Available online at www.derpharmachemica.com

ISSN 0975-413X
CODEN (USA): PCHHAX

Der Pharma Chemica, 2016, 8(10):57-62
(http://derpharmachemica.com/archive.html)

2D and 3D QSAR Analysis of some novel 2-{[2-(*1H*-imidazol-1-yl)ethyl]sulfanyl}-*1H*-benzimidazole derivatives

Mayura Kale*, Rajesh Nawale, Gajanan Sonawane and Ratnamala Survase

Department of Pharmaceutical Chemistry, Government College of Pharmacy, Aurangabad-431005, Maharashtra, India

---

ABSTRACT

QSAR model development was carried out for of 2-{[2-(1H-imidazol-1-yl)ethyl]sulfanyl}-1H-benzimidazole derivatives that had been evaluated for antiprotozoal activity activity. The physicochemical parameters were calculated using VLIFE MDS 4.5 software. Stepwise multiple linear regression analysis was applied to derive QSAR models, which were further evaluated for statistical significance and predictive power by internal and external validation. The best quantitative structure activity relationship model was selected having a correlation coefficient (r^2) of 0.9740, cross-validated correlation coefficient (q^2) of 0.9588 and, r^2 pred of 0.7691. The predictive ability of the selected model was also confirmed by leave one-out cross-validation. The QSAR model indicates that the descriptors (chiV1,chi3Cluster, XAHydrophobicArea) highly influence antiprotozoal activity. The information derived from the present study may be useful in the design of more potent substituted 2-{[2-(1H-imidazol-1-yl)ethyl]sulfanyl}-1H-benzimidazole derivatives.

Key words: QSAR, benzothiazole, multiple linear regression, partial least square, antiprotozoal activity

---

INTRODUCTION

Parasitic infections caused by protozoa still[1] represent a major public health problem in the developing countries. Some important intestinal protozoa include *Giardia intestinalis* and *Entamoeba histolytica* which are the causal agents of giardiasis and amoebiasis; respectively. According to the World Health Organization (WHO)[2], there are estimated 280 million giardiasis cases, 50 million amoebiasis cases each year and and has been classified as one of the most common causes of death from parasitic diseases. In addition to the common symptoms such as diarrhea and dysentery, this protozoan can penetrate the intestinal mucosa and migrate to other organs causing severe damage. In addition to intestinal infections, the genitourinary infection[3] caused by *T. vaginalis* (trichomoniasis) is estimated to be more than 180 million new cases annually. For these three diseases, metronidazole (MTZ) has been successfully used as the drug of choice for more than 40 years; however, its side effects and the development of resistant strains limits its use. Although some additional chemotherapeutic agents are available (e.g. tinidazole and nitazoxanide used in the treatment of giardiasis), it is still important to have more options of treatment, because of different individual response to drugs. During the last years, an important number of benzimidazole derivatives have been synthesized and evaluated as antiprotozoals and studies based on the emerging concept of the activity landscape have been undertaken to find out the structure–activity relationships (SAR) of benzimidazole derivatives as antiprotozoal agents.

Pérez-Villanueva et al. have reported the synthesis of 19 new 2-{[2-(1H-imidazol-1- yl)ethyl]sulfanyl}-1H-benzimidazole derivatives[4], in which they found that the introduction of a 2-{[2-(1H-imidazol-1-yl)ethyl]sulfanyl}-

1H-benzimidazole moiety led to compounds displaying high activity and selectivity. Pursuing these research consequences, we have undertaken QSAR study on these previously reported findings with the aim to identify the molecular properties which influence the antiprotozoal activity the most.

MATERIALS AND METHODS

A total of twenty 2-{[2-(1H-imidazol-1-yl)ethyl]sulfanyl}-1H-benzimidazole derivatives that are reported as antiprotozoal were used as data set in QSAR analysis [5] (Table 1). The IC_{50} (μM) were converted to negative logarithmic values to get pIC_{50} for QSAR study. Molecules were divided into the training set (16 molecules) and test set (5 molecules) by spear exclusion method. All the work was performed by drawing the structure of the molecule in 2D Draw application in Tool menu of QSAR Plus of Molecular Design Suit (MDS) software. Then 2D structures where exported to QSAR Plus window (2D structure converted to 3D structure). After the conversion, structure force field and energy minimization was done with the help of MMF [6] which resulted in optimization and optimized molecules were employed to calculate the physicochemical and alignment descriptors. For model development in 2D-QSAR analysis, three methods Random selection method, Manuel data selection method, Sphere Exclusion method were used and training / test set were created with 10 trials run in each case. After the creation of training and test set, minimum and maximum value of the test and training set was checked, using the QSAR tool and then the different statistical methods like Multiple liner regression (MLR), Partial least squares regression (PLSR) were used for model building.

