



## QSPR Models for Predicting Lipophilicity of Triazole Derivatives

Jain Shubha, Awasthi Ashish Kumar\* and Piplode Satish

School of studies in Chemistry & Biochemistry, Vikram University, Ujjain-456010, INDIA

Available online at: [www.isca.in](http://www.isca.in), [www.isca.me](http://www.isca.me)

Received 8<sup>th</sup> May 2014, revised 13<sup>th</sup> June 2014, accepted 17<sup>th</sup> July 2014

### Abstract

The present study consists modeling the lipophilicity ( $\log P$ ) of 5-(2-Oxo-3-aryl-diazenyl-4-methyl-2H-chromen-8-yl)-3-thio-1,2,4-triazole derivatives using distance-based topological indices & indicator parameter. The MLR analysis reveals that O-atom, ZM1, HI and IDE are the best suitable parameters for modeling the lipophilicity of the compounds have been taken in present study. The obtained models are critically discussed on the basis of cross validation technique.

**Keywords:** QSPR, topological indices, lipophilicity, regression analysis, cross validation, triazole derivatives.

### Introduction

Quantitative Structure - activity relationship and Quantitative Structure - property relationship (QSAR/ QSPR) studies are useful tools in the rational search for bioactive molecules. QSPR models are mathematical equations which attempt to correlate chemical structure to a wide variety of physical, chemical and biological properties<sup>1</sup>. (QSPR/QSAR) represents an attempt to relate structural descriptors of molecules with their physicochemical properties and biological activities<sup>2</sup>. Today, QSARs are being applied in many disciplines with much emphasis on drug design<sup>3</sup>. The lipophilicity expressed by the logarithmic partition coefficient ( $\log P$ ) is a very important physicochemical parameter which explains a partitioning equilibrium of solute molecules between water and an immiscible partitioning solvent<sup>4-6</sup>. Lipophilicity is a physicochemical property of principal importance in drug discovery and development<sup>7</sup>. The aim of the present study is to use topological indices for predicting lipophilicity ( $\log P$ ) of a series of 5-(2-Oxo-3-aryl-diazenyl-4-methyl-2H-chromen-8-yl)-3-thio-1,2,4-triazoles. The use of topological indices is an important stage in QSPR studies. In the present work indicator parameter and topological indices used in the modeling of lipophilicity are O-atom, ZM1 (First Zagreb index M1)<sup>8</sup>, HI (Harary index)<sup>9</sup> and IDE (mean information content on the distance equality information indices)<sup>10</sup>. The aim of this study is to develop a QSPR model to correlate the structural features of this class of compounds with their lipophilicity ( $\log P$ ) using topological indices. In the present QSPR study, the topological indices and structural indicator are used as structural descriptors for 21 derivatives of 5-(2-Oxo-3-aryl-diazenyl-4-methyl-2H-chromen-8-yl)-3-thio-1,2,4-triazoles<sup>11</sup> for modeling of lipophilicity.

### Material and Methods

**ClogP:** The value of ClogP is calculated using chemdraw ultra 11 for the set of 21 derivatives of triazole.

**Topological Indices:** The topological indices ZM1, HI and IDE are used for modeling the lipophilicity. All the topological indices employed in the present study were calculated using hydrogen suppressed graph<sup>12-14</sup> of the compounds used.

**Indicator Parameter:** One indicator parameter has been used to understand the impact of electronegative atom on the lipophilicity of the compounds. Indicator parameter O-atom accounts for the number of oxygen atom in the molecule.

**Statistical Analysis:** For the modeling of lipophilicity, we have used maximum  $R^2$  method<sup>15</sup> in forward direction and finally obtained statistically significant models. The regression analysis was performed by SPSS software.

