Journal of Physical and Theoretical Chemistry
of Islamic Azad University of Iran, 14 (2) 149-164: Summer 2017
(J. Phys. Theor. Chem. IAU Iran)
ISSN 1735-2126

4D-QSAR analysis and pharmacophore modeling: propoxy methylphenyl oxasiazole derivatives by electron conformatitional-genetic algorithm method

Burak Tüzün[1], Sevtap Çağlar Yavuz[2] and Emin Sarıpınar[2,*]

[1] Department of Chemistry, Science Faculty, Cumhuriyet University, Sivas, Turkey
[2] Department of Chemistry, Science Faculty, Erciyes University, Kayseri, Turkey

Received November 2017; Accepted December 2017

ABSTRACT

In this 4D-QSAR study, we obtained pharmacophore identification and biological activity prediction for 50 propoxy methylphenyl oxadiazole derivatives by the Electron Conformational Genetic Algorithm approach. In light of the results given in the data obtained from quantum chemical calculations at HF/3-21 G level, the electron conformational matrices of congruity (ECMC) were built by EMRE software. Considering the pharmacophore atoms, a parameter pool was introduced into the field. To find the theoretical biological activity of the molecules used in this study, the non-linear least squares regression method and genetic algorithm were used to determine the best subset of variables affecting bioactivity. As can be understand from our explanations, it should be noted that the results obtained in this study are in good agreement with the experimental data presented in the literature. The model for the training and test sets attained by the optimum 8 parameters gave highly satisfactory results with $R^2_{training}$= 0.872, q^2=0.836 and $SE_{training}$=0.059, q^2_{ext1} = 0.787, q^2_{ext2} = 0.786, q^2_{ext3}=0.830, ccc_{tr} = 0.933, ccc_{test} = 0.896 and ccc_{all} = 0.926.

Keywords: 4D-QSAR; propoxy methylphenyl oxasiazole derivatives; pharmacophore; electron conformational-genetic algorithm method

INTRODUCTION

The lysophospholipid receptor (LPL-R) group is a member of the G protein-coupled receptor family of integral membrane proteins. The lysophospholipid sphingosine 1-phosphate (S1P) is the natural ligand. The physiological role of S1P receptor signaling is studied in several studies. S1P receptor activation is involved in many pathological situations including autoimmunity, inflammation, cardiovascular disorders and cancer [1].

Biological activities can be determined by quantitative structure activity relationship (QSAR) models of biological activities of new or untested chemicals (such as property, reactivity, etc.), from the chemical structures of similar known compounds. The QSAR theory is based on a consideration of the linear total contribution of different structural and chemical properties of a compound to its biological activity. QSAR techniques are classified into two main categories. These include 2D-QSAR with classical Hansch type analysis and 3D-QSAR methods including CoMFA (comparative molecular field analysis) type techniques. In recent years, new QSAR methods have been developed in addition to 2D-QSAR [2,3] and 3D-QSAR methods [4-7]. 4D-QSAR

*Corresponding author: emin@erciyes.edu.tr

[8], 5D-QSAR [9-10], 6D-QSAR [11], 7D-QSAR [12], Hologram QSAR (HQSAR), Inverse QSAR and Binary QSAR are some of the methods resported in recent studies. More detailed information on QSAR is given in previous studies [13-21].

One of the structure-based pharmacophore identification methods used in 3D-QSAR analysis is the electron topological (ET) metatools developed by Dimoglo and coworkers [22]. In this method, the geometric and electronic properties of the molecule obtained from quantum chemical calculations is represented by a matrix called the Electron-Topological Compliance Matrix (ETMC) and for every molecule of a single conformer selected by conformational analysis the pharmacophore is found by calculating the three-dimensional ETMC.

Bersuker and coworkers developed the Electron Conformational Method (ECM) which finds the pharmacophore group and can calculate the quantitative bioactivity [23-25]. In this method developed by Bersuker, despite the presence of the Pha, activity can be reduced (APS, anti-pharmacophore shielding) or increased (AG, auxiliary group) by atomic groups that partially or completely reduce activity outside the Pha and cause steric hindrance. These groups are both APS groups that prevent the Pha's proper interaction with the bioreceptor and AG groups that provide properties such as molecular hydrophobicity.

Using a mixed method with many different methods can reduce mistakes in 4D-QSAR studies, because the purpose of this application is to obtain the best method. In this study, a hybrid 4D-QSAR approach that combines the electron conformational method and the genetic algorithm method was used to identify the pharmacophore (Pha) and to predict the antibacterial activities of 50 propoxy methylphenyl oxadiazol derivatives. It is very important to note that there is not at present any QSAR study about them in the literature. So, the QSAR studies that will be presented in the present work about the aforementioned molecules are important because they are the first molecular modeling studies on these 50 propoxy methylphenyl oxadiazole derivatives

METHOD

Fifty propoxy methylphenyl oxadiazole derivates were discussed in the 4D-QSAR study with the help of the EC-GA method to identify the pharmacophore group and explain the relationship between the biological activities of these molecules and selected molecular parameters [1].

The structures of the studied compounds and their experimental biological activities including $S1P_1$ values obtained from the literature are presented in Table 1-4. In this table, A^{exp} and A^{cal} represent the experimental and calculated biological activities of the compounds, respectively.

In this study, there is one racemic mixture whose activity is 8.075. This molecule has been calculated for both R- and S- conformers. These conformers were calculated as unknown molecules (Numbers 49 and 50). The R-enantiomer of these conformers was calculated and the activity was found to be 7.757. The S-enantiomer of these conformers was calculated and the activity was found to be 9.172. The R-enantiomer is closer to the experimental value. Therefore, the experimental value contribution of the R-enantiomer is greater than that of the R-enantiomer.

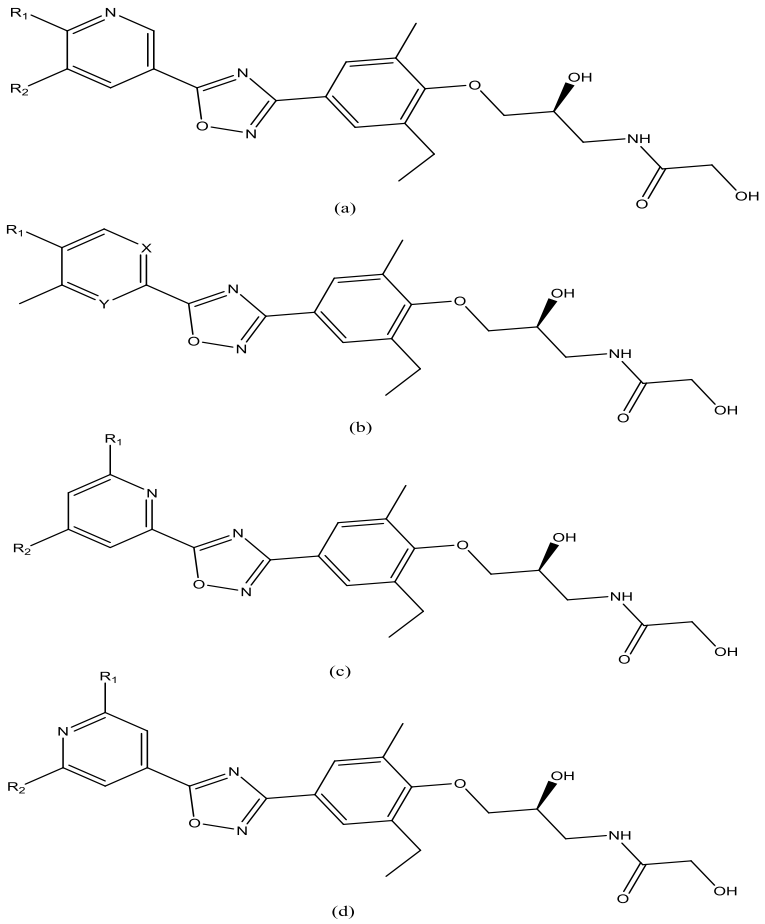

Figure 1. Molecular structures of propoxy methylphenyl oxadiazole acid derivatives.