Sphere Exclusion method: In this method dissimilarity value was obtained which provided an idea to handle training and test set size. This was adjusted by trial and error until a desired division of training and test set was achieved [7]. The increase in dissimilarity value results in increase in number of molecules in the test set.

Partial least square regression (PLSR): PLSR was used for model generation which is an expansion of the multiple linear regression (MLR). PLSR[8] is probably the least restrictive of the various multivariate extensions of the multiple linear regression models. PLSR was used as an exploratory analysis tool to select suitable predictor variables and to identify outliers before classical linear regression. All the calculated descriptors were considered as independent variable and biological activity as dependent variables.

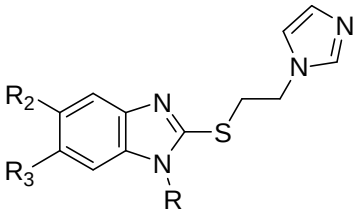


Table 1. Chemical and biological data of 2-{[2-(1H-imidazol-1-yl)ethyl]sulfanyl}-1H-benzimidazole derivatives

Compound no.	R_1	R_2	R_3	pIC_{50}
1	-H	-H	-H	6.7495799976911055
2	-H	-Cl	-H	6.768530409569319
3	-H	-Cl	-Cl	6.8523236757589014
4	$-CH_3$	-H	-H	6.826813731587726
5	$-CH_3$	-H	-Cl	6.892450870255313
6	$-CH_3$	-Cl	-H	6.869988050328096
7	$-CH_3$	-Cl	-Cl	7.016373712875465
8	-H	$-COOCH_3$	-H	6.955460239607589
9	-H	$-COOCH_3$	-Cl	7.138465589140962
10	$-CH_3$	-H	$-COOCH_3$	6.943857737940948
11	$-CH_3$	-Cl	$-COOCH_3$	7.060980223551333
12	$-CH_3$	$-COOCH_3$	-H	7.111259039317107
13	$-CH_3$	$-COOCH_3$	-Cl	7.033389013318065
14	-H	$-OCH_2CH_3$	-H	7.1444808443322
15	-H	$-OCH_2CH_3$	-Cl	7.003926345514724
16	$-CH_3$	-H	$-OCH_2CH_3$	7.1444808443322
17	$-CH_3$	-Cl	$-OCH_2CH_3$	7.156144577376839
18	$-CH_3$	$-OCH_2CH_3$	-H	7.118615343229427
19	$-CH_3$	$-OCH_2CH_3$	-Cl	7.008773924307505
Metronidazole				6.627087997029894
Albendazole				5.798466326556618

RESULTS AND DISCUSSION

The QSAR models (Table 2), depict N as the number of data points used in the model. Pred_ r^2 is the predicted r^2 for external test set and SE is the standard error of estimate (smaller is better). From this table, the equation of the model-1 explains 97.4% (r^2=0.9740) of the total variance in the training set as well as it has internal (q^2= 0.9588) and external (pred_ r^2= 0.7691) with the predictive ability of 95% and 76%; respectively. Model-02 explains 93.46% (r^2 = 0.9346) of the total variance in the training set as well as it has internal (q^2=0.8991) and external (pred_ r^2= 0.5771) predictive ability of 89% and 57%; respectively.

The graphs were plotted between the actual and the predicted biological activities for Model 1 and Model 2 with their r^2 values as shown in Figure 2 and Figure 3; respectively.

