QSAR, Docking and Molecular Dynamics Studies on the Piperidone-grafted Mono- and Bis-spiro-oxindole-hexahydropyrrolizines as Potent Butyrylcholinesterase Inhibitors

M. Amiri[a], M. Fazli[a] and D. Ajloo[b],*

[a]Department of Chemistry, College of Science, Semnan University, Semnan,35131-19111, I.R. Iran
[b]School of Chemistry, Damghan University, Damghan, 36716-41167, I.R. Iran
(Received 15 March 2018, Accepted 28 June 2018)

Quantitative structure-activity relationship (QSAR) study on the piperidone-grafted mono- and bis-spirooxindole-hexahydropyrrolizines as the potent butyrylcholinesterase (BuChE) inhibitors was carried out using statistical methods, molecular dynamics and molecular docking simulation. QSAR methodologies include classification and regression tree (CART), multiple linear regression (MLR), principal component analysis (PCA) and principal component regression analysis (PCRA). Three descriptors in three classes: 3D-Morse, WHIM and GETAWAY descriptors were selected by SPSS software, and then applied in the final tree structure to describe the inhibitory activities. Docking simulations were carried out using AutoDock Vina software for all inhibitors. Docking results showed that the studied BuChE inhibitors have two commons binding modes. Molecular dynamics results obtained by Gromacs showed that the more potent inhibitor has the stronger interaction with the enzyme and higher effect on the enzyme structure.

Keywords: Butyrylcholinesterase inhibitors, Alzheimer, QSAR, Docking, Molecular dynamics

INTRODUCTION

Acetyl and butyrylcholinesterase (AChE and BuChE) are two cholinesterase enzymes in the human brain [1]. The BuChE was discovered about 80 years ago [2]. The soluble form of this enzyme exists in plasma and brain that have a wide range of substrates. The pouch is attached to cell membranes of the brain in a hydrophobic form [3]. BuChE is important in aspects of toxicological and pharmacological, due to its hydrolyzes ester-containing drugs and clean up cholinesterase inhibitors, including potent organophosphorus nerve agents before they act on their synaptic targets [4]. The studies on BuChE in recent years has been growing due to its possible function in Alzheimer's disease and the presentation of anticholinesterase therapy for this disorder. In Alzheimer disease, a cholinergic deficiency and neurological disorder

within the brain have been reported [5-6]. Degeneration of the cholinergic neurons and the loss of cholinergic transmission represent consistent features of Alzheimer's disease. The reduction in choline acetyltransferase causes a decrease in acetylcholine and acetylcholinesterase activity, which commonly leads to an increase in BChE activity [7]. Limited acetylcholine levels that have been protected with potent cholinesterase inhibitor therapeutics act on AChE and BChE as well. Favorite BChE inhibitors prevent the generation of new beta-amyloid plaques that cleave amyloid precursor protein to beta-amyloid protein created by BChE [8]. Computational methods have been developed into effective tools in facilitating and simplifying new drug discovery [9-11]. Using computational techniques, the biological activity of the candidate compounds can be evaluated before experimental trials. These simple and non-expensive methods accelerate the design of compounds with favorable biological activity [12-13]. Two computational methods mostly used in drug design are quantitative

*Corresponding author. E-mail: ajloo@du.ac.ir

structure-activity relationship (QSAR) [14-18] and docking procedure [19-22]. In QSAR methods, a mathematical equation is established between the molecular properties and the biological potency of compounds they elicit when applied to a biological system. Ongoing advances in computing technology has enabled us to estimate the molecular properties of the compounds without the need to synthesize them. Thus, the use of predictive computational (in silico) QSAR models allows the biological properties of virtual structures to be predicted, and a more informed choice of target to be selected for synthesis. QSAR is an appropriate alternative tool to the trail-error based experimental search for better activities in drug discovery. To well perform the QSAR studies of such a complex data set, classification and regression trees (CART) have been considered. This is attributed to the fact that CART holds promising modeling performance and many attractive features, including simplicity, interpretability, high capacity in handling large data sets and modeling nonlinearities, no assumption regarding the data distribution and immunity to outliers, collinearity and heteroscedasticity. In addition, CART is provided with the ability to automatically select the most descriptive variables from a large number of descriptors under investigation. These are the reasons that CART greatly increases in its popularity in QSAR studies. CART analysis is a statistical method explaining the variation of a response variable using a set of explanatory variables, so-called predictors. The method is based on a recursive binary splitting of the data into mutually exclusive subgroups containing objects with similar properties [23]. Modeling and classification by CART were used in medical diagnosis and prognosis [24-26], ecology [27], agriculture [28], and chemistry [29-31]. CART provides a graphical representation that makes the interpretation of the results easier. Accordingly, CART could be an appropriate technique to correlate specific molecular descriptors with targeted properties of the molecule. The CART analysis includes three steps: (I) maximal-tree building, (II) tree ''pruning'' containing the cutting-off of nodes to produce a sequence of simpler trees, and (III) the optimal tree selection that minimizes cross-validation error [23-31]. Classification and regression tree (CART) method can be used for QSAR studies in comparison to multiple linear regression (MLR). Multiple linear regression (MLR) is used to determine a mathematical relationship among a number of random variables. In other words, MLR examines how multiple independent variables are related to one dependent variable. Once each of the independent factors were determined to predict the dependent variable, the information on the multiple variables could be used to create an accurate prediction on the level of effect they have on the outcome variable. The model creates a relationship in the form of a straight line (linear) that approximates all the individual data points. In docking studies, different search algorithms such as genetic algorithm and simulated annealing in composition with a scoring function such as molecular mechanic calculations are used to study the binding of the candidate compounds (ligands) to a protein with known structure. *Via* docking procedures, not only new biologically active ligands are recognized, but also the chemistry of the interactions between ligand and protein is well recognized. The outcomes from this study should be beneficial in modeling new inhibitors for alzheimer's disease.

MATERIAL AND METHODS

Data Set

Experimental data comprising 33 inhibitory activity values (inhibition IC_{50}) for BuChE, were taken from the literature [32] and are converted to $log(IC_{50})$. The structures and experimental data on the compounds used for the study are shown in Table 1.
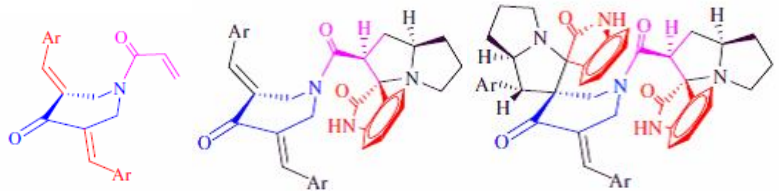
Softwares

Geometry optimization was performed by HyperChem (Version 7.0; Hypercube, Inc, Alberta, Canada). Dragon 2.1 (Milano Chemometrics and QSAR Research Group, Milano, Italy) software was used to calculate the molecular descriptors. The SPSS software (version 16.0) was employed for the CART, MLR and PCA methods. AutoDock Vina software was used to perform automated docking of the ligands to macromolecular protein receptor.

Calculation of Molecular Descriptors

Structures of the molecules were drawn in the HyperChem software. The optimization of the molecular structures was carried out by the semi-empirical Austin

Table 1. Experimental and Predicted Values of log(IC$_{50}$) for the 33 Mono and Bis-spiropyrrolizine Derivatives Based on Subset 3

Compound	R	The amounts of selected descriptors in CART method			log(IC$_{50}$)			
		R3u	Km	Mor15m	Expt.	Calc.$_A$	Calc.$_B$	Calc.$_C$
5a[c]	C$_6$H$_5$	1.78	0.37	-0.27	1.34	1.26	1.30	1.44
5b[c]	2-CH$_3$C$_6$H$_4$	1.69	0.54	-0.35	1.48	1.57	1.46	1.49
5c[p]	2-(OCH$_3$)C$_6$H$_4$	1.42	0.61	0.76	1.46	1.57	1.43	1.58
5d[c]	2-ClC$_6$H$_4$	1.75	0.55	-0.44	1.56	1.57	1.60	1.49
5e[c]	2-FC$_6$H$_4$	1.76	0.55	-0.23	1.73	1.57	1.59	1.54
5f[e]	3-(O$_2$N)C$_6$H$_4$	1.70	0.44	0.13	1.71	1.57	1.89	1.63
5g[p]	2,4-Cl$_2$C$_6$H$_3$	1.69	0.61	-0.18	1.47	1.57	1.65	1.59
5h[c]	4-CH$_3$C$_6$H$_4$	1.66	0.58	-0.55	1.54	1.57	1.59	1.49
5i[c]	4-ClC$_6$H$_4$	1.71	0.64	-0.50	1.58	1.57	1.67	1.56
5j[c]	4-FC$_6$H$_4$	1.73	0.60	-0.30	1.49	1.57	1.34	1.60
5k[p]	1-Napthyl	1.84	0.62	0.08	1.36	1.57	1.17	1.39
8a[c]	C$_6$H$_5$	2.30	0.31	0.74	1.21	1.26	1.13	1.07
8b[c]	2-CH$_3$C$_6$H$_4$	2.22	0.41	0.03	1.25	1.26	1.08	1.07
8c[c]	2-(OCH$_3$)C$_6$H$_4$	2.06	0.41	0.59	0.97	1.26	0.91	1.13
8d[p]	2-ClC$_6$H$_4$	2.16	0.37	0.13	1.30	1.26	1.12	1.13
8e[c]	2-FC$_6$H$_4$	2.24	0.41	0.15	0.54	0.65	0.81	1.15
8f[e]	3-(O$_2$N)C$_6$H$_4$	2.16	0.42	0.04	1.26	1.26	1.24	1.16
8g[c]	2,4-Cl$_2$C$_6$H$_3$	2.19	0.43	0.07	1.33	1.26	1.35	1.11
8h[p]	4-CH$_3$C$_6$H$_4$	2.15	0.42	-0.02	1.40	1.26	1.20	1.05
8i[c]	4-ClC$_6$H$_4$	2.21	0.43	0.01	1.51	1.57	1.32	1.11
8jc	4-FC$_6$H$_4$	2.23	0.42	0.16	0.60	0.65	0.77	1.15

Table 1. Continued

8k[c]	1-Naphthyl	2.30	0.41	1.12	1.07	1.06	0.97	0.98
9a[p]	C_6H_5	2.30	0.31	0.63	1.26	1.06	1.22	1.07
9b[c]	$2\text{-}CH_3C_6H_4$	2.28	0.35	0.79	1.11	1.06	1.15	1.04
9c[c]	$2\text{-}(OCH_3)C_6H_4$	2.21	0.28	0.52	1.28	1.26	1.29	1.09
9d[c]	$2\text{-}ClC_6H_4$	2.32	0.39	0.80	1.07	1.06	1.09	1.06
9e[p]	$2\text{-}FC_6H_4$	2.33	0.32	-0.32	0.96	0.65	0.85	1.07
9f[c]	$3\text{-}(O_2N)C_6H_4$	2.23	0.35	1.09	1.39	1.26	1.20	1.22
9g[c]	$2,4\text{-}Cl_2C_6H_3$	2.30	0.44	0.50	0.98	1.06	0.98	1.08
9h[c]	$4\text{-}CH_3C_6H_4$	2.21	0.31	0.83	1.25	1.26	1.33	1.05
9i[p]	$4\text{-}ClC_6H_4$	2.28	0.33	0.81	1.04	1.06	1.23	1.08
9j[c]	$4\text{-}FC_6H_4$	2.29	0.31	0.90	0.92	1.06	0.80	1.11
9k[c]	1-Naphthyl	2.37	0.32	0.30	0.81	0.65	0.99	1.14

Subscripts A, B and C: Calculated value by CART, MLR and PCR methods. Superscripts c and p shows compounds in the calibration and prediction sets.

Model 1 (AM1) method using the Polak-Ribière algorithm until the root-mean-square gradient obtained was 0.01 kcal mol^{-1}. DRAGON software was used to calculate 1497 molecular descriptors belonging to 18 different types of theoretical descriptors for each molecule. After elimination of descriptors with zero value and descriptors with the same value for all molecules, the remaining descriptors (1,118 descriptors) were used.

