Alzheimer

Supriyo Saha[1*], Prinsa[2], Vikash Jakhmola[1], Arun Kumar Mahato[3], Praveen Kumar Ashok[4], Vishal Warikoo[3], Sayantan Mukhopadhay[5]

[1]*Department of Pharmaceutical Chemistry, Uttaranchal Institute of Pharmaceutical Sciences, Uttaranchal University, Premnagar, Dehradun, Uttarakhand- 248007, INDIA.*
[2]*Siddhartha Institute of Pharmacy, Saharastradhara Road Near IT-Park Dehradun 248001, Uttarakhand India.*
[3]*Sardar Bhagwan Singh University, Balawala, Dehradun 248001, Uttarakhand India.*
[4]*Gyani Inder Singh Institute of Professional Studies, Dehradun 248001, Uttarakhand India.*
[5]*College of Pharmacy, Shivalik Campus, Dehradun-248001.Uttarakhand India.*

**Corresponding author, Email address: supriyo9@gmail.com; Orchid id: https://orcid.org/0000-0003-1365-4698*

Received *27April 2023,*
Revised *26 Aug 2023,*
Accepted *28Aug 2023*

Citation*: Saha S., Prinsa, Jakhmola V., Mahato A.K., Ashok P.K., Warikoo V., Mukhopadhay S. (2023) QSAR model to develop newer generation GSK-3β inhibitors targeting Alzheimer, Mor. J. Chem., 11(4), 1137-1182*

Abstract: In the year 2022 most of the patients affected by the disease was around 65 year age. Among total number of patients, 73% were near 75 year or older age. It was also stated that maximum numbers of patients were women. Black Americans were more affected by Alzheimer than white Americans. GSK-3 has also been linked to the hyperphosphorylation of tau protein, the development of amyloid-beta plaques, other inflammatory responses, activation of microglial cells, the production of neurotoxic inflammatory factors, and a decrease in the level of acetylcholine, all of which together lead to Alzheimer's disease. GSK-3 controlled the inflammatory stress brought on by anomalies in the mitochondria and endoplasmic reticulum. However, none of the compounds utilised in the treatment were particularly helpful in curing the patient completely. The development of newer generation anti-Alzheimer therapeutic compounds was therefore hampered by this curse, and computational approaches were crucial in breaking it. The most effective QSAR model was pIC50 = -5.47052 +2.60572 IC1 +1.64642 GATS2e +2.088 mindssC -0.01441 ATSC7s -13.5191 AVP-0 +0.16712 minssNH -0.15369 minaaN +0.01777 VR2_Dt +1.52684 MATS8s +0.04725 nAtomP with all necessary acceptance criteria Q^2: 0.60111, r^2: 0.65711, |r0^2-r'0^2|: 0.07866, k: 0.99121 [(r^2-r0^2)/r^2] 0.00543 or k': 0.92437 [(r^2-r'0^2)/r^2] 0.12513. It is clear that our QSAR model will be a blessing for humanity if we wish to produce a chemical that works as a GSK-3 inhibitor to treat Alzheimer's disease in the near future.

Keywords: GSK-3β; Alzheimer Disease; Modelability Index; KS Method; GT acceptable criteria; YR Test.

1. Introduction

According to recent statistics, more than 60 lakh Americans had Alzheimer's disease. Most patients with the condition in 2022 were around 65 years old (Tang *et al.*, 2014; Limor & Eldar-Finkelman 2013; El khatabi *et al.*, 2021). A total of 73% of the patients were 75 years of age or older. Additionally, it was noted that women made up the majority of the patients. According to the

Alzheimer's Facts and Figures Report Alzheimer's Association, black Americans were more likely to develop Alzheimer's disease than white Americans. Rivastigmine, Donepezil, galantamine, memantine are some commercially available anti Alzheimer drugs. GSK-3, or glycogen synthase kinase 3, originally attracted attention in year 1980. According to (Pal & Saha 2019; Ling *et al.*, 2013), the enzyme was primarily used to facilitate the production of glycogen from glucose using uridine diphosphate glucose molecules. GSK was a cell-specific enzyme with a serine/threonine amino acid basis (Akihiko 2006; Toral-Rios *et al.*, 2020). The enzyme comes in two varieties, GSK-3 and GSK-3. Through the phosphorylation of serine21 for alpha and serine9 for beta, this enzyme primarily initiated the downregulation process of neurons (Angela *et al.*, 2021). According to (Griebel *et al.*, 2019), GSK-3 controlled the development of beta-amyloid plaques via the Wingless and Int-1/phosphatidylinositol-3 pathway. According to (Kim *et al.*, 2006), GSK-3 has also been linked to the hyperphosphorylation of tau protein (Kareti & Subash 2020; Chtita *et al.*, 2016), the development of amyloid-beta plaques, other inflammatory responses, activation of microglial cells, the production of neurotoxic inflammatory factors, and a decrease in the level of acetylcholine, all of which contribute to the development of Alzheimer's disease (El Alaouy *et al.*, 2021; El-Mernissi *et al.*, 2023). According to Hooper *et al.*, 2008, GSK-3 controlled the inflammatory stress caused by anomalies in the endoplasmic reticulum and the mitochondria (El Alaouy *et al.*, 2023). These pathways have been collectively linked to numerous neurological conditions like Parkinson's, Alzheimer's, mood swings, and other illnesses connected to cognition and behaviour (Ma, 2014; De Strooper & Karran 2016). To anticipate the bioactivity of newer generation GSK-3 inhibitors effective against Alzheimer's disease, we created a QSAR model in this context.

2. Materials and Methodology

2.1. Dataset and Descriptor Calculation

A dataset of 124 GSK-3 inhibitors for Alzheimer's disease was obtained from the database (Thomas *et al.*, 2012). The ACD ChemSketch software was used to create each molecule, which was then saved in MDL Mol format. Then, using the PADEL descriptor, two-dimensional descriptors of the molecules were derived (Saha *et al.*, 2022a). The biological activities associated with each molecular descriptor were tabulated in.CSV format, with IC50 values converted to pIC50 values.

2.2.Modelability Index

Modelability index is a method for estimating feasibility that is defined by the ratio between the total number of pairings and the nearest-neighbour pairs of compounds that belong to the same activity class (Saha *et al.*, 2022b; Alamari *et al.*, 2023). This idea was related to the wasteful efforts of a QSAR dataset involved in the creation of a QSAR model.

2.3. Descriptor Pretreatment

Then, by using a variance cut off and correlation coefficient values of 0.0001 and 0.99, respectively, closely related descriptors found in the dataset were eliminated (Saha *et al.*, 2022).

2.4. Dataset Division

The Kennard Stone (KS), Random Faster, and Euclidean Distance methods were typically used to partition the dataset into training and test sets. We chose the KS approach to partition the dataset of 124 molecules into a training set and a test set in this instance. Following dataset partition, the training set contained 86 molecules, whereas the test set contained 38 molecules (de Moura *et al.*, 2021).

2.5. Suitable Descriptor Selection

Using Stepwise MLR software, a set of appropriate descriptors was chosen with F values ranging from 3.9 to 4.0. Then, the R2 cut off value of 0.6 was shown to be the optimum subset combination (Kumar *et al.,* 2023).

2.6. Stepwise regression

The construction of a stepwise multiple linear regression equation involved three independent processes, including the discovery of an initial model, repeating the previous step to improve the F and R2 value, and calibrating the model. Statistical SPSS software was used to create the stepwise regression equation, which was evaluated using the parameters of explained variance (R2a), correlation coefficient (R), standard error of estimate (s), and variance ratio (F) with a given DF. Finally, using cross validation R2 (Q2), SPRESS, and SDEP parameters, the LOO approach was used to validate the model (Hassan *et al.,* 2022).

2.7. QSAR Equation Development

According to the accuracy of the predictions, the final QSAR model was created by Multiple Linear Regression Plus valid software (Hanieh *et al.,* 2022).

2.8. QSAR Equation Validation

The acceptable model criteria of Golbraikh and Tropsha(GT acceptable criteria) were used to validate the constructed QSAR model. The following were the requirements for an acceptable model (Ambure & Roy 2016).

1. Q2 > 0.5.
2. R2 > 0.6.
3. |r02-r'02| < 0.3.
4. [0.85<k<1.15 and ((r^2-r0^2)/r^2)<0.1 or [0.85<k'<1.15 and ((r^2-r'0^2)/r^2)<0.1].

2.9. QSAR Equation Validation

The LOO procedure was used to cross-validate the QSAR model. By using mahalanobis distance and euclidean distance approaches, the model's applicability domain was examined. A specified application domain threshold value was compared to the distance between a test set and its closest neighbour in the training set (Yap 2011).

2.10. MLR YRandomization (YR)test

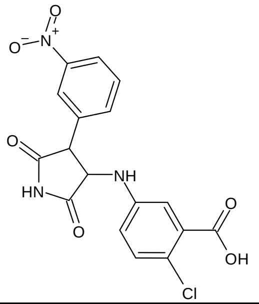
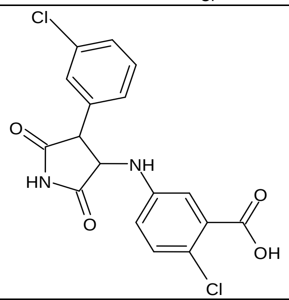
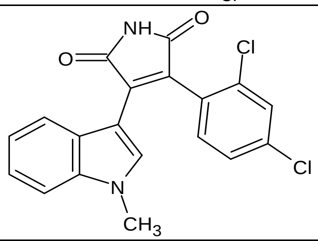
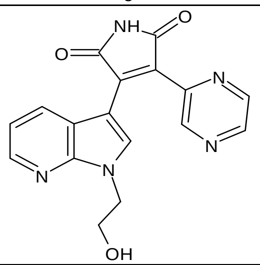
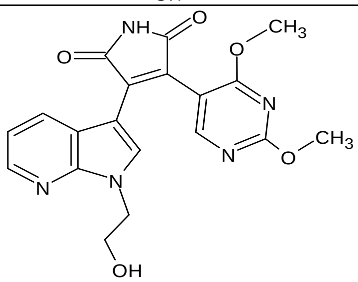
In the YR test, a quicker random technique was used to create a random multiple linear regression model by changing the dependent variable while keeping the independent variable constant. After multiple trials, the model with considerably better R2 and Q2 values demonstrated that the proposed model was reliable and repeatable. In order to pass this test, another parameter, cRp2, which must be more than 0.5, was also calculated (Golbraikh *et al.,* 2014).

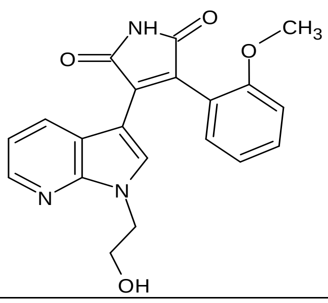
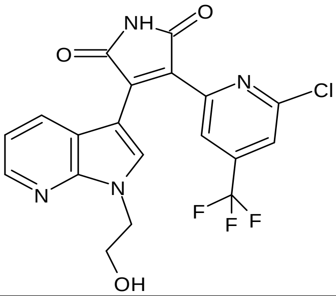
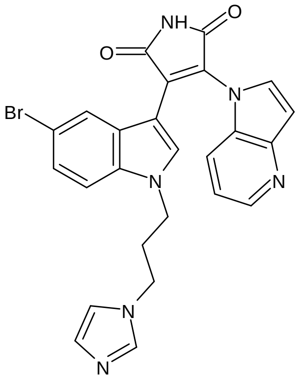
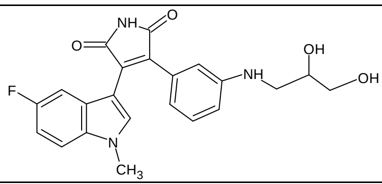
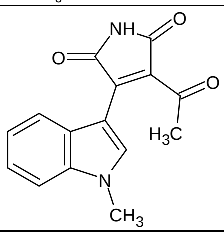
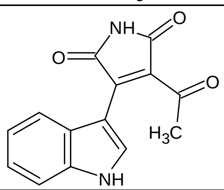
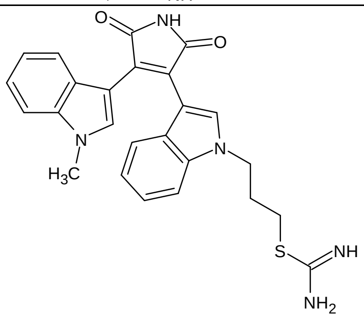
3. Results and Discussion

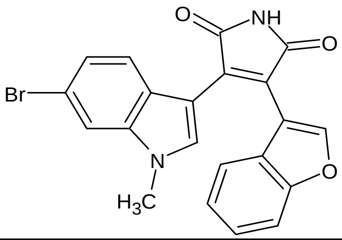
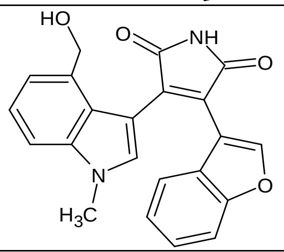
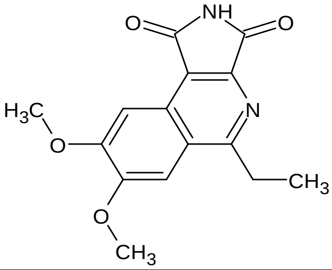
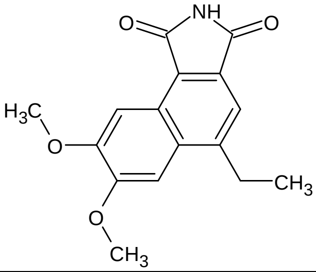
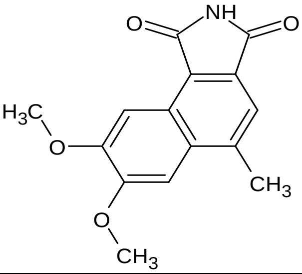
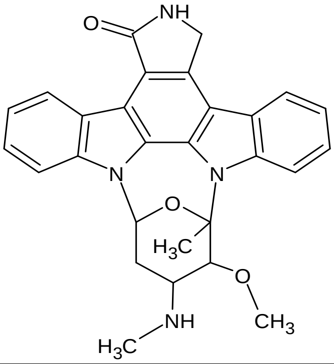
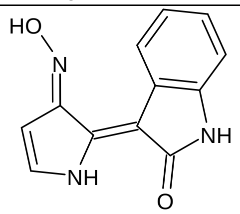
When the modelability index value of the entire dataset was first verified, it was found to be 0.55, with 41 molecules having a high total active/lower activity and 83 molecules having a low total active/toxic activity, with a threshold value of 0.65. Therefore, the model's modelability index score was 0.5926, indicating that the dataset was near to what was needed to create a good QSAR

