



Pergamon

Study of the Binding Affinity for Corticosteroid-Binding Globulin (CBG) Using the Electron Topological Method (ETM) as Three-Dimensional Quantitative Structure–Activity Relationship (3D QSAR)

Yahya Guzel* and Emel Ozturk

Erciyes University, Faculty of Science and Art, Departments of Chemistry, 38039 Kayseri, Turkey

Received 3 January 2003; accepted 10 July 2003

Abstract—The Electron Topological Method, called ETM, is a descriptor for predicting the biological activities of molecules based on three-dimensional quantitative structure–activity relations (3D QSAR). ETM uses a modified electron topological state index to substitute for electronic properties and a topological distance for the relative distance in the molecule. It is shown that the molecular fragments responsible for this activity possess fixed electronic and geometric characteristics associated with a distinct arrangement and the steric accessibility of an oxygen atom and a group of carbon atoms. After that, it is essential to employ a linear regression analysis technique to derive a 3D QSAR model relating the biological activities to the ETM. The ETM is used to study the 3D QSAR of the corticosteroid-binding globulin (CBG) binding affinity to 31 steroids, and resulting models have a comparable to current 3D methods such as CoMFA. Though the ETM is a descriptor based on 3D topological information obtained by quantum chemical derived descriptors, give the best answer for both the similarity analysis and the statistical fitting.

© 2003 Elsevier Ltd. All rights reserved.

Introduction

Strategy involves the identification of the pharmacophoric pattern, by means of a similarity analysis where the structural and electronic characteristics are compared, followed by a quantitative structure–activity relationship (QSAR) study. For the relationship between the chemical structure of a compound and its biological or pharmacological activity, compounds are often classed together because they have structural characteristics in common including shape, size, stereochemical arrangement, and distribution of functional groups. Other factors contributing to structure–activity relationship include steric effects, chemical reactivity, electronic effects, resonance, and inductive effects.

QSAR techniques have become very important in all aspects of research into the molecular interpretation of

biological properties.¹ There are many different methods aimed at this purpose but nearly all of them can be interpreted as searching for similarity among chemical structures. Almost all drugs, if not found by chance, result from the optimization of lead structures. It has become evident that the physical, chemical, or biological properties of a compound depend on the 3D arrangements of the atoms in the molecule. The ability to produce quantitative correlation between 3D properties of the molecules and the biological activity of these compounds is of inestimable value in deciding upon the choice of future synthetic chemistry.² Among the several widespread techniques used for this purpose are comparative molecular field analysis (CoMFA) introduced by Cramer³ and used by several other authors^{4,5} and the electron topological method (ETM) introduced by bersuker.⁶

The ETM offers a significant advance both in understanding a model of a molecule's biological interaction and in predicting the molecule's activity.⁷ A set of molecules may be compared to a single reference molecule to yield a predictive QSAR.^{8,9} Similarity between pairs of molecules can be defined in terms of shape and

*Corresponding author. Tel.: +90-536-315-9256; fax: +90-352-437-4933; e-mail: yahyag@yahoo.com

of electronic properties of molecules, but instead of a large number of values at comparisons of diagonal matrix representing each molecule in ETM it has a single numerical measure of the overall similarity, comprising the active fragment in only active compounds. More information can be obtained from a matrix of the similarity indices between all pairs of molecules in a training set using the statistical analysis to derive a 3D QSAR. Assuming that for the bioactivity under consideration there is only one (unique) Pharmacophore (Pha), it can be defined as a group of specific atoms in a given geometric arrangement which represent the bioactivity under consideration in a series of compounds.¹⁰ To extract the relationship between one fragment in nearly all active molecules and the biological activity, statistical analysis is used by Baroni and Cruciani.^{11,12} The resulting models in ETM have a comparable quality to the current 3D methods such as CoMFA.

