International Journal of Chemistry and Technology

Research Article

4D-QSAR study of flavonoid derivatives with MCET method

Burçin TÜRKMENOĞLU*, Hayriye YILMAZ, Ekrem Mesut SU, Tuğba ALP TOKAT, Yahya GÜZEL

Department of Chemistry, Faculty of Science, Erciyes University, 38039, Kayseri, Turkey

*Corresponding author's e-mail address: bkilic@erciyes.edu.tr (B. Türkmenoğlu)

ABSTRACT

The Molecular Conformer Electron Topological (MCET) method was performed for the identification of the pharmacophore (Pha) group and predicting inhibitory activity of 42 flavonoid ligands on gamma-aminobutyric acid/benzodiazepine receptor complex (GABAA/BZR). In this method, Electron Topological Matrix (ETM) was used to visualize 3D structural descriptors. Multiple comparisons of ETM matrices for all flavonoid compounds allow us to define Pha-structure. Genetic algorithm (GA)- Partial Least-Squares (PLS) methods were performed to construct QSAR model and to select most important descriptors of the training set (32 compounds) and test set (10 compounds). The GA-PLS based model showed good results, $q^2 = 0.808$ and $r^2_{test} = 0.775$ with high internal and external validation. The developed model can help to understand the inhibitory mechanism.

Keywords: Flavonoid derivatives, GABA$_A$, 4D-QSAR, MCET.

MCET Method ile Flavonoid Türevlerinin 4D-QSAR İncelemesi

ÖZ

Moleküler Konformer Elektron Topolojik (MCET) metodu, gamma-aminobütirik asit/benzodiazepin reseptör kompleksi (GABAA / BZR) üzerindeki 42 flavonoid ligandın farmakofor grubunun tanımlanması ve önleyici aktivitesinin öngörülmesi için gerçekleştirildi. Bu metotta, 3D yapısal tanımlayıcıları görselleştirmek için Elektron Topolojik Matris (ETM) kullanılmıştır. Tüm flavonoid bileşikler için ETM matrislerinin çoklu karşılaştırmaları, Pharmacophore (Pha) yapısını tanımlamamızı sağlar. QSAR modelini oluşturmak ve eğitim setinin (32 bileşik) ve test setinin (10 bileşik) en önemli tanımlayıcılarını seçmek için Genetik Algoritma (GA)-Kısmi En Küçük Kareler (PLS) yöntemleri gerçekleştirildi. GA-PLS'ye dayalı model, yüksek iç ve dış doğrulama ile $q^2 = 0.808$ ve $r^2_{test} = 0.775$ iyi sonuçlar verdi. Geliştirilen model inhibitör mekanizmasının anlaşılmasına yardımcı olabilir.

Anahtar Kelimeler: Flavonoid türevleri, GABA$_A$, 4D-QSAR, MCET.

---

1. INTRODUCTION

Flavonoids are widespread in nature, mainly in green plants[1], and exert a protective effect against both UV light and microbial invasion by pathogens in plants.[2,3] Flavonoids[4], which were first isolated from herbal plants and used as tranquilizers in folk medicine, have been shown to possess a selective and relatively mild affinity for the benzodiazepine binding site of γ-amino butyric acid type A receptors (GABA$_A$Rs/BZRs).[5-7] This new family of natural products, along with various synthetic derivatives[8], has an extremely potent anxiolytic effect which is not associated with myorelaxant, amnestic, or sedative actions.[9] Inhibition in the adult mammalian central nervous system (CNS) is mediated by GABA. The fast-inhibitory actions of GABA are mediated by GABA$_A$Rs, which mediate both phasic and tonic inhibition in the brain.[10]

QSAR of flavonoids for the inhibition of cAMP phosphodiesterase has been determined[11] and new inhibitors of xanthine oxidase have also been developed using a rational design approach.[12] 3-dimensional quantitative structure–activity relationship (3D-QSAR) has been applied to explore the structural requisites of flavone derivatives.[13,14] The methods and software used for 4D-QSAR model establishment and analysis (including descriptor calculation and selection, partial least-squares–PLS- analysis, and related software) have been described in our previous publication with MCET.[15,16] Multiple complementary applications of 4D-QSAR paradigm[17] may be a good way to extend our knowledge and understanding of the SARs of flavonoids using this 'quality for quantity' argument. The fourth 'dimension' of the 4D-QSAR paradigm is ensemble sampling of the spatial features of the members of the training set.[17] This sampling process in turn enables the construction of optimized dynamic spatial 4D-QSAR

models, in the form of 3D bioactive structure dependent on the structure and descriptors of the Pha.[18]

Complementary to building 4D-QSAR models that embed 3D bioactive structure is the construction of high-throughput 4D fingerprint models for virtual screening. The 4D-QSAR paradigm has been successfully applied to a variety of chemical classes and biological interaction points.[19] Our study was focused on finding a 4D-QSAR model that able to predict the anti-benzodiazepine receptor activity of 42 flavonoid derivatives and even provides clues for mechanism of drug receptor interaction.

2. EXPERIMENTAL

2.1. Data Set

A series of flavonoids derivatives (42 compounds) with experimental biological activities was taken from the literature.[20] The sum of set molecules was randomly divided a training set (32 compounds) and test set (10 compounds) for confirming 4D-QSAR model. The activities of studied compounds are shown in Table 1.

