Chemical Physics Impact 8 (2024) 100544

Contents lists available at ScienceDirect

Chemical Physics Impact

journal homepage: www.sciencedirect.com/journal/chemical-physics-impact

Full Length Article

QSAR modelling to predict structural features of certain sulfonamide as Urokinase-type Plasminogen Activator inhibitors

Rahul D. Jawarkar [a,*], Magdi E.A. Zaki [b,*], Sami A. Al-Hussain [b], Abdullah Yahya Abdullah Alzahrani [c], Long Chiau Ming [d], Abdul Samad [e], Rahul G. Ingle [f], Suraj N. Mali [g,*]

[a] Department of Medicinal Chemistry, Dr. Rajendra Gode Institute of Pharmacy, University-Mardi Road, Amravati, India
[b] Department of Chemistry, Faculty of Science, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia
[c] Department of Chemistry, Faculty of Science and Arts, King Khalid University, Mohail Assir, Saudi Arabia
[d] School of Medical and Life Sciences, Sunway University, Sunway City, Malaysia
[e] Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Tishk International University, Erbil, Iraq
[f] Datta Meghe College of Pharmacy, DMIHER Deemed University, Wardha, India
[g] Department of Pharmaceutical Chemistry, School of Pharmacy, D.Y. Patil University (Deemed to be University), Sector 7, Nerul, Navi Mumbai 400706, India

ARTICLE INFO

Keywords:
Urokinase-type plasminogen activator inhibitors
QSAR
Pharmacophoric features

ABSTRACT

The serine hydrolase family includes serine proteases. It is essential for hydrolyzing protein peptide bonds and breaking them. The urokinase-type plasminogen activator (uPA) selectively binds to the uPAR on numerous cell types, including cancer cells. Pericellular proteolysis of cell-bound proteins requires this interaction. High uPA and uPAR levels regularly worsen cancer prognoses. Thus, small chemical active-site inhibitors that block uPA may diminish cancer cell invasion and metastasis. In compliance with Organization for Economic Corporation and Development guidelines, this research performed a complete Quantitative structure activity relationship analysis of sulfonamide compounds as Urokinase-type Plasminogen Activator inhibitors. Py-Descriptors were used for this investigation. PyDescriptor uses PyMOL standards and idioms to calculate 11,145 simple molecular descriptors. This plugin calculates molecular descriptors irrespective of molecular representation properties like atom numbering or labelling, spatial reference frame, translational and rotational invariance, etc. The investigation sought to find essential and hidden structural characteristics that regulate sulfonamide-type drugs' Urokinase-type Plasminogen Activator Inhibitory action. Twenty-eight sulfonamide chemicals are used in the Quantitative structure activity relationship study to generate statistically robust and highly predictive univariate and multivariate models. All models were thoroughly evaluated and meet several statistical parameter thresholds (e.g., R^2 =0.9259–0.9280, Q^2_{Loo}= 0.8579–0.8558, Q^2_{LMO}= 0.8013–0.7865). The analysis reveal that occurance of ring carbon atoms exactly at 3 A0 from carbon atom, number of negatively charged atoms from sulphur atoms within 5 bonds, presence of hydrogen atom exactly at 3 bonds from donar atoms, presence of carbon atom exactly at 4 A0 from donar atom, presence of acceptor atom exactly at 5 A0 from sulphur atom and sum of partial charges of lipo atoms within 6 bonds from sulphur atom are important pharmacophoric features for Urokinase-type Plasminogen Activator Inhibition binding affinity. Thus, the developed Quantitative structure activity relationship study has an equilibrium of quantitative and qualitative tactics. The results could be useful for future optimizations of sulfonamide analogues.

1. Introduction

There is a widespread consensus that the capacity of tumour cells to generate and recruit proteolytic enzymes is a critical factor in the progression of cancer cell attack and metastasis. There are many other proteolytic enzyme systems that are produced by human tumours; however, the plasminogen activator-plasmin system is particularly engaged in the process of cancer cell invasion and metastasis by cancer cells {[10] #628}. The plasminogen activators are a group of serine proteolytic enzymes that are mostly found in the form of tissue-type

* Corresponding authors.
E-mail addresses: rahuljawarkar@gmail.com (R.D. Jawarkar), mezaki@imamu.edu.sa (M.E.A. Zaki), mali.suraj1695@gmail.com (S.N. Mali).

https://doi.org/10.1016/j.chphi.2024.100544
Received 21 January 2024; Received in revised form 9 February 2024; Accepted 16 February 2024
Available online 17 February 2024

plasminogen activator (tPA) and urokinase-type plasminogen activator (uPA) {[10] #629. One of the most important initiators of the extracellular proteolytic cascade that is responsible for cellular invasiveness is uPA, which is a trypsin-like serine protease {[9] #630}. Proteolytically active uPA is a two-chain protein that is connected by a disulfide and is formed from proteolytically inactive pro-uPA by the hydrolysis of the Lys158-Ile159 peptide bond {[9] #631.The transformation of plasminogen into plasmin, which is responsible for the digestion of the components of the extracellular matrix and basement membranes, is an essential function of uPA. Plasmin may either directly digest these components or activate the proenzymes of matrix metalloproteinases {[9] #632}.

The urokinase plasminogen activator (uPA) inhibitors are composed of tiny molecules that include amidino or guanidino groups, which are positively charged and mimic the amino acid arginine. These groups establish a salt bridge interaction with the carboxylate of Asp189 in the S1 pocket of the urokinase active site, as well as in other serine proteases that resemble trypsin. The previously documented uPA inhibitors included arylamidines, guanidines with minor modifications, and amiloride. Some examples of novel selective uPA inhibitors are benzo(b)thiophene-2-carboxamidines, indole/benzoimidazole-5-carboxamidines, 2-pyridinylguanidines, 4-aminoarylguanidines/benzamidines, thiophene-2-carboxa midines, 2-naphthamidines, 1-isoquinolinylguanidines, and 2-aminobenzimidazoles {[20] #636}{[30] #641}. Several of these inhibitors have binding affinities in the nanomolar range and are now being evaluated as potential medication candidates. There have been no significant advancements in meeting the necessary criteria for a therapeutic drug that may effectively treat cancer by inhibiting uPA.

The conventional approach to developing new drugs and measuring the inhibitory effects of chemicals involves a trial-and-error method, which is both inefficient and expensive. As a result, significant efforts have been made to estimate activity levels by statistical modelling. Computational techniques are an essential element of the drug development process. Several in silico technologies have been developed to streamline the process of drug development, successfully reducing both the cost and the chance of failure [15].

There is increasing interest in using computational approaches to predict the biological activities of compounds. This is because we cannot create new compounds with stronger inhibitory activities unless we have information about their biological characteristics. Regarding this matter, there exists a well recognised approach that may provide valuable evidence by analysing the biological activities and chemical structures of specifically made molecules. When built and verified correctly, Quantitative Structure-Activity Relationship (QSAR) models are very useful for screening and ranking compounds that lack experimental data or have not yet been synthesised, as part of a safe chemical strategy approach {[15] #649}.

The potential use of QSAR models for screening chemical databases or virtual libraries prior to their synthesis is equally appealing to chemical producers, pharmaceutical businesses, and government bodies, especially during periods of limited resources. The primary objective of this study is to develop a novel QSAR model using a genetic algorithm–multiple linear regression (GA–MLR) approach to predict the binding affinity of known sulfonamide derivatives as inhibitors of urokinase plasminogen activator [1,23,27]. Furthermore, this text outlines many methods for validating QSAR models, such as leave-one-out (LOO), leave-many-out (LMO), cross-validation, external test set, and the Y-randomization test procedures.