Table.1 Experimental and calculated activity values of propoxy methylphenyl oxadiazole acid derivatives for (a)

Compound	R_1	R_2	A^{exp}	A^{calc}	Conformer number
1*	-	-	9.154	9.314	24
2	Isobutyl	H	8.244	8.636	39
3[a]	Isobutyl	H	7.050	7.401	18
4	n-propyl	H	8.337	8.435	26
5	Isopropyl	H	8.408	8.437	8
6	Ethyl	H	7.638	8.172	19
7*	Methyl	H	6.605	6.734	23
8	Isobutyl	$-CH_3$	9.000	9.697	43
9[a]*	Isobutyl	$-CH_3$	8.721	8.711	24
10	İzopropil	$-CH_3$	10.000	10.000	36
11*	Pent-3-yl	$-CH_3$	9.045	9.267	19
12	Cyclopentyl	$-CH_3$	10.000	9.452	17
13	Isobutyl	$-CH_2CH_3$	10.000	9.673	21
14	Cyclopentyl	$-CH_2CH_3$	9.698	9.877	9
15	Diethylamino	$-CH_2CH_3$	9.522	9.177	16
16	N-pyrrolidine	$-CH_3$	7.638	7.830	25
17	N-pyrrolidine	$-CH_2CH_3$	8.552	8.301	8

Table 2. Experimental and calculated activity values of propoxy methylphenyl oxadiazole acid derivatives for (b)

Compound	R_1	X	Y	A^{exp}	A^{calc}	Conformer number
18	Isobutyl	CH	N	8.356	8.172	28
19	Cyclopentyl	N	CH	9.698	9.158	17
20	Cyclopentyl	CH	N	9.096	8.697	14

Table 3. Experimental and calculated activity values of propoxy methylphenyl oxadiazole acid derivatives for (c)

Compound	R_1	R_2	A^{exp}	A^{calc}	Conformer number
21*	Isobutyl	Methyl	9.301	9.446	23
22*	Methyl	Isobutyl	8.866	9.572	38
23	Cyclopentyl	Methyl	9.698	9.238	14
24	Methyl	Cyclopentyl	9.045	8.547	55
25	Diethylamino	Methyl	8.886	8.839	22
26	Methyl	Diethylamino	7.267	7.108	16

Table 4. Experimental and calculated activity values of propoxy methylphenyl oxadiazole acid derivatives for (d)

Compound	R_1	R_2	A^{exp}	A^{calc}	Conformer number
27	H	H	5.954	5.907	15
28	Methyl	H	7.065	7.062	43
29	Ethyl	H	8.677	8.379	34
30*	n-propyl	H	9.154	8.489	38
31	Isopropyl	H	9.522	10.023	21
32	n-butyl	H	8.744	8.858	11
33	Isobutyl	H	9.698	10.189	38
34	Pent-3-yl	H	10.000	9.994	7
35	Ethyl	$-CH_3$	9.522	8.962	25
36	n-propyl	$-CH_3$	9.397	9.637	49
37*	Isobutyl	$-CH_3$	10.000	9.897	23
38	Pent-3-yl	$-CH_3$	9.698	10.064	21
39	Cyclobutyl	$-CH_3$	10.000	9.477	21
40	Cyclopentyl	$-CH_3$	9.698	9.722	17
41	Siklohexyl	$-CH_3$	9.000	9.624	15
42	Pent-3-yl	$-CH_2CH_3$	9.698	10.100	10
43	Cyclobutyl	$-CH_2CH_3$	10.000	10.409	27
44*	Cyclopentyl	$-CH_2CH_3$	9.522	10.002	19
45	Cyclohexyl	$-CH_2CH_3$	8.698	9.274	21
46	Ethylamino	$-CH_3$	9.397	9.060	27
47	Diethylamino	$-CH_3$	9.221	9.375	22
48	N-pyrrolidine	$-CH_3$	8.721	8.878	20
49[a]	Isopentyl	$-CH_3$	-	7.757	31
50	Isopentyl	$-CH_3$	-	9.172	15
51[b]	Isopentyl	$-CH_3$	8.075	-	-

[a](R) enantiomer, [b]racemic mixture, the compounds denoted by "*" are test compounds

Optimization under certain circumstances it is possible to choose the best among the possible alternatives in a problem. The genetic algorithm (GA) gives

the closest solutions when working with the appropriate parameters. The GA can be used to solve problems that are difficult or impossible to solve by traditional methods [26-28]. The equation was solved by using the GA and the EC-GA method.

In the context of the EC-GA method, series have been previously worked on different by our group and this method has not been elaborated here since the basis and details are given in the previous literature. [16,18,21]. For the identification of the pharmacophore, AG and APS groups, and in order to reveal the best parameter group with the greatest effect and contribution to the activity, the EC-GA approach is a new hybrid method which was developed by combining the EC method with the GA optimization technique and it has been found to be very suitable for 4D-QSAR model studies.

In the EC-GA method, the genetic algorithm optimization technique is used to determine the best parameters that affect the activity of the drug molecule and to finish the calculations in a shorter time. Using this method, thousands of parameters to define molecules and conformers can be prepared and it is possible to prepare a model which can give good results by selecting the parameters which are most suitable for both the training and the test set from among them. After creating a model using the genetic algorithm optimization technique, in order to accurately predict the validity of this model and the activities of compounds whose experimental activity is not known, this model is tested with a leave-one-out cross validation (LOO-CV) method.

In these studies, the EMRE package program included the discovery of the Pha group of all the conformers of the compounds and the calculation of the activity calculations. Much better results were obtained when the EC-GA method was used to define the Pha group, AG and APS groups and to calculate biological activities. Spartan 10' software optimizations [30] at Hartree-Fock 3-21G level were performed in the water phase, because the most suitable solvent for any biological system is water. Conformation analysis of the compounds was performed by Spartan 10 'software by means of which the structures of the molecules are created. After the conformational search, the conformers with lower energy which are more responsible for the biological activity were kept but the conformers with Boltzmann distribution under 1/10000 were eliminated. After elimination, the Electron Conformational Matrices of Congruity (ECMC) were created using the remaining conformers. ECMCs have Mullikan charges on their diagonal elements [20,21]. Utilizing quantum chemical calculation data, 1266 ECMCs were generated. The illustration of the sample ECMC for the lowest energy conformer of compound 10 is given in Figure 2. As seen in Figure 2, the Mullikan charge of the N1 atom is -0.748 , the bond order between the C4 and C3 atoms is 1.416 and the distance between the C22 and C26 atoms is 4.122.