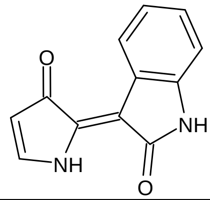
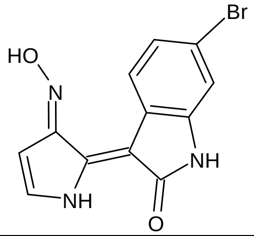
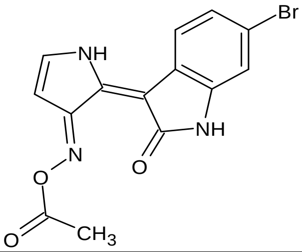
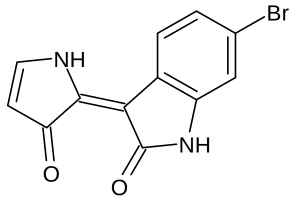
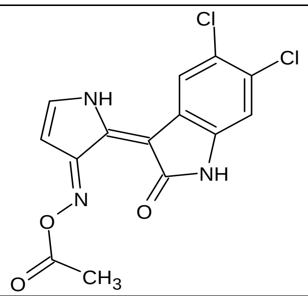
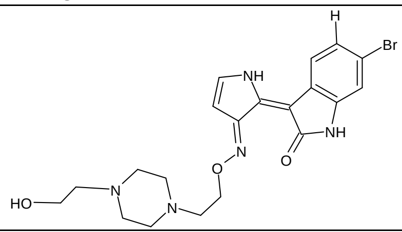
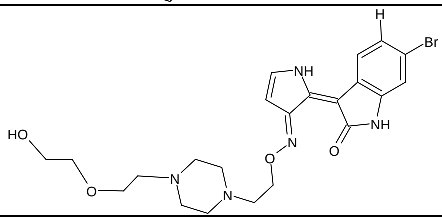
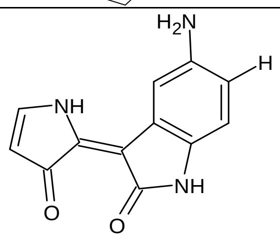
(Quantitative Structure Activity Relationship) model. The Kennard-Stone approach was then used to split the entire dataset into training and test set. In the training and test sets, there were, respectively, 86 and 38 molecules present. The most likely group of descriptors to employ in a successful QSAR model were found using stepwise multiple linear regression analysis. Then, using the best set of descriptors available, the best subset selection process was carried out using an R2 cut-off value of 0.6 and an R2 cut-off value of 0.5 for inter-correlation between descriptors (Ballabio *et al.*, 2014). Following the multiple linear regression analysis, we created five distinct QSAR models (Table 1 and Table 2), ranging from the fewest to the most descriptors.

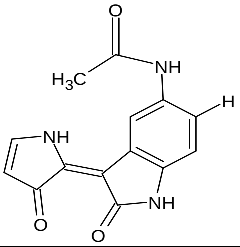
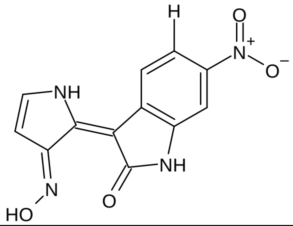
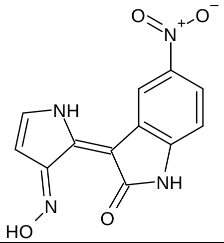
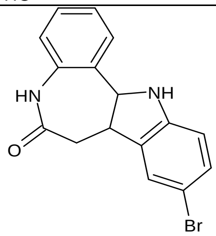
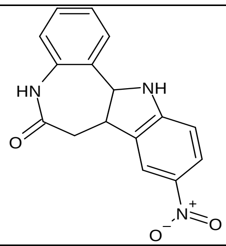
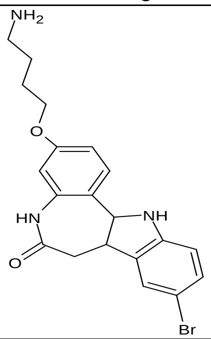
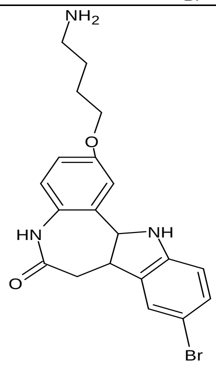
Table 1. Actual pIC50, and Predicted pIC50 Values of Training Set Molecules of the best QSAR Model.

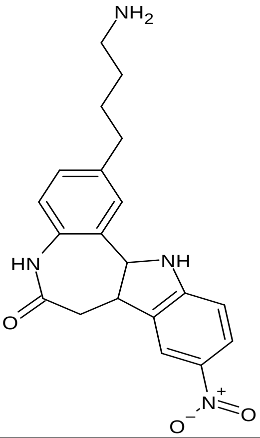
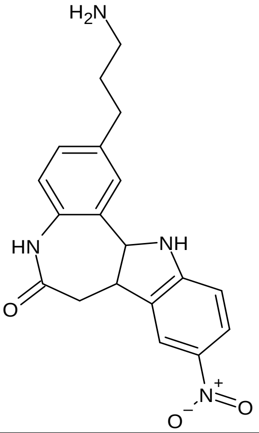
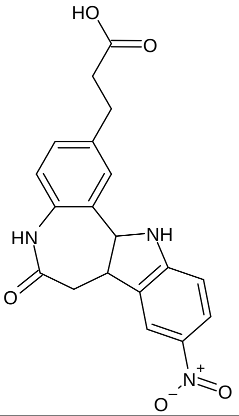
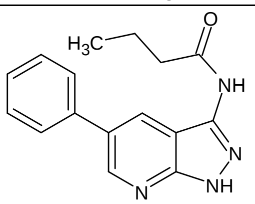
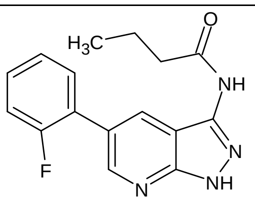
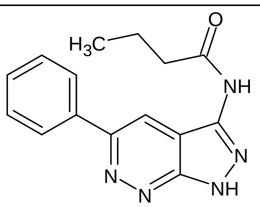
SN	Structure of the compounds	Actual pIC50	Predicted pIC50
1		-1.41497	-2.22094
2		-2.89542	-2.73921
3		-1.53148	-2.18685
4		-1.39794	-0.72689
5		-0.77815	-0.40648

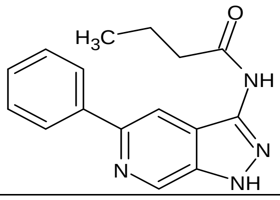
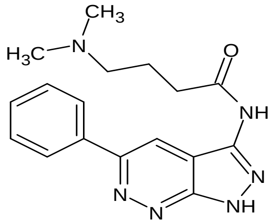
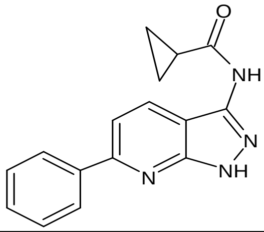
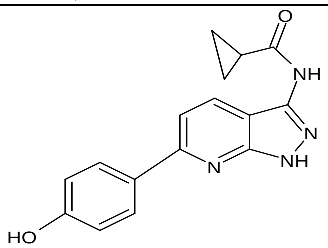
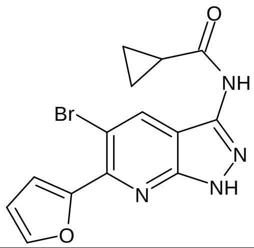
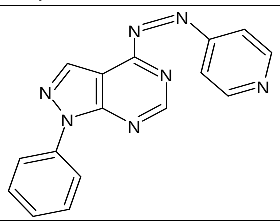
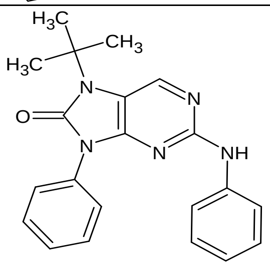
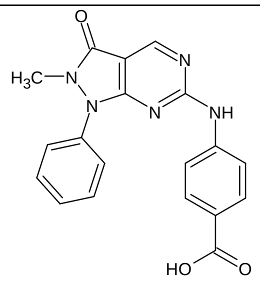
6		-0.60206	-0.58177
7		-1.41497	-1.62494
8		-2.14613	-1.1836
9		0.221849	-0.35942
10		-2.94939	-0.99586
11		-3.65031	-1.34828
12		-0.44716	0.589746

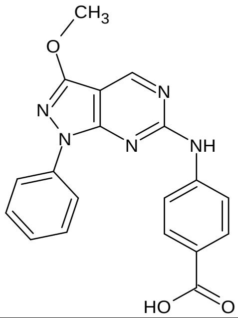
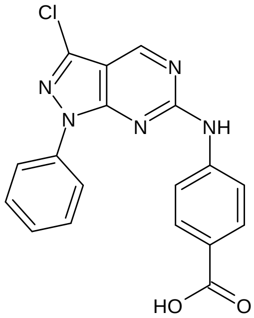
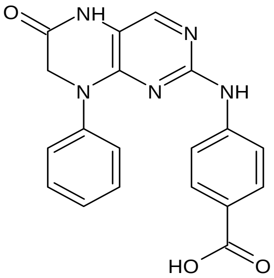
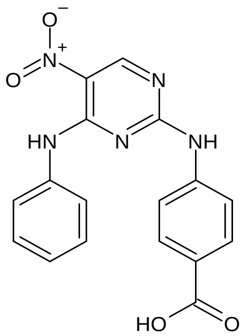
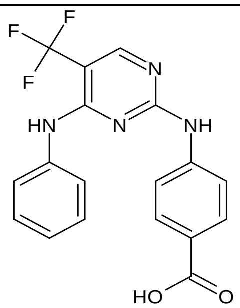
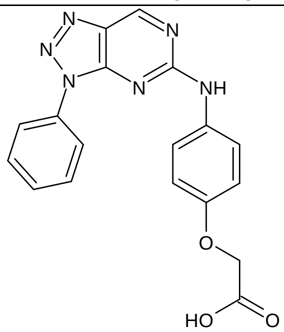
13		-0.8451	-0.54697
14		-0.73239	-0.51165
15		-2.48287	-2.19145
16		-1.96379	-2.13535
17		-2.43136	0.083766
18		-1.17609	-1.86128
19		-1.34242	-1.27194

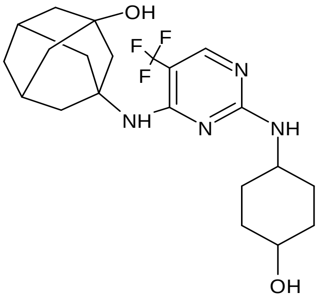
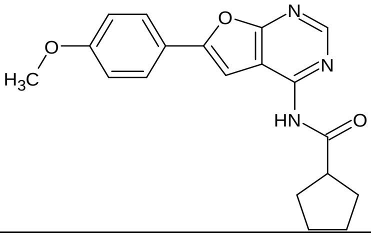
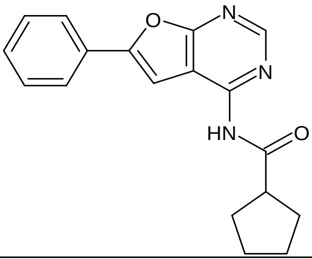
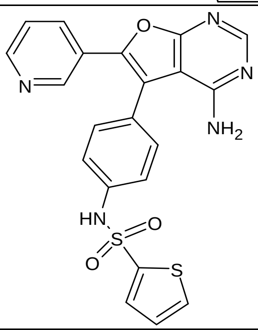
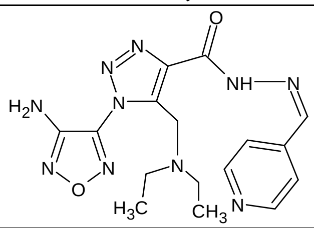
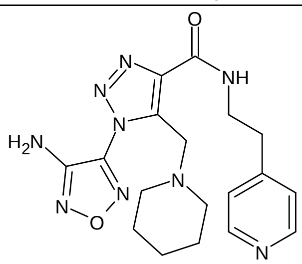
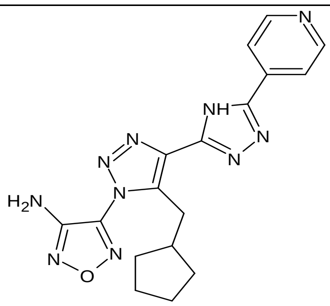
20		-2.77815	-0.62703
21		-0.69897	-0.12701
22		-1	-0.78529
23		-1.65321	-0.87523
24		-0.60206	-0.93966
25		-0.69897	-1.00089
26		-1.14613	-1.03339
27		-1.90309	-0.72496

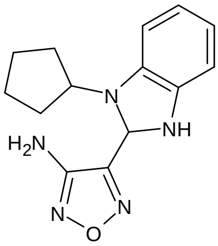
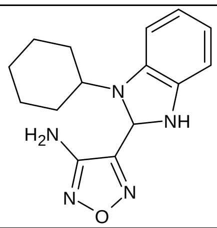
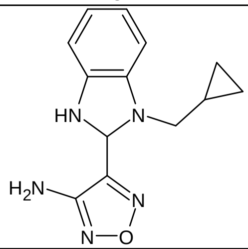
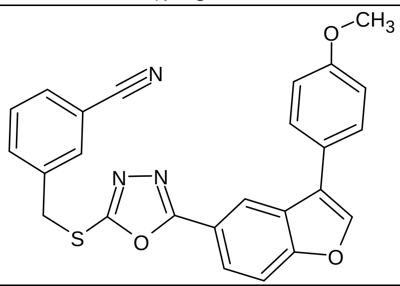
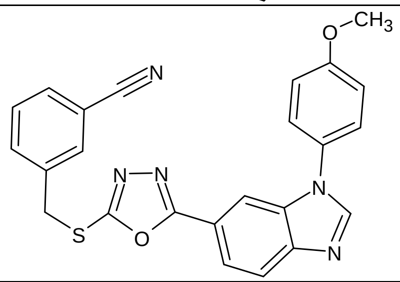
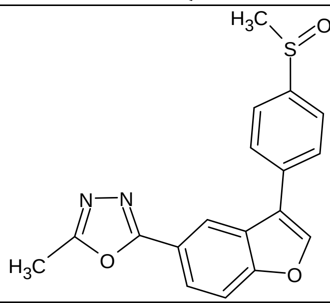
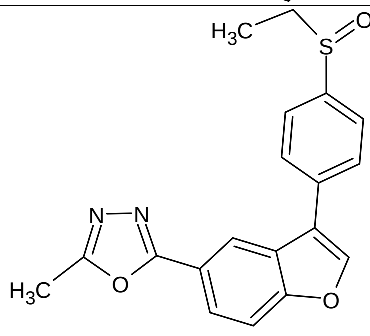
28		-0.87506	-0.6128
29		-1.60206	-2.3758
30		-0.32222	-1.52093
31		-1.36173	-1.96416
32		-0.60206	-0.68383
33		-1.47712	-0.39746
34		-1.60206	-1.70237

