IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 12, December 2019

QSAR Analysis of Novel Dicationic 2-Phenylbenzofurans as Potent Anti-Leishmanial Agents

Ankita Agarwal[1]

Assistant Professor, Department of Chemistry, Keral Verma Subharti College of Science,

Swami Vivekanand Subharti University, Meerut, India[1]

Abstract: The QSAR study was conducted on 41, dicationic 2-Phenylbenzofurans derivatives with *L. donovani* DNA minor groove inhibitory activity using Multiple Linear Regression (MLR) and Partial Least Squares (PLS) methods. The statistical values from both the techniques were analyzed and compared to establish the good predictability of the models obtained. The MLR and PLS generated comparable models with good predictive ability and all other statistical values, r, r^2, r^2_{cv}, r^2 (test set) and F and S values, were 0.874, 0.765, 0.716, 0.718 and 22.804, 0.284, respectively, for MLR and r^2, r^2_{cv}, r^2 (test set) and statistical significance value were 0.758, 0.704, 0.710 and 0.99, respectively, for PLS, were satisfactory. The results obtained from this study indicate that the electronic descriptors play an important role in determining the anti-leishmanial activity of the compounds.

Keywords: TSAR, MLR, PLS, *L. Donovani* DNA Minor Groove

I. INTRODUCTION

Visceral Leishmaniasis (VL) is caused by the protozoan parasite *Leishmania donovani* [1] and transmitted by the bite of around 30 species of phlebotamine sandflies [2-5]. Research over the past decade has identified a number of drugs and formulations that offer improved treatment for this disease [6, 7]. The drugs for leishmaniasis's treatment are sodium stibogluconate (pentostam) and meglumine antimonate (glucantime), but they exhibit renal and cardiac toxicity [8]. Alternative drugs, such as pentamidine, amphotericin B, and some azo-derivatives are also very toxic and exhibits serious side effects [9]. Miltefosine, a phosphocholine analogue is the oral agent effective against both cutaneous and visceral leishmaniasis [10] but presents severe gastrointestinal problems [11]. Pentavalent antimony, the most widely prescribed drug to treat leishmaniasis patients, has serious side effects, requires a prolonged course of treatment and is losing its efficacy in some regions due to increasing parasite resistance. Although treatment for leishmanianis exist, they are not optimal due to problems of toxicity, high price or difficulty in administration [12].

Given the problems of toxicity, need for hospitalization, growing resistance, and high costs associated with the currently available drugs for leishmaniasis, it is clear that patients urgently need new, improved, efficient, and safe drugs treatments to replace or complement these drugs [13]. So it will be beneficial to optimize existing anti-leishmanial agents using QSAR modeling techniques to identify the important molecular properties required for the effective inhibition of parasite. In line to above discussion, we felt that there is need to revaluate the binding requirements of anti-leishmanials by employing computational approach.

One of the most promising technique to set insight into the structural requirements is QSAR, which is a mathematical relationship linking chemical structure and pharmacological activity in a quantitative manner for a series of compounds. QSAR is a prominent tool to explore the relationship between the structures of ligands and their binding affinities. QSAR methodologies save resources and expedite the process of the development of new molecules and drugs. QSAR increases the probability of success and reduce the time and cost involved in the drug discovery process [14].

So it will be beneficial to optimize existing anti-leishmanial agents using QSAR modeling techniques to identify the important molecular properties required for the effective inhibition of parasite. Likewise, molecular similarity has been extensively used in drug design, e.g., in the selection of analogs for substitution, in the estimation of molecular properties, in the rational selection of candidates from large databases, and in some QSAR approaches [15]. The similarity between a pair of molecules is estimated on the bases of overlap of the analogous fields of the two molecules, taken as a sum over all components on a three-dimensional (3D) grid, using various 3D-molecular similarity indices [16, 17]. The electron density, electric field, electrostatic potential, molecular lipophilicity potential, molecular fields, shape, etc., have been used for similarity assessment [18].

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 12, December 2019

II. MATERIALS AND METHODS

Structures of dicationic 2-Phenylbenzofurans derivatives along with their biological activities [19] were taken for present studies in view of high structural diversity and sufficient variation in biological activities.

