IJPSR (2023), Volume 14, Issue 3 (Research Article)

INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES AND RESEARCH

Received on 15 July 2022; received in revised form, 19 August 2022; accepted 01 September 2022; published 01 March 2023

2D QSAR APPROACH TO DEVELOP NEWER GENERATION SMALL MOLECULES ACTIVE AGAINST SARS-COVID

Supriyo Saha [*1] and Dilip Kumar Pal [2]

School of Pharmaceutical Sciences & Technology [1], Sardar Bhagwan Singh University, Dehradun - 248161, Uttarakhand, India.
Department of Pharmaceutical Sciences [2], Guru Ghasidas Vishwavidyalaya (A Central University), Bilaspur - 495009, Chhattisgarh, India.

Keywords:

SARS-CoV-2, Modelability index, Kennard stone method, Multiple linear regression, Golbraikh and tropsha acceptable model criteria, Y Randomization test

Correspondence to Author:
Supriyo Saha

School of Pharmaceutical Sciences & Technology, Sardar Bhagwan Singh University, Dehradun -248161, Uttarakhand, India.

E-mail: supriyo9@gmail.com

ABSTRACT: We are in the half past of 2022, but still, we are facing the coronavirus pandemic situation. When a patient is hospitalized, only some FDA-approved drugs were administered to cure the patient. In treating coronavirus infection, nitazoxanide, granulocyte-macrophage colony-stimulating factor inhibitors, and various monoclonal antibodies are present. But all the molecules used in the treatment were not so effective in fully curing the patient. So, to break this jinx to develop of newer generation anti-SARS-CoV-2 drug molecules, computational approaches played an essential role. 2D QSAR studies related to anti-SARS-CoV-2 molecule development, some QSAR models observed with good statistical parameters such as R^2: 0.748, cross-validated Q^2 (LOO): 0.628, external predicted R^2: 0.723 and another model suggested with R^2: 0.764, Q^2: 0.627 and R_m^2: 0.610, Q^2 (F_1): 0.727, Q^2 (F_1): 0.652, MAE score: 0.127. We developed a new 2D QSAR model with a higher number of molecules and greater statistical parameters. A dataset of 84 anti-SARS-CoV2 molecules was obtained from literature followed by descriptor calculation PADEL software; the QSAR model was generated using the Modelability index, dataset pretreatment, division, MLR equation, validation, and Y randomization test. The model was pIC50 = -1.79268(+/-0.3652) +0.07995(+/-0.03551) naaaC -0.4051(+/-0.09672) nsssN -0.45945(+/-0.11025) SHsOH +1.23189(+/-0.28144) ETA_BetaP with R^2 and Q^2 values were 0.87028 and 0.70493 with MAE fitness score value: 0.14298. Atoms E-state and electronic features of the molecules directly related to anti-SARS-CoV-2 drug activity. It can be easily concluded that we want to develop a small molecule effective against SARS-CoV-2 disease in the near future.

INTRODUCTION: We are in the half past of 2021, but still, we are facing the coronavirus pandemic situation. Till date, globally 19, 2054, 106 cases were registered and among them, 41, 280, 58 deaths were reported (https://www.worldometers.info/coronavirus/ accessed on 25.07.2021).

QUICK RESPONSE CODE

DOI: 10.13040/IJPSR.0975-8232.14(3).1372-91

This article can be accessed online on www.ijpsr.com

DOI link: http://dx.doi.org/10.13040/IJPSR.0975-8232.14(3).1372-91

Presently, FDA approved three vaccines were approved in US such as Pfizer-BioNTech, Moderna, Johnson & Johnson's Janseen; whereas in other countries, oxford AstraZeneca, Sputnik V, Sinopharm, covaxin, *etc*. vaccines were approved [1, 2].

As per world data, 26.8% of the world population was vaccinated with the first dose, and 13.4% of the world population was vaccinated with both doses. As per WHO, we were already in the middle of the third wave of coronavirus pandemic situations [3, 4]. WHO declared SARS-CoV-2 variants into three main categories: variant of

interest, variants of interest and variants of high consequence [5]. In US, variants like B.1.1.7 (α), B.1.351 (β), B.1.617.2 (δ) and P.1 (γ) were concerned as variants of concern without any variants of high consequence [6, 7]. Variants of interest were certain genetic markers that hampered viral transmission, diagnosis, and response towards therapeutic dose; variants of concern created a high impact on the response towards diagnosis, treatment regime with a possible increment of viral transmission, disease condition, and response towards vaccination whereas variants with less response towards diagnosis, treatment and vaccination [8, 9].

B.1.427, B.1.429, B.1.525, B.1.526, B.1.617.1, B.1.617.3 were categorized under variants of interest and B.1.1.7, B.1.351, B.1.617.2, P.1 were categorized under variants variant of concern. In contrast, no variants were considered under variants of high consequence (https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html accessed on 29.07.2021) [10]. When a patient is hospitalized, only some FDA-approved drugs were administered to cure the patient. Nowadays, in the treatment of coronavirus infection, nitazoxanide, granulocyte-macrophage colony-stimulating factor inhibitors (Gimsilumab, lenzilumab, namilumab, and otilimab), chloroquine, hydroxychloroquine, azithromycin, colchicine and various monoclonal antibodies (casirivimab, imdevimab) were used in the treatment (https://www.covid19treatmentguidelines.nih.gov/about-the-guidelines/whats-new/ accessed on 30.07.2021).

But all the molecules used in the treatment were not so effective in fully cure the patient. So, to break this jinx to develop of newer generation anti-SARS-CoV-2 drug molecules, computational approaches such as molecular docking studies, molecular dynamics simulation studies, and QSAR (2D or 3D) studies play an essential role [11, 12]. In the category of 2D QSAR studies related to anti-SARS-CoV-2 drug development [13], some QSAR models observed with good statistical parameters such as R^2: 0.748, cross-validated Q^2 (LOO): 0.628, external predicted R^2: 0.723[14] and another model suggested with R^2: 0.764, Q^2: 0.627 (56 molecules in the training set) and R_m^2: 0.610, Q^2 (F_1): 0.727, Q^2 (F_1): 0.652, MAE score: 0.127 (13 molecules in the test set)[15]. In this context, we developed a new 2D QSAR model with a higher number of molecules and validated statistical parameters.

MATERIALS AND METHODS:

Dataset and Descriptor Calculation: A dataset of 84 anti-SARS-CoV2 drug molecules obtained from different literatures, including the top twenty potential anti-SARS-CoV-2 drugs obtained from 1553 FDA approved drugs [16], top 20 Potential Anti-SARS-CoV-2 drugs separated from 7012 investigational or off-market drug molecules [17], existing protease inhibitors, some polyamines targeting cellular attachment and entry of coronavirus [18], SARS-Cov-2 M^{pro} inhibitors [19], recently developed novel coronavirus (2019-nCoV) inhibitors and inhibitors of coronavirus main protease $3CL_{pro}$[20]. All the molecules were drawn by ACD ChemSketch software followed by saved as MDL Mol format. Then two-dimensional descriptors were calculated of the molecules using PADEL descriptor [21]. All the molecular descriptors along with their corresponding biological activities were tabulated in CSV format, where IC_{50}values were changed into pIC_{50} values [22].

Modelability Index: Modelability index is an estimation feasibility tool defined by the ratio between activity class-weighted ratio of the number of nearest-neighbor pairs of compounds corresponding with same activity class and the total number of pairs [23]. This concept correlated with the unnecessary efforts of a QSAR dataset associated with the development of QSAR model.

Descriptor Pretreatment: Then very closely related descriptors present within the dataset were removed by considering variance cut off and correlation coefficient values of 0.0001 and 0.99, respectively [24].

Dataset Division: Generally, the dataset was divided into training and test sets using Kennard Stone, Random Faster, and Euclidean Distance methods. Among them, here we considered the Kennard Stone method to divide the dataset of 84 molecules into training and test set. After the dataset division, 63 and 21 molecules were present in the training and test set, respectively [25-26].