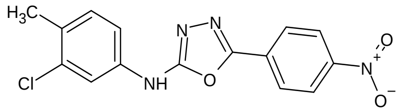
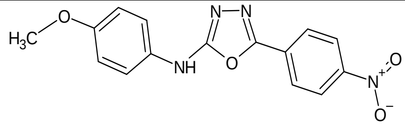
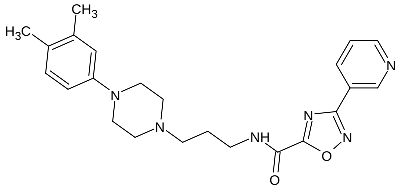
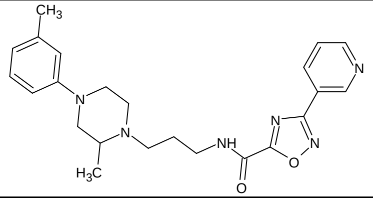
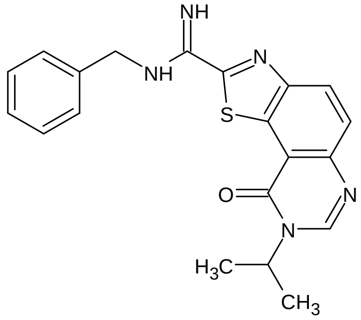
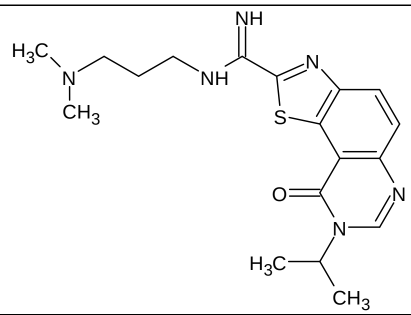
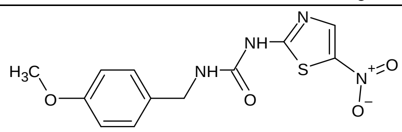
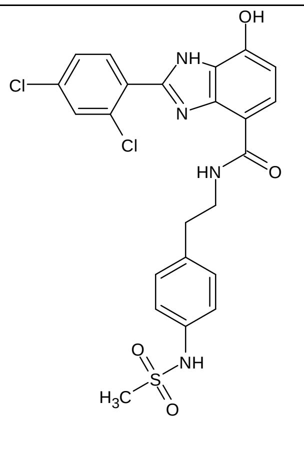
35		-0.77815	-1.90844
36		-0.39794	-1.97428
37		-0.81291	-1.35609
38		-1.74819	-1.75919
39		-1.25527	-1.4052
40		-0.60206	-1.51684

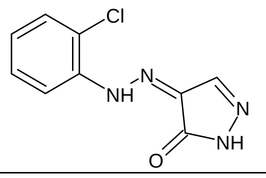
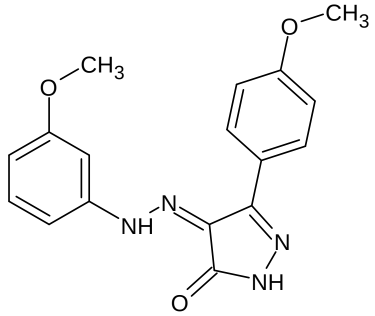
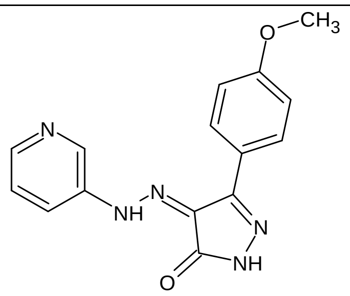
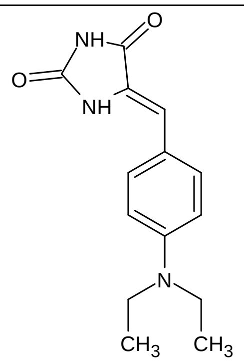
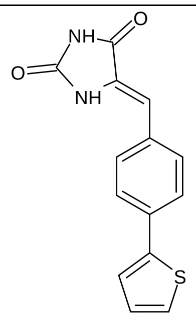
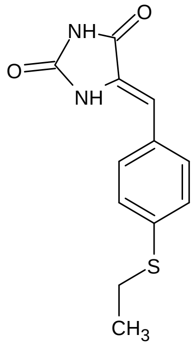
41		-0.8451	-1.7273
42		-1.34242	-0.69354
43		-2.62839	-0.8534
44		-0.90309	-1.86433
45		-0.8451	-0.59118
46		-2	-0.80109
47		-3.38021	-2.50705
48		-2.98227	-0.54668

49		-2.88081	-2.81112
50		-2.25527	-2.13329
51		-3.54407	-2.17863
52		-2.17609	-1.89979
53		-2.6902	-2.88638
54		-1.11394	-1.97767

55		-1.61278	-1.49188
56		-1.50515	-0.97557
57		-1.36173	-1.86114
58		-1.36173	-1.27403
59		-2.61278	-1.54548
60		-3.06446	-1.26618
61		-2.44716	-1.51465

62		-2.32222	-1.99199
63		-2.38021	-2.60846
64		-2.4624	-2.53969
65		-0.54407	-1.49584
66		-0.36173	-1.76997
67		-2.14613	-2.36521
68		-2.27875	-1.50115

69		-1.23045	-0.76386
70		-0.8451	-1.10673
71		-2.54407	-2.92428
72		-2.83885	-2.337
73		-2.11394	-2.5921
74		-2.74819	-2.86505
75		-2.01703	-2.19594
76		-1.17609	-1.9819

77		-3.17319	-1.66181
78		0.09691	-3.19567
79		-0.30103	-2.73399
80		-3.62325	-0.74004
81		-3.80618	-1.02228
82		-3.89209	-1.48276

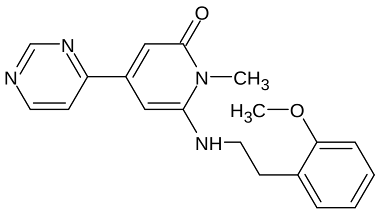
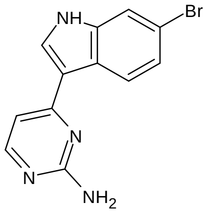
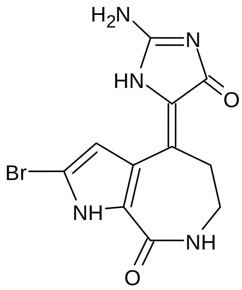
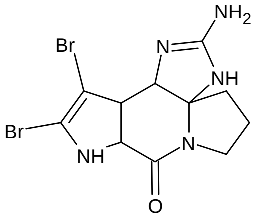
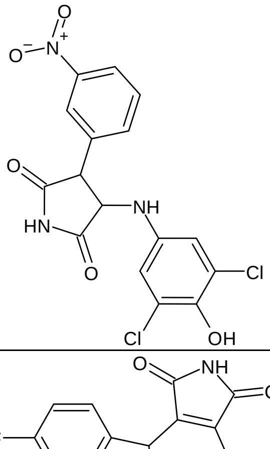
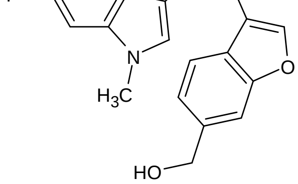
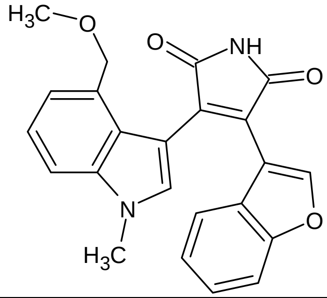
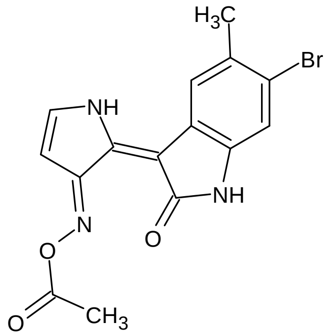
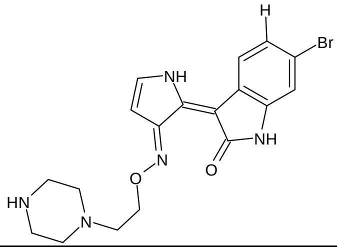
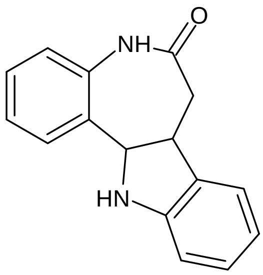
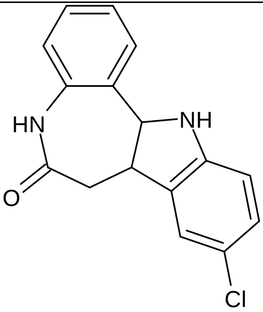
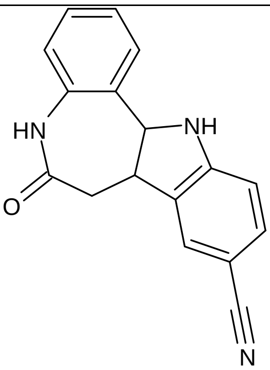
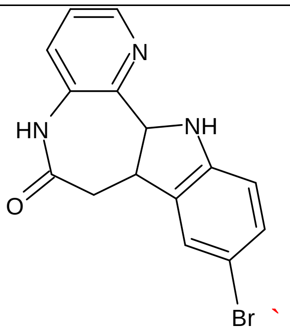
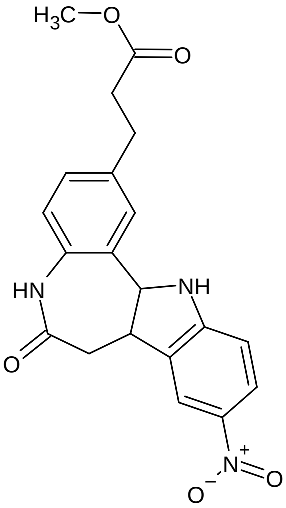
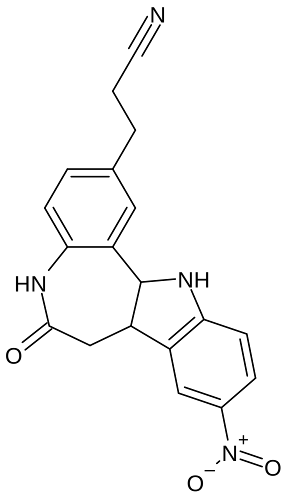
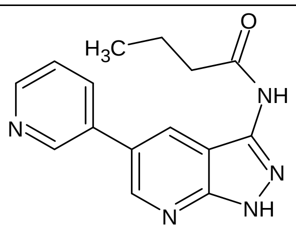
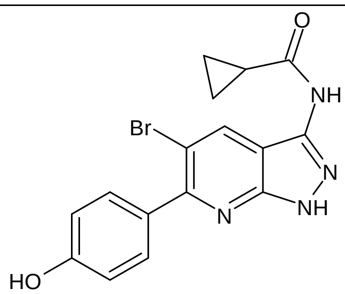
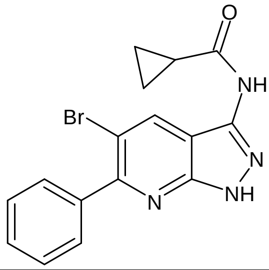
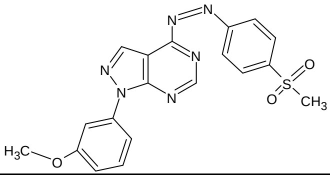
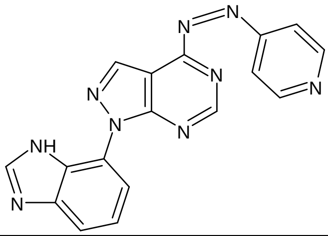
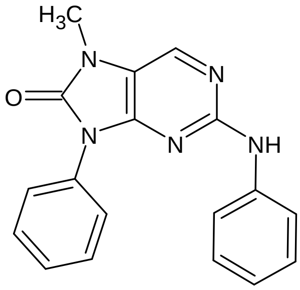
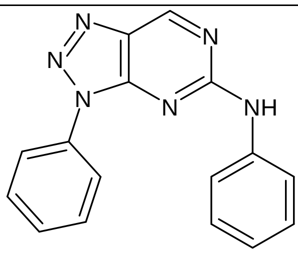
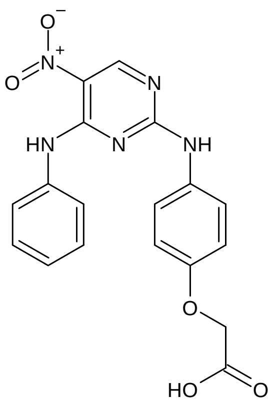
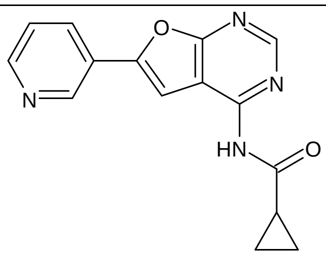
83		-1.24055	-3.39249
84		-3.39794	-3.80155
85		-1	-1.0735
86		-3.47712	-2.29833

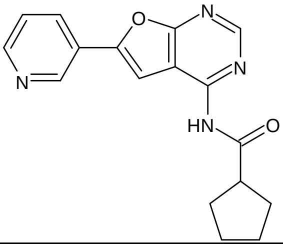
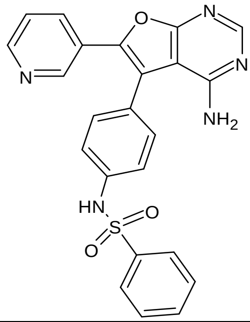
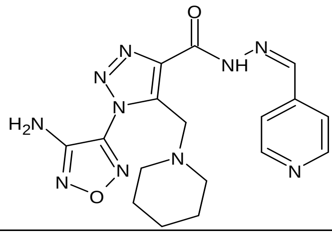
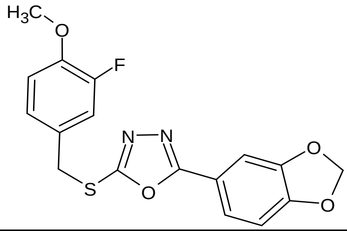
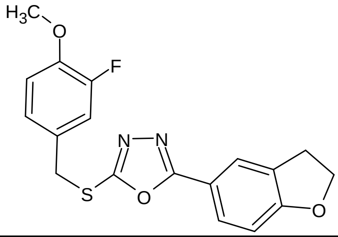
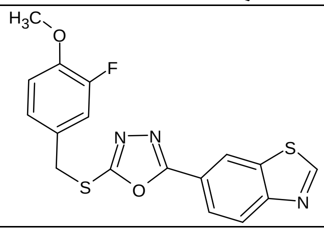
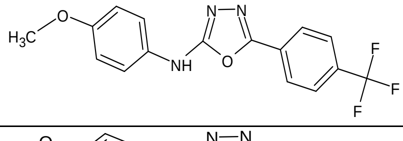
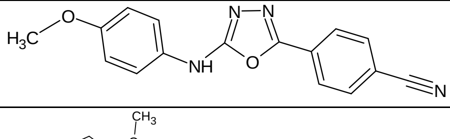
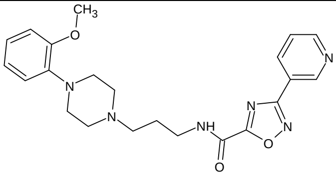
Table 2. Actual pIC50, and Predicted pIC50 Values of Test Set Molecules of the best QSAR Model.