**Cross validation:** Cross-validation parameters which have been estimated are given in table-3 and are described below. Cross-validation correlation coefficient  $R^2_{cv}$ , indicates the performance of the model which is defined as:

$$R^2_{cv} = 1 - \frac{PRESS}{SSY}$$

**PRESS** (predicted residual errors sum of squares) is the sum of squared difference between actual (calculated) and the predicted when the compound is omitted from the fitting process.

$$PRESS = \sum (Y_{cal} - Y_{pre})^2$$

### Uncertainty of Prediction

$$S_{PRESS} = \sqrt{\frac{PRESS}{(N - k - 1)}}$$

The lower value of  $S_{press}$  indicates better model.

### Predictive Square Error

$$PSE = \sqrt{\frac{PRESS}{N}}$$

The lower value of PSE indicates better model.

$$\text{Quality factor: } Q = \frac{R}{\text{SEE}}$$

Higher value of quality factor (Q) indicates better predictivity of model.

#### Sum of square of response values (SSY)

$$\text{SSY} = \sum (Y_{\text{cal}} - Y_{\text{mean}})^2$$

'SSY' suggests the overall predictive performance.

## Results and Discussion

The topological indices ZM1, HI and IDE along with structural indicator (O-atom) of 5-(2-Oxo-3-aryl-diazenyl-4-methyl-2H-chromen-8-yl)-3-thio-1,2,4-triazole derivatives are given in table-1. The table also records the ClogP and the position of substituents (R) on the compound.

The topological indices represent molecular structure in a numerical form. These are obtained by transforming molecular structure into its molecular graph *via* mathematical expression. Such transformation of molecular structure in to its graph is carried out by deleting all the carbon-hydrogen as well as heteroatom hydrogen bonds in the molecular structure. In Chemical Graph Theory, molecular structures are normally represented as hydrogen-suppressed graphs, whose vertex and edges act as atoms and covalent bonds, respectively<sup>16</sup>. For modeling of physico-chemical properties, biological activity and toxicity of organic compounds such topological indices have been used<sup>17,18</sup>. The correlation between the lipophilicity, topological indices and indicator is shown in table -2.

Table 2 shows that a single topological index does not correlate with Clog P to yield statistically significant mono-parametric model is possible for modeling the lipophilicity (ClogP). Thus, it can be concluded that stepwise multivariate regression analysis is required to obtain the statistically significant model. The indices ZM1, HI, IDE and the indicator parameter, are the better parameters for modeling lipophilicity among all the topological indices used by us. We have used the recommendations made by Randic<sup>19</sup> to justify the occurrence of highly correlated parameters in the proposed models. The results show that a bi-parametric regression model containing O-atom, ZM1 gave the best results. This model is given as below:

$$\text{Clog P} = -4.503 (\pm 1.736) - 0.722 (\pm 0.070) * \text{O-atom} + 0.071 (\pm 0.013) * \text{ZM1} \quad (1)$$

$$R = 0.941, R^2 = 0.885, R^2_{\text{adj}} = 0.872, \text{SE} = 0.1465, F = 69.243, K = 2$$

Here, R is the correlation coefficient, R<sup>2</sup> is the squared correlation coefficient, R<sup>2</sup><sub>adj</sub> is the adjustable R<sup>2</sup>, SE is the standard error of estimation, F is the F-statics, K is the number of topological invariants used in the regression and the figures

within the parenthesis are the standard error values of the coefficient.

We found a tri parametric regression expression with improved statistics when HI parameter was added during the stepwise regression analysis. This model is given as below:

$$\text{Clog P} = -2.961 (\pm 1.613) - 0.665 (\pm 0.065) * \text{O-atom} + 0.063 (\pm 0.011) * \text{ZM1} - 0.008 (\pm 0.003) * \text{HI} \quad (2)$$

$$R = 0.958, R^2 = 0.918, R^2_{\text{adj}} = 0.904, \text{SE} = 0.1269, F = 63.807, K = 3$$