Suitable Descriptor Selection: Suitable descriptor combination was selected using Stepwise MLR software with F values ranging from 3.9 to 4.0. The best subset combination was observed with 4 descriptors set with R^2 cut-off value was 0.6 [27-28].

Stepwise Regression: Stepwise multiple linear regression equation was built by multistep equation development involving three distinct steps: identification of an initial model, repetition of the previous step to achieve a better F and R^2 value, and calibration of model [24].

The stepwise regression equation was developed using statistical SPSS software, and parameters were judged as explained variance (R^2_a), correlation coefficient (R), standard error of estimate (s), and variance ratio (F) with a specified DF. Finally, the LOO method validated the model with cross-validation R^2 (Q^2), SPRESS, and SDEP parameters [29].

QSAR Equation Development: The final QSAR model was developed by Multiple Linear Regression Plus valid software with all possible combinations of descriptors based on the quality of prediction and MAE-based fitness score [30-31].

QSAR Equation Validation: The developed QSAR model was validated by the Golbraikh and Tropsha acceptable model criteria. The acceptable model criteria were as follows [32]:

1. Threshold value Q^2 greater than 0.5.

2. Threshold value R^2 greater than 0.6.

3. Threshold value $|r0^2\text{-}r'0^2|$ less than 0.3.

4. Threshold value: [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].

QSAR Equation Validation: QSAR model was cross-validated using LOO process. Applicability domain of the model was checked by euclidean distance and mahalanobis distance methods. The distance of a test set to its nearest neighbor in the training set was compared with predefined applicability domain threshold value [33-34].

MLR Y Randomization Test: In the Y randomization test, a random multiple linear regression model was developed by a faster random technique by changing the dependent variable and making the independent variable static. The model with significantly higher R^2 and Q^2 values after several trials confirmed that the developed model was robust and reproducible [35-36]. Another parameter, ${}^cR_p{}^2$ was also calculated which should be more than 0.5 for passing this test [37].

RESULTS AND DISCUSSION: Initially, the model ability index value was checked. The data showed a 0.5926 value with 29 molecules in high total active/less and 55 molecules in less total active/toxic with a threshold value of model ability index 0.65. So model ability index value of the model was 0.5926, which reflected that the dataset was quite close to developing a good QSAR model [38-39]. Then the dataset was divided into training and test sets using the Kennard-Stone method. Among them, 63 and 21 molecules were present in training and test sets, respectively [40]. Stepwise multiple linear regression was used to identify the most probable set of descriptors to build a good QSAR model based on MAE value [41]. After that, the data pretreatment process was also performed with variance cut-off and inter-correlation cut-off values 0.001 and 0.9, respectively, along with F value within (3.9-4.0) [42]. Then, using the best possible set of descriptors best subset selection process was performed with R^2 cut-off value 0.6 and inter-correlation between descriptors R^2 cut-off value 0.5 [43]. Based on BAD, MODERATE, and GOOD MAE fitness scores, the final QSAR model was generated.

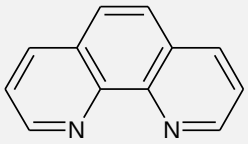
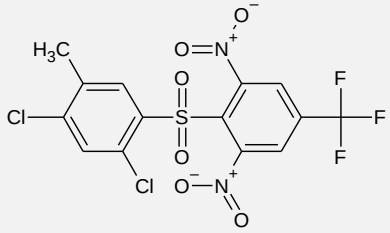
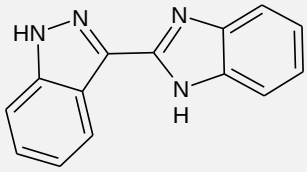
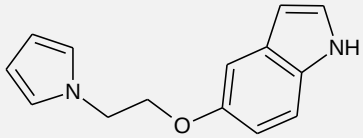
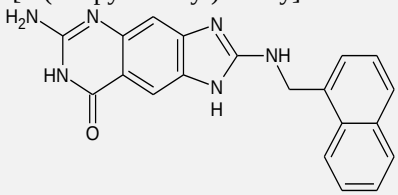
The Final QSAR Model was as follows:

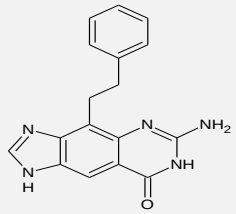
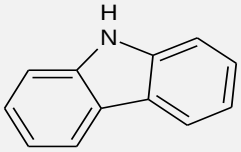
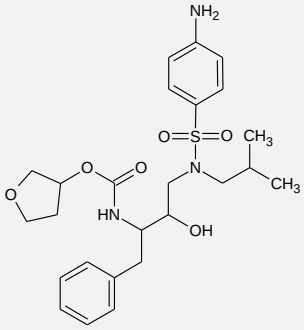
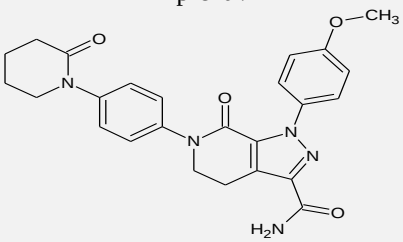
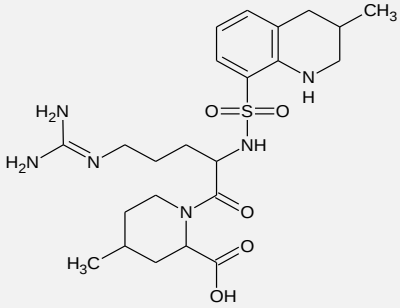
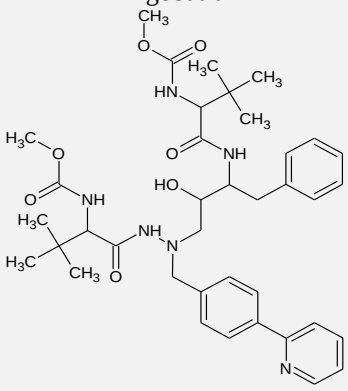
pIC_{50} = -1.79268(+/-0.3652) +0.07995(+/-0.03551) naaaC -0.4051(+/-0.09672) nsssN -0.45945(+/-0.11025) SHsOH +1.23189(+/-0.28144) ETA_BetaP.

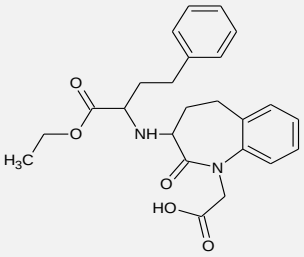
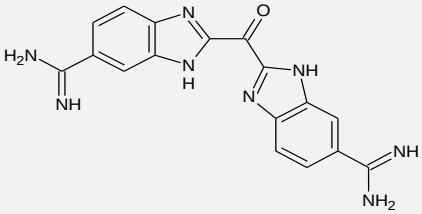
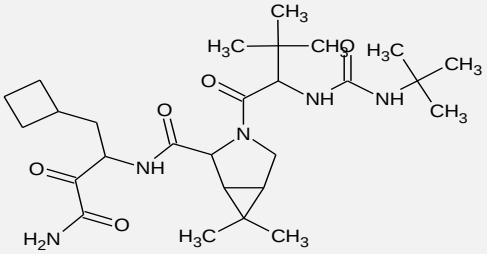
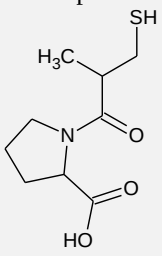
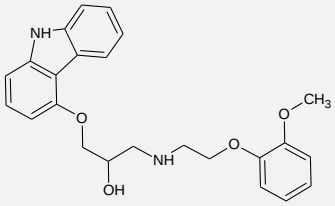
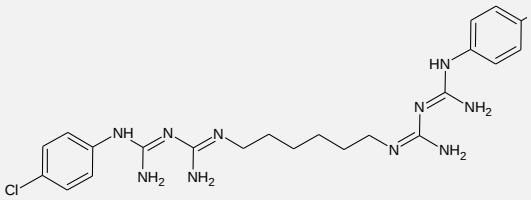
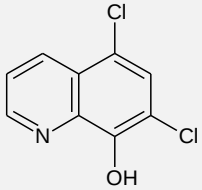
As per the model, naaaC, ETA_BetaP positively contributed to pIC50 value, whereas nsssN, SHsOH negatively contributed to pIC50. As per the internal validation parameter, SEE, R^2, R^2 adjusted, PRESS value, and F values were 0.38202, 0.75862, 0.74197, 8.46426, and 45.57092, respectively. As per LOO values Q^2, average Rm^2 and delta Rm^2 were 0.70493, 0.59478, and 0.1822, respectively. As per external validation parameters (without scaling)

