Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2007) 701–708
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

A DFT-based quantum theoretic QSAR study of aromatic and heterocyclic sulfonamides as carbonic anhydrase inhibitors against isozyme, CA-II

Erol Eroglu [a,*], Hasan Türkmen [b]

[a] Harran University, Physics Department, Osmanbey Kampus, 6300 Sanliurfa, Turkey
[b] Harran University, Chemistry Department, Osmanbey Kampus, 63300, Sanliurfa, Turkey

Received 28 November 2006; received in revised form 26 March 2007; accepted 30 March 2007
Available online 4 April 2007

Abstract

In this study, DFT/B3LYP level of theory with the 6-31G (d) basis set, was used to calculate a set of molecular descriptors of 30 sulfonamide compounds with carbonic anhydrase inhibitory activity. Quantitative structure–activity relationship (QSAR) models of the biological activity (K_i) of 30 inhibitors of carbonic anhydrase (CA-II) were established using some of the following calculated quantum mechanical descriptors, dipole moment (μ), average polarizability (P), ionization potential (I), electron affinity (A), LUMO energy (ε_{LUMO}), HOMO energy (ε_{HOMO}), total energy at 0 K (Te), entropy at 298 K (S), electronegativity (χ), hardness (η), electrophilicity (ω), and differences between HOMO and LUMO energies ($\varepsilon_H - \varepsilon_L$). The QSAR models obtained by employing multiple linear regression techniques are aimed at correlating the structures to their reported experimental inhibitory activity values, K_i. Among the models presented in this study, statistically the most significant one is a four-parameter linear equation with correlation coefficient, R^2 values of ca. 0.847 and the cross-validated correlation coefficient, R values of ca. 0.775. This study also reconstructed the obtained models using AM1-based calculated descriptors to demonstrate the superiority of DFT over AM1 for quantum calculations of mechanical descriptors. The results are discussed in the light of the main factors that influence the inhibitory activity of carbonic anhydrase (CA-II) isozyme.

Keywords: Sulfonamides; Carbonic anhydrase inhibitors; Molecular descriptors; QSAR; DFT; AM1

1. Introduction

The sulfonamides that inhibit the zinc enzyme carbonic anhydrase (CA, EC 4.2.1.1) represent an important class of biologically active compounds. Carbonic anhydrase inhibitors have been extensively studied in the past due to their potential applications as drugs for treating diseases such as cancer, glaucoma, epilepsy, and as diuretics [1–9]. Quantitative structure–activity relationships (QSAR) studies are tools for predicting endpoints of interest in organic molecules acting as drugs [10]. Many physiological activities of a molecule can be related to their composition and structure. Molecular descriptors, which are numerical representations of the molecular structures, are used for performing QSAR analysis [11].

In the literature, there have been a number of QSAR studies of sulfonamides using semi-empirical (AM1) quantum chemical [12–17] and topological [18–21] descriptors. To the best of our knowledge, two QSAR studies [22,23] have been reported. These studies were performed using density functional theory (DFT) method [24] for modeling and estimating CA inhibition by sulfonamides. Some previous comparative QSAR studies [25–28] have shown that employing the descriptors calculated using the DFT method together with B3LYP hybrid functional [29] instead of the semi-empirical methods AM1 [30] or PM3 [31,32], improved the accuracy of the results and led to more reliable QSARs. In many QSAR studies, the energy of LUMO as Koopman's electron affinity (A) and the energy of HOMO as negative of Koopman's ionization potential (I) have been considered measures of electron affinity and ionization potential, respectively. Nevertheless, semi-empirical methods approximate certain properties up to a certain level, and it is obvious that variation between LUMO and HOMO energy values of various congeners is considerably small. This implies that an important error can be introduced if a low-level quantum method is employed in the calculations.

* Corresponding author. Tel.: +90 4143440089; fax: +90 4143440051.
E-mail address: eeroglu@harran.edu.tr (E. Eroglu).

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2007.03.015

The aim of the present study is two-fold. The first is to build QSAR multiple regression models using DFT-based descriptors and to correlate and predict the inhibition constant (K_i) for a diverse set of 30 aromatic and heterocyclic sulfonamide compounds. The second is to envisage reconstruction of the obtained models using AM1-based calculated descriptors to demonstrate the superiority of DFT over AM1 in terms of the calculations. The K_i values of compounds were taken from two different studies [33,8]. Some of these compounds have been recently synthesized and their K_i values have been reported against isozyme, CA-II [8]. They have not been dealt with any QSAR study till now. This study has employed DFT using B3LYP hybrid functional together with 6-31G (d) basis set to calculate various quantum mechanical descriptors and correlated them using the reported experimental inhibition constants employing multilinear regression models.

2. Theory and computational details

2.1. Theory

Several important molecular properties such as chemical hardness (η) and electronegativity (χ) have been defined based on density functional theory [34–36]. Chemical hardness has been used as a tool to understand the chemical reactivity and some other properties of a molecular system [37]. It has also been shown that stability of molecules is related to its chemical hardness [38]. The concept of electronegativity has been introduced as the power of an atom in a molecule to attract electrons onto itself [39]. Chemical hardness (η), and electronegativity (χ) are defined as follows [34–36,40]:

$$\eta = \frac{1}{2}\left[\frac{\partial^2 E}{\partial N^2}\right]_{V(r)} \tag{1}$$

$$\chi = -\left[\frac{\partial E}{\partial N}\right]_{V(r)} \tag{2}$$

where E and $V(r)$ are electronic energy and external potential of an N-electron system, respectively. Using Koopmans' theorem for closed-shell molecules, η and χ can be redefined as