Looking into such an excellent result further regression analysis was not needed. But, with a hope of better results, we have carried out many tetra-parametric regression analysis. When IDE is added to equation 2, great improvement was observed in the statistics, the resulted tetra- parametric model is given below:

$$\text{Clog P} = -4.338 (\pm 1.839) - 0.649 (\pm 0.064) * \text{O-atom} + 0.047 (\pm 0.016) * \text{ZM1} - 0.010 (\pm 0.003) * \text{HI} - 1.073 (\pm 0.751) * \text{IDE} \quad (3)$$

$$R = 0.963, R^2 = 0.927, R^2_{\text{adj}} = 0.910, \text{SE} = 0.1232, F = 51.295, K = 4$$

The parameters contributing to model 3 have both, positive as well as negative contribution in the modeling of lipophilicity. The initial statistics SE, R, R<sup>2</sup><sub>adj</sub> and F indicate that the model as described by eqn.3 is superior than the other proposed models (equation 1 and 2)

In all the models discussed above, we observed that the positive sign associated with ZM1 indicate their positive role towards lipophilicity and negative sign associated with O-atom, HI and IDE indicate their negative role towards lipophilicity.

The value of pogliani Q parameter<sup>20,21</sup> for the model expressed by equation 3 (Q=7.8165) confirms that this model has excellent predictive power. By the use of cross validation parameters like PRESS, SSY, R<sup>2</sup><sub>cv</sub>, PSE and S<sub>PRESS</sub> from which we can test the predictive power of the proposed model<sup>22</sup>. The calculated cross-validation parameters<sup>15</sup> for each of the models are discussed below.

For the model 3, the value of Q is Q=7.8165, which is greater than other proposed model expressed by equations 1&2 respectively.

If the values of PRESS and S<sub>PRESS</sub> are smaller then the model predicts better and can be considered statistically significant<sup>23</sup>. In this regard, the model 3 is the best one. PSE is another cross-validation parameter. The lowest value of PSE for the model 3 supports its highest predictive potential. For a model, PRESS/SSY should be not more than 0.4<sup>24</sup>. In case of our study the ratio PRESS/SSY ranges between 0.0724 - 0.1199.

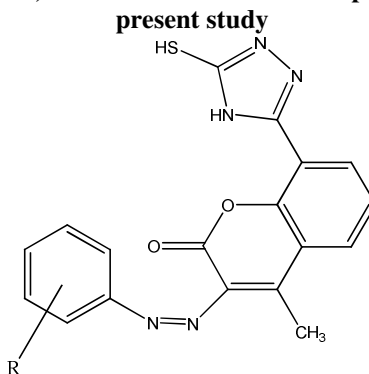
We have predicted the lipophilicity from models expressed by equation (2) and equation (3) which are discussed above. The

predicted lipophilicities are then compared with their calculated values. This comparison is given in table 4. The residue of lipophilicity is the least for the model expressed by equation (3), showing that model 3 is the most appropriate model for modeling the lipophilicity.

We have estimated the predictive correlation coefficients ( $R^2_{pred.}$ ) to examine the relative potential of the proposed models

by plotting graphs between calculated and predicted lipophilicity values using equations (2) and (3). The values  $R^2_{pred.}$  are found as 0.960 and 0.986, respectively, for the models expressed by equations (2) and (3) and shown in figure 1 and 2 respectively. This finally confirms that the model expressed by equation (3) has the best predictive potential.

**Table-1**  
Structural details, calculated lipophilicity value, structural indicator and topological indices for the compounds used in the present study



