Available on line www.jocpr.com

Journal of Chemical and Pharmaceutical Research

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

J. Chem. Pharm. Res., 2010, 2(6):306-315

Computational approaches to the prediction of the 1-octanol/water partition coefficient of benzimidazole derivatives drugs

Z. Bayat*, S. Qanei Nassab

Department of Chemistry, Islamic Azad University-Quchan Branch, Iran

ABSTRACT

*It is important to determine whether a candidate molecule is capable of penetrating the partition coefficient octanol- water ($LogP_{o/w}$) in drug discovery and development. The aim of this paper is to establish a predictive model for $LogP_{o/w}$ penetration using simple descriptors. The usefulness of the quantum chemical descriptors, calculated at the level of the HF theory using 6-31+G** basis set for QSAR study of Benzimidazole derivatives was examined. The QSAR model developed contributed to a mechanistic understanding of the investigated biological effects. In this study was to use a dataset containing 38 drugs with known activity. A multiple linear regression (MLR) procedure was used to model the relationships between molecular descriptors and the partition coefficient ($LogP_{o/w}$) of the Benz imidazole 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. Biological activities contain the logarithm of the ratio of the steady-state concentration of a compound in the 1-octanol to in the water, log P. A multi-parametric equation containing maximum four descriptors at HF/6-31+G** method with good statistical qualities (R_{MAX}= 0 .911, R^2_{MAX}= 0.829 Q^2=0.6877at HF/6-31+G**) was obtained by Multiple Linear Regression using stepwise method. The best QSAR models were further validated by a leave one out technique as well as by the calculation of statistical parameters for the established theoretical models. 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: QSAR; Benzimidazole derivatives; multiple linear regression; HF, Partition coefficient .

INTRODUCTION

The benzimidazoles are a large chemical family used as antimicrobial agents against the wide spectrum of microorganisms [1-9]. Because of its synthetic utility and broad range of pharmacological effects, 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 [10], antitubercular [11], antioxidant [12,13], and ant allergic [14,15] properties. Other reports have revealed that these molecules are also present in a variety of antiparasitic [16,17] and herbicidal agents [18]. Albendazole, fenbendazole and their sulphoxide derivatives are methylcarbamate benzimidazoles with a broad spectrum anthelmintic activity, widely used in human and veterinary medicine [19]. They are used against several systemic parasitoses, including nematodoses, hidatidosis, teniasis and others [20]. They are also used to treat microsporodial and cryptosporodial infections, which can cause lethal diarrhea in patients treated with immunosuppressive drugs, or infected with HIV [21,22]. The *n*-octanol/water partition coefficient is the ratio of the concentration of a chemical in *n*-octanol to that in water in a two-phase system at equilibrium. The logarithm of this coefficient, log $P_{o/w}$, has been shown to be one of the key parameters in quantitative structure activity/property relationship (QSAR/QSPR) studies. The octanol–water partition coefficient is a measure of the hydrophobicity and hydrophilicity of a substance. Hydrophobic "bonding" is actually not bond formation at all, but rather the tendency of hydrophobic molecules or hydrophobic parts of molecules to avoid water because they are not readily accommodated in the highly ordered hydrogen bonded structure of water [23]. Hydrophobic interaction is favored thermodynamically because of increased entropy of the water molecules that accompanies the association of non-polar molecules, which squeeze out water. There are some reports about the applications of MLR [24–27] and artificial neural network [28–31] modeling to predict the *n*-octanol/water partition coefficient of anti-cancer drugs. In our previous papers, we reported on the application of QSAR techniques in the development of a new, simplified approach to prediction of compounds properties [32–36]. In this work a QSAR study is performed, to develop models that relate the structures of a heterogeneous group of 34 drug compounds to their *n*-octanol–water partition coefficients. However, using in vivo methods to measure the logarithmic values of partition coefficient drug concentration ratios (log P) in humans is not possible, and to do so in animal models is expensive and time consuming. Finally, the accuracy of the proposed model was illustrated using the following: leave one out, bootstrapping and external test set, cross-validations and chance correlation techniques.

EXPERIMENTAL SECTION

Methodology

Data set

The properties data used in this study are the $LogP_{o/w}$ of the set of 34 Benz imidazole derivatives [37-50]. The data set was randomly divided into two subsets: the training set containing 38 compounds (80%) and the test set containing ? compounds (20%). The training set was used to build a regression model, and the test set was used to evaluate the predictive ability of the model obtained. The properties data for the complete set of compounds are presented in Table 1, 2.To derive QSAR models, an appropriate representation of the chemical structure is necessary. For this purpose, descriptors of the structure are commonly used.