2.1. Generation and three-dimensional optimization of chemical structures

The molecular structures were drawn and their geometries were cleaned using standalone module of Discovery Studio (Version 2.0) software and were subjected to energy minimization. All the structures were loaded to the worksheet of TSAR (Version 3.3, Accelrys Inc., Oxford, England) and were labeled accordingly. Further, we introduced a chemical encoding scheme according to which each molecule was described as a template with a defined number of substituents attached to this template by a single bond. A single hydrogen atom may also serve as substituents. Certainly, there exist several ways to represent molecules as template with different substituents. All the substituents were numbered according to their positions in molecules (Table 1.). All the structures and their defined substituents were converted into high quality 3D structures using Corina-make 3D option, which includes total energy, valence terms (i.e. bond, bond angle, and torsional potential), and nonbonded terms (electrostatic and vander-waals interactions) [20]. Charges were calculated using charge-2 package available with TSAR.

2.2. Dataset preparation, descriptors calculation and similarity indices

The IC_{50} values of all the 41 compounds used in the present study were converted into negative logarithm of IC_{50} (i.e. $pIC_{50}=\log 1/IC_{50}$). The data set was randomly divided into the training and test set of 25 and 12 compounds, respectively. Four compounds were identified as outliers and were deleted. The training set compounds were used to develop the QSAR model while the test set compounds were used to validate the developed model. The numerical descriptors are responsible for encoding important features of the molecules. In the present study, similarity based descriptors were calculated for each compound in the training set, using the TSAR software.

The concept of bio-isosterism was used as the basis for calculating similarity indices. Similarity indices represent a quantitative measure of the similarity between two molecules on the basis of their size, shape, electronic distribution, lipid solubility, water solubility or chemical reactivity [21]. The molecular similarity indices were computed using Hodgkin index with the ASP similarity program in TSAR software. Among the two approaches (a grid-based method & Gaussian approximation), Gaussian approximation was used for calculating similarity indices because it closely mirrors that of the grid-based calculations but is much faster. Gaussian approximation based N X N similarity matrix were constructed and subjected to data reduction techniques [22]. Correlations shall be derived mainly by relating binding affinity to the similarity data obtained from comparison to a single compound usually that with the highest.

2.3. Descriptor reduction, model development and validation

Since the large pool of similarity descriptors were calculated, there is a significant requirement of data reduction to eliminate the chance correlation. To select the suitable similarity based descriptors for MLR and PLS analysis, Pearson's correlation matrix (pair wise correlation analysis) was constructed upon the larger number of descriptor pool. One of two descriptors with inter-correlation (correlation with two consecutive descriptors) coefficient >0.5, was discarded [23]. After data reduction, three similarity descriptors namely combined similarity vs. molecule (38), charge similarity vs. molecule (47), charge similarity vs. molecule (49) were retrived which exhibited high correlation with the biological activity and also did not had any correlation among each other.

The Multiple Linear Regression (MLR) was carried out to derive best QSAR models. Various MLR models were generated using biological activity data as dependent variable and selected descriptors as the independent variables. These models were used to quantify the relationship between dependent and independent variables. Statistical significance of the regression equations were tested on the basis of conventional regression coefficient (r^2), Fischer's ratio (F), and the standard error of estimate (S).

Both internal and external validation techniques were applied for the assessment of model robustness and its predictive power. Internal validation was performed by applying cross-validation analysis using leave-one-out (LOO) method in which one compound is removed from the training set. Further, the predictive power of the model was assessed by estimating the activity of external test set of compounds not included in the original model. Partial least square (PLS) analysis has been recommended as an alternative approach to enlarge the information content in each model and avoid danger of over fitting [24]. As an approach to check the robustness and the predictive ability of the models generated using multiple linear regression (MLR) analysis, partial least square (PLS) analysis was performed on the same training set of compounds. Similar to the cross-validation method used in MLR, model generated during PLS was also validated using leave out one row [25].