2. Experimental

2.1. Dataset selection

The current study used the Urokinase-type Plasminogen Activator inhibitory activity values of 28 sulfonamide compounds, which were previously published by Ewa Zeslawska and obtained from a binding database, for the purpose of conducting QSAR analysis [6]. The flow of work is depicted in Fig. 1.

2.2. Modelling and molecular descriptors' calculation

The present study adheres to the OECD regulatory concept and follows a standardised process that is supported and used by several academics for efficient QSAR analysis. The structural sketching was performed using the ChemSketch 12 freeware, followed by energy reduction using the MMFF94 force field available in TINKER. The energy-optimized 3D structures were used to compute a substantial number of descriptors (>29,000) using PyDescriptor as a molecular descriptor calculator(Vijay H. [22]). The molecular descriptor pool consists of about 29,000 descriptors, ranging from one-dimensional to three-dimensional, including electro-topological, fingerprints, and other molecular descriptors. However, the estimated descriptors do not include relevant information. Therefore, objective feature selection (OFS) was performed using QSARINS-Chem 2.2.1 to remove a significant number of redundant chemical descriptors. Before doing subjective feature selection (SFS) using QSARINS-Chem 2.2.1, descriptors that were constant, near constant (>98 %), and strongly correlated (|R| >90 %) were excluded [13]. OFS significantly reduced the descriptor pool to just 640 and 185 molecular descriptors for QSAR models, respectively. These descriptors cover a broad range of structural and chemical characteristics, including constitutional (0D), 1D, bi-dimensional (2D), and 3D ([2,7–12]; Vijay H [19]; Vijay H [20]; Vijay H [21]; A. [29]).

2.2.1. Model development

The primary objective of analysing QSAR models is to determine the structural characteristics that influence the activity profile of a series of related compounds, as well as to predict the activity of a compound before it is synthesised or screened for biological activity. In order to get comprehensive information regarding pharmacophoric characteristics, QSAR models were generated utilising both a divided and undivided dataset. Prior to selecting descriptors, the dataset was partitioned into training (80 %) and prediction (20 %) sets in an arbitrary manner. numerous instances of splitting were performed in order to generate numerous Quantitative Structure-Activity Relationship (QSAR) models [25] ([18]; Vijay H [21]). A compound may or may not be included in the training set comprising two distinct splits. The genetic algorithm (GA) module of QSARINS-Chem 2.2.1 was used to choose the optimal number and set of chemical descriptors. The exploration of chemical descriptors was limited to five descriptors using the default parameters in QSARINS-Chem 2.2.1 to avoid overfitting and facilitate the acceptance of the produced QSAR models. The fitness feature of Q2 loo was used to address the issue of raw Q2 ([2–5,9,14,17,26]; Alexander [30]). The chemical descriptor heuristic search was limited to two descriptors using the default parameters in QSARINS-Chem 2.2.1 in order to avoid the issues of over-fitting and complexity [2,8,28].

2.2.2. Model validation

Model validation is an essential component of QSAR model development. Validating a QSAR model ensures that the model has the capacity to accurately predict outcomes outside of the data it was trained on [13]. The statistical capabilities and strengths of the genetic algorithm–multiple linear regression (GA–MLR) equations were assessed through various methods: (a) cross-validation (CV) using leave-one-out (LOO) and leave-many-out (LMO) procedures for internal validation; (b) evaluation using a prediction set; (c) data randomization, specifically Y-scrambling; and (d) verification of the fulfilment of certain criteria: $r^2tr \geq 0.6$, Q^2 loo ≥ 0.5, Q^2 LMO ≥ 0.6, $r^2 > Q^2$, r^2 ex ≥ 0.6, RMSEtr $<$ RMSEcv, $\Delta K \geq 0.05$, CCC ≥ 0.80, Q2 – Fn ≥ 0.60, $r^2 m \geq 0.6$, $(1–r^2/ro^2) < 0.1$, $0.9 \leq k \leq 1.1$ or $(1–r\,2/r'o^2) < 0.1$, $0.9 \leq k' \leq 1.1$, $|ro^2 – r'o^2| <$ 0.3 with RMSE and MAE close to zero (30,31–34]. A genetic algorithm-multiple linear regression (GA-MLR) model that meets the specified threshold values for these restrictions maintains both

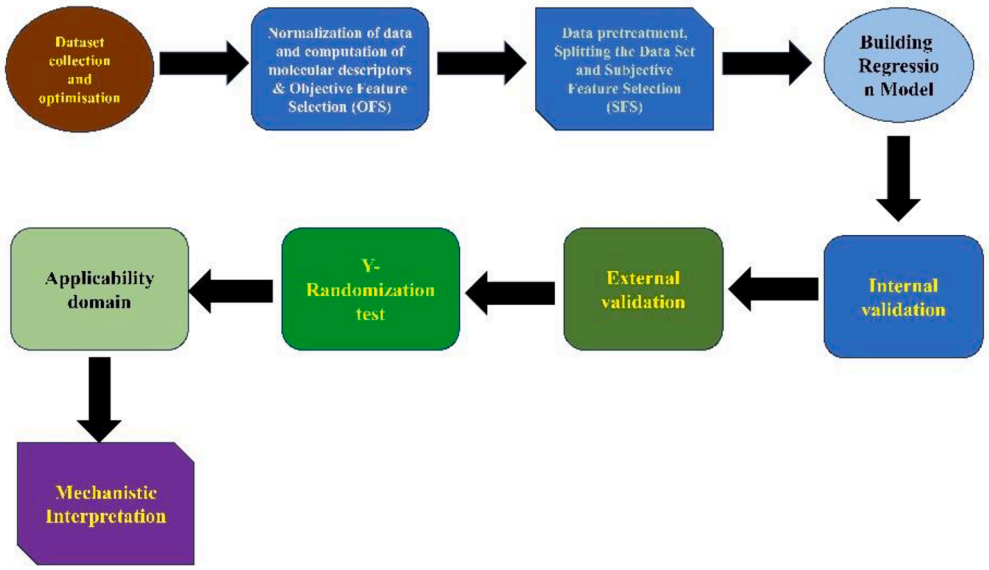


Fig. 1. Display of workflow in the current investigation.

statistical robustness and external predictive capability. Consequently, the models that fail to meet the criteria mentioned before were later eliminated [2,5,16,24].

3. Result and discussion

3.1. QSAR models

While we have compared the activities of the molecules in the dataset using a single characteristic, it is important to note that the combined or opposing influence of unknown factors or other chemical characteristics might significantly affect the activity pattern of the compounds. The QSAR model obtained meets the specified threshold values for a significant number of internal and external validation criteria. Furthermore, the validity of the AD model was verified by the use of the Williams plot, as seen in Fig. 2 and 3. Therefore, the model is statistically reliable and has strong prediction capacity when applied to external data. Furthermore, the achievement of recommended threshold values for several parameters, together with a low r2 value after Y-scrambling, indicates that the model is not a result of random chance. The forecasted pIC50 values obtained from the constructed QSAR models are shown in Figs. 2 and 3, and are also included in Tables 1 and 2. The GA-MLR analysis produces statistically adequate QSAR models 1 and 2, which meet recommended threshold values for various statistical limitations that are critical for assessing the accuracy of the models and their capacity to make predictions on external data. QSAR models may provide insights into the structural characteristics that determine the inhibitory action of UPa.