Classification and Regression Tree

CART, as a binary tree representation, can describe the relationships between the dependent and independent variables with high flexibility and sufficient accuracy. The dependent variable can be either numerical or categorical, respectively, resulting in regression or classification trees. Here, since CART was used for regression tasks, only a concise description of the regression trees is presented. Generally, the configuration of CART consists of three basic steps.

Firstly, the largest tree is grown by applying greedy recursive partitioning. Recursive partitioning is conducted in a top-down fashion, starting from the root node containing the entire training compounds until each node reaches completely homogeneity or a user-specified minimal sample number (*i.e.*, node size) and becomes a terminal or leaf node. Based on the minimal cost-complexity pruning (MCCP) criterion, the largest tree is pruned to yield a sequence of nested subtrees. Ultimately, from such nested subtrees, the latest appropriately-fit CART is selected in terms of its best prediction accuracy, either gained by cross-validation method or pruning set technique. Once the latest appropriately-fit tree is gained, some immanent node information is endowed. Each splittable node is characterized by a splitting rule, including the splitting variable and value. Each node is assigned to the mean bioactivity of the involved compounds as the node output. In addition, the node size, *i.e.,* the number of compounds in the node, is provided in each node. A prediction of the bioactivity of an unseen compound from a given set of descriptors is made by traversing the tree until a leaf node is reached, and this leaf node output acts as the predicted bioactivity.

Molecular Docking

Different BuChE co-crystal structures are available in the protein data bank (www.rcsb.org/pdb). The PDB file of the crystal structure of BuChE downloaded from the PDB bank server (PDB code: 2WIJ) was used in the docking simulations. Automated docking of the ligands is performed by the AutoDock program (www.scripps.edu/pub/olson-web/doc/AutoDock). AutoDock Vina was used to perform automated docking of ligands to their macromolecular protein receptor and study the binding mode of mono and bis-spiropyrrolizine derivatives in the BuChE receptor active site. The docking model, at the molecular level, helped understand the biological activity reported for mono and bis-spiropyrrolizine molecules.

In short, AutoDock performs an automated docking of the ligand with user-specified dihedral flexibility within a protein rigid binding site. The program performs several runs in each docking experiment. Each run provides one predicted binding mode.

To start, all water molecules and ligand, when present, were removed from the original protein data bank files. Polar hydrogen atoms were added, and Kollman charges, atomic solvation parameters, and fragment volumes were assigned to the protein. For all ligands, Gasteiger charges were assigned and non-polar hydrogen atoms merged. All torsions were allowed to rotate during docking. A grid with a spacing distance of 0.375 Å and $80 \times 80 \times 80$ points were created. Energy grid maps for all possible ligand atom types were calculated before docking. For all ligands, fortuitous starting positions, random orientations, and torsions were used. The translation, quaternion, and torsion steps were taken from default values in AutoDock. The Lamarckian genetic algorithm was applied for minimization using default parameters. The number of docking runs was 100. The population in the genetic algorithm was 150, the energy evaluations were 250,000 and the maximum number of iterations 27,000. After docking, 100 solutions were clustered into groups with RMS deviations lower than 1.0 Å. The clusters were ranked by the lowest energy representative of each cluster.

MD Simulation

The structures of two chosen compounds were drawn using Hyperchem7 software and Gaussian 2003 [33]. Force field parameters and geometries of the ligands were generated using PRODRG2 server [34], and they were changed to be adaptable to the preferred force field. The MD simulations were performed by the GROMOS 43a1. Force field and TIPP4 model [35] were used for water molecules. A twin range cutoff was used for long-range interactions: 0.9 nm for van der Waals interactions and 0.9 nm for electrostatic interactions. The PME [36] was used for calculating long-range interaction. The starting structure of BuChE was constructed based upon the X-ray crystal structure of it (PDB ID: 2WIJ). A cubic simulation box of the volume 582 nm^3 was made, and four molecules of an inhibitor were placed randomly in this box, respectively. Then, water molecules were added to the simulation box, and initial configurations were minimized using the steepest descent algorithm with 5000 integration step, and the system was equilibrated for 20 ns at constant pressure (1 atm) and temperature (300 K) using the Parilleno-Rahman procedure. All MD simulations were carried out using the GROMACS 4.5.4 package [37].

The calculations were performed using 5quad core parallel computers, including 40 processor units. The computer applied the Rocks cluster networking and CentOS operating systems. Four molecules of each inhibitor were inserted into a box with size 7.865, 7.764, 9.543 nm^3. Then 16867, 16791 and 16747 water molecules were inserted into that box separately containing enzyme (2wij), enzyme+5e, and enzyme+8e, respectively.

Data Analysis

The conformational changes of the system during MD simulations were monitored by the root-mean-square derivations (RMSD) as follows;

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} m_i (r_i - r_i^o)^2}{\sum_{i=1}^{N} m_i}} \qquad (1)$$

where m_i is the mass of atoms, r_i^o and r_i are the coordinates of i atom at the reference state and during MD simulations, respectively. RMSDs were calculated for the trajectories from the starting structures of the system as a function of time. In all systems, RMSDs reach a stable value within the first nanosecond of all the analyses.

The radial distribution function $g(r)$ is the density probability for finding a particle at distance r from the
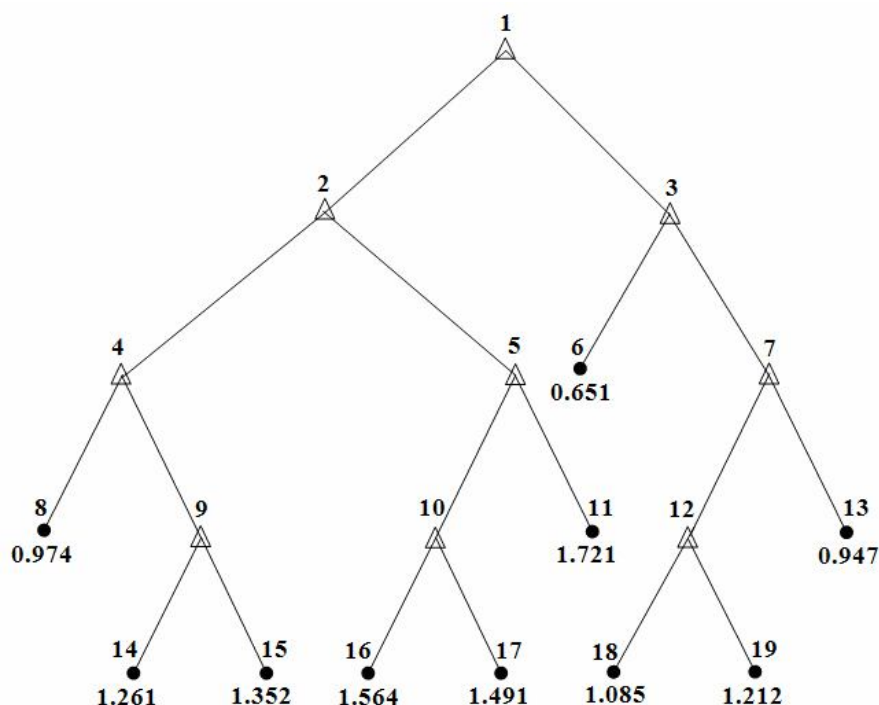
Fig. 1. Maximal regression tree, grown for the $\log(IC_{50})$ values of 25 drugs using 1118 descriptors, where "Δ" represents the internal node and "$\bullet$" refers to the leaf node.

reference particle. The $g(r)$ between particles of type A and B is defined as follows:

$$g(r) = \frac{\langle \rho_{B(r)} \rangle}{\langle \rho_B \rangle_{local}} = \frac{1}{\langle \rho_B \rangle_{local}} \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta(r_{ij} - r)}{4\pi r^2}$$

(2)

where $\langle \rho_{B(r)} \rangle$ is the particle density of type B at a distance r around particles A, and $\langle \rho_B \rangle_{local}$ is the particle density of type B averaged over-all spheres around particles A with radius r_{max}. Solvent accessible surface area for protein was also computed. Accessible surface area, helix percentage, hydrogen bond, angle information and other complementary information were obtained by VADAR [38].

RESULTS AND DISCUSSION

QSAR

To make sure about the absence of a chance correlation, the whole data set was divided into four subsets, and each subset was predicted using other three subsets as the

training set. In CART method, a maximal tree was grown using the binding affinities of 33 imidazobenzodiazepines (IC_{50}). Variables of 1118 descriptors was used as explanatory variables. The plot of maximal regression trees is shown in Fig. 1. Table S1 documents its associated node information, *i.e.*, the splitting variable and value in each split table node, the node output and the node size of each node. To select the optimum tree, 10-fold cross-validation was used. The optimum tree was selected from the maximal tree, which was pruned back with no change in the split limit.

Figure 2 shows the selected tree, indicating the splitting rules, the average response value and the object numbers of the leaves. Additionally, histograms plotted represent the distribution of the response to the objects within each node. For the optimal subtree with four terminal nodes, three molecular descriptors were selected to describe the binding affinity data. The amounts of these descriptors are shown in Table 1. The first selected molecular descriptor is R autocorrelation of lag 3/unweight (R3u) that is a GETAWAY descriptor. The other descriptor is a WHIM
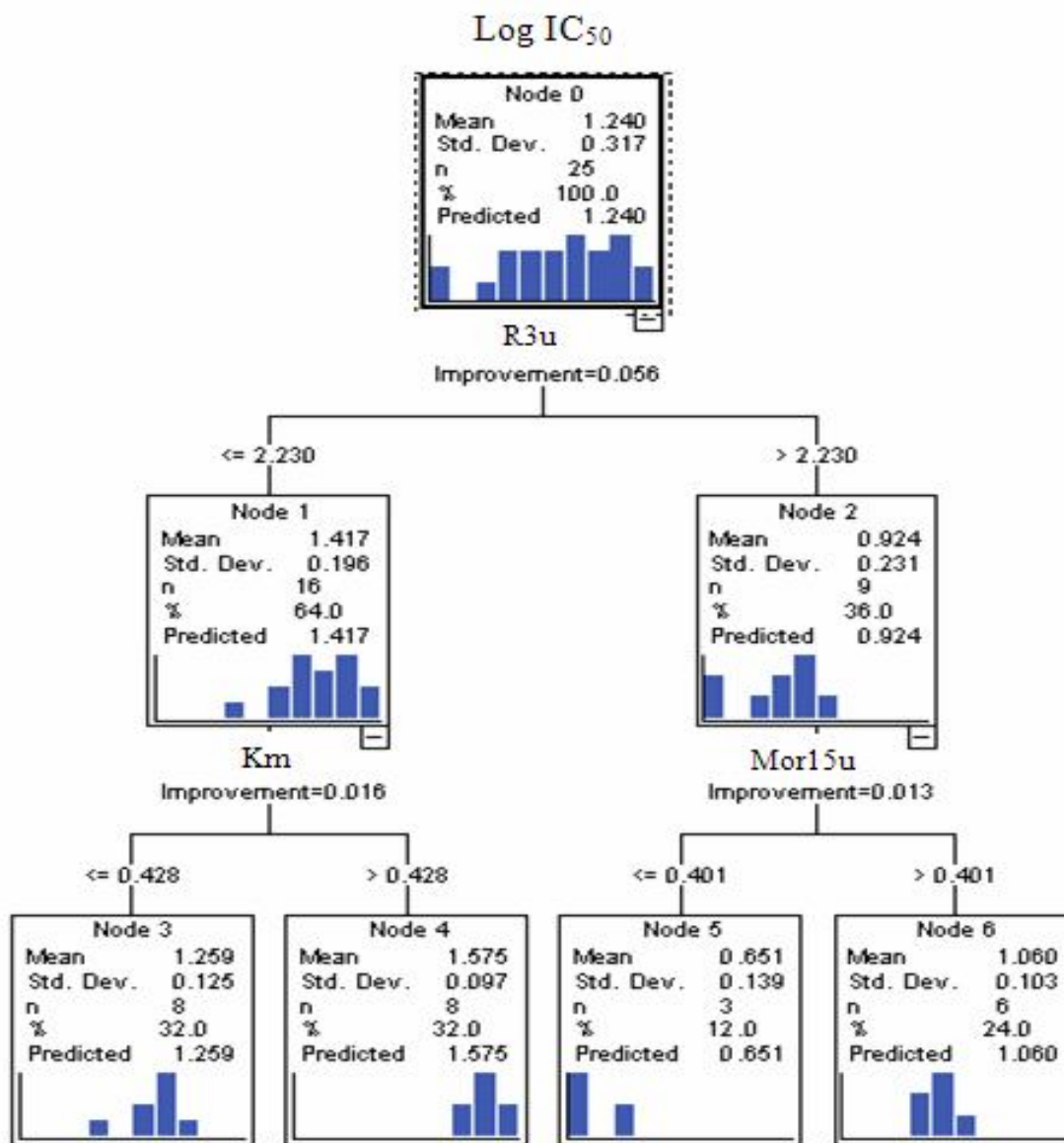
Fig. 2. The optimal tree used for studied inhibitors.

one, and K is global shape index/weighted by atomic masses (K_m). The third selected descriptor is 3D-MoRSE - signal 15/unweighted (Mor15u) from 3D-MoRSE descriptors, 3D-molecule representation of structures based upon electron diffraction. These descriptors are based on the idea of obtaining information from the 3D atomic coordinates by the transform used in electron diffraction studies for preparing theoretical scattering curves. The MoRSE descriptor is defined as follows:

$$Mor\,(s,w) = I(s,w) = \sum_{i=2}^{n} \sum_{j=2}^{i-1} w_i w_j \sin\,(sr_{ij})/(sr_{ij}) \qquad (3)$$

where, r_{ij} is the Euclidean distance between the atoms, and w_i and w_j are the weights of the atoms, i and j, respectively.

