Original Research Article

Pharmacophore and Functional Group Identification of 4,4′-dihydroxydiphenylmethane as Bisphenol-A (BSA) Derivative

Hayriye Yılmaz[1*], Mehmet Boz[2], Burçin Türkmenoğlu[2] and Yahya Güzel[2]

[1]Faculty of Pharmacy, [2]Faculty of Science, Department of Chemistry, Erciyes University, Kayseri, 38039, Turkey.

*For correspondence: Email: hayriyey@erciyes.edu.tr; **Tel:** +90-352-437-91-69

Abstract

Purpose: To predict activity and reveal the pharmacophore (Pha) with certain electronic and topological characteristics for a series of 37 molecules of 4,4'-dihydroxydiphenylmethane, using 4D QSAR (four dimensional Quantitative-Structure Activity Relationships) model.
Methods: We used a computational method called molecular conformer electron topological (MCET) for this study. The quality of Pha and the corresponding quantitative model of activity was validated (and deemed acceptable) by an independent test set of 7 additional analogs with known experimental activities out of 30 molecules of the training set.
Results: The resulting MCET method demonstrated a high statistical capacity for predicting the activity of the molecules under consideration ($R^2 = 0.703$ and $Q^2 = 0.573$).
Conclusion: The model is based on pure computational methods (electronic structure calculations and matrix comparisons) and provides the correct solution within the assumptions of the method, experimental uncertainty, and computational approximations. A different procedure from other QSAR approaches was used to elucidate the interactions between the conformers of the ligand and the target protein.

Keywords: Drug design, estrogenic activity, electron topologic method, 4D-Quantitative-Structure Activity Relationships, 4,4' dihydroxydiphenylmethane.

INTRODUCTION

Bisphenol A (BPA), which is elected in the manufacture of a wide range of consumer products, is a prime candidate for endocrine disruption [1]. BPA is a monomer composed of two unsaturated phenolic rings. The in vitro studies proved that BPA binds to the estrogen receptors, induces estrogen-dependent gene responses [2]. Synthetic endocrine-disrupting chemicals (EDCs) bear the possibility to meddle in the endocrine system by impersonating endogenous hormones such as estrogens and androgens [3].

Computational methods establishing QSARs to predict activity differences within a set of ligands remain a pragmatic alternative. Both classical and 3D QSAR methods have been developed as ligand based approaches [4]. In 3D QSAR such as CoMFA (Comparative Molecular Field Analysis) method, steric and electrostatic features are essentially virtual interaction energies calculated using an appropriate probe atom placed at the intersection of a regularly

spaced grid surrounding the molecule. It could be interpreted as a surrogate representation of the binding site [5]. 4D-QSAR is an effective way for the identification of Pha in action and nature of interactions between conformers of the molecules and target protein especially [6-7]. A pharmacophore map identifies the bioactive sub-structure of each active molecule and indicates how to align. The Pha map represents relationship which types of points match in conformations of the compounds, and represents the common sub-structure of molecules with different structure. In this study the MCET method with 4D QSAR is used to identify ligand substituents needed for high-RBR (Relative Binding Ratio) with ER (estrogen receptor) and explain their orientation. This method for ER is employed to interpret results from affinity studies, and related this information with the design of new ligands. This analysis including ER phenolic molecules is used to obtain information on ligand-receptor interactions that lead to either an increase or decrease in ER affinity for similar or different molecules. In this paper atomic charge and geometric descriptors are used to describe molecular structure of bisphenol A analogs and MCET model based on that their estrogen activities are developed to predict the ligand-receptor interaction points.

METHODS

A set of 4,4′-dihydroxydiphenylmethane derivatives is collected from the literature along with their activity. The collected data are presented in Table 1 [8].

**Table 1:** Specific estrogenic activity of 4,4′-Dihydroxydiphenylmethanes (units/mol) in rats
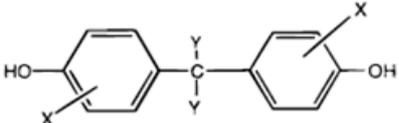
4,4-Dihydroxydiphenylmethanes

Molecule no.	Substituents		log RBR (unit/mol)	
			Observed	Calculated from Eq 3
	X	Y		
01	H	C_3H_7, C_2H_5	0.57	0.57
02	H	C_3H_7, C_3H_7	0.25	0.57
03	H	$i-C_4H_9$, Me	-0.03	-0.14
04	H	C_3H_7, Me	-0.11	-0.43
05	H	Et,Et	-0.11	0.10
06	H	C_4H_9, C_3H_7	-0.17	0.57
07 *	2-Me	$(Et)_2CH,H$	-0.29	-0.38
08	H	Et ,Me	-0.38	-0.25
09	2-Me	$i-C_4H_9$ H	-0.41	-0.45
10	H	$CH_2C_6H_5$,H	-0.46	-0.64
11	2-Me	$(C_3H_7)_2CH$	-0.52	-0.70
12	2-Me	C_3H_7,H	-0.78	-0.74
13 *	H	Me,Me	-0.80	-0.82
14	H	C_4H_9,Me	-0.83	-0.02
15	2-Me	$i-C_3H_7$,H	-0.83	-0.70
16	2-Me	C_6H_{13},H	-0.85	-1.04
17	H	C_6H_5, H	-0.87	-0.80
18	3-Me	H,H	-1.00	-0.60
19 *	H	C_4H_9, C_4H_9	-1.02	-1.51
20	2-Me	Me,H	-1.02	-0.78
21	2-Me	Et,H	-1.15	-0.93
22 *	H	$(C_6H_5)2CH$,H	-1.62	-1.15