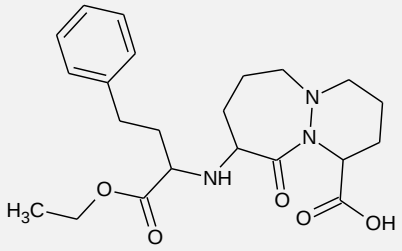
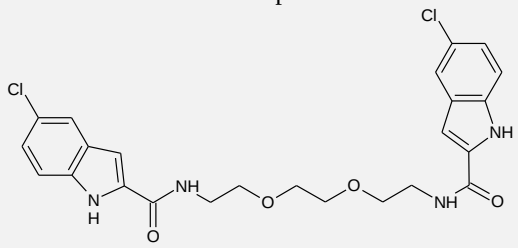
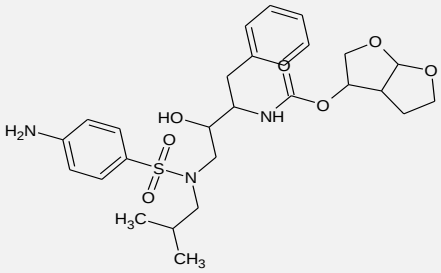
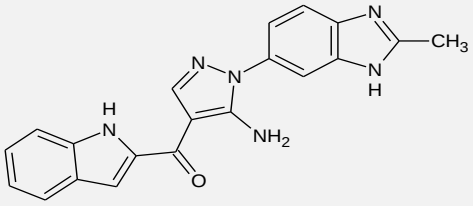
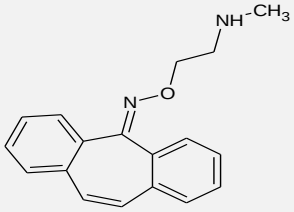
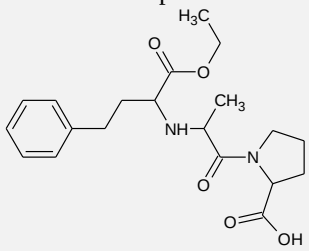
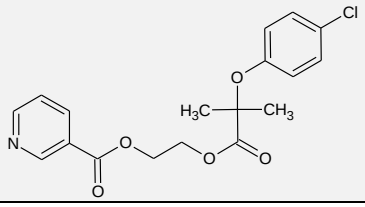
and after scaling R^2, $R0^2$, reverse $R0^2$, RMSEP, $Q2f1/R^2(Pred)$, Q^2f2 were 0.87028, 0.86122, 0.87026, 0.22536, 0.86012, 0.85943 and average Rm^2 (test), delta Rm^2 (test) were 0.82128, 0.00268; respectively. As per the error-based judgment related to testing set prediction: MAE, standard deviation values for 95% of data were 0.14298 and 0.10498, which correspond with GOOD prediction criteria [44]. Then the model was validated using Golbraikh and Tropsha acceptable model criteria. Outcomes showed that Q^2, R^2, $|r0^2-r'0^2|$, k, $[(r^2-r0^2)/r^2]$, k' and $[(r^2-r'0^2)/r^2]$ values were 0.70493, 0.87028, 0.00904, 0.96872, 0.01041, 0.95525 and 0.00002, respectively [45]. The model successfully passed all the validation criteria per the validation parameter. In training set data, the residual values between actual and predicted pIC50 values were between 0.007 to 1.10 Table 1. In the case of the test set, residual values oscillated between 0.001 to 0.5 **Table 2** [46]. The R^2 values observed after plotting actual pIC50 and predicted pIC50 values in training **Fig. 1** and test set were 0.7586 and 0.8703, respectively **Fig 2**. All the molecules in training and test sets were observed within the applicability domain. As per the Y randomization test data of the model, the average R, R^2, Q^2 (LOO), and $^cR_p^2$ values were 0.233071617, 0.070303736, -0.100697065 and 0.733948537, respectively **Table 3**.

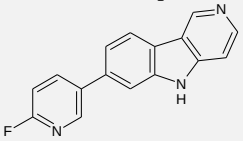
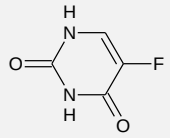
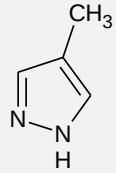
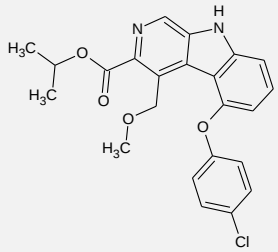
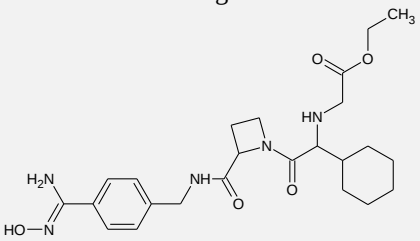
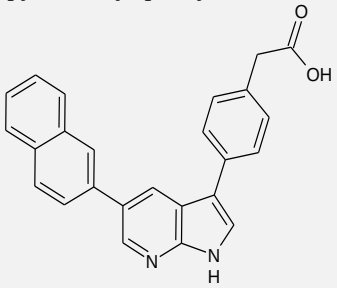
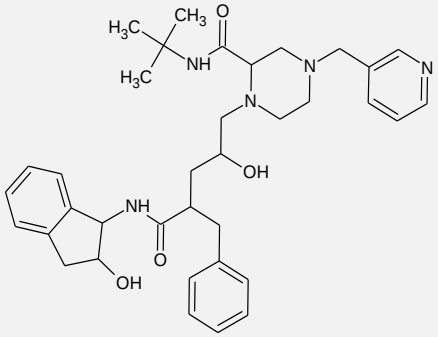
TABLE 1: ACTUAL PIC50, PREDICTED PIC50 AND RESIDUAL VALUES OF TRAINING SET MOLECULES

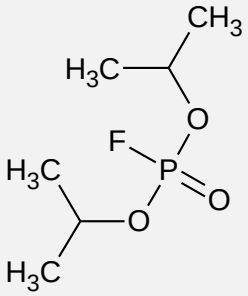
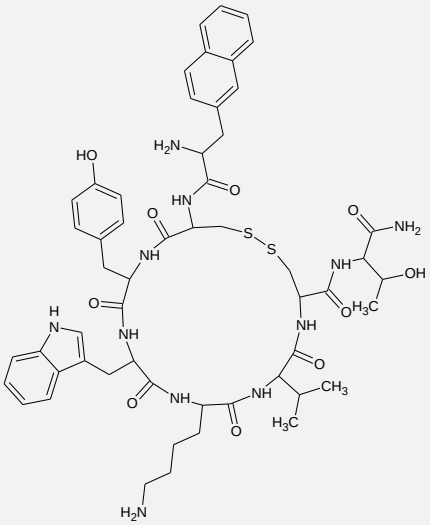
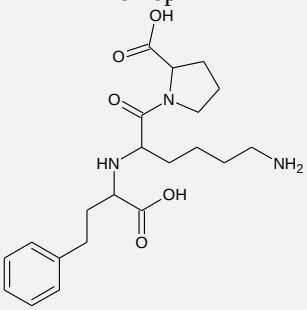
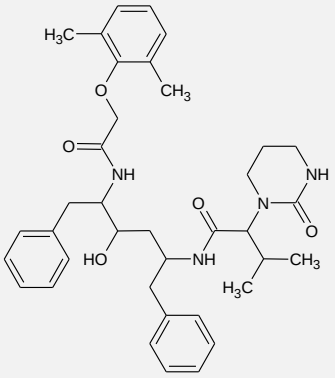
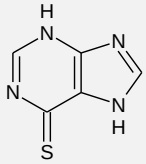
S. no.	Structures of Molecules	Actual pIC_{50} value	Predicted pIC_{50} value	Residual value
1.	 1,10-phenanthroline	0.3467	0.5509	0.20425
2.	 2-(2,4-dichloro-5-methylbenzene-1-sulfonyl)-1,3-dinitro-5-(trifluoromethyl)benzene	0.5228	-0.07682	0.59962
3.	 3-(1H-benzimidazol-2-yl)-1H-indazole	0.6197	0.5802	0.03941
4.	 5-[2-(1H-pyrrol-1-yl)ethoxy]-1H-indole	0.4089	0.2331	0.1757
5.	 6-amino-2-{[(naphthalen-1-yl)methyl]amino}-1,7-dihydro-8H-imidazo[4,5-g]quinazolin-8-one	0.3187	0.39778	0.079083