In the comparison of the ECMC, the lowest energy conformation of the most active compound is used. Compounds with known activity are separated into two groups: active and inactive. Molecules whose activity value is greater than 9.154 are known as active molecules and others are known as inactive molecules. After this step, the Electron Conformational Submatrix of Activity (ECSA) which is the pharmacophore, is described by comparing the ECMC of the pattern compound with all other ECMCs within the given tolerance interval. Many ECSCs obtained from comparison of ECMCs were evaluated using both P_{α} criteria and α_a criteria given in previous studies [13,16,18,19].

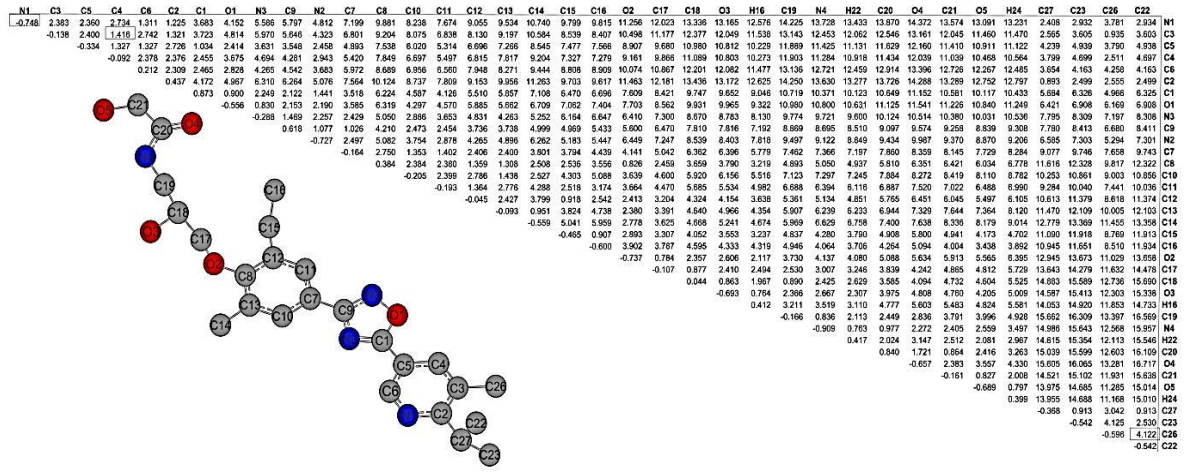

Figure 2. ECMC of the lowest energy conformer for compound 10. Hydrogen atoms bonded to carbon atoms are neglected in the ECMC for clarity.

Many of the methods used in QSAR studies cannot be used for to separate the enantiomers of the molecules from each other. Even though the absolute values of the matrix elements of the two enantiomers are the same, their geometric properties such as angle, distance, and dihedral angle etc. are not the same in these enantiomers. If one of the two enantiomers has a pharmacophore (pha), this does not mean that the other has to have one. Except for the pharmacophore, two important groups are responsible for the activity. The first group is called anti-pharmacophore shielding (APS), which may influence the bioactivity in a decreasing way, the second group is called the auxiliary group (AG), which may influence the bioactivity in an increasing way. Their effects are described by the function S, given below [14,30]:

$$S_{ni} = \sum_{j=1}^{N} \kappa_j a_{ni}^{(j)} \qquad (1)$$

For each conformer, the function S_{ni} is calculated by the sum of the products of the *j*th kind of property in the *i*th conformation of the *n*th compound ($a_{ni}^{(j)}$) and the relative weight of different descriptors (κ_j). N is the number of the selected descriptor.

Based on the function S_{ni}, Bersuker explained the biological activity formula as follows [30]:

$$A_n = A_l \frac{\sum_{i=1}^{m_l} e^{-El_i/RT}}{\sum_{i=1}^{m_n} e^{-E_{ni}/RT}} \frac{\sum_{i=1}^{m_n} \delta_{ni}[Pha] e^{-S_{ni}} e^{-E_{ni}/RT}}{\sum_{i=1}^{m_l} \delta_{li}[Pha] e^{-S_{li}} e^{-E_{li}/RT}} \qquad (2)$$

where δ is a type of Kronecker δ function:

$$\delta ni[pha] = \begin{cases} 0, & Pha\ is\ absent \\ 1, & Pha\ is\ present \end{cases}$$

A_n and A_l, indicate the activity values of the n*th* compound and reference compound, respectively. E_{ni} and E_{li} are the relative energies (in kcal mol^{-1}) of the *i*th conformation for the *n*th compound and reference compound, respectively. R (kcal molK-1) is the gas constant and T (in K) is the temperature. In light of this equation, the variational constant, κj, in Equation (1) was mathematically optimized using the Matlab toolbox function lsqnonlin [31].

In this study, 1172 ECMCs were obtained from the 50 compounds of propoxy methylphenyl oxadiazole derivatives using EMRE software [17]. Eight hundred and four parameters which are include the geometrical,

thermodynamic and topological parameters are prepared making use of Spartan data. Choosing an appropriate subset of data from a large data pool is the most important step of the QSAR model. The GA was used for selecting parameters randomly and generating subgroups for the best parameter selection in the QSAR modeling process. The κ_j values were calculated using the least-squares method taking advantage of the lsqnonlin function in MATLAB [31]. The GA parameters for the optimization were set as follows: number of generations: 100; population size: 100; iteration number: 150; crossover fraction: 80%; mutation rate: 1.5%.

A series of propoxy methylphenyl oxadiazole derivatives containing 50 compounds were divided into a training set consisting of 39 random compounds and a test set consisting of 9 compounds. For the optimal number of parameters, it is necessary to make some calculations about the model's estimated power (q^2) and the number of parameters. First, the compounds were randomly selected, then fixed. Calculation was made from 1 to 8 parameters afterwards. As a result of the calculations made using the MATLAB program for the training and test sets, the theoretical activity (R^2), standard error and q^2 values were obtained for the 1-8 parameters for the training and test sets and are given in Table 6.

PRESS is the sum of the squares of the difference between the observed activity values and the estimated activity values. The cross validation LOO-CV operation in QSAR study is a method that allows you to use the entire data:

$$PRESS_N = \sum_{n=1}^{N} \left| A_n^{exp} - A_n^{calc} \right|^2 \qquad (3)$$

In this equation, N is the total number of training compounds, while A_n^{calc} and A_n^{exp} are the calculated and experimental activity values of the *n*th compound. The quality of the randomly generated EC-GA model was controlled internally based on the training set and externally based on the test set for each of the descriptor subsets. Before checking the model's predictability for an external data set, internal verification was performed to check the conformity of the QSAR model. With external validation, the predicted power of the established model is checked by the test compounds. The stability, quality and prediction capacity of the established model is checked by the numerical value of the cross-validated correlation coefficient (q^2), R^2, q^2_{ext1} and q^2_{ext2} proposed by Schüürman [32], q^2_{ext3} given by Consonni [33] and the concordance correlation coefficient (CCC) proposed by Lin [34,35].