Table 1. Experimental values of $LogP_{o/w}$ for Benzimidazole derivatives training set

No.Train	Name	EXP.	Pred(Train)
1	Albendazole	3.83	4.307526056
2	Fenbendazole	3.93	4.465238092
3	Benzimidazole	1.5	1.205592796
4	Thiabendazole	2.55	2.231039768
5	Pantoprazole	0.5	0.16806728
6	Oxibendazole	2.6	2.944894664
7	Flubendazole	3.32	1.289989084
8	Albendazole sulfoxide	1.24	1.13785822
9	Omeprazole	1.8	1.301367268
10	Droperidol	2.8	9.89199972
11	Compund1	1.97	3.31941264
12	Compund2	2.37	1.732475376
13	Compund4	1.97	1.777871616
14	Compund5	12.5	1.98206018
15	I-CH3	4.85	5.010630124
16	I-F	4.5	4.81403246
17	I-CL	5.05	4.96596986
18	II-CH3	3.6	4.339923672
19	II-F	3.25	3.673920828
20	II-OCH3	3.04	4.020600128
21	III-CH3	4.55	4.604129552
22	III-F	4.2	4.203018192
23	III-CL	4.78	4.1044666
24	III-OCH3	3.97	4.406944408
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

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

No.Test	Name	EXP.	Pred(test)
29	Mebendazole	3.73	4.226592652
30	Rabeprazole	0.6	0.89937306
31	Compund3	1.85	1.425383056
32	I-OCH3	4.28	4.990333504
33	II-CL	3.85	3.735065324
34	Lansoprazole	1.9	1.276033196

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 2. 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. HF method at 6-31+G** were

applied for optimization of Albenzdazole derivatives and calculation of many of the descriptors. At first Benz imidazole 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.

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		Electro philicity (ω=χ^2 /2 η)	ω
	Lowest Unoccupied Molecular Orbital	LUMO		Mullikenl Chargeg	MC
Chemical properties	Partition Coefficient	Log P	Chemical properties	Molecule surface area	SA
	Mass	M		Hydration Energy	HE
	Molecule volume	V		Refractivity	REF

RESULTS AND DISCUSSION

In a QSAR study, generally, the quality of a model is expressed by its fitting ability and prediction ability, and of these the prediction ability is the more important. In order to build and test the model, a data set of 34 compounds was separated into a training set of 28 compounds, which were used to build the model and a test set of 6 compounds, which were applied to test the built model. With the selected descriptors, we have built a linear model using the training set data, and the following equation was obtained:

$$Logp=-17.966(\pm 8.607)\ Ep_{2+}0.317(\pm 0.045)SAG_6+3.871(\pm 0.391)\ He_8+0.521(\pm 0.052)SAG_8-262.949(\pm 126.683)$$

R_{train}=0 .911 R_{train}^2=0.829 R^2_{adj}=0.806 SEE=0.9218 F_{train}=0.35179 N_{train}=28
N_{test}=6 N=34 Q2=0.6877

In this equation, N is the number of compounds, R^2 is the squared correlation coefficient, Q^2 is the squared cross-validation coefficient, and F is the Fisher F statistic. The figures in parentheses the are the standard deviations. The built model was used to predict the test set data and prediction results are given in Table 1.As can be seen from Table 1, the calculated values for the $LogP_{o/w}$ are in good agreement with those of the experimental values. The predicted values for $LogP_{o/w}$ for the compounds in the training and test sets using equation $LogP_{o/w}$ were plotted against the experimental $LogP_{o/w}$ values in Figure 1.and the comparison between $LogP_{o/w}$ using prediction and the experimental .A plot of the residual for the predicted values of RI for both the training and test sets against the experimental $LogP_{o/w}$ values are shown in Figure 2. As can be seen the model did not show any proportional and systematic error, because the propagation of the residuals on both sides of zero are random. The real usefulness of QSAR models is not just their ability to reproduce known data, verified by their fitting power (R^2), but is mainly their potential for predictive application. For this reason the model calculations were performed by maximising the explained variance in prediction, verified by the cross-validated correlation coefficient,Q^2. This indicates that the obtained regression model has a good internal and external predictive power.