2.2. Quantum Chemical Calculations

The first step in developing a QSAR model is a numerical representation of the molecular descriptors for molecules. The minimum energy conformations of the molecular geometries of fullerenes were optimized with 3-21 G* Hartree-Fock method by SPARTAN ′08 (Wavefunction Inc., Irvine, CA, 2000).[21,22] The ETM was constructed using the electronic and geometric features obtained from quantum-chemical calculations which information related to both the topological environment and the electronic character of each atom in a molecule.[23] The quantum chemical calculations were completed in water; conformers with less than 2 kcal mol^{-1} of relative energy were selected and saved as MolFiles. The MolFiles databases were transformed to Electron Topological Matrix (ETM) format using ETM Programmer (ETMP).[16]

2.3. ETM and Conformer Selection

The molecular structures under investigation were represented with many conformers. The number of all total calculated conformers for each molecule was reduced to one hundred. The most active conformers were selected, and the remaining conformers and the duplicate conformers were eliminated. If each corresponding distance between the atoms in the two conformers was less than the threshold (0.4 Å), then the higher energy conformation was rejected.[23] In most populated conformers those which provided an almost perfect fit were selected. The idea underlying 4D-QSAR analysis was related to differences in the Boltzmann

average spatial distribution of conformers with respect to 3D pharmacophores.[24] The resulting activity is averaged over all the selected conformers of the molecule. Considering the Boltzmann population of each conformer, we obtained a mean value which contains the total of terms for all selected conformations as follows:

$$P^i \left[\% \right] = \left[\exp - \left(\Delta H^i / RT \right) \right] * 100 \qquad (1)$$

Where P^i is the probability of a conformer, ΔH^i is the relative heat of the formation of the ith conformer with respect to the lowest energy conformer (J mol^{-1}), R is the gas constant (8.314 J mol^{-1} K^{-1}), T is the temperature in Kelvin, and N is the total number of the selected conformers. ETM is 3D structural descriptors derived from 3D structure of the molecule and are electronic in nature.[23] A detailed description of ETM has been reported in previous publications.[25-30] This contains all the information about the possible action of the molecule. Theoretically, nothing is better than a full electronic structure and topology, which represent molecular ability to interact with other systems.[31]

2.4. Identification of Pha and QSAR Model Generation

In the software, firstly, the Pha structure was extracted by comparing the ETMs of all the conformers with the template conformer, and then the descriptor set in the detailed positions was automatically created and visualized. It is possible to distinguish between the repulsive and attractive effects of the descriptors such as auxiliary groups (AG) and anti-pharmacophore shielding groups (APS). The Pha structure, and subsequently the AG and APS groups should be used for QSAR modeling. The lowest energy conformer of the lead molecule was selected as the template conformer, the starting-structure for the generation of bioactive structures. The MCET method was applied for the detection and interpretation of crucial interaction patterns. The advanced bioactive structure, consisting of the Pha, AG and APS, was defined from the generated models. A bioactive structure is a 3D description developed by specifying the distance (or bond length) and amount of electronic values in the ETM. It may be generated from the superposition of the active molecules by means of their common features. Given a set of active molecules, the 3D approximate model generation of the bioactive structure involves three steps: (1) comparing and matching the molecules to identify the key pharmacophore, (2) aligning three-ordered atoms (based on Pha structure) to superimpose the remaining atoms, and (3) analyzing the various positions to define an AG and APS set, the independent variables.

In specific cases, the analysis of the group of active molecules shows that the Pha does not influence the activity quantity, and is a constant (A$_o$). The atoms

forming the Pha are defined as a basic skeleton. In the presence of the Pha the activity of the molecule may be diminished (partially or completely) by APS which hinders its proper docking with the receptor, or it may be enhanced by AG which provides for other properties (e.g. hydrophobicity or electrostatic attraction between L-R). To determine the AG and APS descriptors, we had to examine the structures (conformations) of the superimposed molecules. Their influence could be used to parameterize the receptor points. This parameterization at the receptor was based on AG and APS described by the electronic and geometric values known from the ETM. We suggested a general scheme for quantitative evaluation to estimate their approximate role in the activity. The main idea (somewhat similar to that involved in Pha identification) was to describe each of them by means of structural and electronic parameters and to reveal their role using a minimization procedure, as is usually done in QSAR problems. Then by processing these descriptors for the training set in comparison with the activities and performing PLS, we obtained the adjustable constants that represent each of these parameters in the activity.

In general, there is not a priority division of the AG and APS in "shielding" and "enhancing" L-R binding, and only the final optimization of the parameter values shows the different kinds of AG and APS contribution. First, we took into account that their contribution reduced or enhanced L-R binding by an amount E which reduced (or increased) the activity by a factor of $\exp(-E/kT)$. We denoted $S_{ni} = E_{ni}/kT$ and introduced the function S in a general way as follows:

$$S_{ni} = \sum_{i=1}^{N} K_j * a_{ni}^j \tag{2}$$

Where a_{ni}^j are the independent variables that describe the jth kind of AG or APS in the ith conformation of nth molecule, N is the number of expected interactions in spite of the different position of scaffold, and κ_j indicates variational parameters (adjustable constant) relative to the interaction point of the receptor. The interaction points were selected from positions of the lowest energy conformers of the most and least active molecule. L-R cross-term descriptors in the positions were also considered, which could be helpful in the identification of AG and APS interactions in determining variation of activity. The interaction between L-R was defined by multiplying the descriptors capturing the properties of the ligand with the descriptors of the receptor and was given by Eq. (2). The magnitude of the individual weighting coefficients within the parameters indicates the relative importance of AG and APS in each position when determining activity. Thus, specific regions in the three-dimensional space, where AG and APS interactions were important, could be identified by superposition, and plotted to derive the pharmacophoric receptor maps. Using this function and taking into account the Boltzmann population of each conformation as a function of its energy and temperature T, we obtained the following general formula of activity: A_0 was a constant

(see below), and for the nth molecule, m_n and m_n^{Pha} were the numbers of the selected conformations and the conformations possessing a Pha (Table 1), respectively.

$$A_n = A_0 \frac{\sum_{i=1}^{m_n^{Pha}} e^{-S_{ni}} e^{-E_{ni}/kT}}{\sum_{i=1}^{m_n} e^{-E_{ni}/kT}} \tag{3}$$

