



Prediction of Activity of Carbonic Anhydrase Inhibitor Drugs Based on QSAR Studies

N. DARZI*, A. MORSALI and S. A. BEYRAMABADI

Department of Chemistry, Mashhad Branch, Islamic Azad University, Mashhad, Iran.

*Correspondence author E-mail: n_darzi138@yahoo.com

<http://dx.doi.org/10.13005/ojc/310264>

(Received: March 02, 2015; Accepted: April 15, 2015)

ABSTRACT

A quantitative structure-activity relationship (QSAR) model, based on three quantum chemical descriptors obtained from the benzene sulphonamide derivatives using the density functional theory (DFT) method. Then this developed model was used to predict the benzene sulphonamide binding constant. The QSAR model has correlation coefficient R of 0.901 and the standard error of 0.646. Also, the predictive power of this model was further examined by leave-7-out cross validation procedure which the obtained statistical parameters were: $Q^2= 0.991$ and $SPRESS= 0.4686$ that giving a good enough predictive power. The selected descriptors are: molecular weight (MW), absolute hardness (AH), HOMO energy (HOMO), respectively.

Key words: Sulphonamides, Binding Constant, QSAR, DFT.

INTRODUCTION

The carbonic anhydrases (CA) form a family of enzymes that catalyze the rapid interconversion of carbon dioxide and water to bicarbonate and protons (or vice versa), a reversible reaction that occurs relatively slowly in the absence of a catalyst. The active site of most carbonic anhydrases contains a zinc ion; they are therefore classified as metalloenzymes. One of the functions of the enzyme is to interconvert carbon dioxide and bicarbonate and in conclusion to maintain acid-base balance in blood, Stomach and other tissues, and also carbon dioxide transmission of red blood cells and of red blood cells to tissues to facilitate the lungs. Change the

function of this enzyme disrupts the body's normal functions.

In between, Benzene sulphonamides important group of drugs that is capable of inhibiting carbonic anhydrase. Benzene sulphonamides are important groups of clinically drugs that be used in the treatment of diseases such as gastric ulcer, duodenal and intestinal disorders, glaucoma, infections and tumors are caused by high altitude¹⁻⁴. Also, additional products of them disarrange normal physiological functions⁴. Whereas, considering the suitable does of drug and binding constant should be determined by experimental method that is limited⁴ therefore, the development of theoretical model to

predict the activities of these drugs are interesting and necessary. Quantitative structure activity relationship (QSAR) methods enable in prediction and interpretation of the activities of a wide range of organic and drug compounds based on the correlation between their activities and molecular characteristics (molecular descriptors). There are several reports about the applications of QSAR in this case. In the probe, Kamal Kumar *et al.* revealed the relation of molecular size and inhabitation of CA by sulfonamides. They developed the theoretical model based on plan of active sites of enzyme and demonstrated the nature of receptor-sulfonamide joint⁵. At the other investigation, Hansch *et al.* provided the QSAR model that consisted of formula for binding constant of benzene sulfonamides as the human carbonic anhydrase⁶. Menziani *et al.* benefited the molecular mechanism Method for studying the interaction between 20 deprotonated benzene sulfonamides and carbonic anhydrase enzyme. Their results shown that the resection of enzyme-inhibitor is prevailed by the short-range van der Waals forces⁷. Vijay *et al.* presented quantitative structure–activity studies on a group of sulphanilamide Schiff's base inhibitors of carbonic anhydrase using distance-based topological indices. The regression analysis consequences indicated that the activities of the inhibiting carbonic anhydrase II (CAII) activity can be modeled excellently in multi parametric model⁸. Gupta *et al.* performed a QSAR study on the inhibition of a few isomers of carbonic anhydrase and some metalloproteinases (MPs). They demonstrated well correlation with Kier's first-order valence molecular connectivity index (1chi (v)) and electrotopological state indices of some atoms in the proposed model⁹. Varma and Wagde, QSAR studies carried out on the basis of 50 substituted benzene sulfonamides. They investigated the effects of Balaban index in predicting activity of carbonic anhydrase inhibitor¹⁰. Also, Felegari, Quantum chemical calculations performed another research based on DFT method for evaluation of anticancer drugs Cyclophosphamide. The results of this work revealed that quantum parameters affecting anticancer activity¹¹.

In the present work we try to predict the logarithm of binding constant of some sulfonamides derivatives from their quantum molecular descriptors by using multiple linear regression (MLR) techniques. For this purpose, we employed the density functional

theory (DFT) method for geometrical optimization.

