Journal of the Serbian Chemical Society
JSCS-info@shd.org.rs • www.shd.org.rs/JSCS

ACCEPTED MANUSCRIPT

This is an early electronic version of an as-received manuscript that has been accepted for publication in the Journal of the Serbian Chemical Society but has not yet been subjected to the editing process and publishing procedure applied by the JSCS Editorial Office.

Please cite this article as P. Mishra, S. Nandi, A. Chatterjee, T. Nayek, S. Basak, A. K. Halder, and A. Mukherjee, *J. Serb. Chem. Soc.* (2024) https://doi.org/10.2298/JSC230221039M

This "raw" version of the manuscript is being provided to the authors and readers for their technical service. It must be stressed that the manuscript still has to be subjected to copyediting, typesetting, English grammar and syntax corrections, professional editing and authors' review of the galley proof before it is published in its final form. Please note that during these publishing processes, many errors may emerge which could affect the final content of the manuscript and all legal disclaimers applied according to the policies of the Journal.

Journal of the Serbian Chemical Society
JSCS-info@shd.org.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. **00(0)** 1-16 (2024)
JSCS–12286
Original scientific paper
Published DD MM, 2024

Accepted manuscript

Development of 2D and 3D QSAR models of pyrazole derivatives as acetylcholine esterase inhibitors

PUJA MISHRA[1]*, SUMIT NANDI[1], ANKIT CHATTERJEE[1], TRIDIB NAYEK[1], SOUVIK BASAK[1], AMIT KUMAR HALDER[1,2], ARUP MUKHERJEE[3]

[1]Dr. B.C. Roy College of Pharmacy & Allied Health Sciences, Durgapur, WB, India, [2]LAQV@REQUIMTE/Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, Porto, Portugal, and [3]Department of Biotechnology, Maulana Abul Kalam Azad University of Technology, WB, India

(Received 21 February 2023; revised 2 August 2023; accepted 31 March 2024)

Abstract: The drugs that are most useful in all stages of Alzheimer's disease (AD) are acetylcholinesterase (AChE) inhibitors. The objectives of this work are to generate various QSAR models and to select robust predictive models from corresponding models. Studies were then focused on finding a range of pyrazole-like AChE inhibitors by 2D and 3D QSAR analysis. Genetic algorithm-based multiple linear regression (GA-MLR) provided the statistically robust 2D-QSAR model that depicted the significance of molecular volume and number of multiple bonds along with the presence/absence of specific atom-centred fragments and topological distance between 2D pharmacophoric features. Furthermore, these results were correlated well with the electrostatic and steric contour maps retrieved from the 3D-QSAR (i.e., alignment-dependent molecular field analysis). The 2D QSAR analysis developed a highly statistical and reliable model which was compared with the mechanistic interpretation of 3D structures and their electrostatic and steric field contributions leading to a predictive 3D QSAR model. The molecule-protein interactions elicited by molecular docking corroborated with the field interactions as revealed by 2D-QSAR. Thus, the developed computational models and simulation analyses in the current work provide valuable information for the future design of pyrazole and spiropyrazoline analogs as potent AChE inhibitors.

Keywords: acetylcholinesterase; QSAR analysis; GA-MLR; contour maps; molecular docking.

INTRODUCTION

Alzheimer's disease (AD) was defined as a progressive neurodegenerative disorder, characterized by gradual loss of cholinergic neurons and accumulation of β-amyloid protein in the brain areas like cortex and hippocampus. AD is

* Corresponding author. E-mail: pujam.phe15@itbhu.ac.in, Ph: +91-8340435004
https://doi.org/10.2298/JSC230221039M

manifested by successive impairment of activities of daily living, cognitive and memory deterioration, and a variety of neuropsychiatric symptoms and disturbances. Loss of memory, cognitive decline, impaired performance of activities of daily life, and behavioral changes are hallmarks of the disease. Every 20 years the cases were estimated to double thus leading to a figure of over 120 million affected from AD in Asia by 2050. World Health Organization (WHO) projects that by 2025, about 3/4th of the estimated 1.2 billion people age 60 years and older will reside in developing countries. Alzheimer's disease onset starts with short-term memory impairment that gradually progresses to complete loss of cognitive function, weak performance of activities of daily life, and ultimately death. Research in this area will be beneficial to AD patients.[1–3]

Alzheimer's disease has no proper cure to date thus a molecule that provides symptomatic relief by inhibiting acetylcholinesterase (AChE) enzyme within the brain is studied. Almost 40 such inhibitors were collected from the literature survey and divided into a test set (20% of total compounds and a training set (80% of total compounds). Validated and predictive quantitative structure-activity relationship (QSAR) models were generated based on various feature selection methods. In addition, molecular docking [4] was also employed to reveal that the ionization state of the compounds had an impact on their interaction with AChE [2,4–7].

EXPERIMENTAL

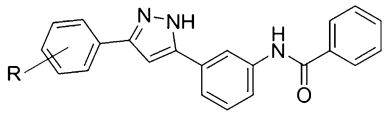
A dataset containing forty pyrazole and spiro pyrazoline analogs with AChE inhibitors was collected from the report of Gutti et al[7]. For the development of the QSAR model, software such as ChemDraw ultra 8.0 April 23, 2003 (Chembridgesoft Corporation, 100 Cambridge Park Drive, Cambridge, MA 02140, USA), Discovery Studio Visualizer v21.1.0.20298, Marvin View tool (Marvin View. Version 18.18.0; ChemAxon: Budapest, Hungary, 2010), alvaDescv.2.0.4 under OCHEM web server,[8,9] MLR plus Validation, Version 1.3 (https://sites.google.com/site/mlrplusvalidation), DTC-QSARv1.0.6[10] and Open3DQSAR (http://www.softsea.com/review/Open3DQSAR.html).[11]

Construction of chemical structures

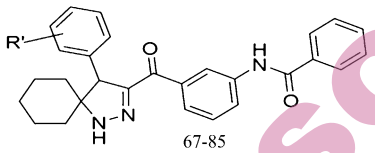
The SMILES notations of 40 compounds (Table I) were first converted to 2D structures by Marvin View tool (Marvin View. Version 18.18.0; ChemAxon: Budapest, Hungary, 2010, https://chemaxon.com/products/marvin), and these structures were subsequently saved as 3D structures (in .sdf file).