Compound	ClogP	R	O-atom	ZM1	HI	IDE
1	3.747	H	2	140	95.74	3.49
2	4.460	2-Cl	2	146	53.51	3.52
3	4.460	3-Cl	2	146	53.12	3.54
4	4.460	4-Cl	2	146	52.84	3.48
5	3.853	2-OH	3	146	53.51	3.52
6	3.853	3-OH	3	146	53.12	3.54
7	3.853	4-OH	3	146	52.84	3.48
8	4.031	3-OCH <sub>3</sub>	3	150	58.04	3.61
9	4.031	4-OCH <sub>3</sub>	3	150	57.54	3.61
10	3.490	2-NO <sub>2</sub>	4	156	64.41	3.59
11	3.490	3-NO <sub>2</sub>	4	156	63.46	3.65
12	3.853	4-NO <sub>2</sub>	4	156	62.56	3.71
13	3.745	2-COOH	4	156	64.41	3.59
14	3.745	3-COOH	4	156	63.46	3.65
15	3.745	4-COOH	4	156	62.56	3.71
16	4.246	3-CH <sub>3</sub>	2	146	53.12	3.54
17	4.246	4-CH <sub>3</sub>	2	146	52.84	3.48
18	4.775	3-C <sub>2</sub> H <sub>5</sub>	2	150	58.04	3.61
19	4.775	4-C <sub>2</sub> H <sub>5</sub>	2	150	57.54	3.61
20	4.610	3-Br	2	146	53.12	3.54
21	4.610	4-Br	2	146	52.84	3.48

**Table-2**  
Intercorrelation matrix of structural descriptor for proposed model

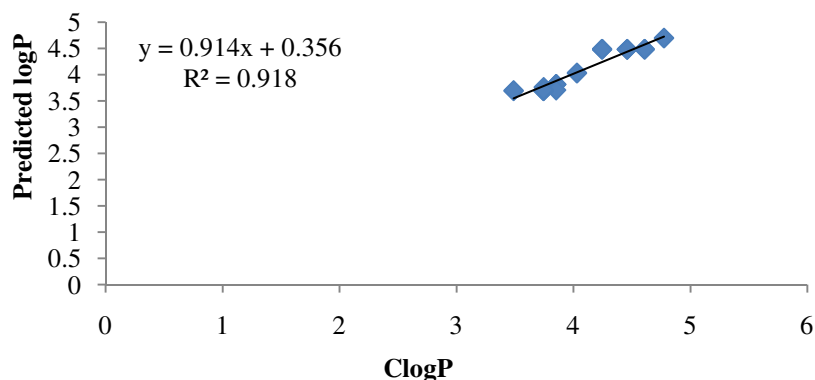
	Clog P	O-atom	ZM1	HI	IDE
Clog P	1				
O-atom	-0.828	1			
ZM1	-0.461	0.846	1		
HI	-0.459	0.204	0.033	1	
IDE	-0.337	0.695	0.849	0.162	1

**Table-3**  
 Values of cross-validation parameters for the proposed models

Model	Parameters Used	PRESS	S <sub>PRESS</sub>	PSE	Q	R <sup>2</sup> <sub>cv</sub>	SSY	PRESS/SSY
1	O-atom, ZM1	0.4030	0.1496	0.1385	6.4232	0.8801	3.3606	0.1199
2	O-atom, ZM1, HI	0.2755	0.1273	0.1145	7.5492	0.9181	3.3606	0.0819
3	O-atom, ZM1, HI, IDE	0.2433	0.1233	0.1076	7.8165	0.9276	3.3606	0.0724

**Table-4**  
 ClogP and Predicted logP values of 5-(2-Oxo-3-aryl-diazenyl-4-methyl-2H-chromen-8-yl)-3-thio-1,2,4-triazole analogues derived from the regression equation 2 and 3

