J. Indian Chem. Soc.,
Vol. 97, No. 9b, September 2020, pp. 1590-1595

DFT based QSAR studies on 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives as an anti-malarial agent

Ananda Sarkar[a], Bikash Kumar Sarkar[b] and Atish Dipankar Jana*[c]

[a]Department of Physics, Acharya Prafulla Chandra College, Kolkata-700 131, India

[b]Department of Physics, Mrinalini Datta Mahavidyapith, Birati, Kolkata-700 051, India

[c]Department of Physics, Behala College, Parnasree, Kolkata-700 060, India

E-mail: adjana2014@gmail.com, atishdipankarjana@yahoo.in

Manuscript received online 01 June 2020, accepted 29 August 2020

In the pursuit of better anti-malarial drugs, a quantitative structure-activity relationship (QSAR) studies have been performed on a series of 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives. The derived QSAR model is based on three molecular descriptors namely highest occupied molecular orbital (HOMO) energy, electrophilic group frontier electron density and nucleus independent chemical shift (NICS) obtained from the density functional theory (DFT) based energy minimized and geometry optimized structures of the molecules. The best QSAR model has a square correlation coefficient q^2 = 0.900 and cross-validated square correlation coefficient. The model is also tested successfully from external validation criteria (r^2_{pred} = 0.920).

Keywords: DFT, QSAR, NICS, group frontier electron density, anti-malaria.

Introduction

Malaria is one of the most devastating tropical diseases with over 660 million reports of infection per year and more than one million deaths annually according to a World Health Organization (WHO) estimate[1]. It is a protozoal infection and the parasite is capable of quick adaptation – becomes drug-resistant[2,3] within a short period of time. Problems associated in controlling vectors are responsible for the continuous rise in the occurrences of malaria infection[4]. The efforts on vaccine development have not yet made any significant success towards controlling this disease[5]. Development of new drugs that are effective against the resistant plasmodium falciparum is still an active area of research[6,7]. Although the discovery of artemisinin, an endoperoxide sesquiterpene lactone and a number of its analogues including trioxane dimer 1,2,4-trioxane have shown high anti-malarial activity[8,9] by destroying plasmodium falciparum, still the search for more potent anti-malarial drugs with increased half-lives and minimum side effects is being vigorously pursued[10,11].

In the present paper, molecular modelling studies of a series of 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives[12] have been undertaken by using Quantitative Structure Activity Relationships (QSAR). The QSAR model is based on DFT based quantum chemical descriptors. Highest occupied molecular orbital (HOMO) energy, electrophilic group frontier electron density (F_g^E)[13] and nucleus independent chemical shift (NICS)[14,15] of the molecules act as effective quantum chemical descriptors giving a very good QSAR model that satisfactorily explain the anti-malarial activity for this class of molecules. This study is likely to provide useful guidelines for the design of new inhibitors with better activities.

Materials and methods

Theoretical background:

In establishing a successful QSAR model, appropriate choice of molecular descriptors is the key factor. In the literature, there are reports of quite a large number of descriptors[13,15,16]. In the QSAR model explored here, the nucleus independent chemical shift (NICS), group frontier electron density (F_g) and HOMO energy have been found to be appropriate descriptors for the chosen set of molecules.

The Nucleus Independent Chemical Shift (NICS):

The nucleus independent chemical shift (NICS) is a widely

used magnetic criterion of aromaticity for chemical systems containing cyclic rings which was proposed by Schleyer *et al.*[14]. For aromatic systems, the NICS parameter is obtained as the negative value of absolute magnetic shielding factors, denoted as NICS(0) and NICS(1). NICS(0) is computed at the ring centroid, whereas NICS(1) is computed 1 Å above the ring centroid. In general, the more negative is the NICS, the more is the aromaticity of the rings and vice versa. A high positive value of NICS indicates anti-aromaticity. The system having NICS values within a short range around zero is non-aromatic. NICS constitute a measure of ring current of π-aromatic systems but the current strength of NICS(0) is contaminated by the ring-current of σ-bonds. Hence, NICS(1) values are often considered to be better, because the ring current 1 Å above the centroid is mainly constituted of π-electrons (Fig. 1). As NICS(1) is a measure of π-electron ring current of aromatic systems, it can be a good descriptor presenting π-interactions between protein and drugs. Recently, for the first time, we have shown that NICS(1) as a quantum chemical descriptor leads to high correlation coefficient in a QSAR model for COX-2 inhibitors[15].