Accepted manuscript

TABLE I. Library of compounds SMILES along with biological activity [7]



43-62



67-85

Comp. no.	SMILES	R/R'	pIC_{50} AChE
43	O=C(Nc1cccc(-c2cc(-c3ccccc3)n[nH]2)c1)c1ccccc1	H	5.393
44	O=C(Nc1cccc(-c2cc(-c3ccc(Cl)cc3)n[nH]2)c1)c1ccccc1	4-Cl	5.713
45	O=C(Nc1cccc(-c2cc(-c3ccccc3Cl)n[nH]2)c1)c1ccccc1	2-Cl	5.517
46	O=C(Nc1cccc(-c2cc(-c3ccc(Cl)cc3Cl)n[nH]2)c1)c1ccccc1	2,4-diCl	5.711
47	O=C(Nc1cccc(-c2cc(-c3ccc(Br)cc3)n[nH]2)c1)c1ccccc1	4-Br	5.665
48	O=C(Nc1cccc(-c2cc(-c3cccc(Br)c3)n[nH]2)c1)c1ccccc1	3-Br	5.521
49	O=C(Nc1cccc(-c2cc(-c3ccc(F)cc3)n[nH]2)c1)c1ccccc1	4-F	5.68
50	O=C(Nc1cccc(-c2cc(-c3cccc(F)c3)n[nH]2)c1)c1ccccc1	3-F	5.589
51	COc1ccc(-c2cc(-c3cccc(NC(=O)c4ccccc4)c3)[nH]n2)cc1	4-OMe	5.073
52	COc1cccc(-c2cc(-c3cccc(NC(=O)c4ccccc4)c3)[nH]n2)c1	3-OMe	5.06
53	COc1ccc(-c2cc(-c3cccc(NC(=O)c4ccccc4)c3)[nH]n2)cc1OC	3,4-diOMe	4.818
54	O=C(Nc1cccc(-c2cc(-c3ccc(C(F)(F)F)cc3)n[nH]2)c1)c1ccccc1	4-CF_3	5.627
55	O=C(Nc1cccc(-c2cc(-c3cccc(C(F)(F)F)c3)n[nH]2)c1)c1ccccc1	3-CF_3	5.495
56	O=C(Nc1cccc(-c2cc(-c3ccc(OC(F)(F)F)cc3)n[nH]2)c1)c1ccccc1	4-OCF_3	5.215
57	N#Cc1ccc(-c2cc(-c3cccc(NC(=O)c4ccccc4)c3)[nH]n2)cc1	4-CN	5.417
58	N#Cc1cccc(-c2cc(-c3cccc(NC(=O)c4ccccc4)c3)[nH]n2)c1	3-CN	5.405
59	Cc1ccc(-c2cc(-c3cccc(NC(=O)c4ccccc4)c3)[nH]n2)cc1	4-Me	4.828
60	Cc1ccccc1-c1cc(-c2cccc(NC(=O)c3ccccc3)c2)[nH]n1	2-Me	4.771
61	CC(C)c1ccc(-c2cc(-c3cccc(NC(=O)c4ccccc4)c3)[nH]n2)cc1	4-iPr	4.81
62	O=C(Nc1cccc(-c2cc(-c3cccc4ccccc34)n[nH]2)c1)c1ccccc1	α-naphthyl	4.556
66	O=C(Nc1cccc(C(=O)C2=NNC3(CCCCC3)C2c2ccccc2)c1)c1ccccc1	H	5.705
67	O=C(Nc1cccc(C(=O)C2=NNC3(CCCCC3)C2c2ccc(Cl)cc2)c1)c1cccc1	4-Cl	6.333
68	O=C(Nc1cccc(C(=O)C2=NNC3(CCCCC3)C2c2ccccc2Cl)c1)c1ccccc1	2-Cl	5.706
69	O=C(Nc1cccc(C(=O)C2=NNC3(CCCCC3)C2c2ccc(Cl)cc2Cl)c1)c1ccccc1	2,4-diCl	5.877
70	O=C(Nc1cccc(C(=O)C2=NNC3(CCCCC3)C2c2ccc(Br)cc2)c1)c1cccc1	4-Br	5.943
71	O=C(Nc1cccc(C(=O)C2=NNC3(CCCCC3)C2c2ccccc(Br)c2)c1)c1cccc1	3-Br	5.752
72	O=C(Nc1cccc(C(=O)C2=NNC3(CCCCC3)C2c2ccc(F)cc2)c1)c1cccc1	4-F	6.023
73	O=C(Nc1cccc(C(=O)C2=NNC3(CCCCC3)C2c2ccccc(F)c2)c1)c1cccc1	3-F	5.789
74	COc1ccc(C2C(C(=O)c3cccc(NC(=O)c4ccccc4)c3)=NNC23CCCCC3)cc1	4-OMe	5.635

Accepted Manuscript

75	COc1cccc(C2C(C(=O)c3cccc(NC(=O)c4ccccc4)c3)=NNC23CCCCC3)c1	3-OMe	5.548
76	COc1ccc(C2C(C(=O)c3cccc(NC(=O)c4ccccc4)c3)=NNC23CCCCC3)cc1OC	3,4-diOMe	5.521
77	O=C(Nc1cccc(C(=O)C2=NNC3(CCCCC3)C2c2ccc(C(F)(F)F)cc2)c1)c1ccccc1	4-CF_3	5.838
78	O=C(Nc1cccc(C(=O)C2=NNC3(CCCCC3)C2c2cccc(C(F)(F)F)c2)c1)c1ccccc1	3-CF_3	5.716
79	O=C(Nc1cccc(C(=O)C2=NNC3(CCCCC3)C2c2ccc(OC(F)(F)F)cc2)c1)c1ccccc1	4-OCF_3	5.65
80	N#Cc1ccc(C2C(C(=O)c3cccc(NC(=O)c4ccccc4)c3)=NNC23CCCCC3)cc1	4-CN	5.75
81	N#Cc1cccc(C2C(C(=O)c3cccc(NC(=O)c4ccccc4)c3)=NNC23CCCCC3)c1	3-CN	5.707
82	Cc1ccc(C2C(C(=O)c3cccc(NC(=O)c4ccccc4)c3)=NNC23CCCCC3)cc1	4-Me	5.329
83	Cc1ccccc1C1C(C(=O)c2cccc(NC(=O)c3ccccc3)c2)=NNC12CCCCC2	2-Me	5.23
84	CC(C)c1ccc(C2C(C(=O)c3cccc(NC(=O)c4ccccc4)c3)=NNC23CCCCC3)cc1	4-iPr	5.146
85	O=C(Nc1cccc(C(=O)C2=NNC3(CCCCC3)C2c2cccc3ccccc23)c1)c1ccccc1	α-naphthyl	4.535