The first split divides the data into two groups including molecules with R3u values below and above 2.230, respectively. The second and third split divides the data into two groups, with molecules with K_m values below and

above 0.428, and Mor15u values below and above 0.401, respectively. The log(IC$_{50}$) values are divided into two groups by these splits very well. The optimal tree was applied to the prediction of the entire data set. In subset 3, the mean relative error and R were obtained as 3.56%, 1% and 0.9423, 0.8983 for calibration and prediction set, respectively. To make sure the demonstration of the absence of a chance correlation, the whole data set was divided into four subsets, and each subset was predicted by using the other three subsets as the training set. The results are shown in Table 2.

We applied the stepwise method to the total molecular descriptors (in subset 3), which in the present analysis included D = 1118 descriptors. The finest relationship achieved for this modeling, regarding to the best predictive power of the equation and the least variables involved, has the following statistics:

log(IC$_{50}$) = -0.1456($\pm$0.39) + 1.524($\pm$0.197)lop − 1.213($\pm$0.227)GATS5e − 1.37($\pm$0.291)E2m − 0.329($\pm$0.096)Mor24e　　　　　　　　　　　(4)
N = 25, R = 0.92, S = 0.136, F = 27.366

Table 3 shows the correlation matrix between log(IC$_{50}$) with every single variable in Eq. (4). It shows that the optimal descriptors are not seriously intercorrelated, thus justifying the inclusion of all variables in the relationship. There is a mixture of different classes of descriptors in Eq. (4).

Lop is a topological descriptor, GATS5e is a 2D autocorrelations descriptor, E2m is a WHIM descriptor and Mor24e is a 3D-MoRSE descriptor. As seen, the log(IC$_{50}$) decreases with Geary autocorrelation - lag 5/weighted by atomic Sanderson electronegativities (GATS5e), 2nd component accessibility directional WHIM index/weighted by atomic masses (E2m), and 3D-MoRSE-signal 24/weighted by atomic Sanderson electronegativities (Mor24e) and increase with Lopping centric index (Lop). We can also rank the descriptors in Eq. (4) according to their effect on increasing the value of S when removed from the model. In this case, the order found is:

Lop > GATS5e > E2m > Mor24e
After removing the descriptors from the model, the

values of S are 0.265, 0.207, 0.193 and 0.167 for Lop, GATS5e, E2m and Mor24e, respectively. So, the Lop and Mor24e have the most and lowest effect on the log(IC$_{50}$) in this model. Therefore, the most important variable is the Lop belonging to the topological descriptors. Lopping centric index (Lop) is an index defined as the mean information, content derived from the pruning partition of a graph. The lopping centric index is calculated as follows:

$$LOP = -\sum_{i=1}^{R} (n_g / A).\log (n_g / A)　　　　(5)$$

where n_g is the number of terminal vertices removed at the gth step, A is the number of graph vertices, and R *is* the number of steps to remove all graph vertices.

The statistical results offered by MLR are presented in Table 4, in addition to the CART results. From Table 4, correlation coefficient of 0.9200 and RMSE of 0.7416 for the training set, and a correlation coefficient of 0.1222 and an RMSE of 0.1547 for the test set were obtained, respectively. Compared to CART, MLR exhibits a considerable shift to the lower correlation coefficient and higher RMSE values, indicating superiorities of the CART in modeling structurally heterogeneous compounds.

In the total descriptors, some of the descriptors are more general chemosense than others. Correlation analysis was performed on the dependent variable (logIC$_{50}$) and 59 chemosense descriptors. The results were shown in Fig. 3. The details for the selected descriptors are given in Table S2.

In the next step, due to the similarity between some descriptors, principal component analysis (PCA) was used for reduction and classification of descriptors. In the current study, 59 chemosense descriptors were divided into five factors; PC1, PC2, PC3, PC4 and PC5 based on Fig. 4 and listed in Table S3. Figure 4 shows that five descriptors represent about 85% of the variations in the systems.

We refer to factor 1 to 5 as a size, hydrophobicity, shape, aromaticity and double bond factors, respectively. The dependence of log(IC$_{50}$) on these factors can be obtained by MLR. The resulting equation is as follows:

log(IC$_{50}$) = 1.24($\pm$0.53) − 0.186($\pm$0.054)PC1 + 0.015($\pm$0.60)PC2 + 0.095($\pm$0.54)PC3 + 0.050($\pm$0.054)PC4 + 0.002($\pm$0.58)PC5　　　(6)

Table 2. Verification of Statistical Validity of the Model

No of predicted data	The mean square error of calibration set	The mean square error of prediction set-b	R (Calibration set)	R (Prediction set)	The mean relative error of prediction set	The mean relative error of calibration set-b	Prediction set
1,5,9,13,17,21,25,29	0.0441	0.0113	0.9197	0.7899	1.14	1.27	Ser 1
2,6,10,14,18,22,26,30	0.0313	0.0086	0.9497	0.7569	7.50	0.22	Ser 2
3,7,11,15,19,23,27,31	0.0107	0.0284	0.9423	0.8983	3.56	1.00	Ser 3
4,8,12,16,20,24,28,32,33	0.0110	0.0476	0.9197	0.8203	1.47	3.49	Ser 4

Table 3. Correlation Matrix for Variables Entered in Eq. (4)

	$logIC_{50}$	Lop	GATS5e	E2m	Mor24e
$logIC_{50}$	1	0.730	-0.406	-0.327	0.428
Lop		1	-0.320	-0.176	0.110
GATS5e			1	0.091	-0.255
E2m				1	0.238
Mor24e					1

Table 4. Results of QSAR Analysis Using CART Coupled with those Obtained by MLR for Subset 3

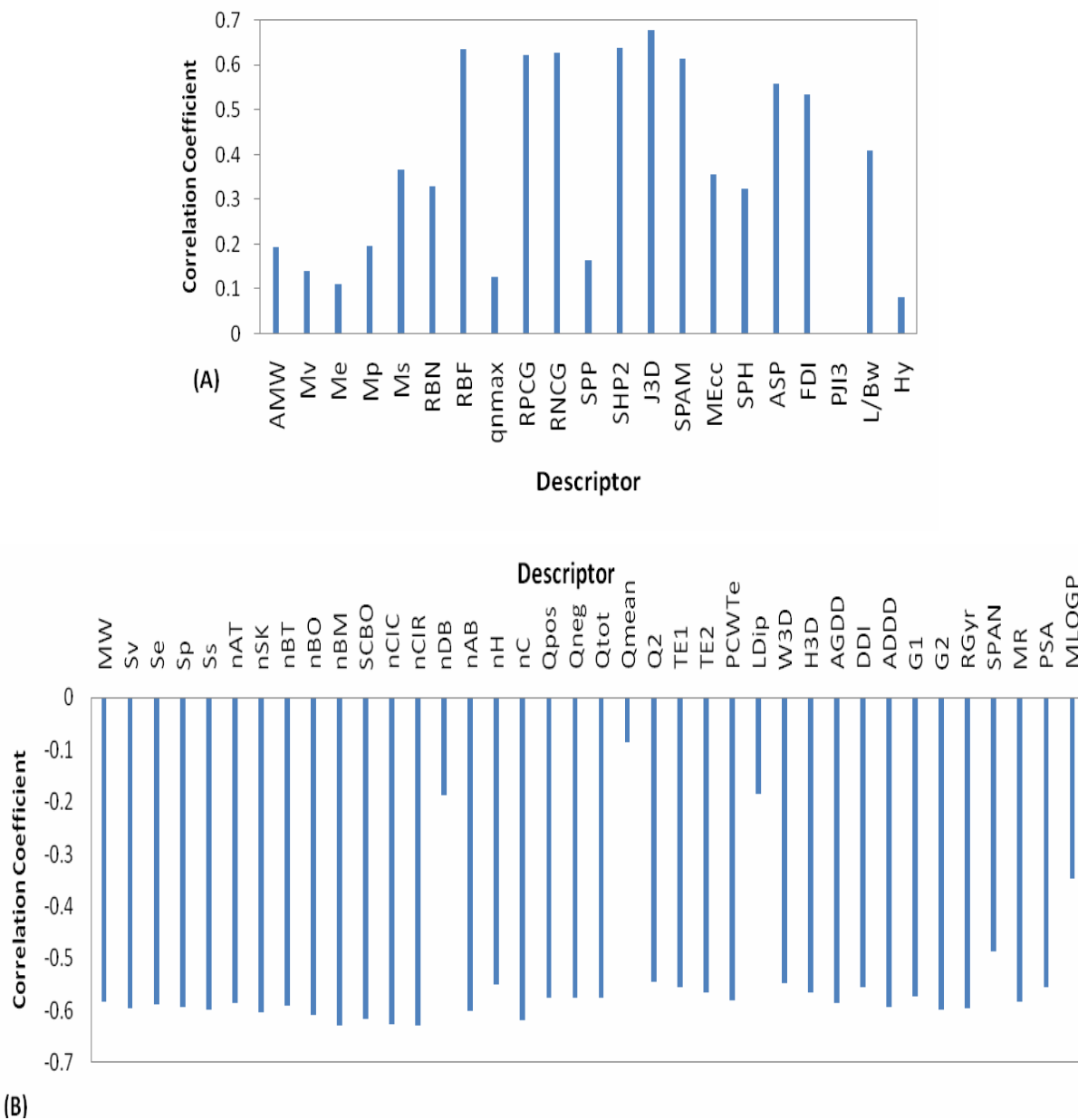| Data set	R (Correlation coefficient)		RMSE (Root mean squared error)	
	CART	MLR	CART	MLR
Training set	0.9423	0.9200	0.1034	0.1222
Test set	0.8983	0.7416	0.1685	0.1547

Fig. 3. Correlation between dependent variable log(IC_{50}) and selected independent variables (A: positive correlation, B: negative correlation), Y axis is the correlation coefficient (R) which obtained by bivariate correlation implemented in SPSS software. Definitions of descriptors in x-axis were listed in Table 7.

