



QSAR modeling of PET imaging agents for the diagnosis of Parkinson's disease targeting dopamine receptor

Priyanka De¹ · Kunal Roy¹

Received: 19 July 2020 / Accepted: 12 October 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Dopamine (D2) receptor has emerged as a potent drug target for the diagnosis and treatment of Parkinson's disease (PD). Radiolabelled imaging such as positron emission tomography (PET) has been recognized as an important tool in medicinal chemistry useful for the early diagnosis of PD. The present study explores quantitative structure—activity relationship analysis of 34 PET imaging agents targeted toward dopamine D2 receptor. The dataset division into training and test sets was done using Euclidean distance division method, while the feature selection was done by double cross-validation-genetic algorithm method. Finally, a five-descriptor partial least squares regression model was derived after carrying out the best subset selection applied on the significant descriptors. The developed model showed robustness in terms of statistical parameters. Finally, the structural information derived from the model descriptors gives an insight for the development of new candidate D2-PET imaging for the use in PD.

Keywords Positron emission tomography · Parkinson's disease · Quantitative structure—activity relationship

1 Introduction

Parkinson's disease (PD) is considered as the second most common progressive neurodegenerative disorder associated with a selective degeneration of the dopaminergic neurons in the substantia nigra pars compacta and loss of projecting nerve fibers in the striatum. It is estimated that more than 10 million people are living with PD worldwide and the occurrence of PD increases with age [1]. About four percent of people with PD are diagnosed before 50 years of age, and men are more prone to this disease than women (about 1.5 times more) [1]. The neurons involved in this disease control the motor movements like resting tremor, muscular

rigidity, bradykinesia, and postural imbalance [2]. Patients with this disease also experience a combination of non-motor symptoms like sleep disturbances, dementia, fatigue, anxiety, depression, apathy, cognitive impairment, olfactory dysfunction, pain, sweating and constipation [3].

Neuroimaging studies are non-invasive methods which help in providing an in vivo image of the nigrostriatal dopaminergic system and further assessment of the extent of neuronal loss associated with PD. Radioactive tracers that selectively bind with dopamine receptors are involved in positron emission tomography (PET) imaging and lately single photon emission computed tomography (SPECT) imaging for research and clinical purposes [4]. PET imaging is a powerful analytical tool which is able to detect in vivo changes in the brain function [5]. PET imaging involves quantification of brain metabolism, abundance of a receptor and its binding in different neurotransmitter systems, and alterations in blood flow in specific region in the brain [5]. PET imaging is considered better than SPECT imaging in terms of accuracy and its regional distributions [6]. Heiss and Hilker (2003) [7] studied that the radiotracer ¹⁸F-fluorodopa (FDOPA) is capable of measuring dopamine deficiency, both its synthesis and storage at the pre-synaptic striatal nerve endings, thus allowing FDOPA-PET in the diagnosis of PD in early disease stages. Wu et al. [8] characterized the clinical

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00214-020-02687-9>) contains supplementary material, which is available to authorized users.

Published as part of the special collection of articles "Festschrift in honour of Prof. Ramon Carbó-Dorca".

✉ Kunal Roy
kunalroy_in@yahoo.com; kunal.roy@jadavpuruniversity.in

¹ Drug Theoretics and Cheminformatics Laboratory,
Department of Pharmaceutical Technology, Jadavpur
University, Kolkata 700032, India

features and associated cerebral glucose metabolism pattern of cognitive impairments in (PD) using ^{18}F -fluorodeoxyglucose (^{18}F -FDG) PET imaging. Glaab et al. integrated blood metabolomics data with PET imaging information which gave better diagnostic discrimination power in understanding cellular processes, including oxidative stress response and inflammation [9].

There is a continuous search of new compounds with improved properties and lowered toxicity which takes enormous human resource and cost into its requirement. Thus, theoretical approaches are gaining more importance among the pharmaceutical and chemical industries enabling logical design of pharmaceutical agents. Currently, quantitative structure-activity relationship (QSAR) has gained great interest in the process of modern drug discovery and design [10, 11]. The study attempts to build a relationship between the chemical properties with a well-defined endpoint as the compounds' activity (QSAR) or property (QSPR) or toxicity (QSTR). QSAR acts as an effective tool in the prediction of biological response (activity/property/toxicity) of existing chemical compounds.

In the present study, we have developed a QSAR model with two-dimensional (2D) molecular descriptors to explore the correlations of the molecular structure of a series of PET tracers against the binding affinity of dopamine (D2) receptor.

2 Materials and methods

2.1 Dataset

Dopamine (D2) receptor binding affinity (K_i) data of 34 PET imaging agents were taken from different literature as mentioned in Table 1. The experimental binding affinity for all the compounds was measured using the same assay protocol, i.e., rat striatal homogenate (RSH) assay method. This datum was applied in the development of a 2D-QSAR model to determine the essential structural features required for good binding to the D2 receptor. The binding affinity (K_i) values for the PET imaging agents were converted to their negative logarithm ($\text{p}K_i$) form and then used for modeling. The compounds were represented using the MarvinSketch software [12] with proper aromatization and addition of hydrogen bond as necessary.

2.2 Molecular descriptors

QSAR models were developed using a selected class of two-dimensional molecular descriptors. The descriptors were *E*-state indices, connectivity, constitutional, functional, 2D atom pairs, ring, atom-centered fragments, and molecular property descriptors. These descriptors were calculated

Table 1 Dataset compounds with their observed binding affinity (in $\text{p}K_i$)

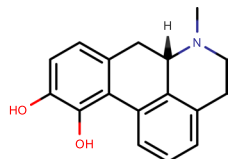
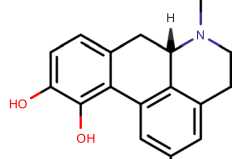
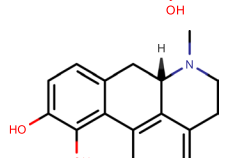
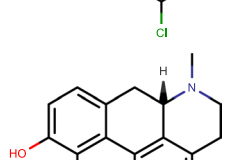
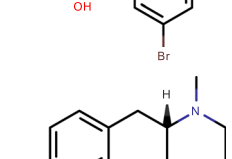
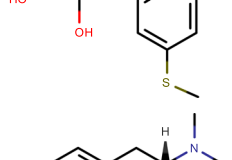
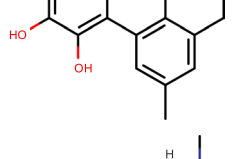
Compound no	Structure	$\text{p}K_i$	Reference
1		2.321	[13]
2		4.420	[14]
3*		2.652	[14]
4		2.752	[14]
5		2.262	[15]
6		2.684	[13]
7*		3.951	[14]