$$\eta \approx -\tfrac{1}{2}(\varepsilon_{\mathrm{LUMO}} + \varepsilon_{\mathrm{HOMO}}) \approx \tfrac{1}{2}(I - A) \tag{3}$$

$$\chi \approx \tfrac{1}{2}(\varepsilon_{\mathrm{HOMO}} - \varepsilon_{\mathrm{LUMO}}) \approx \frac{I + A}{2} \tag{4}$$

$$I \approx -\varepsilon_{\mathrm{HOMO}} \quad \text{and} \quad A \approx \varepsilon_{\mathrm{LUMO}} \tag{5}$$

where I and A are the ionization potential and electron affinity of the molecules, respectively. I characterizes the susceptibility of a molecule, whereas A refers to the capability of a ligand to accept precisely one electron from a donor. Electrophilicity (ω) has been proposed as a measure of lowering of energy due to maximal electron flow between the donor and the acceptor [36]. It can be defined as a ratio of the square of electronegativity to chemical hardness:

$$\omega = -\frac{\chi^2}{2\eta} \tag{6}$$

This study has also included HOMO–LUMO energy gap as a quantum mechanical descriptor. There are numerous applications of HOMO–LUMO energy gap in establishing a correlation between the chemical structure and the biological activity.

It is well known that the polarity of a molecule is important for various physicochemical properties. The dipole moment (μ) thus is the most obvious and the most widely used quantity to describe the polarity of a molecule. Polarization of a molecule by an external electric field is given in terms of the nth order susceptibility tensors of the molecular bulk. The first-order term that contains information about possible inductive interactions in a molecule is referred to as the molecular polarizability. Polarizability (P) value has been shown to be related to hydrophobicity of a molecule and is commonly used as a descriptor in QSAR studies. The other descriptor used in this study is the entropy (S) at 298 K, which is related to the thermodynamic characteristics of molecules.

2.2. Computational details

For all the molecules studied, 3D modeling and calculations were performed using the Gaussian 03 quantum chemistry package [41]. To save computational time, initial geometry optimizations were carried out with the molecular mechanics (MM) method, using the MM+ force fields. The lowest energy confirmations of the molecules obtained by the MM method were further optimized by the DFT [24] method by employing Becke's three-parameter hybrid functional (B3LYP) [29] and the 6-31G (d) basis set. Their fundamental vibrations were also calculated using the same method to check if there were true minima. All the computations were carried out for the ground states of these molecules as single states. In the calculations, the dipole moment (in debye) of the molecules, their polarizability, which was reported as cartesian (x, y, and z) components (in a.u.), and the entropy at 298 K (cal/mol K) was directly extracted from the Gaussian 03 output file. The HOMO and LUMO values taken as molecular orbital coefficients (in a.u.) from the output of Gaussian 03 calculation were converted into energy (in eV). Rest of the descriptors were obtained using Eqs. (3)–(6). CODESSA PRO (Comprehensive Descriptors for Structural and Statistical Analysis), Version 2.7.2 [42] was used for statistical analysis. This code uses diverse statistical structure property/activity correlation techniques for the analysis of experimental data in combination with the calculated molecular descriptors. The heuristic method implemented in CODESSA PRO was employed for selecting the 'best' regression model. By this method [43], a preselection of descriptors is accomplished. All descriptors are checked to ensure that (a) value of each descriptor is available for every structure and (b) there is a variation in these values. The descriptors for which values are not available for every

structure in the data in question are discarded. Descriptors having a constant value for all structures in the data set are also discarded. A printout showing descriptors thus discarded is provided. Thereafter, the one-parameter correlation equations for each descriptor are calculated. To further reduce the number in the "starting set" of descriptors, the following criteria are applied and a descriptor is eliminated if: (a) the F-test's value for the one-parameter correlation with the descriptor is below 1.0, (b) the squared correlation coefficient of the one-parameter equation is less than $R^2_{\min}$ 0.01 by default, (c) the parameter's