Compound	ClogP	Predicted logP		Residual	
		Eq.2	Eq.3	Eq.2	Eq.3
1	3.7470	3.7630	3.7313	-0.0160	0.0156
2	4.4600	4.4789	4.4678	-0.0189	-0.0078
3	4.4600	4.4820	4.4932	-0.0220	-0.0332
4	4.4600	4.4842	4.4316	-0.0242	0.0283
5	3.8538	3.8139	3.8188	0.0398	0.0349
6	3.8538	3.8170	3.8442	0.0367	0.0095
7	3.8538	3.8192	3.7826	0.0345	0.0711
8	4.0310	4.0296	4.0581	0.0013	-0.0271
9	4.0310	4.0336	4.0631	-0.0026	-0.0321
10	3.4900	3.6917	3.6059	-0.2017	-0.1159
11	3.4900	3.6993	3.6798	-0.2093	-0.1898
12	3.8538	3.7065	3.7532	0.1472	0.1005
13	3.7455	3.6917	3.6059	0.0537	0.1395
14	3.7455	3.6993	3.6798	0.0461	0.0656
15	3.7455	3.7065	3.7532	0.0389	-0.0077
16	4.2460	4.4820	4.4932	-0.2360	-0.2472
17	4.2460	4.4842	4.4316	-0.2382	-0.1856
18	4.7750	4.6946	4.7071	0.0803	0.0678
19	4.7750	4.6986	4.7121	0.0763	0.0628
20	4.6100	4.4820	4.4932	0.1279	0.1167
21	4.6100	4.4842	4.4316	0.1257	0.1783



**Figure-1**  
 Correlation between calculated and predicted lipophilicity of 21 derivatives of 5-(2-Oxo-3-aryl-diazenyl-4-methyl-2H-chromen-8-yl)-3-thio-1,2,4-triazole using equation 2

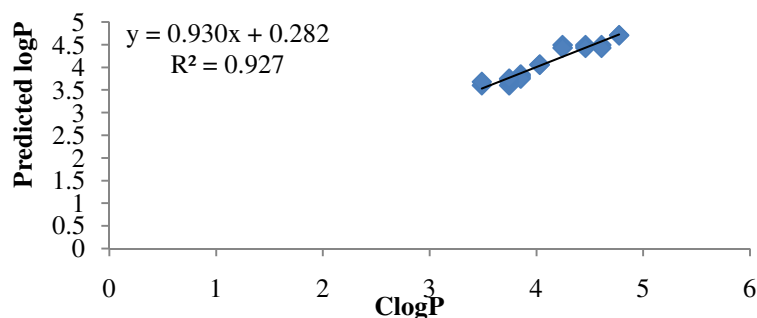


Figure-2

Correlation between calculated and predicted lipophilicity of 21 derivatives of 5-(2-Oxo-3-aryl-diazenyl-4-methyl-2H-chromen-8-yl)-3-thio-1, 2, 4-triazole using equation 3

## Conclusion

The lipophilicity of Triazole derivatives can be modeled using topological indices along with indicator parameter. The model constituted by the ZM1, HI and IDE as molecular descriptors and O-atom as Indicator parameter is the best model having best ability to predict the lipophilicity expressed as ClogP of the triazole. The use of structural indicators, based on the number of electronegative atoms, gave better results with topological indices and thus elaborated the role of electronegative atoms in the modeling of lipophilicity. From the results, as discussed above, it is concluded that the model obtained by combination of topological indices and structural indicators have better quality and predictivity.