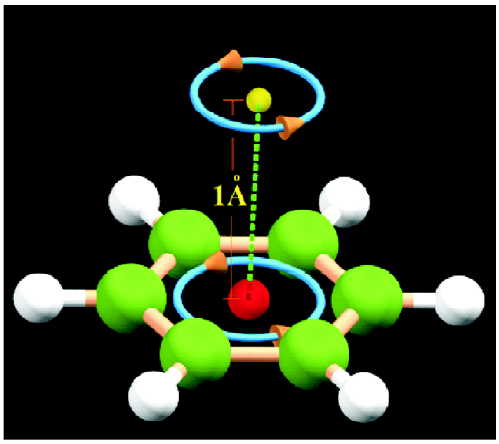


Fig. 1. Ring currents on and 1 Å above an aromatic ring plane. NICS(0) and NICS(1) are related to the strength of the ring current in the plane and 1 Å above plane respectively.

Group frontier electron density (F_g):

Frontier electron density refers to the electron distribution associated with the two frontier orbitals, namely the HOMO and the LUMO. Over the years it has become clear that these two molecular orbitals play a very important role in a wide range of chemical reactions of saturated and unsaturated molecules determining their reactivity. To quantify the role of electron density of the frontier orbitals, Karelson *et al.*[20] introduced two reactivity indices, namely electrophilic frontier electron density (F_k^E) and nucleophilic frontier electron density (F_k^N), defined respectively as

$$(F_k^E) = \frac{\Sigma(C_k^{HOMO})^2}{\Delta E} \times 100 \quad (1)$$

$$(F_k^N) = \frac{\Sigma(C_k^{LUMO})^2}{\Delta E} \times 100 \quad (2)$$

Here (C_k^{HOMO}) and (C_k^{LUMO}) are the coefficients of the atomic orbital of a particular atom (*k*-th) in the HOMO and LUMO respectively. ΔE is the HOMO-LUMO energy gap.

The above definition of the frontier electron density is local, in the sense that, it takes into account the contribution of a single atom in the frontier orbital electron density. Recently we have extended Karelson's definition of frontier electron density to a logically related set of atoms (such as an aromatic ring) that is part of a bigger molecule by defining the '*group frontier electron density*' which is the sum of frontier electron densities over the relevant group of atoms[14]. In a sense, this new reactivity index is semi-global, having been defined for a part of the molecule. Neither it is atom based nor it is defined for the whole molecule and is expected to capture the relevance of a correlated set of atoms in the intermolecular interactions. Based on the above concept of the 'group frontier electron density', two secondary reactivity indices have been further introduced to characterize the electrophilic and nucleophilic attacks, named respectively as '*electrophilic group frontier electron density*' (F_g^E) and '*nucleophilic group frontier electron density*' (F_g^N). These are defined as

$$F_g^E = \Sigma_{i=1}^n F_i^E \quad (3)$$

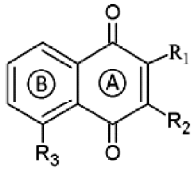
$$F_g^N = \Sigma_{i=1}^n F_i^N \quad (4)$$

where the summation is taken over a group of *n* relevant atoms.

Choice of dataset and computation of descriptors:

A series of 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives with their value against plasmodium falciparum, were taken from the literature[12] and is listed in Table 1. IC_{50} values of the molecules were transformed into log ($1/IC_{50}$) values, abbreviated as pIC_{50}. A test set of seven molecules have been used

Table 1. Chemical structure of 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives and their observed anti-malarial activities