For internal validation of the models, the value of q2 was found by the following formula:

$$q^2 = 1 - \frac{\sum_{n=1}^{N} \left| A_n^{exp} - A_n^{pre} \right|^2}{\sum_{n=1}^{N} \left| A_n^{exp} - \bar{A}_n^{hes} \right|^2} \equiv 1 - \frac{PRESS}{SSY} \qquad (4)$$

$$q^2_{ext1} = 1 - \frac{\sum_{n=1}^{N_{test}} \left| A_{n_{test}}^{exp} - A_{n_{test}}^{pre} \right|^2}{\sum_{n=1}^{N_{test}} \left| A_{n_{test}}^{exp} - \bar{A}_{n_{training}}^{exp} \right|^2} \qquad (5)$$

$$q^2_{ext2} = 1 - \frac{\sum_{n=1}^{N_{test}} \left| A_{n_{test}}^{exp} - A_{n_{test}}^{pre} \right|^2}{\sum_{n=1}^{N_{test}} \left| A_{n_{test}}^{exp} - \bar{A}_{n_{test}}^{pre} \right|^2} \qquad (6)$$

$$q^2_{ext3} = 1 - \frac{\left[\sum_{n=1}^{N_{test}} \left|A^{exp}_{n_{test}} - A^{pre}_{n_{test}}\right|^2\right]/N_{test}}{\left[\sum_{n=1}^{N_{training}} \left|A^{exp}_{n_{training}} - \bar{A}^{exp}_{n_{training}}\right|^2\right]/N_{training}} \qquad (7)$$

where N stands for the count of molecules to be tested. A^{exp}_n and A^{pre}_n are the experimental and the predicted activities of the nth compound in the test set. $\bar{A}^{pre}_{n_{training}}$ and $\bar{A}^{pre}_{n_{test}}$ are the average experimental activity values of the training and test sets, respectively. $A^{exp}_{n_{test}}$ and $A^{pre}_{n_{test}}$ represent the experimental and estimated activities for the nth compound of the test set. N_{test} and $N_{training}$ are the numbers of the test and training molecules. $A^{exp}_{n_{training}}$ denotes the experimental activity of the nth compound in the training set. $\bar{A}^{exp}_{n_{training}}$ is the average of the experimental activities of the training set.

Another external validation, the "concordance correlation coefficient (CCC)" used for the first time by Lin[34,35]. In QSAR model, the CCC value gives information about the accuracy and precision of the model. In this study, we used this equation for the training set, test set and all compounds:

$$CCC = \hat{P}_{training} = \frac{2\sum_{i=1}^{n_{training}}\left(A^{pre}_i - \hat{A}^{pre}\right)\left(A^{exp}_i - \hat{A}^{exp}\right)}{\sum_{i=1}^{n_{training}}\left(A^{pre}_i - \hat{A}^{pre}\right)^2 + \sum_{i=1}^{n_{training}}(A^{exp}_i - \hat{A}^{exp})^2 + n_{training}(\hat{A}^{exp} - \hat{A}^{exp})^2}$$

$$CCC = \hat{P}_{test} = \frac{2\sum_{i=1}^{n_{test}}\left(A^{pre}_i - \hat{A}^{pre}\right)\left(A^{exp}_i - \hat{A}^{exp}\right)}{\sum_{i=1}^{n_{test}}\left(A^{pre}_i - \hat{A}^{pre}\right)^2 + \sum_{i=1}^{n_{test}}(A^{exp}_i - \hat{A}^{exp})^2 + n_{test}(\hat{A}^{pre} - \hat{A}^{exp})^2}$$

$$CCC = \hat{P}_{all} = \frac{2\sum_{i=1}^{n_{all}}\left(A^{pre}_i - \hat{A}^{pre}\right)\left(A^{exp}_i - \hat{A}^{exp}\right)}{\sum_{i=1}^{n_{all}}\left(A^{pre}_i - \hat{A}^{pre}\right)^2 + \sum_{i=1}^{n_{all}}(A^{exp}_i - \hat{A}^{exp})^2 + n_{all}(\hat{A}^{pre} - \hat{A}^{exp})^2}$$

In the QSAR model, the CCC is the numerical value that gives the model its reliability and predictive power. In this way, we could verify the model-data fit of both the training and test sets separately.

RESULTS AND DISCUSSION

The chemical structures, experimental activities (IC50) and conformer numbers of the propoxy methylphenyl oxadiazole derivatives were taken from the literature [1] and are listed in Table 1-4. The geometry optimizations of all the molecular structures were carried out at the HF/3-21g level. Since 51 compounds have a total of 2850 conformation, calculation of the upper basis sets cannot be performed because HF calculations take too long.

The ECMCs for 1339 conformers that used quantum chemical data, were built with the EMRE program. The most active of the propoxy methylphenyl oxadiazole derivatives is compound 10. After this, we set the cut-off activity (IC_{50}) value to 9.154. Twenty four compounds with activity values above this value are classified as highly active compounds and the others as low activity compounds. Following the completion of the creation of the ECMCs, the comparison of ECMCs, which has been described in previous publications, is used to identify the pharmacophore. With the help of this

theory, we obtained the pharmacophore (ECSA) which is containing seven atoms namely O1, N3, C11, C15, C16, C27 and C22 in Figure 3. The pharmacophore atoms are shown in yellow. The highest values of P_α (0.8750) and α_a (0.7800) were obtained from the obtained ECSA.

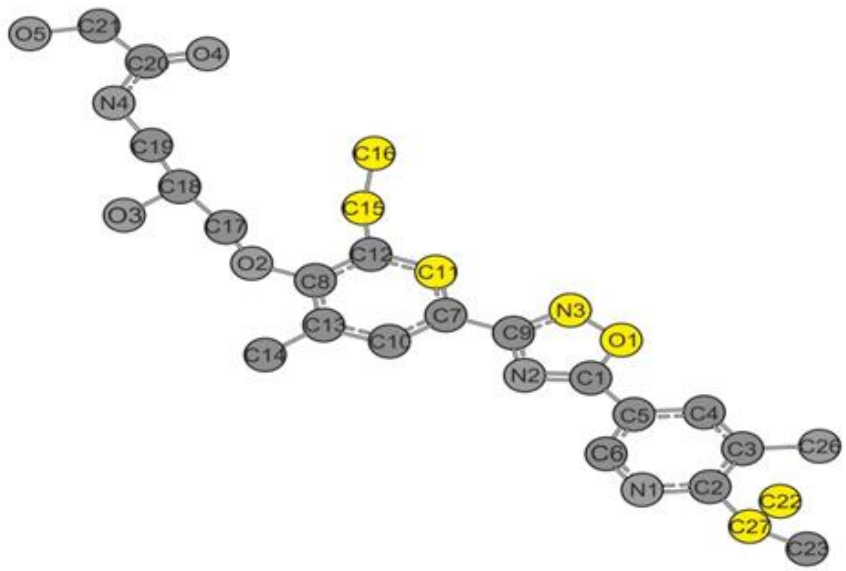

Figure 3. It is shown the pharmacophore (ECSA) which is containing seven atoms namely O1, N3, C11, C15, C16, C27 and C22.