ETM Study Approach

The interactions between structure and activity can be quickly identified. More precisely, m -dimensional $n \times n$ electron-topological matrices of contiguity (ETMC) are taken as a language for compound description (one matrix = one compound), where n is the number of atoms in the corresponding molecule (note that any m -dimensional $n \times n$ matrix may be viewed as a set of m ordinary $n \times n$ matrices).

ETMC formation proceeds in the following way;

1. Let A_i be an atom of a molecule being described, then the corresponding diagonal matrix element

a_{ii} is one of m_1 local atomic characteristics (for example, charge, polarizability, HOMO- and LUMO-orbital coefficients, etc.)

2. If A_i and A_j are any two atoms of the molecule 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 parameters (bond length, bond energy, etc.). Suppose that we have m_2 such characteristics.

As a result, the number of all ordinary $n \times n$ matrices which can be formed on the basis of the data known is $m = m_1 \times m_2$ (m -dimensional ETMC). It has been assumed that a single electronic matrix in ETMC represents a configuration of each molecule. m is taken as one-dimensional, containing charge on the diagonal, bond length and distance on the non-diagonal as shown in Figure 1.

The computational part of the ETM may be considered as consisting of the following steps:

1. Conformational analysis.
2. Quantum-chemical calculations.
3. ETMC formation.
4. ETMC processing (the search of the structural features of activity/inactivity or weak activity).

In the first step, Spartan's *molecular mechanics force field* (MMFF) is used, because it presently provides for the calculation of equilibrium geometries, strain energies and normal-mode vibrational frequencies, as well as for searching of conformation space for both cyclic and

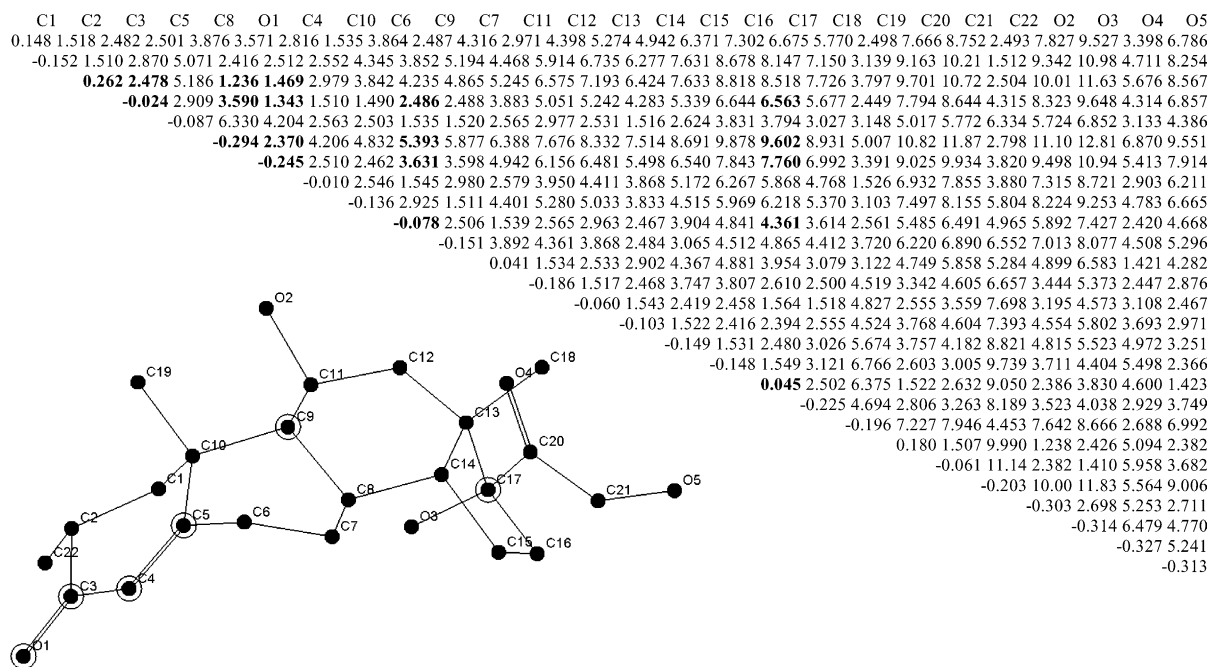
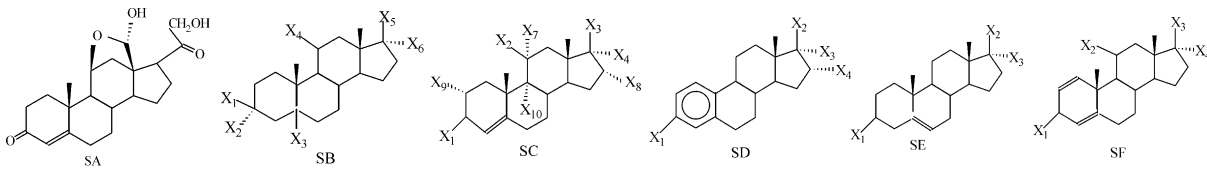