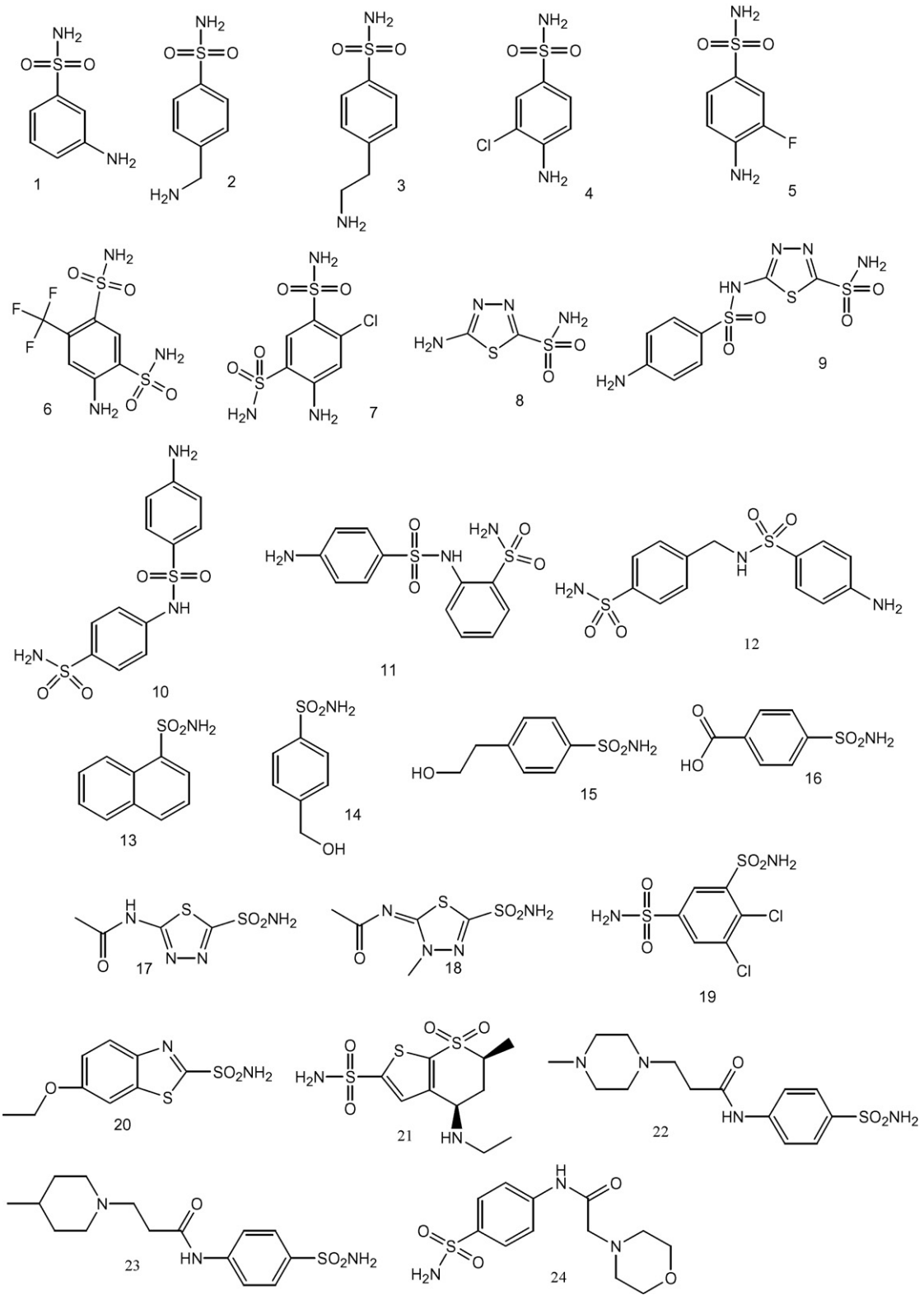

Fig. 1. Structure of the 30 sulfonamides studied.

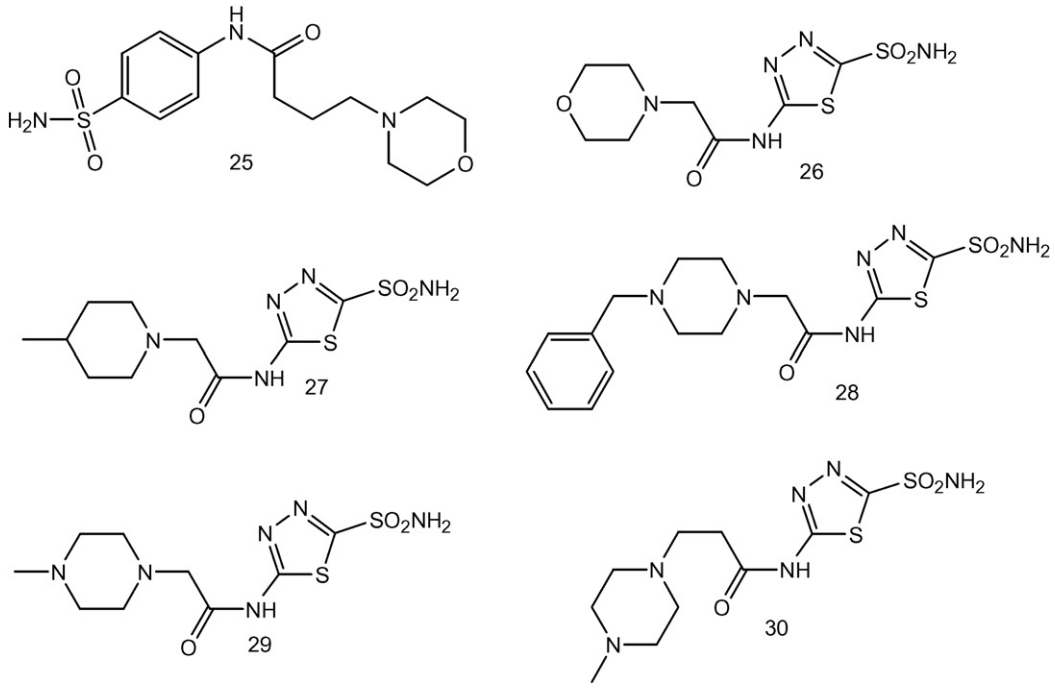

Fig. 1. (*Continued*).

t-value is less than t_1 (where R^2_{min} 0.1 by default and t_1 1.5 by default are user-defined values), and (d) the descriptor is highly intercorrelated (above r_{full}, where r_{full} is a user-specified value by default 0.99), with another descriptor. All the remaining descriptors are then listed in decreasing order according to the correlation coefficient of the corresponding one-parameter correlation equation. All two parameter regression models with remaining descriptors are developed and ranked by the regression correlation coefficient R^2. A stepwise addition of further descriptors scales is performed to find the best multiparameter regression models with optimum values of statistical criteria (highest values of R^2, the cross-validated, R^2_{CV} and the F value).