Table 1 (continued): Specific estrogenic activity of 4,4′-Dihydroxydiphenylmethanes (units/mol) in rats

Molecule no.	Substituents	log RBR (unit/mol)		
	X	Y	Observed	Calculated from Eq 3
23	H	C_6H_5,Me	-1.86	-1.07
24	3-Me	C_3H_7,C_3H_7	1.20	1.22
25 *	3-Me	C_3H_7,Et	-0.63	-0.59
26	3-Me	Et,Et	0.55	0.71
27	3-Me	C_4H_9,C_3H_7	0.49	0.37
28	3-Me	C_3H_7,Me	-0.06	0.19
29	3-Me	Me,i-C_4H_9	-0.22	-0.03
30 *	3-Me	Me,Me	-0.81	-1.51
31	3-Me	Me,Et	-0.83	-0.37
32	3-Me	Me,C_4H_9	-0.87	0.18
33	3-Me	H,i-C_4H_9	-1.00	-1.02
34 *	3-Me	C_4H_9,C_4H_9	-1.23	-0.70
35	3-Me	Et,H	-1.50	-0.62
36	3-Me	Me, C_6H_5	-1.72	-0.65
37 *	3-Me	H, C_6H_{13}	-1.89	-1.88

*Test set molecules
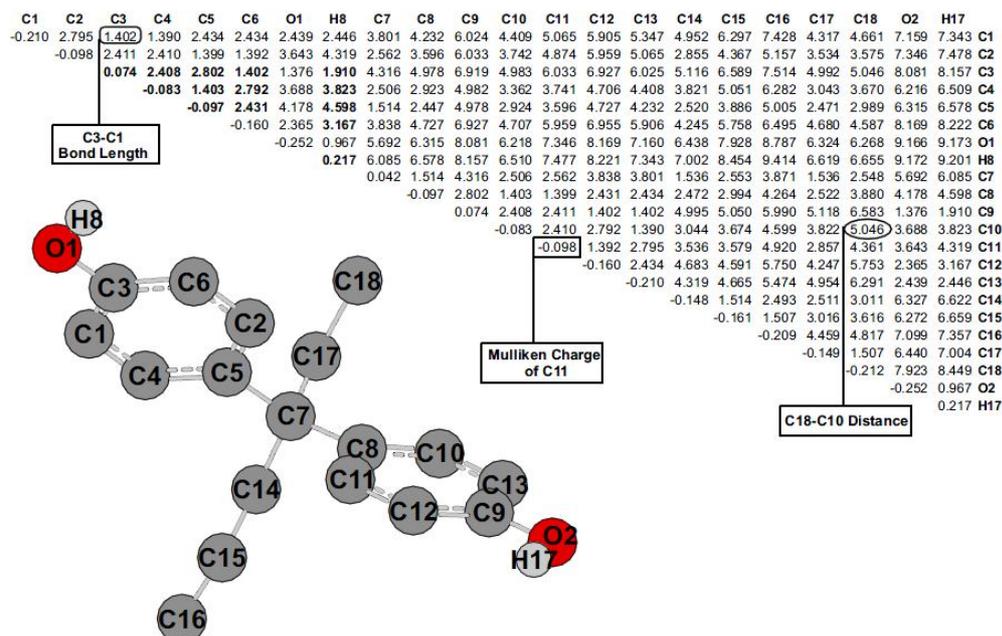
Figure 1: Illustration of ETM and three dimensional structure of reference molecule N01; Pha-group is shown in bold letters.

Conformational analysis

With respect to the molecular conformations, quantum chemical calculations are obtained via "Spartan'08" software [9]. In the first step, Spartan's molecular mechanics force field (MMFF) is used, because it presently provides the calculation of equilibrium geometries, strain energies and normal-mode vibrational frequencies, as well as searches for conformation space for both cyclic and acyclic molecules. In the second step, 3-21G(*) is selected in Spartan's Hartree-Fock module which provides the calculation of the heat of formation, equilibrium and transition-state geometries, the charge on each atom and other atomic properties [10]. After energy minimizations and molecular dynamics calculations are performed, the conformers with the lowest energy for each molecule were selected as the acceptable conformers. These conformers have energies<1 kcal/mol above the ground state conformation. Since acceptable conformers have a larger

population than other conformers, according to Boltzmann distribution, if these conformers possess Pha, they are more responsible than others for activity [11].

Computation of electron topological matrix

We calculate the electronic structure of each selected conformation and arrange the corresponding electronic and geometrical parameters in a matrix $k*k$ (k is the number of atoms), called electron topologic matrix (ETM) The matrix of acceptable conformers is kept in the memory of the computer and processed by MCET software which was written using the C# programming language by our research group.