Figure 1. The ETMC of the N30 compound consisting of 27-atoms without hydrogen in Table 1. The bold entries in the matrix are the significance showing Pha. The electronic value of hydrogen bonding with carbon is not taken since in each molecule it has an equivalent effect and the H-substituent value at each position is set equal to zero.¹⁶

Table 1. Structure and CBG affinity data for steroid series 1-31 of the SA-SF structure^a


N	Steroid	S	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀
1	Aldosterone	SA										
2	Androstenediol	SB	OH	H	H		OH	H				
3	Androstenediol	SE	OH	OH	H ^a							
4	Androstenedione	SC	=O	H	=O				H	H	H	H
5	Androsterone	SB	H	OH	H ^a		=O					
6	Corticosterone	SC	=O	OH	COCH ₂ OH				H	H	H	H
7	Cortisol	SC	=O	OH	COCH ₂ OH				H	H	H	H
8	Cortisone	SC	=O	=O	COCH ₂ OH					H	H	H
9	Dehydroepiandrosterone	SE	OH	=O								
10	Deoxycorticosterone	SC	=O	H	COCH ₂ OH				H	H	H	H
11	Deoxycortisol	SC	=O	H	COCH ₂ OH				H	H	H	H
12	Dihydrotestosterone	SB	=O		H ^a		OH	H				
13	Estradiol	SD	OH	OH	H							
14	Estriol	SD	OH	OH	H							
15	Estrone	SD	OH	=O								
16	Etiocolanolone	SB	H	OH	H ^b		=O					
17	Pregnenolone	SE	OH	COMe	H							
18	17-Hydroxypregnenolone	SE	OH	COMe	OH							
19	Progesterone	SC	=O	H	COMe				H	H	H	H
20	17-Hydroxyprogesterone	SC	=O	H	COMe				H	H	H	H
21	Testosterone	SC	=O	H	OH				H	H	H	H
22	Prednisolone	SF	=O	OH	COCH ₂ OH							
23	Cortisol-21-acetate	SC	=O	OH	COCH ₂ OCOMe							
24	4-Pregnene-3,11,20-trione	SC	=O	=O	COMe							
25	Epicorticoesterone	SC	=O	H	COCH ₂ OH				OH	H	H	H
26	19-Nortestosterone	SC ^c	=O	H	OH				H	H	H	H
27	16 α ,17-Dihydroxy-4-pregnene-3,20-dione	SC	=O	H	COMe				H	OH	H	H
28	16 α -Methyl-4-pregnene-3,20-dione	SC	=O	H	COMe				H	Me	H	H
29	19-Norprogesterone	SC ^c	=O	H	COMe				H	H	H	H
30	11 β ,17,21-Trihydroxy-2 α -methyl-4-pregnene-3,20-dione	SC	=O	OH	COCH ₂ OH				H	H	Me	H
31	11 β ,17,21-Trihydroxy-2 α -methyl-9 α -fluoro-4-pregnene-3,20-dione	SC	=O	OH	COCH ₂ OH				H	H	Me	F