No.	R_1	R_2	R_3	pIC_{50}
1	CH_3	H	OH	5.620
2	CH_3	Aziridin-1-yl	OH	5.824
3	$NHCH_3$	H	OH	5.432
4	$NH(CH_2)_2Cl$	H	OH	5.367
5	Aziridin-1-yl	H	$C_6H_5SO_3$	6.638
6	Aziridin-1-yl	H	CH_3-4- $C_6H_4SO_3$	6.699
7	Aziridin-1-yl	H	C_2H_5-4-$C_6H_4SO_3$	7.018
8#	Aziridin-1-yl	H	CH_3O-C_6H_4-SO_3	6.796
9	Aziridin-1-yl	H	F-4-$C_6H_4S0_3$	6.854
10	Aziridin-1-yl	H	I-4-$C_6H_4SO_3$	6.481
11	Aziridin-1-yl	H	$C_6H_5CHCHSO_3$	6.770
12#	Aziridin-1-yl	H	$(CH_3)_2CH_2$, 4,6-$C_6H_2SO_3$	6.553
13	Aziridin-1-yl	H	Naphtalene-1-sulfonyloxy	6.886
14	Aziridin-1-yl	H	$(CH_3)_2$N-5-Naphtalenel sulfonyloxy	6.886
15#	Aziridin-1-yl	H	Quinolin-8-sulfonyloxy	6.367
16	Aziridin-1-yl	H	Thiophene-2-sulfonyloxy	6.328
17	Aziridin-1-yl	Aziridin-1-yl	$C_6H_5SO_3$	5.237
18	Aziridin-1-yl	Aziridin-1-yl	CH_3-4-$C_6H_4SO_3$	5.337
19	Aziridin-1-yl	Aziridin-1-yl	C_2H_5-4-$C_6H_4SO_3$	5.398
20#	Aziridin-1-yl	Aziridin-1-yl	$(CH_3)_3C$-4-$C_6H_4SO_3$	5.337
21	Aziridin-1-yl	Aziridin-1-yl	Cl-4-$C_6H_4SO_3$	5.495
22	Aziridin-1-yl	Aziridin-1-yl	Br-4-$C_6H_4SO_3$	5.432
23	Aziridin-1-yl	Aziridin-1-yl	NO_2-4-$C_6H_4SO_3$	5.119
24	Aziridin-1-yl	Aziridin-1-yl	$C_6H_3(NO_2)_2SO_3$	5.620
25	Aziridin-1-yl	Aziridin-1-yl	$C_6H_3(CH_3)_3SO_3$	5.409
26#	Aziridin-1-yl	Aziridin-1-yl	$C_6H_5CH_2SO_3$	5.620
27	Aziridin-1-yl	Aziridin-1-yl	$C_6H_5CH=CHSO_3$	5.237
28	Aziridin-1-yl	Aziridin-1-yl	Naphtalene-1-sulfonyloxy	5.469
29	Aziridin-1-yl	Aziridin-1-yl	CH_3SO_3	5.495
30	Aziridin-1-yl	Aziridin-1-yl	$ClCH_2CH_2CH2SO_3$	5.523
31#	Aziridin-1-yl	H	$C_6H_5CO_2$	6.347
32	Aziridin-1-yl	H	CH_3-4-$C_6H_4CO_2$	6.208
33	Aziridin-1-yl	H	F-4-$C_6H_4CO_2$	6.310
34	Aziridin-1-yl	H	Cl-4-$C_6H_4CO_2$	6.276
35	Aziridin-1-yl	H	CH_3O-3, 4,6-$C_6H_2CO_2$	6.796
36	Aziridin-1-yl	H	Furan-2-carbonloxy	6.310
37#	Aziridin-1-yl	H	Thiophene-2-cabonyloxy	6.328

#test set of molecules.

to determine the external predictivity of the resulting QSAR model and these were removed from the original data set. The chosen test set evenly spanned the anti-malarial activity range as well as the chemical structural diversity of the database. The remaining data set, known as training set, is used to develop QSAR model.

All the geometries of selected molecules have been minimized using the DFT method[17–21] with the help of Becke's three parameter hybrid density functional, B3LYP/6-31G (d, p), which include both Hartree-Fock exchange and DFT exchange correlation functional using Gaussin03 program[22]. The optimized geometries are characterized by harmonic-vibrational frequencies, which confirmed that the structures obtained are minimum on the potential energy surface. Various global and local reactivity descriptors are calculated from the Gaussian03 output file by customized software developed by our group.

The nucleus independent chemical shifts (NICS) were calculated with optimized geometries of the molecules by GIAO method as implemented in Gaussian03. The group frontier electron densities (F_A^E, F_A^N, F_B^E, F_B^N) have been calculated using eq. (3) and eq. (4) by summing the frontier electron densities of the constituent atoms of the aromatic rings A and B.