Table 5. (a) ECSA (pharmacophore) of reference compound (compound 10) for (S)-3-(3-ethyl-4-(2-hydroxy-3-(2-hydroxyacetamido) propoxy)-5-methyl phenyl)- 1,2,4 -oxadiazole and its derivatives; (b) tolerance matrix of ECSA for 30 compounds with high activity; (c) Tolerance matrix of ECSA for 18 compounds with low activity (d) tolerance values for all conformers (706). Pharmacophore atoms are shown in yellow

a) ECSA of reference compound (pharmacophore group)

O1	**N3**	**C11**	**C15**	**C16**	**C27**	**C22**	**Pha Atoms**
-0.556	+0.830	+2.153	+7.062	+7.404	+6.421	+6.908	O1
	-0.288	+1.469	+6.164	+6.647	+7.795	+8.308	N3
		+0.618	+4.969	+5.433	+7.780	+8.411	C11
			-0.465	+0.907	+11.090	+11.913	C15
				-0.600	+10.945	+11.934	C16
					0.368	+0.913	C27
						-0.542	C22

b) Tolerance matrix of ECSA for 30 compounds with high activity

O1	**N3**	**C11**	**C15**	**C16**	**C27**	**C22**	**Pha Atoms**
±0.028	±0.020	±0.274	±0.367	±0.425	±1.737	±1.729	O1
	±0.007	±0.767	±0.924	±0.793	±1.465	±1.669	N3
		±0.013	±0.008	±0.294	±1.486	±1.293	C11
			±0.046	±0.003	±1.747	±1.797	C15
				±0.028	±1.648	±1.473	C16
					±0.365	±1.612	C27
						±0.235	C22

c) Tolerance matrix of ECSA for 18 compounds with low activity

O1	N3	C11	C15	C16	C27	C22	Pha Atoms
±0.029	±0.020	±0.274	±0.372	±0.426	±2.223	±1.942	O1
	±0.013	±0.766	±0.932	±0.993	±2.169	±1.969	N3
		±0.013	±0.008	±0.293	±1.599	±2.230	C11
			±0.048	±0.004	±1.465	±2.232	C15
				±0.032	±1.458	±2.373	C16
					±0.804	±1.616	C27
						±0.407	C22

d) Tolerance matrix of ECSA for 706 conformations of 50 compounds

O1	N3	C11	C15	C16	C27	C22	Pha Atoms
±0.030	±0.021	±0.276	±0.386	±0.676	±1.739	±1.945	O1
	±0.011	±0.769	±0.945	±1.336	±1.713	±1.971	N3
		±0.013	±0.019	±0.463	±1.532	±2.232	C11
			±0.050	±0.010	±1.859	±2.233	C15
				±0.033	±1.922	±2.155	C16
					±1.229	±2.116	C27
						±0.440	C22

The ECSA for the tolerance values of propoxy methylphenyl oxadiazole derivatives are given in Table 5 for active and inactive compounds. The ECSA of the minimum energy conformer is demonstrated by the first matrix. The second and third submatrices show the tolerance values of 24 high activity compounds and 24 low activity compounds, respectively. The fourth submatrix for all conformers was obtained without any tolerance constraints applied for 706 conformations of the 50 compounds. In light of the results given above, the tolerance values of the compounds with low activity are higher than those of the compounds with high activity. The first submatrix shows the tolerance values of the pharmacophoric atoms of the lowest energy conformer of the template compound. As can be understood from (b) and (c) of Table 5, the atomic charge tolerances of the O1 atom are ±0.028 and ±0.029 and the tolerances of the distance between the C15 and C16 atoms are ±0.046 and ±0.048 for high and low active compounds, respectively.

The O1, N3, C11, C15, C16, C27 and C22 atoms of the series of propoxy methylphenyl oxadiazole derivatives have an important role in the interaction between the receptor and ligand. Prior to parameter selection and calculation of bioactivity, we created a parameter pool consisting of 804 parameters such as thermodynamic, geometric, electronic, topological and quantum chemistry, which are often used for pharmacophore atoms with the EMRE program for each conformer. In the next step, the data set was randomly divided into two groups: the first group is the training set of 39 compounds which is determined to develop the model, while the second group is the test set of 9 compounds which is created to evaluate the validity of the model. Although all parameters are thought to be effective on biological activity, a very small portion of all parameters is a significant contributor to activity. Other parameters are deleted because other parameters are not related to them. The GA technique is used to eliminate irrelevant parameters and to obtain the optimal parameter set that gives the most appropriate model. The GA technique [36] is used to eliminate irrelevant parameters and to obtain the optimal parameter set that gives the most appropriate model. By running the GA

technique in the MatLab environment [31], the optimal parameter set was obtained. With the help of this process, we calculated the κ_j values in Equation 1 with the GA technique by means of the lsqnonlin function, which is a non-linear least square optimization technique in the MatLab environment [31]. After the κ_j values were calculated, the theoretical activity values were calculated using the obtained values of κ_j. By evaluating the conformance value of each subset of parameters, we completed the best subset of parameters for a given number of parameters giving the optimal EC-GA model.

As is well known, it is important in this study to create the best and most predictive model with the minimum number of parameters. We investigated the relationship between the numbers of parameters for a given model and the estimated power (q^2) in order to define the optimal number of parameters of propoxy methylphenyl oxadiazole derivatives. Using a number between 1-15 parameters for this model, we run the GA. As seen in Figure 4, increasing the number of parameters increases the performance of this model. However, using 8 parameters which seemed to indicate the start of the stabilization point, appears to show that the values remained constant. In light of the result given in Figure 4, the number of compounds should be five times greater than the number of parameters [37]. The optimal number of parameters is 8. Explanation of the 8 parameters selected using the GA and the κ_j values are shown in Table 6. $a^{(5)}$ and $a^{(7)}$ are angle parameters. $a^{(6)}$ and $a^{(8)}$ are orthogonal distance parameters. $a^{(4)}$ is the orthogonal distances plus van der Waals radius parameters. $a^{(3)}$, $a^{(2)}$ and $a^{(1)}$ are other parameters. It is seen in Table 6 that the geometrical and electronic parameters have a crucial role in the biological activity of propoxy methylphenyl oxadiazole derivatives. It is clear that one or more pharmacophore atoms are found in the parameters. Therefore, the presence of the pharmacophore is important for biological activity.