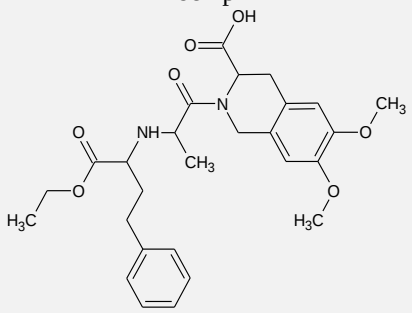
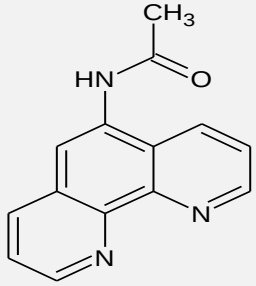
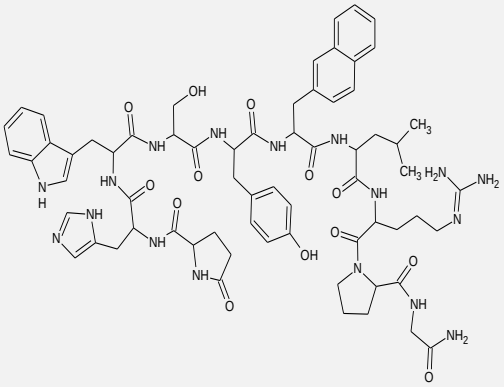
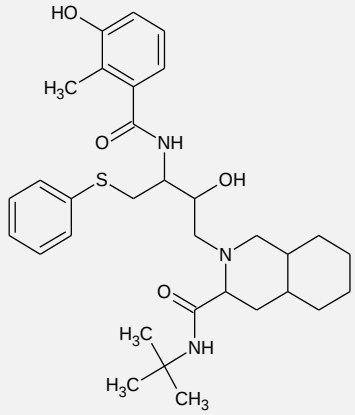
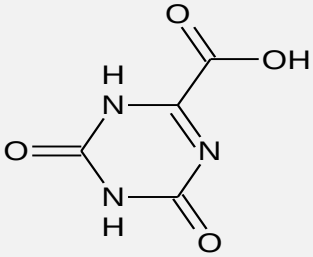
6.	 6-amino-4-(2-phenylethyl)-1,7-dihydro-8H-imidazo[4,5-g]quinazolin-8-one	0.3187	0.16150	0.15719
7.	 9H-carbazole	0.5686	0.517104	0.051495
8.	 Amprenavir	-0.9925	-0.9352	0.05727
9.	 Apixaban	-1.275	-0.98146	0.293506
10.	 Argobatran	-0.8438	-1.1904	0.34660
11.	 Atazanavir	-0.8426	-0.8826	0.04005

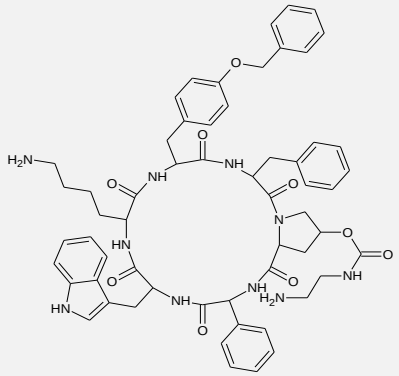
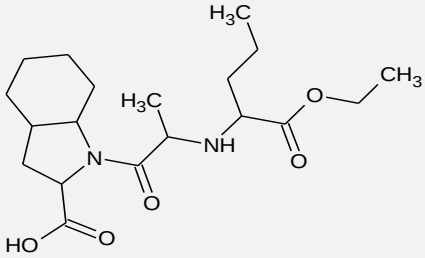
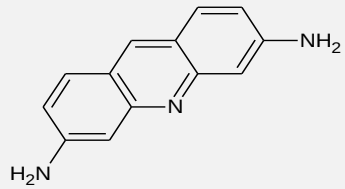
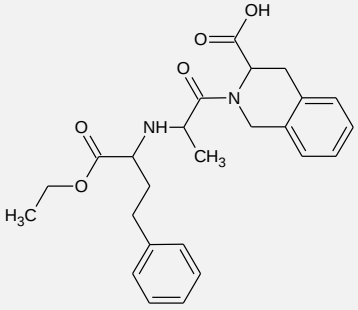
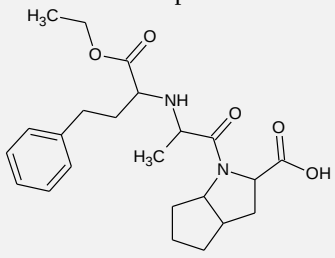
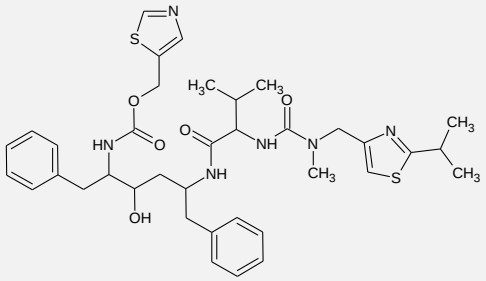
12.	 Benazepril	-1.002	-1.0494	0.047493
13.	 Bis-carbonyldi(1H-benzimidazole-6-carboximidamide)	0.3098	0.43419	0.12439
14.	 Boceprevir	-1.5928	-1.1739	0.41881
15.	 Captopril	-1.0674	-1.4833	0.41590
16.	 Carvedilol	-0.0864	-0.01634	0.07005
17.	 Chlorhexidine	-0.1303	-0.27093	0.14063
18.	 Chlorhexine	0.0506	-0.00618	0.056787

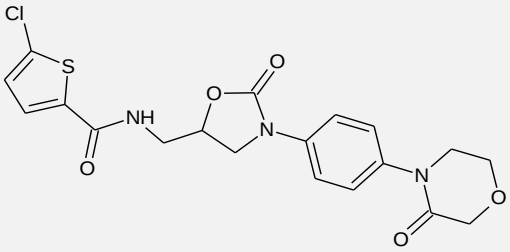
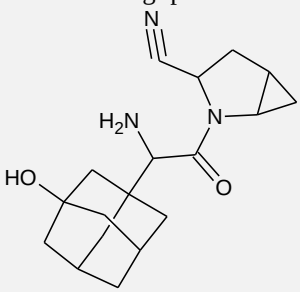
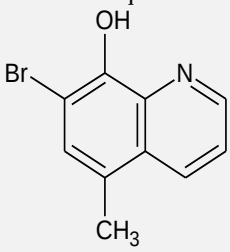
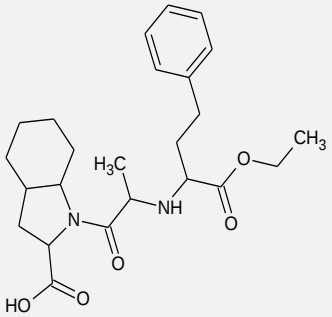
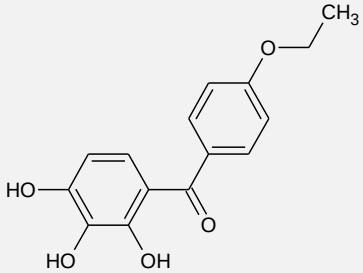
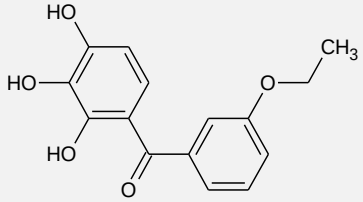
19.	 Cilazapril	-1.3685	-1.6344	0.26596
20.	 CP-526423	0.301	0.1394	0.161544
21.	 Darunavir	-0.7443	-0.9494	0.20519
22.	 Debio-1347	0.6198	0.50043	0.11936
23.	 Demexiptiline	-0.0253	-0.164820	0.13952
24.	 Enalapril	-1.1116	-1.2231	0.11152
25.	Etofibrate 	-0.1703	-0.2281	0.05787