Model-1 (Full Dataset Model)

pKi = 10.116 (± 1.151) + −0.462 (± 0.122) * minus_S_5B + −0.182 (± 0.042) * fCringC3A + −0.108 (± 0.033) * fdonH3B + 0.053 (± 0.023) * fdonC4A + 0.341 (± 0.078) * fSacc5A + −3.283 (± 0.569) * lipo_S_6Bc+0 (± 0) *

(Fitting criteria)

R^2: 0.9246, R^2adj: 0.9030, R^2-R^2adj: 0.0216, LOF: 0.0344, Kxx: 0.3246, Delta K: 0.0592, RMSE tr: 0.1060, MAE tr: 0.0824, RSS tr: 0.3146, CCC tr: 0.9608, s: 0.1224, F: 42.9003

(Internal validation criteria)

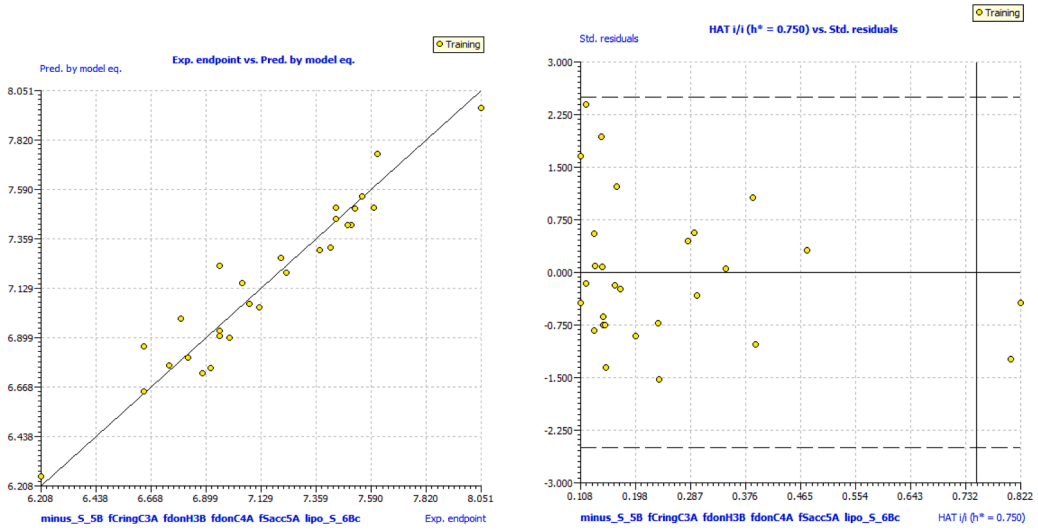


Fig. 2. (a) Graph for Experimental vs Predicted Ki (b) William plot for applicability domain of the developed Full Dataset QSAR model.

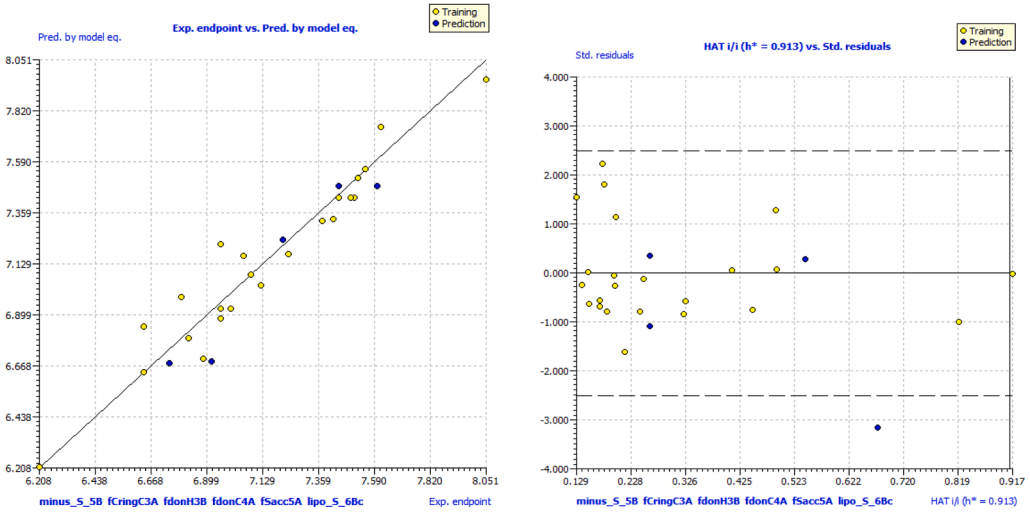


Fig. 3. (a) Graph for Experimental vs Predicted Ki (b) William plot for applicability domain of the developed Multivariate QSAR model.

Table 1
Experimental and predicted pKi by Full Dataset Model along with the status of the molecule.

sn	Binding Database id	Status	Ki (nM)	pKi	Pred. pKi	Pred.Mod. Eq.Res.
1	214,731	Training	8.9	8.0506	7.9737	−0.0769
2	214,730	Training	24	7.6198	7.7569	0.1371
3	214,717	Training	25	7.6021	7.5076	−0.0945
4	214,718	Training	28	7.5528	7.5585	0.0057
5	214,714	Training	30	7.5229	7.5036	−0.0193
6	214,727	Training	31	7.5086	7.4248	−0.0838
7	214,725	Training	32	7.4949	7.4248	−0.07
8	214,719	Training	36	7.4437	7.5076	0.0639
9	214,706	Training	36	7.4437	7.4549	0.0112
10	214,713	Training	38	7.4202	7.3213	−0.0989
11	214,732	Training	42	7.3768	7.3103	−0.0664
12	214,726	Training	58	7.2366	7.2035	−0.0331
13	214,721	Training	61	7.2147	7.2732	0.0586
14	214,722	Training	75	7.1249	7.0411	−0.0838
15	214,711	Training	83	7.0809	7.059	−0.0219
16	214,715	Training	89	7.0506	7.1535	0.1028
17	214,729	Training	100	7	6.9024	−0.0976
18	214,734	Training	110	6.9586	6.9099	−0.0487
19	214,720	Training	110	6.9586	7.2349	0.2762
20	214,728	Training	110	6.9586	6.9325	−0.0261
21	214,716	Training	120	6.9208	6.7578	−0.163
22	214,735	Training	130	6.8861	6.7343	−0.1518
23	214,710	Training	150	6.8239	6.8062	−0.0177
24	214,708	Training	160	6.7959	6.9885	0.1926
25	214,733	Training	180	6.7447	6.7727	0.028
26	214,723	Training	230	6.6383	6.8588	0.2205
27	214,724	Training	230	6.6383	6.6483	0.01
28	214,709	Training	620	6.2076	6.2546	0.047

Table 2
Experimental and predicted pKi by Divided dataset model along with the status of the molecule.