In this formula, the S_{ni} value for conformations that have a Pha contributed to the activity, and these contributions were weighted in accordance with the relative numbers of conformations in the active molecules. These numbers decreased rapidly with the energy of the conformation E_{ni} (at $E_{ni} \sim 2$ kcal mol^{-1} the number of conformations became lower than a 0.02 part of those in the lowest conformation at $E_{n0} = 0$). In the next section we describe how we handled the multi-conformation problem. To determine the A_0 constant, we chose a reference molecule (l) from the training set for which the activity was known and calculated A_l after Eq. (3) by determining A_0 from this equation and substituting it in Eq. (3), we obtained:

$$A_\ell = A_0 \frac{\sum_{i=1}^{m_\ell^{Pha}} e^{-S_{\ell i}} e^{-E_{\ell i}/kT}}{\sum_{i=1}^{m_\ell} e^{-E_{\ell i}/kT}} \tag{4}$$

$$A_n = A_\ell \frac{\sum_{i=1}^{m_l} e^{-E_{\ell i}/kT} \sum_{i=1}^{m_n^{Pha}} e^{-S_{ni}} e^{-E_{ni}/kT}}{\sum_{i=1}^{m_n} e^{-E_{ni}/kT} \sum_{i=1}^{m_\ell^{Pha}} e^{-S_{\ell i}} e^{-E_{\ell i}/kT}} \tag{5}$$

Using the experimental data for the activities of the molecules in the training set, we estimated the adjustable constants κ_j in Eq. (2) by performing a least-squares minimization operation on the function $|A_n^{pred} - A_n^{exp}|^2$. With the constants κ_j determined in this way, we could evaluate the expected activity of any molecular system using Eq. (5). In this formula, only the choice of the $a_{ni}^{(j)}$ independent variable remained uncertain. It required some experience and skill. Bersuker and co-workers have shown how to handle the multiconformation problem successfully by using electron-conformational (EC) method.[32-34] In comparison with the EC method, the MCET method automatically took into account a set of 3D structural descriptors for all compounds in the training set. This had a significant role for a method that considers the L-R binding, consisting of a large number of positions of the different conformers. By

simultaneously solving the rather complex problem of NLME, we demonstrated that the κ_j constant parameters for all the compounds in the training set could be defined. To do this we have presented an improved MCET method and shown its application to the problem of the binding affinities of flavonoids. As we have previously how to employ the MCET method in an earlier paper, we have not repeated it again here.[16]

We applied the variable selection method based on multi-objective GA to the flavonoids data and constructed the nonlinear QSAR model using Eq. (5) of NLME. The best QSAR model was selected according to the correlation coefficient R^2 and the Leave-One-Out Cross Validation (LOO-CV) correlation coefficient Q^2. The obtained results were accurate and interpretable.

3. RESULTS AND DISCUSSION

The chemical structures of 42 flavonoids given in Table 1 were chosen from the literature.[20] In the MCET method, all the conformers' ETMs were compared with that of the template conformer to define Pha structure firstly, then the molecular activities were computed through 3D structural descriptors, which were analyzed using PLS regression in combination with GA. Both ETMP for transforming the data and MCET for computing the activity were written by us in C#. Electron

Topological Sub-Matrix (ETSM) representing the Pha structure were given in bold letters in Figure 1.

Using cluster analysis, the molecules were first divided into two subsets: one training set composed of 32 molecules, and one external test set (marked as *) composed of 10 molecules (N05, 06, 07, 10, 20, 22, 32, 34, 36, 38). The test set molecules were not included in 4D-QSAR model development but rather employed to analyze predictive performance. The size of the test set comprised about 22% of the whole set, ensuring that the test set contained representative samples of the training group.