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 12, December 2019

III. RESULTS AND DISSCUSSION

For all the compounds mentioned in table 1, the shape, electronic, refractivity, lipophilicity and combined similarity matrices were computed. MLR analysis was performed with these similarity descriptors derived from the similarity matrix. Initially, MLR analysis performed on training set compounds with 220 generated similarity descriptors showed very poor internal predictive ability. In the next phase, data reduction was performed first by pair wise correlation method and then using backward elimination on the basis of low T-value. Finally a model was developed ($r = 0.682$, $r^2 = 0.465$, $r^2_{cv} = 0.377$) when 3 similarity descriptors were retrieved as shown in equation 1.

$$Y = 1.356 \times X1 - 0.735 \times X2 - 23.464 \times X3 + 21.351 \quad \text{- (Equation 1)}$$

Owing to unsatisfactory values for r^2 and r^2_{cv}, outliers were detected. Standard statistical values such as residual values were used to find out the possible outliers. Four molecules namely 23, 24, 28 and 39 were found away from the regression line because of their high residual value and therefore were deleted as outliers. The final regression equation obtained from MLR analysis (final 25 molecules in training set) after deleting the outliers is represented as equation 2

$$Y = 2.653 \times X1 - 1.083 \times X2 - 23.844 \times X3 + 20.971 \quad \text{- (Equation 2)}$$

$r = 0.874$, $r^2 = 0.765$, $r^2_{cv} = 0.716$, $F = 22.804$, $S = 0.284$, r^2 (test set) $= 0.718$

To further confirm the soundness and predictive ability of the model, PLS analysis was performed using the same data set. For a well-defined problem, both MLR and PLS should generate comparable results. The results of the PLS as shown in equation 3 also were evaluated on the basis of r^2_{cv} and statistical significance of the model.

$$Y = 2.579 \times X1 - 0.987 \times X2 - 26.342 \times X3 + 23.465 \quad \text{- (Equation 3)}$$

Where, Y = Predictive biological activity, X1 = combined similarity vs. molecule (38), X2 = charge similarity vs. molecule (47) and X3 = charge similarity vs. molecule (49)

Statistical significance = 0.99, $r^2_{cv} = 0.704$, $r^2 = 0.758$, r^2 (test set) $= 0.710$

Since for a well defined problem, both MLR and PLS should generate comparable results the r^2_{cv} values of MLR and the PLS models were evaluated and it was found that both the models have comparable r^2_{cv} value of 0.716 and 0.704 for MLR and PLS respectively. The predictive ability of the model was also validated using the external test set of 12 compounds in context of minimum difference between the actual and predicted biological activity values of MLR and PLS analysis for training and test which is shown in table 1 and their respective plots are depicted in figure 1 and 2.

3.1. Interpretation of entered similarity descriptors

Three similarity descriptors namely combined similarity vs. molecule (38)**,** charge similarity vs. molecule (47) and charge similarity vs. molecule (49) were retrieved.

It may be expected that, in certain cases, the overall similarity will produce the similar activity, (combined similarity), whereas in other cases, only the similarity of certain active regions of the molecules will give rise to similar activities. The basic idea underlying on similarity-based QSAR approaches was "*molecules that are structurally similar likely will have similar properties*". Thus, when the activity of a given molecule is unknown, we can predict it by taking into account similarity values between the molecule under study and the molecules of a data set whose activities are known.

As the similarity data using MLR analysis reports the importance of charge of the molecule. The charge similarity can be explained in terms of the electrostatic force between ligand and receptor that helps to define the affinity of the interaction. Electrostatic interactions are long range meaning that electric fields can be sensed several angstroms away from the point charge. The strength (and effective distance) of these interactions is a function of the dielectric property of the environment. Water molecules are able to shield locale charges and dipoles reducing the range of their electric field forces.

Combined similarity of the molecule 38 and Charge similarity of compound 47 and 49 respectively, indicates that similar activity and charge properties of these molecules are important in imparting biological activity.