## References

1. Minu M., Thangadurai A., S Wakode. R., AgrawaS. S. I and Narasimhan B., Synthesis, antimicrobial activity and QSAR studies of new 2,3-disubstituted-3,3a,4,5,6,7-hexahydro-2H-indazoles, *Bioorg. Med. Chem. Lett.*, **19** (11), 2960-2964 (2009)
2. Kumar A., Narasimhanb B. and D. Kumar, Synthesis, Antimicrobial, and QSAR Studies of Substituted Benzamides, *Bioorg. Med. Chem.*, **15**(12), 4113-4124 (2007)
3. Ghasemi G., Arshadi S., Rashtehroodi A. N., Nirouei M., Shariati S. and Rastgoo Z., QSAR investigation on quinolizidinyl derivatives in Alzheimer's disease, *Journal of Computational Medicine*, **2013**, 3-18 (2013)
4. Raevsky O. A., Schaper K. J. and Seydel J. K., H-Bond Contribution to Octanol-Water Partition Coefficients of Polar Compounds, *Quant. Struct.-Act. Relat.*, **14**(5), 433-436, (1995)
5. Schaper K.J., Zhange H. and Raevsky O.A., pH-Dependent Partitioning of Acidic and Basic Drugs into Liposomes—A Quantitative StructureActivity Relationship Analysis, *Quant. Struct.-Act. Relat.*, **20** (1), 46-54 (2001)
6. Khadikar P.V., Singh S. and Shrivastava A., Novel estimation of lipophilic behaviour of polychlorinated biphenyls, *Bioorg. Med. Chem. Lett.*, **12**(7), 1125-1128 (2002)
7. Rutkowska E., Pajak K. and Jozwiak K., Lipophilicity-methods of determination and its role in medicinal chemistry, *Acta Pol. Pharm -Drug Research*, **70**(1), 3-18 (2013)
8. Gutman I., Ruscic B., Trinajstic N., and Wilcox Jr. C.F., Graph theory and molecular orbitals. XII. Acyclic Polyenes, *J. Chem. Phys.*, **62**(9), 3399-3405 (1975)
9. Diudea M.V., Indices of reciprocal properties or Harary indices, *J. Chem. Inf. Comput. Sci.*, **37**(2), 292-299 (1997)
10. Bonchev D., Information Theoretic indices for Characterization of Chemical Structures, Rsp-Wiley, Chichester, (1983)
11. Upadhyay S., Synthetic and Electrochemical studies on some biologically significant hetrocycles: Aziridines and Triazoles, Ph.D. Thesis, D.A.V.V. Indore (2007)
12. Todeschini R., Cosonni V., Handbook of Molecular Descriptors, Wiley-VCH: Weinheim, (2000)
13. N Trinajstic., Chemical Graph Theory, CRC Press: Boca Raton, Florida, (1992)
14. Karelson M., Molecular Descriptors in QSAR/QSPR, John Wiley & Sons, New York (2000)
15. Chatterjee S., Hadi A.S., Price B., Regression Analysis by Examples, Wiley, New York, 3<sup>rd</sup> Ed (2000)

16. Nunez M.B., Maguna F.P., Okulik N.B. and Castro E.A., QSAR modeling of the MAO inhibitory activity of xanthenes derivatives, *Bioorg. Med. Chem. Lett.*, **14**, 5611–5617 (2004)
17. Karcher I.N. and Devillers J., Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology, Kluwer Academic, Dordrecht, **1990**
18. Diudea M.V., QSPR/QSAR Studies by Molecular Descriptors, Babes-Bolyai University, Cluj, Romania, **2000**
19. Randic M., Comparative structure-property studies: Regressions using a single descriptor, *Croat. Chem. Acta*, **66**, 289-312 (1993)
20. Pogliani L., Structure property relationships of amino acids and some dipeptides, *Amino Acids*, **6**, 141-153 (1994)
21. Pogliani L., Modeling with Special Descriptors Derived from a Medium-Sized Set of Connectivity Indices, *J. Phys. Chem.*, **100**, 18065-18077 (1996)
22. Srivastava A.K., A Pandey., Nath A., S Chaurasia., QSAR based modeling of inhibitory activity of alkenyldiarylmethane derivatives, *J. Saudi Chem. Soc.*, **13** (3), 263–267 (2009)
23. Bhagwat V.W., Khadikar P.V., Tiwari A., Solanki A., Choubey A., Neel Kamal, Manana P. and Lowlekar V., QSPR Evidences for Topological indices to mimic lipophilicity of Thia and Aza-Crown Ethers, *Asian J. Chem.*, **21** (7), 5212-5220 (2009)
24. Thakur A., Thakur M., Kakani N., Joshi A., Thakur S. and Gupta A., “Application of topological and physicochemical descriptors: QSAR study of phenylamino-acridine derivatives”, *Arkivoc*, **xiv**, 36-43 (2004)