Pha structures were created by combining four to six atoms in the template conformer, which was the lowest energy conformer of the N01 reference molecule. In the Pha hypotheses scoring process, each Pha and its associated bioactive structure were treated temporarily as a reference in order to assign a score, and the hypotheses were ranked according to the following scores: 1) the alignment score of Pha atoms, 2) the superimposed score of oriented atoms related to Pha 3) a superposition score consisting of Pha, AG and APS. After analysis of the alignment Pha structure in the training set, the hypothesis was generated, and then the best hypothesis consisting of 3D structural descriptors determined by the relative arrangement according to Pha structure was selected for further research.
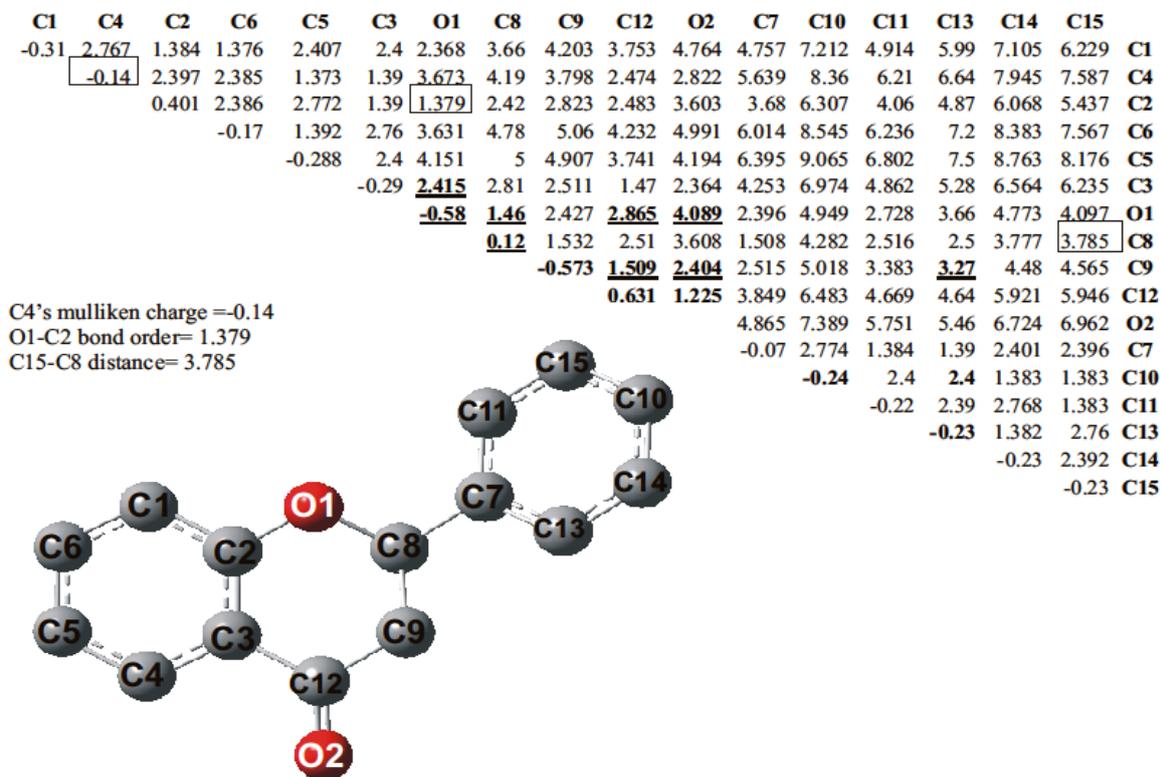
C1	C4	C2	C6	C5	C3	O1	C8	C9	C12	O2	C7	C10	C11	C13	C14	C15	
-0.31	2.767	1.384	1.376	2.407	2.4	2.368	3.66	4.203	3.753	4.764	4.757	7.212	4.914	5.99	7.105	6.229	C1
	-0.14	2.397	2.385	1.373	1.39	3.673	4.19	3.798	2.474	2.822	5.639	8.36	6.21	6.64	7.945	7.587	C4
		0.401	2.386	2.772	1.39	1.379	2.42	2.823	2.483	3.603	3.68	6.307	4.06	4.87	6.068	5.437	C2
			-0.17	1.392	2.76	3.631	4.78	5.06	4.232	4.991	6.014	8.545	6.236	7.2	8.383	7.567	C6
				-0.288	2.4	4.151	5	4.907	3.741	4.194	6.395	9.065	6.802	7.5	8.763	8.176	C5
					-0.29	2.415	2.81	2.511	1.47	2.364	4.253	6.974	4.862	5.28	6.564	6.235	C3
						-0.58	1.46	2.427	2.865	4.089	2.396	4.949	2.728	3.66	4.773	4.097	O1
							0.12	1.532	2.51	3.608	1.508	4.282	2.516	2.5	3.777	3.785	C8
								-0.573	1.509	2.404	2.515	5.018	3.383	3.27	4.48	4.565	C9
									0.631	1.225	3.849	6.483	4.669	4.64	5.921	5.946	C12
										4.865	7.389	5.751	5.46	6.724	6.962	O2
											-0.07	2.774	1.384	1.39	2.401	2.396	C7
												-0.24	2.4	2.4	1.383	1.383	C10
													-0.22	2.39	2.768	1.383	C11
														-0.23	1.382	2.76	C13
															-0.23	2.392	C14
																-0.23	C15

C4's mulliken charge = -0.14
O1-C2 bond order = 1.379
C15-C8 distance = 3.785

Figure 1. Reference molecule's (N01) ETM and Pha's ETSM.

Table 1. Chemical structures and observed and predicted -log Ki values by MCET method

	Compounds							-log K_i		
No	R5	R6	R7	R8	R2'	R3'	R4'	Obsd.	Pred.	Residual
1	H	H	H	H	H	H	H	6.00	6.00	0
2	H	F	H	H	H	OH	H	5.60	5.90	-0.30
3	H	Cl	H	H	H	OH	H	6.07	6.29	-0.22
4	H	Br	H	H	H	OH	H	6.22	6.29	-0.70
5	H	F	H	H	H	NO_2	H	6.74	6.61	0.13
6	H	Cl	H	H	H	NO_2	H	8.10	7.10	1.00
7	H	Cl	H	H	H	H	OCH_3	5.90	5.80	0.10
8	H	Br	H	H	H	H	OCH_3	5.68	5.76	-0.08
9	H	Br	H	H	NO_2	H	H	6.68	6.51	0.17
10	H	NO_2	H	H	H	H	Br	7.60	6.80	0.80
11	H	Cl	H	H	F	H	H	6.38	5.82	0.56
12	H	Br	H	H	F	H	H	6.42	5.82	0.60
13	H	H	H	H	F	H	H	5.45	6.17	-0.72
14	H	F	H	H	H	F	H	6.04	6.18	-0.14
15	H	Cl	H	H	H	F	H	6.93	6.50	0.43
16	H	Br	H	H	H	F	H	7.38	6.50	0.88
17	H	H	H	H	H	H	F	5.44	5.70	-0.26
18	H	F	H	H	H	H	F	5.60	5.70	-0.10
19	H	Cl	H	H	H	H	F	6.74	6.45	0.29
20	H	Br	H	H	H	H	F	6.94	6.45	0.49
21	H	H	H	H	H	Cl	H	6.21	6.30	-0.09
22	H	F	H	H	H	Cl	H	6.70	6.31	0.39
23	H	Cl	H	H	H	Cl	H	7.64	6.73	0.91
24	H	Br	H	H	H	Cl	H	7.77	7.73	0.04
25	H	H	H	H	H	Br	H	6.38	6.31	0.07
26	H	F	H	H	H	Br	H	6.63	6.32	0.31
27	H	Cl	H	H	H	Br	H	7.64	6.73	0.91
28	H	Br	H	H	H	Br	H	7.72	6.73	0.99
29	H	Br	H	H	H	F	H	7.72	6.73	0.99
30	H	Br	H	H	H	H	NO_2	7.15	6.73	0.42
31	H	NO_2	H	H	H	NO_2	H	6.70	6.70	0.00
32	H	Br	H	H	H	NO_2	H	7.92	7.08	0.84
33	OH	Br	OH	Br	H	H	H	9.00	9.10	0.10
34	OH	H	OH	H	H	H	H	6.15	6.19	-0.04
35	OH	H	OH	H	H	H	OH	5.52	5.98	-0.46
36	OH	H	OH	H	Cl	H	H	5.52	5.58	-0.06
37	OH	H	OH	H	F	H	H	5.10	5.80	-0.70
38	OH	OCH_3	OH	H	H	H	OH	5.10	5.80	-0.70
39	OH	OH	$C_6H_9O_7$	H	H	H	H	6.00	5.50	-0.50
40	OH	OH	OH	H	H	H	H	4.11	4.98	-0.87
41	OH	OH	OH	H	H	H	OH	5.25	6.01	-0.76
42	OH	H	OH	OCH_3	H	H	H	4.92	5.62	-0.70