^a5- α .^b5- β .^cH instead of Me at the C₁₀.

acyclic molecules.^{13,14} In the second step, AM1 is selected in Spartan's semi-empirical modul which provides for the calculation of heats of formation, equilibrium and transition-state geometries, the charge on each atom and so on.¹⁵ The next two steps represent the essential part of the ETM. Let C₁, C₂, ..., C_k be a series of compounds under investigation, their numbers of atoms being n_1, n_2, \dots, n_k , correspondingly. For each compound its activity is to be known and evaluated quantitatively so that the compound could be classified as belonging to a definite class of activity (there must be two such classes, at least). Suppose that the first two kinds of calculations were done and that for every compound C_j, a corresponding m -dimensional ETMC was formed.

Then, the procedure for searching the fragment of activity (Pha) is:

- To form ordinary $n_j \times n_j$ matrices after choosing and fixing the type of characteristics both for the diagonal and non-diagonal elements. We select out charge as diagonal and bond length between the bonded atoms and distance between the unbounded as non-diagonal.

- To set the ranges of variation of the diagonal (d1) and non-diagonal (d2) elements.
- To set the level of activity which classifies all compounds into two classes (active and weakly active compounds).
- To chose one compound as the template compound (as a rule, the most active and simple common one) and to compare it with other compounds from the series given. The result is a set of ETM sub-matrices (ETMS) common to all (or nearly all) active compounds.
- To evaluate the probabilities (α_a and P_a —see Fig. 2) of the features' realization in the class of active compounds. If they are not high enough, to repeat the procedure described with other parameters (and with other types of characteristics, possibly) until the values of α_a and P_a for the fragments newly found will be satisfactorily high.

As an example, the ETMC of compound N30 is shown in Figure 1. It is formed from effective charges on atoms (Q_{ii}), the bond length (B_{ij}) and optimized distances between atoms in the molecule (R_{ij}) (H-atoms are not

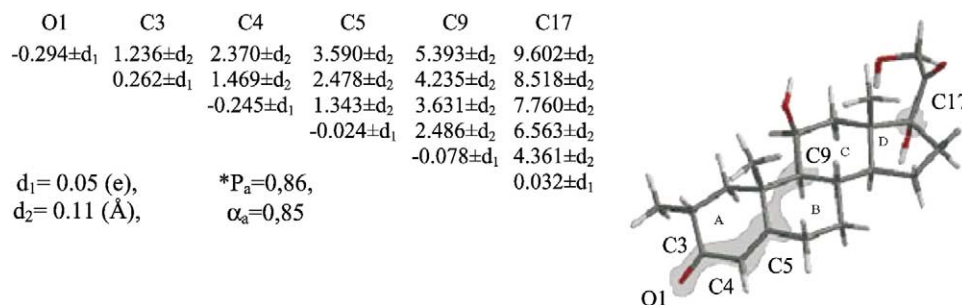


Figure 2. The active fragment (Pha = O1, C3, C4, C5, C9 and C17) responsible for CBG affinity, contains one oxygen and five carbon, more precisely, six reactivity points with limited values of charge, bond length and distance shown in the submatrix of affinity (ECSA). *P_a and α_a are the probabilities of the feature realization: $P_a = (v_1 + 1)/(v_1 + v_2 + 2)$, $\alpha_a = (v_1 \cdot v_4 - v_2 \cdot v_3)/(\mu_1 \cdot \mu_2 \cdot \mu_3 \cdot \mu_4)^{1/2}$, where v_1 and v_2 are the numbers of molecules possessing and not possessing, respectively, the feature of activity (predicted by the ETM) in the class of compounds; v_3 and v_4 have analogous meaning in the weak active compounds; μ_1 and μ_2 are the numbers of molecules in the class of active and weak active compounds; $\mu_3 = v_1 + v_3$; $\mu_4 = v_2 + v_4$.

given). The electronic characteristics are given in electron charge units (e), the bond lengths are given in Å.