The formation of ETM as a language for conformer description (one matrix=one compound) proceeds in the following way [12]:
Lets consider A_i be an atom of a conformer being described, then the corresponding diagonal matrix element a_{ii} will be one of m1, which is the number of local atomic properties.

If A_i and A_j are any two atoms of the conformer then two cases may occur:
a) A_i and A_j are not chemically bonded. In this case the distance between these two atoms represents the corresponding non-diagonal matrix element a_{ij}.
b) A_i and A_j are chemically bonded, and a_{ij} describes this bond by means of electronic properties (bond length, bond order, bond energy, etc.).
Suppose that we have m2, which is the number of such characteristics, then the number of all ordinary $k*k$ matrices is m=m1*m2 where is dimensional ETM. The matrices, based on the data known from quantum chemical calculations, can be formed as well as the number of atoms, k, in the corresponding conformer. It has been assumed that a single electronic matrix in the ETM represents a configuration of each molecule. The matrix contains charge on the diagonal, bond length, and distance on the non-diagonal, and it is one-dimensional. The ETM of N01 in the 4,4′-dihydroxydiphenylmethane series is shown in Figure 1**.**

Pha is commonly defined as an arrangement of molecular features or fragments forming a necessary but not sufficient condition for biological activity [13]. To find Pha, a template conformer of active molecule and the acceptable conformer of the rest of the molecule set are compared as weighted graphs. One conformer of a molecule (high active and simple structure) is taken as a template, and its ETM is compared with the rest of the molecules' ETMs within

tolerances [14]. Flexibility limits are quite important for the realization of Pha. Larger tolerances yield large common submatrices while small tolerances give no Pha within the training set. Therefore, the issue of tolerances should be handled in a delicate manner [11].

Comparison of Topological Matrices and identification of Pha

The ETM is used in both the electron topologic (ET) [15-18] and the electron conformer (EC) methods [22]. The ET method accounts only one conformer with the lowest energy of each ligand and emphasizes only Pha for 3D-QSAR analysis. Likewise for 4D QSAR analysis, using the acceptable conformers of each molecule, the EC method takes into account the other side groups such as AG and APS besides Pha [11]. While Pha is responsible for the existence or non-existence of activity, APS and AG are characteristic variable and responsible for the change in activity values of molecules. A criterion that is commonly used in structural methods for evaluating probable Pha in the series under study is represented by the following formulas Eq.1:

$$P_\alpha = (\ v1+1)/(\ v1+\ v3+2); \quad \alpha_a=(\ v1*\ v4-\ v2*\ v3)/(\mu1*\ \mu2*\ \mu3*\ \mu4)^{1/2} \qquad \text{Eq.1}$$

where v1 and v2 are the numbers of molecules possessing and not possessing, respectively, the feature of activity in the class of molecules; v3 and v4 have analogous meaning in low active molecules; and μ1 and μ2 are the numbers of molecules in the class of active and low active molecules μ3= v1+ v3 and μ4= v2+ v4. This way P_α evaluates only the deposit of active molecules, while α_a reflects the deposit of both active and low active molecules in the feature of the activity found. Then, without setting any constraints on tolerance values, maximum tolerance values are defined for all active molecules [19].

Formula for Quantitative Structure-Activity Relationships

After Pha was identified by comparing the ETM matrices of conformers, independent variables using the properties of atoms acting like AG or APS on the ligand were estimated from the positions consisting of the defined torsion angles, angles, and distances in regard to Pha. The adjustable constants of interaction points on the receptor, which correspond to all the possible positions of the ligand, are calculated by using Newton-Raphson approach [20].

Bersuker et al. introduce the function S to take into account the independent variables ($a_{ni}^{(j)}$) and parameters (κ_j) of J-positions in Eq.2 as follows [11]:

$$S_{ni} = \sum_{j=1}^{J} \kappa_j a_{ni}^{(j)} \qquad Eq.2$$

where $a_{ni}^{(j)}$ describes the functional group of jth position in the ith conformation of the nth molecule and it is also taken as an atomic charge. κ_j (j=1,2…J) indicates the magnitude of values at the interaction points of receptor. For each of the side groups of J-number used in the model, κ_j is based upon magnitude of the corresponding partial least square (PLS) regression coefficients, and it is calculated using multiple nonlinear Eq.3 for n^{th} molecule.

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

A_0 in Eq. 3 is a constant (see below), m_n^{Pha} and m_n are numbers of the conformers containing Pha and the acceptable conformer, respectively. E is the heat of formation; k is Boltzmann constant, and T is a room temperature.

A reference molecule (ℓ) is chosen from the training set, for which the activity is known as A_ℓ in Eq. 4:

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

It is difficult to find the A_0 constant because A_0 constant is not directly determined and known. To find A_0 we divide Eq.3 by Eq.4 by eliminating A_0 in the both equations. The following formula is used to find the activity prediction through different conformations of the same molecule. With the knowledge that the activity depends on the tolerances in the values of the sub-matrix of Pha, we can proceed to find an approximate model for the quantitative value of the activity using Eq.5. The following conditions will need to be fulfilled:
- Using the Boltzmann distribution formula in related to the conformation energy E_{ni}
- Selecting the conformers possessing Pha,
- Taking into account values of functional groups of each conformer.

