Available on line www.jocpr.com

Journal of Chemical and Pharmaceutical Research

ISSN No: 0975-7384
CODEN(USA): JCPRC5

J. Chem. Pharm. Res., 2011, 3(1):64-71

A DFT-Based QSARs Study of Benzimidazoles Drugs Derivatives

S. Qanei Nassab[a], *Z. Bayat [a]*, J. Movaffagh[b]*

[a]*Department of Chemistry, Islamic Azad University Quchan Branch, Iran*
[b]*Pharmaceutics School of Pharmacy, University of Medical Sciences(MUMS), Mashhad, Iran*

ABSTRACT

*A set of benzimidazole derivatives were tested for their the partition coefficient octanol- water ($LogP_{o/w}$) . Quantitative structure activity relationship (QSAR) analysis was applied to 28 of the above mentioned derivatives using a combination of various physicochemical, steric, electronic, and structural molecular descriptors obtained by Density Functional Theory (DFT) method by employing Becke's three-parameter hybrid functional (B3LYP) and 6-31+G** basis set. By using the multiple linear regression (MLR) technique several QSAR models have been drown up with the help these calculated descriptors and the antibacterial activity of the benzimidazole derivatives. The stepwise regression method was used to derive the most significant models as a calibration model for predicting the $LogP_{o/w}$ of this class of molecules. Among the obtained QSAR models presented in the study, statistically the most significant one is a five parameters linear equation with the squared correlation coefficient R^2 value of 0.831. To confirm the predictive power of the models, an external set of molecules was used. High agreement between experimental and predicted ($LogP_{o/w}$) values, obtained in the validation procedure, indicated the good quality of the derived QSAR models.*

Keywords: Partition coefficient, QSAR; benzimidazole derivatives; MLR, DFT, descriptor.

INTRODUCTION

The benzimidazole nucleus is an important heterocyclic ring, and interest in the chemistry,synthesis and microbiology of this pharmacophore continues to be fuelled by its antifungal [1],antitubercular [2], antioxidant [3, 4], and antiallergic [5, 6] properties. Other reports have revealed that these molecules are also present in a variety of antiparasitic [7, 8] and herbicidal agents. Albendazole, fenbendazole and their sulphoxide derivatives are

methylcarbamate benzimidazoles with a broad spectrum anthelmintic activity, widely used in human and veterinary medicine [9]. They are used against several systemic parasitoses, including nematodoses, hidatidosis, teniasis and others [10]. Although a variety of benzimidazole derivatives are known, the development of new andconvenient strategies to synthesize new biologically active benzimidazoles is of considerable interest. QSAR studies are useful tools in the rational search for bioactive molecules. The main success of the QSAR method is the possibility to estimate the characteristics of new chemical compounds without the need to synthesize and test them. This analysis represents an attempt to relate structural descriptors of compounds with their physicochemical properties and biological activities. This is widely used for the prediction of physicochemical properties in the chemical, pharmaceutical, and environmental spheres. This method included data collection, molecular descriptor selection, correlation model development, and finally model evaluation. QSAR studies have predictive ability and simultaneously provide deeper insight into mechanism of drug receptor interactions [11,12]. In view of the above and in continuation of our studies on the partition coefficient octanol- water ($LogP_{o/w}$) of benzimidazole derivatives, as well as on correlation of molecular properties with activity [13-20], the objective of this investigation was to study the usefulness of QSAR in the prediction of the antibacterial activity of benzimidazole derivatives antiparasitic . Multiple linear regression (MLR) models have been developed as a mathematical equation which can relate chemical structure to the ($LogP_{o/w}$).