Derivation and validation of the model:

QSAR models were derived from the training set by multiple linear regression (MLR) using observed anti-malarial activities as the dependent variables and various combinations of the chosen descriptors as the independent variables. The quality of the model was considered as statistically satisfactory on the basis of number of data points (n), square of correlation coefficient (r^2), standard error estimate (SEE), population (p), F-statistics (F) and T-statistics (T).

A large F value indicates that the QSAR model is quite reliable and it is not a chance occurrence. The T-test measures the statistical significance of the regression coefficients. The higher T-test values correspond to the relatively more significant regression coefficients.

The models obtained were validated by calculating the cross-validated squared correlation coefficient (q^2), which are calculated from "leave-one-out" (LOO) test[23,24]. Many authors[25,26] consider that higher q^2 value (> 0.5) as an indicator of highly predictive QSAR model.

In order to evaluate the external predictive potential of the QSAR model derived from the training set, the model was used to predict the biological activities of the external test set of seven molecules according to the procedure given in Roy *et al.*[27].

Results and discussion

To ascertain the relationship between chemical structures of selected naphthoquinonyl sulfonate and acylate derivatives and their pIC_{50} values against malaria, we have generated various equations through different combinations of DFT based local and global reactivity descriptors. It was kept in mind that for the best QSAR model the number of descriptors should be as small as possible and should have maximum correlation coefficient for the measured activities. In the present case, the best model was obtained by using three descriptors – (i) energy of highest occupied molecular orbital (HOMO), (ii) nucleus independent chemical shift (NICS) at aromatic ring B and (iii) electrophilic group frontier electron density at the aromatic ring A (see Table 1 head). The model having the highest correlation coefficient is:

$$pIC_{50} = -4.749 - 0.0397\ HOMO - 0.626\ NICS(1) - 0.00411\ F_A^E \quad (5)$$

with $n = 30$, $r^2 = 0.929$, $q^2 = 0.900$, $P = 0.000$, $F = 113.19$ and SEE = 0.178

Other relevant statistical parameters have been listed in Table 2. The pearson correlation matrix (Table 3) shows that the descriptors are independent. The predicted pIC_{50} values of the training set and test set from the QSAR model is given in Table 4 along with actual measured activity values. A graph

Table 2. Uncertainties, T-test and P-values of the QSAR model

Variables	Uncertainties	T-test values	P-values
Constant	0.90	–5.83	0.000
HOMO	0.0030	–12.72	0.000
NICS(1)	0.082	–9.18	0.000
F_A^E	0.00041	–9.61	0.000

Table 3. Pearson correlation matrix

	pIC_{50}	HOMO	NICS(1)	F_A^E
pIC_{50}	1.000			
HOMO	–0.680	1.000		
NICS(1)	–0.460	0.004	1.000	
F_A^E	–0.508	0.032	-0.036	1.000

Table 4. Observed and predicted values according to the QSAR model

Molecule No.	Observed activity	Predicted activity	Residual	Molecule No.	Observed activity	Predicted activity	Residual
1	5.620	5.818	–0.198	20#	5.337	5.309	0.028
2	5.824	6.113	–0.289	21	5.495	5.392	0.103
3	5.432	5.260	0.172	22	5.432	5.392	0.040
4	5.367	5.499	–0.132	23	5.119	5.422	–0.303
5	6.638	6.731	–0.093	24	5.620	5.395	0.225
6	6.699	6.732	–0.033	25	5.409	5.557	–0.148
7	7.018	6.727	0.291	26#	5.620	5.446	0.174
8#	6.796	6.742	0.054	27	5.237	5.235	0.002
9	6.854	6.778	0.076	28	5.469	5.370	0.099
10	6.481	6.661	–0.180	29	5.495	5.405	0.090
11	6.770	6.723	0.047	30	5.523	5.539	–0.016
12#	6.553	6.690	–0.137	31#	6.347	6.217	0.130
13	6.886	6.803	0.083	32	6.208	6.163	0.045
14	6.886	6.924	–0.038	33	6.310	6.248	0.062
15#	6.367	6.324	0.043	34	6.276	6.294	–0.018
16	6.328	6.690	–0.362	35	6.796	6.670	0.126
17	5.237	5.577	–0.340	36	6.310	6.299	0.011
18	5.337	5.619	–0.282	37#	6.328	6.421	–0.093
19	5.398	5.307	0.091				

#test set of molecules.

of actual activity versus predicted of the training set and test set has been provided in Fig. 2.