Based on these conditions we can calculate the activity of the n-th compound by the different contributions of conformers in Eq.5:

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

The functions of S_{ni} and $S_{\ell i}$ on the right of Eq.5 are containing substituents such as the AG or APS group, are formulated as an exponential such as the Boltzmann factor. Boltzmann factor is a weighting factor that determines the relative probability of a conformer to be in a multi-state molecule in thermodynamic equilibrium at temperature T.

The MCET is modified in such a similar way to the EC, but the applied procedure and the computer programming are different from the EC. Some of the most important progresses and differences in the MCET are the following:

(a) To develop a pharmacophore model, several approaches have been proposed [21]. The structure of Pha should include at least one conformer of an active and simple molecule. If there is more than one conformer in the template molecule, each of conformers can be sequentially used in producing Pha which is a common structure for the active molecules. Pha used as a useful common template in our method is a hypothetical approach, and the process of superimposition is based on it. Even if there is a different molecular structure, it is conformationally directed to assume the shape obligatory for its sub-molecular map. The superimposition is done for matching of the acceptable conformers of other molecules with the template conformer. After all possible pharmacophores are generated with the combination of the atoms of the template molecular conformer, the conformers with Pha may be aligned and superimposed possible on the atoms of Pha with a great degree of accuracy.

(b) An algorithm for an automatic extraction of the APS and AG variables from the ETM with the purpose to evaluate them has not yet found. Therefore the extraction has been done individually for each conformation. However, such an algorithm has been constructed and employed in the MCET method, and the variables of a_{ni}^j and the values of the corresponding parameter of κ_j in Eq.2 have been

simultaneously estimated using Newton-Raphson approximation [20]. The predicted κ_j in the S function of Eq.5 is controlled by the minimization of the variables. The minimization can be realized by means of the PLS procedure. The predicted κ_j is determined from the condition of minimum of the sum $\sum_n^N |A_n^{calc} - A_n^{exp}|^2$, where N is the number of molecules in a training set. The value of κ_j obtained in this way characterizes the weights of each kind of the $a_{ni}^{(j)}$ parameters in the overall APS/AG influence [14].

For many κ_j-problems, we would need values of all the other κ_j [22] to solve one κ_j for j = 1, 2,...J. Such a calculation requires a self-consistent approach as outlined below.
1. Start with trial κ_j as number J-1 (except one κ_i, initial value of j equals 1).
2. To calculate one κ_j, use the defined value of all remained κ_j.
3. Solve the Newton-Raphson Equation derived from Eq.5 for one κ_j.
4. Define the new value of one κ_j.
5. Repeat the calculation for another κ_j coming after the defined ones.
6. Continue iteration until self-consistent results for all the κ_j are obtained, i.e., the activities of all molecules do not change from one iteration to the next.

With this approach, all the κ_j are calculated according to Eq.5 to form the 4D QSAR model.

(c) While the CoMFA uses only probes of a different charge to pick out the descriptors distributed in various regions around the ligand, the MCET employs only the atomic properties (for example charge, polarizability, HOMO- and LUMO-orbital coefficients) to select the independent variables oriented around Pha. Since the purpose of this study is to identify the interaction points of bio-structure in three dimensional spaces, the properties of atoms in positions interacting with a receptor are used as an independent variable. According to the variable of distance, angle or dihedral-angle, only the positions of atoms in related Pha for all conformers can be classified.

d)   The positions in Eq.2through GA are varied in order to determine the optimal number of descriptors [23-24]. GA can not only automatically select the optimum number of descriptors in linear regression analysis but also construct the model of multiple nonlinear equations (5), which includes nonlinear exponential functions. Therefore, GA is employed with linear regression analysis to derive and validate the model taken from Eq.5. Thus, the relationship between structure and

activity can be quickly identified from the results of the model constructed using MCET.

Results of the model which explain the observed biological activity are controlled basing on PLS analysis; data fitting for the training set. A procedure which involves iterative selection of bio-structures can be repeated until squared correlation coefficient (Q^2) reaches to big value. For the test set, model predictive ability is often quantified in terms of the predictive squared correlation coefficient (R^2) which is calculated by methods based on some form of sample reuse, such as leave-one-out cross-validation (LOO-CV) [25]:

$$R^2 = 1 - \frac{[\sum_{i=1}^{n_{TR}}(\hat{y}_i - y_i)^2]/n_{TR}}{[\sum_{i=1}^{n_{TR}}(y_i - \bar{y}_{TR})^2]/n_{TR}} = 1 - \frac{RSS/n_{TR}}{TSS/n_{TR}} \quad (5)$$

where n_{Test} is the total number of objects in the entire test set, $\hat{y}_i$ is the predicted value for the ith test molecule and y_i is observed value and TSS is the total sum of squares, that is, the sum of squared deviations from the data set mean, and PRESS is the sum of squares of the prediction errors. Common definition of the parameter Q^2 used for assessing the model fit from the training set (TR) objects [25]:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_{Test}}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n_{Test}}(y_i - \bar{y}_{Test})^2} = 1 - \frac{PRESS}{TSS} \quad (6)$$

where RSS is the residual sum of squares, that is, the sum of the squared deviations between experimental and calculated response values over the training set. TSS is the total sum of squares, which is related to the total response variance of the training cases and n_{TR} is the total number of objects belonging to the training set.