The active molecules were employed to generate regression models utilizing the GA. According to the GA, all Pha hypotheses produced in the previous step were used to build the 4D-QSAR models. The 4D-QSAR model for binding affinity in the training set was constructed with the LOO-CV technique from the superposition, and then the predictive power of this model was validated by the test set molecules. For the interpretation of the results, only the relative magnitudes and signs of 3D structural descriptors at the ligand are important, not their absolute values. The degree of increase or decrease in binding affinity is strongly dependent on the values of descriptors. We studied the global structure (backbone-traces) as well as the local structure (binding sites) by comparing pair-wise the related atoms. Potentials created by atoms of the receptor and the ligand were computed on each point of the interaction interface. The number of interaction points in the ligand and receptor were considered as measures of the complexity of the system. The adjustable constants with related to points in the binding site were calculated by Eq. (5) using values of the ligand atoms. For a molecule, a set of descriptors was obtained by computing the interaction energy between each surface point from the virtual receptor and the atoms in the molecule.

A Pha structure consisting of four atoms stabilized the L-R complex. In the Pha structure which was depicted in Figure 2, O1 and O2 atoms made hydrogen bonds with the receptor and the other two carbon atoms made hydrophobic bonds. The two O atoms were essential for BzR site binding affinities together with other two C atoms on Pha structure.
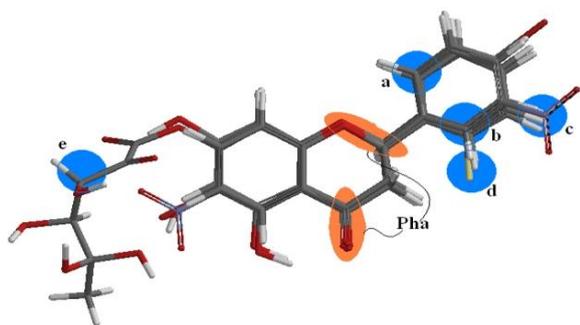

Figure 2. Common Pha structure and descriptors (AG and APS) with letters a, b, c etc. were circled.

The five descriptors shown in Figure 2 were based on simple statistics. We marked these positions with letters (a, b, c, etc.) as given in Table 2. 4D-QSAR visualization was indicated by regions of Pha structure and favorable and unfavorable descriptors of AG and APS for binding affinity, respectively.

The values of κ_j, the receptor binding parameters, were the adjustable constants corresponding to the jth positions given in Table 2. The positions were properly determined by the dihedral angle, angle and distance according to the first 3 atoms (O1, C8 and C12) of the Pha. One can describe the distance between 2 atoms, the

angle among 3 atoms and the dihedral angle among 4 atoms. As could be seen in Table 2, for instance, the distance between the C13 atom and O1 in N32 molecule was 2.71 Å; in the same orientation in atomic order the angle among C13-O1-C8 was about 66.72 degrees and the dihedral angle among C13-O1-C8-C12 was 42.61 degrees, where the Pha atoms were shown as underlined.

The quantitative explanation of the binding affinity is impossible due to the cooperative effects of the interacting points. It is appropriate to explain each interaction by discarding the relationships among different types of interaction. Since quantitative explanation required all the interactions of the model in Eq. (5), qualitative analysis was done for each interaction. The qualitative analysis occurred when the simple treatment effect in at least one position had a different sign or magnitude from that in other positions: this interaction was important. Only qualitative interpretation of the data obtained from the lowest energy conformers seemed to be easy because their effect on activity was greater. Therefore, the use of descriptors for only these conformers might be appropriate.

In this study, the atomic charge was used as a descriptor. Systematic studies on the influence of descriptors in various positions in Table 3 revealed the introduction of their sign and magnitude change activity. Discussing the magnitude and sign of the different charge at each position, we could qualitatively interpret the effect of a-, b-, c- descriptors on the predicted binding affinities in the following text.

To show the effect of charge exchange in the a- and b-positions of the interaction points, the instances of the molecules N01, N10, N29, N30 and N40 in Table 3, in which the first of the two conformers included a- and b-positions, were taken. Although atomic charge values in b-position were almost constant, the difference of those in a-position helped us to interpret its effect on the activity. A smaller negative value of the charges in a-position correlated with a lower activity value for the molecules. The correlation between the charge values in a-position and activities is due to the negative interaction between L-R. The higher magnitude of negative charge in N01 and N40 (-0.573 and -0.568 au, respectively) dramatically decreased the activity more than in N10, N29 and N30 (-0.204, -0.218 and -0.230 au). This result indicates that the negative charge in the a-position of the ligands interacts with the negative charge on the corresponding receptor site. Hence, an atom of negative charge in a-position for any ligand would act as APS with this point of receptor, and leads to the negative interaction.