Table 2. Predictive QSAR models with equation generated from regression methods

<table>
<tr><th>Model no.</th><th>Equation</th></tr>
<tr><td>1. Multiple linear Regression (MLR)[9]</td><td>Multiple Regression
Training Set Size = 16, Test Set Size = 5
Selected Descriptors: chiV1, chi3Cluster, XAHydrophobicArea
Coefficient: 0.5054(±0.0163), -0.5118(±0.0961), -0.0034(±0.0000)
Constant: 4.8114
Statistics:
<table>
<tr><td>N</td><td>16</td></tr>
<tr><td>Degree_of_freedom</td><td>12</td></tr>
<tr><td>r2</td><td>0.9740</td></tr>
<tr><td>q2</td><td>0.9588</td></tr>
<tr><td>F_test</td><td>149.6514</td></tr>
<tr><td>r2_se</td><td>0.0594</td></tr>
<tr><td>q2_se</td><td>0.0747</td></tr>
<tr><td>pred_r2</td><td>0.7691</td></tr>
<tr><td>pred_r2se</td><td>0.0933</td></tr>
</table></td></tr>
<tr><td>2. Principle Component Regression (PCR)[10]</td><td>Training Set Size = 16, Test Set Size = 5
Selected Descriptors: chiV1
Coefficient:0.2769
Constant: 4.8685
Statistics:
<table>
<tr><td>Optimum Components</td><td>1</td></tr>
<tr><td>N</td><td>16</td></tr>
<tr><td>Degree_of_freedom</td><td>14</td></tr>
<tr><td>r2</td><td>0.9346</td></tr>
<tr><td>q2</td><td>0.8991</td></tr>
<tr><td>F_test</td><td>199.9467</td></tr>
<tr><td>r2_se</td><td>0.0872</td></tr>
<tr><td>q2_se</td><td>0.1083</td></tr>
<tr><td>pred_r2</td><td>0.5771</td></tr>
<tr><td>pred_r2se</td><td>0.1262</td></tr>
</table></td></tr>
</table>

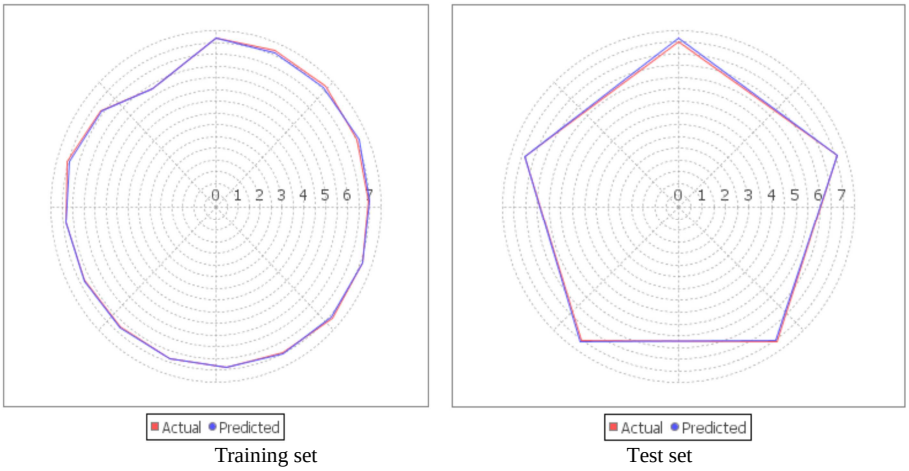
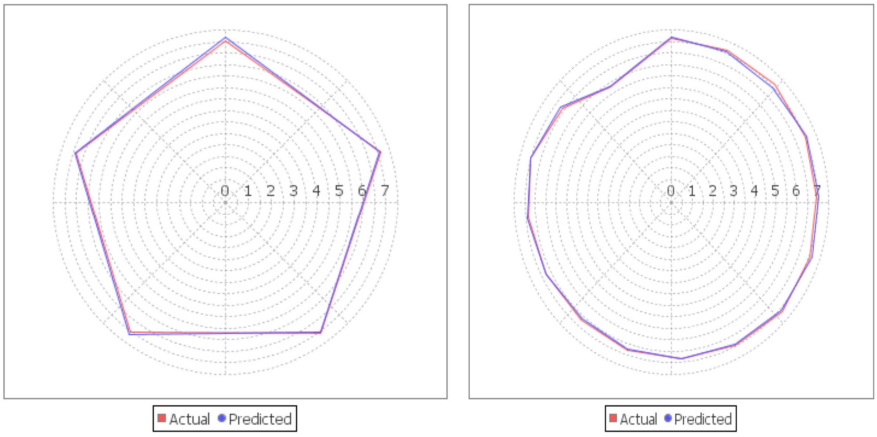


Figure 2. Graph between actual and predicted biological activity for training and test set of Model-01



Training set **Test set**

Figure 3. Graph between actual and predicted biological activity for training and test set of Model-02

The Figures represent models which are having $r^2 > 0.93$, indicating good predictive ability in predicting the activity of the test set molecules.