3. Results and discussion

The structures of 30 aromatic and heterocyclic sulfonamides are shown in Fig. 1. The experimental inhibitory activities of a set of 30 sulfonamides were taken from two references [8,33]. Table 1 shows the following information: (i) DFT-based calculated molecular descriptor values, (ii) experimental K_i (nM) values taken from the original references, and (iii) the calculated K_i (nM) values using the best DFT-based model I obtained in this study. The plot of observed versus calculated K_i for CA II using DFT-based model I is shown in Fig. 2.

Using the heuristic method, several regression equations were obtained in this study. Among the regression results, three equations were selected as models and are given in Table 2. The statistical quality of model I is good, as evident from $R^2 = 0.847$ and $R^2_{CV} = 0.775$. The other models II and III are only statistically satisfactory as evident from ($R^2 = 0.829$, $R^2_{CV} = 0.774$ model II) and ($R^2 = 0.811$ and $R^2_{CV} = 0.724$ model III). In these models, the correlation coefficient, R^2, is a measure of the fit of the regression equation. F, the Fisher test value, reflects the ratio of the variance explained by the model and the variance due to the error in the model. Higher values of F-test indicate the significance of the equation. s^2 is the standard deviation of the regression. R^2_{CV}, the 'leave one out' (LOO) cross-validated coefficient, is a practical and reliable method for testing the predictive performance and stability of a regression model. LOO approach involves developing a number of models with one sample omitted at a time. After developing each model, the omitted data are predicted and the differences between the experimental and predicted activity values are calculated. R^2_C values are then calculated according to the following formula [44]:

$$R^2_{CV} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (7)$$

where y_i is the actual experimental activity, $\bar{y}$ the average actual experimental activity and $\hat{y}_i$ is the predicted activity of compound i computed by the new regression equation obtained each time after leaving out one datum point (No. i). In addition to LOO method, we also used another internal validation method. General algorithm of this method involved the following steps: (i) the parent data points ($N = 30$) are divided into three subsets (A, B and C): the first, fourth, seventh, etc., data points go into the first subset (A), the second, fifth, eight, etc., go into the subset (B), and the third, sixth, ninth, etc., into the subset (C); (ii) the three sets A–C were prepared as the combination of two training subsets (A + B), (A + C) and (B + C). The remaining subsets (A, B and C, respectively) then become the corresponding test sets; and (iii) a correlation equation is derived for each of the training set with the same descriptors as models I, II, and III. Finally, the equation obtained is used to predict the

Table 1
DFT-based calculated molecular descriptors and predicted inhibition activities of 30 sulfonamides compound by the model I

C	μ	P	ε_{HOMO}	ε_{LUMO}	$\varepsilon_H - \varepsilon_L$	η	χ	ω	S	o-K_i (nM)	p-K_i (nM)
1	3.7871	95.477	−5.9169	−0.7758	−5.1411	−2.5705	−3.3463	−2.1781	103.405	240[a]	239.2
2	5.0935	106.588	−6.6919	−0.8738	−5.8181	−2.9091	−3.7828	−2.4595	111.249	170[a]	123.5
3	2.7705	117.844	−6.4608	−1.0215	−5.4393	−2.7197	−3.7412	−2.5732	118.090	160[a]	157.5
4	6.1070	108.866	−6.1928	−0.7668	−5.4260	−2.7130	−3.4798	−2.2317	108.506	110[a]	172.0
5	5.9615	97.3013	−5.7392	−0.3050	−5.4342	−2.7171	−3.0221	−1.6807	102.182	300[a]	247.1
6	6.4272	140.856	−6.6301	−1.8177	−4.8124	−2.4062	−4.2239	−3.7074	139.163	63[a]	29.4
7	6.2949	140.683	−6.5449	−1.4896	−5.0554	−2.5277	−4.0173	−3.1923	129.216	75[a]	69.7
8	5.6204	84.836	−7.0103	−1.6564	−5.3539	−2.6769	−4.3334	−3.5073	99.729	60[a]	85.8
9	6.5397	183.882	−6.3381	−1.7097	−4.6284	−2.3142	−4.0239	−3.4984	149.252	2[a]	40.6
10	4.4446	191.175	−6.1602	−1.1579	−5.0023	−2.5012	−3.6590	−2.6764	154.237	46[a]	89.7
11	8.9923	185.893	−6.2772	−1.2528	−5.0244	−2.5122	−3.7650	−2.8213	142.047	50[a]	44.9
12	6.8844	204.978	−6.1411	−0.9723	−5.1688	−2.5844	−3.5567	−2.4474	161.273	33[a]	52.7
13	6.2759	129.771	−6.4584	−1.7889	−4.6695	−2.3348	−4.1236	−3.6416	104.591	133[a]	121.7
14	2.7365	102.256	−7.0274	−0.9304	−6.0970	−3.0485	−3.9789	−2.5966	111.100	125[a]	115.8
15	3.9792	114.230	−6.8990	−0.8376	−6.0614	−3.0307	−3.8683	−2.4687	118.226	110[a]	101.9
16	3.1498	104.194	−7.5988	−2.0899	−5.5090	−2.7545	−4.8443	−4.2599	112.144	40[a]	16.5
17	7.5833	111.654	−7.2557	−1.8050	−5.4507	−2.7253	−4.5304	−3.7654	118.946	12[a]	−7.6
18	7.1901	126.661	−6.4274	−1.6839	−4.7435	−2.3718	−4.0556	−3.4675	127.973	14[a]	68.3
19	0.9875	142.139	−7.5292	−2.0210	−5.5082	−2.7541	−4.7751	−4.1396	133.056	8[a]	8.6
20	8.1089	158.641	−6.2546	−1.6602	−4.5944	−2.2972	−3.9574	−3.4087	126.835	38[a]	78.2
21	6.8271	180.531	−6.4666	−1.7010	−4.7656	−2.3828	−4.0838	−3.4995	151.464	9[a]	19.4
22	5.0308	202.086	−5.4516	−1.1638	−4.2877	−2.1439	−3.3077	−2.5517	162.310	165[b]	139.7
23	5.2198	204.558	−5.2652	−1.1238	−4.1413	−2.0707	−3.1945	−2.4642	163.503	173[b]	155.5
24	7.6717	125.364	−5.7833	−0.8988	−4.8845	−2.4422	−3.3410	−2.2853	146.795	87[b]	112.6
25	5.9426	196.910	−5.6219	−2.4855	−3.1364	−1.5682	−4.0537	−5.2393	164.413	104[b]	69.5
26	9.2740	164.751	−6.6162	−1.6942	−4.9220	−2.4610	−4.1552	−3.5079	145.338	0.9[b]	−13.0
27	5.8951	184.304	−6.2878	−1.6858	−4.6020	−2.3010	−3.9868	−3.4538	152.643	3.8[b]	47.3
28	10.2529	240.385	−6.0951	−1.6120	−4.4831	−2.2415	−3.8535	−3.3124	175.902	1.8[b]	−31.5
29	10.1516	181.302	−6.2687	−1.6218	−4.6469	−2.3235	−3.9453	−3.3495	151.829	1.6[b]	1.1
30	11.1580	156.00	−6.2848	−1.4860	−4.7988	−2.3994	−3.8854	−3.1459	157.091	1.9[b]	−19.8