EXPERIMENTAL

The studied materials of this research work are 29 derivatives of benzene sulfonamide (12). Structure of these compounds and their experimental and predicted values of binding constant (log K) are available in table 1. Before anything else, the structures of all the derivatives of benzene sulfonamide have been optimized and then encoded the structural features (descriptors) of the molecules. We used of Gaussian (03) program (13) with B₃lyp function (14, 15) and 6-311++G (d, p) basis set for geometrical optimization. Finally, quantum chemical descriptors were revealing for each molecule were obtained from the Gaussian output. Then the data set was separated into two groups: training and test sets. All molecules were placed by Y-ranking method in these sets. The training set, consisted of 22 molecules, was used for the model generation and the test set, consisted of 7 molecules, was used to take care of the overtraining and evaluate the prediction power of the generated model.

Descriptor selection and model development

For suitable descriptor selection, it should be considered the correlation between the calculated structural descriptors and the experimental binding constant values. Multiple linear regression is one of the most used modeling methods for this case. At first, for screening the calculated descriptors, highly correlated descriptors were removed and the descriptors with constant or almost constant values for all molecules were eliminated from the collection of descriptors. At the end of this step a total of 7 descriptors were reminded to further investigations. These descriptors are: Heat of Formation (HF), Molecular Weight (MW), Total Energy (TE), HOMO Energy (HOMO), LUMO Energy (LUMO), Absolute Hardness (AH) and Electronegativity (ELEC). In order to successive decision for developing QSAR model, it should be stopped adding descriptors to the model during the selection procedure. Therefore, we benefited from 'break-point' procedure to control the model expansion¹⁶. In this technique, improvement of the statistical quality of the models is analyzed by plotting the squared correlation coefficient values of the obtained models versus the number of descriptors that were involved in each model. Thus, stepwise-MLR

procedure was applied to the training set and multi linear regression equations of up to 7 descriptors were developed. In this procedure we employed SPSS (V. 17)¹⁷. Then variations of square of correlation coefficient (R^2) against the number of descriptors in the models according to break-point procedure were recorded and are shown in fig 1. Pursuant to this figure led to the conclusion that the best model had 3 descriptors. These descriptors were listed in table 2 with their correlation matrix. As can be seen in this table there are not any high correlation between these descriptors.

RESULTS AND DISCUSSION

The best model was selected based on the statistics of correlation coefficient (R), standard error (SE) and Fisher-statistics value (F). Consequently, among different models the three-parameter model was chosen based on the break point procedure basis on the fig 1. A model result has been shown below.

$$\log k = 0.024(\pm 0.004)MW + 9.836(\pm 1.763)AH + 12.536(\pm 1.638)HOMO + 143.506(\pm 19.15)$$

R=0.901, SE=0.646, F=36.047, R_{cv} =0.87, Q^2 =0.991, SPRESS = 0.4686

The MLR predicted values of log K were shown in table 1. Finally, the leave 7-out cross-validation (L-7-O) was used to evaluate credibility and robustness of this model. For examining this test, the statistical parameters of Q^2 and SPRESS were calculated¹⁸.

$$Q^2 = 1 - \frac{\sum(\hat{y}_i - y_i)^2}{\sum(y_i - \bar{y})^2} \quad \dots(1)$$

$$SPRESS = \sqrt{\frac{\sum(\hat{y}_i - y_i)^2}{n - k - 1}} \quad \dots(2)$$

In the above expressions, k is the number of independent variables (descriptors) in regression equation, n is the number of observations (data set), \bar{y} is the mean of dependent variable, \hat{y}_i is the predicted value and y_i is experimental value of depended variable.

The resulted statistical values indicate that the proposed has stability, robustness, satisfactory fitting and prediction ability to predict the binding constant of benzene sulphonamide derivatives. We plotted the observed log K versus the calculated values of them in the fig 2. The excellent agreement between the experimental and calculated values of binding constant are demonstrate in this figure. The residuals of the MLR calculated values of their experimental values of benzene sulphonamide binding constant are plotted against their experimental values in fig 3. The random distribution of residuals about zero

Table 1: Structure of the benzene sulphonamide and their experimental and predicted values