Binding Database id	sn	Status	Ki (nM)	pKi	Pred. pKi	Pred.Mod. Eq.Res.
214,731	1	Training	8.9	8.0506	7.9625	−0.0881
214,730	2	Training	24	7.6198	7.7514	0.1316
214,717	3	**Prediction**	25	7.6021	7.4832	−0.1189
214,718	4	Training	28	7.5528	7.5602	0.0074
214,714	5	Training	30	7.5229	7.5185	−0.0044
214,727	6	Training	31	7.5086	7.4304	−0.0782
214,725	7	Training	32	7.4949	7.4304	−0.0644
214,719	8	**Prediction**	36	7.4437	7.4832	0.0395
214,706	9	Training	36	7.4437	7.4311	−0.0126
214,713	10	Training	38	7.4202	7.3336	−0.0866
214,732	11	Training	42	7.3768	7.3233	−0.0535
214,726	12	Training	58	7.2366	7.1769	−0.0597
214,721	13	**Prediction**	61	7.2147	7.2394	0.0247
214,722	14	Training	75	7.1249	7.0338	−0.0912
214,711	15	Training	83	7.0809	7.0806	−0.0003
214,715	16	Training	89	7.0506	7.1687	0.1181
214,729	17	Training	100	7	6.9288	−0.0712
214,734	18	Training	110	6.9586	6.886	−0.0726
214,720	19	Training	110	6.9586	7.2185	0.2599
214,728	20	Training	110	6.9586	6.9295	−0.0291
214,716	21	**Prediction**	120	6.9208	6.6903	−0.2305
214,735	22	Training	130	6.8861	6.7031	−0.1829
214,710	23	Training	150	6.8239	6.7968	−0.0272
214,708	24	Training	160	6.7959	6.9817	0.1858
214,733	25	**Prediction**	180	6.7447	6.6825	−0.0623
214,723	26	Training	230	6.6383	6.8488	0.2106
214,724	27	Training	230	6.6383	6.6405	0.0023
214,709	28	Training	620	6.2076	6.2141	0.0065

Q^2loo: 0.8579, R^2-Q2loo: 0.0667, RMSE cv: 0.1455, MAE cv: 0.1155, PRESS cv: 0.5927, CCC cv: 0.9265, Q^2LMO: 0.8013, R^2Yscr: 0.2219, Q^2Yscr: −0.6319, RMSE AV Yscr: 0.3396

Predictions by LOO:

Exp(x) vs. Pred(y): R^2: 0.8606, R'^{2}o: 0.8494, k': 0.9978, Clos': 0.0130, r'^{2}m: 0.7696, Pred(x) vs. Exp(y): R^2: 0.8606, R^2o: 0.8590, k: 1.0018, Clos: 0.0018, r^2m: 0.8263

Model 2 (Divided dataset model) pKi= 10.395 (± 1.415) + −0.5 (± 0.167) * minus_S_5B + −0.185 (± 0.049) * fCringC3A + −0.106 (± 0.04) * fdonH3B + 0.052 (± 0.028) * fdonC4A + 0.374 (± 0.129) * fSacc5A + −3.34 (± 0.649) * lipo_S_6Bc+0 (± 0) *

R^2: 0.9280, R^2adj: 0.9010, R^2-R^2adj: 0.0270, LOF: 0.0501, Kxx: 0.3917, Delta K: 0.0430, RMSE tr: 0.1071, MAE tr: 0.0802, RSS tr: 0.2636, CCC tr: 0.9626, s: 0.1283, F: 34.3553, Q^2loo: 0.8558, R^2-Q^2loo: 0.0722, RMSE cv: 0.1515, MAE cv: 0.1160, PRESS cv: 0.5277, CCC cv: 0.9253, Q^2LMO: 0.7865, R^2Yscr: 0.2724, Q^2Yscr: −1.1128, RMSE AV Yscr: 0.3387, R^2Yrnd: 0.2757, Q^2Yrnd: −1.1086, RMSE ext: 0.1211, MAE ext: 0.0952, PRESS ext: 0.0733, R^2ext: 0.9323, Q^2-F1: 0.8571, Q^2-F^2: 0.8549, Q^2-F3: 0.9078, CCC ext: 0.9380, r^2m aver.: 0.8028, r^2m delta: 0.0821

3.2. Interpretation of descriptor

The used descriptors in univariate and multivariate QSAR analysis

are briefly described below.

fCringC3A

The molecular descriptor fCringC3A indicates the frequency of ring carbon atoms occurring precisely at a distance of 3 A0 from a carbon atom. The Urokinase plasminogen activator binding affinity shows a strong and positive connection with a coefficient of determination (R2) of 0.924. The coefficient of this descriptor is positive, and hence, it should be maximised. In order to accomplish this goal, it is necessary to increase the number of carbon atoms around the aromatic ring by a distance of 3°A. This is connected to the presence of steric hindrance in the vicinity of the aromatic rings.

The compound 214,731 (pKi-8.05) has a methyl carbon atom located next to the sulphone moiety. The aromatic C1 carbon is positioned at a topological distance of 3A0 from this methyl carbon. Hence, we may conclude that the presence of the C1 carbon close to the sulphone is advantageous for the binding affinity of molecule 214,731. Furthermore, there is an additional benzyl ring connected to the amide linkage via a carbon atom, positioned at a topological distance of 3 A0. This indicates that the presence of an aromatic benzyl ring at a topological distance of 3 A0 is advantageous for the binding affinity of molecule 214,731 (as seen in Fig. 4).

Both the phenyl and benzyl rings are connected to the molecule's structure by a Sp3 hybridised carbon atom. It is important to note that the presence of this Sp3 hybridised carbon atom, which is located at a topological distance of 3A0, is crucial for the molecule's binding affinity.

minus_S_5B

The descriptor "minus_S_5B" indicates the count of negatively charged atoms that are connected to sulphur atoms within a distance of 5 bonds. The presence of a positive coefficient in this chemical descriptor indicates that raising its value will result in an increase in the Ki value for the compounds employed in this study. The presence of a negatively charged group may strengthen the electrostatic contact with the receptor, hence increasing the binding affinity to the receptor.

In compound 214,730 (with a pKi value of −7.61), there is an NH group adjacent to a sulphur atom, separated by only one bond. There is also an OH group located four bonds away from the sulphur atom, a CO group located five bonds away, and an NH group in an amide linkage located four bonds away from the sulphur atom (as depicted in Fig. 2). In compound 214,733 (pKi-6.74 nM), the hydroxy group is missing a negatively charged oxygen atom. The amide group has a negatively charged oxygen atom, which is located six bonds away in terms of topology. The sulfonamide group has a negatively charged NH atom, which is located one bond away in terms of topology. The NH2 substituent at the C2 carbon atom is located four bonds away in terms of topology. The research reveals that the lack of an OH substituent and the presence of an amide oxygen at a topological distance of six bonds may decrease the activity of compound 214,133 (as seen in Fig. 5).