The predictive r^2_{pred} value using the test set of molecules has been found to be 0.920. This indicates that the QSAR model given by eq. (5) has high predictive ability.

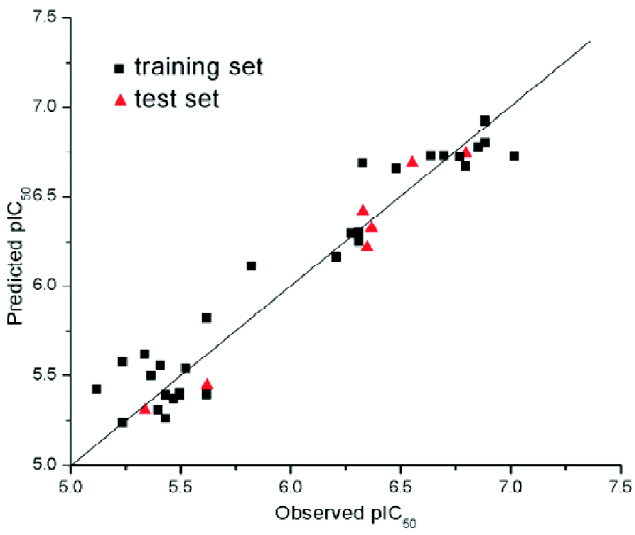


Fig. 2. The plot of observed versus predicted pIC_{50} values following QSAR eq. (5).

From eq. (5), it is obvious that NICS(1) of ring B is the most important determining factor of the anti-malarial activity. The NICS(1) values being itself negative and the coefficient of the NICS(1) term being the largest in eq. (5), a higher magnitude of NICS(1) creates a positive contribution to the pIC_{50} values and thus is likely to be responsible for the anti-malarial activity of the present set of molecules. Also, electron releasing groups at R_3 is likely to increase the NICS(1) value of ring B. From Table 1, one can see that all sulfonyloxy group bearing molecules show comparatively higher activities. Sulfonyloxy group being electron releasing group increases the electron density on ring B and thus a higher value of the NICS(1) is induced by this group.

The group frontier electron density of ring A is the least important among the descriptors as the coefficient multiplying it in eq. (5) is hundred times smaller than the other coefficients. Group frontier electron density itself being positive, a high value of it tends to decrease the pIC_{50}. The electron density of ring A can be increased by substituting the elec-

tron releasing groups at R_1 and R_2.

The HOMO and LUMO of the molecules are mainly located on rings A and B (Fig. 3) and partly on R_1 and R_2. HOMO energy being negative, a high value of HOMO energy will make a positive contribution to the activity.

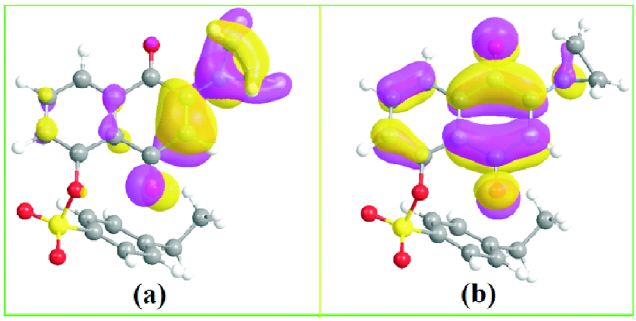


Fig. 3. (a) HOMO and (b) LUMO of molecule 7.

Conclusion

In the present paper, the effectiveness of naphthoquinoly based bis-aziridinyl and sulfonyl molecules as potential anti-malarial active molecules has been studied. QSAR studies based on DFT optimized structures of the molecules reveal that NICS(1) on ring B, electrophilic group frontier electron density and HOMO energy of the molecules are appropriate descriptors. In summary, the present QSAR analysis nicely explain the observed anti-malarial activity of the napthoquinolin based aziridinyl and sulfonyl molecules.

Acknowledgement

Atish Dipankar Jana acknowledges the financial support from the Department of Science, Technology and Biotechnology, Government of West Bengal, India through the Project No. ST/P/S & T/16G-47/2017.