Descriptors generation

The 3D chemical structures of 40 AChE inhibitors were subjected to descriptors calculation using alvaDesc v.2.0.4 (https://www.alvascience.com/alvadesc/) under OCHEM web server. The 3D descriptors were calculated after geometrical optimization performed separately with Corina (molecular mechanical) and ULYSSES (quantum chemical PM6[7,12]). The calculated descriptors for the dataset compounds were combined with their respective pIC_{50} values to generate a data-matrix for 2D-QSAR model generation.

Model generation

To initiate the model development procedure the data set (n=40) was divided into a training set (consisting of 80% of the total number of compounds) and a test set (20% of the total number of compounds) using a Java-based platform DTC-QSAR[8,10,13–15] tool (https://dtclab.webs.com/software-tools) using random division technique with multiple random state values (i.e., by applying different seed values like 5, 10, 14, etc). Since the major objective of the current investigation is to generate interpretable 2D-QSAR models, we employed selected categories of alvaDesc descriptors and these belong to constitutional descriptors, functional group counts, 2D-atom pairs, drug-like indices, ring descriptors, atom-centered fragments, pharmacophore descriptors, and molecular properties. The linear interpretable 2D-QSAR models were also generated using the DTC-QSAR tool, freely accessed from http://teqip.jdvu.ac.in/QSAR_Tools/. The Genetic Algorithm-Multiple Linear Regression (GA-MLR) method, implemented in DTC-QSAR-tool, was employed for regression-based 2D-QSAR model generation. Data treatment was carried out by setting a variance cut-off of 0.001 (to remove constant and near-constant descriptors) and a correlation cut-off of 0.99 (to eliminate

Accepted manuscript

highly inter-correlated descriptors). During model development, a maximum of 5 descriptors were allowed.

Model validation procedures

Model validation was performed using both external and internal validation parameters which included Leave-One-Out (LOO) cross-validated determination coefficient (Q^2_{LOO}) and predcited R^2 (R^2_{pred}) as well as related parameters which have been described in detail in supplementary information.

3D-QSAR Modelling

The 3D QSAR is used to explore the relationship between three-dimensional molecular structures and their measured biological activity. The 3D-QSAR depends mainly on the biologically active conformers of the ligands (compound **67**) and their structural alignments.[19] The atom-based alignment or unsupervised rigid-body molecular alignment method[16] was used for aligning the dataset compounds. The 3D structures of the ligands were first minimized using the OpenBabel 'obminimize' tool by the steepest descent method with a maximum of 10000 runs. After minimization, the structures were eventually allowed to generate 100 conformations by superimpositions followed by alignment rdMolAlign.GetCrippenO3A program of Rdkit. The Python scripts used for atom-based alignment are provided in the GitHub repository. (https://github.com/rdkit/rdkit/blob/master/Code/GraphMol/MolAlign/Wrap/testMolAlign.py)

Model Development

Open 3D-QSAR[16] is an open software used for model development. This open software calculates electrostatic fields by using a volume less positively (+1) charged probe of the query chemicals whereas steric fields were calculated using a sp^3 carbon probe. The steric and electrostatic energies were considered independent CoMFA variables. Comparative Molecular Field Analysis (CoMFA) is an alignment-dependent and molecular field-based method used for developing a quantitative structural activity relationship with the response of steric and electrostatic fields. A generalized Smart Region Definition (SRD) cut-off of 2 was set by removing N-level variables. This SRD was based on the closeness of variables in 3D space. The Open3DQSAR uses two different variable selection algorithms, Uninformative Variable Elimination Partial Least Square (UVE-PLS) and Fractional Factorial design-based variable SELection (FFD-SEL). The contour maps were visualized with iso-contour values at PLS coefficients of +0.005 (green) and −0.005 (yellow) for the steric field and +0.003 (blue) and −0.003 (red) for the electrostatic field.

The dataset was randomly divided into a training set (nTr = 33) and a test set (nTs=7) for 3D-QSAR model generation. The 3D-QSAR-PLS models were generated for the best fit or active compound and low active compound. The statistical values generated from the compounds was examined using the R^2/SDEC which is the coefficient R^2 and its Standardized Errors of Calibration (SDEC), F -test results, Q^2_{LOO}/SDEP (leave-one-out Q^2_{LOO}), Q^2_{LTO}/SDEP (leave-two-out Q^2_{LTO}), Q^2_{LMO}/SDEP (leave-many-out Q^2_{LMO}) with associated Standardized Errors of Prediction (SDEP) values. However, R^2_{Pred} values obtained from 3D-QSAR can be compared with a 2D-QSAR model to generate the predictivity of the model. In addition, the uniqueness of the model was justified by progressive scrambling methods, and the following criteria were followed: critical value: 0.80, type: LMO groups = 5, runs = 20, and scrambling = 20. The robustness of the model is justified by Q^2_{LMO}.