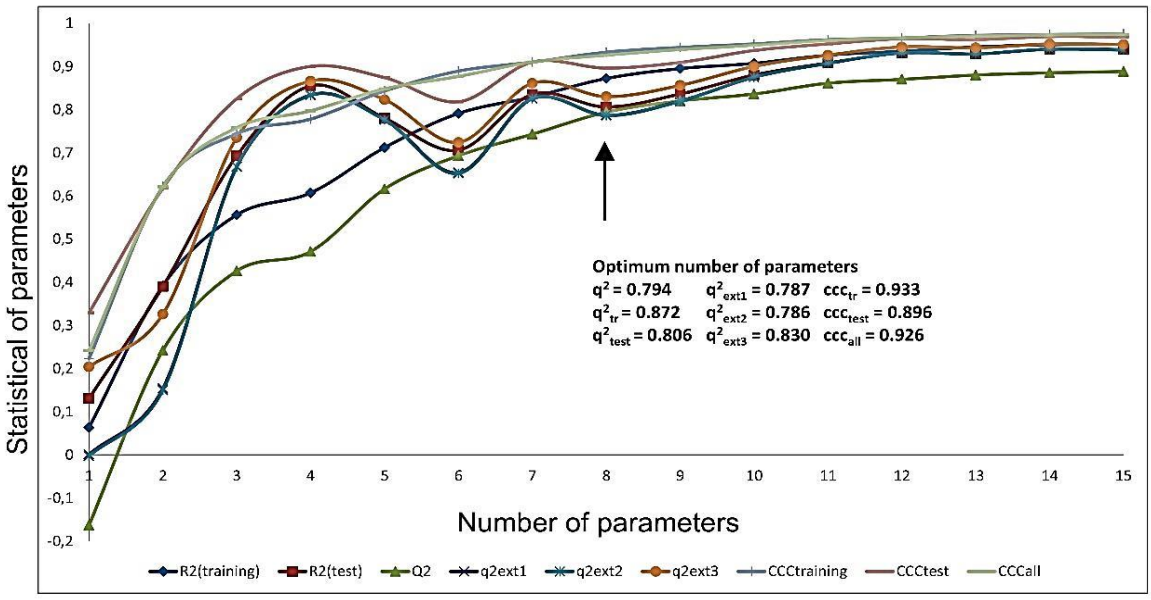

Figure 4. Correlation chart between some statistical parameters and the number of molecular parameters.

To understand that APs and AG have an increasing or decreasing effect on biological activity, we need to look at what is the product of the parameter value and the κj values. The S function given in Equation 2 has an effect on the activity decreasing (APS) if the product of the numerical value of the relevant parameter and the κ_j value is positive and the activity increasing effect (AG) if the product is negative. As can be understood from Table 6, four parameters, namely $a^{(1)}$, $a^{(2)}$, $a^{(3)}$ and $a^{(8)}$, indicate an increasing effect on activity as AG, although parameters $a^{(4)}$, $a^{(5)}$, $a^{(6)}$ and $a^{(7)}$ indicate a reducing effect on activity as APS. In Figure 5, the $a^{(6)}$ and $a^{(8)}$ parameters which show the orthogonal distance of the C2 atom to the N2 – C9 – N3 plane and the orthogonal distance of the C6 atom to the N2 – C9 – N3 plane are presented.

Table 6. Description of optimum 8 parameters chosen by GA and their κj values employed in the calculation of the activity

$a_{ni}^{(j)}$	Name of parameter	Kj value
$a^{(1)}$	Solvent energy	-2.861
$a^{(2)}$	Cpk quality	-0.005
$a^{(3)}$	Mulliken of C8 atom (e^-)	-18.776
$a^{(4)}$	Orthogonal distance of O3 atom to C16 – C15 – C12 plane + Van der Walls radius of H16 atom (Å)	0.076
$a^{(5)}$	C10 C7 C9 N2 dihedral angle	2.246
$a^{(6)}$	Orthogonal distance of C2 atom to N2 – C9 – N3 plane	10.123
$a^{(7)}$	N3 N2 H16 angle	0.256
$a^{(8)}$	Orthogonal distance of C6 atom to N2 – C9 – N3 plane	-10.311

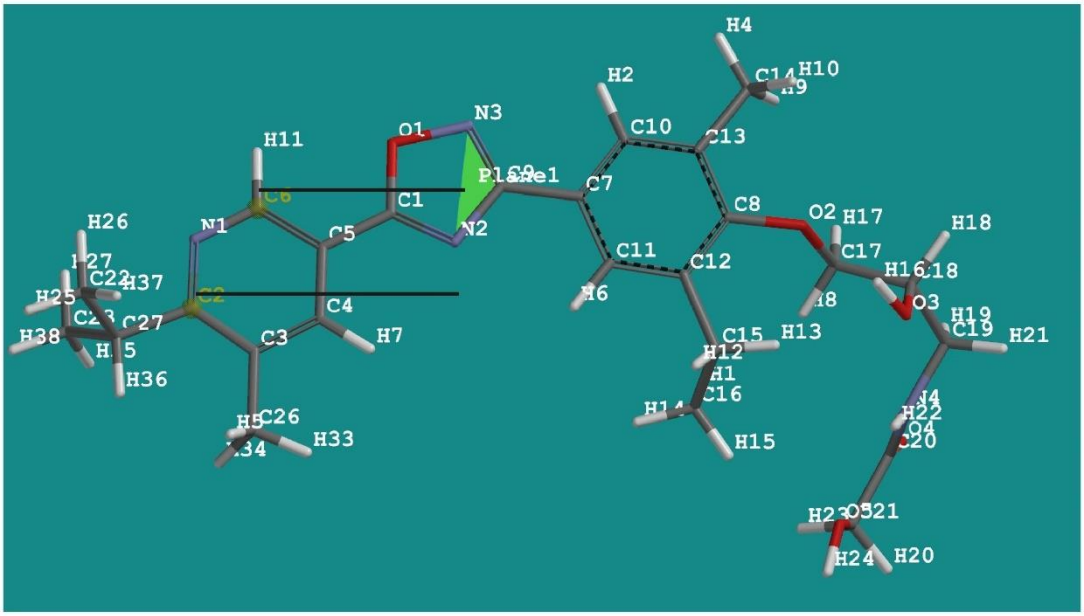

Figure 5. Presentation of the orthogonal distance of C2 atom to N2 – C9 – N3 plane and orthogonal distance of C6 atom to N2 – C9 – N3 plane ($a^{(6)}$ and $a^{(8)}$ parameter) in Table 6.

When all the combined conformers and 8 parameters are considered, the best EC-GA model was created using 39 compounds for the training set and 9 compounds for the test set. The training and test sets were randomly selected from 50 compounds. The experimental efficiency values and the calculated efficiency values are shown in Table 1-4. In Table 1-4, the test compounds are indicated by "a" index. The lsqnonlin function for the training set takes the values of the Kj value that was used to calculate the activity value of the test set. For the training and test sets in which 8 parameters are used, Figure 6 shows the correlation between the experimental activity values and the estimated activity values. The correlation coefficient R^2 indicates the power of the model. For a model to be at the optimum level, it must have a high R^2 and high internal and external q^2 values close to 1. We created the EC-GA model with on R^2 value of 0.872 with a standard deviation of 0.059 for the test set and a R^2 value of 0.836 with a SE of 0.155 for the test set. Internal and external verification values were obtained for both the training and test sets. The EC-GA model had a q^2 value of 0.794 for the training set. The q^2 values of external validation for the test set are q^2_{ext1} = 0.787, q^2_{ext2} = 0.786 and q^2_{ext3} =0.830. Although the numerical values obtained with the parameters used for a QSAR model are satisfactory, the CCC values were calculated to obtain more accurate and better results with external verification for training, testing and other data. The CCC_{tr} (0.933), CCC_{test}, (0.896) and CCC_{all} (0.926) values are close to 1, indicating the sensitivity of the model. We can say that the model obtained is a predictive and robust model when all statistical data are taken into consideration. With this model, it is possible to predict the activities of unknown compounds.