IV. CONCLUSION

Similarity based multivariate analysis of 2-Phenylbenzofurans derivatives was successfully carried out to build a statistically significant model possessing a good correlative and predictive capability for *L. donovani* DNA minor groove inhibitory activity. The goal of this study was to develop a model for prediction of anti-leishmanial activity of 2-Phenylbenzofurans derivatives. According to the developed model presented in the current work, similarity based parameters encoding the combined similarity index of the entire compounds vs compound 38 positively contributes towards activity which means that the combined charge, shape, lipophilicity and refractivity indices are favorable for the inhibitory activity. Our study reveals that charge similarity index of the entire compounds vs compound 47 and 49 negatively contributes towards activity which means that the high electronegative character similar to 47 and 49 will have negative effect on biological activity.

ACKNOWLEDGMENT

We gratefully acknowledge support for this research from Rameshwardasji Birla Smarak Kosh, Medical research center, Mumbai and Indian Council of Medical Research (ICMR), New Delhi.

REFERENCES

[1]. S. K. Bhattacharya, D. Sur and K. Juntra, "Childhood visceral leishmaniasis", Indian Journal of Medical Research, 123 (2006): 353-356.
[2]. R. S. Barnetson, R. S. Ridley and H. W. Wheate, "A form of muco-cutaneous leishmaniasis in the Old World", Transactions of the Royal Society of Tropical Medicine and Hygiene, 72 (1970): 516-8.
[3]. P. A. Bates, "Transmission of Leishmania metacyclic promastigotes by phlebotomine sand flies", International Journal for Parasitology, 37 (2007): 1097-1106.
[4]. P. Desjeux, "Leishmaniasis. Public health aspects and control", Clinics in Dermatology, 14 (1996): 417-423.
[5]. B. Veress, E. R. Abdalla and M. A. Hassan, "Vascular lesions in mucosal leishmaniasis", Morphol Igazsagugyi Orv Sz, 17(1977): 226-30.
[6]. H. W. Murray, "Treatment of visceral leishmaniasis (kala-azar): a decade of progress and future approaches", International Journal of Infectious Diseases, 4 (2004): 158-77.
[7]. S. L. Croft and V. Yardley, "Chemotherapy of leishmaniasis", Current Pharmaceutical Design, 8 (2002): 319-42.
[8]. S. Raht, A. Trivellin and T. R. "Imbrunito, Antimoniais Empregados no Tratamento da Leishmaniose: Estado da Arte", Química Nova, 26 (2003): 550-557.
[9]. A. Mcgregor, "Who Warms of Epidemic Leishmania", Lancet, 351 (1998): 575-575.
[10]. J. Soto, B. A. Arana & J. Toledo, "Miltefosine for New World Cutaneous Leishmaniasis", Clinical Infectious Diseases , 38 (2004): 1266-1272.
[11]. H. Sangraula, K. K. Sharma, S. Rijal, S. Dwivedi and S. Koirala, "Orally Effective Drugs for Kala-azar (Visceral Leishmaniasis): Focus on Miltefosine and Sitamaquine", Journal of the Association of Physicians of India, 51 (2003): 686-690.
[12]. N. Singh, "Drug resistance mechanisms in clinical isolates of Leishmania donovani", Indian Journal of Medical Research, 123 (2006): 411-422.
[13]. P. B. Carvalho, M. A. G. Arribas and E. I. Ferreira, "Leishmaniasis. What do We Know About its Chemotherapy?", Brazilian Journal of Pharmaceutical Sciences, 36 (2000): 69-96.
[14]. M. M. Neaz, M. Muddassar, F. A. Pasha and J. C. Seung, "2D-QSAR of non benzodiazepines to benzodiazepines receptor", Medicinal Chemistry Research, 18 (2008): 98-111.
[15]. H. Eckert and J. Bajorath, "Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches", *Drug Discovery Today*, 12 (2007): 225-33.
[16]. H. Kubinyi, "Molecular similarity. 1. Chemical structure and biological action", Pharmazie in unserer Zeit, 27 (1998): 92-106.
[17]. H. Kubinyi, "Molecular similarity. 2. The structural basis of drug design", Pharmazie in unserer Zeit., 27 (1998): 158-72.
[18]. E. E. Hodgkin, W. G. Richards, "Molecular similarity based on electrostatic potential and electric field", International journal of quantum chemistry, Quantum biology symposium, 14 (1987): 105-110.
[19]. S.A.Bakunov, et al, "Synthesis & antiprotozoal activity of cationic 2phenylbenzofurans", Journal of Medicinal Chemistry, 51(2008):6927–6944
[20]. N. Smith and M. Spittle, "ABC of AIDS. Tumours", British Medical Journal, 294 (1987): 1274-7.
[21]. V. Burt, G. W. Richards, "Molecular similarity: The introduction of flexible fitting", Journal of Computer-Aided Molecular Design, 4 (1990): 231-238.
[22]. H. Kubinyi, F. A. Hamprecht and T. Mietzner, "Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices", Journal of Medicinal Chemistry, 41 (1998): 2553-64.
[23]. A. Agarwal, R. Mishra and S. Paliwal, "Chemistry of Phytopotentials: Health, Energy and Environmental Perspectives", Springer-Verlag, Berlin, DOI:10.1007/978–3-642–23394-4_30 (2012).
[24]. R. Mishra, A. Agarwal and S. Paliwal, "Chemistry of Phytopotentials: Health, Energy and Environmental Perspectives", Springer-Verlag, Berlin, DOI:10.1007/978–3–642–23394–4_29 (2012).
[25]. A. Agarwal, S. Paliwal and R. Mishra, et al, "Discovery of a selective, safe and novel anti-malarial compound with activity against chloroquine resistant strain of *Plasmodium falciparum*", Scientific Report, 5 (2015): 13838.

