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

Hayriye Yılmaz¹, Mehmet Boz², Burçin Türkmenoğlu² and Yahya Güzel²
¹Faculty of Pharmacy, ²Faculty of Science, Department of Chemistry, Erciyes University, Kayseri, 38039, Turkey.

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

Received: 4 December 2012 Revised accepted: 19 October 2013

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>
<thead>
<tr>
<th>Molecule no.</th>
<th>Substituents</th>
<th>log RBR (unit/mol)</th>
<th>Observed</th>
<th>Calculated from Eq 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01</td>
<td>H</td>
<td>C_2H_5,C_2H_5</td>
<td>0.57</td>
<td>0.57</td>
</tr>
<tr>
<td>02</td>
<td>H</td>
<td>C_3H_7,C_3H_7</td>
<td>0.25</td>
<td>0.57</td>
</tr>
<tr>
<td>03</td>
<td>H</td>
<td>i-C_4H_9,Me</td>
<td>-0.03</td>
<td>-0.14</td>
</tr>
<tr>
<td>04</td>
<td>H</td>
<td>C_3H_7,Me</td>
<td>-0.11</td>
<td>-0.43</td>
</tr>
<tr>
<td>05</td>
<td>H</td>
<td>Et,Et</td>
<td>-0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>06</td>
<td>H</td>
<td>C_4H_9,C_3H_7</td>
<td>-0.17</td>
<td>0.57</td>
</tr>
<tr>
<td>07 *</td>
<td>2-Me</td>
<td>(Et)_2CH,H</td>
<td>-0.29</td>
<td>-0.38</td>
</tr>
<tr>
<td>08</td>
<td>H</td>
<td>Et,Me</td>
<td>-0.38</td>
<td>-0.25</td>
</tr>
<tr>
<td>09</td>
<td>2-Me</td>
<td>i-C_4H_9,H</td>
<td>-0.41</td>
<td>-0.45</td>
</tr>
<tr>
<td>10</td>
<td>H</td>
<td>CH_2C_6H_5,H</td>
<td>-0.46</td>
<td>-0.64</td>
</tr>
<tr>
<td>11</td>
<td>2-Me</td>
<td>(C_3H_7)_2CH</td>
<td>-0.52</td>
<td>-0.70</td>
</tr>
<tr>
<td>12</td>
<td>2-Me</td>
<td>C_3H_7,H</td>
<td>-0.78</td>
<td>-0.74</td>
</tr>
<tr>
<td>13 *</td>
<td>H</td>
<td>Me,Me</td>
<td>-0.80</td>
<td>-0.82</td>
</tr>
<tr>
<td>14</td>
<td>H</td>
<td>C_2H_5,Me</td>
<td>-0.83</td>
<td>-0.02</td>
</tr>
<tr>
<td>15</td>
<td>2-Me</td>
<td>i-C_3H_7,H</td>
<td>-0.83</td>
<td>-0.70</td>
</tr>
<tr>
<td>16</td>
<td>2-Me</td>
<td>C_6H_13,H</td>
<td>-0.85</td>
<td>-1.04</td>
</tr>
<tr>
<td>17</td>
<td>H</td>
<td>C_6H_5,H</td>
<td>-0.87</td>
<td>-0.80</td>
</tr>
<tr>
<td>18</td>
<td>3-Me</td>
<td>H,H</td>
<td>-1.00</td>
<td>-0.60</td>
</tr>
<tr>
<td>19 *</td>
<td>H</td>
<td>C_4H_9,C_4H_9</td>
<td>-1.02</td>
<td>-1.51</td>
</tr>
<tr>
<td>20</td>
<td>2-Me</td>
<td>Me,H</td>
<td>-1.02</td>
<td>-0.78</td>
</tr>
<tr>
<td>21</td>
<td>2-Me</td>
<td>Et,H</td>
<td>-1.15</td>
<td>-0.93</td>
</tr>
<tr>
<td>22 *</td>
<td>H</td>
<td>(C_6H_5)_2CH,H</td>
<td>-1.62</td>
<td>-1.15</td>
</tr>
</tbody>
</table>
Table 1 (continued): Specific estrogenic activity of 4,4′-Dihydroxydiphenylmethanes (units/mol) in rats