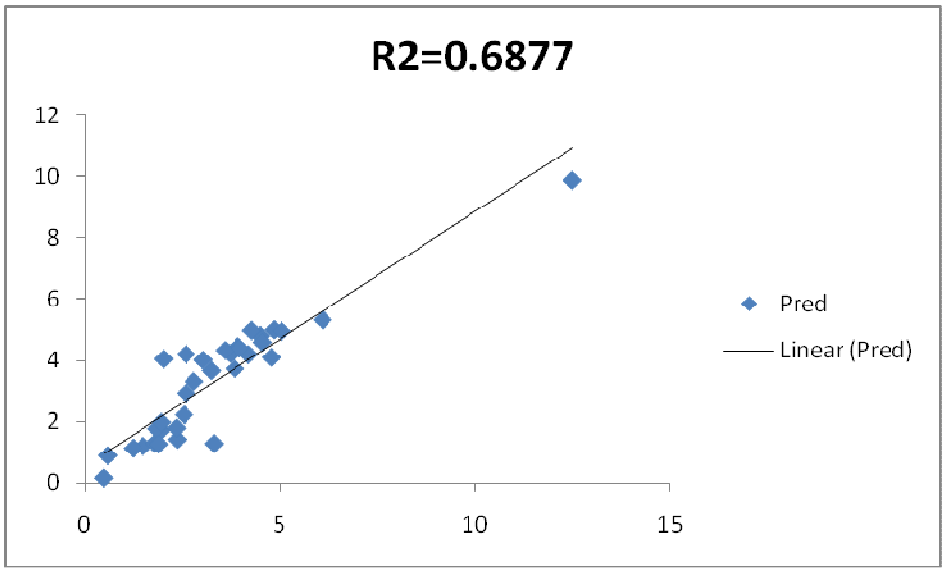


Figure1. The predicted versus the experimental $LogP_{o/w}$ by MLR

Also, in order to assess the robustness of the model, the chance correlation test was applied in this study. The dependent variable vector ($LogP_{o/w}$) was randomly shuffled and The new QSAR models (after several repetitions) would be expected to have low R^2 and R values (Table 3). If the opposite happens then an acceptable QSAR model cannot be obtained for the specific modeling method and data.

The MLR analysis was employed to derive the QSAR models for different Benzimidazole derivatives. MLR and correlation analyses were carried out by the statistics software SPSS (Table 4).

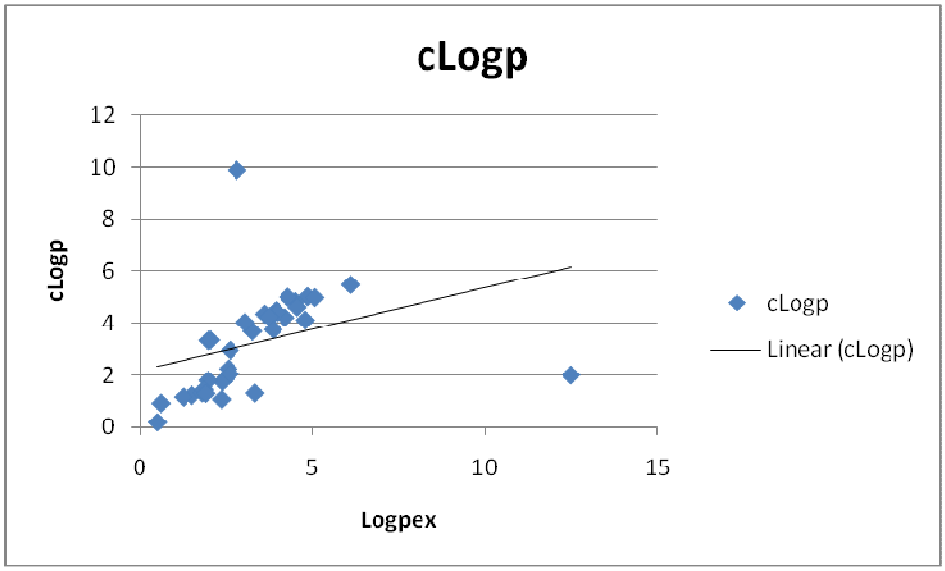


Figure 2. The residual versus the experimental LogP byMLR.
(See colour version of this figure online at www.informahealthcare.com/enz)