IARJSET

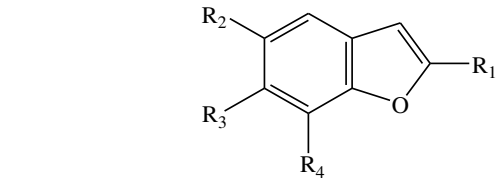
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

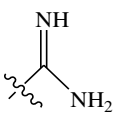
International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 12, December 2019

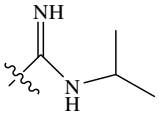
APPENDICES

Table 1. Series of 2-Phenylbenzofurans derivatives along with their actual & predicted biological values

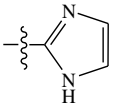




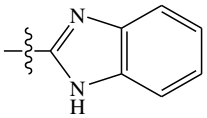
Am



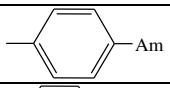
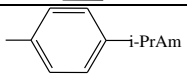
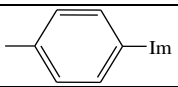
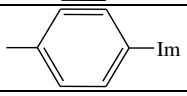
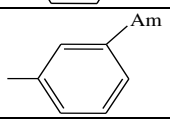
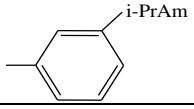
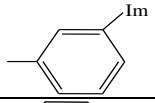
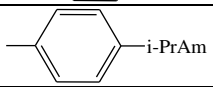
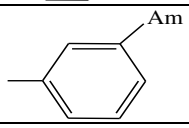
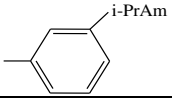
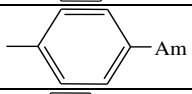
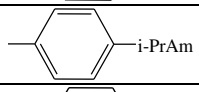
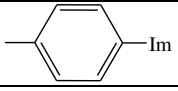
i-PrAm