Table 1 (continued)

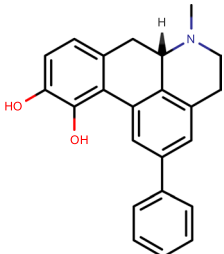
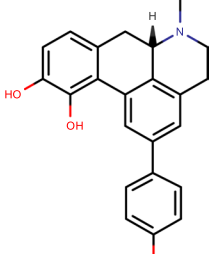
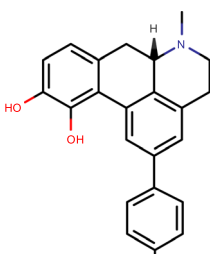
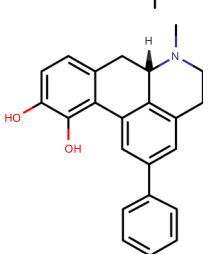
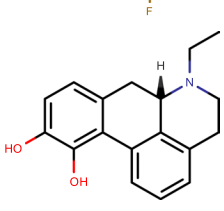
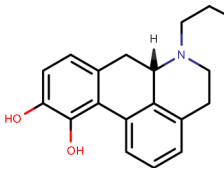
Compound no	Structure	pKi	Reference
8		2.932	[13]
9		3.401	[13]
10		1.460	[16]
11		1.839	[16]
12		4.658	[17]
13		4.097	[17]

Table 1 (continued)

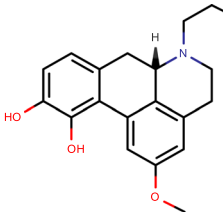
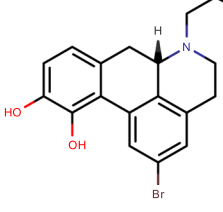
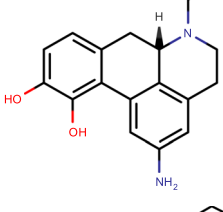
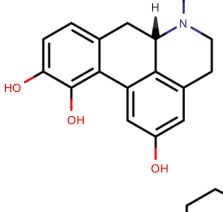
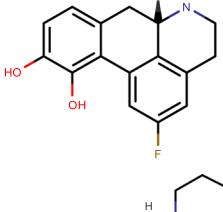
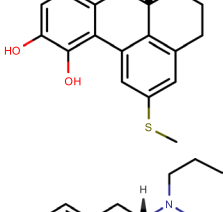
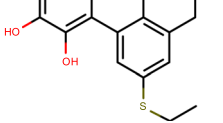
Compound no	Structure	pKi	Reference
14		4.770	[14]
15		4.824	[18]
16		4.036	[18]
17		5.276	[14]
18		5.921	[18]
19*		3.428	[15]
20		3.108	[15]

Table 1 (continued)

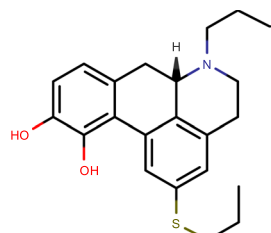
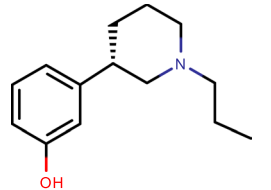
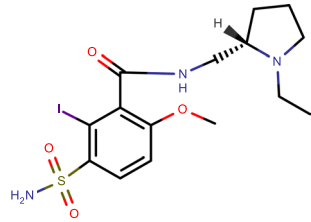
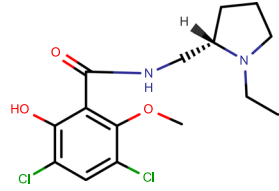
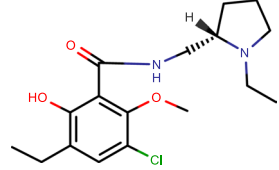
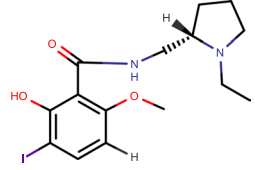
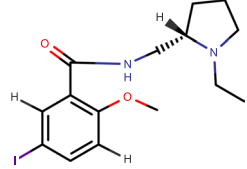
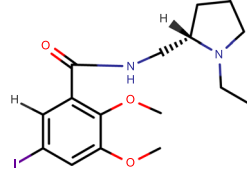
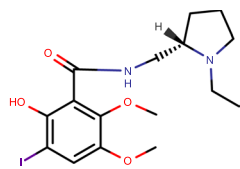
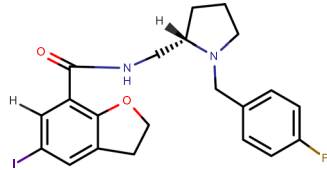
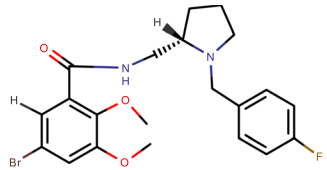
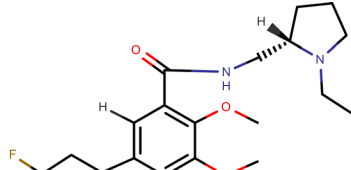
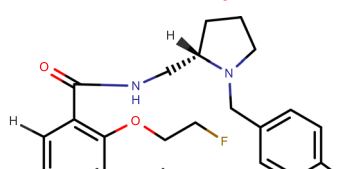
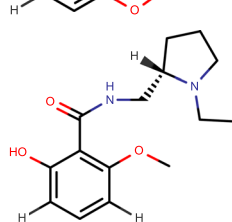
Compound no	Structure	pKi	Reference
21		2.807	[15]
22*		3.114	[19]
23		3.824	[20]
24		3.959	[20]
25*		5.046	[20]
26*		4.367	[20]
27		3.523	[20]
28*		5.602	[20]

Table 1 (continued)

Compound no	Structure	pKi	Reference
29		5.721	[20]
30		4.975	[20]
31		4.833	[20]
32		5.699	[20]
33		2.886	[20]
34		2.507	[21]

Compounds marked with "*" are test set compounds

using Dragon 7 [22] descriptor calculator. A total of 403 Dragon descriptors were calculated. Before the development of the QSAR model, the data were curated [23] by removing intercorrelated ($|r| > 0.95$), constant (variance < 0.0001), and other noisy and redundant data by using data pretreatment software developed in our laboratory and available from <https://dtclab.webs.com/software-tools>. After data pretreatment, the number of descriptors was reduced to 179.