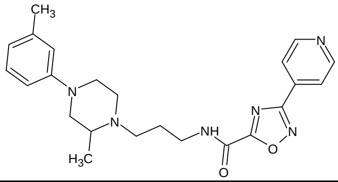
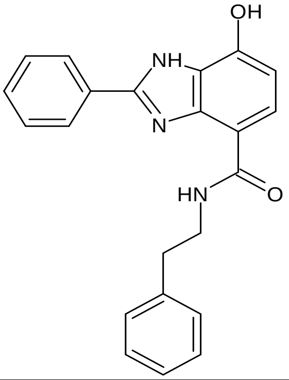
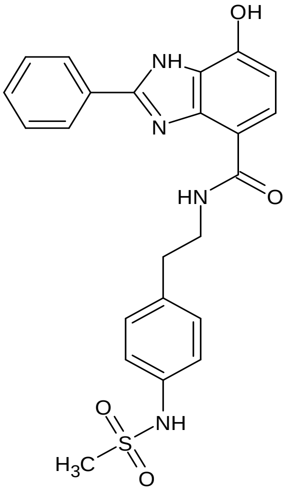
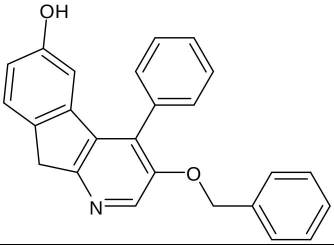
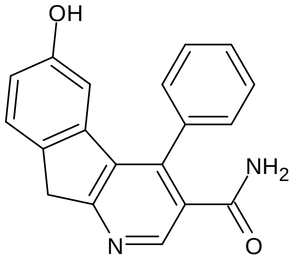
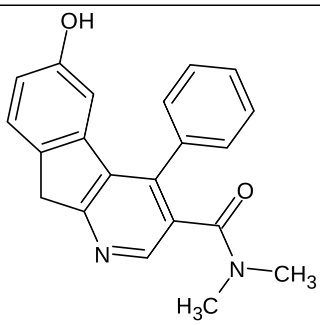
SN	Structure of the compounds	Actual pIC50	Predicted pIC50
1		-1.30103	-1.27507
2		0.455932	-2.94906

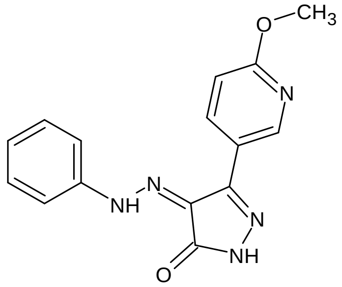
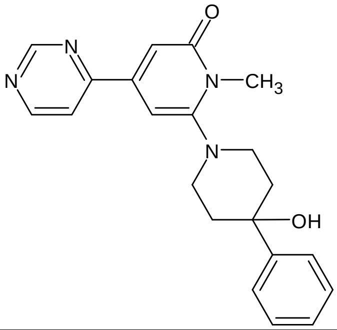
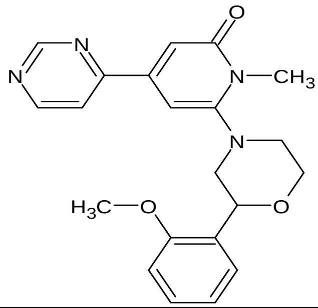
3		0.638272	-2.59976
4		-0.8451	-2.57599
5		-0.51851	-1.04211
6		-2.79239	-1.98943
7		-1.38021	-1.9894
8		-1	-1.7943

9		-1.25527	-2.29929
10		-1.53148	-0.10967
11		0.09691	0.108851
12		-1.04139	-1.35735
13		0.09691	-0.66613

14		-1.87506	-0.14005
15		-0.39794	-2.33035
16		-0.60206	-2.47494
17		-3.23045	-3.2014
18		-2.79796	-2.8862
19		-1.96379	-2.2231
20		-0.69897	-1.86114

21		-0.69897	-1.23586
22		-1.47712	-0.77889
23		-2	-1.98689
24		-1.81291	-1.39531
25		-1.64345	-1.59345
26		-0.49136	-1.03709
27		-1.72428	-2.42491
28		-1.25527	-1.79302
29		-2.61278	-2.72557

30		-3.05308	-2.05047
31		-2.76343	-2.62762
32		-1.39794	-3.01513
33		-3.76343	-2.46821
34		-3.61278	-3.70334
35		-3.17609	-0.90886

36		-0.65321	-1.0735
37		-0.92942	-3.74375
38		-1.20683	-1.47041

3.1. Model 1 (with 10 descriptors)

pIC50 = -5.47052 +2.60572 IC1 +1.64642 GATS2e +2.088 mindssC -0.01441 ATSC7s -13.5191 AVP-0 +0.16712 minssNH -0.15369 minaaN +0.01777 VR2_Dt +1.52684 MATS8s +0.04725 nAtomP.

[Internal Validation Parameters (IVP): SEE :0.58767, r^2 :0.70183, r^2 adjusted :0.66207, PRESS :25.90197, F :17.6534 (DF (Degree of Freedom) :10, 75); Leave-One-Out(LOO) Result: Q2 :0.60111, Average rm^2(LOO) :0.47316, Delta rm^2(LOO) :0.19941; External Validation Parameters (EVP)(Without Scaling):: r^2 :0.65711, r0^2 :0.65354, reverse r0^2:0.57488, RMSEP (Root Mean Square Error of Prediction) :0.63121, Q2f1/R^2(Pred) :0.68626, Q2f2 :0.65324; EVP (After Scaling): Average rm^2(test) :0.54109, Delta rm^2(test) :0.1653]

{GT acceptable criteria : Q^2: 0.60111 Passed (Q^2>0.5), r^2: 0.65711Passed (r^2>0.6), |r0^2-r'0^2|: 0.07866 Passed (|r0^2-r'0^2|<0.3), k: 0.99121 [(r^2-r0^2)/r^2] 0.00543 or k': 0.92437 [(r^2-r'0^2)/r^2] 0.12513 Passed}

3.2.Model 2 (with 11 descriptors)(Roy and Mitra 2011)

pIC50 = -6.19416 + 2.64534 IC1 +1.80112 GATS2e +2.20934 mindssC -0.01408 ATSC7s -12.84456 AVP-0 +0.19708 minssNH -0.1799 minaaN +0.02228 VR2_Dt +1.61577 MATS8s +0.05144 nAtomP -0.25802 nF12Ring.

[IVP:: SEE :0.55403, r^2 :0.73853, r^2 adjusted :0.69966, PRESS :22.71409, F :19.00109 (DF :11, 74); Leave-One-Out(LOO) Result :: Q2 :0.59013, Average rm^2(LOO) :0.47465, Delta rm^2(LOO) :0.12766; EVP(Without Scaling):: r^2 :0.66916, r0^2 :0.6592, reverse r0^2:0.61934, RMSEP:0.62738, Q2f1/R^2(Pred) :0.69006, Q2f2 :0.65743; EVP (After Scaling):: Average rm^2(test) :0.5594,Delta rm^2(test) :0.09565]

{GT acceptable criteria:: Q^2 : 0.59013 Passed, r^2: 0.66916 Passed, |r0^2-r'0^2|: 0.03986 Passed), k: 0.97891 [(r^2-r0^2)/r^2] 0.01489 or k': 0.93738 [(r^2-r'0^2)/r^2] : 0.07445Passed}

3.3.Model 3 (with 12 descriptors)

pIC50 = -7.11297 + 2.59644 IC1 -0.23252 minsssN +2.11949 GATS2e +2.03621 mindssC -0.01331 ATSC7s -10.98537 AVP-0 +0.24422 minssNH -0.16546 minaaN +0.02086 VR2_Dt +1.21546 MATS8s +0.03775 nAtomP -0.2656 nF12Ring.

[IVP:: SEE :0.53557, r^2 :0.75896, r^2 adjusted :0.71933, PRESS :20.93933, F :19.15421 (DF :12, 73); Leave-One-Out(LOO) Result :: Q2 :0.64714, Average rm^2(LOO) :0.53284, Delta rm^2(LOO) :0.14785; EVP (Without Scaling):: r^2 :0.72485, r0^2 :0.72259, reverse r0^2:0.67595, RMSEP:0.56733, Q2f1/R^2(Pred) :0.74655, Q2f2 :0.71988; EVP (After Scaling):: Average rm^2(test) :0.62788, Delta rm^2(test) :0.09232]

{GT acceptable criteria:: Q^2 : 0.64714 Passed, r^2: 0.72485 Passed, |r0^2-r'0^2| : 0.04664 Passed, k: 0.97418 [(r^2-r0^2)/r^2] 0.00311 or k' : 0.95766 [(r^2-r'0^2)/r^2] 0.06745 Passed}

3.4.Model 4 (with 13 descriptors)(Roy *et al.*, 2014)

pIC50 = -6.022 +2.58977 IC1 -0.24611 minsssN +1.93576 GATS2e -1.28529 GATS7v +2.08453 mindssC -0.0132 ATSC7s -10.29187 AVP-0 +0.24584 minssNH -0.15455 minaaN +0.02051 VR2_Dt +1.09995 MATS8s +0.03865 nAtomP -0.20946 nF12Ring.

[IVP::SEE :0.51741, r^2 :0.77811, r^2 adjusted :0.73804, PRESS :19.27568, F :19.42172 (DF :13, 72); Leave-One-Out(LOO) Result :: Q2 :0.69747, Average rm^2(LOO) :0.5901, Delta rm^2(LOO) :0.16211; EVP(Without Scaling):: r^2 :0.72304, r0^2 :0.71848, reverse r0^2:0.68331, RMSEP:0.57069, Q2f1/R^2(Pred) :0.74354, Q2f2 :0.71655; EVP(After Scaling):: Average rm^2(test) :0.62587, Delta rm^2(test) :0.08572]

{GT acceptable criteria : Q^2 : 0.69747 Passed, r^2: 0.72304 Passed, |r0^2-r'0^2| : 0.03516 Passed, k: 0.97811 [(r^2-r0^2)/r^2] 0.00632 or k' : 0.9528 [(r^2-r'0^2)/r^2]: 0.05495 Passed}

3.5.Model 5 (with 24 descriptors)(Dimitrov *et al.*, 2002)

pIC50 = -5.31625 - 0.09091 ALogP +2.23424 IC1 -0.21727 minsssN +0.98009 GATS2e +1.21443 mindssC -0.0104 ATSC7s -8.07662 AVP-0 +0.56286 MATS8s +0.15643 minssNH -2.0612 GATS7v -0.17167 C1SP2 +0.03561 nAtomP +0.01557 VR2_Dt -0.13713 minaaN -0.28466 nF12Ring +0.03612 AATSC4v -5.72469 VE1_D +3.51384 VE1_Dze +0.46693 ndsN +0.83305 SP-5 -0.11812 MDEC-22 -0.02118 MPC6 -0.17558 AATS4s -0.24231 nCl.

[IVP:: SEE :0.35722, r^2 :0.91039, r^2 adjusted :0.87514, PRESS :7.78406, F :25.82319 (DF :24, 61); Leave-One-Out(LOO) Result :: Q2 :0.80158, Average rm^2(LOO):0.73688, Delta rm^2(LOO):0.01789; EVP(Without Scaling):: r^2 :0.88284, r0^2 :0.86732, reverse r0^2:0.81182, RMSEP:0.39046, Q2f1/R^2(Pred) :0.87995, Q2f2 :0.86731; EVP (After Scaling):: Average rm^2(test) :0.75257, Delta rm^2(test) :0.11285]

{GT acceptable criteria:: Q^2: 0.80158 Passed, r^2: 0.88284 Passed, |r0^2-r'0^2|: 0.0555 Passed, k : 1.00114 [(r^2-r0^2)/r^2] 0.01758 or k' : 0.96682 [(r^2-r'0^2)/r^2] : 0.08045 Passed}.

The findings revealed that all models met the criteria for acceptance, however only Model 1 had the bare minimum of descriptors (Paola 2013). Model 5 showed the highest Q2 (0.80158) and R2 (0.88284) values. There were 24 descriptors in this model. More descriptors increase prediction noise. One description was assigned for every ten molecules according to a rule. In this case, Model 1 was regarded as the best model (Krstajic*et al.*, 2014). In Model 1, the model was predicted using just 10 descriptors. In this model IC1 (Information Content index), GATS2e (Geary autocorrelation of lag 2 weighted by Sanderson electronegativity), mindssC (Minimum atom-type E-State =C), minssNH (Minimum atom-type E-State: -NH-), VR2_Dt (normalised Randic-like eigenvector-based index from detour matrix), MATS8s (Moran autocorrelation of lag 8 weighted by I-state) and nAtomP (Number of atoms in the largest pi system) were positively contributed in the bioactivity (Zhang *et al.*, 2006) The model meets every requirement for validation, according to the validation parameter. Figure 1's curve between the actual and projected pIC50 values in the training and test set demonstrated that the difference between the two values was within acceptable bounds. According on the applicability domain study, two compounds were found to be outliers (Rucker *et al.*, 2007).The average R2, Q2 (LOO), and cRp2 values for the model are 0.15, -0.16, and 0.64, respectively, according to the YR test results (Table 3).

Table 3.Y Randomization (YR) Data of the best QSAR Model.