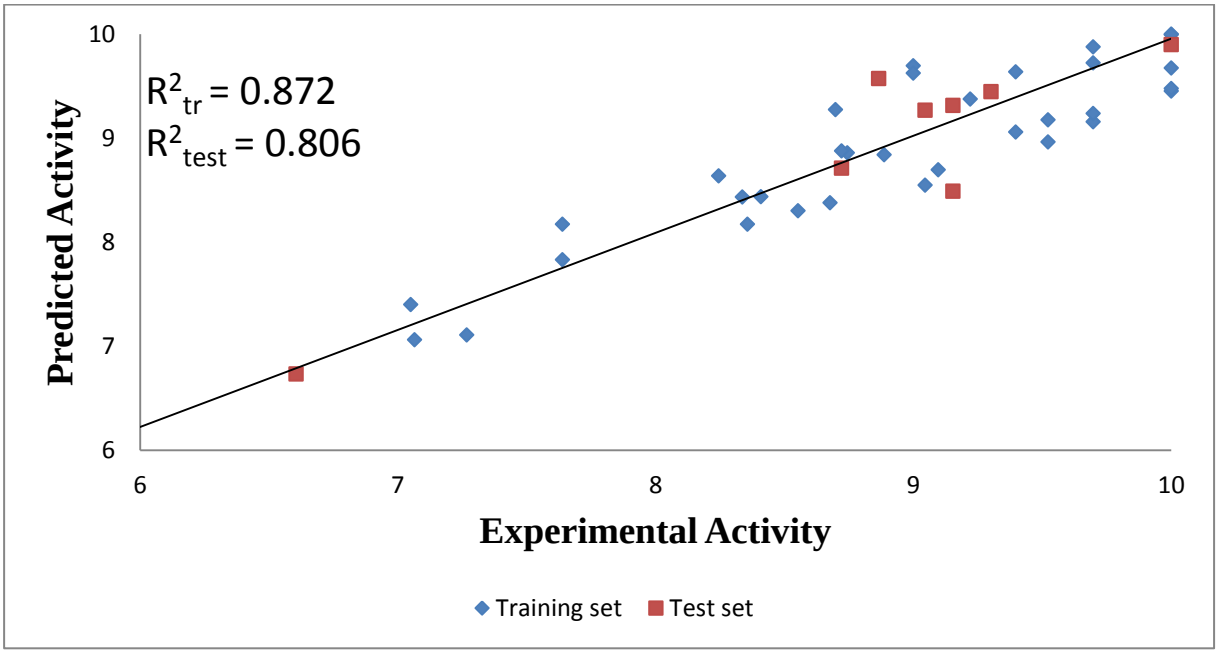

Figure 6. Correlation graph between experimental and predicted activities for training and test sets with 12 parameters.

Table 7 shows the comparison of the results with the use of the minimum energy conformers and all conformers of the propoxy methylphenyl oxadiazole derivatives. As can be understand from Table 7, the q^2= *0.538*, q^2_{ext1} = 0.584, q^2_{ext2} = 0.582 and q^2_{ext3} = 0.669 values indicated a less effective model with lower predictive ability. On the other hand, the CCC_{tr}, CCC_{test} and CCC_{all} values are in good agreement.

The resulting model, which is obtained from more than one conformation, is a function of the contribution at different rates with 8 descriptors. The EC-GA model, in which the best results were obtained, was quantified by the E-statistics of the independent contribution of each of the 8 parameters [13-21]. After each parameter was skipped once, we calculated the E values as reported in the literature [38] and other statistical parameters with the remaining parameters. The E-statistics of 8 parameters are given in Table 8. The magnitude of the numerical value of E indicates that the parameter is important for the model. If the numerical value of E is lower, then that parameter is very important for the model. Within of the 8 parameters, the $a^{(2)}$ parameter with the lowest E value (0.1868) is the most effective parameter. When this parameter (Cpk quality) is eliminated, a remarkable decrease in the R^2_{tr} (from 0.872 to 0.364), R^2_{test} (from 0.836 to 0.507) and especially the q^2 (from 0.794 to -0.097), q^2_{ext1} (from 0.787 to 0.373), q^2_{ext2} (from 0.786 to 0.370) and q^2_{ext3} (from 0.830 to 0.501) values is observed. When we look at the numerical values of E, the three most important parameters are $a^{(2)}$, $a^{(3)}$ *and* $a^{(4)}$. It is seen that the R^2_{tr}, R^2_{test} and q^2 values of these parameters are greatly reduced. The $a^{(8)}$ parameter with the highest E value is the parameter that has the most effect on the model power. A slight decrease in the R^2_{tr} (from 0.872 to 0.829), R^2_{test} (from 0.836 to 0.836) and q^2 (from 0.794 to 0.743) values in the $a^{(8)}$ parameter proves that there is not much change in the numerical value of the model when it is excluded from the model. For the $a^{(7)}$ parameter with the second highest E value (0.7991), a slight decrease in R^2_{tr} and q^2 values is seen. Therefore, this parameter has little effect on activity. As shown in table 8, the parameters used for this model are listed from highest to lowest: $a^{(2)} > a^{(1)} > a^{(3)} > a^{(4)} > a^{(6)} > a^{(5)} > a^{(7)} > a^{(8)}$. As can be understand from this model, $a^{(2)}$ (Cpk quality) is the most important parameter for the bioactivity of the propoxy methylphenyl oxadiazole derivatives.

CONCLUSION

In this study, a model obtained from 50 propoxy methylphenyl oxadiazole derivatives was developed for the prediction of the activity and identification of pharmacophores for the treatment of autoimmune diseases of the central nervous system (CNS) by the 4D-QSAR EC-GA method. The effect of the conformational ensemble of the compounds related with Boltzmann distribution was included in all stages of the study. Pharmacophores are formed from the composition of the O1, N3, C11, C15, C16, C27 and C22 atoms in this model.