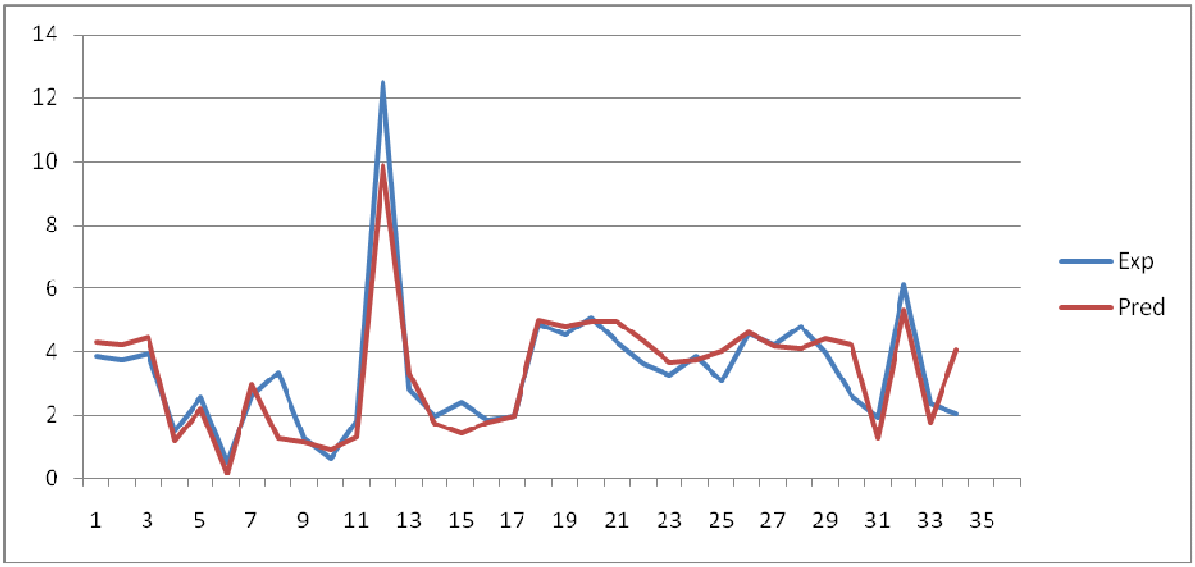
Table 3. The R^2_{train} and R values after several chance correlation tests

No	R^2	R
1	0.486	0.697
2	0.442	0.665
3	0.326	0.571
4	0.488	0.699
5	0.38	0.616
6	0.461	0.679
7	0.425	0.652
8	0.42	0.684
9	0.41	0.641
10	0.429	0.655

Table 4. The correlation coefficient existing between the variables used in different MLR and equations with HF/6-31+G method**

	logpex	hf	sag6	sag8	he8	pot2
logpex	1	0	0	0	0	0
hf	0.283	1	0	0	0	0
sag6	-0.353	-0.097	1	0	0	0
sag8	0.466	0.447	-0.792	1	0	0
he8	0.171	-0.444	0.151	-0.597	1	0
pot2	-0.386	-0.48	0.553	-0.514	0.164	1

Figure 2 has showed that results were obtained from equation HF/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 2. The comparison between properties ($LogP_{o/w}$) using experimental and prediction

Interpretation of descriptors

The QSAR developed indicated that Nuclear magnetic Resonance (σ_5), mass (M), Mullikan charge (MC_4) 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 σ_5, M and M C_4 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.

CONCLUSION

In this article, a QSAR study of 34 benzimidazole derivatives. was performed based on the theoretical molecular descriptors calculated by the GAUSSIAN software and selected. The built model was assessed comprehensively (internal and external validation) and all the validations indicated that the QSPR model built was robust and satisfactory, and that the selected descriptors could account for the structural features responsible for the benzimidazole derivatives properties of the compounds. The QSPR model developed in this study can provide a useful tool to predict the $LogP_{o/w}$ of new compounds and also to design new compounds with high $LogP_{o/w}$.