Molecular descriptor generation

To derive QSAR models, an appropriate representation of the chemical structure is necessary. For this purpose, descriptors of the structure are commonly used. These descriptors are generally understood as being any term, index or parameter conveying structure information. Commonly used descriptors in the QSAR analysis are presented in Table 1. Some of the descriptors are obtained directly from the chemical structure, e. g. constitutional, geometrical, and topological descriptors. Other chemical and physicochemical properties were determined by the chemical structure (lipophilicity, hydrophilicity descriptors, electronic descriptors, energies of interaction). In this work, we used Gaussian 03 for ab initio calculations. B3LYP method at 6-31+G** were applied for optimization of Albenzdazole derivatives and calculation of many of the descriptors. At first Benzimidazole derivatives were built by Hyperchem software and some o the descriptors such as surface area, hydration energy, and refractivity were calculated through it. The rest of the descriptors were obtained of Gaussian calculations. A large number of descriptors were calculated by Gaussian package and Hyperchem software. One way to avoid data redundancy is to exclude descriptors that are highly intercorrelated with each other before performing statistical analysis. Reduced multi collinearity and redundancy in the data will facilitate selection of relevant variables and models for the investigated endpoint. Variable-selection for the QSAR modeling was carried out by stepwise linear regression method. A stepwise technique was employed that only one parameter at a time was added to a model and always in the order of most significant to least significant in terms of F-test values. Statistical parameters were calculated subsequently for each step in the process, so the significance of the added parameter could be verified. The goodness of the correlation is tested by the regression coefficient (R^2), the F-test and the standard error of the estimate (SEE). The test and the level of significance, as well as the confidence limits of the regression coefficient, are also reported. The squared correlation

coefficient, R^2, is a measure of the fit of the regression model. Correspondingly, it represents the part of the variation in the observed (experimental) data that is explained by the model.

Table 1. The calculated descriptors used in this study

Descriptors	Symbol	Abbreviation	Descriptors	Symbol	Abbreviation
Quantum chemical descriptors	Molecular Dipole Moment	MDP	Quantum chemical descriptors	difference between LUMO and HOMO	E_{GAP}
	Molecular Polarizability	MP		Hardness [η=1/2 (HOMO+LUMO)]	H
	Natural Population Analysis	NPA		Softness (S=1/ η)	S
	Electrostatic Potentialc	EP		Electro negativity [χ= -1/2 (HOMO–LUMO)]	*X*
	Highest Occupied Molecular Orbital	HOMO		El Electro philicity ($\omega=\chi^2/2\eta$)	Ω
	Lowest Unoccupied Molecular Orbital	LUMO		MullikenlChargeg	MC
	Partition Coefficient	Log P		Molecule surface area	SA
Chemical properties	Mass	M	Chemical properties	Hydration Energy	HE
	Molecule volume	V		Refractivity	REF

EXPERIMENTAL SECTION

Molecular Modeling

All molecular modeling studies were performed by using HyperChem 7.5 software (HyperCube Inc,Version 7.5) running on a P-III processor [26]. HyperChem includes a model builder that turns a rough2D sketch of a molecule into a 3D one. In this article, a QSAR study of 28 benzimidazole derivatives. was performed based on the theoretical molecular descriptors calculated by the GAUSSIAN software and selected.

Descriptor Generation

The QSAR developed indicated that Natural Population Analysis (NPA_8), Electro philicity(η), Electro negativity(X) Mullikan charge (MC_7) and Partition coefficient (LogP) compound $LogP_{o/w}$. negative values in the regression coefficients indicate that the indicated descriptor contributes positively to the value of $LogP_{o/w}$, whereas positive values indicate that the greater the value of the descriptor the lower the value of $LogP_{o/w}$. In other words, increasing the X, η and NPA_8 will decrease $LogP_{o/w}$ and increasing the LogP increases extent of $LogP_{o/w}$ of the benzimidazole derivatives. The standardized regression coefficient reveals the significance of an individual descriptor presented in the regression model.

RESULTS AND DISCUSSION

In the first step of the present study, different substituted benzimidazoles (Figure 1) were evaluated for *in vivo* antiparasitic Benzimidazole drugs derivatives . The $LogP_{o/w}$ effects of compounds **1 – 28**, expressed as $LogP_{o/w}$, are summarized in Table 2.

Figure1. Chemical structures of the drugs

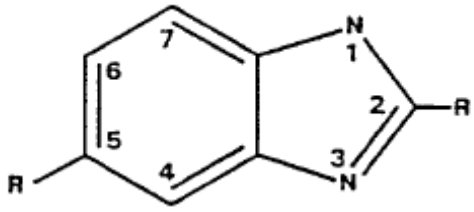


Table 2. Experimental values of $LogP_{o/w}$ for Benzimidazole derivatives training set.