2.3 Dataset splitting

Splitting of the dataset into training and test sets is a vital step in QSAR modeling, and it enables the development of a robust and well-validated model. Data division must be done in such a way that the points representing both training and test set are well scattered within the whole descriptor space defined by the entire dataset. The training set is used for model development and the test set for model validation. The division of the dataset was executed by one of the most extensively used methods, Euclidean distance division method, where the Euclidean distances for all of the compounds in the dataset are calculated and the compounds are then sorted, based on the Euclidean distance [24].

2.4 Variable selection and model development

The main aim of the present study is to develop a well-validated QSAR model to understand the binding of PET imaging agents toward dopamine (D2) receptor for the diagnosis of Parkinson's disease. Critical selection of statistically significant descriptors ensures improvement in the quality of the model. Prior to development of the QSAR model, we have extracted a number of significant descriptors using double cross-validation-genetic algorithm (DCV-GA) approach applied on the training set compounds [25–27]. Finally, a partial least squares (PLS) [28] regression model was generated using descriptors selected from the best subset selection (BSS).

Double cross-validation (DCV) is an attractive statistical design which combines both model generation and model assessment with the aim to produce better models [25, 29]. Sometimes the fixed composition of a training set can lead to biased descriptor selection. DCV method helps in better descriptor selection by dividing the training set into 'n' calibration and validation sets. This results in diverse compositions of the modeling set, thus removing any bias in descriptor selection. DCV technique consists of two nested cross-validation loops commonly known as internal and external cross-validation loops. In the external loop, the data objects are split randomly into disjoint subsets known as training set compounds and test set compounds. The training set compounds are involved in the internal loop for the purpose of model development and model selection, and the test set is used solely for the intention of checking model predictivity. Further, in the internal loop, the training set compounds are repetitively split into calibration (construction) and validation sets by employing the *k*-fold cross-validation technique (here, *k* = 10) [29] and producing *k* iterations to construct calibration and validation sets.

The calibration objects are used to derive different models by altering the tuning parameter(s) of the model (i.e., the descriptors), whereas the validation objects are used to guess the models' error. The model with the lowest cross-validated error is selected. The test compounds in the outer loop are employed to assess the predictive performance of the selected model.

In the current study, descriptor selection in the DCV platform was done using genetic algorithm (GA) approach. GA is a model optimization approach with an algorithm inspired by the theory of evolution [26]. GA has five basic steps: (i) coding of variables; (ii) initiation of population; (iii) evaluation of the response; (iv) reproduction; and (v) mutation. Steps (iii) to (v) are repeated until a termination criterion is reached. The criterion can be based on a lack of improvement in the response or simply on a maximum number of generations or on the total time allowed for the elaboration.

2.5 Statistical validation metrics

Validation of the robustness and predictive ability of the developed models is a very crucial step in a QSAR study. A meticulous examination of the statistical quality of the developed model has been done to judge the robustness in terms of reliability and predictivity measures using various internal and external validation parameters. For determining the quality of the developed model, statistical parameters like determination coefficient R^2 and explained variance R_a^2 were calculated. Other parameters including internal predictivity parameters such as predicted residual sum of squares (PRESS) and leave-one-out cross-validated correlation coefficient (Q^2_{LOO}) were also calculated along with external predictivity parameters like R^2_{pred} or Q^2_{F1} , Q^2_{F2} , and concordance correlation coefficient (CCC) [30]. Further, we have also calculated r_m^2 metrics (i.e., r_m^2 and Δr_m^2) for both training and test set compounds [31]. Validation using mean absolute error (MAE)-based criteria for both external and internal validation was done [32]. The Q^2_{ext} -based criteria do not always interpret the correct prediction quality because of the impact of the response range as well as the distribution of the values of the response in both the training and test set compounds; so MAE was calculated to check the average error [32]. Figure 1 shows the flowchart of the present work methodology.

3 Results and discussion

3.1 Modeling binding affinity of PET tracers toward dopamine (D2) receptor

The final PLS model of three latent variables (LVs) consisted of five descriptors that explains the binding properties of the

PET radioligands toward dopamine receptor. The final model is given below:

$$\text{pKi} = 4.512 - 0.184 \times \text{SaaCH} - 1.554 \times \text{B08}[C-S] + 0.060 \times \text{SsF} - 2.350 \times \text{B10}[N-F] + 1.425 \times \text{B10}[C-O]$$