o-K_i, observed inhibition constant (nM); p-K_i, predicted inhibition constant (nM) by model I; μ, dipole moment (debye); P, polarizability (a.u.); ε_{LUMO}, energy of LUMO (eV); ε_{HOMO}, energy of HOMO (eV); $\varepsilon_H - \varepsilon_L$, energy difference between HOMO and LUMO (eV); η, chemical hardness (eV); χ, electronegativity (eV); ω, electrophilicity (eV); S, entropy at 298 K (cal/mol K).

[a] From Ref. [31].

[b] From Ref. [8].

inhibition activity of the corresponding test set. Table 3 shows the results of internal validations of the three models. This internal validation method shows that the average predictive ability of the three models is statistically satisfactory.

A perusal of Table 2 shows that six types of descriptors are involved in all the three models. In the evaluation of the regression equations in all the three models, μ plays an important role in exhibiting CA-II inhibition activities of sulfonamides molecules. The negative coefficient of μ indicates that the higher the dipole moment, the greater is the activity. The dipole moment is a measure of the molecular polarity, which seems to have an important effect on the inhibition activity of CA-II isozyme. According to models I and III, χ has the highest positive coefficients; highlighting the fact that strength of molecular association by charge transfer plays an important role between sulfonamide molecules and the receptor CA-II isozyme. Also, electronegativity is the most statistically significant descriptor in models I and III, as is evident from the fact that it has the highest t-test value (Table 2). By analyzing model II, it can be seen that the most representative descriptors are ε_{HOMO} with the positive coefficient and the highest t-test value. According to Eq. (5), I equals to $-\varepsilon_{HOMO}$ and the electron affinity A is the ε_{LUMO}. This shows that the energies of the frontier molecular orbital (HOMO) affect the inhibition activity. ε_{HOMO} is responsible for the formation of charge transfer in a chemical reaction and characterizes the susceptibility of the molecule towards attack by electrophiles. In much the same role ε_{LUMO} also characterizes the susceptibility of the molecule, but towards attack by nucleophiles. In model II, the most important descriptor is ε_{HOMO} which has the highest coefficient and

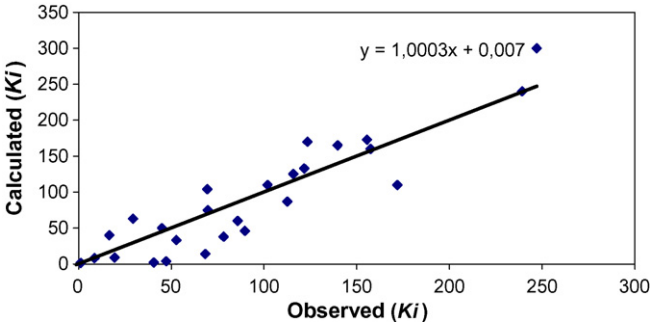

Fig. 2. Plot of observed vs. calculated K_i for C II using model I.

Table 2
DFT-based regression models built for 30 sulfonamide molecules by using heuristic method

Model I	Descriptor	Coefficient	*t*-Test	Number of descriptor
$R^2 = 0.847$	Intercept	1.139×10^3	9.185	4
$F = 34.67$	χ-Electronegativity	2.088×10^2	6.747	
$s^2 = 1072.44$	μ-Dipole moment	-1.214×10^1	−4.169	
$N = 30$	*S*-Entropy at 298 K.	-2.073×10^0	−5.445	
$R^2_{CV} = 0.775$	ε_{LUMO}	-7.644×10^1	−2.724	
Model II	**Descriptor**	**Coefficient**	***t*-Test**	**Number of descriptor**
$R^2 = 0.829$	Intercept	1.247×10^3	11.215	3
$F = 41.97$	ε_{HOMO}	1.199×10^2	9.163	
$s^2 = 1155.53$	μ-Dipole moment	-1.274×10^1	−4.244	
$N = 30$	*S*-Entropy at 298 K	-2.403×10^0	−7.015	
$R^2_{CV} = 0.7746$				
Model III	**Descriptor**	**Coefficient**	***t*-test**	**Number of descriptor**
$R^2 = 0.811$	Intercept	1.007×10^3	8.056	4
$F = 26.86$	ε_{LUMO}	-7.121×10^1	−2.238	
$s^2 = 1325.26$	χ-Electronegativity	2.044×10^2	5.813	
$N = 30$	μ-Dipole moment	-1.317×10^1	−4.109	
$R^2_{CV} = 0.724$	*P*-Polarizability	-1.010×10^0	−4.384	