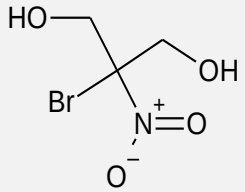
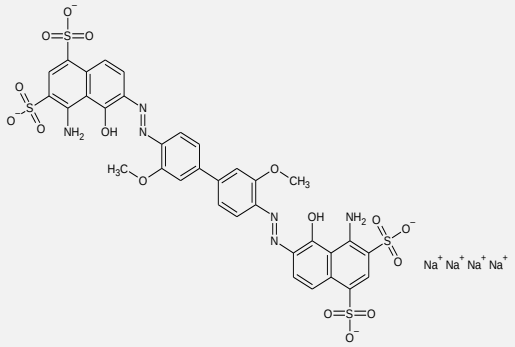
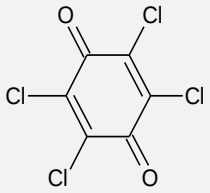
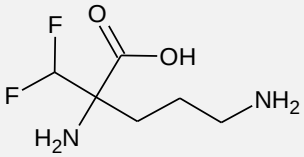
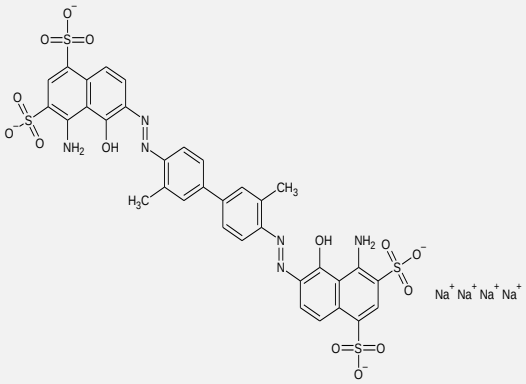
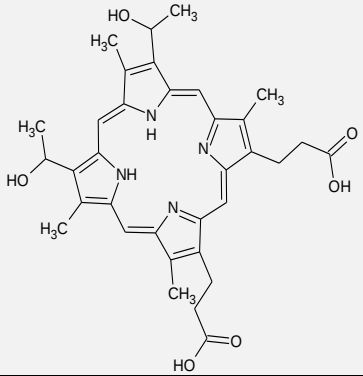
26.	Flortaucipir 	0.3767	0.544354	0.167654
27.	Flurouracil 	-0.0453	-0.32126	0.275958
28.	Fomepizole 	-0.143	-0.04749	0.095505
29.	Gedocarnil 	0.3564	0.210721	0.145679
30.	Ximelagatran 	-1.4173	-1.32573	0.091571
31.	{4-[5-(naphthalen-2-yl)-1H-pyrrolo[2,3-b]pyridin-3-yl]phenyl}acetic acid 	0.2596	0.112025	0.147575
32.	Indinavir 	-1.5173	-1.40489	0.112411

33.	Isoflurophate 	-0.7979	-0.89676	0.098861
34.	Lanreotide 	-0.1703	-0.44481	0.274513
35.	Lisinopril 	-1.3399	-1.48444	0.144543
36.	Lopinavir 	-1.2284	-0.85492	0.37348
37.	Mercaptopurine 	-0.1003	0.116753	0.217053

38.	Moexipril 	-1.1939	-1.04484	0.149065
39.	N-(1,10-phenanthrolin-5-yl)acetamide 	0.4815	0.426302	0.055198
40.	Nafarelin 	-0.1303	-0.75815	0.627854
41.	Nelfinavir 	-1.5591	-1.17597	0.383127
42.	Oteracil 	-0.0645	-0.61538	0.550878

43.	Pasireotide	-0.1673	-0.542	0.374703
				
44.	Perindopril	-1.3444	-1.44868	0.104275
				
45.	Proflavine	0.1427	0.451962	0.309262
				
46.	Quinapril	-0.9269	-1.03693	0.110031
				
47.	Ramipril	-1.1086	-1.23043	0.121833
				
48.	Ritonavir	-1.4065	-0.81131	0.595189
				

49.	Rivaroxaban 	-1.0523	-1.07364	0.021339
50.	Saxagliptin 	-1.-5474	-1.21112	0.336283
51.	Tilbroquinol 	-0.0719	-0.10064	0.02874
52.	Trandolapril 	-1.2479	-1.23993	0.007971
53.	2,3,4-Trihydroxy-4'-ethoxybenzophenone 	-0.95424	-0.91727	0.036971
54.	3,4-Didesmethyl-5-deshydroxy-3'-ethoxyscleroin 	-1.0523	-0.91291	0.139388
55.	Bronopol	-0.6435	-1.30644	0.662937