With regards to b-position, there is no a clear clue in Table 3 since there was no ligand possessing only b-position, and since the charge value in the b-position for the aforementioned molecules was not changed. However, we could interpret the effect of b-position by comparing the values of κ_j in Table 2. The negative charges of conformers in a-position had a negative interaction with the negative value (-0.354) of κ_j, and

Table 2. The positions defined from the proper places of the atoms within 3 molecules (N32, N37 and N39), among the most and least active molecules

Mol. No Conf. No	Atom No	j-nth Position	Dihedral Angle	Angle	Distance	κ_j
N32_01	C13	a	42.61	66.72	2.71	-0.354
N32_01	C11	b	16.78	29.98	3.65	0.116
N32_01	N1	c	16.70	35.71	6.03	-0.156
N37_01	F1	d	356.11	11.08	4.14	-0.256
N39_01	C18	e	347.64	138.44	7.81	-17.622

Table 3. Exp. and pred. binding affinities, relative energy and structural properties of molecules N01, N10, N29, N30 and N40, which had two conformers of a- and b-position

Mol.	Exp.	Pred.	Position	Atomic charge (au)	Conformer-1	Conformer-2	
					Heat of form. (kJ mol^{-1})	Heat of form. (kJ mol^{-1})	$\Delta E_{(conf2-conf1)}$
N01	6.0	6.0	a	-0.573	-1892260.32	-1892254.34	5.98
			b	-0.232			
N10	7.6	6.8	a	-0.204	-9143856.68	-9143852.22	4.46
			b	-0.217			
N29	7.15	6.73	a	-0.218	-8612723.14	-8612717.8	5.34
			b	-0.231			
N30	6.70	6.70	a	-0.230	-9143853.63	-9143846.65	6.98
			b	-0.243			
N40	5.25	6.01	a	-0.568	-2478674.71	-2478668.64	6.07
			b	-0.231			

worked as an APS group. In the same way, the negative charge in b-position acts as an AG group with the positive value (0.116) of κ_j. In this case, the higher negative charge in b-position had more effect for AG group.

As for N17-N20, they also had substituents in c-position together with a- and b- positions at their lowest energy conformer. Likewise, their second conformers did not contain an atom in any positions like N01, N10, N29, N30 and N40. Furthermore, the addition of negative charge in c-position of ligands N17 and N18 reduced activity; on the other hand, the addition of positive charge in ligands N19 and N20 enhanced activity. The activity of ligands N19 and N20 was the much more potent than that of ligands N17 and N18 due to the change in charge from negative to positive in c-position. The negative charge in c-position may have been responsible for their lower binding affinities.

To explain of charge effect in d-position there was no districted molecule. We could interpret it as for the b-position by looking at the value of κ_j. The interaction of the negative value (-0.256) of κ_d with positive charges of conformers in d-position enhanced activity, and acted as an AG group.

For the e-position in which only ligand N39 had a substituent, the activity was decreased dramatically to 4.98 with a κ_e value of -17.622. The big change in interaction shows that there might be steric hindrance between the large size of groups within receptor and groups crowded around the C18-atom at N39. The repulsion in e-position between the receptor and ligand may occur.

This method also aided in understanding the structure–activity relationship revealed by 4D-QSAR. After a series of hierarchical filters in the MCET was systematically used to search for possible locations of the ligand in the active-site region, the study was applied to show the functional group effects. The molecules were analyzed with the binding sites to study the possible binding mode. The resultant activity revealed that a total of five functional groups were formed together with a Pha structure. The negative charges in a-, c- and d-positions of the ligand destabilized the L-R complex due to the negative interaction; on the other hand, the positive interaction in the b-position stabilized it. In addition, the active site of the receptor may be surrounded by some groups of carbon atoms in the e-position.

The statistical parameters associated in the generated QSAR model are as follows: leave-one-out $q^2 = 0.808$ (for internal validation), and predictive $r^2 = 0.775$ (for external validation). These values indicated a high degree of confidence. The regression scheme of the observed and predicted activities was shown in Figure 3. The predicted activities of the training and test set molecules were also listed in Table 4. The MCET method has been shown to be both useful and reliable for the construction of 4D QSAR models with a 3D bioactive structure.
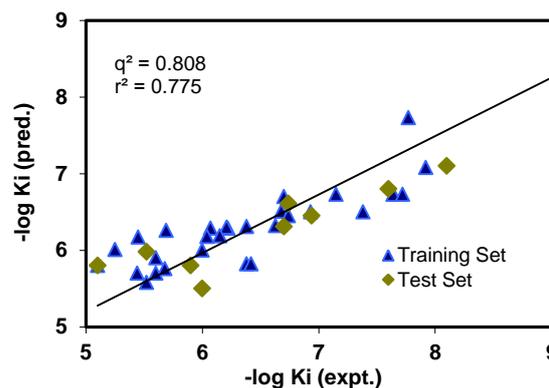

Figure 3. Fitness graph between observed and predicted binding affinity for training and test set molecules.

Table 4. Experimental and predicted activities of molecules, relative energy of the conformers and each atomic position showed by a letter a, b, c etc. in conformers