RESULTS

The values and positions of the atoms were unique for activities of the molecules under consideration. As shown in Figure 1 the sub matrix was marked with the bold letter to show Pha placed in the ETM of template conformer.

All the conformers containing Pha are matched and aligned by placing the coordinate values of the first three atoms of Pha using a Cartesian coordinate system, providing x1=0; y1=0; z1=0, x2=0; y2=0; z2≠0 and x3=0; y3≠0; z3≠0, respectively. A Pha needs at least three points of connection, but there can be more. A complete Pha can be represented geometrically as a

triangle if three points are used; or as some other geometric figures if more points are available. To compare the compounds with the various structures in 3D, first three atoms of Pha structure, instead of atoms in the common skeleton of compounds, is aligned, and then the remaining atoms are again oriented according to Pha. After the structural alignment of the considered Pha, the functional AG and APS groups from several variations of the oriented atoms of the molecules were determined by GA, and the best statistical results were obtained by taking them into account together with Pha. Interactions between a receptor and a ligand are complex and depend upon the context in which the groups of ligands are presented to the receptor. These interactions are also more responsible for the activity than other factors. Local independent variables, which are the electronic properties of the oriented atoms around Pha, describe the ability of atomic sites and fragments to take part mostly in interactions dependent upon the present Pha, and subsequently should be taken into account for the modeling activity. These parameters, however, can also be used for assessing the propensity of chemicals to take part in weak interactions; hence, they are also included in the list of parameters available for the description of non-specific interactions. Finally, specific parameters of Pha between a ligand and a receptor have also shown to be critically important for deriving mechanistically reasonable models. These molecules might become more or less active due to their conformations and functional groups. As some atoms of the conformers in the molecules undergo

Table 2: The observed and calculated data related to the molecules of training set understudy

Mol	Obs.	Calc.	The heat of formation (kcal/mol) and AG/APS positions of conformers.								
n01	0.57	0.57	-265.53ab	-265.53abc							
n02	0.25	0.57	-293.78abc	-293.78ab							
n03	-0.03	-0.14	-249.41ab	-250.33ab	-249.29ab						
n04	-0.11	-0.43	-247.70ab	-247.70ab	-247.52ab	-247.52ab	§-247.34	§-247.34	§-245.88		
n05	-0.11	0.10	-237.22abc	§-237.08	§-236.91						
n06	-0.17	0.57	-322.37abc	-322.37ab							
n07*	-0.29	-0.38	-311.59ab	-311.42ab	§-311.42	§-311.42	-311.38ab	-311.19ab	§-310.86	-308.53abc	-307.49abd
n08	-0.38	-0.25	-219.46ab	-219.46ab	§-216.31						
n09	-0.41	-0.45	-288.57abc	-288.56abc	§-285.65						
n10	-0.46	-0.64	§-90.53	§-90.53	§-90.45	-90.38ab	-89.01ab				
n11	-0.52	-0.70	-370.31abc	-370.31abc	-370.08abc	-370.07abc	§-370.00	§-369.85	-369.82abc	-369.76abc	-369.76ab
n12	-0.78	-0.74	-284.41abc	-284.09ab	-284.06abc	§-283.90	-283.76ab	§-283.59	§-283.55	§-283.55	-282.34abc
n13*	-0.80	-0.82	-201.67abd	-201.48abd	§-201.33						
n14	-0.83	-0.02	-276.29ab	-276.11abd	-276.07abd	§-275.93	-275.67ab	-275.64abc	-275.50abd	-275.42abd	-275.28abd
n15	-0.80	-0.70	-267.55ab	-267.49ab							
n16	-0.85	-1.04	-370.00abd	-369.67abd	§-369.50	§-369.45	-369.45abd	-367.92abd	-367.41abc	-367.01abd	-366.69abd
n17	-0.87	-0.80	-57.14ab	-57.14abcd	§-57.11	§-57.11	-57.11abd	-57.10abc	§-57.04	-56.09abc	-55.84abc
n18	-1.00	-0.60	§-223.99	§-222.40	§-222.40	§-222.39	§-222.38	-222.18ab	§-222.17	§-222.16	§-222.16
n19*	-1.02	-1.51	-350.96ab	-350.96abc	-350.78abc	-350.77abc	-350.75abd	§-347.61			
n20	-1.02	-0.78	-230.89ab	-230.69ab	-230.23abcd						
n21	-1.15	-0.93	-255.16abc	-255.12abcd	-254.96abcd	§-254.94	-254.89abc				
n22	-1.62	-1.15	-244.12	-244.53abcd	-244.57abd	-244.45d	-244.12abc	-244.31abd			
n23	-1.86	-1.07	§-44.53	-44.53abc	-44.47abd	§-44.45	-44.43abc	-44.31abd	-44.31abd	§-44.31	-44.30ab
n24	1.20	1.22	-321.77abd	§-321.66	§-321.66	§-321.66	§-321.66	§-321.66	§-321.66	§-321.66	§-321.66
n25*	1.13	-0.59	-295.55b	§-295.32	§-295.32	-295.16ab	-293.52ab	-293.52abc	§-293.42	§-293.41	-293.32abd
n26	0.55	0.71	-267.32abd	§-267.08	-266.92abd	§-266.76	-265.28ab	§-265.17	§-265.07	§-264.99	
n27	0.49	0.37	-352.39ab	§-352.16	-352.16ab	-352.15abd	§-352.15	-352.00ab	-351.99abd	§-351.83	§-351.83
n28	-0.06	0.19	-277.73abc	-277.73abd	§-277.52	-277.50ab	-277.36ab	-277.35abd	§-277.21	§-277.18	-276.68abd
n29	-0.22	-0.03	-280.92ab	-280.67ab	-280.48abd	§-280.35	-280.24abc	§-279.95	-279.84ab	§-279.74	-278.88abd
n30*	-0.81	-1.51	-231.71abd	-231.50abd	-231.34abd	§-231.19	-229.68abd	§-229.58	-229.49abd	§-229.42	
n31	-0.83	-0.37	-249.48ab	-249.44abc	-249.11abd	-249.10ab	-247.84abd	§-247.77	§-247.75	§-247.27	-246.42ab
n32	-0.87	0.18	-306.33abc	-305.92abd	§-305.77	-304.95abc	§-304.17	§-304.12	-303.38abc	-303.17abd	-303.14ab
n33	-1.00	-1.02	-297.07abc	-296.82ab	§-296.72	-296.70ab	-296.46abd	-295.12abc	§-294.92	-294.90ab	§-294.88
n34*	-1.23	-0.70	-380.75abd	-380.59abd	§-378.84	-378.02abd	-378.01abc	§-377.78	§-377.78	-377.74ab	-377.74ab
n35	-1.50	-0.62	-259.47abc	-259.45ab	-259.17ab	-259.13ab	§-259.09	§-259.09	-257.82ab	-257.81abc	§-257.50
n36	-1.72	-0.65	§-74.42	-74.42abc	-74.35ab	§-74.32	§-74.24	-74.22abd	§-74.14	§-73.95	§-72.47
n37*	-1.89	-1.88	-374.01abd	-372.09abd	-372.08abd	§-371.82	-371.37abd	-371.03ab	-370.23abd		