Accepted manuscript

Molecular docking analyses

The X-ray crystal structure of human AChE (PDB ID: 4EY7) [7] was retrieved from the Protein Data Bank (https://www.rcsb.org/). The protein structure was prepared by removing all water molecules and ligands. Subsequently, the hydrogen atoms were added. The partial atomic charges were assigned using the Gasteiger–Marsili method. Initially, a blind docking analysis was carried out using the Autodock Vina tool locating [17] the grid at the center of the macromolecule and extending the grid to cover the whole protein structure. After ensuring that the ligands can bind at the catalytic site of the enzyme, defined by the location-bound ligand donepezil, the Autodock 4.2 tool was employed for final docking.[18] A grid size of 40 Å × 40 Å × 40 Å with a grid-point spacing of 0.375 Å was defined at X=-11.14, Y = -45.85, Z = 23.65. The 3D structures of the input ligands were protonated at pH 7.4 and were subsequently minimized using the OpenBabel 'obminimize' tool by the steepest descent method with a maximum of 10,000 runs. A genetic algorithm-based conformational search was carried out for the semi-rigid docking setting the maximum number of evaluations as 2,500,000. Other important genetic algorithm parameters employed for docking are as follows: (a) number of runs: 10, (b) population size: 150, (c) maximum number of generations: 27,000, (d) rate of gene mutation: 0.02, (e) rate of cross-over: 0.8. Default docking parameter settings found in Autodock 4.2 were used. Analysis of the 2D ligand protein interactions was conducted using the Discovery Studio Visualizer 2017 R2.

RESULTS AND DISCUSSION

Collection of Dataset and dataset preparation:

The compounds of the dataset consisted of pyrazole derivatives[7] (compounds **43-62**) and spiropyrazolines derivatives[7] (compounds **66-85**) **(**Table I**).** The GA-MLR generates the most predictive 2D-QSAR model based on interpretable descriptors. A summary of the obtained statistical result of the model is presented in Equation 1 whereas the observed vs predicted activity plot of the model is depicted in Fig. 1.

$$pIC_{50} = -0.275(\pm 0.022) * nBM - 0.416(\pm 0.061) * C\text{-}001 + 13.255\ (\pm 1.579) * Mv + 0.232(\pm 0.061) * CATS2D\_08\_DL - 0.325(\pm 0.091) * C\text{-}014 + 2.019(\pm 0.827) \quad (1)$$

The model was based on five molecular descriptors, namely *nBM, C-001, Mv, CATS2D_08_DL*, and *C-014*. The descriptors belong to constitutional descriptors (*nBM, Mv***)**, atom-centered fragments (*C-001*, *C-014*), and pharmacophore descriptors (*CATS2D_08_DL***)**. We initially developed this GA-MLR model with the Corina (molecular mechanical) optimized structures. It is noteworthy that all these five descriptors are 2D in nature and therefore, the values of these are not dependent on the 3D structures of the compounds. Nevertheless, we calculated the alvaDesc descriptors after optimizing these structures with the semi-empirical method ULYSSES (quantum chemical PM6)[7] but GA-MLR failed to develop any better model. The relative significance of these five descriptors was estimated based on the standardized coefficients that are presented in Fig. 1. The *Mv* has the maximum relative significance on the model. The value of *Mv* or mean atomic van

Accepted manuscript

der Waals volume (scaled on carbon atom) depends on the formula provided below:

$$V_{vdW} = \sum \text{all-atom contributions} - 5.92N_B - 14.7R_A - 3.8R_{NA} \quad (2)$$

Where V_{vdW} is the mean atomic van der Waals volume, N_B is the number of bonds, R_A is the number of aromatic rings, and R_{NA} is the number of non-aromatic rings). The formula describes that the contribution of aromatic ring or R_A decreases the van der Waals volume as is observed in our QSAR results. The low active compound (compound **85**) exhibits mean atomic van der Waals volume (scaled on carbon atom), *Mv* of 0.6567 due to the presence of a large aromatic group naphthalene. However, *nBM* is decisive in distinguishing the most active and the least active substances. The presence of polyaromatic rings like naphthalene is responsible for decreased *Mv*. However, as shown in Table II, the highly active compound (compound **67**) exhibits a greater value of 0.6607 than that of compound **85**. The structural features of both the compounds are elucidated in Fig. 2. Compound **67** has an additional chlorophenyl group which is replaced by naphthalene in compound **85**. Chlorophenyl contributes less to the V_{vdW} value as compared to naphthalene. The least contributing descriptor also gets correlated to our results. Compound **67** has less *nBM* value than compound **85**. Chlorophenyl has less number of multiple bonds (*nBM* value of 21 in compound **67**) in comparison to a greater number of multiple bonds as in naphthalene (*nBM* value of 26 in compound **85**). Furthermore, two atom-centered fragments *C-001* and *C-014* (which refer to CH_3R/CH_4 and CX_4, respectively) contributed unfavorably to the biological activity. Indeed, in some lower active compounds such as **59**, **60**, **82**, **83**, and **84**, the substitution of phenyl ring with methyl or isopropyl groups was found to be detrimental to the biological activity. Similarly, the substitution of benzene with trifluoromethyloxy group in compounds like **56** and **79** was also found to have a negative influence on the biological activity. Finally, *CATS2D_08_DL* stands for hydrogen bond donor and lipophilic features located at a topological distance of 8. Two (*Mv, CATS2D_08_DL*) out of five selected descriptors influence the predicted IC_{50} value positively as illustrated in Equation 1. The *CATS2D_08_DL* is responsible for the increase in IC_{50} whereas *nBM, C-001,* and *C-014* are negatively correlated to the above equation.

Accepted manuscript

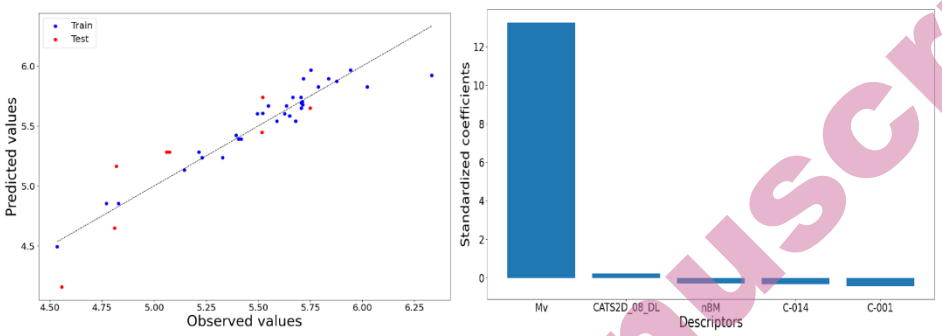


Fig. 1. Observed vs. predicted biological activity plot (left) and relative significance of the descriptors of the most predictive 2D-QSAR model.