<table>
<thead>
<tr>
<th>Molecule no.</th>
<th>Substituents</th>
<th>log RBR (unit/mol)</th>
<th>Observed</th>
<th>Calculated from Eq 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>H</td>
<td>C₆H₅,Me</td>
<td>-1.86</td>
<td>-1.07</td>
</tr>
<tr>
<td>24</td>
<td>3-Me</td>
<td>C₃H₇,C₃H₇</td>
<td>1.20</td>
<td>1.22</td>
</tr>
<tr>
<td>25 *</td>
<td>3-Me</td>
<td>C₃H₇,Et</td>
<td>-0.63</td>
<td>-0.59</td>
</tr>
<tr>
<td>26</td>
<td>3-Me</td>
<td>Et,Et</td>
<td>0.55</td>
<td>0.71</td>
</tr>
<tr>
<td>27</td>
<td>3-Me</td>
<td>C₄H₉,C₃H₇</td>
<td>0.49</td>
<td>0.37</td>
</tr>
<tr>
<td>28</td>
<td>3-Me</td>
<td>C₃H₇</td>
<td>-0.06</td>
<td>0.19</td>
</tr>
<tr>
<td>29</td>
<td>3-Me</td>
<td>Me,i-C₄H₉</td>
<td>-0.22</td>
<td>-0.03</td>
</tr>
<tr>
<td>30 *</td>
<td>3-Me</td>
<td>Me,Me</td>
<td>-0.81</td>
<td>-1.51</td>
</tr>
<tr>
<td>31</td>
<td>3-Me</td>
<td>Me,Et</td>
<td>-0.83</td>
<td>-0.37</td>
</tr>
<tr>
<td>32</td>
<td>3-Me</td>
<td>Me,C₄H₉</td>
<td>-0.87</td>
<td>0.18</td>
</tr>
<tr>
<td>33</td>
<td>3-Me</td>
<td>H,i-C₄H₉</td>
<td>-1.00</td>
<td>-1.02</td>
</tr>
<tr>
<td>34 *</td>
<td>3-Me</td>
<td>C₄H₉,C₃H₉</td>
<td>-1.23</td>
<td>-0.70</td>
</tr>
<tr>
<td>35</td>
<td>3-Me</td>
<td>Et,H</td>
<td>-1.50</td>
<td>-0.62</td>
</tr>
<tr>
<td>36</td>
<td>3-Me</td>
<td>Me,C₆H₅</td>
<td>-1.72</td>
<td>-0.65</td>
</tr>
<tr>
<td>37 *</td>
<td>3-Me</td>
<td>H,C₆H₁₃</td>
<td>-1.89</td>
<td>-1.88</td>
</tr>
</tbody>
</table>

*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]:

1. Suppose that we have $m_2$, which is the number of local atomic properties.
2. 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 $m_2$, which is the number of such characteristics, then the number of all ordinary $k^k$ matrices is $m=m_1*m_2$ where $k$ is the number of atoms, $m_1$ is the number of all acceptable conformers of each molecule, 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:

$$P_a = \frac{v_1+1}{v_1+3}$$

where $v_1$ and $v_2$ are the numbers of molecules possessing and not possessing, respectively, the feature of activity in the class of molecules; $v_3$ and $v_4$ have analogous meaning in low active molecules; and $\mu_1$ and $\mu_2$ are the numbers of molecules in the class of active and low active molecules $\mu_3= v_1 + v_3$ and $\mu_4= v_2 + v_4$. This way $P_a$ evaluates only the deposit of active molecules, while $P_a$ 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 $S_\text{ni}$ to take into account the independent variables ($a_{ni}^{(j)}$) and parameters ($\kappa_i$) of $J$-positions in Eq.2 as follows [11]:

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

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

$$A_n = A_0 \frac{\sum_{i=1}^{m_n} e^{-S_i} e^{-E_{ai}/kT}}{\sum_{i=1}^{m_n} e^{-E_{ai}/kT}} \quad \text{Eq.3}$$

$a_0$ in Eq. 3 is a constant (see below), $m_n^\text{Pha}$ 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 (I) is chosen from the training set, for which the activity is known as $A_i$ in Eq. 4:

$$A_i = A_0 \frac{\sum_{i=1}^{m_i^\text{Pha}} e^{-S_i} e^{-E_{ai}/kT}}{\sum_{i=1}^{m_i} e^{-E_{ai}/kT}} \quad \text{Eq.4}$$

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:

- 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:

$$A_n = \frac{\sum_{i=1}^{m_n} e^{-E_{ai}/kT} \sum_{i=1}^{m_n} e^{-S_i} e^{-E_{ai}/kT}}{\sum_{i=1}^{m_n} e^{-E_{ai}/kT} \sum_{i=1}^{m_n} e^{-E_{ai}/kT}} \quad \text{Eq.5}$$

The functions of $S_\text{ni}$ and $S_\text{ni}$ 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 $\kappa_i$ in Eq.2 have been...
simultaneously estimated using Newton-Raphson approximation [20]. The predicted $\kappa_j$ 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 $\kappa_j$ is determined from the condition of minimum of the sum $\sum_{n=1}^{N} | A_n^{\text{calc}} - A_n^{\text{exp}} |^2$, where $N$ is the number of molecules in a training set. 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]:

$$R^2 = 1 - \frac{[\sum_{i=1}^{n_{\text{test}}} (\hat{y}_i - y_i)^2] / n_{\text{test}}}{[\sum_{i=1}^{n_{\text{test}}} (y_i - \bar{y}_{\text{test}})^2] / n_{\text{test}}} = 1 - \frac{\text{PRESS}}{\text{TSS}}$$  

where $n_{\text{test}}$ is the total number of objects in the entire test set, $\hat{y}_i$ is the predicted value for the $i^{th}$ 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]:

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

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 $n_{\text{test}}$ 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 $x_1=0; y_1=0; z_1=0; x_2=0; y_2=0; z_2=0$ and $x_3=0; y_3=0; z_3=0$, respectively. A Pha needs at least three points of connection, but there can be more. A complete Pha can be represented geometrically as a
Table 2: The observed and calculated data related to the molecules of training set understudy