N = 5, R = 0.670, S = 0.263, F = 3.111

The plot of PC2 versus PC1 (Fig. 5) displays distributing compounds over the first two principal component space. The predicted values of log(IC_{50}) for the 33 mono and bis-

spiropyrrolizdine derivatives based on the CART, MLR and PCR methods are shown in Table 1. The correlation plots between the calculated and observed log(IC_{50}) values and dispersion plots of residual based on QSAR methods are shown in Fig. 6.
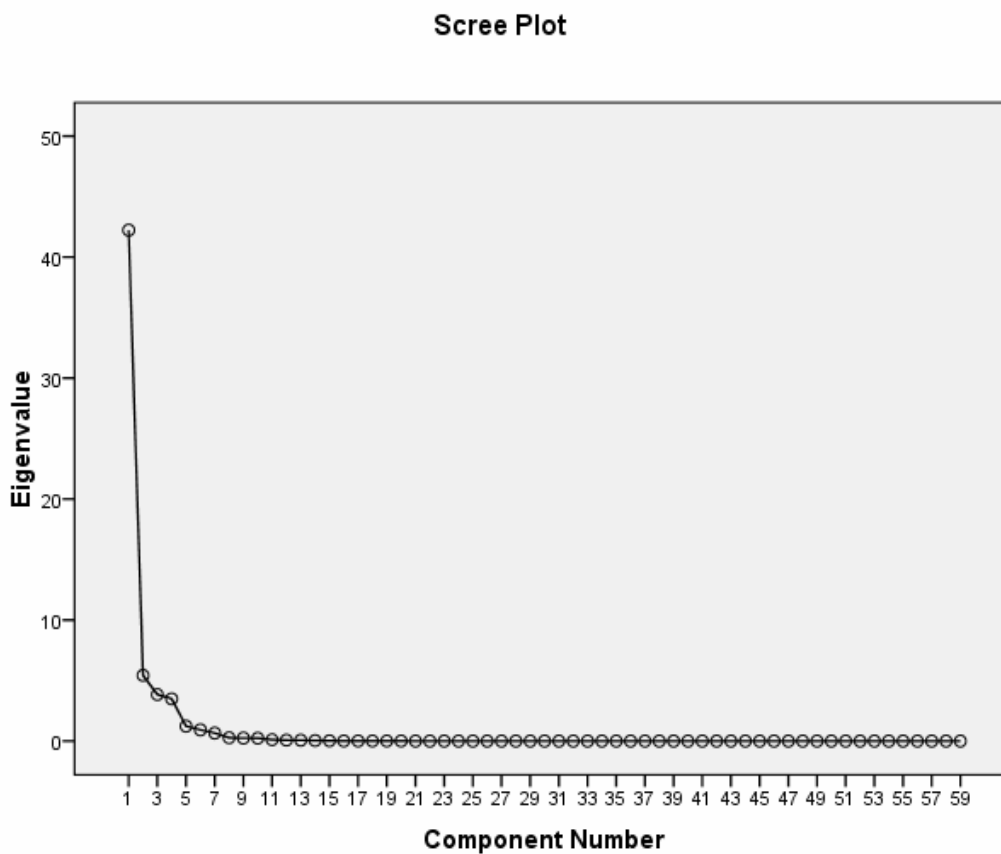
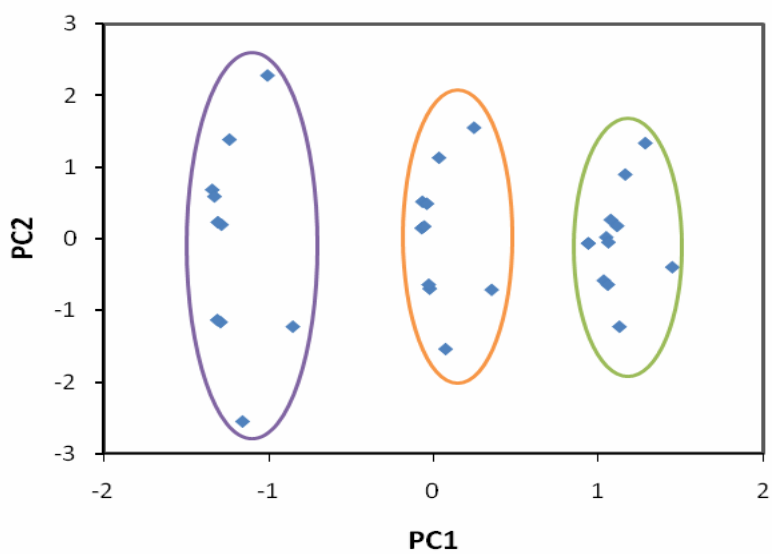**Fig. 4.** Scree plot for representation the effect of number of PCs on discussion of the system.

Fig. 5. Plot of PC2 against PC1: group 1 (5a-5k), group 2 (8a-8k), group 3 (9a-9k).
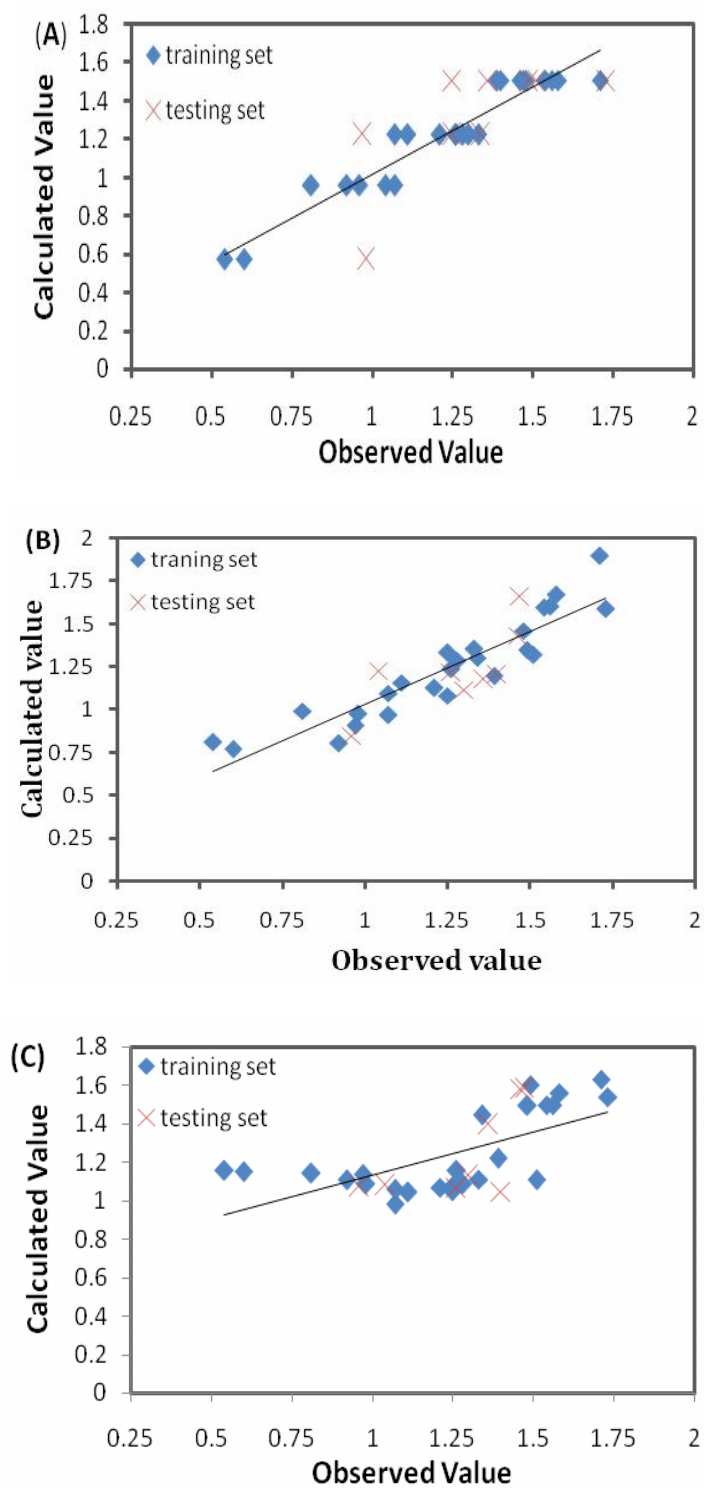
Fig. 6. Plots of calculated *vs*. observed activity (A) CART, (B) MLR and (C) PCR and dispersion plots of residual for three QSAR models based on (D) CART, (E) MLR and (F) PCR.
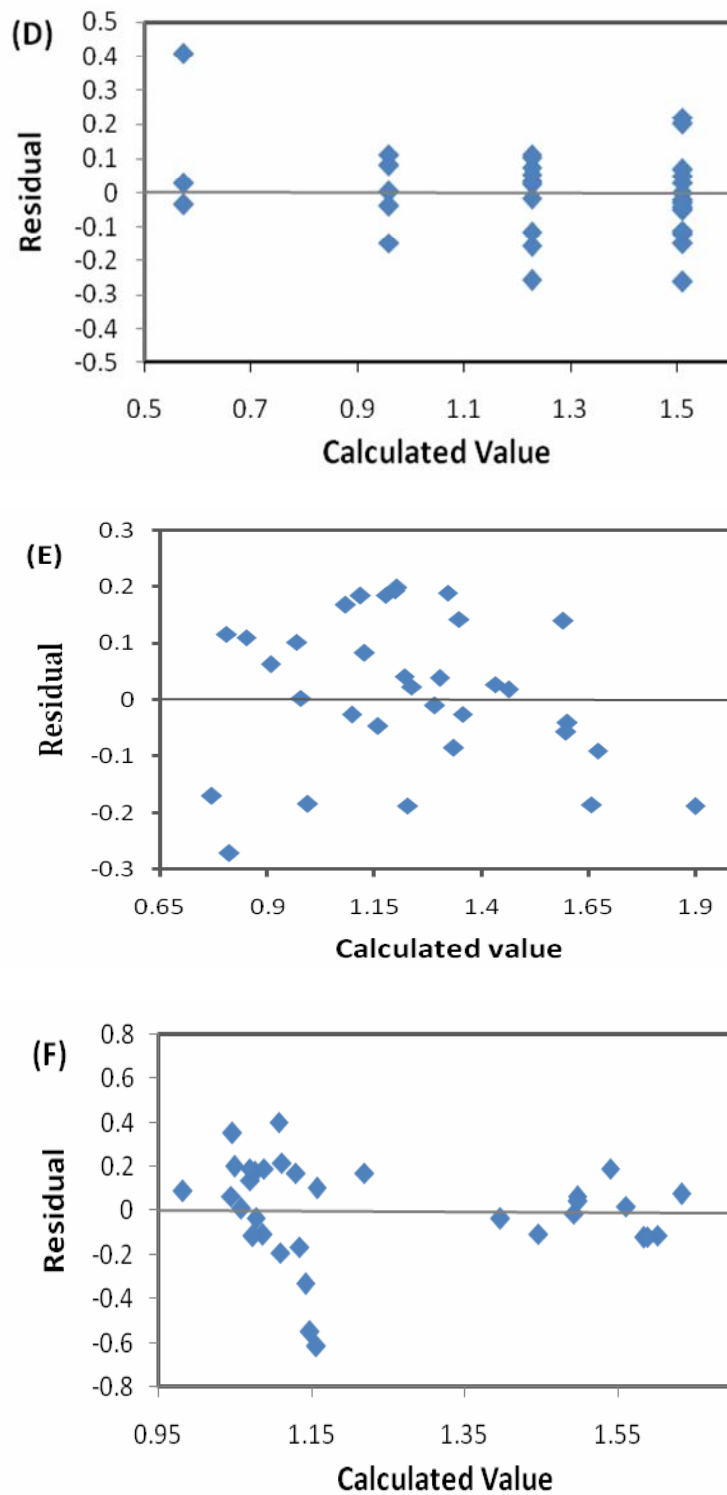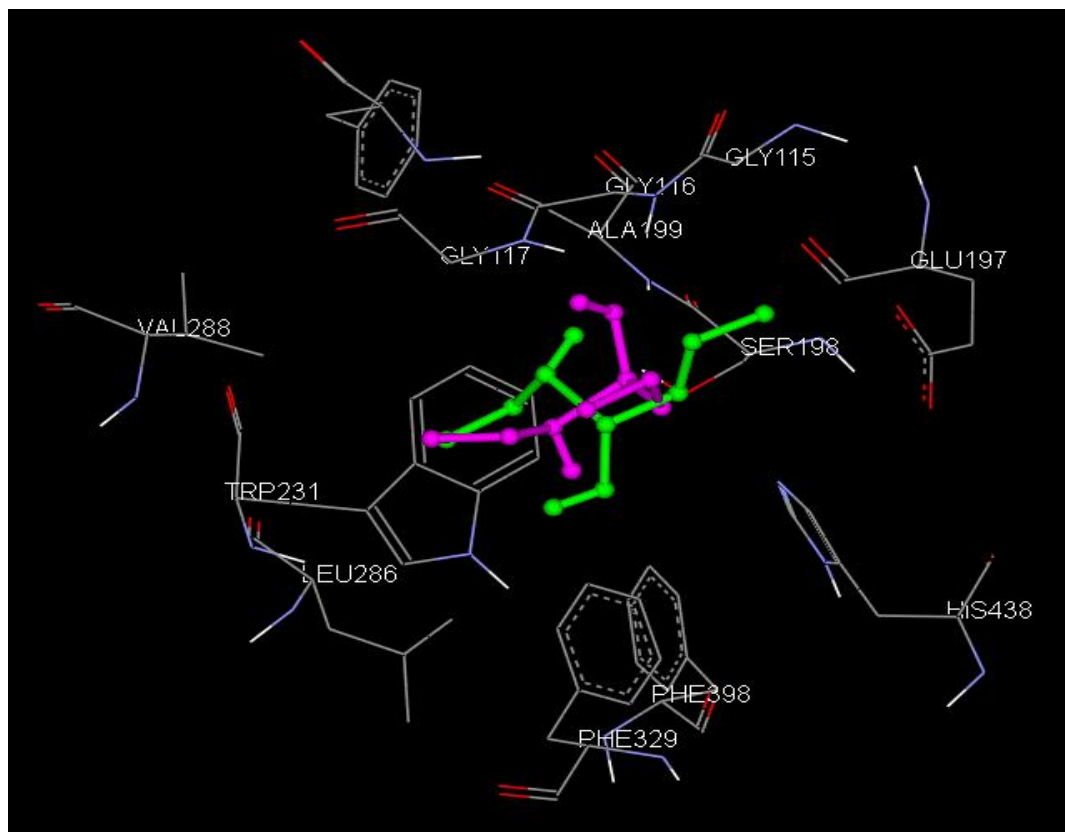
Fig. 6. Continued.