Avg R^2	Avg Q^2 (LOO)	cRp^2
0.15	-0.16	0.64

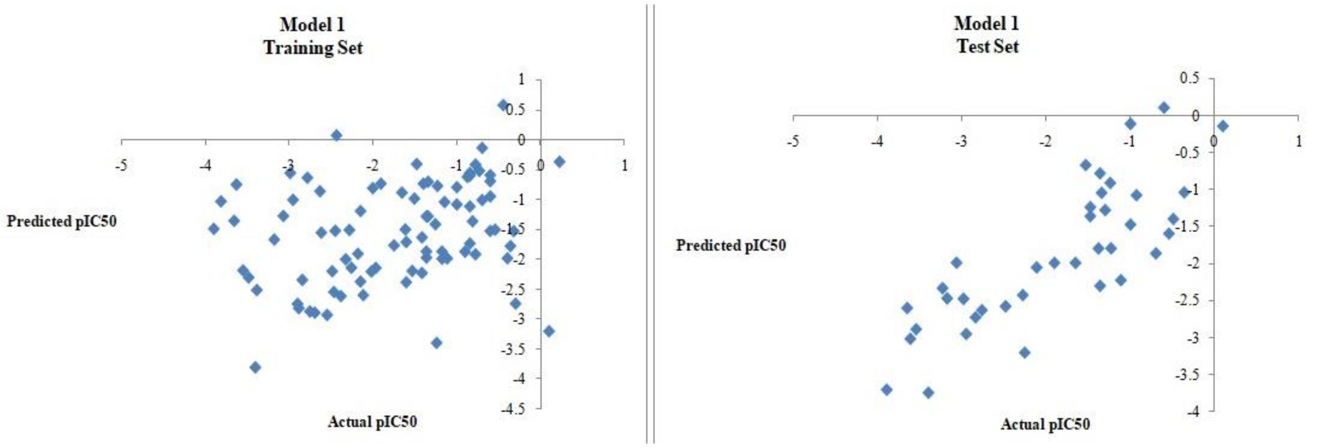


Figure 1 Graph between Actual and Predicted pIC50 values in Training Set and Test Set.

Conclusion

Here, we draw the conclusion that the established QSAR model will function as a good predictor with any chemical scaffold and descriptor combination in order to develop newer generation GSK-3 inhibitors targeting Alzheimer's disease.

List of Abbreviations

QSAR: Quantitative Structure Activity Relationship, DF: Degree of Freedom, SPRESS: Standard Deviation based on Predicted Residual Error Sum of Squares, SDEP: Standard Deviation of Error of Prediction, RMSEP: Root Mean Square Error of Prediction, LOO: Leave One Out, MLR: Multiple Linear Regression.

Acknowledgment: Authors conveyed special thanks to Mr. Jitender Joshi, Chancellor, and Prof. (Dr.) Dharam Buddhi, Vice Chancellor of Uttaranchal University-Dehradun encourages publishing this research work. We are grateful to the Division of Research and Innovation (DRI) and Central Instrumentation Facility (CIF), Uttaranchal University-Dehradun, India for providing all facilities during the experimental work.

Author Contributions: Data collection - Vikash Jakhmola, Arun Kumar Mahato, and Praveen Kumar Ashok; Data interpretation: Vishal Warikoo and Sayantan Mukhopadhay; QSAR study:;Prinsa; Supervision – Supriyo Saha.

Funding: The work is not funded by any organization.

Availability of Data and Material: All data used to support the findings of this study are included within the main text. Actual pIC50, Predicted pIC50 values and Y-Randomization Data of other models (Model 2, 3, 4 and 5) are included in Supplementary files (Table S1, Table S2 and Table S3).

Declaration: This is an original work and no part is submitted to other Journal.

Conflicts of Interest: There is no conflict of interest to declare.

References

Alamri M.A., Alawam A.S., Alshahrani, M.M., *et al.*, (2023) Establishing the Role of Iridoids as Potential Kirsten Rat Sarcoma Viral Oncogene Homolog G12C Inhibitors Using Molecular Docking; Molecular Docking Simulation; Molecular Mechanics Poisson–Boltzmann Surface Area; Frontier Molecular Orbital Theory; Molecular Electrostatic Potential; and Absorption, Distribution, Metabolism, Excretion, and Toxicity Analysis. *Molecules*., 28, 5050. https://doi.org/10.3390/molecules28135050

Akihiko T. (2006) GSK-3 is essential in the pathogenesis of Alzheimer's disease. *J. Alzheimers. Dis.*, 9(3), 309-317. https://doi.org/10.3233/jad-2006-9s335

Ambure P., and Roy K. (2016) Understanding the structural requirements of cyclic sulfone hydroxyethylamines as hBACE1 inhibitors against Aβ plaques in Alzheimer's disease: a predictive QSAR approach. *RSC. Adv.*, 6, 28171-28186. https://doi.org/10.1039/C6RA04104C

Angela De S., Vincenzo T., Vincenza, A., *et al.*,(2021) Glycogen Synthase Kinase 3β: A New Gold Rush in Anti-Alzheimer’s Disease Multitarget Drug Discovery? *J. Med. Chem.*, 64, 26-41. https://doi.org/10.1021/acs.jmedchem.0c00931

Ballabio D., Consonni V., Claeys-Bruno M., *et al.*, (2014) A novel variable reduction method adapted from space-filling designs. *Chemometric. Intelligent. Lab. Systems.*, 136, 147-154. https://dx.doi.org/10.1016/j.chemolab.2014.05.010

Chtita S., Hmamouchi R., Majdouline L.,*et al.*, (2016). QSPR studies of 9-aniliioacridine derivatives for their DNA drug binding properties based on density functional theory using statistical methods: Model, validation and influencing factors. *J. Taibah. University. Science.*, 10(6), 868-876. https://doi.org/10.1016/j.jtusci.2015.04.007

De Moura É.P., Fernandes N.D., Monteiro A.F.M.,*et al.*, (2021) Machine Learning, Molecular Modeling, and QSAR Studies on Natural Products Against Alzheimer's Disease. *Curr. Med. Chem.*, 28(38), 7808-7829. https://doi.org/10.2174/0929867328666210603104749

De Strooper B., and Karran E. (2016) The Cellular Phase of Alzheimer's Disease. *Cell.*, 164(4):603-15. doi: 10.1016/j.cell.2015.12.056.

Dimitrov S.D., Dimitrova N.C., Walker J.D., *et al.*, (2002) Predicting bioconcentration factors of highly hydrophobic chemicals. Effects of molecular size. *Pure. Appl. Chem.*,74, 1823-1830. https://doi.org/10.1351/pac200274101823

El Alaouy M. A., Moukhliss Y.,Boutalaka M.,*et al.*, (2021). The 2D-QSAR study, Drug likeness and in-silico ADMET prediction of about 3,5-diaryl-1H-pyrazole derivatives as multifunctional agents for the treatment of Alzheimer's disease. *RHAZES: Green. Appl. Chem.*,13, 09-28.

El Alaouy M. A., El Bahi S.,Boutalaka M.,*et al.*, (2023) Organic compounds based on 1-(prop-2-yn-1-ylamino)-2,3-dihydro-1H-indene-4-thiol as selective human monoamine oxidase B inhibitors. Quantitativeanalysis of structure-activity relationships and in-silico investigations. *Mor. J. Chem.*, 14(3), 802-817.

El khatabi K., Aanouz I., El-mernissi R.,*et al.*, (2021) In silico analysis of 3D QSAR and Molecular Docking studies to discover new thiadiazole-thiazolone derivatives as mitotic kinesin Eg5 inhibition. *Mor. J. Chem.*, 9(3), 394-405. https://doi.org/10.48317/IMIST.PRSM/morjchem-v9i2.18721

El-Mernissi R., El Khatabi K., Khaldan A.,*et al.*, (2023) New compounds based on 1H-pyrrolo[2,3-b] pyridine as potent TNIK inhibitors against colorectal cancer cells. Molecular modeling studies. *Mor. J. Chem.*, 11(01), 20-33. Doi: https://doi.org/10.48317/IMIST.PRSM/morjchemv11i1.35380

Golbraikh A., Muratov E., Fourches D.,*et al.*, (2014) Data set modelability by QSAR. *J. Chem. Inf. Model.*, 54, 1-4. https://doi.org/10.1021/ci400572x

Griebel G., Stemmelin J., Lopez-Grancha M., *et al.*, (2019) The selective GSK3 inhibitor, SAR502250, displays neuroprotective activity and attenuates behavioral impairments in models of neuropsychiatric symptoms of Alzheimer’s disease in rodents. *Sci. Rep.*, 9, 18045. https://doi.org/10.1038/s41598-019-54557-5

Hanieh A., Masood F., Zohreh M. (2022) Vilazodone-Tacrine Hybrids as Potential Anti-Alzheimer Agents: QSAR, Molecular Docking, and Molecular Dynamic (MD) Simulation Studies. *Biointerface. Research. Applied. Chemistry.*, 12(1), 588-607. https://doi.org/10.33263/BRIAC121.588607

Hassan N., Ossama D., Oussama A., *et al.*, (2022) Combined computational approaches for developing new anti-Alzheimer drug candidates: 3D-QSAR, molecular docking and molecular dynamics studies of liquiritigenin derivatives. *Heliyon*, 8(12), e11991. https://doi.org/10.1016/j.heliyon.2022.e11991

Hooper C., Killick R., Lovestone S. (2008) The GSK3 hypothesis of Alzheimer disease. *J. Neurochem.*, 104(6), 1433-1439. https://doi.org/10.1111/j.1471-4159.2007.05194.x

Kareti S.R., and Subash, P. (2020). In silico exploration of anti-Alzheimer’s compounds present in methanolic extract of NeolamarckiaCadamba bark using GC–MS/MS. *Arabian J Chemistry.*, 13(7), 6246–6255. https://doi.org/10.1016/j.arabjc.2020.05.035

Kim W.Y., Zhou F.Q., Zhou J.,*et al.*, (2006) Essential Roles for GSK-3s and GSK-3-Primed Substrates in Neurotrophin-Induced and Hippocampal Axon Growth. *Neuron.*, 52, 981-996. https://doi.org/10.1016/j.neuron.2006.10.031

Krstajic D., Buturovic L.J., Leahy D.E., *et al.*,(2014) Cross-validation pitfalls when selecting and assessing regression and classification models.*J. Cheminform.*,6, 1-15. https://doi.org/10.1186/1758-2946-6-10.

Kumar V., Saha, A., Roy K. (2023) Multi-target QSAR modeling for the identification of novel inhibitors against Alzheimer's disease. *Chemometrics. Intelligent. Laboratory. Systems.*, 233, 104734. https://doi.org/10.1016/j.chemolab.2022.104734

Limor A., and Eldar-Finkelman H. (2013) GSK-3 and lysosomes meet in Alzheimer's disease. *Communicative. Integrative. Biol.*, 6, 5. https://doi.org/10.4161/cib.25179

Ling Z., Shuli S., Chuan Q. (2013) The Role of HDAC6 in Alzheimer's Disease. *J. Alzheimers. Dis.*, 33, 283-295. https://doi.org/10.3233/JAD-2012-120727

Ma T. (2014) GSK3 in Alzheimer disease: mind the isoforms.*J. Alzheimers. Dis.*, 39(4), 707-710. https://doi.org/10.3233/JAD-131661

Pal D.K., Saha S. (2019) Chondroitin: a natural biomarker with immense biomedical applications. *RSC. Adv.*, 9(48), 28061-28077. https://doi.org/10.1039/C9RA05546K

Paola G. (2013) On the development and validation of QSAR models. *Methods. Mol. Biol.*, 930, 499-526. https://doi.org/10.1007/978-1-62703-059-5_21

Roy K., Das R.N., Paul L.A.P. (2014) Quantitative structure–activity relationship for toxicity of ionic liquids to Daphnia magna: Aromaticity vs. lipophilicity. *Chemosphere.*,112, 120-127. https://doi.org/10.1016/j.chemosphere.2014.04.002

Roy K., and Mitra I. (2011) On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Comb. Chem. High. Throughput. Screen.*,14, 450-474, https://doi.org/10.2174/138620711795767893

Rucker C., Rucker G., Meringer M. (2007) y-Randomization and its variants in QSPR/QSAR. *J. Chem. Inf. Model.*, 47, 2345-2357. https://doi.org/10.1021/ci700157b

Saha S., Pal D., Nimse S.B. (2022) Indazole Derivatives Effective against Gastrointestinal Diseases. *Curr. Top. Med. Chem.*, 22(14), 1189-1214. https://doi.org/10.2174/1568026621666211209155933

Saha S., Pal D., Nimse S.B. (2022) Indazole-Based Microtubule-Targeting Agents as Potential Candidates for Anticancer Drugs Discovery. *Bioorg. Chem.*, 122(3), 105735. https://doi.org/10.1016/j.bioorg.2022.105735

Saha S., Yeom G.S., Nimse S.B., Pal D. (2022) Combination Therapy of Ledipasvir and Itraconazole in the Treatment of COVID-19 Patients Coinfected with Black Fungus: An In-silico Statement.*Biomed. Res. Int.*, Article ID 5904261, 10 2022. https://doi.org/10.1155/2022/5904261

Tang M., Shi S., Guo Y., *et al.*, (2014) GSK-3/CREB pathway involved in the gx-50's effect on Alzheimer disease. *Neuropharmacology.*, 81, 256-266. https://doi.org/10.1016/j.neuropharm.2014.02.008

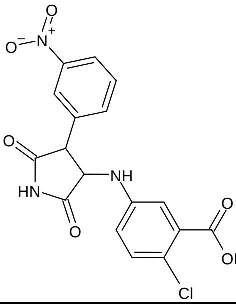
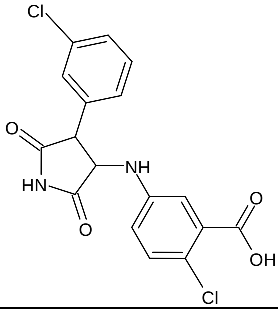
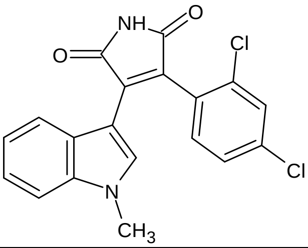
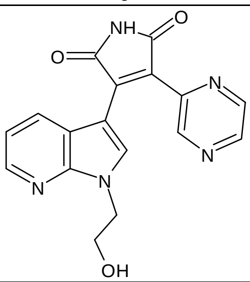
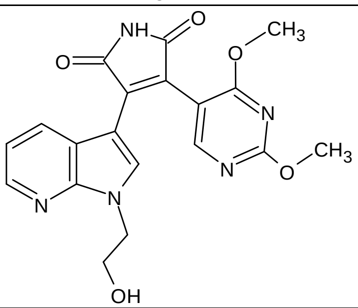
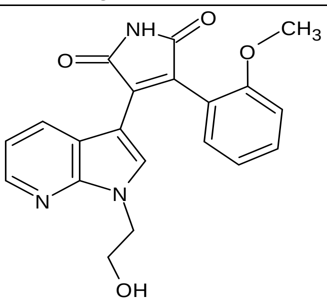
Thomas K., Boris S., Fabio Lo M. (2012) Small-Molecule Inhibitors of GSK-3: Structural Insights and Their Application to Alzheimer's Disease Models. *Int. J. Alzheimers. Dis.*, Article ID 381029, https://doi.org/10.1155/2012/381029