We can single out 3D QSARs in statistical calculations related to the corticosteroid affinity and the properties of electron configuration (Q_{iii}, B_{ij}, R_{ij}, HOMO, LUMO, etc.)

Results and Discussion of 3D QSAR Studies on 31 Steroids

The Pha is usually defined as a group of specific atoms in a given geometric arrangement that is deemed to exert the activity under consideration in a series of compounds. To locate the Pha means to be able to identify the presence or absence of the bioactivity.

The main-skeleton of the series under investigation comprises the four rings on which molecules are superimposed. The Pha consists of O1, C3, C4, C5, C9, C17. Although O1 is bonding on A-ring, C3, C4, C5, C9 and C17 are respectively in the A-, B-, C- and D-rings (see Fig. 2). Since the Pha is found in the lowest energy conformation rather than in some other conformer for all compounds, not all conformers of the molecule are account for in the evaluation, but only lowest energy conformer is. In applying this strategy, one must recognize that one is assuming that it is the minimum energy conformers that will bind most favorably in the receptor site. In fact, there is no a priori reason to exclude higher energy conformers as the source of activity.

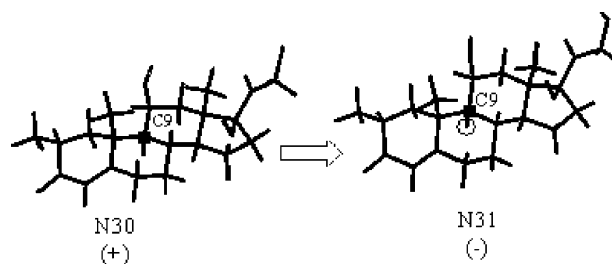
Compounds can be divided into two groups as actives (under pK = -6.0) and weak-actives (above pK = -6.0). Although nearly all-active molecules have the Pha only within given tolerance values (d₁ = 0.05, d₂ = 0.11), none of the weak-actives do. Thus, the qualitatively obtained result is exact. Electronic (charge and bond length) and geometrical (distance) properties in the Pha and other parameters (surface area, volume, electronical-, total-, HOMO- and LUMO-energy, etc.) are quantitatively investigated in statistical analysis.

For each weak-active or inactive compound, since there is at least one reason to decrease the activity such as the pharmacophore shielding, resonance effect and position

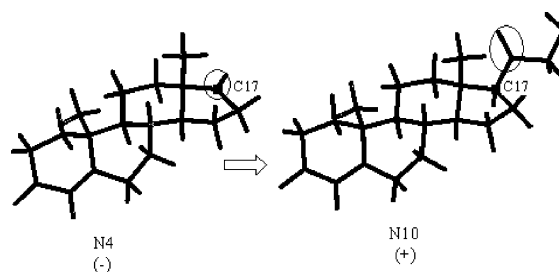
effect of each group or the substituent, none of them is taken in the quantitative calculation. Let us discuss these features on the substituent around the pharmacophore in a few examples (Schemes 1–4).

This study indicates that for maximum CBG affinity, the groups of (O1, C3, C4, C6), C9 and C17 are a requirement (see Fig. 1).

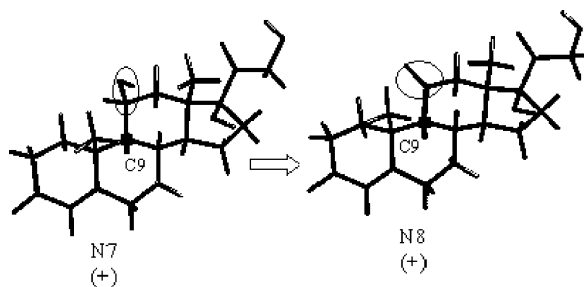
The system of prediction described (identifying the active fragment and the properties of its constituents) may be used for the preliminary selection of active compounds with CBG affinity. To continue SAR investigations, we carried out 3D QSAR analyses on the basis of quantum-chemical calculations and on the results of the ETM. The data resulting from the elec-