Based on the pIC_{50} value, the highly active compound and the lowest active compound were selected for comparison with the molecular descriptor as depicted in Fig. 2. The highly active compound **67** with pIC_{50} values of 6.333 is our best lead and the values of each descriptor are compared to the lowest active compound **85** with pIC_{50} values of 4.535. Maximum contributing descriptors in both high and low active compounds give no difference as observed in Table II. Similarly, the second major descriptor also plays no role in their activity. However, the third highest contributing factor *Mv* influences the IC_{50} value.

Table II. Comparison of high and low active compounds

Compound No.	Observed pIC_{50}	Predicted pIC_{50}	Mv	nBM
67	6.333	5.923	0.6607	21
85	4.535	4.492	0.6567	26

Both the compounds have a structural resemblance with the marketed drug for Donepezil (as depicted in Fig. 2). Donepezil,[19] is an N-benzyl piperidine derivative with Indanone (Site I), piperidine (Site II) and benzyl segments (Site III). All three segments of Donepezil interact with AChE in such an orientation that finds a pattern for all AChE interacting agents.[20]

Accepted manuscript

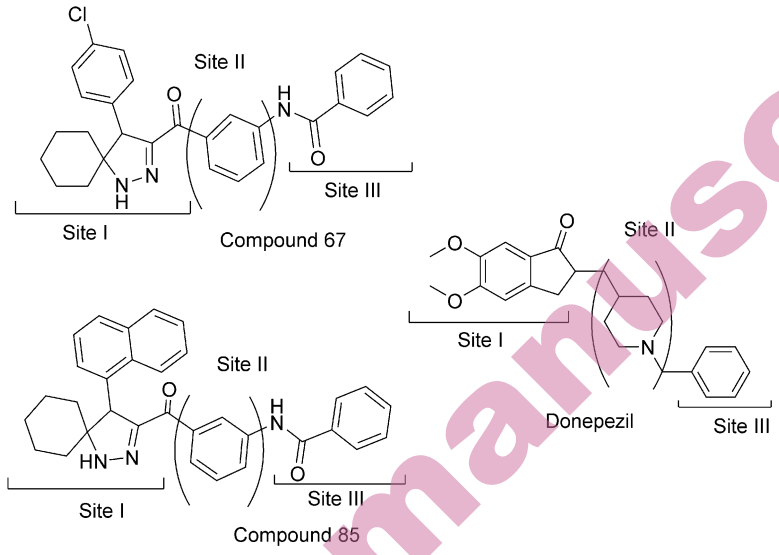


Fig. 2. Structural correlation of three sites of compound **67** and compound **85** with Donepezil

Model validation procedures: Leave-one-out method:

Overall statistical quality of each model was justified by internal validation parameter Q^2 and external validation parameter R^2_{Pred}. The best model was generated with a random division-based training set and test set using seed value 10.

The GA-MLR generated the generated most predictive model based on interpretable descriptors. The best model was generated with 32 training sets and 8 test set compounds. The results obtained for the training set (Table III) and test set (Table IV) are presented below. A good QSAR model is represented by a minimal coefficient of determination (R^2), and an adjusted R^2 (R^2_A) of nearly one (obtained values are $R^2 = 0.906$, $R^2_A = 0.888$). The MAE (mean absolute error) was used to generate the goodness fit of the model (obtained values of (obtained values of MAE are 0.074). Thus the QSAR model can be said to be a good fit for the model. Internal cross-validation coefficient Q^2_{LOO} was obtained when all the descriptors were used for model development and to check the robustness and internal predictivity. The Q^2_{LOO} of 0.876 exhibits that the model is not over-fitted for the training set. A high Q^2 value that is $Q^2 > 0.5$ is considered proof of the high predictive ability of the model (here, $Q^2 = 0.876$, $r_m^2 = 0.825$). 0.825). Furthermore, values of r_m^2 are 0.825 and $r_m^2{}_{LOO}$ is 0.042 are internal validation parameters and suggests that the training set is validated.

Table III. Training set: Statistical results

No.	R^2	R^2_{adj}	Q^2_{LOO}	MAE	MSE	$R_m{}^2{}_{Train}$	$\Delta R_m{}^2{}_{Train}$
32	0.906	0.888	0.876	0.074	0.11	0.825	0.042

Accepted manuscript

Maximum inter-correlation (Pearson r) between two descriptors: 0.766.

Table IV: Test set: Statistical results

No.	R^2_{Pred}/Q^2_{F1}	Q^2_{F2}	RMSEP	$R_m{}^2_{Test}$	$\Delta R_m{}^2_{Test}$
8	0.821	0.628	0.24	0.668	0.161

The maximum inter-correlation (Pearson r) between two descriptors is 0.766 which suggests that the descriptors of the regression model are devoid of high inter-co linearity and each descriptor of the model is thus unique.

External Validation: Calculation of R^2_{pred}

The R^2pred was calculated by fitting the test set descriptors into the developed 2D-QSAR model equation and thereafter comparing the predicted bioactivity of the compounds with their observed bioactivity. As far as the external validation is concerned, a satisfactory R^2_{Pred} of 0.821. The value of $R^2_{pred} > 0.6$ indicates a good external predictability of a model. The GA-MLR model yields a highly predictive QSAR model with a Q^2_{LOO} value greater than 0.5 and an R^2_{pred} value greater than 0.6. The satisfactory internal and external predictivity of the model was demonstrated. Furthermore, the maximum inter-correlation (Pearson r) between two descriptors is 0.766 which suggests that the descriptors of each model are independent of each other.

The applicability domain of the model

The applicability domain of the model was found by availing the Williams plot (generated with one of our tools named SFS-QSAR-tools_v2, https://github.com/ncordeirfcup/SFS-QSAR-tool_v2) which is a plot drawn between standardized residuals in the y-axis and the leverage values in the x-axis. If the leverage value of any compound is more than the hat value (h*) which is calculated by the formula: $h^* = 3p'/n$ (where, p' is several model descriptors +1, whereas n is several data in the training set), then the compound is assumed to be a structural outlier. However, if the standardized residuals are greater than ± 3.0 then it is considered a response outlier.