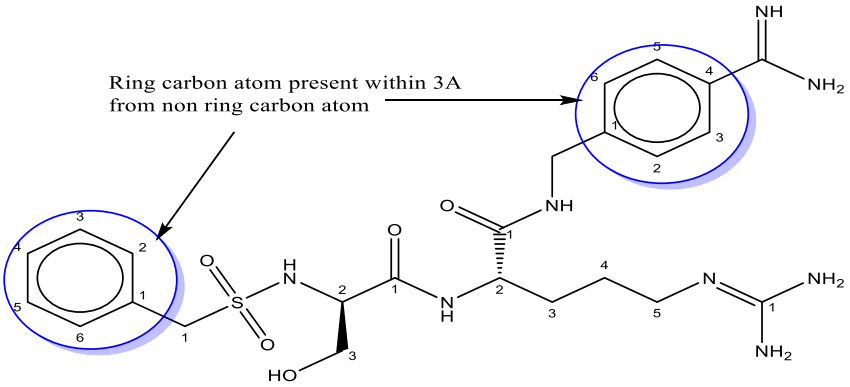
fdonh3b

The "fdonh3b descriptor" refers to the frequency at which a hydrogen atom occurs precisely three bonds away from donor atoms. The coefficient of this descriptor is positive, hence it is desirable to maximize its value. The compound 214,717 (with a pKi value of 7.60) has a sulfonamide group (NH group), an OH group attached to the C3 carbon atom, an amide group NH group near the phenyl methyl carbon, and NH and NH2 groups attached to the C4 carbon of the phenyl ring. Additionally, the molecule has a donor group in the form of a (S) 3-OH group attached to the C3 carbon atom, as seen in Fig. 4. In 214,717, there are approximately six donor groups. The hydrogen atom (C1 carbon) of the phenyl ring carbon atom is separated by three bonds from the NH donor group of sulfonamides. For the ®OH at the C3 carbon, the hydrogen atom of the NH group in the sulfonamide is separated by three bonds. In the amide linkage, there are two hydrogen atoms located exactly three bonds away. One hydrogen atom is in the OH group, while the other is in the subsequent NH group near the right phenyl methyl moiety. The NH and NH2 groups act as donors at the C4 carbon of the phenyl ring. The hydrogen atoms at the C3 and C5 positions of the phenyl methyl ring are located at a topological distance of three bonds. The current research demonstrates that the existence of hydrogen at a topological distance of three bonds plays a vital role in the binding affinity of Urokinase inhibitors. Furthermore The presence of a donor is crucial in the electrostatic interaction with a receptor. In compound 214,710 (pKi-6.82 nm), all the donor properties are present except for the absence of the (R) OH group at the C1 carbon atom. This absence may be the main cause for the difference in activity between compounds 214,717 (pKi-7.60 nM) and 214,710 (pKi-6.82 nm). The current study concludes that the presence of a greater number of donors with hydrogen atoms positioned at a topological distance of three bonds is crucial for the binding affinity of Urokinase (as seen in Fig. 6).

The molecule's binding affinity may be increased if there are successively more hydrogen atoms present at a topological distance of three bonds from the donor. Hydrogen atoms are essential to the structural makeup of molecules because an increase or reduction in the amount of hydrogen atoms close to a certain characteristic might affect the biological activity profile of the molecule.

fdonC4A

This description shows how often the carbon atom occurs precisely at 4 A0 from the donor atom. The QSAR model's positive coefficient for fsp3Osp3C3B suggests that raising this descriptor's value might improve the target enzyme's affinity. This finding is supported by a straightforward assessment of 214,730(pKi-7.62 nM) with 214,723(pKi-6.63Nm). About 18 carbon atoms, three of which are sp2 hybridised and the other fifteen of which are sp3 hybridised, are positioned precisely at 4A0 from donor atoms in compound 214,730. Additionally, compound 214,730 has one oxygen atom and six donar nitrogen atoms. Of the



Binding Database id-214731, pKi-8.05 nM

Fig. 4. Pictorial demonstration of minus_S_5B compound 214,731(pKi-8.05 nm).

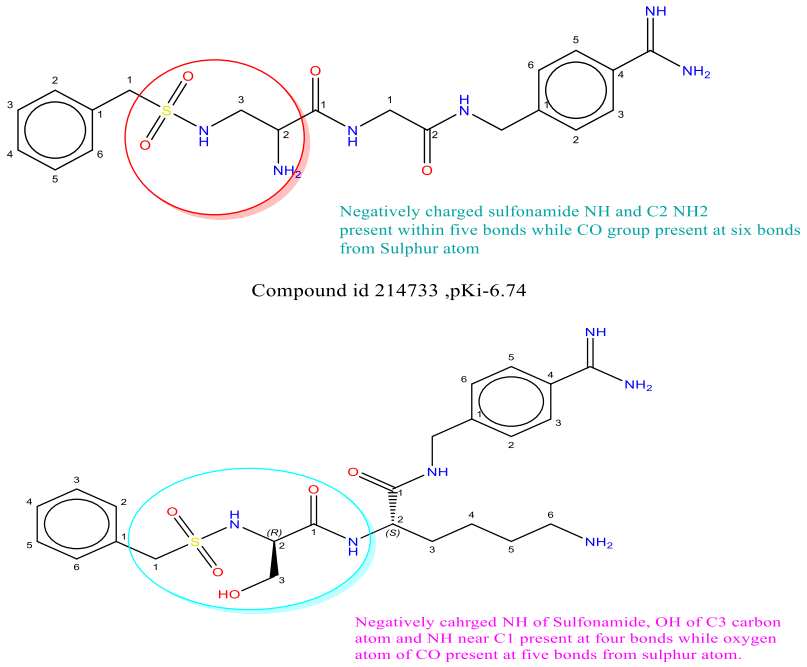


Fig. 5. Pictorial demonstration of minus_S_5B compound 214,733(pKi-6.74) and 214,730(pKi-6.74Nm) only.

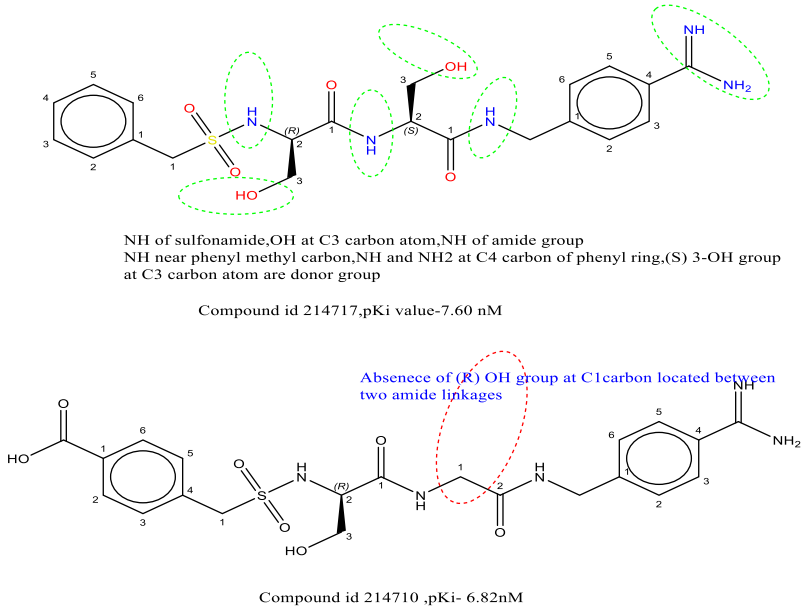


Fig. 6. Pictorial representation of **Fdonh3b** compound 214,717 (pKi-7.60 nM) and 214,710(pKi-6.82 nm) only.

nitrogen-containing donars, three belong to the amide component; one nitrogen is Sp3 hybridised, another is Sp2 hybridised, and the last nitrogen is Sp3 hybridised, which is a planer kind of nitrogen.

There are about 22 carbon atoms in compound 214,723 that are precisely 4A0 away from the donar atom. In compound 214,723, every carbon atom in the structure is precisely 4A0 away from the donor atom. Compound 214,723 consists of five donar features, three of which belong to the amide group. Of these, one nitrogen has undergone planar Sp2 hybridization, while the other nitrogen has undergone Sp3 hybridization.as seen in Fig. 7).