Mol	Exp.	Pred.	The heat of formation (kJ mol⁻¹) and positions of conformers♣				
N01	6.00	6.00	-1892260.3ab	-1892254.3			-
N02	5.60	5.90	-2345858.9ab	-2345857.9abc	-2345857.6ab	-2345856.4abc	-
N03	6.07	6.29	-3287100.8ab	-3287099.9abc	-3287099.5ab	-3287098.2abc	-
N04	6.22	6.29	-8808180.4ab	-8808179.4abc	-8808178.9ab	-8808177.8abc	-
N05*	6.74	6.61	-2681531.3ab	-2681531.3abc	-	-	-
N06*	8.10	7.10	-3622773.4abc	-3622773.2ab	-	-	-
N07*	5.90	5.80	-3388991.9abc	-3388991.6ab	-3388988.3	-	-
N08	5.68	5.76	-8910071.6abc	-8910071.1ab	-8910067.8	-	-
N09	6.68	6.51	-9143844.9abd	-9143832.3	-9143821.5a	-9143820.0ab	-9143817.4a
N10*	7.60	6.80	-9143856.7ab	-9143852.2	-	-	-
N11	6.38	5.82	-3349802.2abd	-3349795.5	-3349790.7ab	-3349787.4a	-
N12	6.42	5.82	-8870881.78abd	-8870875.1	-8870870.3ab	-8870866.9a	-
N13	5.45	6.17	-2150410.1ab	-2150409.4abc	-2150403.9	-2150403.2	-
N14	6.04	6.18	-2408551.0ab	-2408550.2abc	-2408545.3	-2408544.4	-
N15	6.93	6.50	-3349792.9ab	-3349792.2abc	-3349787.3	-3349786.5	-
N16	7.38	6.50	-8870872.5ab	-8870871.7abc	-8870866.8	-8870865.9	-
N17	5.44	5.70	-2150409.3abc	-2150403.6	-	-	-
N18	5.60	5.70	-2408550.2abc	-2408544.9	-	-	-
N19	6.74	6.45	-3349792.2abc	-3349787.1	-	-	-
N20*	6.94	6.45	-8870871.8abc	-8870866.6	-	-	-
N21	6.21	6.30	-3091649.1ab	-3091648.4ab	-3091642.6	-3091641.9	-
N22*	6.70	6.31	-3349790.1ab	-3349789.3ab	-3349784.0	-3349783.2	-
N23	7.64	6.73	-4291032.1ab	-4291031.4ab	-4291026.2	-4291025.5	-
N24	7.77	7.73	-9812111.6ab	-9812110.9ab	-9812105.6	-9812104.9	-
N25	6.38	6.31	-8612727.9ab	-8612727.4ab	-8612721.6	-8612720.7	-
N26	6.63	6.32	-8870868.9ab	-8870868.3ab	-8870863.1	-8870862.0	-
N27	7.64	6.73	-9812110.9ab	-9812110.3ab	-9812105.2	-9812104.2	-
N28	7.72	6.73	-15333190.4ab	-15333189.9ab	-15333184.7	-15333183.7	-
N29	7.15	6.73	-8612723.1ab	-8612717.8	-	-	-
N30	6.70	6.70	-9143853.6ab	-9143846.7	-	-	-
N31	7.92	7.08	-2954518.4ab	-2954518.3abc	-2954513.1	-2954512.8	-
N32*	9.00	9.10	-9143852.9abc	-9143852.8ab	-9143846.8	-9143846.8	-
N33	6.15	6.19	-15724126.4ab	-15724124.2	-15724123.4ab	-	-
N34*	5.52	5.98	-2283217.7ab	-2283215.4ab	-2283211.6	-	-
N35	5.52	5.58	-2478673.9abc	-2478671.4abc	-2478670.9ab	-2478668.8§	-2478668.2
N36*	5.10	5.80	-3482602.9ab	-3482591.3	-3482588.8	-3482586.2§	-348258.9ab
N37	5.10	5.80	-2541376.7abd	-2541370.2	-2541363.0ab	-2541360.2§	-
N38*	6.00	5.50	-2776015.1abc	-2776010.1	-2775985.7abc	-	-
N40	5.25	6.01	-2478674.7ab	-2478668.6	-2478646.0ab	-	-
N41	4.92	5.62	-2674131.0abc	-2674130.2ab	-2674125.6	-2674102.1abc	-
N42	5.69	6.26	-2580553.6ab	-2580551.5	-2580545.4a	-2580531.8ab	-2580531.4

♣More than 5 conformers were not shown
§Conformers, which not possessing Pha structure, were not affecting the binding affinity
*Test set compounds

4. CONCLUSIONS

The MCET method used in this work is related to the class of ligand-based approaches in drug design. Considering the properties of a ligand rather than its biological target (biological receptor), such methods are especially useful for the prediction of activity when the structure of the receptor is unknown. As the main part of this work, we have constructed a model showing the interaction between the receptor and the Pha, AG and APS groups in the multiple conformers of the ligand.

In summary, for a series of flavonoids, a significant 4D-QSAR model was applied to 32 ligands in the training set and validated with satisfactory predictions with 10 ligands in the test set. The results of the model suggest that a Pha structure consisting of four atoms is required for the activity, depending on the interactions of the 3D structural identifiers. In addition, the model shows the contribution of the different AG and APS values of each ligand conformer to activity. Furthermore, it is explained how the AG and APS atoms have influence in which positions due to the alignment of the Pha structure in the model. It has been shown how the activity of flavone derivatives can be influenced by positive (AG) or negative (APS) interactions in the defined positions. The present study provided useful guidelines through 3D structural identifiers to develop flavone derivatives as potent active molecules in ligand-based drug design approaches.

Conflict of interest

Authors declare that there is no a conflict of interest with any person, institute, company, etc.

REFERENCES

1. Martens, S.; Mithöfer A. *Phytochem.* **2005**, 66, 2399-2407.

2. Harborne, J.B.; Williams C.A. *Phytochem.* **200**0, 55, 481–504.

3. Treutter, D. *Environ. Chem. Lett.* **2006**, 4: 147–157.

4. Medina, J.H.; Viola, H.; Wolfman, C.; Marder, M.; Wasowski, C.; Calvo, D. Paladini AC. *Neurochem. Res.* **1997**, 22, 419-425.

5. Marder, M.; Viola, H.; Wasowski, C.; Wolfman, C.; Waterman, P.G.; Cassels, B.K.; Medina, J.H.; Paladini, A.C. *Biochem. Biophy. Res. Commun.* **1996**, 223, 384-389.

6. Marder, M.; Zinczuk, J.; Colombo, M.I.; Wasowski, C.; Viola, H.; Wolfman, C.; Medina, J.H.; Ruveda, E.A.; Paladini, A.C. *Bioorg. Med. Chem. Lett.* **1997**, 7, 2003-2008.

7. Marder, M.; Viola, H.; Bacigaluppo, J.A.; Colombo, M.I.; Wasowski, C.; Wolfman, C.; Medina, J.H.; Ruveda, E.A.; Paladini, A.C. *Biochem. Biophy. Res. Commun.* **1998**, 249, 481-485.