Im



BzIM

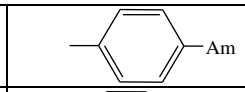
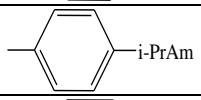
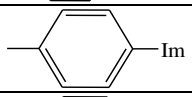
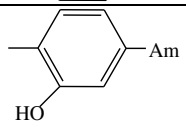
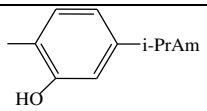
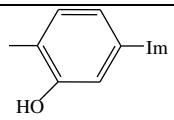
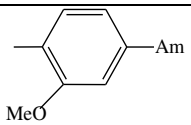
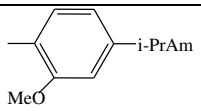
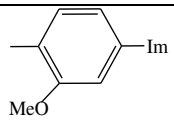
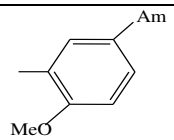
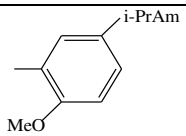
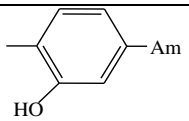
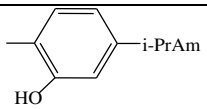
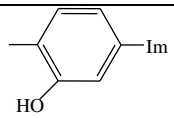
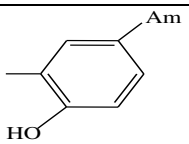
Comp. Name	Subst. R_1	Subst. R_2	Subst. R_3	Subst. R_4	Actual Value (Log 1/c	Predicted Value	
						MLR	PLS
1		Am	H	H	0.004	0.054	0.163
2		i-PrAm	H	H	-0.204	-0.329	-0.279
3*		Im	H	H	-1.322	-0.821	-0.782
6*		BzIM	H	H	-0.755	-0.449	-0.455
8*		Am	H	H	-1.763	-1.243	-1.227
9		i-PrAm	H	H	-1.113	-1.364	-1.366
10		Im	H	H	-1.380	-1.492	-1.503
11		H	Am	H	-0.491	-0.816	-0.805
12*		H	i-PrAm	H	-1.079	-0.997	-1.002
14		H	Am	H	-1.146	-1.350	-1.290
15		H	i-PrAm	H	-1.380	-1.571	-1.533
17		Am	H	OMe	-0.477	-0.428	-0.446
18*		i-PrAm	H	OMe	-1.176	-0.701	-0.748
19		Im	H	OMe	-0.653	-0.671	-0.727

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 12, December 2019

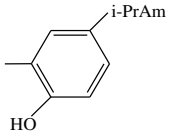
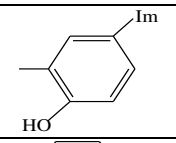
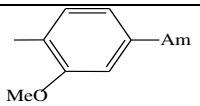
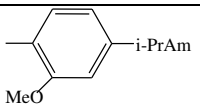
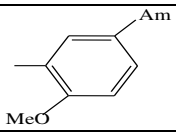
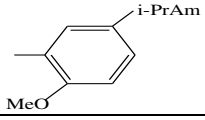
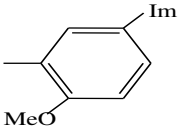
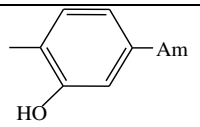
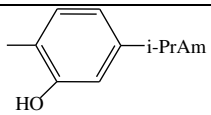
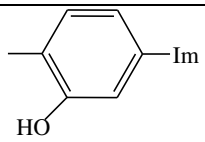
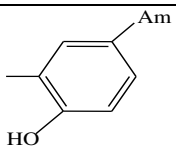
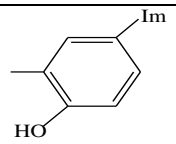
20*		Am	H	OH	-0.724	-0.352	-0.368
21*		i-PrAm	H	OH	-1.301	-0.737	-0.780
22		Im	H	OH	-0.869	-0.718	-0.770
23**		Am	H	OH	-1.154	--	--
24**		i-PrAm	H	OH	-1.677	--	--
25*		Im	H	OH	-1.204	-0.827	-0.868
26		Am	H	H	-0.301	-0.278	-0.295
27		i-PrAm	H	H	-0.176	-0.498	-0.551
28**		Im	H	H	-1.795	--	--
29*		Am	H	H	-0.755	-0.744	-0.730
30		i-PrAm	H	H	-1.397	-1.087	-1.114
32		Am	H	H	-0.255	-0.510	-0.526
33		i-PrAm	H	H	-0.255	-0.483	-0.538
34		Im	H	H	-0.763	-0.855	-0.923
35		Am	H	H	-1.301	-0.931	-0.927