According to the current study, the presence of a carbon atom precisely at this distance is advantageous for the molecule's binding affinity. Therefore, if a molecule has more than 18 carbon atoms, its activity profile may be reduced. The difference between compound 214,730′s (pKi-7.62 nM) and compound 214,723′s (pKi-6.63Nm) activity is supported by this investigation. Additionally, the presence of more donars is advantageous for activity; in compound 214,730, nitrogen that is Sp3 hybridised, one nitrogen that is Sp2 hybridised, and one nitrogen that is Sp3 hybridised planer play critical roles in molecule binding affinity, thereby enhancing drug receptor interaction.

fSacc5A

The frequency of acceptor atoms occurring precisely at 5 A° from sulphur atoms is described by fSacc5A. Because this descriptor has a positive coefficient, it positively affects the molecule's biological

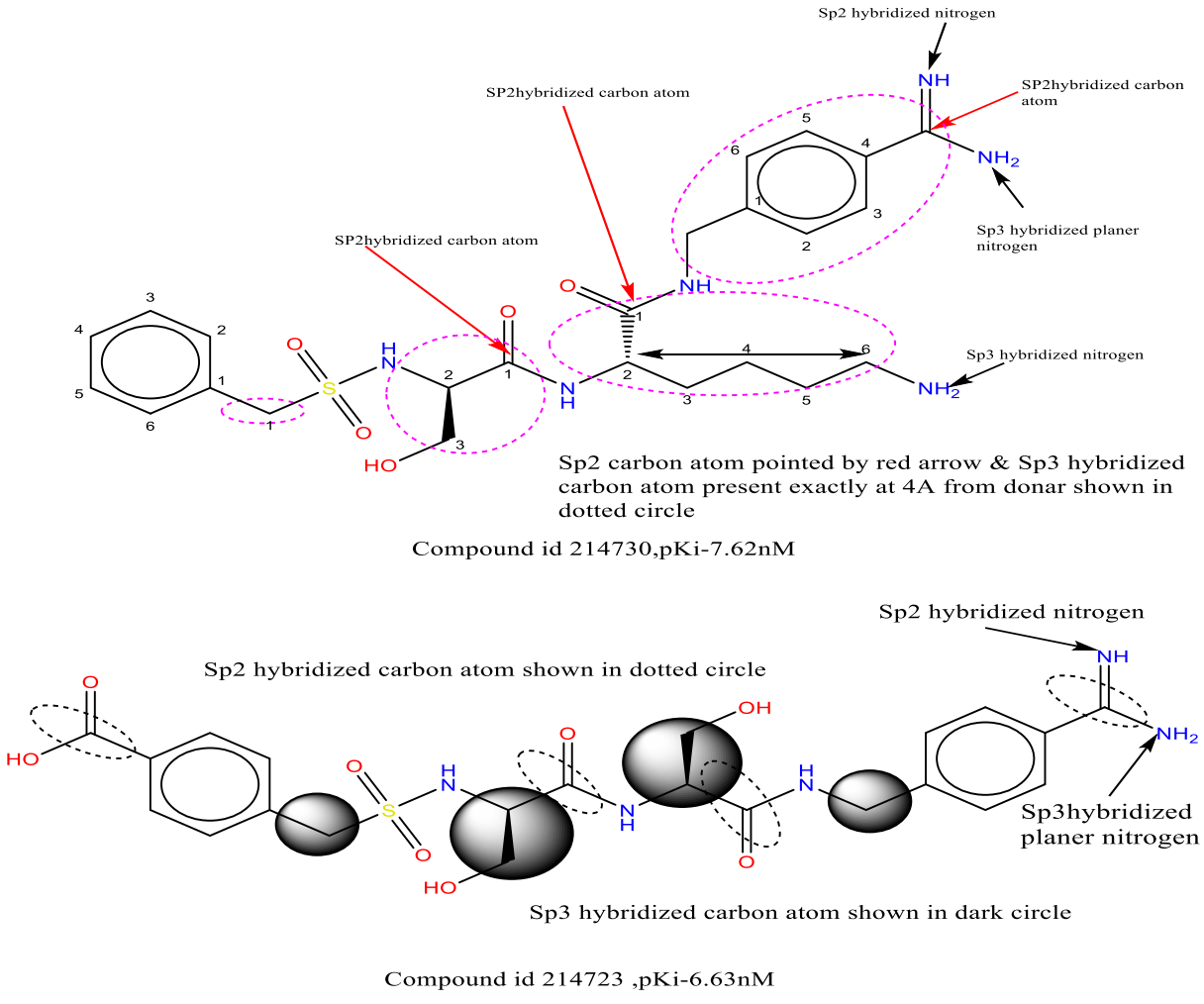


Fig. 7. Pictorial representation of **fdonC4A** compound 214,730 (pKi-7.62 nM) and 214,723(pKi-6.63 nm) only.

activity profile. Both sulphone acceptor oxygen atoms are precisely positioned at position 5 A° from the sulphur atom in compound 214,718 (pKi-7.55Nm). Similarly, the carbonyl and hydroxy acceptor groups are also precisely located at 5 A0 from the sulphur atom. Compound 214,718 (pKi-7.55 Nm) exhibits Sp2 hybridization of carbonyl oxygen and sulphone atoms, and Sp3 hybridization of hydroxy oxygen. Furthermore, both Sp2 hybridised sulphone oxygen atoms are identified as acceptor atoms in compound 214,733 (pKi-6.74Nm), where they are present precisely at 5A0 from the sulphur atom. In contrast, carbonyl acceptor oxygen and hydroxy acceptor atoms are absent (According to Fig. 8).

This finding is corroborated by the variation in affinity between 214,718 (pKi = 7.55 nM) and 214,733 (pKi = 6.74 nM). Thus, it can be inferred that the presence of four or more acceptors, which are required to be precisely at a distance of five A0, is beneficial for raising the molecule's binding affinity against the receptor site. Furthermore, a molecule's biological activity profile may be enhanced by the addition of two or more acceptor groups.

lipo_S_6Bc

Another molecular descriptor with a positive coefficient is lipo_S_6Bc, which represents the total sum of partial charges of lipophilic atoms located within a distance of 6 chemical bonds from sulphur (S) atoms. The presence of a positive coefficient indicates that an increase in the amount of lipophilic carbon atoms at a six-bond distance from the sulphur atom might result in a higher Ki value. The Fig. 9 displays the partial charge of the lipophilic carbon atom located six bonds away from the sulphur atom.

Fig. 5 displays the partial charge of the lipophilic carbon atom located six bonds away from the sulphur atom. In compound 214,714, the presence of an aromatic ring with a methyl chain on the left side of the sulphur atom, a carbon chain ending with a hydroxy group, and one methyl substituent precisely positioned between two amide linkages represents lipophilic atoms located within six bonds from the sulphur atom (as depicted in Fig. 10). These lipophilic atoms contribute to the molecule's lipophilic character and can strongly engage in hydrophobic interactions with a receptor, thereby increasing the molecule's binding affinity.

The lack of a methyl substituent between two amide linkages in compound 214,710 suggests that the difference in activity between compound 214,714 (pKi-7.532 M) and compound 214,710 (pKi-6.82 nm) may be attributed to the missing methyl group. Moreover, the molecule's binding affinity may be enhanced by introducing a methyl substituent in the amide bond. The present investigation has provided valuable insights into the specific factors that contribute to the variance in the binding affinity of sulfonamides to the Urokinase plasminogen activator. This was achieved by examining several chemical descriptors inside the QSAR model.