*Test set molecules. "§"the conformers not possessing Pha. For the conformers possessing Pha, Atomic positions are shown with signs of a, b, c, d.

conformational change, they might catch the appropriate position affecting the activity. All the positions were distinguished from each other according to the distance and the dihedral, and angle values. The Pha and the functional groups of the conformers in these positions were signed as symbols a, b, c, etc.

According to the data set shown in Table 2, the position of one conformer might be different from that of the other conformers. To calculate the activity, these groups, together with Pha, were used in Eq.5by separating and combining their electrostatic and shape fields.

Table 2 presents the experimental and theoretical (calculated) activity, the heat of formation of the conformers, and the signs of each conformer's positions; conformers not possessing Pha were marked with the "§" symbol.

Unfortunately, since there were more variables than the defined side groups, such as the APS or AG, this extensive study of electronic and geometric similarity was flawed by the apparent incorporation of numerous errors in the molecular structure as well as accounting errors in the measurements of the activity. In the current study, we did a quantitative account of the model in order to define the most appropriate AG and APS groups together with Pha. For correct classification of the positions, these characteristic properties, given in Table 4, were then subjected to quantitative analyses via multiple linear regressions. To demonstrate the receptor-ligand interaction, the results in the data set were compared with the corresponding MCET. These local parameters in this study were composed of three different positions affecting the activity and were marked with blue circles in Figure 2.
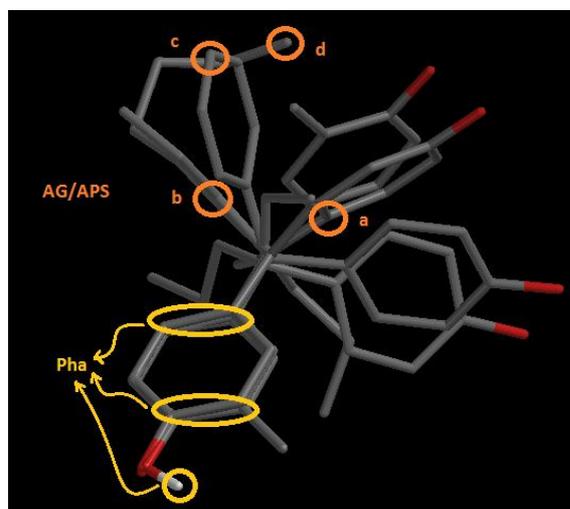

Figure 2. Pha (yellow circles) and AG or APS groups of a,b,c,d (orange circles)

Using LOO-CV, κ-values corresponding to the interaction points of the receptor were simultaneously calculated from the experimental activity of the molecules in the training set and then used to predict the activity of those in the test set. The positioned local atoms in each conformer of the ligand contributed differently to the activity. According to Eq.2 S_{ni} is the sum of the multiplication of the κ-value on the receptor by the related atomic charge on the ligand; its value is changed and affected negatively or positively, depending on the signs of the κ-value and the atomic charge. A negative value of S_{ni} corresponds to $\exp(-S_{ni})$ being greater than 1, while a positive value corresponds to S_{ni} being less than 1. If the multiplication has the negative result, the atom under consideration supports the ligand binding, and acts as the AG group; otherwise, the positive result impedes the ligand binding and acts as the APS group.