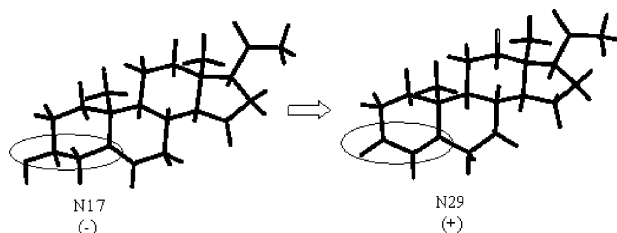
Scheme 1. In comparing between N30 (+, active) and N31 (-, weak-active) since fluorine increases the charge of C9 in N30, and has the steric effect of being a shielding group in the lipophile region, it does weak-active the N31.



Scheme 2. The effects of replacing the CO-group in molecules N1, N6–8, N10, 11, N19, 20, N22–25, N27–30 with a more electronegative oxygen in N4, 5, N9, N15, 16 will result in a change in the charge of C17 in the Pha. As shown in the following discussion, since increasing the charge of C17 decreases the activity of the molecules, this oxygen destroys the Pha.



Scheme 3. While a CO group is appended to C9 atom, included in the Pha, in N8 and N24, a COH group in N6, 7, N22, 23, N25, N30–31 appears. In place of the COH group, the CO decreases the charge of C9 from $\sim(-0.079)$ to $\sim(-0.136)$ and causes deviation from the values of C9 in the reference molecule (N30).



Scheme 4. Since the group shape of (O1, C3, C4, C6) in N27 and N29 circled, is deformed in N17 and 18 it leads to decreasing activity.

tronic structure calculations made for the series of CBG affinities were used for constructing the quantitative SAR model. Such molecular characteristics as V_M , charge of C17, $\Delta QO1$ and Pha are defined as responsible for activity from binding affinity. Numerical values of the parameters mentioned are given in Table 2.

Multistep regression analyses (with no constant) are used to provide empirical links among the observed values of pK and parameters presently calculated. The best results are obtained for Q17, $\Delta QO1$, Pha, V_M and ΔE .

$$\begin{aligned} \text{pK} = & -1.58(\pm 0.63)\text{Q17} - 6.132(\pm 2.09)\Delta\text{QO1} + 1.65 \\ & \times (\pm 0.17)\text{Pha} - 1.67\text{E-}02(\pm \text{E-}04)V_M - 3.44\text{E-}04 \\ & \times (\pm \text{E-}05)\Delta E \end{aligned}$$

Here, $R^2=0.96$, $\text{SE}=0.36$ (R is the correlation coefficient; SE is the standard error of estimate). The correlation among the experimental values of pK and the values calculated is given in Figure 3. The Pha enters with substantial weight into the equation of regression, because it is related to the realization of the electron-topological fragment of activity. The absence of the fragment (Pha=0) causes the decrease of pK. When the values of Q17, V_M , ΔE increase, the activity grows. The decrease in $\Delta QO1$ yields the growth of pK too. Results

Table 2. Experimental and calculated binding activities of chemical compounds under investigation