S. No.	R	LogK (exp.)	LogK (pred.)	Residuals
1*	H	6.69	6.41	0.28
2*	4-CH ₃	7.09	7.41	-0.32
3	4-C ₂ H ₅	7.53	7.58	-0.05
4	4-C ₃ H ₇	7.77	7.82	-0.05
5*	4-C ₄ H ₉	8.30	8.01	0.29
6	4-C ₅ H ₁₁	8.86	8.55	0.31
7	4-CO ₂ CH ₃	7.99	7.12	0.87
8	4-CO ₂ C ₂ H ₅	8.50	7.39	1.11
9	4-CO ₂ C ₃ H ₇	8.77	8.42	0.35
10	4-CO ₂ C ₄ H ₉	9.11	8.09	1.02
11	4-CO ₂ C ₅ H ₁₁	9.39	8.43	0.96
12	4-CO ₂ C ₆ H ₁₃	9.39	8.94	0.45
13*	4-CONHCH ₃	7.08	7.32	-0.24
14	4-ONHC ₂ H ₅	7.53	7.71	-0.18
15	4-ONHC ₃ H ₇	8.08	8.09	-0.01
16*	4-ONHC ₄ H ₉	8.49	8.24	0.25
17	4-ONHC ₅ H ₁₁	8.75	8.98	-0.23
18	4-ONHC ₆ H ₁₃	8.88	9.03	-0.15
19	4-ONHC ₇ H ₁₅	8.93	8.79	0.14
20*	3-CO ₂ CH ₃	5.87	6.74	-0.87
21	3-CO ₂ C ₂ H ₅	6.21	7.01	-0.80
22	3-CO ₂ C ₃ H ₇	6.44	7.65	-1.21
23	3-CO ₂ C ₄ H ₉	6.95	7.87	-0.92
24	3-CO ₂ C ₅ H ₁₁	6.86	8.45	-1.59
25	2-CO ₂ CH ₃	4.41	4.31	0.10
26	2-CO ₂ C ₂ H ₅	4.80	4.35	0.45
27	2-CO ₂ C ₃ H ₇	5.28	6.20	-0.92
28	2-CO ₂ C ₄ H ₉	5.76	6.02	-0.26
29*	2-CO ₂ C ₅ H ₁₁	6.18	5.02	1.16

* The molecules used in test set

line verified that there is no systematic error in developed model.

Also, we profited Y-scrambling test to investigate any chance correlation in our modeling¹⁹. In this case, the dependent variable (log K) was randomly admixed and the original descriptors matrix is kept fixed then a new QSAR model was evolved. The results of this study are summarized in table 3. According to these results, it can be concluded that there was not any chance correlation between our data.

Descriptor analysis and interpretation

By the explanation of descriptors of this model, obtaining the effected factors (MF) on the

Table 2: Descriptors correlation matrix and mean effects

	MW	HOMO	AH	M E
MW	1	-0.522	-0.432	-5.313
HOMO	-0.522	1	0.701	0.999
AH	-0.432	0.701	1	3.450

Table 3: The results of Y-scrambling test

Iterations	R	SE
1	0.316	1.415
2	0.213	1.457
3	0.302	1.423
4	0.203	1.460
5	0.257	1.441
6	0.503	1.289
7	0.225	1.453
8	0.350	1.397
9	0.407	1.362
10	0.273	1.436
11	0.492	1.298
12	0.452	1.329
13	0.385	1.376
14	0.153	1.473
15	0.511	1.282
16	0.234	1.449
17	0.501	1.291
18	0.579	1.196
19	0.151	1.474
20	0.462	1.323

activity of benzenesulphonamides is well possible. Here, a brief interpretation of these factors is given based on the results of mean effect analysis²⁰. This parameter calculated for each descriptor that was shown in table 2.

$$MF_j = \frac{\beta_j \sum_{i=1}^n d_{ij}}{\sum_{j=1}^m \beta_j \sum_{i=1}^n d_{ij}} \quad \dots(3)$$

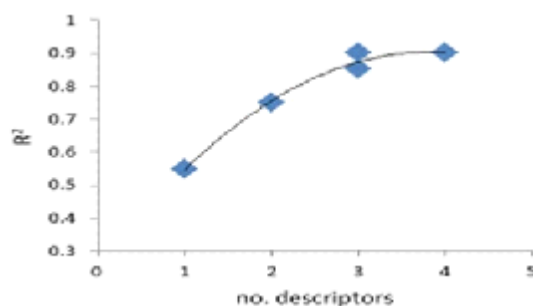


Fig. 1: plot of R² for the obtained models versus the number of descriptors involved

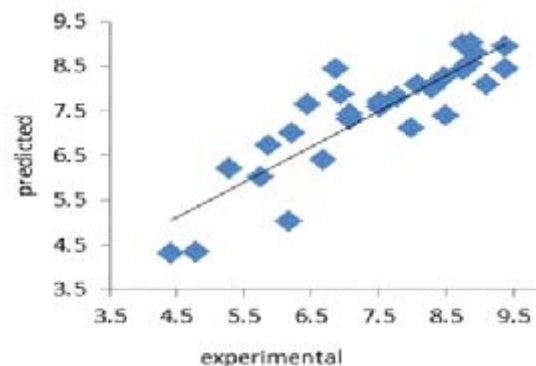


Fig. 2: The plot of log k versus the calculated values of them

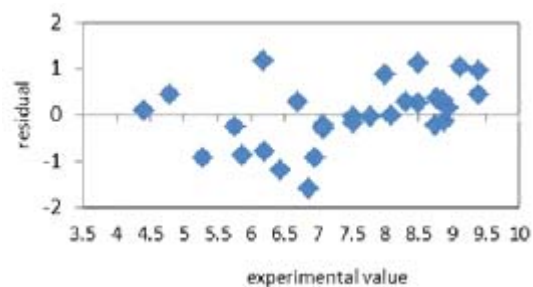


Fig. 3: The residuals of MLR calculated and experimental values of log k versus experimental

where; MF_j is the mean effect for considered descriptor j , β_j is the coefficient of descriptor j and d_{ij} is the value of interested descriptors for each molecule, n is the number of molecules in the data set and m is the number of descriptors in the model.