DISCUSSION

According to this approach, the activity of the molecule in one row of Table 4 was quantitatively calculated in Eq.5 depending on the heat of formation of the conformers and the values of the atomic charge in the signed positions of each conformer. In this way although we could calculate the activities, could not easily show the attribution of the signed positions in the Table 4. Even if it was difficult to show quantitatively the contribution of each group to the activity in this calculation, we were able to make the following qualitative interpretation: If conformers of a molecule involved a functional group in only one position its activity was to be increase or decrease according to the activity. This functional group could be interpreted as the AG or the APS, respectively. But if there were more than one position in the molecule, we were not able to determine their effects. In the same way, the contribution of a- and b-positions was not defined because they were not studied separately in one molecule. If some molecules had only one of these positions, it was possible to say whether they increased or decreased the activity. If any position was presented in the conformers of only one molecule, the charge of this atom of the conformer gave some information about the interaction point of the receptor corresponding to the position. By comparing the activities of the reference molecule (n01) and the nth molecule, it might be possible to understand the contribution of different groups in these molecules. Increasing or decreasing the activities depends on the atomic charge in the defined positions. According to this approach, we were able to explain the contribution of the positions as follows: Since the conformers of the molecules such as n01, n02

and n06 had the functional groups consisting of a-, b- and c-positions together with Pha, their activities were calculated as the same value of 0.57. Since the calculated activities of the molecule n03 and n15 were less than that of the previous molecules whose all conformers possessed a- and b-positions (not c-position), we could give information about c-position as a reason for different activity between two groups of molecules. The atoms of c-position acted as an AG in the molecules. For the molecules (n04-n05, n07-n10 etc.), since some conformers of them did not have Pha, their activities should be lower than the others. The activities of the molecules such as n13-n14 which contain d-position together with a- and b-positions decreased with respect to the activity of the reference. It could be interpreted that the atoms in the d-position might act as an APS group.

The 4D-QSAR model was resulted through 30 ligands in the training set and validated by their LOO-CV errors, and was used to predict the activity of 7 samples in the test set. The relationships between the calculated and experimental RBR for molecules in both training and test set were shown in Figure 3.
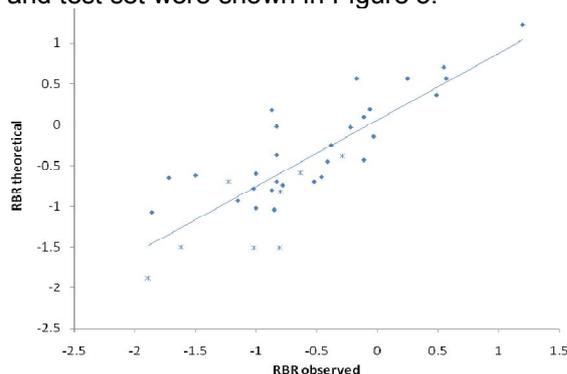

Figure 3: Correlation between experimental and calculated data sets on estrogenic activity given in training set▪ and test set*

Using the electronic and geometric properties in the conformers, the results obtained from 4D-QSAR analysis are statically interpreted. The study showed that taking into account the conformers according to the descriptors (Pha, AG and APS) was a useful approach. Only topological descriptors were used to produce an objective description of the fit. Under these conditions it might be assumed that appropriately chosen and well-evaluated molecular features were unique for the estrogenic activity. The defined topological descriptors were indeed responsible for the activity and it was well-established. The model arising from Pha, AG and APS was compatible with three dimensional structure of the receptor environment, and could summarize the information of closely related descriptors. The correlations coefficients of the

model for the training and test set (respectively, R^2 = 0.703 and Q^2 = 0.573) show that it can be useful for computer aided drug design of the understudy molecules. In our study, three successful strategies of the MCET method were designed and applied to build a 4D-QSAR model. These are (1) the prediction of activities, (2) a simple illustration of the structure-properties relationships for the activity; and (3) the description of the ligand-receptor interaction points.

CONCLUSION

We have developed the model based on the atomic charge of the ligand, and it's the best-fit topology. The model refers to pure computational methods (electronic structure calculations and matrix comparisons) and provides the correct solution (within the assumptions of the method, experimental uncertainty, and computational approximations). The computational method establishing 4D-QSAR in respect to the three-dimensional orientation of substituents, which increase or decrease the activity, in each molecular conformer has been analyzed and discussed. A different procedure from other QSAR approaches was used in the nature of the interactions between the conformers of the ligand and the target protein.