REFERENCES

[1] Nguyen, P.T.M.; Baldeck, J.D.; Olsson, J.; Marquis, R.E. *Oral Microbiol. Immunol*. **2005**, 20, 93-99.
[2] Kazimierczuk, Z.; Upcroft, J.A.; Upcroft, P.; Gorska, A.; Starosciak, B.; Laudy, A. *Acta Biochim. Polon.* **2002**, 49, 185-195.
[3] Goker, H.; Alp, M.; Yildiz, S. *Molecules* **2000**, 10, 1377-1386.
[4] Podunavac-Kuzmanović, S.O.; Cvetković, D.D. *Centr. Eur. J. Occupat. Environ. Med*. **2006**, 12, 55-60.
[5] Perišić-Janjić, N.U.; Podunavac-Kuzmanović, S.O.; Balaž, J.S.; Vlaović, Đ. *J. Planar. Chromatogr*. **2000**, 13, 123-129.
[6] Podunavac-Kuzmanović, S.O.; Leovac, V.M.; Perišić-Janjić, N.U.; Rogan, J.; Balaž, J., *J. Serb. Chem. Soc*. **1999**, 64, 381-388.
[7] Podunavac-Kuzmanović, S.O.; Markov, S.L. *Centr. Eur. J. Occupat. Environ. Med*. **2006**, 12, 61-66.
[8] Podunavac-Kuzmanović, S.O.; Cvetković, D. *J. Serb. Chem. Soc*. **2007**, 75, 459-466.
[9] Ates-Alagoz, Z.; Yildiz, S.; Buyukbingol, *E. Chemotherapy* **2007**, 53, 110-113.
[10] Ayhan-Kilcigil, G.; Altanlar, N. *Turk. J. Chem*. **2006**, 30, 223-228.
[11] Mohamed, B.G.; Hussein, M.A.; Abdel-Alim, A.M.; Hashem, M. *Arch. Pharm. Res*. **2006**, 29, 26-33.
[12] Kus, C.; Ayhan-Kilcigil, G.; Can-Eke, B.; Iscan, M. *Arch. Pharm. Res*. **2004**, 27, 156-163.
[13] Ayhan-Kilcigil, G.; Kus, C.; Coban, T.; Can-Eke, B.; Iscan, M. *J. Enz. Inhib. Med. Chem*. **2004**, 19, 129-135.
[14] 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.
[15] Fukuda, T.; Saito, T.; Tajima, S.; Shimohara, K.; Ito, K. Arzneim. *Forsch./Drug Res*. **1984**, 34, 805-810.
[16] 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.
[17] 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.
[18] Ward, C.; Berthold, R.; Koerwer, J.; Tomlin, J.; Manning, T. *J. Agric. Food Chem*. **1986**, 34, 1005- 1010.
[19] Lipkowitz, K.B.; McCracken, R.O. *Parasitol. Res.* **1993**, 79, 475-479.
[20] Campbell, W.C. *Parasitol. Today* **1990**, 6, 130-133.
[21] Didier, E.S.; Stovall, M.E.; Green, L.C.; Brindley, P.J.; Sestak, K.; Didier, P.J. Vet. *Parasitol*. **2004**, 126, 145-166.
[22] Zulu, I.; Veitch, A.; Sianongo, S.; McPhail, G.; Feakins, R.; Farthing, M.J.G.; Kelly, P. *Alim. Pharmacol. Ther*. **2002**, 16, 595-601.
[23] Pandey, V.K.; Upadhay, M.; Dev Gupta, V.; Tandon, M. *Acta Pharm*. **2005**, 55, 47-56.
[24] Beaulieu, P.L.; Bousquet, Y.; Gauthier, J.; Gillard, J.; Marquis, M.; McKercher, G.; Pellerin, C.; Valois, S.; Kukolj, G. *J. Med. Chem*. **2004**, 47, 6884-6892.
[25] Patel, P.D.; Patel, M.R.; Kaushik-Basu, N.; Talele, T.T. *J. Chem. Inf. Model* **2008**, 48, 42-55.