The value of MF manifests the relative importance of a descriptor in comparison with the other descriptors in the model. Also, the sign of descriptor represents the direction of variation in the values of activities resulted by increasing or decreasing the values of this descriptor. In this work, the results of mean effect analysis indicate that the order of importance of descriptors is: MW, AH and HOMO respectively.

MW with maximum mean effect is related to molecular size and is atom-type sensitive. The negative sign of this descriptor shows the inverse relation white log K (binding constant)²¹. AH whit middle mean effect is an important quantum chemical descriptor that relates to energy levels and wave function of molecule and donates the stability of

molecule as well as HOMO-LOMO energy gap. The positive mark of this descriptor shows the direct relation white log K²¹. HOMO with minimum mean effect is related to information about reactivity-stability of specific region of the molecule. Its positive mark donates the direct relation white log K²¹.

CONCLUSION

We have used quantum chemical descriptors basis of DFT calculations for QSAR studies of derivatives of benzene sulphonamides. The selected descriptors in proposed model can encode perfectly features of molecules which were responsible in benzene sulphonamide binding constant. Moreover, Q² and SPRESS in leave-7-out cross-validation of MLR were used as a measurement of generalization performance. The obtained statistical parameters of this test demonstrate devised QSAR model is a simple, robust and significant tool to predict the activity of new sulphonamides drugs via main factors of structure and the quantum chemical indices.

REFERENCES

- Bhat, M. A.; Imran, M.; Khan, S. A.; Siddiqui, N. *Ind. J. pharmaceutical sci.* **2005**, *67*, 151-159.
- Siliveru, S.; Parthasarathy, T. *Int. res. J. pharmacy.* **2012**, *3*, 213-215.
- Villar, R.; Encio, I.; Migliaccio, M.; Gil, M. J.; Martinez, M. V. *Bioorga and Med Chem.* **2004**, *12*, 963-968.
- Supuran, T. C.; Scozzafava, A.; Casini, A. *Med. Res. Rev.* **2003**, *23*, 146-189.
- Kumar, K.; Bindal, M. C.; Singh, P.; Gupta, S. P. *Int. J. Quant. Chem.* **1981**, *20*, 123-129.
- Hansch, C.; Mclarin, J.; Klein, J.; Langridge, R. *Mol. Pharmacol.* **1985**, *27*, 493-498.
- Menziani, M. C.; Debenedetti, P. G.; Gago, F.; Richards, W. G. *J. Med. Chem.* **1989**, *32*, 951-956.
- Agrawal, V. K.; Srivastava, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3287-3293.
- Gupta, S. P.; Maheswaran, V.; Pande, V.; Kumar, D. *J. Enzyme Inhib. Med. Chem.* **2003**, *18*, 7-13.
- Varma, R. G.; Wagde, r. *Orient J. Chem.* **2013**, *29*, 1621-1626.
- Felegari, Z. *Orient J. Chem.* **2014**, *30*, 1865-1875.
- Thakur, A.; Thakur, M.; Khadikar, P. V.; Supuran, C. T.; Sudele, P. *Bioorg. Med. Chem.* **2004**, *12*, 789-793.
- Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Montgomery, J.; Vreven, T.; Kudin, K.; Burant, J. Gaussian 03, Revision B.05 Gaussian Inc., Pittsburgh, PA, 2003.
- Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B.* **1988**, *37*, 785-789.
- Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- Fatemi, M. H.; Baher, E. *Mol. Divers.* **2009**, *13*, 343-352.
- SPSS/PC, Statistical Package for IBMPC, Quiad software, Ontario, 1986.
- Fatemi, M. H.; Baher, E. *SAR & QSAR Environ. Res.* **2009**, *20*, 77-90.
- Fatemi, M. H.; Baher, E.; Ghorbanzade'h, M. J. *Sep. Sci.* **2009**, *32*, 1-10.
- Fatemi, M. H.; Gharaghani, S.; Mohammad khani, S.; Rezaie, Z. *Electrochimica Acta.* **2008**, *53*, 4276-4282.
- Tadeschini R. and Consonni V., Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, (2000).