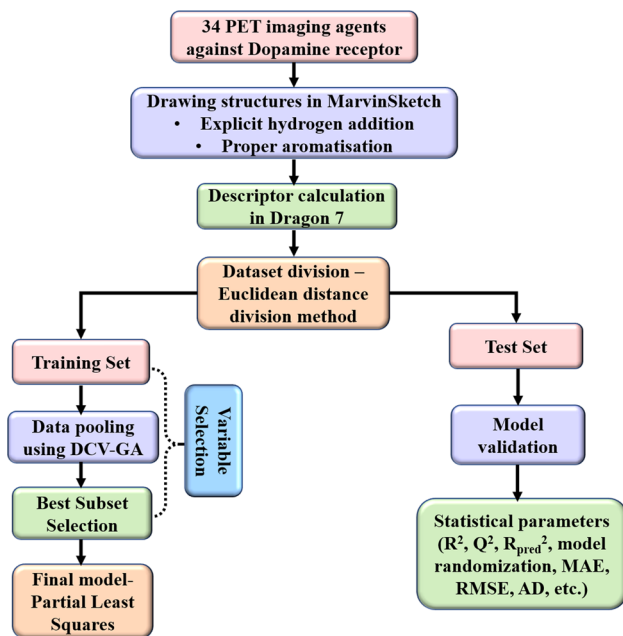


Fig. 1 Flowchart of the present work methodology

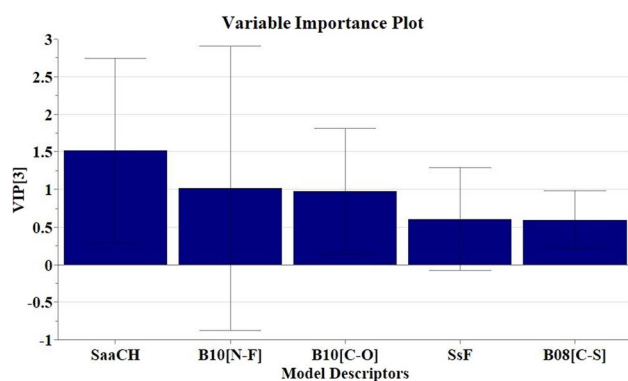


Fig. 2 Variable importance plot of the PLS model

Table 2 Descriptor meaning and their contribution

Serial no	Descriptor	Descriptor type	Contribution	Discussion
1	SaaCH	Atom-type E-state	-ve	Sum of the atom-type E-state values for aromatic -CH groups
2	B08[C-S]	2D atom pairs	-ve	Presence or absence of carbon and sulfur at the topological distance 8
3	SsF	Atom-type E-state	+ve	Sum of the atom-type E-state values for -F fragments
4	B10[N-F]	2D atom pairs	-ve	Presence or absence of nitrogen and fluorine at the topological distance 10
5	B10[C-O]	2D atom pairs	+ve	Presence or absence of carbon and oxygen at the topological distance 10

$$\begin{aligned} n_{\text{training}} &= 27, R^2 = 0.731, R_{\text{adj}}^2 = 0.696, Q^2 = 0.623, \overline{r_{m(\text{LOO})}^2} \\ &= 0.507, \Delta r_{m(\text{LOO})}^2 = 0.159, \text{MAE}(\text{train}) \\ &= 0.528, \text{SD}(\text{train}) = 0.550, \text{PRESS} = 15.392 \end{aligned}$$

$$\begin{aligned} n_{\text{test}} &= 7, Q_{F1}^2 = 0.687, Q_{F2}^2 = 0.664, \overline{r_{m(\text{test})}^2} = 0.742, \Delta r_{m(\text{test})}^2 = 0.116, \\ \text{MAE}(\text{test}) &= 0.505, \text{SD}(\text{test}) = 0.280, \text{CCC}(\text{Test}) = 0.812 \end{aligned}$$

3.2 Mechanistic interpretation

The variable importance plot (VIP) (Fig. 2) gives an idea about the influence of the individual descriptors on the model and thereby on the binding affinity [33]. The order of importance of the descriptors was found as follows: SaaCH, B10[N-F], B10[C-O], SsF, and B08[C-S]. The VIP gives an understanding that descriptors SaaCH and B10[N-F] are highly influential due to their VIP scores being more than one. The regression coefficient plot (not shown) provides a basic understanding about the contribution of the individual descriptor on the model [28]. It is seen that the descriptors SaaCH, B08[C-S], and B10[N-F] negatively contribute to the response, while the descriptors SsF and B10[C-O] positively contribute to the response. The details of the descriptors and their contributions are given in Table 2 and also explained below in detail. The observed vs predicted scatter plot is shown in Fig. 3.

The E-state indices descriptor SaaCH gives idea on the sum of the atom-type E-state values for aromatic -CH groups. From the regression coefficient of the descriptor, it can be inferred that aromaticity hinders the binding of the PET compounds to the D2 receptor as in compounds **8** (SaaCH = 18.392) (Fig. 4), **10** (SaaCH = 16.63), and **11** (SaaCH = 14.214). These compounds are aromatic and have high SaaCH values, and they have lower binding affinity values (pKi = 2.931, 1.460, and 1.839). Further, in compounds like **29** and **32**, aromaticity is less as compared to the previously mentioned compounds, thus having lower values for the descriptor (SaaCH = 3.583 and 1.640, respectively). These compounds have better binding affinity (compound **29** (pKi = 5.700) and compound **32** (pKi = 5.721)) toward dopamine receptor.

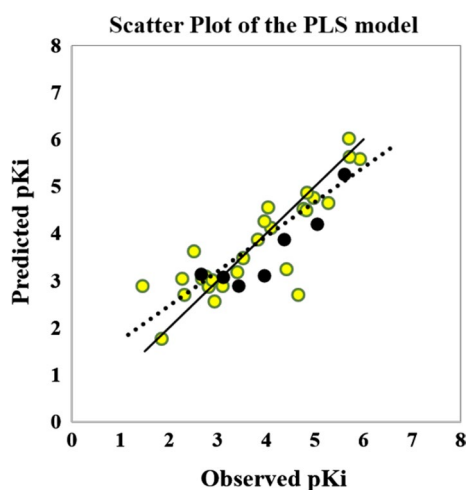


Fig. 3 Observed versus predicted pKi plot

The next important descriptor is B10[N-F] (2D atom pair type), and the negative contribution implies that the presence of nitrogen and fluorine at the topological distance 10 will hinder the binding affinity seen in compounds **11** (B10[N-F] = 1; pKi = 1.838) (Fig. 4) and **33** (B10[N-F] = 1; pKi = 2.886). Further, the absence of this fragment will increase the binding affinity as observed in compounds

29 (B10[N-F] = 0; pKi = 5.700) and **32** (B10[N-F] = 0; pKi = 5.721). The effect of the electronegativity of fluorine atom on nitrogen is a determining factor for the good binding which is latter explained while studying the descriptor SsF. The closeness between nitrogen and fluorine atom explains how the binding will occur.

B10[C-O] is another 2D atom pair descriptor representing the presence or absence of C-O fragment at the topological distance 10. The descriptor positively influences the binding affinity of the PET tracers toward dopamine receptor as seen in compounds **18** (B10[C-O] = 1; pKi = 5.921) (Fig. 4), **29** (B10[C-O] = 1; pKi = 5.721), and **32** (B10[C-O] = 1; pKi = 5.700). The presence of this kind of fragment affects the electronegativity of the compounds essential for binding. The absence of this fragment on the other hand decreases the dopamine binding affinity observed in compounds like **1** (pKi = 2.321) and **5** (pKi = 2.262).