It is known that if the leverage value of any compound is more than the hat value (h*), here 0.55 then the compound is assumed to be a structural outlier. Similarly, if the standardized residuals (here greater than ± 3) then it is considered a response outlier. Two structural outliers (one training set and one test set) and one response outlier were found (Fig. 3). The relative significance of the descriptors could be described as most of the data remains within the standardized residuals.

Accepted manuscript

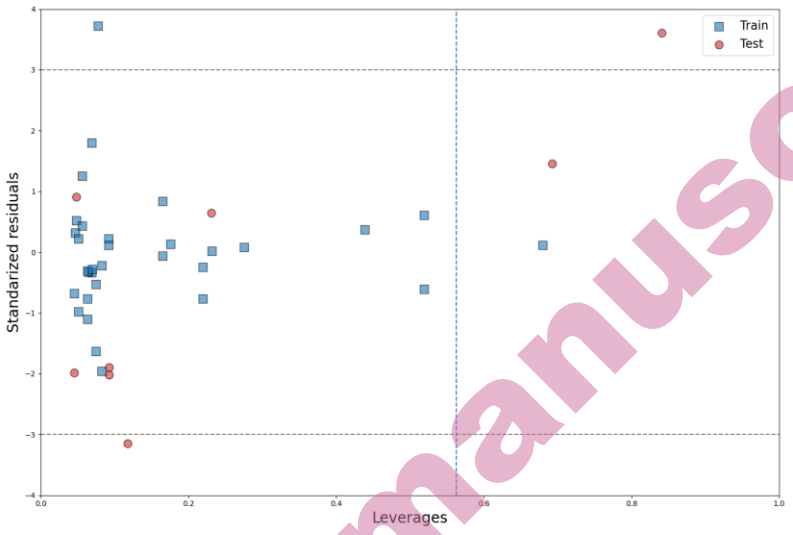


Fig. 3. Applicability domain of the 2D-QSAR model

3D QSAR: Alignment Method

The Open3DQSAR model was generated by two methods namely, Fractional Factorial Design based variable SELection (FFD-SEL) and Uninformative Variable Elimination based Partial Least Square (UVE-PLS). The statistical results of both the FFD-SEL and UVE-PLS models are tabulated in Table V**.** The results of 3D QSAR analysis infer that the UVE-PLS model as compared to FFD-SEL is more helpful in drawing a successful conclusion. The UVE-PLS model exhibits a Q^2_{LOO} (leave-one-out Q^2_{LOO}) of 0.693 and an R^2_{Pred} value of 0.692. The uniqueness of the model was suggested by Q^2_{LMO} (leave-many-out Q^2_{LMO}) values which happen to be 0.662. Thus, it can be concluded that the current 3D-QSAR model can explain the structural requirements of these dataset compounds.

Table V. 3D-QSAR Statistical results in FFD-SEL and UVE-PLS

Parameter	FFD-SEL	UVE-PLS
$N_{training}$	33	33
NC	4	4
R^2/SDEC	0.898/0.134	0.920/0.119
F	61.41	80.23
Q^2_{LOO}/SDEP	0.718/0.222	0.693/0.232
Q^2_{LTO}/SDEP	0.711/0.225	0.687/0.234
Q^2_{LMO}/SDEP	0.679/0.237	0.662/0.243
N_{test}	7	7
R^2_{Pred}	0.658/0.184	0.692/0.175

Accepted manuscript

The above results were also confirmed from alignment-dependent molecular field analysis which correlates with the results obtained from 2D-QSAR analysis.[20] The contour maps were visualized with iso-contour values at steric (green and yellow regions) and electrostatic field (blue and red regions) at PLS coefficients as depicted in Fig. 4. The electrostatic and steric fields of the two compounds (best active compound **67** and least active compound **85**) were elucidated in Fig. 4. The steric contribution was 43% whereas the electrostatic contribution was 57%. Steric regions of the best active compound **67** contain a bulky chloro benzene moiety inserted into the steric favorable field (Fig. 4A). The bulky naphthalene group is found close to the steric unfavorable field which also confirms the results obtained from that of 2D QSAR analysis (Fig. 4C). The van der Waals volume of compound **67** was 0.661 due to the presence of less number of aromatic groups as compared to compound **85** which consists of bulky poly aromatic naphthalene group. The steric effects of compound **67** mainly contribute to the model development. The steric favorable group was found near the positive field which happens to lead to greater biological activity. Fig. 4A and Fig. 4C depict that the side chain variation in the steric fields leads to a change in biological activity. The electrostatic contribution mainly contributes to the model development as compared to steric field contributions. The Electrostatic contribution of the best active compound **67** elucidates that electron-deficient moiety is inserted into the electropositive field (Fig. 4B). The electron-rich naphthalene moiety is inserted in the electropositive field thus explaining the reason for the decreased biological activity of compound **85** (Fig. 4D). The naphthalene ring being highly electron-rich gets embedded into the electropositive field. The electron-rich moieties favor lower potency.

To gain a more complete scientific significance, a molecular docking study should be performed on the AChE with two analyzed compounds (**67** and **85**). AChE was selected and downloaded from Protein Data Bank (www.rcsb.org, PDB ID: 4EY7). The two compounds were docked using AutoDock v4.2 (The Scripps Research Institute, La Jolla, California). The docking poses are shown in Fig. 5.