The correlation matrix is used to see the mutual correlation among the parameters used in the model. This matrix shows that descriptors have low inter-correlation value. The contribution charts for Model 1 and Model 2 are shown in Figure 4 and Figure 5; respectively.

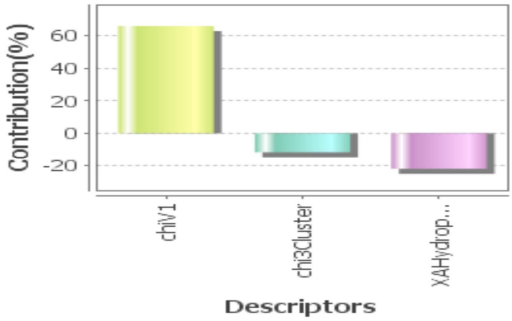


Figure 4. Contribution chart of various descriptors in biological activity for Model-01

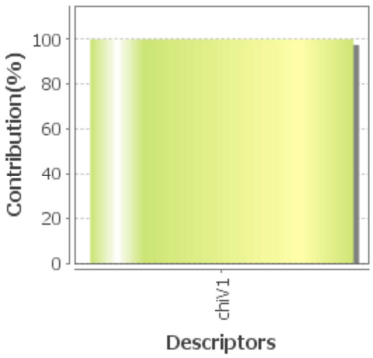
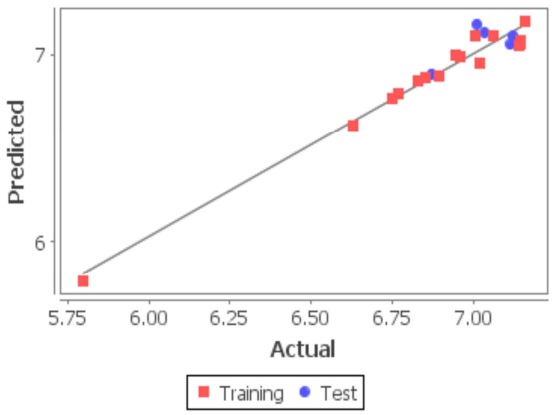


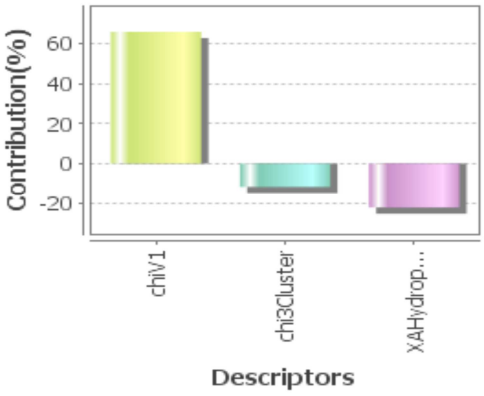
Figure 5. Contribution chart of various descriptors in biological activity for Model-02

In the present study, MLR (coupled with stepwise forward variable selection); led to the development of a statistically significant model. The developed Model 1 reveals that descriptors chiV, chi3 cluster, play an important role (≈62,) in determining antiprotozoal activity. The other descriptor i.e. chi3cluster - index[11] is inversely proportional to the biological activity (≈-12%) & third descriptor XAHydrophobicArea[12] is also inversely proportional to activity (≈-22). In Model 2 the, descriptor chiV1 influences antiprotozoal activity the most (≈100%).



Fitness plot

Figure 6



Contribution plot

Figure 7

Table 3. Actual and predicted activity for Training set and test set

Sr.no	Compound Name	Actual value	Predicted value
1	35	6.7495799976911055	6.765543
2	36	6.768530409569319	6.786989
3	37	6.8523236757589014	6.876633
4	38	6.826813731587726	6.86102
5	39	6.892450870255313	6.888315
6	40	6.869988050328096	6.896134*
7	41	7.016373712875465	6.952883
8	42	6.955460239607589	6.987804
9	43	7.138465589140962	7.046264
10	44	6.943857737940948	6.994669
11	45	7.060980223551333	7.101201
12	46	7.111259039317107	7.051818*
13	47	7.033389013318065	7.11577*
14	48	7.1444808443322	7.057057
15	49	7.003926345514724	7.09817
16	50	7.1444808443322	7.073742
17	51	7.156144577376839	7.178927
18	52	7.118615343229427	7.098996*
19	53	7.008773924307505	7.161729*
20	54	6.627087997029894	6.616283
21	55	5.798466326556618	5.793847

*indicates compounds are in the test set for the corresponding model and rest are in the training set.