				
56.	Chicago blue	-0.8865	0.013249	0.899749
				
57.	Chloranil	-0.6129	-0.4068	0.206099
				
58.	DFMO	-2.3617	-1.25776	1.103941
				
59.	Evans blue	0.6989	0.015249	0.683651
				
60.	Hematoporphyrin	-0.5911	-0.7285	0.137398
				

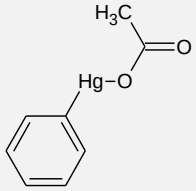
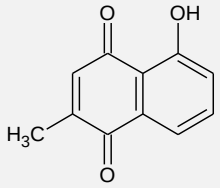
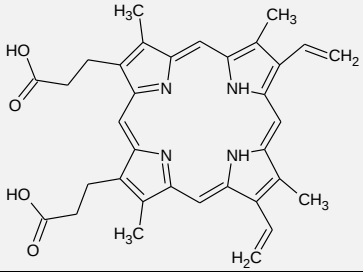
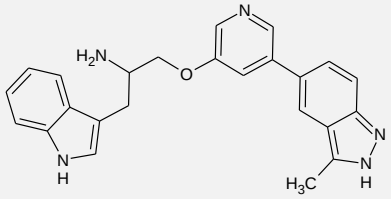
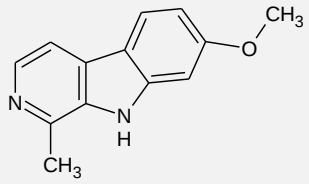
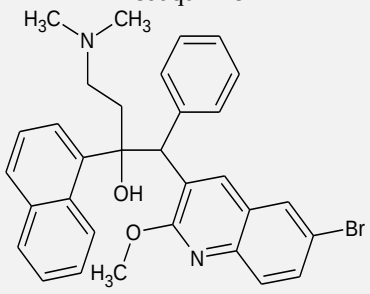
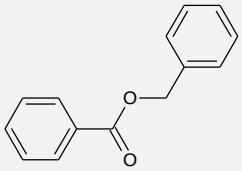
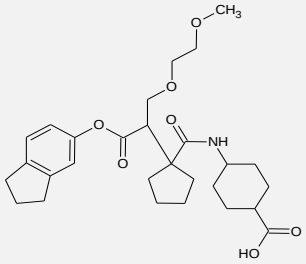
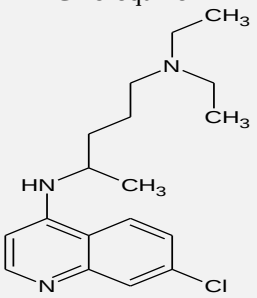
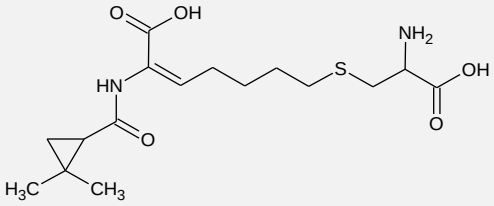
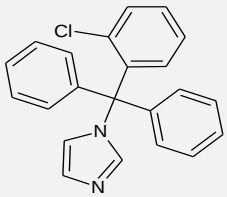
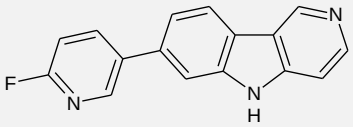
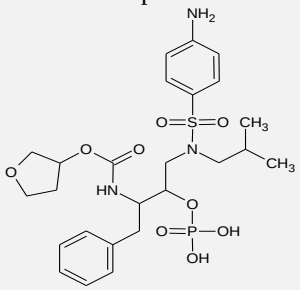
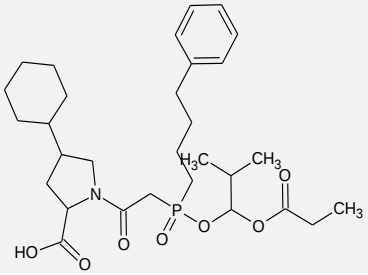
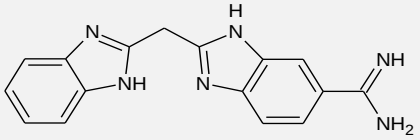
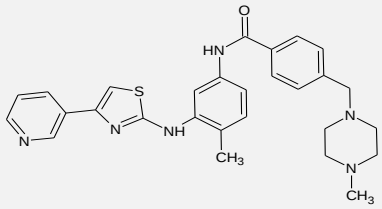
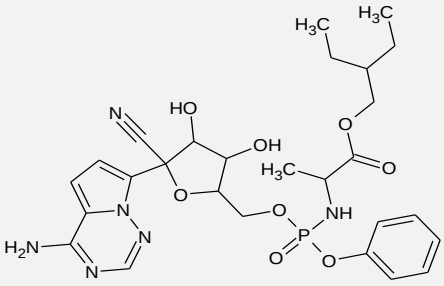
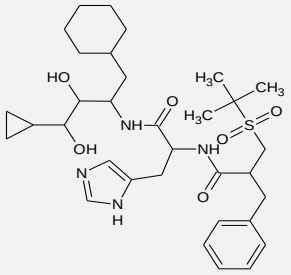
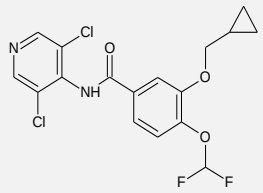
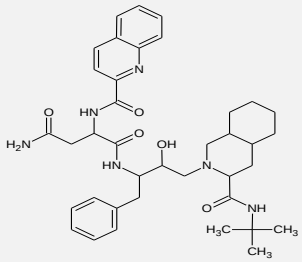
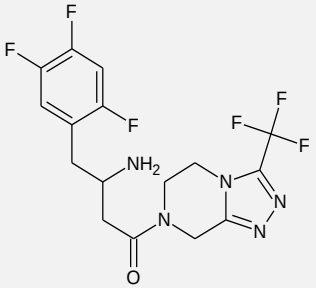
61.	Phenylmercuric acetate 	0.3979	-0.22481	0.622713
62.	Plumbagin 	-1.233	-0.37032	0.862678
63.	Protoporphyrin IX 	-1.3617	-0.37591	0.98579

TABLE 2: ACTUAL PIC50, PREDICTED PIC50 AND RESIDUAL VALUES OF TEST SET MOLECULES

S. no.	Structures of Molecules	Actual pIC_{50} value	Predicted pIC_{50} value	Residual value
1.	(2S)-1-(1H-indol-3-yl)-3-{[5-(3-methyl-1H-indazol-5-yl)pyridin-3-yl]oxy}propan-2-amine 	0.259637	0.32363	0.06399
2.	7-methoxy-1-methyl-9H-β-carboline 	0.3098	0.3749	0.06516
3.	Bedaquiline 	-0.1106	-0.1608	0.0502
4.	Benzyl Benzoate 	-0.1613	-0.0410	0.12022

5.	Candoxatril 	-0.7033	-0.9171	0.2138
6.	Chloroquine 	-1.0	-0.6659	0.3340
7.	Cilastatin 	-1.4828	-1.3541	0.1286
8.	Clotrimazole 	-0.1303	0.1291	0.25937
9.	Flortaucipir 	0.3767	0.54435	0.1676
10.	Fosamprenavir 	-0.9169	-1.2518	0.3349
11.	Fosinopril 	-1.0523	-1.2788	0.2265

12.	Hemi-Babam 	0.4559	0.45896	0.0031
13.	Masitinib 	-0.3979	-0.9176	0.5197
14.	Remdesivir 	-0.5682	-0.5086	0.0596
15.	Remikiren 	-0.5527	-0.7034	0.1507
16.	Roflumilast 	-0.1492	-0.2173	0.0681
17.	Saquinavir 	-1.1836	-0.7257	0.4578
18.	Sitagliptin 	-0.9149	-0.7459	0.1689

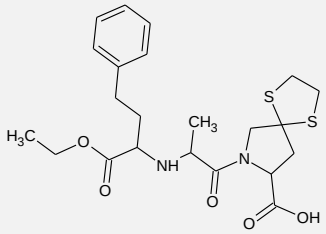
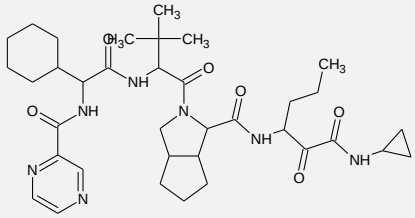
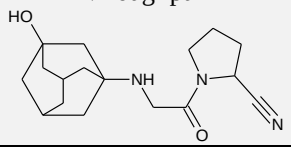
19.	Spirapril 	-1.2345	-1.2361	0.00165
20.	Telaprevir 	-1.0622	-1.0225	0.0397
21.	Vildagliptin 	-1.4826	-1.2223	0.2603

TABLE 3: Y RANDOMIZATION DATA OF THE QSAR MODEL

Model Type	R	R^2	Q^2(LOO)
Original	0.870987	0.758619	0.704927
Random 1	0.304665	0.092821	-0.08696
Random 2	0.254802	0.064924	-0.13797
Random 3	0.282466	0.079787	-0.05948
Random 4	0.241007	0.058084	-0.14795
Random 5	0.282468	0.079788	-0.13114
Random 6	0.268701	0.0722	-0.08202
Random 7	0.218067	0.047553	-0.11477
Random 8	0.178099	0.031719	-0.13693
Random 9	0.093867	0.008811	-0.18697
Random 10	0.211966	0.04493	-0.12703
Random 11	0.199883	0.039953	-0.14289
Random 12	0.130252	0.016966	-0.17501
Random 13	0.133566	0.01784	-0.15175
Random 14	0.354096	0.125384	-0.02601
Random 15	0.248325	0.061665	-0.10308
Random 16	0.297476	0.088492	-0.07256
Random 17	0.286555	0.082114	-0.09196
Random 18	0.343171	0.117766	-0.03885
Random 19	0.192178	0.036932	-0.14273
Random 20	0.132698	0.017609	-0.16777
Random 21	0.192432	0.03703	-0.11045
Random 22	0.31235	0.097562	-0.08334
Random 23	0.069428	0.00482	-0.19237
Random 24	0.395448	0.156379	0.021469
Random 25	0.076507	0.005853	-0.18681
Random 26	0.391448	0.153232	-0.00983
Random 27	0.202658	0.04107	-0.14581
Random 28	0.312214	0.097478	-0.08703
Random 29	0.235215	0.055326	-0.0895
Random 30	0.397565	0.158058	0.016638
Random 31	0.288741	0.083372	-0.11538
Random 32	0.376591	0.14182	0.005174
Random 33	0.26917	0.072452	-0.10687
Random 34	0.200702	0.040281	-0.12166