Fig. 7. Comparison between X-ray crystal structure (purple) and docking structure (green) obtained from AutoDock Vina.

Docking Calculation

Comparison between structure from docking and X-ray crystal structure. We performed a test to ensure the validity of docking calculations and the conditions as well as parameters of docking. The ligand was removed from the complex crystal structure (PDB code: 2WIJ) and then docking calculation was done into the evacuated 2WIJ enzyme. The docking procedure (100 runs) yielded only one cluster with conformations differing by less than 2 Å. The docked conformation of ligand chosen for this study was the lowest energy one. It was close to the crystal structure since the RMSD between two conformations was just 1.72 Å (by AutoDock Vina), which is quite satisfactory. Relative to the crystal conformation, ligand is docked correctly to active site with only a slightly different conformation. According to this result, AutoDock Vina significantly improves the average accuracy of the binding mode predictions. Figure 7 shows the comparison between the X-ray crystal structure of the ligand bound to 2WIJ enzyme (in purple color) and cluster conformation of the ligand docked to 2WIJ enzyme (in green color) by AutoDock Vina.

Docking of mono and bis-spiropyrrolizine derivatives. After validation of docking calculations, the docking was extended to the 33 ligands. The docking results on BuChE receptor were analyzed using Accelrys DS Visualizer v2.0.1 software for this inhibitor to their respective receptor. The docking results revealed that all the studied mono and bis-spiropyrrolizine derivatives have two binding sites (binding site 1 and binding site 2). The predicted binding conformations of all investigated inhibitors with the lowest binding energy and their alignment are shown in Fig. 8.

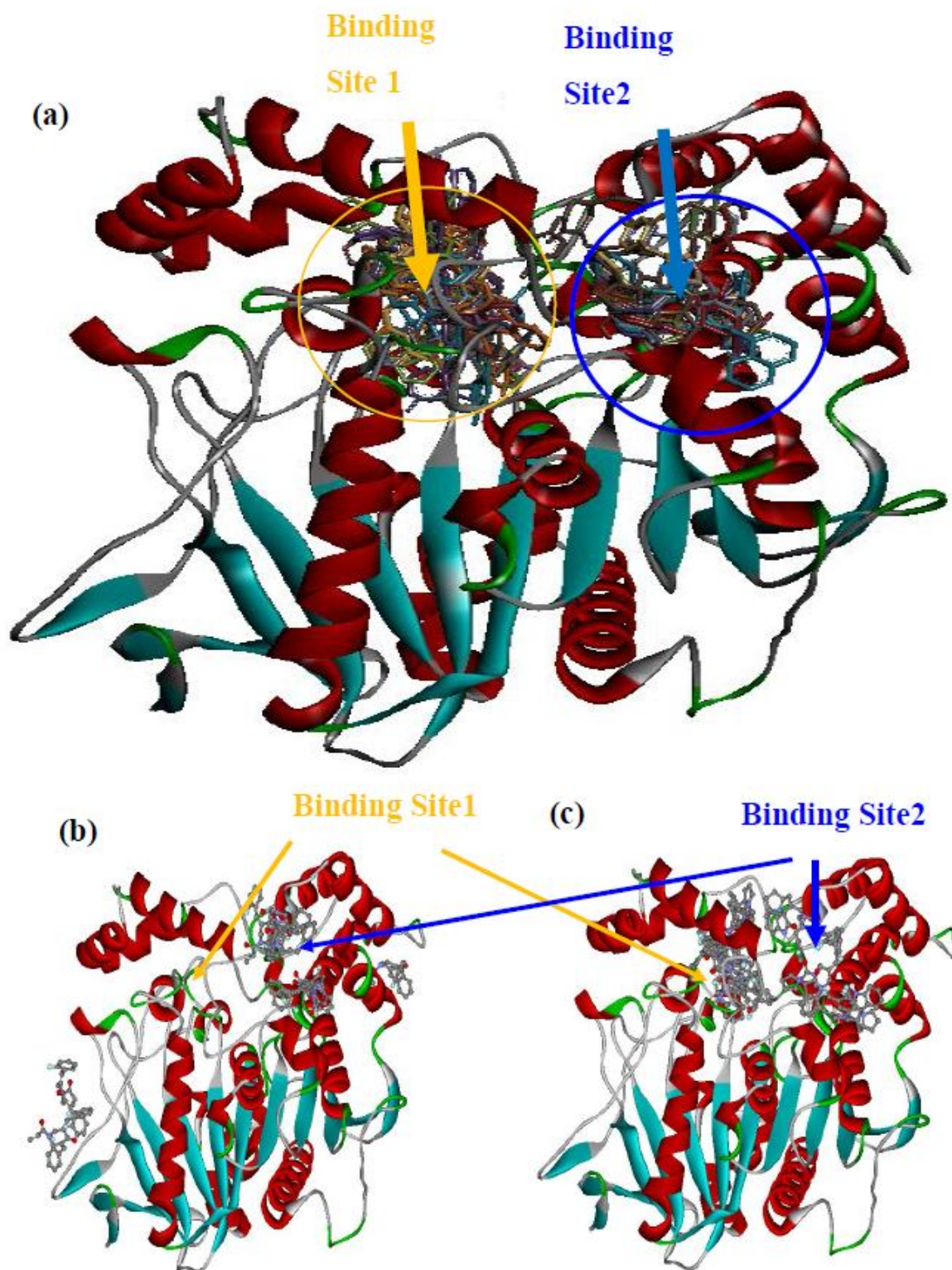Interaction modes of BuChE receptors with the most (8e compound) and lowest (5e compound) active molecules

Fig. 8. Binding sites for the lowest docking energies obtained for (a) all 33 investigated inhibitors, for the first ten negative docking energies belong to (b) 5e and (c) 8e.
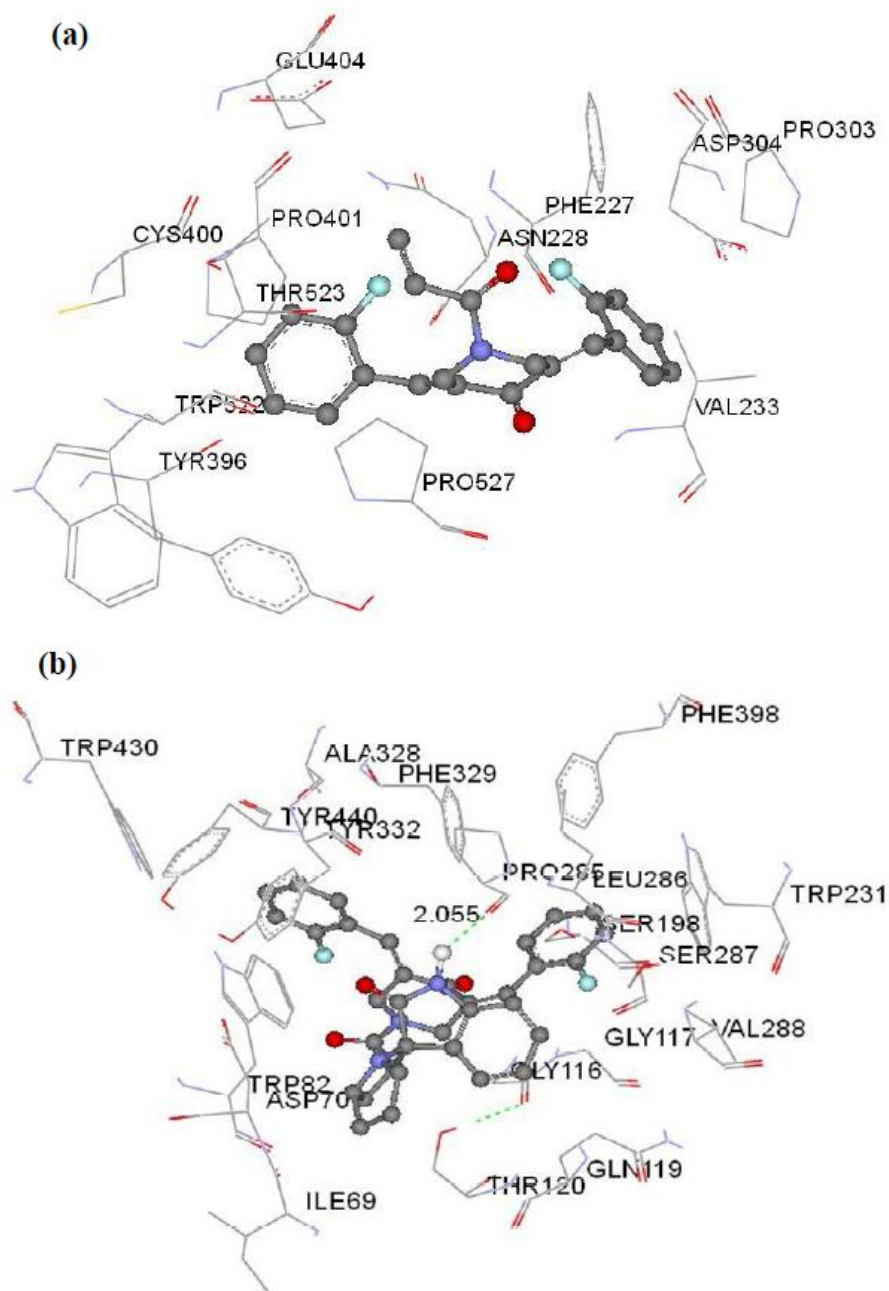
Fig. 9. Binding sites for (a) 5e and (b) 8e obtained by AutoDock Vina. Binding site amino acid residues are represented by lines, while the inhibitor is shown as ball and stick model with the atoms colored as carbon: gray, hydrogen: white, nitrogen: blue, and oxygen: red. The green dashed lines show the hydrogen bond and distance.

were investigated and are shown in Fig. 9. Compound 8e selected active site 1 while 5e was observed in active site 2. Docking studies on the most potent inhibitor, 8e, showed that the binding pocket includes Ile69, Asp70, Trp82, Gly116, Gly117, Gln119, Thr120, Ser198, Trp231, Pro285, Leu286, Ser287, Val288, Ala328, Phe329, Tyr332, Phe398,