Table 3
Internal validation of the DFT-based QSAR models

Training set	*N*	R^2	*F*	s^2	Test set	*N*	R^2 (pred)
Model I							
A + B	20	0.878	27.04	1097.75	C	10	0.711
A + C	20	0.819	16.97	1044.15	B	10	0.897
B + C	20	0.898	47.33	776.93	A	10	0.683
Model II							
A + B	20	0.862	33.40	1163.47	C	10	0.637
A + C	20	0.824	13.14	1085.74	B	10	0.886
B + C	20	0.824	24.99	1349.36	A	10	0.872
Model III							
A + B	20	0.881	27.97	1065.86	C	10	0.656
A + C	20	0.767	12.41	1339.33	B	10	0.909
B + C	20	0.898	47.33	776.93	A	10	0.683

highlights that the higher the energy of HOMO orbital, the greater is the activity. It may be said that due to the existence of ε_{HOMO} or ε_{LUMO} in all the obtained models, charge transfer between sulfonamide molecules and the receptor CA-II is a dominant factor for modeling the inhibition activity in all three models. Also, the polarity of molecules affects the activity as evident from μ, which is involved in all three models with relatively large coefficients. This can be referred from Table 2. *S* (entropy at 298 K) in model I and II and *P* (polarizability) in

Table 4
AM1-based calculated molecular descriptors and predicted inhibition activities of 30 sulfonamides compound by the model I

C	μ	*P*	ε_{HOMO}	ε_{LUMO}	$\varepsilon_H - \varepsilon_L$	η	χ	ω	*S*	o-K_i (nM)	p-K_i (nM)
1	4.6250	82.336	−8.9920	−0.4800	8.5120	−4.2560	−4.7360	−2.6351	101.910	240[a]	192.9
2	4.7200	88.230	−9.9350	−0.6090	9.3260	−4.6630	−5.2720	−2.9803	111.332	170[a]	154.3
3	4.0070	96.583	−9.9710	−0.6770	9.2940	−4.6470	−5.3240	−3.0498	117.619	160[a]	137.9
4	6.0760	90.917	−9.1720	−0.4950	8.6770	−4.3385	−4.8335	−2.6925	109.342	110[a]	179.4
5	5.6560	86.336	−9.1940	−0.5780	8.6160	−4.3080	−4.8860	−2.7708	106.261	300[a]	170.7
6	8.4280	117.728	−9.9180	−1.3090	8.6090	−4.3045	−5.6135	−3.6603	143.568	63[a]	12.8
7	8.6100	113.047	−9.7440	−0.9640	8.7800	−4.3900	−5.3540	−3.2648	130.939	75[a]	78.7
8	5.4540	76.001	−9.6990	−1.2840	8.4150	−4.2075	−5.4915	−3.5837	98.7619	60[a]	70.0
9	4.2370	160.177	−9.3460	−1.3620	7.9840	−3.9920	−5.3540	−3.5903	147.910	2[a]	10.4
10	4.4280	165.749	−9.3150	−0.7760	8.5390	−4.2695	−5.0455	−2.9813	154.581	46[a]	89.7
11	6.7080	157.168	−9.2160	−0.6910	8.5250	−4.2625	−4.9535	−2.8783	143.869	50[a]	112.3
12	4.5270	169.900	−9.2690	−0.7990	8.4700	−4.2350	−5.0340	−2.9919	155.788	33[a]	85.3
13	5.3460	112.571	−9.3050	−1.0260	8.2790	−4.1395	−5.1655	−3.2229	106.743	133[a]	103.4
14	4.1100	85.738	−10.257	−0.7160	9.5410	−4.7705	−5.4865	−3.1550	110.218	125[a]	137.2
15	5.0170	94.558	−10.053	−0.5190	9.5340	−4.7670	−5.2860	−2.9308	118.076	110[a]	158.8
16	3.0850	88.493	−10.695	−1.3650	9.3300	−4.6650	−6.0300	−3.8972	112.017	40[a]	36.4
17	7.2370	96.874	−9.8610	−1.3630	8.4980	−4.2490	−5.6120	−3.7061	118.138	12[a]	34.0
18	6.1320	110.492	−9.2470	−0.9590	8.2880	−4.1440	−5.1030	−3.1420	127.737	14[a]	90.5
19	2.2340	110.184	−10.433	−1.4660	8.9670	−4.4835	−5.9495	−3.9474	132.278	8[a]	3.2
20	6.3960	139.514	−9.0110	−1.1220	7.8890	−3.9445	−5.0665	−3.2538	126.969	38[a]	69.4
21	6.1520	146.344	−9.8900	−1.5800	8.3100	−4.1550	−5.7350	−3.9579	150.133	9[a]	−31.4
22	6.3150	169.971	−9.1440	−0.6460	8.4980	−4.2490	−4.8950	−2.8196	164.275	165[b]	98.1
23	6.9460	170.354	−9.2300	−0.6060	8.6240	−4.3120	−4.9180	−2.8046	164.694	173[b]	102.0
24	7.9020	151.365	−9.3570	−0.5120	8.8450	−4.4225	−4.9345	−2.7529	147.996	87[b]	131.5
25	7.4260	164.914	−9.4170	−0.6010	8.8160	−4.4080	−5.0090	−2.8460	167.235	104[b]	97.6
26	7.9480	141.855	−9.7810	−1.3180	8.4630	−4.2315	−5.5495	−3.6390	149.800	0.9[b]	6.6
27	8.1990	155.146	−9.7300	−1.2710	8.4590	−4.2295	−5.5005	−3.5767	154.781	3.8[b]	8.4
28	8.8720	205.125	−9.5830	−1.2450	8.3380	−4.1690	−5.4140	−3.5154	180.925	1.8[b]	−15.0
29	9.2690	154.364	−9.6370	−1.2470	8.3900	−4.1950	−5.4420	−3.5298	156.816	1.6[b]	9.6
30	8.4740	160.727	−9.5400	−1.1860	8.3540	−4.1770	−5.3630	−3.4429	162.839	1.9[b]	13.9