Random 35	0.144258	0.02081	-0.15697
Random 36	0.324662	0.105405	-0.05196
Random 37	0.144908	0.020998	-0.181
Random 38	0.153895	0.023684	-0.1413
Random 39	0.225588	0.05089	-0.12262
Random 40	0.109666	0.012027	-0.17064
Random 41	0.23167	0.053671	-0.09958
Random 42	0.114539	0.013119	-0.16721
Random 43	0.088168	0.007774	-0.16394
Random 44	0.223768	0.050072	-0.12645
Random 45	0.178214	0.03176	-0.14034
Random 46	0.158131	0.025006	-0.14051
Random 47	0.120291	0.01447	-0.16084
Random 48	0.209177	0.043755	-0.10589
Random 49	0.122297	0.014957	-0.19632
Random 50	0.091656	0.008401	-0.18533

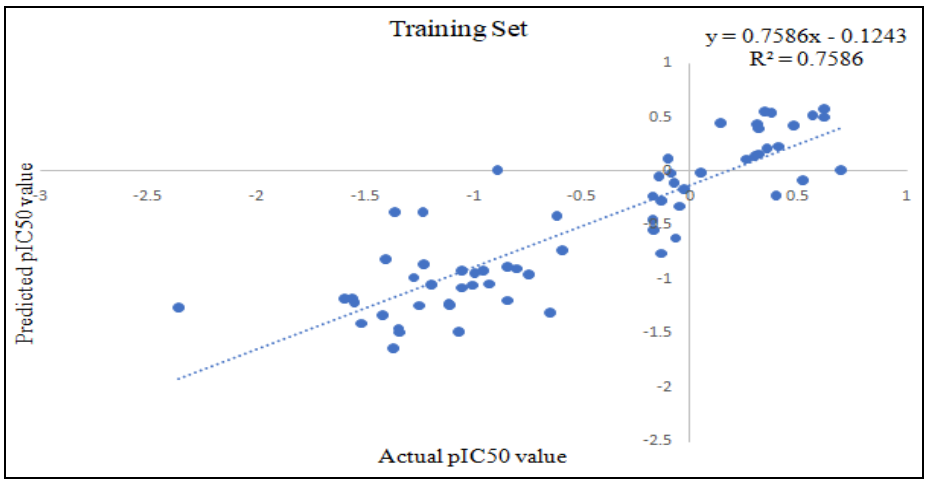


FIG. 1: GRAPH BETWEEN ACTUAL AND PREDICTED PIC50 VALUES IN THE TRAINING SET

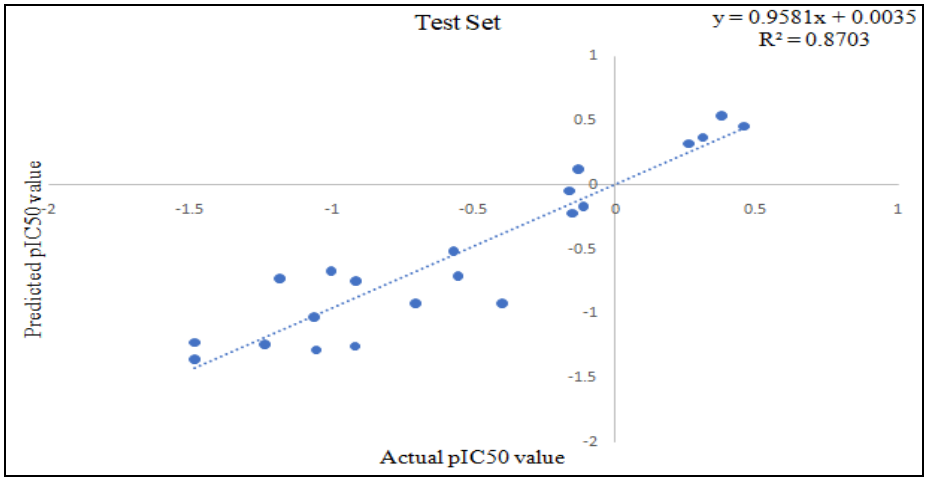


FIG. 2: GRAPH BETWEEN ACTUAL AND PREDICTED PIC50 VALUES IN THE TEST SET

As per the QSAR model, four two-dimensional descriptors such as naaaC (count of atom-type E-State: C:), nsssN (count of atom-type E-State: >N), SHsOH (sum of atom-type H E-State: -OH) and ETA_BetaP (measurement of electronic features of the molecule relative to molecular size) were significantly contributed [47]. As per the previous QSAR models related to anti-SARS-CoV-2 drug development, maximum R^2 and Q^2 values were 0.764 and 0.652 with MAE fitness score of 0.127, whereas our developed QSAR model was observed with higher statistical parameter values such as R^2 and Q^2 values were 0.87028 and 0.70493 with MAE fitness score value: 0.14298.

CONCLUSION: So, it was quite obvious that the newly developed QSAR model was statistically more validated than previous models with a large number of molecules present in the dataset. Also, atoms' E-state and electronic features of the molecules were directly related to anti-SARS-CoV-2 drug activity.

It can be easily concluded that if in the near future we want to develop a small molecule effective against SARS-CoV-2, the developed QSAR model will work as a good predictor of the activity profile with any chemical scaffold with possible descriptor combination.

ACKNOWLEDGEMENT: Declared none.

CONFLICTS OF INTEREST: The authors have no conflicts of interest, financial or otherwise.

REFERENCES:

1. Bennur T, Khan Z, Kshirsagar R, Javdekar V and Zinjarde S: Biogenic gold nanoparticles from the Actinomycete Gordoniaamarae: Application in rapid sensing of copper ions. Sensor Actuators B: Chemical 2016; 233(5): 684-90.
2. Gralinski LE and Menachery VD: Return of the Coronavirus: 2019-NCoV. Viruses 2020; 12(2): 135. https://doi.org/10.3390/v12020135.
3. Tripathi N, Tripathi N and Goshisht MK: COVID-19: inflammatory responses, structure-based drug design and potential therapeutics. Mol Divers 2021; 5: 1-17.
4. Cui W, Yang K and Yang H: Recent Progress in the Drug Development Targeting SARS-CoV-2 Main Protease as Treatment for COVID-19. Front Mol Biosci 2020; 7: 616341.
5. Saha S, Pal DK and Kumar S: Design, synthesis and antiproliferative activity of hydroxyacetamide derivatives against HeLa cervical carcinoma cell and breast cancer cell line. Trop J Pharm Res 2016; 15(7): 1319-26.
6. Saha S, Pal DK and Kumar S: Antifungal and Antibacterial Activities of Phenyl and Ortho-Hydroxy Phenyl Linked Imidazolyl Triazolo Hydroxamic Acid Derivatives. Inventi Rapid Med Chem 2017; 2017(2): 42-9.
7. Saha S and Pal D: Pyrazole and its derivatives, preparation, SAR and uses as antioxidative agent In: Pyrazole Prepararion and Uses; Pal D, Ed.; Nova Publisher: New York 2020; 211-43 ISBN: 9781536182507.
8. Saha S and Pal D: Role of pyrazole ring in neurological drug discovery. in: pyrazole prepararion and uses; pal d, ed.; nova publisher. New York 2020; 245-64. ISBN: 9781536182507.
9. Saha S, Pal DK and Kumar S: Hydroxyacetamide derivatives: cytotoxicity, antioxidative and metal chelating studies. Indian J Exp Biol 2017; 55: 831-7.
10. Pal DK, Kumar S and Saha S: Antihyperglycemic activity of phenyl and ortho-hydroxy phenyl linked imidazolyl triazolo hydroxamic acid derivatives. Int J Pharm Pharm Sci 2017; 9(12): 247-51.
11. Pal DK and Saha S: Chondroitin: a natural biomarker with immense biomedical applications. RSC Adv 2019; 9(48): 28061-77.
12. Saha S, Pal D and Nimse SB: Recent advances in the discovery of gsk-3 inhibitors from synthetic origin in the treatment of neurological disorders. Current Drug Target 2021; 22: 1-26.
13. Kaushik B, Pal D and Saha S: Gamma Secretase Inhibitor: Therapeutic Target via NOTCH Signaling in T cell Acute Lymphoblastic Leukemia. Curr Drug Targ 2021; 22: 1-10.
14. Batiha GE, Zayed MA, Awad AA, Shaheen HM, Mustapha S, Calderon OH,Pagnossa JP, Algammal AM, Zahoor M, Adhikari A, Pandey I, Elazab ST, Rengasamy KRR, Martins NC and Hetta HF: Management of SARS-CoV-2 Infection: Key Focus in Macrolides Efficacy for COVID-19. Front Med (Lausanne) 2021; 8: 642313.
15. Khan PM, Kumar V and Roy K: *In-silico* modeling of small molecule carboxamides as inhibitors of SARS-CoV 3CL protease: An approach towards combating COVID-19. Comb Chem High Throughput Screen 2020; doi: 10.2174/1386207323666200914094712.
16. Kumar V and Roy K: Development of a simple, interpretable and easily transferable QSAR model for quick screening antiviral databases in search of novel 3C-like protease (3CLpro) enzyme inhibitors against SARS-CoV diseases. SAR QSAR Environ Res 2020; 31(7): 511-26.
17. Firpo MR, Mastrodomenico V, Hawkins GM, Prot M, Levillayer L, Gallagher T, Loriere ES and Mounce BC: Targeting Polyamines Inhibits Coronavirus Infection by Reducing Cellular Attachment and Entry. ACS Infect Dis 2021; 7: 1423-32.
18. Gao K, Nguyen DD, Chen J, Wang R and Wei GW: Repositioning of 8565 Existing Drugs for COVID-19. J Phys Chem Lett 2020; 11: 5373-82.
19. Coelho C, Gallo G, Campos CB, Hardy L and Wurtele M: Biochemical screening for SARS-CoV-2 main protease inhibitors. PLoS ONE 2020; 15(10): 0240079.
20. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, Shi Z, Hu Z, Zhong W and Xiao G: Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019- nCoV) *in-vitro*. Cell Res 2020; 30: 269-71.
21. Drayman N, DeMarco JK, Jones KA, Azizi SA, Froggatt HM and Tan K: Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2. Science 2021; 20: eabg5827.
22. Yap CW: PaDEL-descriptor: an open-source software to calculate molecular descriptors and fingerprints. J Comput Chem 2011; 32: 1466-74.
23. Garcia J, Marrero-Ponce CR, Acevedo MY,Barigye L, Valdes SJ and Contreras TJR: QuBiLSMIDAS: A parallel free-software for molecular descriptors computation based on multi-linear algebraic maps. J Comput Chem 2014; 35: 1395-409.
24. Golbraikh A, Muratov E, Fourches D and Tropsha A: Data set modelability by QSAR. J Chem Inf Model 2014; 54: 1-4.
25. Ballabio D, Consonni V, Claeys- Bruno M, Sergent M and Todeschini R: A novel variable reduction method adapted from space-filling designs. Chemometric Intelligent Lab Systems 2014; 136: 147-54.
26. Roy K and Kar S: The rm2 metrics and regression through origin approach: reliable and useful validation tools for predictive QSAR models (Commentary on 'Is regression through origin useful in external validation of QSAR models?'). Eur J Pharm Sci 2014; 62: 111-14.

27. Roy K and Mitra I: On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. Comb Chem High Throughput Screen 2011; 14: 450-74.
28. Roy K, Das RN and Paul LA: Quantitative structure-activity relationship for toxicity of ionic liquids to Daphnia magna Aromaticity vs. lipophilicity. Chemosphere 2014; 112: 120-27.
29. Garg R and Carr JS: Predicting the bioconcentration factor of highly hydrophobic organic chemicals. Food Chem Toxicol 2014; 69: 252-59.
30. Dearden JC and Hewitt M: QSAR modeling of bioconcentration factor using hydrophobicity, hydrogen bonding and topological descriptor. SAR QSAR Environ Res 2010; 21: 671-80.
31. Dimitrov SD, Dimitrova NC, Walker JD, Veith GD and Mekenyan OG: Predicting bioconcentration factors of highly hydrophobic chemicals. Effects of molecular size. Pure Appl Chem 2002; 74: 1823-30.
32. Ravichandran V, Rajak H, Jain A, Sivadasan S, Varghese CP and Agrawal RK: Validation of QSAR models-strategies and importance. Int J Drug Design Discovery 2011; 3: 511-19.
33. Krstajic D, Buturovic LJ, Leahy DE and Thomas S: Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform 2014; 6: 1-15.
34. Zhang S, Golbraikh A, Oloff S, Kohn H and Tropsha A: J Chem Inf Model 2006; 46: 1984-95.
35. Jaworska J, Nikolova-Jeliazkova N and Aldenberg T: QSAR applicability domain estimation by projection of the training set in descriptor space: A review. ATLA 2005; 33: 445-59.
36. Rucker C, Rucker G and Meringer M: Y Randomization and its variants in QSPR/QSAR. J Chem Inf Model 2007; 47: 2345-57.
37. Humberto FF, Tania FB and Marcelo SC: 2D chemometric studies of a series of azole derivatives active against fluconazole-resistant Cryptococcus gattii. J Braz Chem Soc 2013; 24: 962-72.
38. Kim JH: Is it possible to predict the ADI of pesticides using the QSAR approach. J Environ Health Sci 2012; 38: 550-60.
39. Saha S, Pal D and Nimse SB: Indazole Derivatives Effective against Gastrointestinal Diseases. Curr Top Med Chem 2022; 22(14): 1189-1214.
40. Saha S, Yeom GS, Nimse SB and Pal D: Combination Therapy of Ledipasvir and Itraconazole in the Treatment of COVID-19 Patients Coinfected with Black Fungus: An *In-silico* Statement. Biomed Res Int 2022; Article ID 5904261, 10 2022. https://doi.org/10.1155/2022/5904261.
41. Saha S, Pal D and Nimse SB: Indazole-Based Microtubule-Targeting Agents as Potential Candidates for Anticancer Drugs Discovery. Bioorganic Chemistry 2022; 122(3):105735.
42. Saha S, Pal D and Kumar S: Anti-inflammatory and analgesic activities of imidazolyl triazolo hydroxamic acid derivatives. Indian J Exper Biology 2022; 60(4): 280-85.
43. Krishnan V, Verma P, Saha S, Singh B, Vinutha T, Kumar RR, Kulshreshta A, Singh SP, Sathyavathi T, Sachdev A and Praveen S: Polyphenol-enriched extract from pearl millet (*Pennisetum glaucum*) inhibits key enzymes involved in post prandial hyper glycemia (α-amylase, α-glucosidase) and regulates hepatic glucose uptake. Biocatalysis and Agricultural Biotechnology 2022; 43: 102411.
44. Saha S and Pal D: Computational Approaches related to Drug Disposition. International J Pharm Pharmaceutical Sci 2021; 13(7): 19-27.
45. Pal D and Saha S: Chondroitin: a natural biomarker with immense biomedical applications. RSC Adv 2019; 9(48): 28061-77.
46. Kaushik B, Pal D and Saha S: Gamma Secretase Inhibitor: Therapeutic Target via NOTCH Signaling in T cell Acute Lymphoblastic Leukemia. Curr Drug Target 2021; 22(15): 1789-98.
47. Saha S, Pal D and Nimse SB: Recent Advances in the Discoveryof GSK-3 Inhibitors from Synthetic Origin in the Treatment of Neurological Disorders. Curr Drug Target 2021; 22(12): 1437-61.
48. Poduria R, Joshi G and Jagadeesh G: Drugs targeting various stages of the SARS-CoV-2 life cycle: Exploring promising drugs for the treatment of Covid-19. Cellular Signal 2020; 74: 109721.

How to cite this article:
Saha S and Pal DK: 2D QSAR approach to develop newer generation small molecules active against SARS-Covid. Int J Pharm Sci & Res 2023; 14(3): 1372-91. doi: 10.13040/IJPSR.0975-8232.14(3).1372-91.

All © 2023 are reserved by International Journal of Pharmaceutical Sciences and Research. This Journal licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

This article can be downloaded to **Android OS** based mobile. Scan QR Code using Code/Bar Scanner from your mobile. (Scanners are available on Google Playstore)