Table 5. Docking Energy for the 33 Mono and Bis-spiropyrrolizine Derivatives

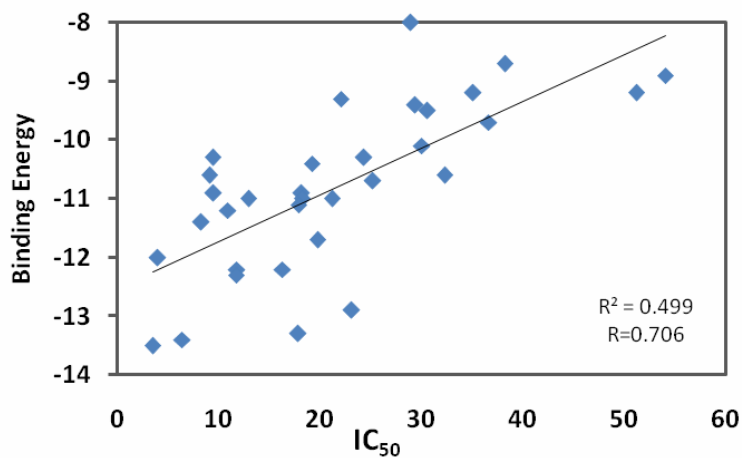| Compound	Docked energy (AutoDock Vina) (kcal mol^{-1})	Compound	Docked energy (AutoDock Vina) (kcal mol^{-1})
5b	-10.1	8h	-10.7
5c	-8	8i	-10.6
5d	-9.7	8j	-12
5e	-8.9	8k	-12.2
5f	-9.2	9a	-11
5g	-9.4	9b	-11
5h	-9.2	9c	-10.4
5i	-8.7	9d	-12.3
5j	-9.5	9e	-10.6
5k	-12.9	9f	-10.3
8a	-12.2	9g	-10.9
8b	-13.3	9h	-11.1
8c	-10.3	9i	-11.2
8d	-11.7	9j	-11.4
8e	-13.5	9k	-13.4
8f	-10.9		

Fig. 10. Correlation between binding energy and (IC$_{50}$) of 33 inhibitors from AutoDock Vina.
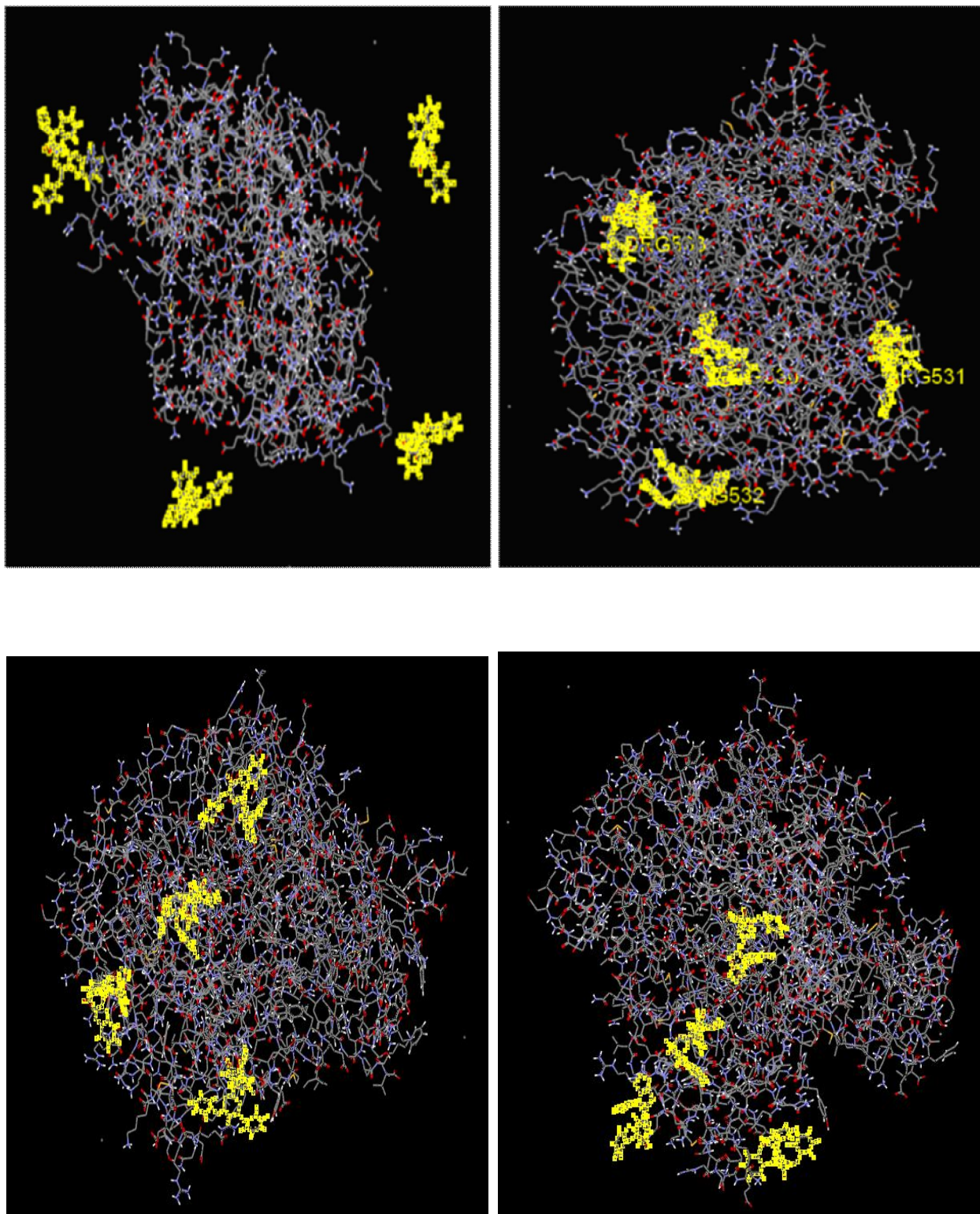
Fig. 11. Molecular snapshot in different MD simulation times for BuChE in the presence of 8e.
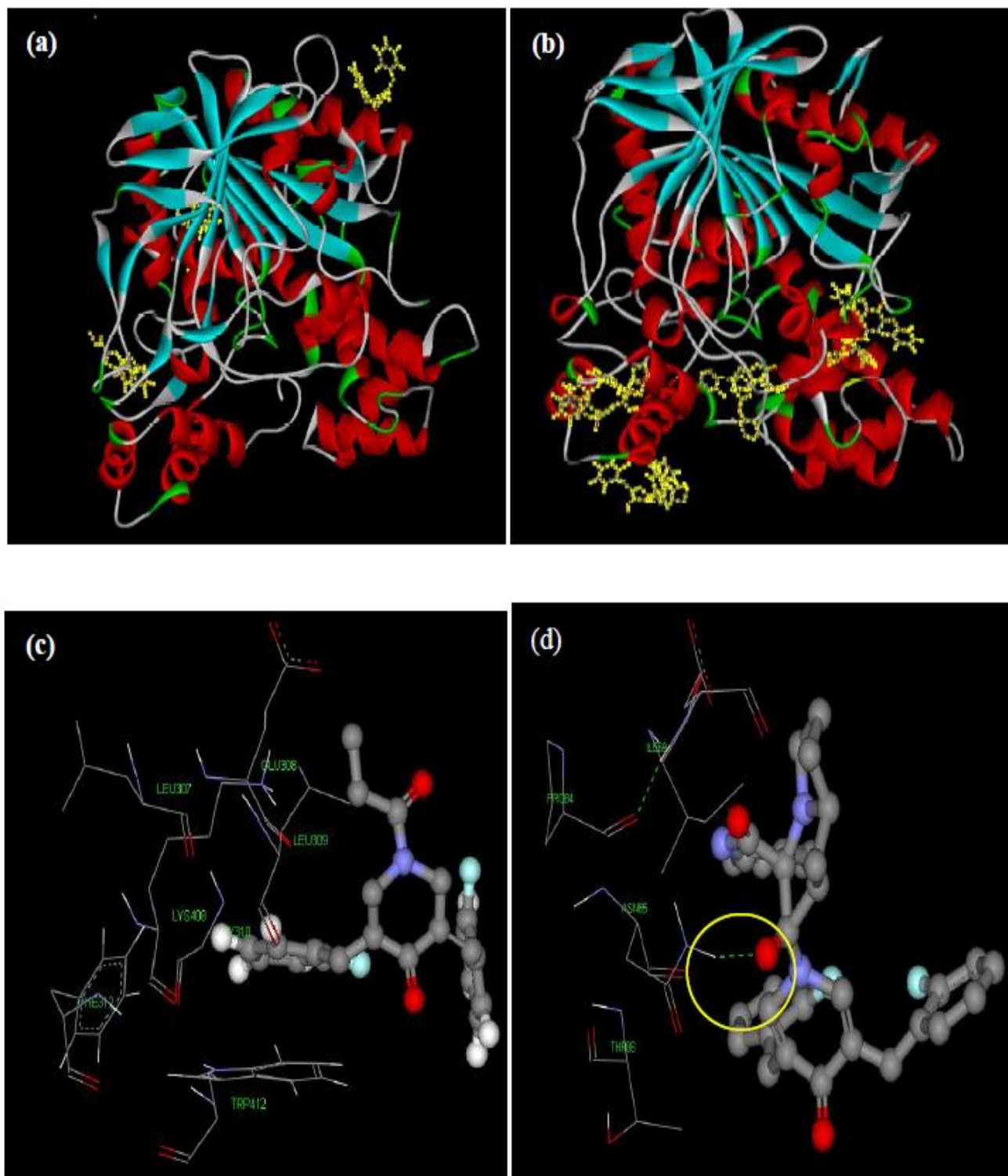
Fig. 12. The final structure of BuChE in the presence of (a) 5e and (b) 8e and close view of 5e (c) and 8e (d).

Table 6. Comparison between Binding Sites Obtained by Autodock and Molecular Dynamics Calculations

Inhibitor	Inhibitor number	Closest amino acids to the inhibitors obtained by MD	Docking site obtained by Autodock
5e	Ligand 1	Trp412, Lue309, Gly310, Phe312, Glu308, Lys408, Leu307	Val233, Phe227, Asn228, Pro303, Asp304, Tyr396, Cys400, Pro401, Glu404, Trp522, Thr523, Pro527
	Ligand 2	Gly296, Asp297, Val294, Asp295, Pro160, Asn150, Gly158, Pro157, Leu156, Ala162,Gly164,	
	Ligand 3	Ile462, Ser466, Thr506, Glu506, Thr508, Ser507, Arg465,	
	Ligand 4	No interaction	
8e	Ligand 1	Ile69, Asp70, Gln71, Asn85, Gln270	Ile69, Asp70, Trp82, Gly116, Gly117, Gln119, Thr120, Ser198, Trp231, Pro285, Leu286, Ser287, Val288, Ala328, Phe329, Tyr332, Phe398, Trp430, Tyr440
	Ligand 2	Tyr237, Asn241, Asn245, Leu244, Lys248, Arg254,	
	Ligand 3	Asn266, Asp268, Gln270, Glu271	
	Ligand 4	Pro 429, Pro431, Trp433, Gly75, Phe76, His77, Glu80, Met81, Asn341	

Trp430 and Tyr440. The results for the lowest potent inhibitor, 5e, showed that the binding pocket comprises of Val233, Phe227, Asn228, Pro303, Asp304, Tyr396, Cys400, Pro401, Glu404, Trp522, Thr523, and Pro527. The hydrogen bond is the interaction between polar hydrogen and electron density of electronegative elements (such as oxygen and nitrogen). Hydrogen bonding plays a significant role in physiological or biochemical performance of the molecule. For 8e-BuChE receptor, there is an important hydrogen bond between the carbonyl oxygen of Pro285 and one amino hydrogen ($>C=O\ldots H-N$) with an O$\ldots$H distance 2.055 Å (AutoDock Vina). For 5e-BuChE receptor, there was not any hydrogen bond.

The AutoDock software is also employed to determine docking energy for the 33 mono and bis-spiropyrrolizine derivatives toward the BuChE receptor. The results are listed in Table 5. By the investigation of the correlation

between inhibitory activity values (IC_{50}) and docking energies, it was clear that there is a rational correlation. The calculated docking energy and experimental (IC_{50}) were obtained, and are shown in Fig. 10.

Molecular Dynamics Simulation

The MD simulation gives us a microscopic insight into ligand and macromolecule interactions. In this method, an initial velocity is estimated regarding temperature, kinetic energy, and Maxwell distribution. The position of atoms was obtained using Verlett or Leap-frog integration algorithms through the MD simulation. After stability of the system (smoothing the energy and RMSD curves), analyzing, and extraction of structural and thermodynamic parameters were obtained by time averaging from the last period of MD simulation.