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 12, December 2019

36*		i-PrAm	H	H	-1.602	-1.183	-1.203
37		Im	H	H	-1.755	-1.563	-1.568
38		H	Am	H	-0.568	-0.388	-0.377
39**		H	i-PrAm	H	-1.508	--	--
41		H	Am	H	-0.556	-0.836	-0.806
42*		H	i-PrAm	H	-1.732	-1.223	-1.206
43		H	Im	H	-1.531	-1.065	-1.073
44*		H	Am	H	-1.462	-0.759	-0.734
45		H	i-PrAm	H	-1.591	-0.915	-0.915
46		H	Im	H	-1.113	-1.134	-1.156
47		H	Am	H	-1.612	-1.400	-1.329
49		H	Im	H	-1.544	-1.797	-1.769

*Compounds included in Test set, **Outliers (not included in the final model)

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 12, December 2019

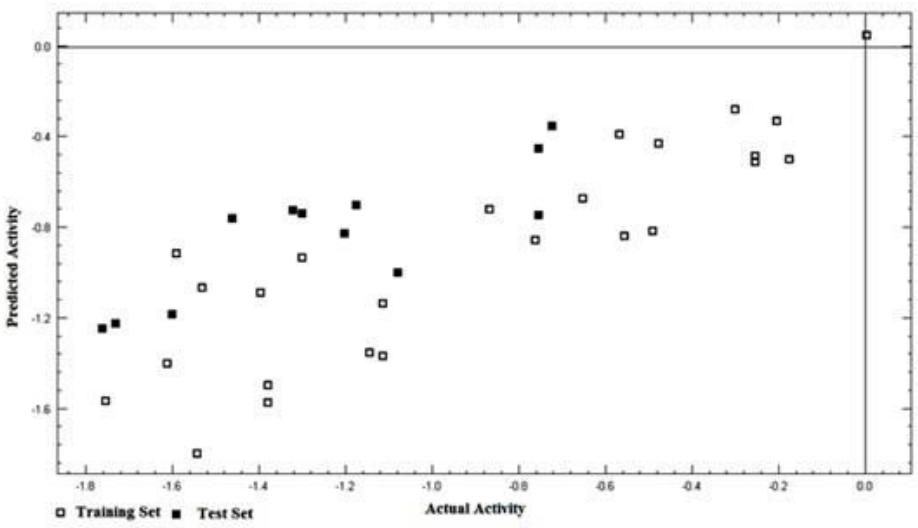


Fig. 1. Graph plotted between actual and predicted activity using similarity based MLR analysis for training and test set of compounds

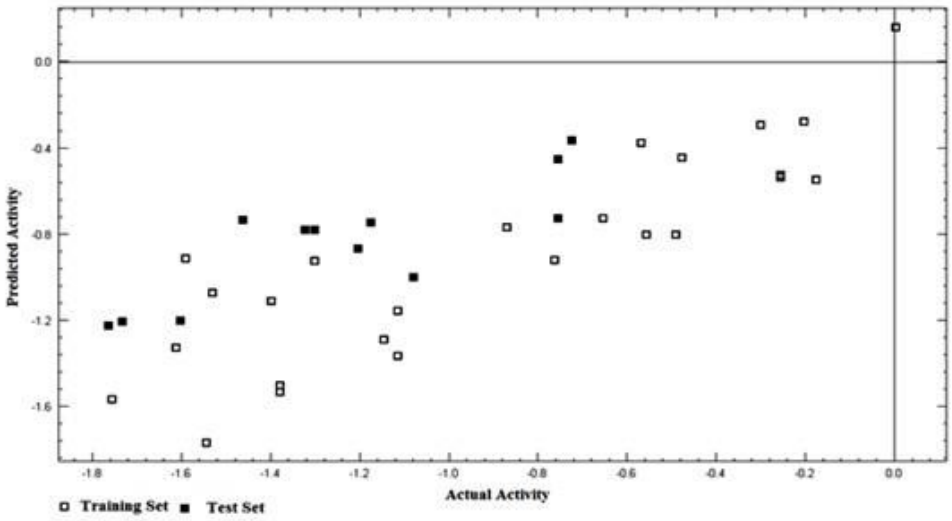


Fig. 2. Graph plotted between actual and predicted activity using similarity based PLS analysis for training and test set of compounds

DOI 10.17148/IARJSET.2019.61208