The *E*-state values for the descriptor SsF depend on the number of fluorine atoms present in a PET tracer molecule. From the regression coefficient, it can be understood that with increasing fluorine atoms the binding affinity also increases as observed in **18** (SsF = 14.107; pKi = 5.921), **32** (SsF = 12.490; pKi = 5.698) (Fig. 4), and **31** (SsF = 13.108; pKi = 4.833). The electronegative fluorine atom is presumed to decrease electron charge density on nitrogen atoms. This

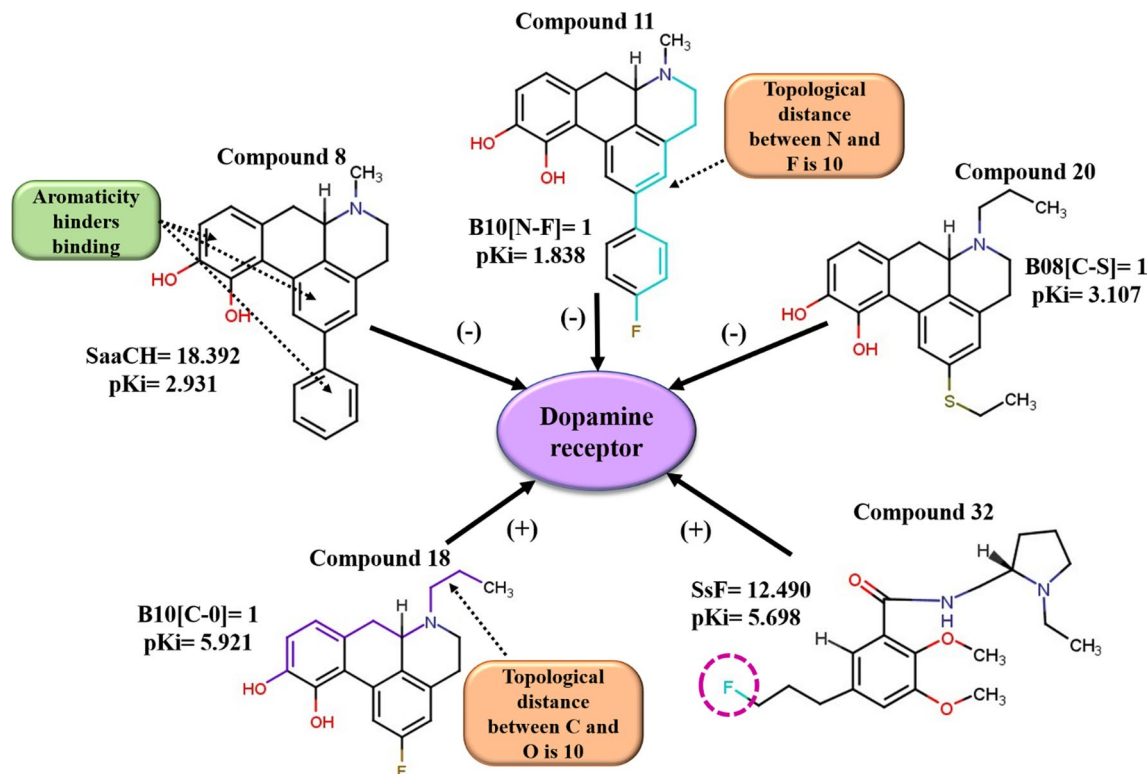


Fig. 4 Descriptors appearing in the PLS model and their contribution

reduces nitrogen basicity and its prospect to get protonated at physiological pH which is a basic requirement for good binding to dopamine receptors [34].

The least important descriptor is B08[C-S], which is also a 2D atom pair descriptor and gives an idea of the presence or absence of C-S fragment at a topological distance 8. The negative contribution suggests that the presence of this fragment will result in a decreased binding affinity toward the dopamine receptor which is observed in compounds **21** ($pK_i = 2.807$) and **20** ($pK_i = 3.107$) (Fig. 4). Alternatively, compounds like **18** ($pK_i = 5.921$), **29** ($pK_i = 5.721$) and **32** ($pK_i = 5.698$) have no such fragment, thus having higher binding affinity.

From the descriptors and their contributions, we can draw an inference that the oxygen for B10[C-O] and fluorine for SsF impart an electronegative character to the PET ligands which plays an essential role for the good dopamine (D2) binding.

3.3 Plot Interpretation

1 *Loading Plot*— This plot gives a relationship between the X-variables (i.e., the descriptors) and Y-variable (i.e., response) [35]. In Fig. 5, five X-variables and one Y-variable are shown. Generally, the plot is developed with the first and second components. A loading plot provides an insight about how much a variable contributes to a model and which variable provides the maximum footprint. For interpretation, the distance from the origin is taken under consideration. Descriptors which are similar in nature and providing similar contribution are correlated and grouped together. Descriptors which are situ-

ated far away from the plot origin are supposed to have greater impact on the Y-response. From the loading plot it, is seen that descriptors SaaCH and B10[N-F] are far away from the plot origin supporting their higher influence also explained by the VIP. The positive or negative algebraic symbol is also taken under consideration in a PLS plot. Features explained by descriptors SsF and B10[C-O] are beneficial for binding because of their closeness to pK_i in the plot. On the other hand, SaaCH, B10[N-F] and B08[C-S] are present in the negative side of the plot origin and are detrimental for good binding.

2 *Score Plot*— Figure 6 shows the distribution of the compounds in the latent variable space as defined by the scores. We have plotted the scores of the first two components t_1 and t_2 . The applicability domain of the model is designated by the ellipse, as defined by Hotelling's t^2 . Hotelling's t^2 defines multivariate generalization of Student's t test. The method offers a check for compounds adhering to multivariate normality [36]. Compounds which are situated near each other in the plot have similar properties, whereas compounds which are far from each other have dissimilar properties with respect to their binding affinity toward dopamine receptor. As an example, we can take compounds **14**, **15**, **16**, and **17** which are clubbed together as a group on the plot space and can be considered to be with similar properties. On the other hand, compounds **18** and **12** are completely located on the opposite side of the origin and far from each other and they represent heterogeneity in their properties. Since there are no compounds out of the ellipse, we can conclude that there are no outliers according to this method.