No.Train	Name	EXP.	Pred(Train)
1	Albendazole	3.83	3.015267793
2	Benzimidazole	1.5	2.083179653
3	Thiabendazole	2.55	1.570662517
4	Pantoprazole	0.5	1.055525925
5	Oxibendazole	2.6	2.639761207
6	Fenbendazole	3.93	2.50023484
7	Flubendazole	3.32	3.387168456
8	Albendazole sulfoxide	1.24	0.671945438
9	Omeprazole	1.8	1.632912279
10	Droperidol	2.8	9.566195596
11	Compound1	1.97	3.497098754
12	Compound2	2.37	1.923352162
13	Compound4	1.97	1.734427922
14	Compound5	12.5	1.037985647
15	I-CH3	4.85	4.144283077
16	I-F	4.5	5.436503837
17	I-CL	5.05	4.162770217
18	II-CH3	3.6	3.004541643
19	II-F	3.25	4.24416231
20	II-OCH3	3.04	2.610421924
21	III-CH3	4.55	4.0391149
22	III-F	4.2	4.279382571
23	III-CL	4.78	3.396029489
24	III-OCH3	3.97	3.445398993
25	Emedastine	2.6	2.032101212
26	Candesartan	6.1	5.470055474
27	5,6-dimethylbenzimidazole	2.35	1.050705671
28	Oxfendazole	2.03	3.335288725

In the second step, we focused our efforts on developing the QSAR models of compounds **1** – **28** as antiparasitic. A set of benzimidazoles was used for MLR model generation. Developing a QSAR model requires a diverse set of data, and, thereby a large number of descriptors have to be considered. Descriptors are numerical values that encode different structural features of the molecules. Selection of a set of appropriate descriptors from a large number of them requires a method, which is able to discriminate between the parameters. Pearson's correlation matrix has been performed on all descriptors by using SPSS Software. The analysis of the matrix revealed nine descriptors for the development of MLR model. Linear models were then formed by a stepwise addition of terms. A delition process was then employed, whereby each variable in the model was held out in turn and using the remaining parameters models were generated. Each descriptor was chosen as input for the statistical software package and then the stepwise addition method implemented in the software was used for choosing the descriptors contributing to the antibacterial activity of benzimidazole derivatives.

The specifications for the best-selected MLR models are shown in under.

$$Logp=0.003(\pm0.001)hf+711.706(\pm146.308)\ \eta+4.813(\pm0.794)MC_7-36.321(\pm4.188)NPA_8+195.483(\pm33.959)\ X+287.985(\pm31.469)$$

N=28 R= 0.911 $R^2_{adj=}0.8$ SEE=0.9338 F=0.27474 $R^2=0.831$

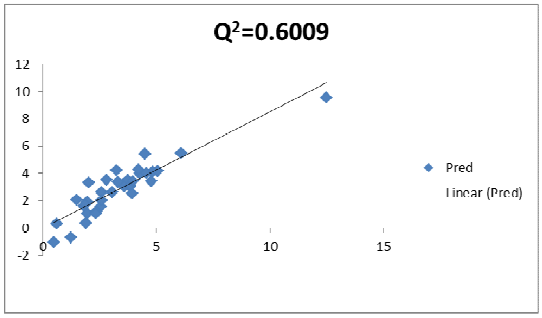
It is well known that there are three important components in any QSAR study: development of models, validation of models and utility of developed models. Validation is a crucial aspect of any QSAR analysis [21]. The statistical quality of the resulting models, as depicted in up, is determined by *r*, *s*, and *F* [22-24]. It is noteworthy that all these equations were derived using the entire data set of compounds (n = 28) and no outliers were identified. The *F*-value presented in up is found tatistically significant at 99% level since all the calculated *F* values are higher as compared to tabulated values. For the testing the validity of the predictive power of selected MLR models the LOO technique was used.