Accepted manuscript

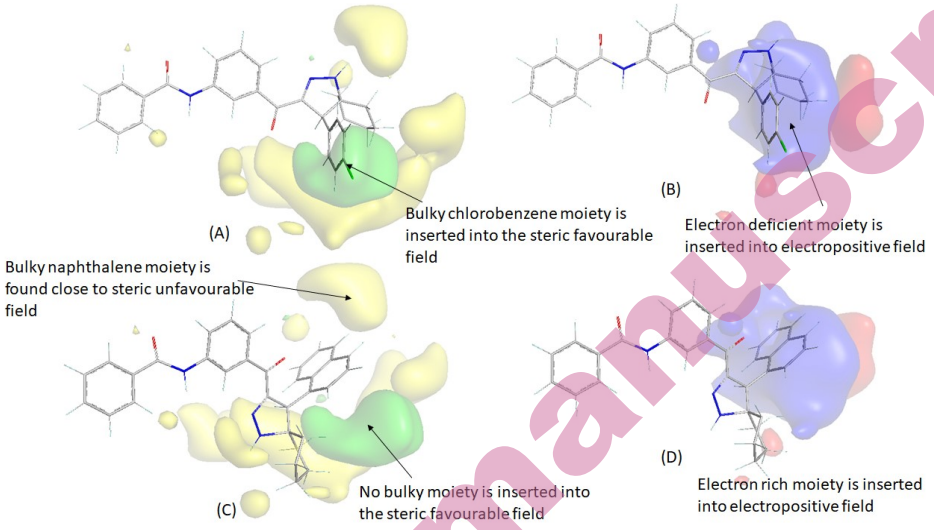


Fig. 4. Contour maps visualized with PLS coefficients (A) Steric regions of best active compound **67**, (B) Electrostatic region of the best active compound **67**, (C) Steric regions of least active compound **85** and (D) Electrostatic region of the least active compound **85** (Green: Steric favorable, Yellow: Steric unfavorable Blue: Electropositive favorable; Red: Electronegative favorable).

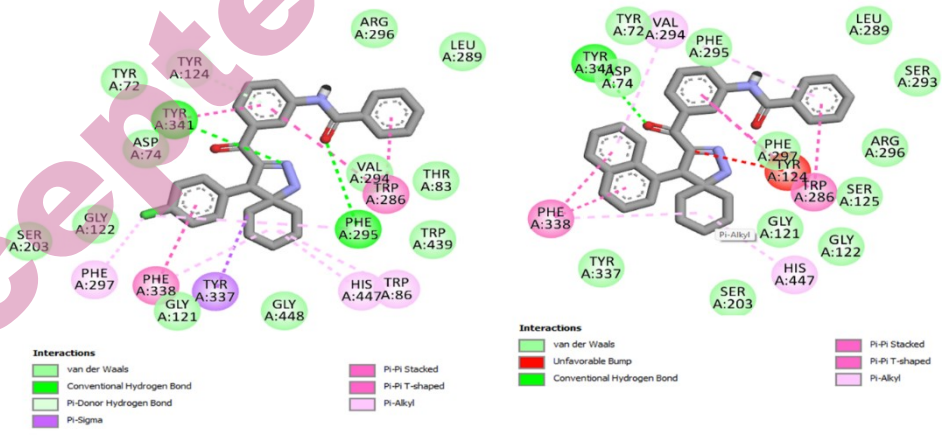


Fig. 5. The docking poses of **67** (left) and **85** (right)

First of all both **67** and **85** were docked at the same binding site though there is a large difference between their binding affinities. The best poses of compounds **67** and **85** were found to have binding affinities of -5.64 and -0.62 kcal/mol, respectively. Both 2D- and 3D-QSAR analyses highlighted the importance of the chlorobenzene residue of **67** for higher activity as compared to the naphthalene moiety of **85**. The 3D-QSAR highlighted that unfavorable steric and electrostatic

Accepted manuscript

interactions with the naphthalene moiety of **85** are responsible for significantly low activity of **85**. Now, from docking analyses, it is evident that similar to the chlorobenzene residue of **67**, the naphthalene residue forms a pi-pi interaction with Phe338. However, the former is more effective as its chlorine is associated with pi-alkyl interactions with Phe295 and Phe297. Noticeably, the possibility of such interactions was highlighted in 3D-QSAR analyses. However, more interestingly, the docking analyses revealed that the pi-pi interactions between the naphthalene residue of **85** and Phe338 caused it to have an unfavorable interaction clash between **85** and Tyr124. Therefore, it may be inferred that the interpretations obtained from our 2D-QSAR and 3D-QSAR analyses are consistent with the docking results.

CONCLUSION

The current work investigates the structural requirements of a series of pyrazole and spiropyrazoline analogs for higher AChE inhibitory potential. We performed 2D-QSAR and 3D-QSAR analyses in a systematic manner to find out the most crucial structural attributes. The 2D-QSAR model was constructed with selected interpretable alvaDesc descriptors and the most predictive and evaluative GA-MLR model was developed with satisfactory statistical predictivity. The 2D-QSAR model highlighted that molecular volume and number of multiple bonds along with presence/absence of specific atom centered fragments and topological distance between 2D pharmacophoric features determine the potency of these compounds against this enzyme. The 3D-QSAR model, on the other hand, depicted the importance of specific electrostatic and steric contours regions which favorably or unfavorably influence the activity of the compounds. We compared the most potent (**67**, N-(3-(4-(4-chlorophenyl)-1,2-diazaspiro[4.5]dec-2-ene-3-carbonyl)phenyl)benzamide) and the least potent (**85**) compounds of the dataset. In spite of having high structural similarity, their biological activities varied to a considerable extent. The 3D-QSAR model was able to explain this phenomenon by showing the importance of chlorophenyl moiety of **67** for higher activity as compared to naphthyl moiety of **85**. More importantly, the inference was largely supported by the molecular docking methodology performed with these. The current work provides valuable information for future design of pyrazole and spiropyrazoline analogs as potent AChE inhibitors.

Acknowledgment: The Authors are thankful to Dr. B.C. Roy College of Pharmacy & Allied Health Sciences, Durgapur, WB, and India for providing all the support for the execution of the work. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Accepted manuscript

И З В О Д

РАЗВОЈ 2D И 3D QSAR МОДЕЛА ПИРАЗОЛСКИХ ДЕРИВАТА КАО ИНХИБИТОРА АЦЕТИЛХОЛИН ЕСТЕРАЗЕ

PUJA MISHRA[1], SUMIT NANDI[1], ANKIT CHATTERJEE[1], TRIDIB NAYEK[1], SOUVIK BASAK[1], AMIT KUMAR HALDER[1,2], ARUP MUKHERJEE[3]

[1]Dr. B.C. Roy College of Pharmacy & Allied Health Sciences, Durgapur, WB, India, [2]LAQV@REQUIMTE/Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, Porto, Portugal, [3]Department of Biotechnology, Maulana Abul Kalam Azad University of Technology, WB, India