References

1. WHO Report, 2012, http://www.who.int.
2. L. M. Ursos and P. D. Roepe, *Med. Res. Rev.,* 2002, **22(5)**, 465.
3. J. F. Trape, *Am. J. Trop. Med. Hyg.,* 2001,**64(1)**, 12.
4. A. Ghosh, M. J. Edwards and M. Jacobs-Lorena, *Parasitol Today*, 2000, **16(5)**, 196.
5. W. Campbell and R. S. Rew, "Chemotherapy of parasitic diseases", Plenum, New York.
6. P. A. Winstanley, *Parasitol Today*, 2000, **16(4)**, 146.
7. A. F. David, J. Philip, S. L. Rosenthal, R. B. Croft and N. Solomon, *Nat. Rev. Drug Discov.*, 2004, **3(6)**, 509.
8. B. Mekonnen, E. Weiss, E. Katz, J. Ziffer, H. Ma and D. E. Kyle, *Bioorg. Med. Chem.*, 2000, **8(5)**, 1111.
9. G. H. Posner, J. P. Maxwell, H. O'Dowd, M. Krasavin, S. Xie and T. A. Shapiro, *Bioorg. Med. Chem.*, 2000, **8(6)**, 1361.
10. P. M. O'Neill, A. Miller, L. P. D. Bishop, S. Hindley, J. L. Maggs, S. W. Ward, S. M. Roberts, F. Scheinmann, A. V. Stachulski, G. H. Posner and B. K. Park, *J. Med. Chem.*, 2001, **44(1)**, 58.
11. M. Calas, M. L. Ancelin, G. Cordina, P. Portefaix, G. Piquet, V. Vidal-Sailhan and H. Vial, *J. Med. Chem.*, 2000, **43(3)**, 505.
12. T. S. Lin, L. Y. Zhu, S. P. Xu, A. A. Divo and A. C. Sartorelli, *J. Med. Chem.*, 1991, **34(5)**, 1634.
13. A. Sarkar, T. R. Middya and A. D. Jana, *J. Mol. Model.*, 2012, **18**, 2621.
14. R. SchleyerPv, C. Maerker, A. Dransfeld, H. Jiao and N. J. HommesRv E, *J. Am. Chem. Soc.*, 1996, **118(26)**, 6317.
15. A. Sarkar and G. Mostafa, *J. Mol. Model.*, 2009, **15**, 1221.
16. M. Karelson, V. S. Lobanov and A. R. Katritzky, *Chem. Rev*, 1996, **96(3)**, 1027.
17. A. D. Becke, *Phys. Rev. A*, 1988, **8(63)**, 3098.
18. A. D. Becke, *J. Chem. Phys.*, 1993, **98(2)**, 372.
19. A. D. Becke, *J. Chem. Phys.*, 1993, **98(7)**, 5648.
20. S. Hirata, C. G. Zhan, E. Apra, T. L. Windus and D. A. Dixon, *J. Phys. Chem. A,* 2003, **107(47)**, 10154.
21. M. Sulpizi, G. Folkers, U. Rothlisberger, P. Carloni and L. Scapozza,*Quant. Struct. Act. Relat.*, 2002, **21(2)**, 173.
22. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. V. Tomasi Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Sefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. A-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andes, C. Gonzalez, M. Head-Gordon, E. S. Replogle and J. A. Pople, Gaussian 03 - Revision B.03, *Gaussian* Inc, Pittsburgh PA, 2003.
23. A. Golbraik, M. Shen, Z. Xiao, Y. D. Xiao and K. H. Lee, *J. Comput. Aid. Mol. Des.*, 2003, **17(2-4)**, 241.
24. D. M. Hawkins, S. C. Basak and D. Mills, *J. Chem. Inf. Comput. Sci.*, 2003, **43(2)**, 579.
25. B. F. Thomas, D. R. Compton, B. R. Martin and S. F. Semus, *Mol. Pharmacol.*, 1991, **40(5)**, 656.
26. A. Agarwal, P. P. Pearson, E. W. Taylor, H. B. Li, T. Dahlgren, M. Herslof, Y. Yang, G. Lambert, D. L. Nelson, J. W. Regan and A. R. Martin, *J. Med. Chem.,* 1993, **36(25)**, 4006.
27. P. P. Roy and K. Roy, *QSAR Comb. Sci.*, 2008, **27(3)**, 302.