The PRESS value above can be used to compute an r^2CV statistic, called r^2 cross validated, which reflects the prediction ability of the model. This is a good way to validate the prediction of a regression model without selecting another sample or splitting your data. It is very possible to have a high r^2 and a very low r^2 CV. When this occurs, it implies that the fitted model is data dependent. This r^2 CV ranges from below zero to above one. When outside the range of zero to one, it is truncated to stay within this range. Adjusted *r*-squared (r^2 adj) is an adjusted version of r^2. The adjustment seeks to remove the distortion due to a small sample size In many cases r^2CV and *r*2adj is taken as a proof of the high predictive ability of QSAR models. A high value of these statistical characteristic (> 0.5) is considered as a proof of the high predictive ability of the model, although recent reports have proven the opposite [25]. Although a low value of *r*2CV for the training set can indeed serve as an indicator of a low predictive ability of a model, the opposite is not necessarily true. Indeed, the high r^2CV does not imply automatically a high predictive ability of the model. Thus, the high value of LOO *r*2 CV is the *necessary* condition for a model to have a high predictive power, it is not a *sufficient* condition. It is proven that the only way to estimate the true predictive power of a model is to test it on a sufficiently large collection of compounds from an external test set. The test set must include no less than five compounds, whose activities and structuresmust cover the range of activities and structures of compounds from the training set. This application is necessary for obtaining trustful statistics for comparison between the observed and predictedactivities for these compounds. Besides high r^2 CV, a reliable model should be also characterized by a high correlation coefficient between the predicted and observed activities of compounds from a test set of molecules that was not used to develop the models. To confirm the predictive power of the QSAR models, an external set of benzimidazoles was used. Six benzimidazole derivatives which were tested in our previous paper for their the partition coefficient octanol- water $LogP_{o/w}$ were used as the external set of molecules [18]. The values of $LogP_{o/w}$ of a test set of molecules was calculated with the models 1.These data are compared with experimentally obtained values of $LogP_{o/w}$.

Table 3. Experimental values of $LogP_{o/w}$ for benzimidazole derivatives test set.

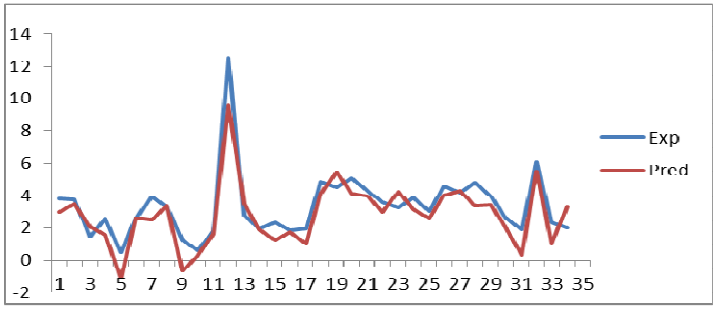
No.Test	Name	EXP.	Pred(test)
29	Mebendazole	3.73	3.507075465
30	Rabeprazole	0.6	0.285986583
31	Compound3	1.85	1.23393818
32	I-OCH3	4.28	3.965537466
33	II-CL	3.85	3.138764322
34	Lansoprazole	1.9	0.338254018

Figure 2. The predicted versus the experimental $LogP_{o/w}$ by MLR.



From the data presented in Table 3 and Figure3, it is shown that high agreement between experimental and predicted $LogP_{o/w}$ values was obtained (the residual values are small, indicating the good predictability of the established models.

Figure 3 has showed that results were obtained from equation B3LYP /6-31+G** to the experimental values.



Series 1: the values of log P were obtained by using prediction.
Series 2: the values of log P were obtained by using Experimental methods

Figure 3. The comparison between properties ($LogP_{o/w}$) using experimental and prediction

CONCLUSION

QSAR analysis have been used to study the quantitative effects of the molecular structure of the benzimidazoles on their inhibitory activity. Accurate mathematical models were developed for predicting the antibacterial activity of this class of compounds. The validity of the models have been established by the determination of suitable statistical parameters. From the established QSAR models, it was calculated of the partition coefficient octanol- water ($LogP_{o/w}$)of the external set of benzimidazoles and close agreement between experimental and predicted values

was obtained. The low residual activity and high cross-validated *r*2 values (*r*2 CV) observed indicated the predictive ability of the developed QSAR models.