Table 7. Statistical parameters of the 8-parameter EC-GA model obtained by using both minimum energy conformer and multiple conformers

	R_{Tr}	R_{Test}	q^2	q^2_{ext1}	q^2_{ext2}	q^2_{ext3}	CCC_{tr}	CCC_{test}	CCC_{all}
Single conformer	0.711	0.713	0.538	0.584	0.582	0.669	0.840	0.822	0.836
Multiple conformers	0.872	0.836	0.794	0.787	0.786	0.830	0.933	0.896	0.926

Table 8. in the propoxy methylphenyl oxadiazole derivatives the effect of model performance of each of 8 parameters show E, $R^2_{training}$, R^2_{test}, q^2, q^2_{ext1}, q^2_{ext2}

Parameters	E	R_{Tr}	R_{Test}	q^2	q^2_{ext1}	q^2_{ext2}	q^2_{ext3}	CCC_{tr}	CCC_{test}	CCC_{all}
$a^{(1)}$	0.1871	0.525	0.388	-0.096	-0.256	-0.262	-0.001	0.696	0.542	0.666
$a^{(2)}$	0.1868	0.364	0.507	-0.097	0.373	0.370	0.501	0.587	0.619	0.592
$a^{(3)}$	0.3463	0.643	0.344	0.407	-0.424	-0.431	-0.135	0.800	0.529	0.735
$a^{(4)}$	0.5684	0.769	0.544	0.639	0.185	0.182	0.351	0.874	0.682	0.835
$a^{(5)}$	0.7430	0.817	0.853	0.735	0.840	0.839	0.873	0.902	0.921	0.906
$a^{(6)}$	0.5731	0.768	0.892	0.642	0.891	0.891	0.913	0.875	0.942	0.886
$a^{(7)}$	0.7991	0.836	0.681	0.743	0.615	0.613	0.693	0.914	0.811	0.896
$a^{(8)}$	0.8001	0.829	0.835	0.743	0.826	0.825	0.861	0.910	0.910	0.910

The QSAR model obtained with LOO cross-validated R^2 and q^2 values showed high internal and external validation of activity and proved robustness by dividing the data set into training and test sets. By using the GA method, the geometrical and electronic parameters were obtained by selecting variables. With the obtained 4D-QSAR EC-GA model and the internal and external validity, the agreement values between the experimental and predicted activities are over 0.750. The prediction power shown for both the training and test sets by the q^2, q^2_{ext1} and q^2_{ext2} values was greater than 0.750.

In consideration of previous explanations, the results of the model show that the QSAR model of propoxy methylphenyl oxadiazole derivatives using the EC-GA model is a promising tool for the future design of new propoxy methylphenyl oxadiazole derivatives as receptors for the treatment of autoimmune diseases of the central nervous system (CNS).

ACKNOWLEDGMENTS

This study was supported by the Scientific Research Projects Chairmanship (BAP) of Cumhuriyet University (Project Number: CÜBAP F-427).

The authors also report no conflicts of interest.

REFERENCES

[1] M.H. Bolli, S. Abele, M. Birker, R. Bravo, R. Bur and et al., J. of Medicinal Chemistry 57 (2013) 110-130.

[2] S.J. Free, J. Wilson, J. Med. Chem. 7 (1964) 395–399.

[3] C. Hansch, T. Fujita, J. Am. Chem. Soc. 86 (1964) 1616–1626.

[4] R. Cramer, D. Patterson, J. Bunce, J. Am. Chem. Soc. 110 (1988) 5959–5967.

[5] G. Klebe, Springer Science & Business Media. 3 (1998) 87–104.

[6] A. Özalp, S.Ç. Yavuz, N. Sabancı, F. Çopur, Z. Kökbudak, E. Sarıpınar, SAR and QSAR in Environmental Research. 27(4) (2016) 317-42

[7] H. Kubinyi, Drug discovery today. 2(11) (1997) 457-467

[8] A.J. Hopfinger, S. Wang, J.S. Tokarski, B. Jin, M. Albuquerque, P.J. Madhav, C. Duraiswami, J. Am. Chem. Soc. 119 (1997) 10509-10524

[9] C. Oberdorf, T.J. Schmidt, B. Wünsch, European Journal of Medicinal Chemistry. 45 (2010) 3116-3124

[10] A. Vedani, M. Dobler, J. Med. Chem. 45 (2002) 2139–2149.

[11] A. Vedani, M. Dobler, M.A. Lill, J. Med. Chem. 48 (2005) 3700-3703

[12] J. Polanski, Curr. Med. Chem., 16 (2009) 3243–3257.

[13] E. Sarıpınar, N. Geçen, K. Şahin, E. Yanmaz, Eur. J. Med. Chem. 45 (2010) 4157-4168
[14] E. Yanmaz, E. Sarıpınar, K. Şahin, N. Geçen, F. Çopur, Bioorg. Med. Chem. 19 (2011) 2199-2210
[15] L. Akyüz, E. Sarıpınar, E. Kaya, E. Yanmaz, SAR and QSAR in Environmental Research, 23 (2012) 409-433
[16] K. Şahin, E. Sarıpınar, E. Yanmaz, N. Geçen, SAR and QSAR in Environmental Research. 22 (2011) 217-238
[17] S.Ç. Yavuz, N. Sabancı, E. Sarıpınar, Curr. Comput. Aided Drug Des. doi: 10.2174/1573409913666170529103206
[18] A. Özalp, S.Ç. Yavuz, N. Sabancı, F. Çopur, Z. Kökbudak, E. Sarıpınar, SAR and QSAR in Environmental Research. 2016; 27(4): 317-42
[19] N. Geçen, E. Sarıpınar, E. Yanmaz, K. Şahin, J. Mol. Model., 18 (2012) 65-82
[20] E. Sarıpınar, Y. Güzel, S. Patat, I. Yıldırım , Y. Akçamur, A.S. Dimoglo, Arznemttel-Forschung/ Drug Research. 1996; 46: 824-28
[21] Y. Güzel, E. Sarıpınar, İ. Yıldırım, J. Mol. Struct. (Theochem). 418 (1997) 83-91
[22] A.S. Dimoglo, N.M. Shvets, I.V. Tetko, D.J. Livingstone, Quant. Struct.-Act. Relat. 20 (2001) 31-45
[23] I.B. Bersuker, Current Pharmaceutical Design. 9 (2003) 1575-1606
[24] A.V. Marenich, P. Yong, I.B. Bersuker, J.E. Boggs, J. Chem. Inf. Model. 48 (2008) 556-568
[25] P. Yong, A.V. Marenich, I.B. Bersuker, J.E. Boggs, Abstracts of Papers of The American Chemical Society. 18(2) (2004) 81-89
[26] N. Sukumar, G. Prabhu, P. Saha, In Applications of Metaheuristics in Process Engineering. 1 (2014) 315-324
[27] J. Devillers, Genetic Algorithms in Molecular Modeling, Academic Press Inc, San Diego, (1996) 327
[28] I.B. Bersuker, S. Bahçeci, J.E. Boggs, J. Chem. Inf. Comput. Sci. 40 (2000) 1363–1376
[29] Spartan'10; Wavefunction, Inc. Irvine, CA, 2011.
[30] [30] I.B. Bersuker, J. Comput. Aided. Mol. Des. 22 (2008) 423–430.
[31] MATLAB (ver 7.0), The MathWorks Inc, 3 Apple Hill Drive, Natick, MA 01760-2098.
[32] G. Schüürmann, R.U. Ebert, J. Chen, B. Wang, R. Kuhne, J. Chem. Inf. Model. 48 (2008) 2140-2145
[33] V. Consonni, D. Ballabio, R. Todeschini, J. Chem. Inf. Model. 49 (2009) 1669-1678
[34] L.I. Lin, Biometrics. 45 (1989) 255-268.
[35] L.I. Lin, Biometrics. 48 (1992) 599-604
[36] J. Devillers, Academic Press, Lyon, 1996
[37] J.G. Topliss, J. Med. Chem. 22 (1979) 1238–1244
[38] D. Livingstone, Data Analysis for Chemists; Oxford University: Oxford, 1995