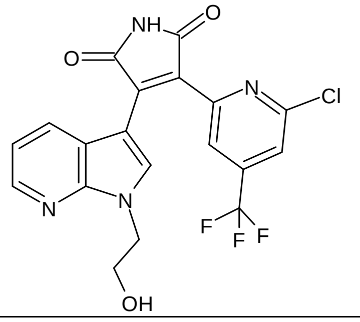
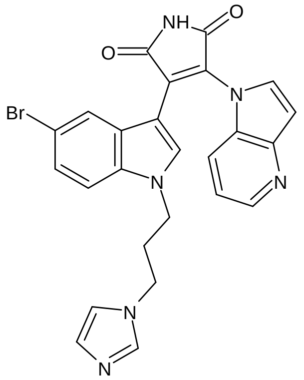
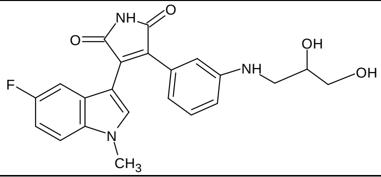
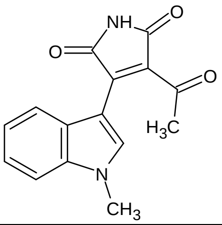
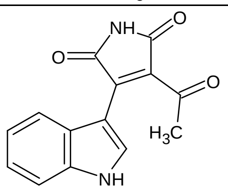
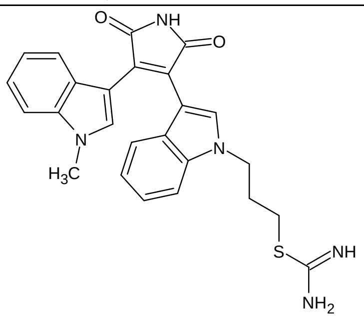
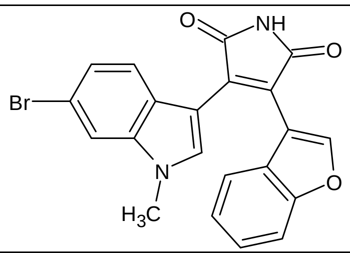
Toral-Rios D., Pichardo-Rojas P.S., Alonso-Vanegas M., *et al.*, (2020) GSK3β and Tau Protein in Alzheimer's Disease and Epilepsy. *Front. Cell. Neurosci.*, 14, 1-9. https://doi.org/10.3389/fncel.2020.00019

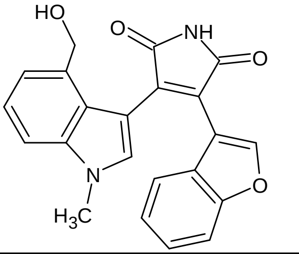
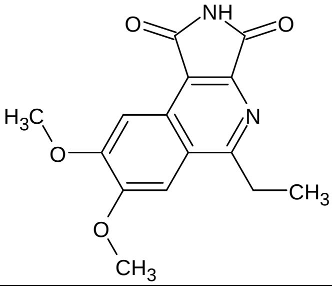
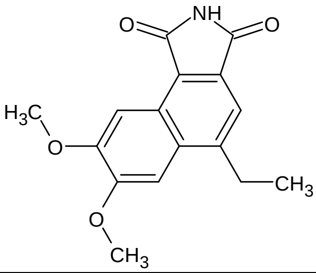
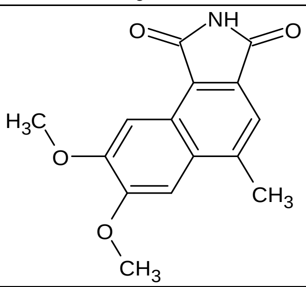
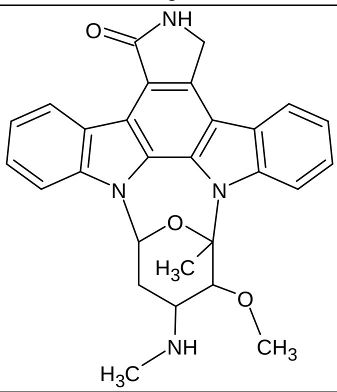
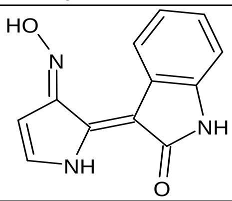
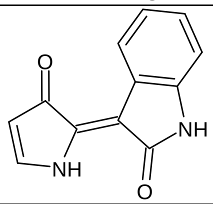
Yap, C.W. (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.*, 32, 1466-1474. https://doi.org/10.1002/jcc.21707

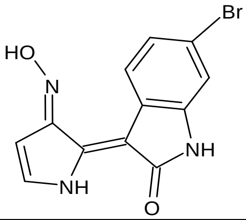
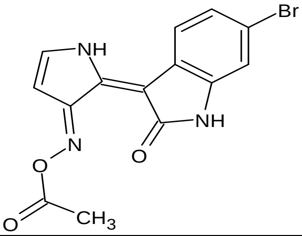
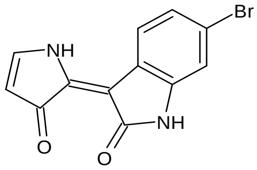
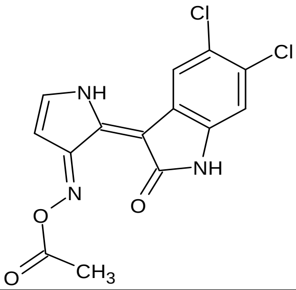
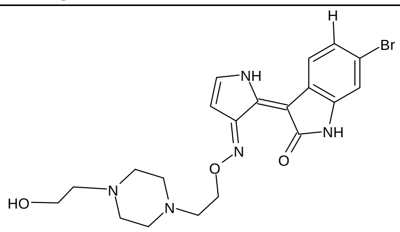
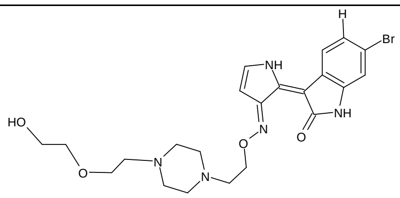
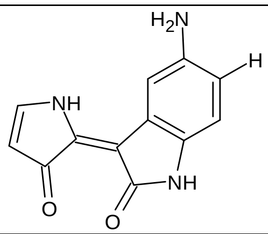
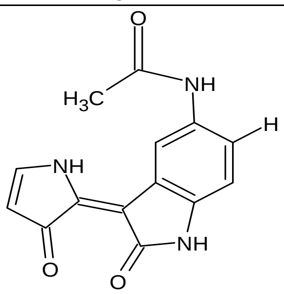
Zhang S., Golbraikh A., Oloff S.,*et al.*,(2006) A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. *J. Chem. Inf. Model.*,46, 1984-1995. https://doi.org/10.1021/ci060132x

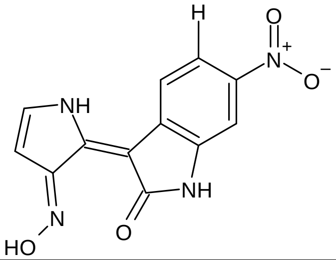
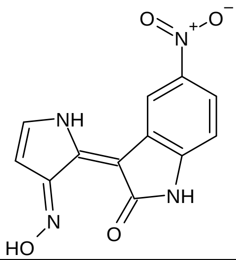
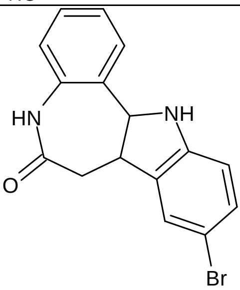
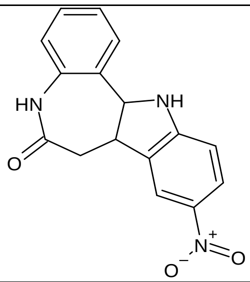
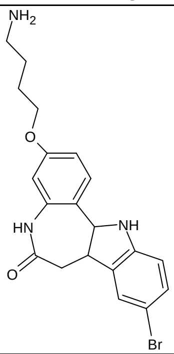
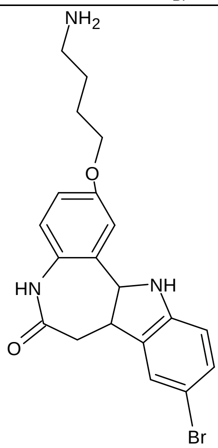
Table S1. Actual pIC50, and Predicted pIC50 Values of Training Set Molecules of different QSAR Models.

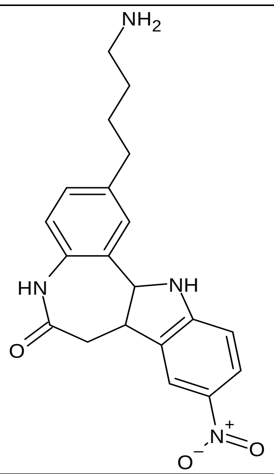
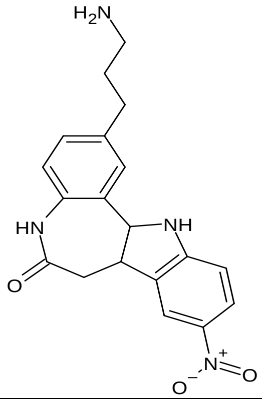
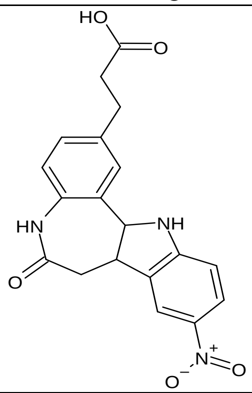
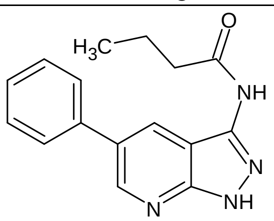
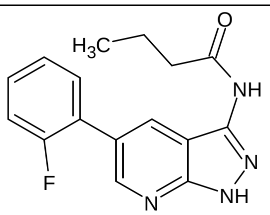
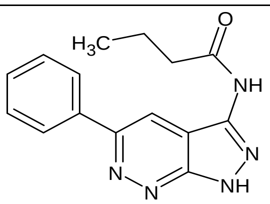
SN	Structure of the compounds	Actual pIC50	Predicted pIC50 (Model 2)	Predicted pIC50 (Model 3)	Predicted pIC50 (Model 4)	Predicted pIC50 (Model 5)
1		-1.41497	-2.30004	-2.23258	-2.14573	-1.40867
2		-2.89542	-2.82135	-2.6834	-2.66375	-2.5139
3		-1.53148	-2.13956	-2.03201	-1.96968	-1.85593
4		-1.39794	-0.79598	-0.72437	-0.85654	-1.45258
5		-0.77815	-0.43003	-0.33837	-0.54359	-0.68705
6		-0.60206	-0.5895	-0.48583	-0.55114	-0.85404

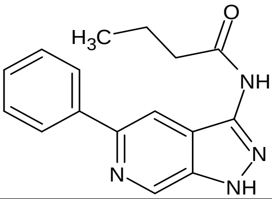
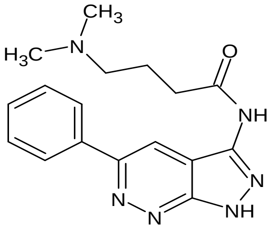
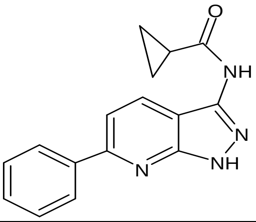
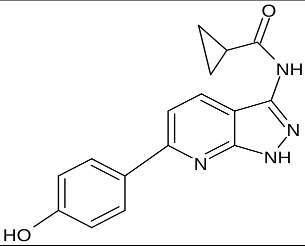
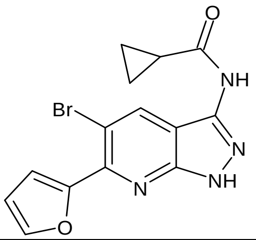
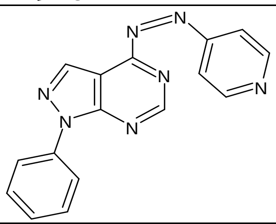
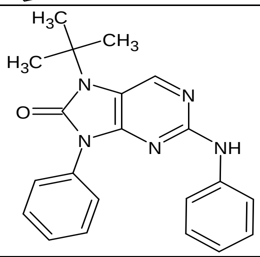
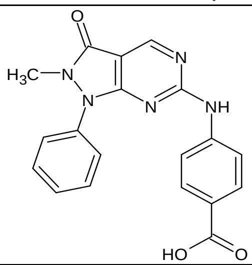
7		-1.41497	-1.72622	-1.73665	-1.77401	-1.22522
8		-2.14613	-1.15536	-0.99847	-0.93181	-1.75658
9		0.221849	-0.27469	-0.28134	-0.35449	0.109629
10		-2.94939	-0.91179	-0.86128	-0.73648	-0.72958
11		-3.65031	-1.1996	-1.08469	-1.00174	-0.5754
12		-0.44716	0.697029	0.59284	0.537598	0.654765
13		-0.8451	-0.47302	-0.4567	-0.43978	-0.34366

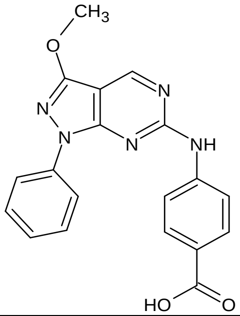
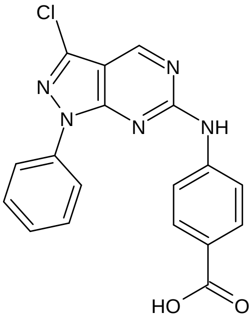
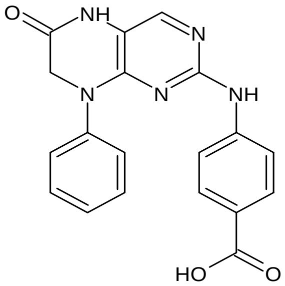
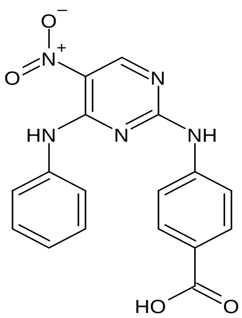
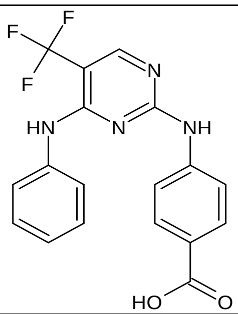
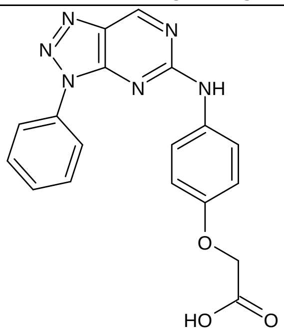
14		-0.73239	-0.40635	-0.29248	-0.33784	-0.19527
15		-2.48287	-2.04245	-1.81369	-1.8969	-2.36854
16		-1.96379	-2.01599	-1.83776	-1.94506	-2.3328
17		-2.43136	-1.44855	-1.36137	-1.18429	-1.09335
18		-1.17609	-1.86729	-1.86815	-2.13011	-2.45918
19		-1.34242	-1.20029	-1.07961	-1.13321	-0.92177
20		-2.77815	-0.51826	-0.58883	-0.60891	-0.84507