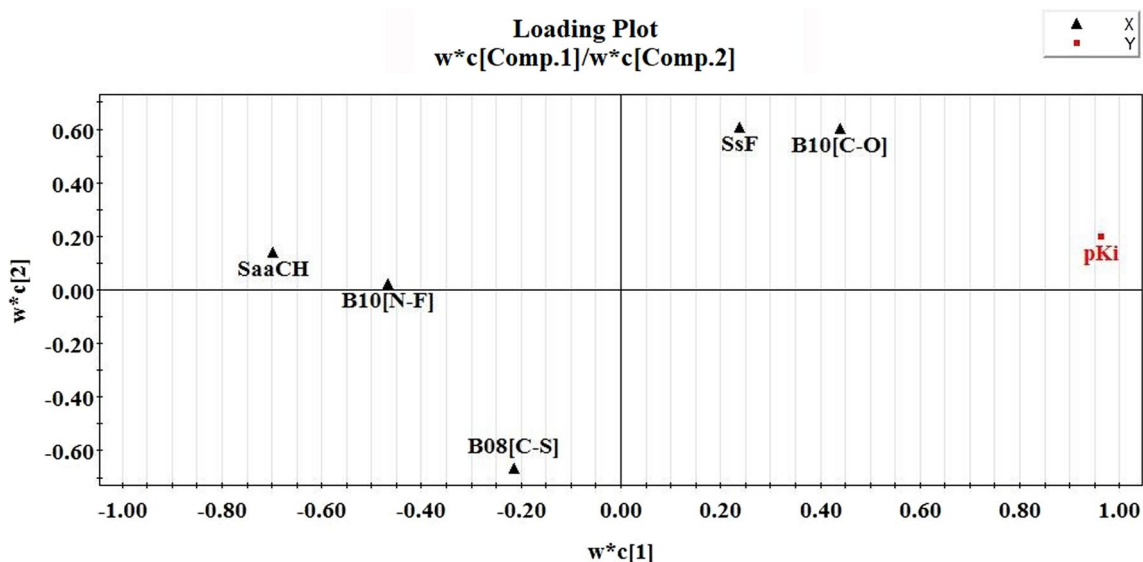
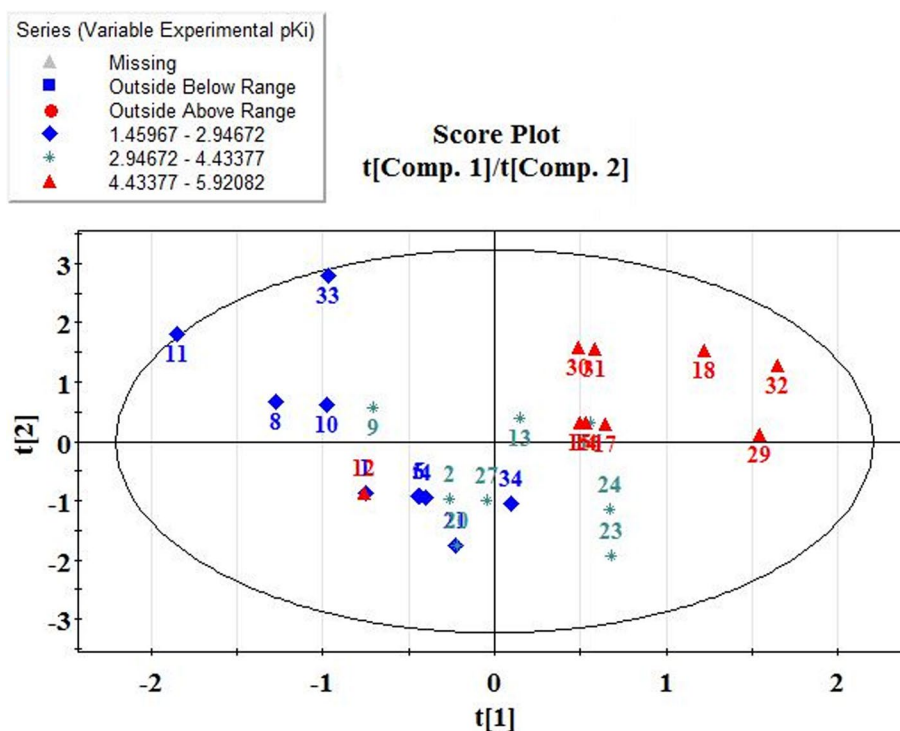


Fig. 5 Loading plot of the PLS model

Fig. 6 Score plot of the PLS model

Y-Randomization Plot— Model randomization gives a notion about the model significance and ensures that the model is not an outcome of a chance correlation [37]. A randomized model is generated by the development of multiple models by shuffling or reordering different combinations of *X*-or *Y*-variables (here *Y*-variable only) and based on the fit of the reordered model. In the present study, we have used 100 permutations which can be changed according to the choice of the user. A randomized model should have very poor statistics. The R^2

and Q^2 values for the random models (*Y*-axis) are plotted against correlation coefficient between the original *Y* values and the permuted *Y* values (*X*-axis); the R_y^2 intercept should not exceed 0.3, and the Q_y^2 intercept should not exceed 0.05. Figure 7 shows the correlation between original *Y*-vector and permuted *Y*-vector versus cumulative R_y^2 , cumulative Q_y^2 plot where R_y^2 intercept = 0.09 and Q_y^2 intercept = -0.393 proving the model is robust and non-random.

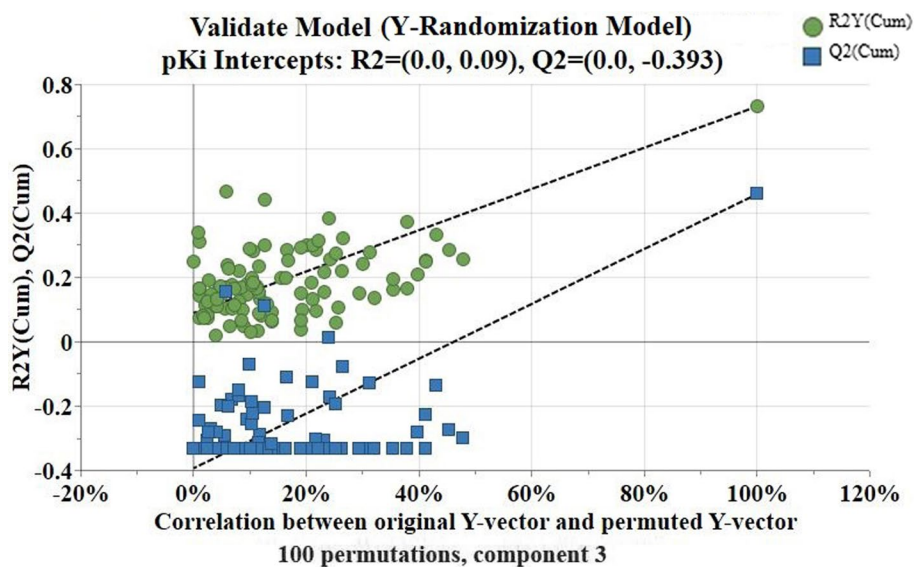
Fig. 7 *Y*-randomization plot of the PLS model