8. Teuber, L.; Watjen, F.; Jensen, L.H. *Curr. Pharm. Des.* **1999**, 5, 317- 343.

9. Argyropoulos, S.V.; Nutt, D. *Eur. Neuropsychophar harmacol.* **1999**, 6, 407-412.

10. Tija, C.J.; Stephen, J.M.; Rachel, J. *Nat. Rev. Neurosci.* **2008**, 9, 331-343.

11. Amić, D.; Davidović-Amić, D.; Jurić, A.; Lučić, B.; Trinajstić, N. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 1034-1038.

12. Costantino, L.; Rastelli, G.; Albasini, A. *Eur. J. Med. Chem.* **1996**, 31, 693–699.

13. Li, G.; Mian, Z.; Song, W.; Ai-Lin, L.; Guan-Hua, D. *Bioorg. Med. Chem. Lett*. **2011**, 21, 5964-5970.

14. Duchowicz, P.R.; Vitale, M.G.; Castro, E.A.; Autino, J.C.; Romanelli, G.P.; Bennardi, D.O. *Eur. J. Med. Chem.* **2008**, 43, 1593-1602.

15. Yilmaz, H.; Güzel, Y.; Önal, Z.; Altiparmak, G.; Özhan-Kocakaya, Ş. *Bull. Korean Chem. Soc.* **2012**, 32, 4352-4360.

16. Yilmaz, H.; Güzel, Y.; Boz, M.; Türkmenoğlu, B. *Trop. J. Pharm. Res.* **2014**, 13,117-129.

17. Hopfinger, A.J.; Wang, S.; Tokarski, J.S.; Jin, B.; Albuquerque, M.; Madhav, P.J.; Duraiswami, C. *J Am Chem Soc.* **1997**, 119, 10509-10524.

18. Hopfinger, A.J.; Reaka, A.; Venkatarangan, P.; Duca, J.S.; Wang, S. *J. Chem Inf Comput Sci.* **1999**, 39: 1151-1160.

19. Senese, C.L.; Duca, J.; Pan, D.; Hopfinger, A.J.; Tseng, Y.J. *J. Chem. Inf. Comput. Sci.* **2004**, 44: 1526-1539.

20. Huang, X.; Liu, T.; Gu, J.; Luo, X.; Ji, R.; Cao, Y.; Xue, H.; Wong, J.T.; Wong, B.L.; Pei, G.; Jiang, H.; Chen, K. *J. Med. Chem.* **2001**, 44, 1883-1891.

21. Bernardino, A.M.R.; Castro, H.C.; Frugulhetti, L.C.P.P.; Loureiro, N.I.V.; Azevedo, A.R.; Pinheiro, L.C.S.; Souza, T.M.L.; Giongo, V.; Passamani, F.; Magalhaes, U.O.; Albuquerque, M.G.; Cabral, L.M.; Rodrigues, C.R. *Bioorg. Med. Chem.* **2008**, 16, 8670-8675.

22. Bernardino, A.M.R.; Pinheiro, L.C.S.; Rodrigues, C.R.; Loureiro, N.I.V.; Castro, H.C.; Lanfredi-Rangel, A.; Sabatini-Lopes, J.; Borges, J.C.; Carvalho, J.M.; Romeiro, G.A.; Ferreira, V.F.; Frugulhetti, I.C.P.P.; Vannier-Santos, M.A. *Bioorg. Med. Chem.* **2006**, 14, 5765-5770.

23. Bharatam, P.V. Modeling and Informatics in Drug Design. *Precilinical Development Handbook.* **2008**, 1-47.

24. Hopfinger, A.J.; Wang, S.; Tokarski, J.S.; Jin, B.; Albuquerque, M.; Madhav, P.J.; Duraiswami, C. *J. Am. Chem. Soc.* **1997**, 119, 10509-10524.

25. Karelson, M.; Lobanov, V.S.; Katritzky, A.R. *Chem Rev.* **1996**, 96, 1027-1044.

26. Bersuker, I. B.; Dimoglo, A.S. In *Reviews in Computational Chemistry*, Lipkowitz, K. B., Boyd, D.B., Eds.; VCH: New York, NY, U.S.A., 1991; Vol 2, pp 423-460.

27. Bersuker, I.B.; Dimoglo, A.S.; Yu, M.; Gorbachov, P.; Vlad, F.; Pesaro, M. *New J. Chem.* **1991**, 15, 307-320.

28. Dimoglo, A.S.; Beda, A.A.; Shvets, N.M.; Gorvachov, M.Y.; Kheifits, L.A.; Aulchenko, I.S. *New. J. Chem.* **1995**, 19, 149-154.

29. Guzel, Y.; Saripinar, E.; Yildirim, I. *J. Mol. Struct.* **1997**, 418, 83-91.

30. Guzel, Y.; Ozturk, E. *Bioorg. Med. Chem.* **2003**, 11, 4383-4388.

31. Bersuker, I.B. *J Comput. Aided Mol. Des.* **2008**, 22, 423–430.

32. Bersuker, I.B.; Bahceci, S.; Boggs, E.J. *The Electron Conformational Method of Identification of Pharmacophore and Anti-Pharmacophore Shielding.* In *Pharmacophore Perception. Development, and Use in Drug Design.* Guner OF., Ed.; International University Line: La Jolla, 2000, 457-474.

33. Bersuker, I.B.; Bahceci, S.; Boggs, E.J.; Pearlman, R.S. *J. Comput. Aided Mol. Des.* **1999**, 13, 419-434.

34. Bersuker, I.B.; Bahceci, S.; Boggs, E.J.; Pearlman, R.S. *EnViron. Res.* **1999**, 10, 157-173.

ORCİD

ID 0000-0002-5770-0847 (B. Türkmenoğlu)

ID 0000-0001-9305-8364 (H. Yılmaz)

ID 0000-0002-4519-914 3 (E.M. Su)

ID 0000-0003-2713-7093 (T. Alp Tokat)

ID 0000-0002-4698-1866 (Y. Güzel)