4. Conclusion

Using the Genetic algorithm and multiple linear regressions, the QSAR analysis of a number of compounds that were shown to be UPa inhibitors was carried out in this study. Using the algorithm genetic technique, the descriptors that were deemed to be the most relevant

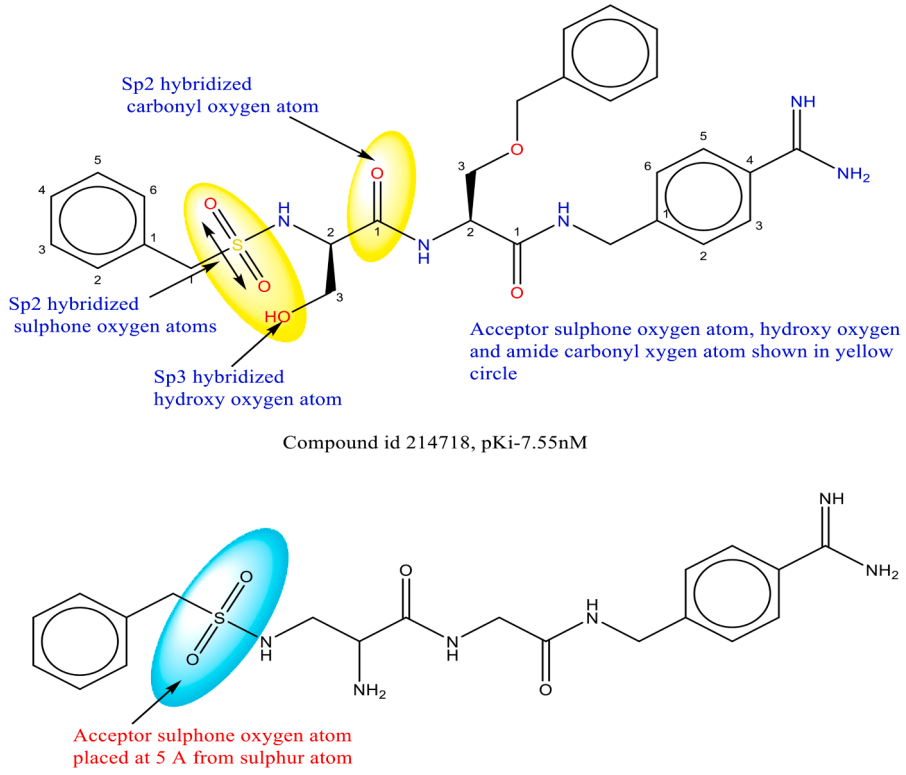


Fig. 8. Pictorial illustration of **fSacc5A** compound 214,718 (pKi-7.55 nM) and 214,733(pKi-6.74 nm) only.

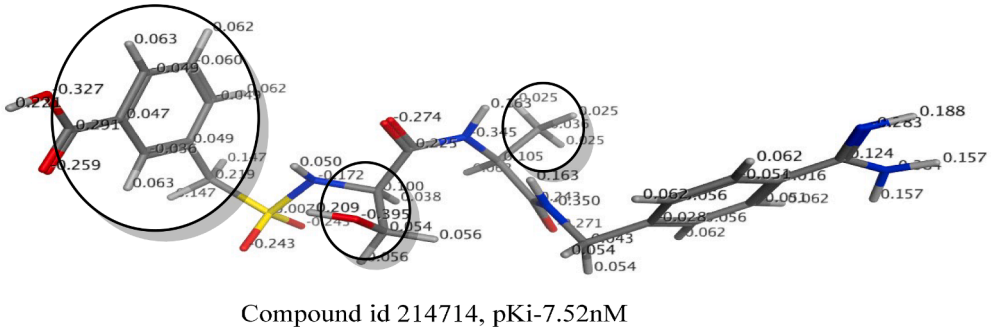


Fig. 9. 3D pictorial depiction of partial charge of Compound 214,714 (pKi-7.52 nm).

were chosen. The correctness and robustness of the constructed model are shown by the validation procedures that were carried out, which include Y-randomization and cross-validation. For the purpose of this investigation, the QSAR models that were constructed have the potential to be useful for predicting the activity of novel compounds as UPa inhibitors, and they can also give a better understanding for the creation of potent UPa inhibitors.

Funding

This work was supported by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU).

CRediT authorship contribution statement

Rahul D. Jawarkar: Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Magdi E.A. Zaki:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Investigation, Funding acquisition. **Sami A. Al-Hussain:** Validation, Supervision. **Abdullah Yahya Abdullah Alzahrani:** Conceptualization. **Long Chiau Ming:** Conceptualization. **Abdul Samad:** Conceptualization. **Rahul G. Ingle:** Conceptualization. **Suraj N. Mali:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

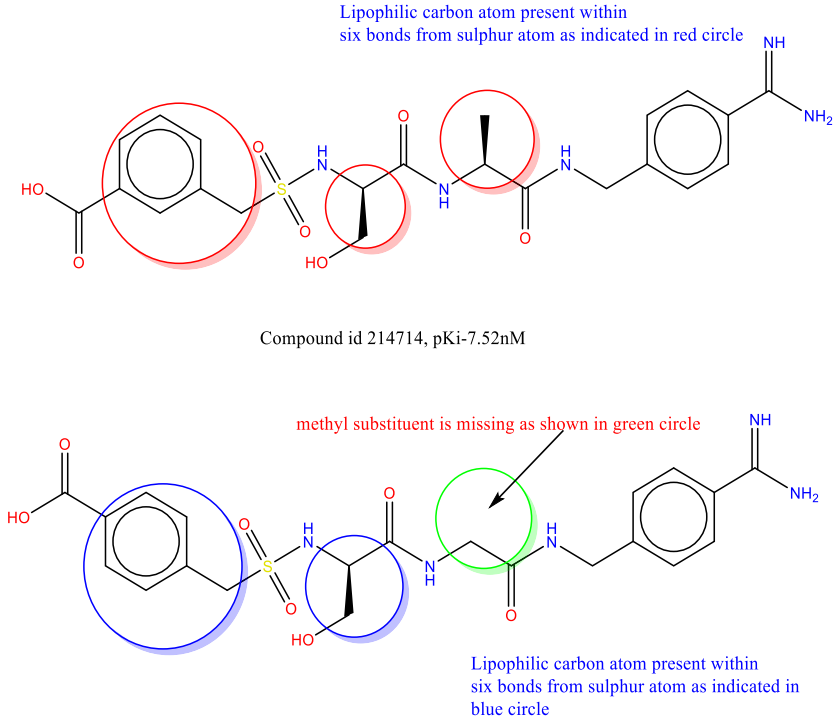


Fig. 10. Pictorial representation of **lipo_S_6Bc** compound 214,714 (pKi-7.532 M) and 214,710 (pKi-6.82 nm) only.

Acknowledgements

The author Rahul D. Jawarkar is thankful to Dr. Paola Gramatica and her team for providing the free copy of QSARINS- 2.2.4. The authors acknowledge the president; Honorable Yogendraji Gode Sir, President, IBBSS, Buldhana, for providing research facilities during entire course of research work. This work was supported by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU).