Fig. 8 DModX applicability domain of the training set

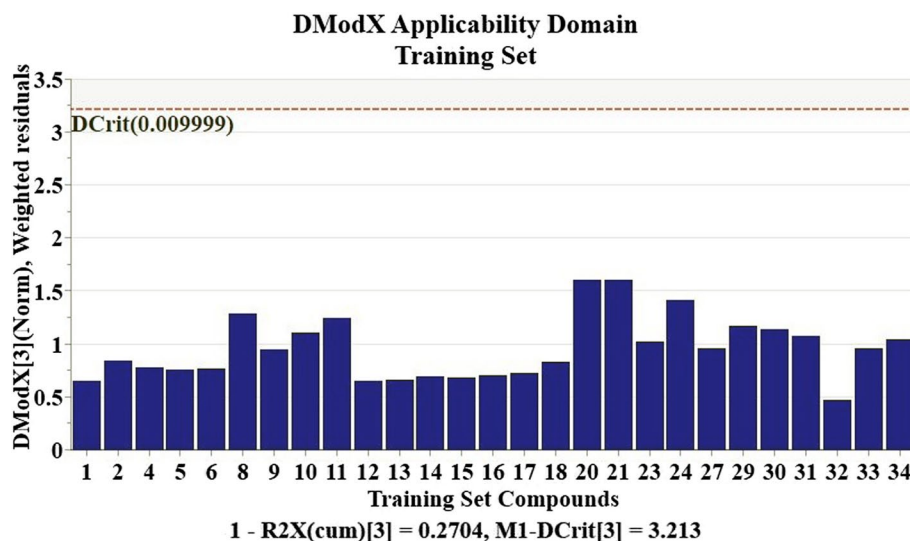
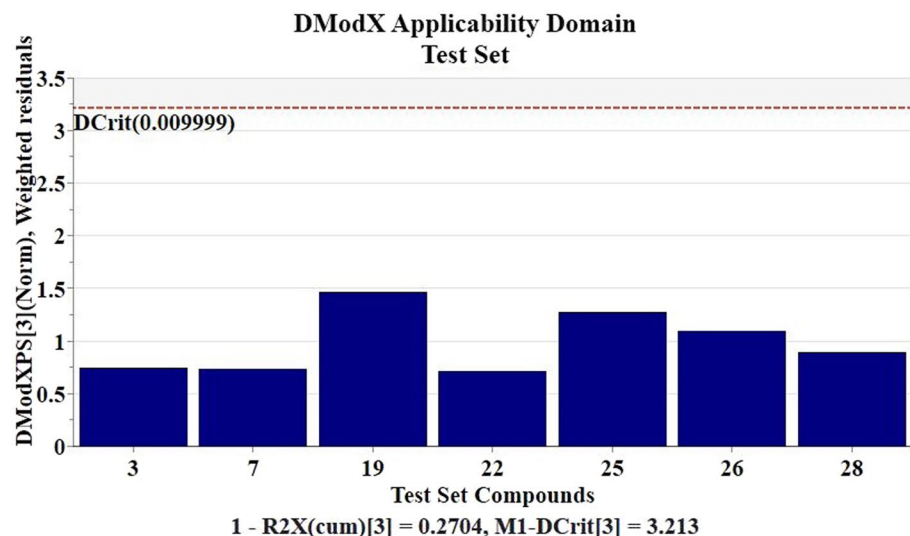


Fig. 9 DModX applicability domain of the test set



3 *Applicability Domain (AD)*— The prediction reliability of a particular model is dependent on its applicability domain (AD) assessment. Applicability domain (AD) “represents a chemical space from which a model is derived and where a prediction is considered to be reliable” [38]. The AD evaluation was done using the DModX (distance to model) in the X -space using SIMCA 16.0.2 software available at <https://landi.ng.umetrics.com/downloads-simca>. The AD plots are given in Figs. 8 and 9 and for training and test sets, respectively, and it is found that there are no outliers in case of training set, and none of the compounds are outside AD in case of the test set at 99% confidence level (D -crit = 0.009999, M -Dcrit [3] = 3.213).

4 Conclusion

In vivo imaging targeting dopamine receptor is a subject of extensive studies nowadays. Dopamine plays a vital role in controlling the pathophysiology of Parkinson’s disease. Hence, it can be treated as a suitable target in controlling the disease. The present study aims in the development of a 2D QSAR model of a group of 34 PET imaging agents having affinity toward dopamine D2 receptor. The 2D QSAR model developed is simple and interpretable and provides knowledge about the basic structural features required for good dopamine binding. The use of simple two-dimensional descriptors reduces the need of time-consuming computational approaches of conformational analysis or energy

minimization; thus, the developed model may be suitable for the quick screening purposes.

Acknowledgements Special issue to Celebrate 80th Birthday of Prof Ramon Carbó-Dorca