In MD calculation, four molecules having the highest
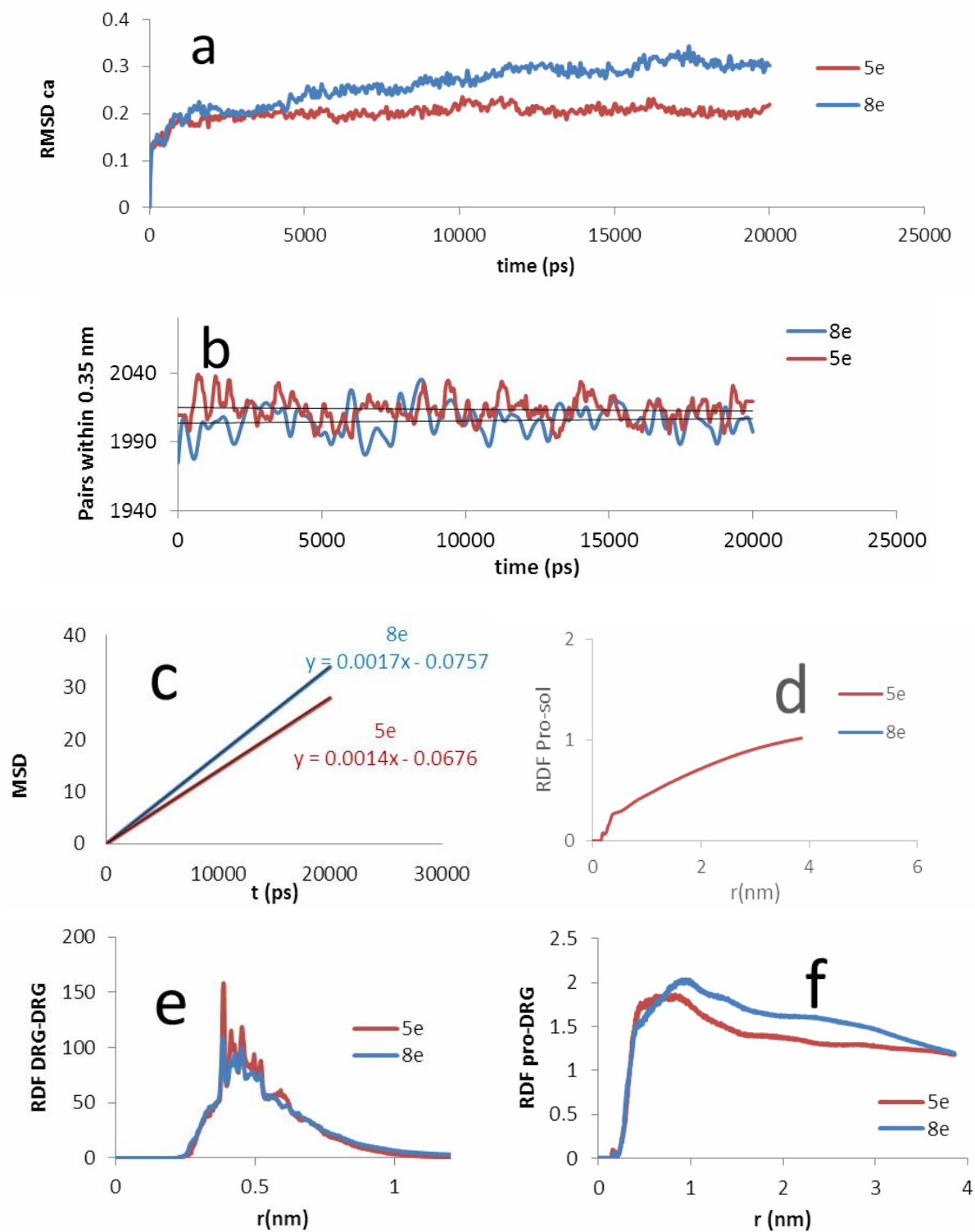
Fig. 13. Variation of carbon alpha RMSD (a) BuChE intramolecular hydrogen bond (b) MSD (c) radial distribution function for BuChE-solvent (d) radial distribution function for inhibitor-inhibitor (e) and BuChE-inhibitor (f).

Table 7. Structural Data for Buthylcholinesterase in the Absence and Presence of Two Inhibitors Extracted
from MD Simulation and VADAR

	BuChE	BuChE+5e	BuChE+8e
Helix	176	98	100
Beta	119	109	95
Coil	232	261	273
Turn	128	80	80
Mean hbond distance	2.2	2.3	2.3
Mean hbond energy	-1.7	-1.6	-1.5
#Res with hbonds	376	315	322
Mean Helix Phi	-66.9	-61.5	-60.2
Mean Helix Psi	-36.0	-39.9	-41.8
#Res with Gauche+Chi	232	171	174
#Res with Gauche-Chi	73	80	82
#Res with Trans Chi	116	112	107
Mean Chi Gauche+	-64.3	-69	-67.2
Mean Chi Gauche-	63.4	59.2	60
Mean Chi Trans	169.1	165	166.7
Std. dev of chi pooled	13.57	13.15	13.44
Mean Omega ($|omega| > 90$)	179.6	178.1	177.5
#Res with $|omega| < 90$	7	46	45
#Res in phipsi core	462	311	316
#Res in phipsi allowed	58	112	105
#Res in phipsigenerous	2	22	24
#Res in phipsi outside	5	23	23
#Res in omega core	450	306	312
#Res in omega allowed	56	73	62
#Res in omega generous	10	25	31
#Res in omega outside	11	64	63
#Packing defects	51	83	176

Table 7. Continued

Free energy of folding	542	466	463
#Res 95% buried	201	130	129
#Buried charges	13	7	11
Total ASA	20400.7	21040.5	21662.6
ASA of backbone	2408.5	3424.8	3475.1
ASA of sidechains	17992.2	17615.7	18187.5
ASA of C	12077.6	11932.3	12276.6
ASA of N	1746.8	2118.7	2399.0
ASA of N^+	1353.5	619.1	580.3
ASA of O	4203.3	5068.8	4904.7
ASA of O^-	947.0	1214.6	1394.6
ASA of S	72.5	86.9	107.3
Exposed nonpolar ASA	11887.4	11718.8	11942.1
Exposed polar ASA	4771.7	5588.5	5464.5
Exposed charged ASA	3741.6	3733.3	4256.0
Side exposed nonpolar ASA	11890.5	11422.6	11624.6
Side exposed polar ASA	2449.4	2542.5	2374.9
Side exposed charged ASA	3652.3	3650.6	4188.0
Fraction nonpolar ASA	0.58	0.56	0.55
Fraction polar ASA	0.23	0.27	0.25
Fraction charged ASA	0.18	0.18	0.2
Mean residue ASA	38.7	45.0	46.3
Mean frac ASA	0.2	0.3	0.3
%Side ASA hydrophobic	25.41	30.69	30.9

IC_{50} (5e) and lowest IC_{50} (8e) were selected, and MD calculation was done same as method section. Molecular snapshots for the enzyme in the presence of four molecules of an inhibitor (8e in here) are shown in Fig. 11.

Structure of the enzyme in the presence of one of the studied inhibitors at the last MD simulation time for both 5e and 8e is shown in Fig. 12 (a and b). Close view of parts a and b was depicted in part c and d in which the closet amino acids and hydrogen bond (green dotted line) are also represented. As shown, the hydrogen bond was only observed between 8e and enzyme. This may justify the more tendency of 8e to interact with BuChE. Comparison
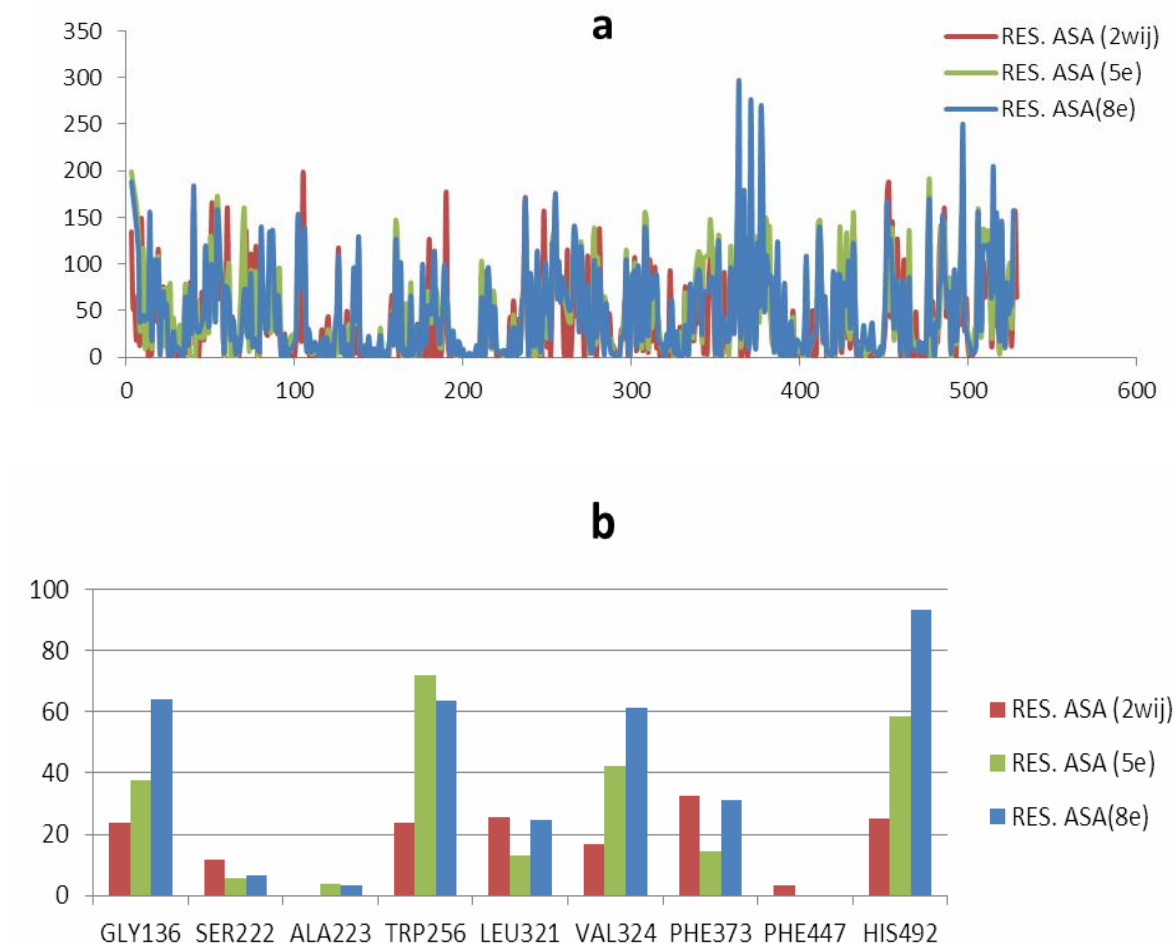
Fig. 14. Accessible surface area for (a) total residues of BuChE (b) for active site amino acids in the absence (2wij) and the presence of 5e and 8e.

between a binding site obtained from docking calculation and MD simulation shows that binding site of a ligand such as 8e in MD is similar to the docking site (Ile 69 and Asp70 in Table 6). Some of the structural and energetic parameters such as RMSD, surface area, hydrogen bond, and radial distribution functions (RDF) can be obtained by analyzing trajectory file. Variation of RMSD for carbon alpha and hydrogen bond was obtained and plotted in Fig. 13. It shows that RMSD carbon alpha is higher in the presence of 8e while intera protein hydrogen bond is lower and RMSD ligand is lower than 5e.

Protein-ligand, protein-solvent, and ligand-ligand radial distribution functions (RDF) were also obtained. The results show that 8e is more distributed around protein relative to 5e, while 5e intends to do self-interaction. Protein water interaction does not differ in the presence of two ligands (Fig. 13d). Structure of enzyme at the last MD simulation time was extracted (Fig. 11) and compared with initial time. Then, final structure entered into VADAR and some of the useful parameters were calculated. The secondary structure, such as helix, coil, beta percentage, surface area, and angles were calculated and listed in Table 7.