REFERENCES

[1] Ayhan-Kilcigil, G.; Altanlar, N. *Turk. J. Chem.* **2006**, *30*, 223-228.
[2] Mohamed, B.G.; Hussein, M.A.; Abdel-Alim, A.M.; Hashem, M. *Arch. Pharm. Res.* **2006**, *29*, 26-33.
[3] Kus, C.; Ayhan-Kilcigil, G.; Can-Eke, B.; Iscan, M. *Arch. Pharm. Res.* **2004**, *27*, 156-163.
[4] Ayhan-Kilcigil, G.; Kus, C.; Coban, T.; Can-Eke, B.; Iscan, M. *J. Enz. Inhib. Med. Chem.* **2004**, *19*, 129-135.
[5] Nakano, H.; Inoue, T.; Kawasaki, N.; Miyataka, H.; Matsumoto, H.; Taguchi, T.; Inagaki, N.; Nagai, H.; Satoh, T. *Chem. Pharm. Bull.* **1999**, *47*, 1573-1578.
[6] Fukuda, T.; Saito, T.; Tajima, S.; Shimohara, K.; Ito, K. *Arzneim. Forsch./Drug Res.* **1984**, *34*, 805-810.
[7] Valdez, J.; Cedillo, R.; Hernandez-Campos, A.; Yepez, L.; Hernandez-Luis, F.; Navarrete-Vazquez, G.; Tapia, A.; Cortes, R.; Hernandez, M.; Castillo, R. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2221-2224.
[8] Navarrete-Vazquez, G.; Cedillo, R.; Hernandez-Campos, A.; Yepez, L.; Hernandez-Luis, F.; Valdez, J.; Morales, R.; Cortes, R.; Hernandez, M.; Castillo, R. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 187-190.
[9] Lipkowitz, K.B.; McCracken, R.O. *Parasitol. Res.* **1993**, *79*, 475-479.
[10] Campbell, W.C. *Parasitol. Today* **1990**, *6*, 130-133.
[11] Hansch, C.; Leo, A.; Hoekman, D.H. *Exploring QSAR, Fundamentals and Application in Chemistry and Biology*; American Chemical Society: Washington, DC, USA, 1995.
[12] Hansch, C.; Leo, A.; Hoekman, D.H. *Exploring QSAR, Hydrophobic, Electronic and Steric Constants*; American Chemical Society: Washington, DC, USA, 1995.
[13] Podunavac-Kuzmanović, S.O.; Markov, S.L.; Barna, D.J. *J. Theor. Comp. Chem*. **2007**, *6*, 687-698.
[14] Podunavac-Kuzmanović, S.O.; Cvetković, D.D. *CICEQ* **2007**, *13*, 68-71.
[15] Perišić-Janjić, N.U.; Podunavac-Kuzmanović, S.O. *J. Planar. Chromatogr*. **2008**, *21*, 135-141.
[16] Podunavac-Kuzmanović, S.O.;.Cvetković, D.D.; Barna, D.J. *J. Serb. Chem. Soc.* **2008**, *73*, 967-978.
[17] Podunavac-Kuzmanović, S.O.;.Cvetković, D.D.; Barna, D.J. *Chem. Listy* **2008**, *102*, 753-756.
[18] Podunavac-Kuzmanović, S.O.;.Cvetković, D.D.; Barna, D. *Chem. Listy* **2008**, *102*, 757-761.
[19] Podunavac-Kuzmanović, S.O.;.Cvetković, D.D. *Chem. Listy* **2008**, *102*, 762-764.
[20] Podunavac-Kuzmanović, S.O.; Leovac, V.M.; Cvetković, D.D. *J. Serb. Chem. Soc*. **2008**, *73*, 1153-1160.
[21] Topliss, J.G.; Edwards, R.P. *J. Med. Chem.* **1979**, *22*, 1238-1244.
[22] Snedecor, G.W.; Cochran, W.G. *Statistical Methods*; Oxford and IBH: New Delhi, India, 1967; p. 381.
[23] Chaltterjee, S.; Hadi, A.S.; Price, B. *Regression Analysis by Examples*; Wiley VCH: New York, USA, 2000.

[24] Diudea, M.V. *QSPR/QSAR Studies for Molecular Descriptors*; Nova Science: Huntingdon, New York, USA, 2000.
[25] Golbraikh, A.; Tropsha, J. *J. Mol. Graph. Mod.* **2002**, *20*, 269-276.
[26] *HyperChem 7.5*. Hypercube, Inc.: Waterloo, ON, Canada.