Инхибитори ацетилхолинестеразе (AChE) су лекови који су најкориснији за лечење Алцхајмерове болести (AD) у свим стадијумима. Циљеви овог рада су генерисање различитих QSAR модела и одабир робусних предиктивних модела. Фокус даљег истраживања је на проналажењу низа AChE инхибитора сличних пиразолу помоћу 2D и 3D QSAR анализе. Вишеструка линеарна регресија заснована на генетском алгоритму (GA-MLR) обезбедила је статистички робустан 2D QSAR модел који је приказао значај запремине молекула волумена и броја вишеструких веза заједно са присуством/одсуством специфичних фрагмената центрираних на атому и тополошке удаљености између 2D фармакофора. Штавише, ови резултати су били у доброј корелацији са електростатичким и стерним мапама контура преузетим из 3D QSAR (тј. анализа молекуларног поља зависна од поравнања). 2D QSAR анализа развила је високо статистички и поуздан модел који је упоређен са механичком интерпретацијом 3D структура и њиховим доприносима електростатичком и стеричном пољу што је довело до предиктивног 3D QSAR модела. Интеракције молекул-протеин изазване молекуларним спајањем поткрепиле су интеракције поља које је открио 2D QSAR. Дакле, развијени рачунарски модели и симулационе анализе у овом раду дају драгоцене информације за будући дизајн аналога пиразола и спиропиразолина као моћних инхибитора AChE.

(Примљено 21. фебруара 2023; ревидирано 2. августа 2023; прихваћено 31. март 2024.)

REFERENCES

1. M. Hernández-Rodríguez, J. Correa-Basurto, F. Martínez-Ramos, I. I. Padilla-Martínez, C. G. Benítez-Cardoza, E. Mera-Jiménez, M. C. Rosales-Hernández, *J. Alzheimer's Dis.* **41** (2014) 1073 (https://doi.org/10.3233/JAD-140471)
2. S. Habtemariam, *Molecules* **24** (2019) 1519 (https://doi.org/10.3390/molecules24081519)
3. Y. Yamaguchi, H. Miyashita, H. Tsunekawa, A. Mouri, H. C. Kim, K. Saito, T. Matsuno, S. Kawashima, T. Nabeshima, *J. Pharmacol. Exp. Ther.* **317** (2006) 1079 (https://doi.org/10.1124/jpet.105.098640)
4. P. Mishra, S. Basak, A. Mukherjee, A. Basu, *Lett. Drug Des. Discov.* **19** (2021) 192 (https://doi.org/10.2174/1570180818666210813120444)
5. L. Lecanu, V. Papadopoulos, *Recent Pat. CNS Drug Discov.* **2** (2007) 113 (https://doi.org/10.2174/157488907780832715)
6. R. Chawla, A. Arora, M. K. Parameswaran, P. Chan, D. Sharma, S. Michael, & T. K. Ravi, *Acta Pol. Pharm.* **67** (2010) 247 (https://www.ptfarm.pl/pub/File/acta_pol_2010/3_2010/247-253.pdf)

Accepted manuscript

7. G. Gutti, D. Kumar, P. Paliwal, A. Ganeshpurkar, K. Lahre, A. Kumar, S. Krishnamurthy, S. K. Singh, *Bioorg. Chem.* **90** (2019) 103080 (https://doi.org/10.1016/j.bioorg.2019.103080)
8. *Ecotoxicological QSARs*, Ed. K. Roy Editor, Springer 2020, ISBN 978-1071601495,
9. I. Sushko, A. K. Pandey, S. Novotarskyi, R. Körner, M. Rupp, W. Teetz, S. Brandmaier, A. Abdelaziz, V. V. Prokopenko, V. Y. Tanchuk, R. Todeschini, A. Varnek, G. Marcou, P. Ertl, V. Potemkin, M. Grishina, J. Gasteiger, I. I. Baskin, V. A. Palyulin, E. V. Radchenko, W. J. Welsh, V. Kholodovych, D. Chekmarev, A. Cherkasov, J. Aires-De-Sousa, Q. Y. Zhang, A. Bender, F. Nigsch, L. Patiny, A. Williams, V. Tkachenko, I. V. Tetko, *J. Cheminform.* **3** (Suppl 1) 2011 P20 (https://doi.org/10.1186/1758-2946-3-S1-P20)
10. P. Ambure, R. B. Aher, A. Gajewicz, T. Puzyn, K. Roy, *Chemom. Intell. Lab. Syst.* **147** (2015) 1 (https://doi.org/10.1016/j.chemolab.2015.07.007)
11. P. Tosco, T. Balle, *J. Mol. Model.* **17** (2011) 201 (https://doi.org/10.1007/s00894-010-0684-x)
12. J. Sadowski, J. Gasteiger, G. Klebe, *Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures*, 1994
13. A. K. Halder, M. N. D. S. Cordeiro, *Biomolecules* **11** (2021) 1670 (https://doi.org/10.3390/biom11111670)
14. G. Kryger, I. Silman, J. L. Sussman, *Structure* **7** (1999) 297–307 (https://doi.org/10.1016/S0969-2126(99)80040-9)
15. H. Safarizadeh, Z. Garkani-Nejad, *J. Mol. Graph. Model.* **87** (2019) 129 (https://doi.org/10.1016/j.jmgm.2018.11.019)
16. T. K. Shameera Ahamed, V. K. Rajan, K. Muraleedharan, *Food Sci. Hum. Wellness* **8** (2019) 53–62 (https://doi.org/10.1016/j.fshw.2019.02.001)
17. O. Trott, A. J. Olson, *J. Comput. Chem.* **31** (2009) 455 (https://doi.org/10.1002/jcc.21334)
18. G. M. Morris, H. Ruth, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *J. Comput. Chem.* **30** (2009) 2785 (https://doi.org/10.1002/jcc.21256)
19. H. Sugimoto, *Pure Appl. Chem.* **71** (2009) 2031 (https://doi.org/10.1351/pac199971112031)
20. P. Mishra, P. Sharma, P. N. Tripathi, S. K. Gupta, P. Srivastava, A. Seth, A. Tripathi, S. Krishnamurthy, S. K. Shrivastava, *Bioorg. Chem.* **89** (2019) 103025 (https://doi.org/10.1016/j.bioorg.2019.103025)

Accepted manuscript