References

[1] K. Bukowski, M. Kciuk, R. Kontek, Mechanisms of multidrug resistance in cancer chemotherapy, Int. J. Mol. Sci. 21 (9) (2020), https://doi.org/10.3390/ijms21093233.
[2] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, A. Tropsha, QSAR modeling: where have you been? Where are you going to? J. Med. Chem. 57 (12) (2014) 4977–5010, https://doi.org/10.1021/jm4004285.
[3] N. Chirico, P. Gramatica, Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient, J. Chem. Inf. Model. 51 (9) (2011) 2320–2335, https://doi.org/10.1021/ci200211n.
[4] N. Chirico, P. Gramatica, Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection, J. Chem. Inf. Model. 52 (8) (2012) 2044–2058, https://doi.org/10.1021/ci300084j.
[5] V. Consonni, D. Ballabio, R. Todeschini, Comments on the definition of the Q2 parameter for QSAR validation, J. Chem. Inf. Model. 49 (7) (2009) 1669–1678, https://doi.org/10.1021/ci900115y.
[6] D. Fourches, E. Muratov, A. Tropsha, Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research, J. Chem. Inf. Model. 50 (7) (2010) 1189–1204, https://doi.org/10.1021/ci100176x.
[7] T. Fujita, D.A. Winkler, Understanding the Roles of the "Two QSARs", J. Chem. Inf. Model. 56 (2) (2016) 269–274, https://doi.org/10.1021/acs.jcim.5b00229.
[8] A. Golbraikh, E. Muratov, D. Fourches, A. Tropsha, Data set modelability by QSAR, J. Chem. Inf. Model. 54 (1) (2014) 1–4, https://doi.org/10.1021/ci400572x.
[9] P. Gramatica, On the development and validation of QSAR models, Methods Mol. Biol. 930 (2013) 499–526, https://doi.org/10.1007/978-1-62703-059-5_21.
[10] P. Gramatica, External evaluation of QSAR models, in addition to cross-validation verification of predictive capability on totally new chemicals, Mol. Inform. 33 (4) (2014) 311–314.
[11] P. Gramatica, Principles of QSAR modeling, Int. J. Quant. Struct. Prop. Relatsh. 5 (3) (2020) 61–97, https://doi.org/10.4018/IJQSPR.20200701.oa1.
[12] P. Gramatica, S. Cassani, P.P. Roy, S. Kovarich, C.W. Yap, E. Papa, QSAR modeling is not push a button and find a correlation: a case study of toxicity of (Benzo-) triazoles on algae, Mol. Inform. 31 (11–12) (2012) 817–835.
[13] P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, QSARINS: a new software for the development, analysis, and validation of QSAR MLR models, J. Comput. Chem. 34 (24) (2013) 2121–2132, https://doi.org/10.1002/jcc.23361.
[14] P. Gramatica, E. Giani, E. Papa, Statistical external validation and consensus modeling: a QSPR case study for Koc prediction, J. Mol. Graph. Model. 25 (6) (2007) 755–766, https://doi.org/10.1016/j.jmgm.2006.06.005.
[15] R.D. Jawarkar, M.E.A. Zaki, S.A. Al-Hussain, A.A. Al-Mutairi, A. Samad, N. Mukerjee, S. Rashid, QSAR modeling approaches to identify a novel ACE2 inhibitor that selectively bind with the C and N terminals of the ectodomain, J. Biomol. Struct. Dyn. (2023) 1–20, https://doi.org/10.1080/07391102.2023.2205948.
[16] S. Kar, K. Roy, J. Leszczynski, Applicability domain: a step toward confident predictions and decidability for QSAR modeling, in: , 2018, pp. 141–169.
[17] D. Krstajic, L.J. Buturovic, D.E. Leahy, S. Thomas, Cross-validation pitfalls when selecting and assessing regression and classification models, J. Cheminform. 6 (1) (2014) 10, https://doi.org/10.1186/1758-2946-6-10.
[18] T.M. Martin, P. Harten, D.M. Young, E.N. Muratov, A. Golbraikh, H. Zhu, A. Tropsha, Does rational selection of training and test sets improve the outcome of QSAR modeling? J. Chem. Inf. Model. 52 (10) (2012) 2570–2578, https://doi.org/10.1021/ci300338w.
[19] V.H. Masand, D.T. Mahajan, P. Gramatica, J. Barlow, Tautomerism and multiple modelling enhance the efficacy of QSAR: antimalarial activity of phosphoramidate and phosphorothioamidate analogues of amiprophos methyl, Med. Chem. Research 23 (11) (2014) 4825–4835.
[20] V.H. Masand, D.T. Mahajan, T.B. Hadda, R.D. Jawarkar, A.M. Alafeefy, V. Rastija, M.A. Ali, Does tautomerism influence the outcome of QSAR modeling? Med. Chem. Res. 23 (4) (2014) 1742–1757.
[21] V.H. Masand, D.T. Mahajan, G.M. Nazeruddin, T.B. Hadda, V. Rastija, A.M. Alfeefy, Effect of information leakage and method of splitting (rational and random) on external predictive ability and behavior of different statistical parameters of QSAR model, Med. Chem. Res. 24 (3) (2015) 1241–1264.
[22] V.H. Masand, V. Rastija, PyDescriptor: a new PyMOL plugin for calculating thousands of easily understandable molecular descriptors, Chemom. Intell. Lab. Syst. 169 (2017) 12–18, https://doi.org/10.1016/j.chemolab.2017.08.003.
[23] C. Mottini, F. Napolitano, Z. Li, X. Gao, L. Cardone, Computer-aided drug repurposing for cancer therapy: approaches and opportunities to challenge anticancer targets, Semin. Cancer Biol. 68 (2021) 59–74, https://doi.org/10.1016/j.semcancer.2019.09.023.
[24] T.I. Netzeva, A. Worth, T. Aldenberg, R. Benigni, M.T. Cronin, P. Gramatica, C. Yang, Current status of methods for defining the applicability domain of

(quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52, Altern. Lab. Anim. 33 (2) (2005) 155–173. Retrieved from, http://www.ncbi.nlm.nih.gov/pubmed/16180989.
[25] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open Babel: an open chemical toolbox, J. Cheminform. 3 (1) (2011) 33, https://doi.org/10.1186/1758-2946-3-33.
[26] R.B. Rao, G. Fung, R. Rosales, On the dangers of cross-validation, Exp. Eval. (2008) 588–596, https://doi.org/10.1137/1.9781611972788.54.
[27] L. Sleire, H.E. Førde, I.A. Netland, L. Leiss, B.S. Skeie, P.Ø. Enger, Drug repurposing in cancer, Pharmacol. Res. 124 (2017) 74–91, https://doi.org/10.1016/j.phrs.2017.07.013.
[28] J.J.P. Stewart, MOPAC: a semiempirical molecular orbital program, J. Comput. Aided Mol. Des. 4 (1) (1990) 1–103, https://doi.org/10.1007/bf00128336.
[29] A. Tropsha, Recent trends in statistical QSAR modeling of environmental chemical toxicity, EXS 101 (2012) 381–411, https://doi.org/10.1007/978-3-7643-8340-4_13.
[30] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest validation is the absolute essential for successful application and interpretation of QSPR models, QSAR Comb. Sci. 22 (1) (2003) 69–77. Retrieved from, http://onlinelibrary.wiley.com/doi/10.1002/qsar.200390007/abstract.