N ^a	Q17	$\Delta QO1$	Pha	V_M	ΔE	pK (exp) ^b	pK (calcd ETM)	ΔpK (ETM)	pK CoMFA ^c	ΔpK (CoMFA)
1	-0.128	0	0	403.000	654.450	-6.279	-6.224	-0.055		
2	0.031	-0.032	1	360.010	1815.250	-5.000	-5.104	0.104		
3	0.033	-0.323	1	355.570	1843.300	-5.000	-5.051	0.051		
4	0.238	-0.295	1	344.410	1899.700	-5.763	-5.364	-0.399		
5	0.236	-0.325	1	354.080	1843.460	-5.613	-5.314	-0.299		
6	-0.152	-0.151	0	402.690	947.140	-7.881	-7.883	0.002		
7	0.045	-0.297	0	411.740	626.760	-7.881	-7.802	-0.079		
8	0.050	-0.294	0	406.820	654.790	-6.892	-7.257	0.365		
9	0.238	-0.324	1	350.020	1871.650	-5.000	-5.267	0.267		
10	-0.150	-0.295	0	393.890	1267.520	-7.653	-7.374	-0.279		
11	0.048	-0.295	0	405.080	947.150	-7.881	-7.831	-0.05		
12	0.031	-0.295	1	353.980	1843.450	-5.919	-5.994	0.075		
13	0.032	-0.253	1	324.940	2054.620	-5.000	-4.954	-0.046	-5.367	-0.367
14	-0.008	-0.252	1	335.790	1734.010	-5.000	-5.059	0.059	-5.342	-0.342
15	0.238	-0.253	1	319.800	2082.750	-5.000	-5.285	0.285	-5.307	-0.307
16	0.236	-0.325	1	354.080	1843.460	-5.255	-5.314	0.059	-5.354	-0.099
17	-0.154	-0.326	1	389.160	1560.120	-5.255	-5.203	-0.052	-5.661	-0.406
18	0.042	-0.324	1	397.760	1239.640	-5.000	-5.026	0.026	-5.620	-0.62
19	0.091	-0.296	0	383.200	1588.220	-7.380	-7.270	-0.11	-6.957	0.423
20	0.030	-0.297	0	392.580	1267.850	-7.740	-7.707	-0.033	-6.637	1.103
21	0.034	-0.296	0	349.750	1871.500	-6.724	-6.740	0.016	-6.144	0.58
22	0.046	-0.296	0	405.570	655.290	-7.512	-7.218	-0.294	-7.040	0.472
23	0.043	-0.298	0	456.880	22.740	-7.553	-7.810	0.257	-7.104	0.449
24	-0.151	-0.295	0	387.130	1295.840	-6.779	-6.871	0.092	-6.970	-0.191
25	-0.149	-0.296	0	403.520	946.980	-7.200	-7.007	-0.193	-7.865	-0.665
26	0.034	-0.296	0	349.750	1871.500	-6.144	-6.140	-0.004	-6.383	-0.239
27	0.042	-0.296	0	392.360	1267.820	-6.247	-6.227	-0.02	-6.848	-0.601
28	-0.151	-0.296	0	403.450	1432.470	-7.120	-7.188	0.068	-6.702	0.418
29	-0.153	-0.296	0	383.220	1588.220	-6.817	-6.909	0.092	-6.212	0.605
30	0.033	-0.295	0	430.020	471.010	-7.688	-7.540	-0.148	-7.634	0.054
31	0.035	-0.292	1	435.560	0.000	-5.797	-5.867	0.07	-7.574	-1.777

pK, CBG affinity; V_M , molecular volume; Pha, is defined as the identification module in a given geometric arrangement (Pha = 1 when the activity feature enters into a compound, otherwise Pha = 0); Q, atomic charge, ΔQ and ΔE , the difference of atomic charge and total energy between the reference molecule and other molecules.

^aSee Table 1 for name and structure of molecules.

^bStructures and data according to refs 17 and 18.

^cpK CoMFA data according to ref 19.