[26] Hansch, C.; Leo, A.; Hoekman, D.H. Exploring QSAR, Fundamentals and Application in Chemistry and Biology; American Chemical Society: Washington, DC, USA, **1995**.
[27] Hansch, C.; Leo, A.; Hoekman, D.H. Exploring QSAR, Hydrophobic, Electronic and Steric Constants; American Chemical Society: Washington, DC, USA, **1995**.
[28] Podunavac-Kuzmanović, S.O.; Markov, S.L.; Barna, D.J. *J. Theor. Comp. Chem*. **2007**, 6, 687-698.
[29] Podunavac-Kuzmanović, S.O.; Cvetković, D.D. CICEQ 2007, 13, 68-71.
[30] Perišić-Janjić, N.U.; Podunavac-Kuzmanović, S.O. *J. Planar. Chromatogr*. **2008**, 21, 135-141.
[31] Podunavac-Kuzmanović, S.O.;.Cvetković, D.D.; Barna, D.J. *J. Serb. Chem. Soc*. **2008**, 73, 967-978.
[32] Podunavac-Kuzmanović, S.O.;.Cvetković, D.D.; Barna, D.J., *Chem. Listy* **2008**, 102, 753-756.
[33] Podunavac-Kuzmanović, S.O.;.Cvetković, D.D.; Barna, D.J, *Chem. Listy* **2008**, 102, 757-761.
[34] Podunavac-Kuzmanović, S.O.;.Cvetković, D.D. *Chem. Listy* **2008**, 102, 762-764.
[35] Podunavac-Kuzmanović, S.O.; Leovac, V.M.; Cvetković, D.D. *J. Serb. Chem. Soc*. **2008**, 73, 1153-1160.
[36] Topliss, J.G.; Edwards, R.P. *J. Med. Chem*. **1979**, 22, 1238-1244.
[37] M.L. Mottier, L.I. Alvarez, M.A. Pis, and C.E. Lanusse, *Experimental Parasitology* 103 (**2003**) 1–7
[38] 1.Oxberry ME, Reynoldson JA, Thompson RC: *J Vet Pharmacol Ther*. **2000**, Jun; 23(3): 113-20.
2.Ramirez T, Benitez-Bribiesca L, Ostrosky-Wegman P, Herrera LA: *Arch Med Res*. **2001** Mar-Apr;32(2):119-22.
[39] Julio C. Rivera, Lili´an Y´epez-Mulia, Alicia Hern´andez-Campos, Rafael Moreno-Esparza, Rafael Castillo, Gabriel Navarrete-V´azquez, In´es Fuentes-Noriega, Helgi Jung-Cook, *International Journal of Pharmaceutics* 343 (**2007**) 159–165
[40] S. O. Podunavac-Kuzmanovic, D. D. Cvetkovic and D. J. Barna, *J. Serb. Chem. Soc*. 73 (10) 967–978 (**2008**) JSCS–3778
[41] Martin Danaher ,Hendrik De Ruyck, Steven R.H. Crooks,Geraldine Dowling, Michael O'Keeffe , *Journal of Chromatography* B, 845 (**2007**) 1–37
[42] 1.Molina AJ, Merino G, Prieto JG, Real R, Mendoza G, Alvarez AI: *Eur J Pharm Sci*. **2007** May;31(1):16-24. Epub 2007 Feb 6.
2.Chemistry portal
[43] Nobuyuki Okamura, Takahiro Suemoto, Shozo Furumoto, Masako Suzuki, Hiroshi Shimadzu, Hiroyasu Akatsu,Takayuki Yamamoto, Hironori Fujiwara, Miyako Nemoto, Masahiro Maruyama, Hiroyuki Arai, Kazuhiko Yanai,Tohru Sawada, and Yukitsuka Kudo, The *Journal of Neuroscience*, November 23, **2005** • 25(47):10857–10862.
[44] 1.Newman PR, Greeb J, Keeton TP, Reyes AA, Shull GE: *DNA Cell Biol*. **1990** Dec;9(10):749-62.
2. .www. Drugs.com
[45] 1.Solana HD, Sallovitz JM, Lanusse CE, Rodriguez JA: *Methods Find Exp Clin Pharmacol*. **2002** Jan-Feb;24(1):7-13
2. .www. Drugs.com

[46] 1.Ma JY, Song YH, Sjostrand SE, Rask L, Mardh S: *Biochem Biophys Res Commun*. **1991** Oct 15;180(1):39-45.
2. www. Drugs.com
[47] 1. Guenette S, Prichard RK, Klein RD, Matlashewski G: *Mol Biochem Parasitol*. **1991** Feb;44(2):153-64.
2. www. Drugs.com
[48] 1.Helm R, Selkirk ME, Bradley JE, Burns RG, Hamilton AJ, Croft S, Maizels RM: *Parasite Immunol*. **1989** Sep;11(5):479-502
2. www. Drugs.com
[49] 1. Shull MM, Lingrel JB: Proc Natl Acad Sci U S A. **1987** Jun;84(12):4039-43.
2. www. Drugs.com
[50] 1. Maeda M, Oshiman K, Tamura S, Futai M: *J Biol Chem*. **1990** Jun 5;265(16):9027-32.
2. www. Drugs.com