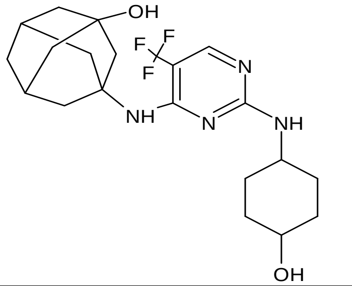
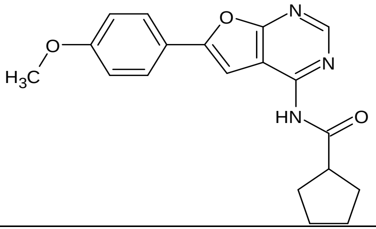
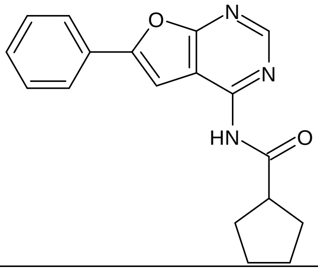
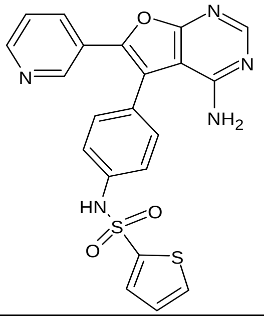
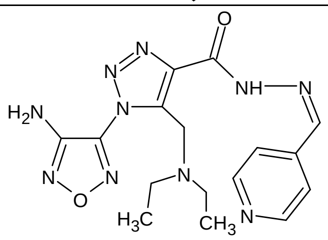
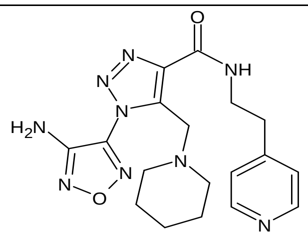
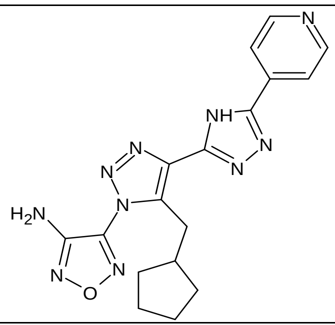
21		-0.69897	-0.02092	-0.11468	-0.17541	-0.83205
22		-1	-0.68463	-0.70411	-0.59629	-0.65531
23		-1.65321	-0.68596	-0.91526	-0.75859	-0.36744
24		-0.60206	-0.7294	-0.9344	-0.83984	-0.65749
25		-0.69897	-0.79755	-0.97665	-1.03631	-1.25777
26		-1.14613	-1.00397	-1.07686	-1.0799	-1.27856
27		-1.90309	-0.73288	-0.8216	-1.01674	-1.18456
28		-0.87506	-0.6296	-0.70327	-0.76289	-0.63109

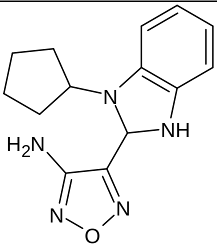
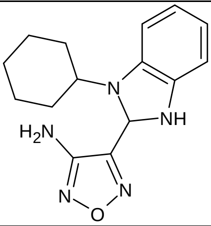
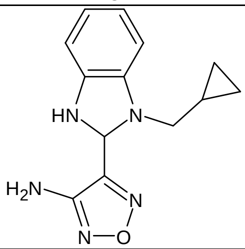
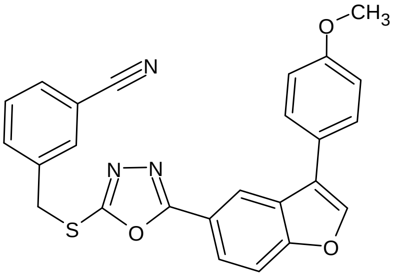
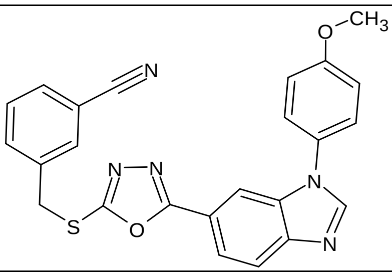
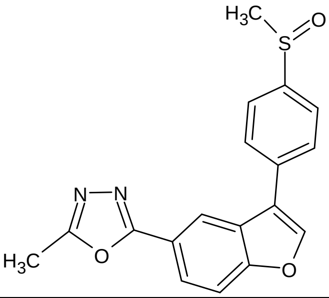
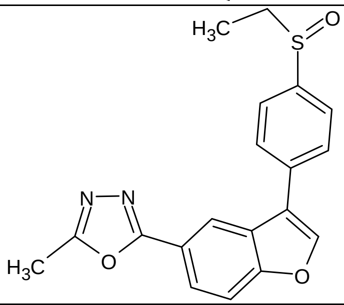
29		-1.60206	-2.32736	-2.17871	-2.16424	-2.3064
30		-0.32222	-1.36718	-1.04414	-0.95986	-1.68108
31		-1.36173	-1.89785	-1.59642	-1.38647	-1.28267
32		-0.60206	-0.63048	-0.70215	-0.70853	-0.89312
33		-1.47712	-0.32444	-0.42398	-0.40326	-0.28892
34		-1.60206	-1.78293	-1.77262	-1.76719	-1.21492

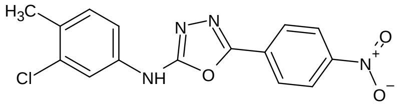
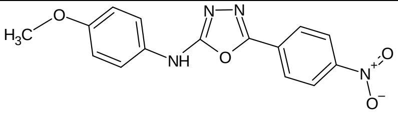
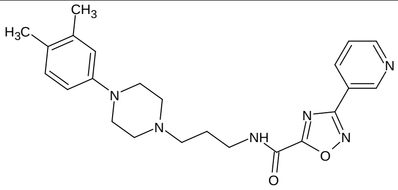
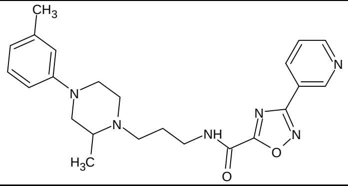
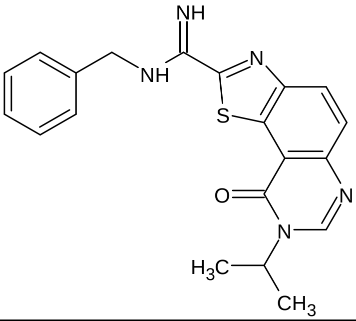
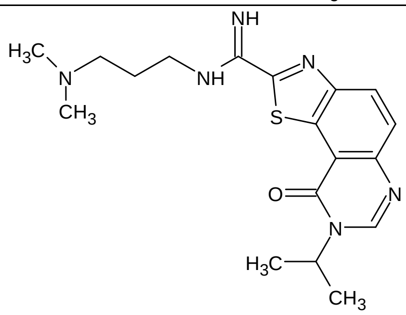
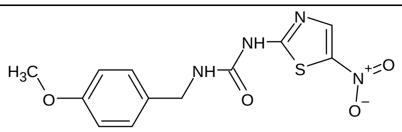
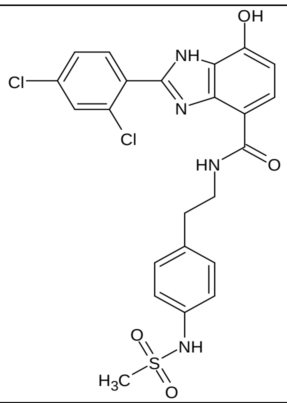
35		-0.77815	-1.85024	-1.78648	-1.70026	-1.66258
36		-0.39794	-1.99331	-1.88426	-1.78381	-1.54936
37		-0.81291	-1.28162	-1.20721	-1.11966	-1.19969
38		-1.74819	-1.70293	-1.66593	-1.64999	-1.16942
39		-1.25527	-1.32476	-1.27739	-1.29819	-1.20384
40		-0.60206	-1.50572	-1.8366	-1.79823	-1.26106

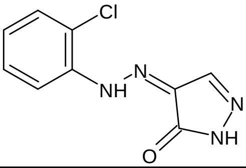
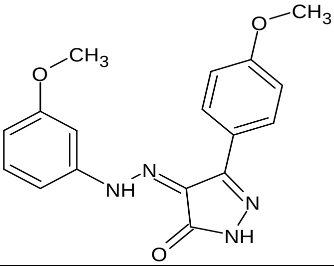
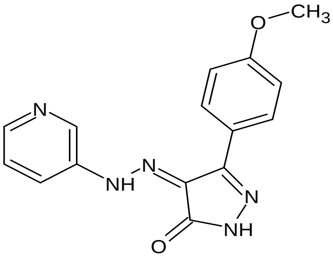
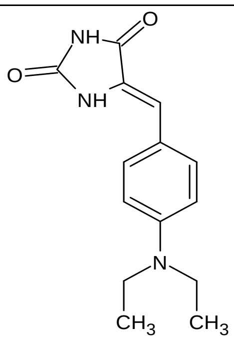
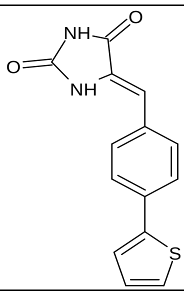
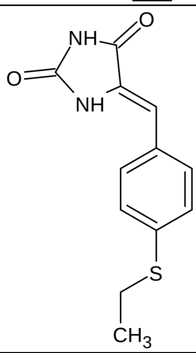
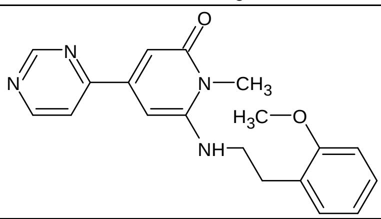
41		-0.8451	-1.73161	-1.72174	-1.83262	-2.13273
42		-1.34242	-0.6871	-0.72391	-0.84536	-0.91578
43		-2.62839	-0.7903	-0.72911	-0.72825	-0.17177
44		-0.90309	-1.81215	-1.70345	-1.77353	-1.40783
45		-0.8451	-0.51051	-0.34936	-0.43911	-0.64388
46		-2	-0.79856	-0.87639	-0.78181	-0.0336
47		-3.38021	-2.57317	-2.56155	-2.73292	-2.40363
48		-2.98227	-0.50783	-0.65366	-0.9334	-0.6203

49		-2.88081	-2.81615	-3.16829	-3.42202	-3.71252
50		-2.25527	-2.2175	-2.19284	-2.4026	-2.19672
51		-3.54407	-2.20601	-2.30905	-2.3306	-2.52371
52		-2.17609	-1.9437	-2.02975	-2.03443	-2.5492
53		-2.6902	-2.88344	-2.83131	-2.7947	-2.52929
54		-1.11394	-2.12724	-2.06828	-2.10328	-2.14187

55		-1.61278	-1.61625	-1.56539	-1.40164	-1.59834
56		-1.50515	-0.84986	-0.72522	-0.86441	-0.80867
57		-1.36173	-1.85979	-1.79061	-1.90179	-1.38761
58		-1.36173	-1.25793	-1.1872	-1.34339	-0.96918
59		-2.61278	-1.5973	-1.79574	-1.82646	-2.47975
60		-3.06446	-1.33031	-1.61328	-1.56254	-1.92922
61		-2.44716	-1.61252	-1.63734	-1.52088	-2.17555

62		-2.32222	-1.85875	-2.06846	-2.12803	-2.33823
63		-2.38021	-2.55919	-2.74464	-2.81479	-2.17512
64		-2.4624	-2.61129	-2.67446	-2.69233	-2.69243
65		-0.54407	-1.50621	-1.464	-1.50507	-1.30741
66		-0.36173	-1.78588	-1.73863	-1.74121	-1.70141
67		-2.14613	-2.41121	-2.48671	-2.38873	-2.67373
68		-2.27875	-1.42259	-1.50215	-1.15754	-1.59148

69		-1.23045	-0.67261	-0.68706	-0.44822	-1.04778
70		-0.8451	-1.07109	-1.14375	-0.87355	-1.11622
71		-2.54407	-2.93449	-3.13921	-3.1271	-3.17411
72		-2.83885	-2.31807	-2.455	-2.28096	-2.52259
73		-2.11394	-2.6011	-2.83325	-2.65245	-2.63491
74		-2.74819	-2.81163	-2.74266	-3.01938	-2.53783
75		-2.01703	-2.23854	-2.01232	-2.0085	-2.44876
76		-1.17609	-2.0615	-1.94474	-1.86403	-1.55231

77		-3.17319	-1.72148	-1.55145	-1.3515	-1.21064
78		0.09691	-3.22932	-3.16331	-3.19711	-3.84867
79		-0.30103	-2.70091	-2.4354	-2.73939	-2.95431
80		-3.62325	-0.59672	-0.52066	-0.35065	0.246198
81		-3.80618	-0.97199	-0.83399	-0.76394	-0.55527
82		-3.89209	-1.46444	-1.316	-1.30007	-0.86088
83		-1.24055	-3.40718	-3.76988	-3.57262	-3.54276