ACKNOWLEDGEMENT

REFERENCES

1. Schönfelder G, Friedrich K, Paul M, Chahoud I, Developmental effects of prenatal exposure to bisphenol A on the uterus of rat offspring. Neoplasia, 2004; 6: 584–594.
2. Gaido KW, Leonard LS, Lovell S, Gould JC, Babai D, Portier CJ, Mc Donnell D.P, Evaluation of chemicals with endocrine modulating activity in a yeast-based steroid hormone receptor gene transcription assay. Toxicol Appl Pharmacol, 1997; 143: 205–212.
3. Kavlock RJ, Daston GP, DeRosa C, Fenner-Crisp P, Gray LE, Kaattari S, Lucier, G, Luster M, Mac MJ, Maczka C, Miller R, Moore J, Rolland R, Scott G, Sheehan DM, Sinks T, Tilson HA. Research needs for the risk assessment of health and environmental effects of endocrine disruptors. Environ. Health Perspect. 1996; 104: 715-740.
4. Sippl W, Holtje H. Structure-based 3D-QSAR emerging the accuracy of structure-based alignments with the computational efficiency of ligand- based methods. J Mol Struct. 2000; 503: 31–50.
5. Vyas V, Jaın A, Gupta A. Virtual Screening: A Fast Tool For. Drug Design Sci Pharm. 2008; 6: 33–360.
6. Hopfinger AJ, Wang S, Tokarski JS, Jin B, Albuquerque M, Madhav PJ, Duraiswami C. Construction of 3D-QSAR Models Using the 4D-QSAR Analysis

Formalism. *J. Am. Chem. Soc.* 1997; 119: 10509-10524.

7. Yilmaz H, Güzel Y, Önal Z, Altıparmak G, Kocakaya OS. 4D-QSAR Study of p56lck Protein Tyrosine Kinase Inhibitory Activity of Flavonoid Derivatives Using MCET Method, Bull. Korean Chem. Soc., 201; 12: 4352-4360.

8. Gao H, Katzenellenbogen J, Garg R, Hansch C. Comparative QSAR analysis of estrogen receptor ligandsChem. Rev. 1999; 99: 723-744.

9. SPARTAN, Version 08.Wavefunction,Inc. 2008.

10. Bersuker IB, Bahçeci S, Boggs, JE, Pearlman RS. A novel electron-conformational approach to molecular modeling for QSAR by identification of the pharmacophore and anti-pharmacophore screening, SAR and QSAR in Env. 1999; 10: 157-173.

11. Bersuker IB. Pharmacophore Identification and Quantitative Bioactivity Prediction Using the Electron-Conformational Method, Curr. Pharm. Design. 2003; 9: 1575-1606.

12. Guzel Y, Ozturk E. Study of the Binding Affinity for Corticosteroid-Binding Globulin (CBG) Using the Electron Topological Method (ETM) as Three-Dimensional Quantitative Structure–Activity Relationship.Bioorganic & Medicinal Chem. 2003; 11: 4383–4388.

13. Guner O.F. Pharmacophore Perception Development and Use in Drug Design; International University Line, La Jolla, California. 2000.

14. Güzel Y, Saripinar E, Yildirim I. Electron Topological Investigation of structure antagonist activity of a series of di benzo[a,d]cycloalkenimines, J. Mol. Struc.-Theochem 1997; 418: 83–91.

15. Bersuker I.B, Dimoglo A.S. Reviews in Computational Chemistry. In Lipkowits K.B. and Body D.B Eds. 1991; 2: 423-460.

16. Dimoglo AS, Beda AA, Shvets NM, Gorbachov MYu, Kheifits LA, Aulchenko IS. Investigation of the relationship between sandalwood odor and chemical structure: electron-topological approach , New. J. Chem. 1995; 19: 149-154.

17. Dimoglo AS, Gorbachov MYu, Bersuker IB, Greni AI, Vysotskaya LE, Stepanova OV, Lukash EYu. Structural and electronic origin of meat odour of organic hetero-atomic compounds, Die Nahrung. 1988; 32: 461-473.

18. Guzel Y. Investigation of the relationship between the inhibitory activity of glycolic acid oxidase (GAO) and its chemical structure: electron-topological approach. J. Molecular Structure. 1996; 366: 131-137.

19. Yanmaz E, Sarıpınar E, Şahin K, Geçen N, Çopur F. 4D-QSAR Analysis and Pharmacophore Modeling: Electron Conformational-Genetic Algorithm Approach for Penicillins, Bioorganic & Medicinal Chemistry.2011; 19: 2199-210.

20. Tjalling JY. Historical development of the Newton-Raphson method, SIAM Review 1995; 37: 531–551.

21. Verma J, Khedkar V.M, Coutinho EC. 3D-QSAR in Drug Design, Current Topics in Medicinal Chemistry, 2010; 10: 95-115.

22. Ponomarev IV, Deych LI, Shuvayev VA, Lisyansky AA. Self-consistent approach for calculations of exciton binding energy in quantum wells., Physica E. 2005; 25: 539–553

23. Melanie M, An Introduction to Genetic Algorithms, MIT Press, London. 1999.

24. Cho DH, Lee SK, Kim BT. Quantitative structure-activity relationship (QSAR) study of new fluorovinyloxyacetamides, Bull. Korean Chem. Soc. 2001; 22: 388 –394.

25. Consonni V, Ballabio D, Todeschini R. Comments on the definition of the Q2 parameter for QSAR validation, J. Chem. Inf. Model. 2009; 49: 1669–1678.