o-K_i, observed inhibition constant (nM); p-K_i, predicted inhibition constant (nM) by model I; μ, dipole moment (debye); *P*, polarizability (a.u.); ε_{LUMO}, energy of LUMO (eV); ε_{HOMO}, energy of HOMO (eV); $\varepsilon_H - \varepsilon_L$, energy difference between HOMO and LUMO (eV); η, chemical hardness (eV); χ, electronegativity (eV); ω, electrophilicity (eV); *S*, entropy at 298 K (cal/mol K).

[a] From Ref. [31].

[b] From Ref. [8].

Table 5
The comparison of statistical quality between the DFT and AM1 models

Models	Statistical characteristic							
	DFT				AM1			
	R^2	F	s^2	R^2_{CV}	R^2	F	s^2	R^2_{CV}
I	0.847	34.67	1072	0.775	0.661	12.24	2360	0.543
II	0.828	41.97	1155	0.774	0.346	4.60	4386	0.173
III	0.811	26.86	1325	0.724	0.675	13.03	2263	0.557

model III, with relatively small coefficients also contribute to the inhibition activity.

Finally, the quantum mechanical descriptors, which were used to build all three models in Table 2, and were also calculated by using the semi-empirical AM1 method. The results of the calculation are presented in Table 4. The models I, II, and III were reconstructed using AM1-based descriptors. A comparison of the DFT-based models with the AM1-based QSAR models is given in Table 5, which clearly demonstrates that the DFT-based quantum mechanical descriptors led to better correlation relationships than the corresponding descriptors based on the AM1 method. This result is not surprising as the semi-empirical methods like AM1 approximate some molecular properties only up to a certain level, so an error can be introduced if a low-level quantum method is employed in the calculations. The results of this study are in accordance with the studies [25–28] of comparisons of DFT-based QSARs and semi-empirical methods-based QSARs.

4. Conclusions

In this study, the calculation of quantum mechanical descriptor of 30 aromatic and heterocyclic sulfonamide compounds were presented at the DFT/B3LYP/6-31G (d) and semi-empirical AM1 levels of the theory. The DFT-based descriptors were used to drive the QSAR models by which the calculated quantum mechanical parameters were correlated to inhibition activity of compounds taken from the literature [8,33]. The heuristic method, implemented in CODESSA PRO was employed to obtain the QSAR models. Three of the obtained multiple regression equations, which had the highest statistical quality, were chosen as models (model I, II and III). This study attempted to analyze the models to rationalize the mechanism of inhibition activity of the compounds against the CA-II. Model I is statistically the most significant four-parameter equation with the correlation coefficient, R^2 values of ca. 0.847 and the cross-validated correlation coefficient, R^2_{CV} values of ca. 0.775. Model II is statistically satisfactory as evident from its R^2 values of ca. 0.829, and model III is also only statistically satisfactory with its R^2 values of ca. 0.811. The obtained models showed not only the statistical significance but also the predictive ability as their R^2_{CV} values are ca. 0.775 for model I, ca. 0.774 for model II, and ca. 0.724 for model III. The predictive ability of the models was further demonstrated with an internal validation method by dividing 30 compounds into subgroups as training and test sets. This validation method also confirmed the predictive ability of the models. To demonstrate the superiority of DFT method over AM1 method in the calculations of molecular descriptor, the above-mentioned models were reconstructed using the AM1-based calculated descriptors. This resulted in the dramatic reduction of the statistical quality of the models from $R^2 = 0.847$ to 0.661 for model I, from $R^2 = 0.828$ to 0.346 for model II, and from $R^2 = 0.811$ to 0.675 for model III.

The descriptors used in this study are global values of the molecules, which allowed the authors to give a physical explanation to quantum mechanical properties contributing to the inhibition activity. The models described in this study showed that the charge transfer between the sulfonamide molecules and the receptor CA-II enzyme is the most important factor contributing to the inhibition activity. Also, the inhibition activity is highly influenced by the polarity of the molecules. The authors believe that the models introduced in this study can be used to estimate the inhibition activity of novel sulfonamide compounds of this series, prior to synthesis by calculating the descriptors involved in these equations.

Acknowledgments

This work has been supported in part by Harran University Research Council (HUBAK) Project no: 426 and by the Turkish Government Planning Organization (DPT) Project no: 2003K120590.