CONCLUSION

A quantitative structure activity relationship study was performed on a series of 2-{[2-(1H-imidazol-1-yl)ethyl]sulfanyl}-1H-benzimidazole derivative possessing antiprotozoal activity. It was done to establish quantitative relationship between biological activity and their physicochemical /structural properties. Two dimensional quantitative structure activity relationship (2D QSAR) study by means of multiple regression (MR) method was performed on a series of 2-{[2-(1H-imidazol-1-yl)ethyl]sulfanyl}-1H-benzimidazole derivatives possessing antiprotozoal activity using molecular design suite (VLifeMDS 4.5). This study was performed with 21 compounds (data set) using sphere exclusion (SE) algorithm, random and manual selection methods for the division of the data set into training and test set. MR methodology with stepwise (SW) forward variable selection method was used for building the QSAR models. Statistically significant QSAR models were generated. Among them most significant model has squared correlation coefficient (r^2), cross validated correlation coefficient (q^2) and predictive correlation coefficient (pred_r^2) 0.9740, 0.9588and 0.7691; respectively. The second model generated by using partial least squre method (PLS) Among them most significant model has squared correlation coefficient (r^2), cross validated correlation coefficient (q^2) and predictive correlation coefficient (pred_r^2) 0.9346, 0.8991 and 0.5771 respectivaly The QSAR model indicates that the descriptors chiV1 chi3 ClusterXA Hydrophobic Area contributed 50%, 51%, and 13 %; respectively and in second model chiV1 contributed 100%; respectively to biological activity. The positive coefficient value of chiV1 on the biological activity indicated that higher value leads to better antiprotozoal activity whereas lower value leads todecrease activity. Negative coefficient value of chi3Cluster XA

Hydrophobic Area indicates that lower value leads to better antiprotozoal activity and vice-versa is also true. In present study, an attempt has been made to identify the necessary structural and substituent requirements that influence the biological activity. From the present QSAR analysis, two best models were generated among which any one can be used for predicting the activity of the newly designed compounds for finding some more potent molecules. Finally, it is concluded that the work presented here will play an important role in understanding the relationship of physiochemical parameters with structure and biological activity. By studying the QSAR model, one can select the suitable substituent and design new compounds with improved biological activity.

Acknowledgement
The authors are thankful to the Principal, Government College of Pharmacy, Aurangabad for providing necessary facilities for this research work.

REFERENCES

[1] S. Montaner, A. Galiano, M. Trelis, L. Martin-Jaular, H. del Portillo, D. Bernal, A. Marcilla, *Front. Immunol.*, **2014**, 5, 433.
[2]P. E. Petersen, *Int. Dent. J.*, **2008**, 58 (3), 115–121.
[3] M.J. Matos, M.T. Bacelar, P. Pinto, I. Ramos, *Eur. J. of Radiology*. **2005**, 5, 181–187.
[4]J. Pérez-Villanueva, A. Hernández-Campos, L. Yépez-Mulia,C. Méndez-Cuesta, O. Méndez-Lucio, F. Hernández-Luis, R. Castillo, *Bioorg. Med. Chem. Lett.*, **2013**, 23 (14), 4221–4224.
[5] B. Testa; QSAR: Hansch Analysis and Related Approaches; 16, VCH Publishers, New York, **1995**.
[6] T. Halgren, *J. Comput. Chem.*, **1996**, 17, 490–519.
[7] A. Gobbi, M. Lee, *J. Chem. Inf. Comput. Sci.*, **2003**, 43 (1), 317–323.
[8]L. Carrascal, I. Galván, O. Gordo, *Oikos*, **2009**, 118 (5), 681–690.
[9]J. He, A. Zelikovsky, *Bioinformatics*, **2006**, 22 (20), 2558–2561.
[10] B. Mevik, R. Wehrens, *J. Stat. Softw.*, **2007**, 18 (2), 1–24.
[11] Y. Fang, J. Xie, G. Dai, M. Wang, F. Zhu, T. Xu, E. Wong, E, 3D Deep Shape Descriptor, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **2015,** 7 (12), 2319–2328.
[12] K. Singh, M. Sharma, S. Jain, S. Kad, D. Raghuvanshi, *J. Pharm. Res.*, **2012**, 55 (88), 4127–4132.