Diffusion coefficient (D_A) is the slope of MSD curve versus *t* obtained from Eq. (7). As shown in Fig. 13c, slope and so D_A are higher for 8e compared to 5e indicating stronger interaction of 8e with BuChE.

$$MSD = (||\, r_i(t) - r_i(0)\, ||^2)$$

$$\lim \left\langle \left\| r_i(t)_{r \to \infty} - r_i(0) \right\|^2 \right\rangle_i = 6D_A t$$

(7)

The results show that both ligands decrease the helix and increase the coil (decreasing the secondary structure), and 8e is more effective. It also shows that enzyme becomes more exposed in the presence of 8e and so its structure shifts to higher compared to 5e. Accessible surface areas for individual amino acids and active sites are shown in Fig. 14. As indicated in this figure, individual and active site amino acids have higher values of surface area in the presence of 8e that is comparable with other results.

CONCLUSIONS

The aim of this work was to develop the QSAR method using classification and regression tree methodology and molecular docking as well as molecular dynamics for 33 butyrylcholinesterase inhibitors. The generated tree was evaluated and applied for the prediction of inhibitory activity values of mono and bis-spiropyrrolizine derivatives. The application of CART to this data set has demonstrated that the CART analysis can perform a better prediction than MLR method in terms of prediction accuracy. Docking results showed that the studied 33 mono and bis-spiropyrrolizine derivatives are docked into common binding sites of butyrylcholinesterase. The calculated docking energies by molecular docking showed a rather good correlation with experimental (IC_{50}) values. The results of this study can be used to design new compounds with better biological activity. Molecular dynamics results also showed that more effective inhibitor has higher tendency to interact with the enzyme and changing its structure.

ACKNOWLEDGMENTS

REFERENCES

[1]   Mesulam, M.; Guillozet, A.; Shaw, P.; Quinn, B., Widely spread butyrylcholinesterase can hydrolyze acetylcholine in the normal and Alzheimer brain. *Neurobiol. Dis.* **2002**, *9*, 88-93, DOI: 10.1006/nbdi.2001.0462.

[2]   Stedman, E.; Stedman, E.; Easson, L. H., Choline-esterase. An enzyme present in the blood-serum of the horse. *Biochem. J.* **1932**, *26*, 2056-2066, DOI: 10.1042/bj0262056.

[3]   Darvesh, S.; Hopkins, D. A.; Geula, C., Neurobiology of butyrylcholinesterase. *Nat. Rev. Neurosci.* **2003**, *4*, 131-138, DOI: 10.1038/nrn1035.

[4]   Raveh, L; Grauver, E; Grunwald, J; Cohen, E.; Ashani, Y., The stoichiometry of protection against soman and VX toxicity in monkeys pretreated with human butyrylcholinesterase. *Toxicol. Appl. Pharm.* **1997**, *145*, 43-53, DOI: 10.1006/taap.1997.8160.

[5]   Mesulam, M.; Giacobini, E., Neuroanatomy of cholinesterase in the normal human brain and in Alzheimer's disease, in: Cholinesterase and Cholinesterase Inhibitors. Martin Donitz, London: 2000; p. 121-137.

[6]   Giacobini, E., Cholinergic function and Alzheimer's disease. *Int. J. Geriatr. Psychiatry.* **2003**, *18*, S1-S5, DOI: 10.1002/gps.935.

[7]   Allam, A. R.; Sridhar, G. R.; Thota, H.; Babu, C. S.; Prasad, A. S.; Divakar, C., Alzheimer's disease and Type 2 Diabetes mellitus: the cholinesterase connection?. *Lipids. Health. Dis.* **2006**, *5*, 28-32, DOI: 10.1186/1476-511X-5-28.

[8]   Greig, N. H; Utsuki, T.; Ingram, D, K.; Wang, Y.; Pepeu, G.; Scali, C,; Yu, QS.; Mamczarz, J.; Holloway, H. W.; Giordano, T.; Chen, D.; Furukawa, K.; Sambamurti, K.; Brossi, A.; Lahiri, D. K., Selective Butyrylcholinesterase inhibition elevates brain acetylcholine, augments learning and lowers Alzheimer beta-amyloid peptide in rodent. *Proc. Natl. Acad. Sci. USA.* **2005**, *102*, 17213-17218, DOI: 10.1073/pnas.0508575102.

[9]   Kuntz, I. D., Structure-based strategies for drug design and discovery. *Science.* **1992**, *257*, 1078-

1082, DOI: 10.1126/science.257.5073.1078.

[10] Blundell, T. L., Structure-based drug design. *Nature.* 1996, *384*, 23-36, DOI: 10.1038/384023a0.

[11] Wang, N. C. Y.; Venkatapathy, R.; Bruce, R. M.; Moudgal, C. J., Development of quantitative structure-activity relationship models to predict the carcinogenic potency of chemicals. II: Using oral slope factor as a measure of carcinogenic potency. regul. *Toxicol. Pharmacol.* **2011**, *59*, 215-226, DOI: 10.1016/j.yrtph.2010.09.019.

[12] Lorber, D. M., Computational drug design. *Chem. Biol.* **1999**, *6*, R227-R228, DOI: http://dx.doi.org/10.1016/S1074-5521 (99) 80093-3.

[13] Pooyafar, M.; Ajloo, D., Tertiary structure prediction of a-glucosidase and inhibition Properties of N(phenoxydecyl) phthalimide derivatives. *Acta. Chim. Slov.* **2012**, *59*, 233-241.

[14] Hemmateenejad, B.; Akhond, M.; Miri, R.; Shamsipur, M., Genetic algorithm applied to the selection of factors in principal component-artificial neural networks: Application to QSAR study of calcium channel antagonist activity of 1,4-dihydropyridines (nifedipine analogous). *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1328-1334, DOI: 10.1021/ci025661p.

[15] Moosavi-Movahedi, A. A.; Safarian, S.; Hakimelahi, G. H.; Ataei, G.; Ajloo, D.; *et al.*, QSAR analysis for ADA upon interaction with a series of adenine derivatives as inhibitors. *Nucleos. Nucleot. Nucleic Acids.* **2004**, *23*, 613-624, DOI: 10.1081/NCN-120030719.

[16] Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P. A.; Markopoulos, J.; Igglessi-Markopoulou, O., A novel simple QSAR model for the prediction of anti-HIV activity using multiple linear regression analysis. *Mol. Divers.* **2006**, *10*, 405-414, DOI: 10.1007/s11030-005-9012-2.

[17] Ajloo, D.; Saboury, A. A.; Haghi-Asli, N.; Ataei-Jafarai, G.; Moosavi-Movahedi, A. A.; Ahmadi, M.; Mahnam, K.; Namaki, S., Kinetic, thermodynamic and statistical studies on the inhibition of adenosine deaminase by aspirin and diclofenac. *Enz. Inhib. Med. Chem.* **2007**, *22*, 395-406, DOI: 10.1080/

14756360701229085.

[18] Sahebjamee, H.; Yaghmaei, P.; Abdolmaleki, P.; Foroumadi, A. R., Quantitative structure-activity relationships study of carbonic anhydrase inhibitors using logistic regression model. *Iran. J. Chem. Chem. Eng. (IJCCE),* **2013**, *32*, 19-29.

[19] Nandi, S.; Bagchi, M. C., 3D-QSAR and molecular docking studies of 4-anilino quinazoline derivatives: a rational approach to anticancer drug design. *Mol. Divers.* **2010**, *14*, 27-38, DOI: 10.1007/s11030-009-9137-9.

[20] Karthikeyan, S.; Bharanidharan, G.; Kesherwani, M.; Mani, K. A.; Srinivasan, N.; Velmurugan, D.; Aruna, P.; Ganesan, S., Insights into the binding of thiosemicarbazone derivatives with human serum albumin: spectroscopy and molecular modelling studies. *J. Biomol. Struct. Dyn.* **2016**, *34*, 1264-1281, DOI: 10.1080/07391102.2015.1075905.

[21] Balasubramanian, P. K.; Balupuri, A.; Cho, S. J., Molecular modeling studies of trisubstituted thiazoles as Cdc7 kinase inhibitors through 3D-QSAR and molecular docking simulation. *Bull. Korean Chem. Soc.* **2015**, *36*, 1599-1612, DOI: 10.1002/bkcs.10304.

[22] Chiou, S. Y.; Weng, T. T.; Lin, G. Z.; Lu, R. J.; Jian, S. Y.; Lin, G., Molecular docking of different inhibitors and activators to butyrylcholinesterase. *J. Biomol. Struct. Dyn.* **2015**, *33*, 563-572, DOI: 10.1080/07391102.2014.896749.

[23] Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J., Classification and Regression Trees. Wadsworth, Monterey, 1984.

[24] Put, C.; Perrin-Questier, F.; Coomans, D.; Massart, D.L.; Vander-Heyden, Y., Classification and regression tree analysis for molecular descriptor selection and retention prediction in chromatographic quantitative structure-retention relationship studies. *J. Chromatogr.* **2003**, *988*, 261-276, 10.1016/S0021-9673(03)00004-9.

[25] Lavrac, N., Selected techniques for data mining in medicine. *Artif. Intell. Med.* **1999**, *16*, 3-23, DOI: 10.1016/S0933-3657(98)00062-1.

[26] Marshall, R. J., The use of classification and

regression trees in clinical epidemiology. *J. Clin. Epidemiol.* **2001**, *54*, 603-609, DOI: 10.1016/S0895-4356(00)00344-9.

[27] De'Ath, G.; Fabricius, K. E., Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology.* **2000**, *81*, 3178-3192, 10.1890/0012-9658(2000)081[3178: CARTAP]2.0.CO;2.

[28] Tittonell, P.; Shepherd, K. D.; Vanlauwe, B.; Giller, K. E., Unravelling the effects of soil and crop management on maize productivity in smallholder agricultural systems of western kenya: An-application of classification and regression tree analysis. *Agric. Ecosyst. Environ.* **2008**, *123*, 137-150, DOI: 10.1016/j.agee.2007.05.005.

[29] Questier, F.; Put, R.; Coomans, D.; Walczak, B.; Vander-Heyden, Y., The use of CART and multivariate regression trees for supervised and unsupervised feature selection. *Chemom. Intell. Lab.* **2005**, *76*, 45-54, DOI: 10.1016/j.chemolab. 2004.09.003.

[30] Jalali-Heravi, M.; Shahbazikhah, P., Quantitative structure-mobility relationship study of a diverse set of organic acids using classification and regression trees and adaptive neuro-fuzzy inference systems. *Electrophoresis.* **2008**, *29*, 363-374, DOI: 10.1002/ elps.200700136.

[31] Asadollahi-Baboli, M., Quantitative structure-activity relationship analysis of human neutrophil elastase inhibitors using shuffling classification and regression trees and adaptive neuro-fuzzy inference systems. *SAR. QSAR. Environ. Res.* **2012**, *23*, 505-520, DOI: 10.1080/1062936X.2012.665811.

[32] Kia, Y.; Osman, H.; Kumar, R. S.; Murugaiyah, V.; Basiri, A.; Perumal, S.; Wahab, H. A.; Bing, C. S., Synthesis and discovery of novel piperidone-grafted mono- and bis-spirooxindole-hexahydropyrrolizines as potent cholinesterase inhibitors. *Bioorg. Med. Chem.* **2013**, *21*, 1696-1707, DOI: 10.1016/ j.bmc.2013.01.066.

[33] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, R. J.; Montgomery, Jr J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; and *et al.*, Gaussian 03, RevisionE.01. Gaussian, Inc, Wallingford, CT, 2004.

[34] Schüttelkopf, A. W.; van-Aalten, D. M. F., PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta. Crystallogr.* **2004**, *D60*, 1355-1363, DOI: 10.1107/ S0907444904011679.

[35] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; MerzJr, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197, DOI: 10.1021/ja00124a002.

[36] Ewald, P., Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369*, 253-287, DOI: 10.1002/andp.19213690304.

[37] van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., GROMACS: fast, flexible and free. *J. Comp. Chem.* **2005**, *26*, 1701-1718, DOI: 10.1002/jcc.20291.

[38] Willard, L.; Ranjan, A.; Zhang, H.; Monzavi, H.; Boyko, R. F.; Sykes, B. D.; Wishart, D. S., "VADAR: a web server for quantitative evaluation of protein structure quality". *Nucleic. Acids. Res.* **2003**, *31*, 3316-3319.