References

[1] C.T. Supuran, A. Scozzafava, A. Casini, Med. Res. Rev. 23 (2003) 146.
[2] C.T. Supuran, A. Scozzafava, Exp. Opin. Ther. Pat. 12 (2002) 217.
[3] C.T. Supuran, Exp. Opin. Invest. Drugs 12 (2003) 283.
[4] T.H. Maren, G.C. Wynns, P.J. Wistrand, Mol. Pharmacol. 44 (1993) 901.
[5] A. Casini, F. Abbate, A. Scozzafava, S. David, A. Mastrolorenzo, C.T. Supuran, Bioorg. Med. Chem. Lett. 13 (2003) 841.
[6] T.H. Maren, B.W. Clare, C.T. Supuran, Roum. Chem. Quart. Rev. 2 (1994) 259.
[7] G. Liang, J.P. Baysb, J.P. Bowen, J. Mol. Struct. (Theochem) 401 (1997) l65.
[8] H. Turkmen, M. Durgun, S. Yilmaztekin, M. Emul, A. Innoocenti, D. Vullo, A. Scozzafava, C.T. Supuran, Bioorg. Med. Chem. Lett. 15 (2005) 367.
[9] T.H. Maren, Drug Dev. Res. 10 (1987) 255.
[10] M. Karelson, Molecular Descriptors in QSAR/QSPR, Wiley, New York, 2000.
[11] R. Tadeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley–VCH, Weinheim, 2000.
[12] B.W. Clare, C.T. Supuran, J. Mol. Struct. (Theochem) 428 (1998) 109.

[13] B.W. Clare, C.T. Supuran, Eur. J. Med. Chem. 32 (1997) 311.
[14] B.W. Clare, C.T. Supuran, Eur. J. Med. Chem. 34 (1999) 4643.
[15] C.T. Supuran, B.W. Clare, Eur. J. Med. Chem. 33 (1998) 489.
[16] C.T. Supuran, B.W. Clare, Eur. J. Med. Chem. 30 (1995) 687.
[17] C.T. Supuran, B.W. Clare, Eur. J. Med. Chem. 34 (1999) 41.
[18] V.K. Agrawal, S. Shrivastava, P.V. Khadikar, C.T. Supuran, Bioorg. Med. Chem. 11 (2003) 5353.
[19] V.K. Agrawal, P.V. Khadikar, Bioorg. Med. Chem. Lett. 13 (2003) 447.
[20] V.K. Agrawal, R. Sharma, P.V. Khadikar, Bioorg. Med. Chem. 10 (2002) 2993.
[21] M. Jaiswal, P.V. Khadikar, A. Scozzafava, C.T. Supuran, Bioorg. Med. Chem. Lett. 14 (2004) 3283.
[22] B.W. Clare, C.T. Supuran, Bioorg. Med. Chem. 13 (2005) 2197.
[23] B.W. Clare, C.T. Supuran, J. Chem. Inf. Model. 45 (2005) 1385.
[24] R.G. Parr, W. Yang, Density Functional Theory of Atoms and Molecules, Oxford University Press, New York, 1989.
[25] M. Reis, B. Lobato, J. Lameira, A.S. Santos, C.N. Alves, Eur. J. Med. Chem. 41 (2006) 1.
[26] F.A. Pasha, H.K. Srivastava, P.P. Singh, Bioorg. Med. Chem. 13 (2005) 6823.
[27] L. Zhang, J. Wan, G. Yang, Bioorg. Med. Chem. 12 (2004) 6183.
[28] S. Trohalaki, E. Gifford, R. Pachter, Comp. Chem. 24 (2000) 421.
[29] A.D. Becke, J. Chem. Phys. 98 (1993) 5648.
[30] M. Dewar, W. Thiel, J. Am. Chem. Soc. 99 (1977) 4499.
[31] J.J.B. Stewart, J. Comp. Chem. 10 (1989) 209.
[32] J.J.B. Stewart, J. Comp. Chem. 10 (1989) 221.
[33] D. Vullo, M. Franchi, E. Gallori, J. Pastorek, A. Scozzafava, S. Pastorekovac, C.T. Supurana, Bioorg. Med. Chem. Lett. 13 (2003) 1005.
[34] R.G. Parr, R.A. Donnelly, M. Levy, W.E. Palke, J. Chem. Phys. 68 (1978) 3801.
[35] R.G. Parr, R.G. Pearson, J. Am. Chem. Soc. 105 (1983) 7512.
[36] R.G. Parr, L.V. Szentpaly, S. Liu, J. Am. Chem. Soc. 121 (1999) 1922.
[37] R.G. Parr, P.K. Chattaraj, J. Am. Chem. Soc. 113 (1991) 1854.
[38] Z. Zhou, R.G. Parr, J. Am. Chem. Soc. 112 (1990) 5720.
[39] I. Pauling, The Nature of the Chemical Bond, 3rd ed., Cornell University Press, Ithaca, NY, 1960.
[40] R.G. Pearson, in: Z.B. Maksic (Ed.), Theoretical Models of Chemical Bonding Part II, Springer, Berlin, 1990, p. 45.
[41] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision B.05, Gaussian, Inc., Pittsburgh, PA, 2003.
[42] CODESSA™, Semichem, 7204, Mullen, Shawnee, KS, USA, Copyright© Semichem and the University of Florida, 2002.
[43] CODESSA™, References Manual, V. 2.13 (PC). Semichem, 7204, Mullen, Shawnee, KS, USA, Copyright© Semichem and the University of Florida, 2002.
[44] R.H. Mayers, Classical and Modern Regression With Applications, PWS-KENT Publ. Co., Boston, 1990.