Funding PD thanks Indian Council of Medical Research, New Delhi, for awarding with a Senior Research Fellowship. KR thanks Science and Engineering Research Board (SERB), New Delhi, for financial assistance under the MATRICS scheme (File number MTR/2019/000008). Financial assistance from DAE-BRNS under the scheme 36 (3)/14/08/2017-BRNS is also thankfully acknowledged.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Parkinson's Foundation (2020) Understanding Parkinson's, Statistics. <https://www.parkinson.org/Understanding-Parkinsons/Statistics>. Accessed on 02 July 2020
- Jankovic J (2008) Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry* 79(4):368–376
- Barone P (2010) Neurotransmission in Parkinson's disease: beyond dopamine. *Eur J Neurol* 17(3):364–376
- Antonini A, Moresco R, Gobbo C, De Notaris R, Panzacchi A, Barone P, Calzetti S, Negrotti A, Pezzoli G, Fazio F (2001) The status of dopamine nerve terminals in Parkinson's disease and essential tremor: a PET study with the tracer [11-C] FE-CIT. *Neurol Sci* 22(1):47–48
- Politis M, Piccini P (2012) Positron emission tomography imaging in neurological disorders. *J Neurol* 259(9):1769–1780
- De P, Roy J, Bhattacharyya D, Roy K (2020) Chemometric modeling of PET imaging agents for diagnosis of Parkinson's disease: a QSAR approach. *Struct Chem*. <https://doi.org/10.1007/s11224-020-01560-6>
- Heiss WD, Hilker R (2004) The sensitivity of 18-fluorodopa positron emission tomography and magnetic resonance imaging in Parkinson's disease. *Eur J Neurol* 11(1):5–12
- Wu L, Liu FT, Ge JJ, Zhao J, Tang YL, Yu WB, Yu H, Anderson T, Zuo CT, Chen L (2018) Clinical characteristics of cognitive impairment in patients with Parkinson's disease and its related pattern in 18F-FDG PET imaging. *Hum Brain Mapp* 39(12):4652–4662
- Glaab E, Trezzi JP, Greuel A, Jäger C, Hodak Z, Drzezga A, Timmermann L, Tittgemeyer M, Diederich NJ, Eggers C (2019) Integrative analysis of blood metabolomics and PET brain neuroimaging data for Parkinson's disease. *Neurobiol Dis* 124:555–556
- Roy K (2018) Quantitative structure-activity relationships (QSARs): a few validation methods and software tools developed at the DTC laboratory. *J Indian Chem Soc* 95(12):1497–2150
- Gramatica P (2020) Principles of QSAR modeling: comments and suggestions from personal experience. *IJQSPR* 5(3):61–97
- MarvinSketch software (2020). <https://www.chemaxon.com> Accessed on 25 May 2020
- Sipos A, Kiss B, Schmidt É, Greiner I, Berényi S (2008) Synthesis and neuropharmacological evaluation of 2-aryl-and alkylapomorphines. *Bioorg Med Chem* 16(7):3773–3779
- Gao Y, Baldessarini RJ, Kula NS, Neumeyer JL (1990) Synthesis and dopamine receptor affinities of enantiomers of 2-substituted apomorphines and their N-n-propyl analogs. *J Med Chem* 33(6):1800–1805
- Tóth M, Berényi S, Csutorás C, Kula NS, Zhang K, Baldessarini RJ, Neumeyer JL (2006) Synthesis and dopamine receptor binding of sulfur-containing apomorphines. *Bioorg Med Chem* 14(6):1918–1923
- Søndergaard K, Kristensen JL, Palner M, Gillings N, Knudsen GM, Roth BL, Begtrup M (2005) Synthesis and binding studies of 2-arylapomorphines. *Org Biomol Chem* 3(22):4077–4081
- Gao Y, Ram VJ, Campbell A, Kula NS, Baldessarini RJ, Neumeyer JL (1990) Synthesis and structural requirements of N-substituted norapomorphines for affinity and activity at dopamine D-1, D-2, and agonist receptor sites in rat brain. *J Med Chem* 33(1):39–44
- Baldessarini R, Kula N, Gao Y, Campbell A, Neumeyer J (1991) R (–) 2-fluoro-nn-propylnorapomorphine: a very potent and D2-selective dopamine agonist. *Neuropharmacology* 30(1):97–99
- Vasdev N, Natesan S, Galineau L, Garcia A, Stableford WT, McCormick P, Seeman P, Houle S, Wilson AA (2006) Radiosynthesis, ex vivo and in vivo evaluation of [11C] preclamol as a partial dopamine D2 agonist radioligand for positron emission tomography. *Synapse* 60(4):314–331
- Chumpradit S, Kung M, Billings J, Mach R, Kung H (1993) Fluorinated and iodinated dopamine agents: D2 imaging agents for PET and SPECT. *J Med Chem* 36(2):221–228
- Murphy RA, Kung HF, Kung MP, Billings J (1990) Synthesis and characterization of iodobenzamide analogs: potential D-2 dopamine receptor imaging agents. *J Med Chem* 33(1):171–178
- Dragon version 7 (2016) Kodesrl, Milan, Italy. <https://www.talet.e.mi.it/index.htm>. Accessed on 26 May 2020
- Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. *Mol Inform* 29(6–7):476–488
- Golmohammadi H, Dashtbozorgi Z, Acree WE Jr (2012) Quantitative structure–activity relationship prediction of blood-to-brain partitioning behavior using support vector machine. *Eur J Pharm Sci* 47(2):421–429
- Roy K, Ambure P (2016) The “double cross-validation” software tool for MLR QSAR model development. *Chemom Intell Lab Syst* 159:108–126
- Devillers J (1996) Genetic algorithms in molecular modeling. Academic Press, Cornwall, Great Britain
- Khan PM, Roy K (2018) Current approaches for choosing feature selection and learning algorithms in quantitative structure–activity relationships (QSAR). *Expert Opin Drug Discov* 13(12):1075–1089
- Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* 58(2):109–130
- Baumann D, Baumann K (2014) Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation. *J Cheminform* 6(1):47
- Roy K, Mitra I (2011) On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Comb Chem High Throughput Screen* 14(6):450–474
- Ojha PK, Mitra I, Das RN, Roy K (2011) Further exploring rm2 metrics for validation of QSPR models. *Chemom Intell Lab Syst* 107(1):194–205
- Roy K, Das RN, Ambure P, Aher RB (2016) Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemom Intell Lab Syst* 152:18–33
- Akarachantachote N, Chadcham S, Saithanu K (2014) Cutoff threshold of variable importance in projection for variable selection. *Int J Pure Appl Math* 94(3):307–322

34. Finnema SJ, Bang-Andersen B, Wikstrom HV, Halldin C (2010) Current state of agonist radioligands for imaging of brain dopamine D2/D3 receptors in vivo with positron emission tomography. *Curr Top Med Chem* 10(15):1477–1498
35. De P, Aher RB, Roy K (2018) Chemometric modeling of larvicidal activity of plant derived compounds against zika virus vector *Aedes aegypti*: application of ETA indices. *RSC Adv* 8(9):4662–5467
36. Jackson JE (2005) A user's guide to principal components, vol 587. Wiley, United States of America
37. Topliss JG, Edwards RP (1979) Chance factors in studies of quantitative structure-activity relationships. *J Med Chem* 22(10):1238–1244
38. Gadaleta D, Mangiatordi GF, Catto M, Carotti A, Nicolotti O (2016) Applicability domain for QSAR models: where theory meets reality. *IJQSPR* 1(1):45–63

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.