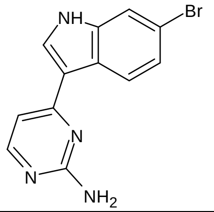
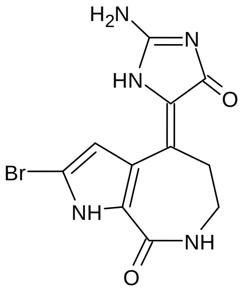
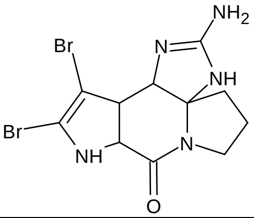
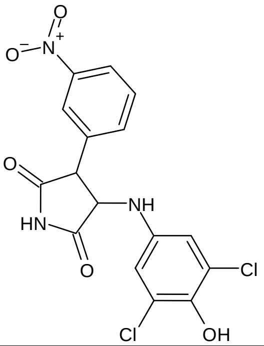
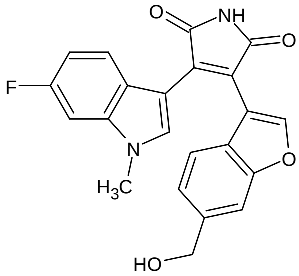
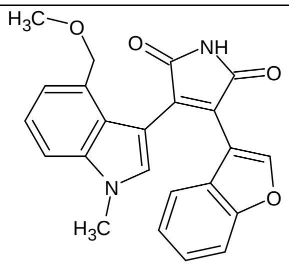
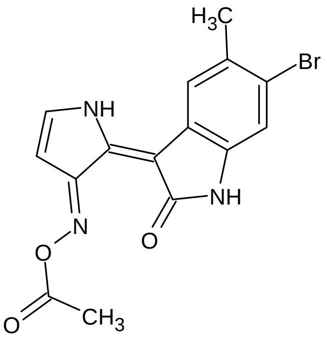
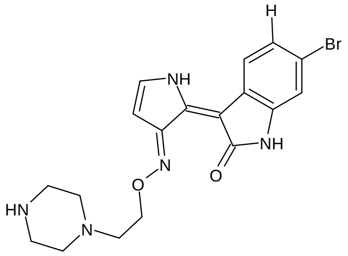
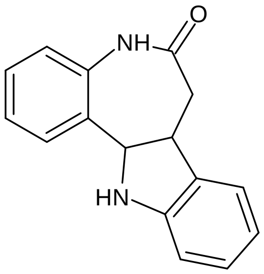
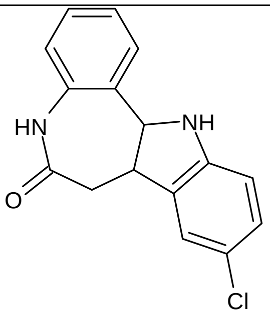
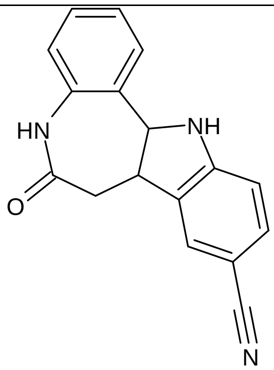
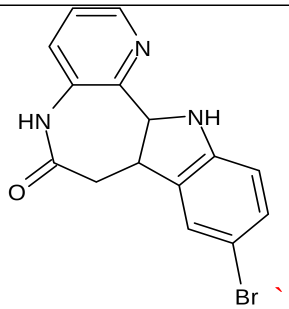
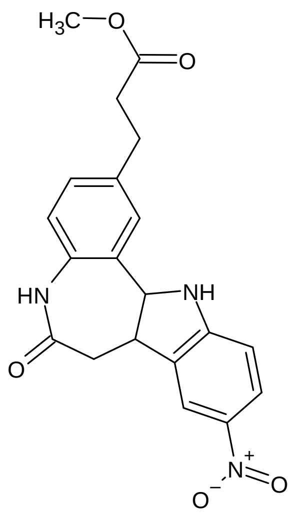
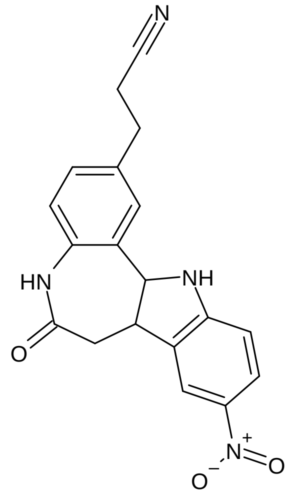
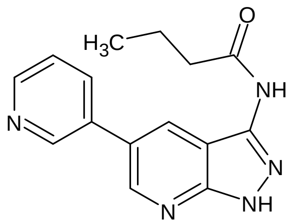
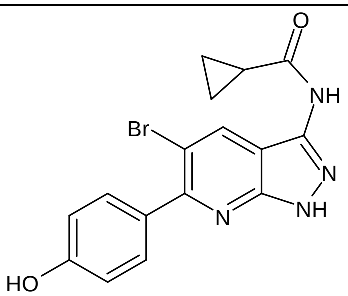
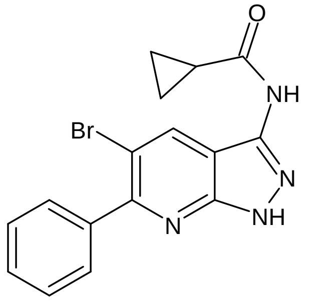
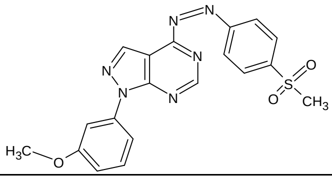
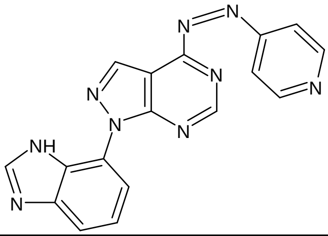
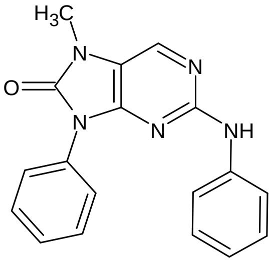
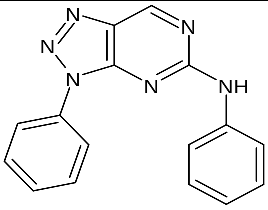
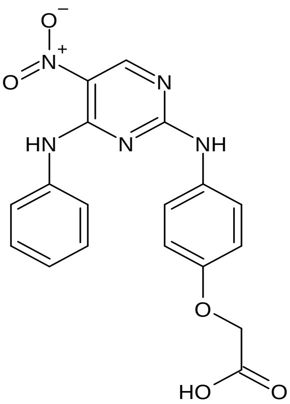
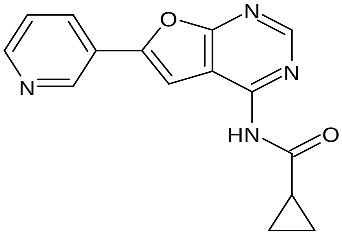
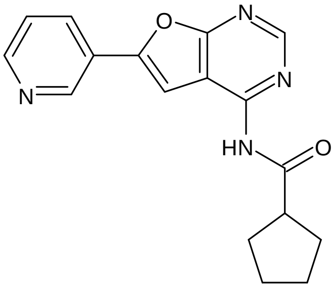
84		-3.39794	-3.86152	-3.7999	-3.51148	-3.4637
85		-1	-1.07088	-1.25888	-1.1567	-1.215
86		-3.47712	-2.84139	-3.0448	-3.45799	-3.67018

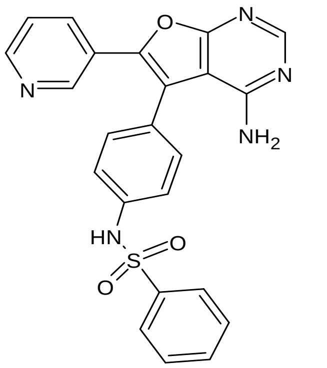
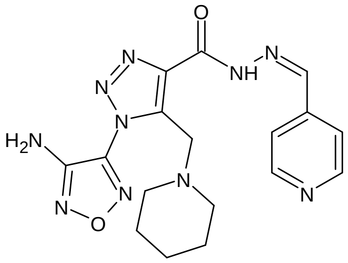
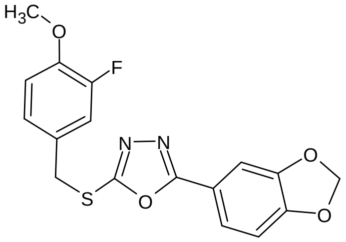
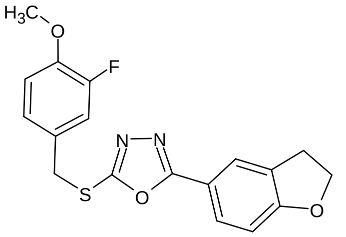
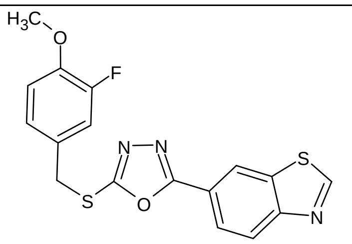
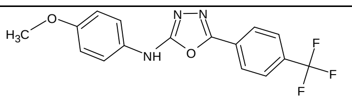
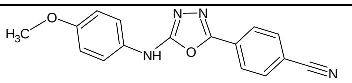
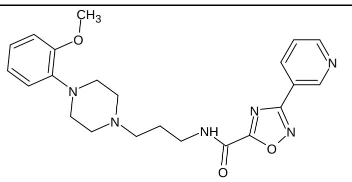
Table S2. Actual pIC50, and Predicted pIC50 Values of Test Set Molecules of different QSAR Models.

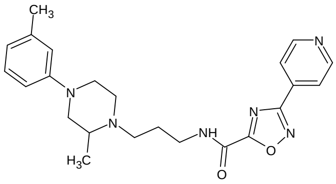
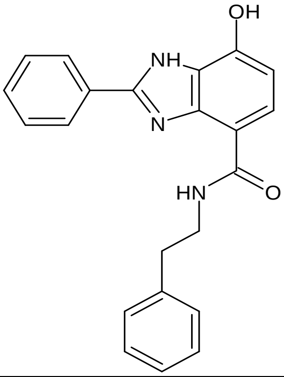
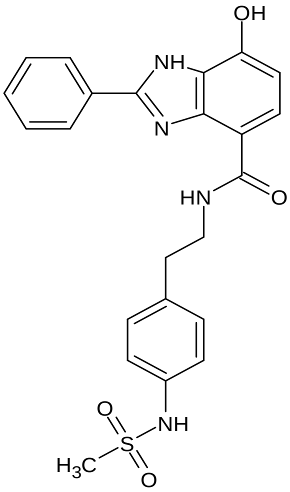
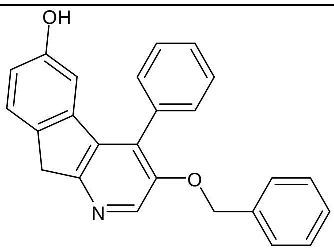
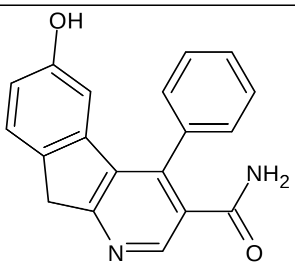
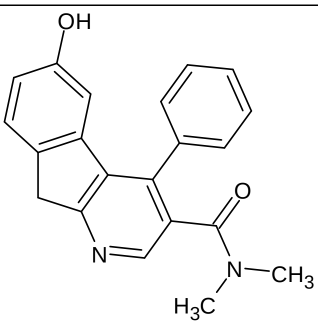
SN	Structure of the compounds	Actual pIC50	Predicted pIC50 (Model 2)	Predicted pIC50 (Model 3)	Predicted pIC50 (Model 4)	Predicted pIC50 (Model 5)
1		-1.30103	-1.24499	-1.18999	-1.05578	-0.90191
2		0.455932	-3.02138	-2.79518	-2.94851	-2.7294
3		0.638272	-2.67948	-2.60372	-2.81586	-3.19567

4		-0.8451	-2.55217	-2.23694	-2.25148	-2.35823
5		-0.51851	-1.04091	-1.06345	-1.31319	-1.41323
6		-2.79239	-1.91673	-1.79193	-1.84845	-2.06799
7		-1.38021	-2.01779	-1.90029	-1.91985	-1.39755
8		-1	-1.65137	-1.4939	-1.31735	-1.5387
9		-1.25527	-2.12677	-1.89394	-1.69193	-1.57186

10		-1.53148	0.094285	-0.06494	0.007229	-0.77778
11		0.09691	0.178822	-0.10486	-0.05533	-0.89018
12		-1.04139	-1.21624	-0.88358	-0.81981	-1.61719
13		0.09691	-0.64458	-0.65798	-0.7253	-1.0553
14		-1.87506	-0.09312	-0.21414	-0.17276	-0.39538

15		-0.39794	-2.31872	-2.72759	-2.70075	-3.17392
16		-0.60206	-2.58918	-2.96401	-2.94737	-2.89934
17		-3.23045	-3.31936	-3.35144	-3.44064	-3.05172
18		-2.79796	-3.02355	-3.4654	-3.42876	-3.05327
19		-1.96379	-2.33595	-2.23061	-2.2877	-1.46029
20		-0.69897	-1.85979	-1.79061	-1.90179	-1.38761
21		-0.69897	-1.22213	-1.40194	-1.55047	-1.73706

22		-1.47712	-0.74298	-0.87667	-1.08011	-1.76131
23		-2	-2.04707	-2.22989	-2.1214	-2.70784
24		-1.81291	-1.38933	-1.37071	-1.29737	-0.88844
25		-1.64345	-1.5615	-1.59767	-1.51346	-1.28357
26		-0.49136	-1.00044	-1.0599	-0.93604	-0.43445
27		-1.72428	-2.47081	-2.54255	-2.33287	-2.4352
28		-1.25527	-1.78717	-1.81222	-1.38524	-1.4032
29		-2.61278	-2.7362	-2.95038	-2.75703	-2.64634

30		-3.05308	-2.04621	-2.17407	-2.43196	-2.41212
31		-2.76343	-2.71455	-2.56637	-2.48332	-2.57254
32		-1.39794	-3.20082	-3.38694	-3.53726	-3.28303
33		-3.76343	-2.53459	-2.99086	-3.22618	-2.69573
34		-3.61278	-3.75633	-3.54664	-3.19798	-3.0588
35		-3.17609	-0.84491	-0.90655	-0.63841	-1.53777

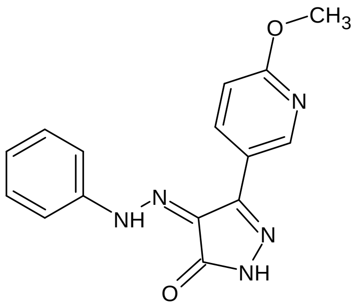
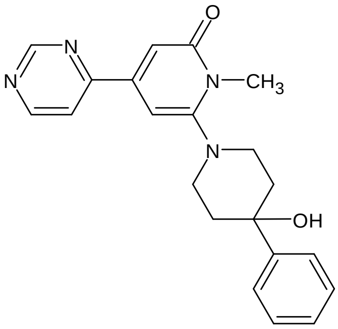
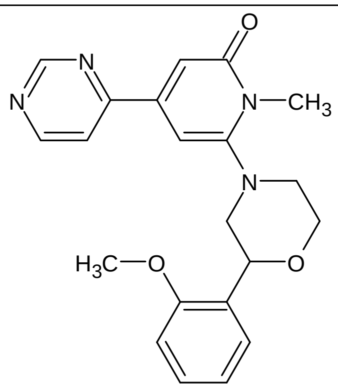
36		-0.65321	-1.07088	-1.25888	-1.1567	-1.215
37		-0.92942	-3.85206	-3.67773	-3.54236	-3.79259
38		-1.20683	-1.40269	-1.37178	-1.62418	-1.34722

Table S3. YR Data of different QSAR Models.

Model 2			Model 3			Model 4			Model 5		
Avg R^2	Avg Q^2 (LOO)	cRp^2	Avg R^2	Avg Q^2 (LOO)	cRp^2	Avg R^2	Avg Q^2 (LOO)	cRp^2	Avg R^2	Avg Q^2 (LOO)	cRp^2
0.14	-0.23	0.66	0.15	-0.24	0.68	0.15	-0.32	0.70	0.29	-0.68	0.75

(2023) ; https://revues.imist.ma/index.php/morjchem/index