<table>
<thead>
<tr>
<th>Mol Obs. Calc.</th>
<th>The heat of formation (kcal/mol) and AG/APS positions of conformers.</th>
</tr>
</thead>
<tbody>
<tr>
<td>n01</td>
<td>0.57  0.57 -265.53ab -265.53abc</td>
</tr>
<tr>
<td>n02</td>
<td>0.25  0.57 -293.78abc -293.78ab</td>
</tr>
<tr>
<td>n03</td>
<td>-0.03 -0.14 -249.41ab -249.41abc</td>
</tr>
<tr>
<td>n04</td>
<td>-0.11 -0.43 -247.70ab -247.70abc</td>
</tr>
<tr>
<td>n05</td>
<td>-0.11 0.10 -237.22abc §-237.08 §-236.91</td>
</tr>
<tr>
<td>n06</td>
<td>-0.17 0.57 -322.37abc §-322.37ab</td>
</tr>
<tr>
<td>n07</td>
<td>-0.29 -0.38 -311.59ab §-311.42ab §-311.42ab -311.38ab -311.19ab</td>
</tr>
<tr>
<td>n08</td>
<td>-0.38 -0.25 -219.46abc -219.46abc §-216.31</td>
</tr>
<tr>
<td>n09</td>
<td>-0.41 -0.45 -288.57abc §-288.56abc §-285.65</td>
</tr>
<tr>
<td>n10</td>
<td>-0.46 -0.64 §-90.53 §-90.53 §-90.45 §-90.38ab -89.01ab</td>
</tr>
<tr>
<td>n11</td>
<td>-0.52 -0.70 -370.31abc -370.31abc §-370.07ab §-370.00 §-369.85</td>
</tr>
<tr>
<td>n12</td>
<td>-0.78 -0.74 -284.41abc §-284.09ab §-283.90 §-283.76ab §-283.59</td>
</tr>
<tr>
<td>n13</td>
<td>-0.80 -0.82 -201.67abd -201.48abd §-201.33</td>
</tr>
<tr>
<td>n14</td>
<td>-0.83 -0.02 -276.29ab §-276.11ab §-276.07abd §-275.93 §-275.87 ab</td>
</tr>
<tr>
<td>n15</td>
<td>-0.80 -0.70 -267.55ab §-267.49ab</td>
</tr>
<tr>
<td>n16</td>
<td>0.65 -1.04 -370.00abd -369.67abd §-369.50 §-369.45 §-369.45ab</td>
</tr>
<tr>
<td>n17</td>
<td>-0.67 -0.80 -57.14ab -57.14abcd §-57.11 §-57.11 §-57.11ab §-57.09ab</td>
</tr>
<tr>
<td>n18</td>
<td>-1.00 -0.60 §-223.99 §-222.40 §-222.39 §-222.38 §-222.16 §-222.16</td>
</tr>
<tr>
<td>n19</td>
<td>-1.02 -1.51 §-350.96abc §-350.78abc §-350.77abc §-350.75abd §-347.61</td>
</tr>
<tr>
<td>n20</td>
<td>-1.02 -0.78 §-230.89ab §-230.89ab §-230.23abcd §-230.23abcd</td>
</tr>
<tr>
<td>n21</td>
<td>-1.15 -0.93 §-255.16abc §-255.12abcd §-254.94ab §-254.89abc</td>
</tr>
<tr>
<td>n22</td>
<td>-1.16 -1.15 §-244.12 §-244.53abc §-244.57abd §-244.45 §-244.12ab</td>
</tr>
<tr>
<td>n23</td>
<td>-1.86 -1.07 §-44.53 §-44.53ab §-44.47ab §-44.45 §-44.43ab §-44.31ab §-44.31ab §-43.41ab §-43.40ab</td>
</tr>
<tr>
<td>n24</td>
<td>1.20 1.22 §-321.77abd §-321.66 §-321.66 §-321.66 §-321.66 §-321.66</td>
</tr>
<tr>
<td>n25</td>
<td>1.13 -0.59 §-295.55 §-295.32 §-295.32 §-295.32 §-295.32 §-295.32</td>
</tr>
<tr>
<td>n26</td>
<td>0.55 0.71 §-267.32abd §-267.08 §-266.92ab §-266.76 §-265.28ab §-265.17 §-265.07 §-264.99</td>
</tr>
<tr>
<td>n27</td>
<td>0.49 0.37 §-352.39ab §-352.16 §-352.16 §-352.16 §-352.16 §-352.16</td>
</tr>
<tr>
<td>n28</td>
<td>0.06 0.19 §-277.73ab §-277.73ab §-277.52 §-277.50ab §-277.35ab §-277.21 §-277.18 §-276.68ab</td>
</tr>
<tr>
<td>n29</td>
<td>-0.22 -0.03 §-280.92ab §-280.67ab §-280.48ab §-280.35 §-280.24abc §-279.95 §-279.84ab §-279.74 §-278.88ab</td>
</tr>
<tr>
<td>n30</td>
<td>-0.81 -1.51 §-231.71abd §-231.50ab §-231.34abcd §-231.19 §-229.68ab §-229.58 §-229.49ab §-229.42</td>
</tr>
<tr>
<td>n31</td>
<td>-0.83 -0.37 §-249.48ab §-249.44abc §-249.11ab §-249.10ab §-248.84abc §-247.77 §-247.75 §-247.27 §-246.42ab</td>
</tr>
<tr>
<td>n32</td>
<td>-0.87 0.18 §-306.33abc §-305.92abc §-305.77 §-305.49abc §-304.71 §-304.12 §-303.88abc §-303.14ab</td>
</tr>
<tr>
<td>n33</td>
<td>-1.00 -1.02 §-297.07ab §-296.82ab §-296.72 §-296.70ab §-296.46ab §-295.12abc §-294.92 §-294.90ab §-294.88</td>
</tr>
<tr>
<td>n34</td>
<td>-1.23 -0.70 §-380.75abd §-380.95ab §-378.84 §-378.02abcd §-378.01abc §-377.78 §-377.78ab §-377.74ab §-377.74ab</td>
</tr>
<tr>
<td>n35</td>
<td>-1.50 -0.62 §-259.47abc §-259.45ab §-259.17ab §-259.13ab §-259.09 §-259.09 §-257.82ab §-257.81abc §-257.50</td>
</tr>
<tr>
<td>n36</td>
<td>-1.72 -0.65 §-74.42 §-74.42ab §-74.35ab §-74.32 §-74.24 §-74.22ab §-74.14 §-73.95 §-72.47</td>
</tr>
<tr>
<td>n37</td>
<td>-1.89 -1.88 §-374.01ab §-372.09abd §-372.09abd §-371.82 §-371.37ab §-371.03ab §-370.23abd</td>
</tr>
</tbody>
</table>

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.5 by 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.

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.2Sni 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

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

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, $R = 0.703$ 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**

This work was financially supported by Erciyes University Scientific Research Projects (BAP) of Turkey (Grant no. FBD-10-2983).

**REFERENCES**