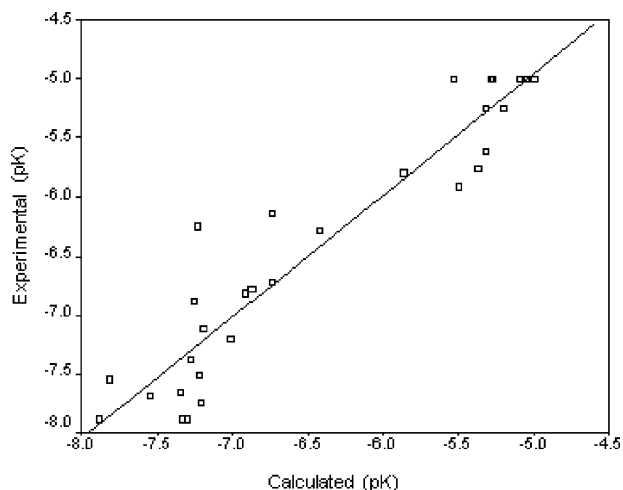


Figure 3. Correlation between experimental and theoretical data sets on pK.

obtained enable us to apply the rules stated above to the purposeful search of new compounds capable of demonstrating CBG affinity.

The most important feature of the ETM models resulting from these investigations is their potential as predictive tools for estimating the corticosteroid binding affinity and mitogenic potential of steroidal estrogens. While reasonable predictions of mitogenic potential are expected for steroids without corticosteroid binding data, it will be seen that in predictions made from ETM and CoMFA models, pK_a values are more convenient in the ETM model. So ETM should provide the most realistic estimates. These 3D QSAR models derived from ETM might also contribute to the elucidation of the ligand-binding site of the estrogen receptor. Considering the structural constraints inherent to the steroid estrogens of this study may provide a realistic steric and electrostatic mirror of portions of the estrogen receptors' hormone interaction site.

Acknowledgements

The Scientific Technical Research Council of Turkey (TUBITAK) financially supported this project (Grant No. TBAG-2113).

References and Notes

- Katritzky, A. R.; Maran, U.; Lobanov, M. J. *Chem. Inf. Comput. Sci.* **2000**, *40*, 1.
- Robinson, D. D.; Winn, P. J.; Lyne, P. D.; Richards, W. G. *J. Med. Chem.* **1999**, *42*, 573.
- Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
- DePriest, S. A. *J. Am. Chem. Soc.* **1993**, *115*, 5372.
- Klebe, G.; Abraham, U. *J. Med. Chem.* **1993**, *36*, 70.
- Bersuker, I. B.; Dimoglo, A. S.; Gorbachov, M. Yu.; Vlad, P. F.; Koltza, M. N. In Hadzi, D., Jerman-Blazic; Elsevier: Amsterdam, 1987; p 43.
- Dimoglo, A. S.; Gorbachov, M. Y.; Lesnik, T. I.; Sarıaçoğlu, M.; Güzel, Y.; Yıldırm, Y. *Curr. Med. Chem.* **1997**, *4*, 23.
- Burt, G.; Huxley, P.; Richards, W. G. *J. Comput. Chem.* **1990**, *11*, 1139.
- Seri, L. A.; Salter, R.; West, S.; Richards, W. G. *J. Med. Chem.* **1994**, *29*, 687.
- Bersuker, I. B.; Bahçeci, S.; Boggs, J. E.; Pearlman, R. S. *J. Comput. Aided Mol. Des.* **1999**, *13*, 407.
- Baroni, M.; Clementi, S.; Cruciani, G.; Costantino, G.; Riganeli, P. *J. Chemometr.* **1992**, *6*, 347.
- Cruciani, G.; Watson, K. A. *J. Med. Chem.* **1994**, *37*, 2589.
- Clark, M.; Cramer, R. D., III; van Opdensh, N. *J. Comp. Chem.* **1989**, *10*, 982.
- Halgren, T. A. *J. Comp. Chem.* **1996**, *17*, 490.
- Dewar, M. J. S.; Zebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
- Fujita, L.; Ban, C. *J. Med. Chem.* **1971**, *14*, 148.
- Good, A. C.; Peterson, S. J.; Richards, W. G. *J. Med. Chem.* **1993**, *36*, 2929.
- Waganer, M.; Sadowski, J.; Gasteiger, J. *J. Am. Chem. Soc.* **1995**, *117*, 7769.
- Polanski, J.; Walczak, B. *Comp. Chem.* **2000**, *24*, 615.