Digest Journal of Nanomaterials and Biostructures Vol. 5, No 2, May 2010, p. 519 – 533

3D QSAR STUDIES ON A SERIES OF 1,3,4-THIADIAZOLE-2-ARYLHYDRAZONE DERIVATIVES AS ANTITRYPANOSOMAL AGENTS. THE k-NEAREST NEIGHBOR MOLECULAR FIELD ANALYSIS APPROACH

MALLESHAPPA N. NOOLVI*, H. M. PATEL, NAVDEEP SINGH SETHI
Department of Pharmaceutical Chemistry, ASBASJS Memorial College of Pharmacy, Bela (Ropar)-140111, Punjab, India

Megazol is a highly active compound against *Trypanosoma cruzi*, and has become a core structure for the design of new trypanocidal agent. Recently, a new potent trypanocide agent Brazilizone A (a derivatives of megazol), was identified which presents an IC_{50} two fold more potent than the prototype megazol. This result has encouraged us to perform QSAR study on structurally-related 1,3,4-thiadiazole-2-arylhydrazone derivatives (Brazilizones), in order to get a better understanding of their structural features and antiprotozoal activity. The k-Nearest Neighbor Molecular Field Analysis (kNN-MFA), a three dimensional quantitative structure activity relationship (3D-QSAR) method has been used in the present case to study the correlation between the molecular properties and the *Trypanosoma cruzi* inhibitory activities on a series of 1,3,4-thiadiazole-2-arylhydrazone derivatives. kNN-MFA calculations for both electrostatic and steric field were carried out. The master grid maps derived from the best model has been used to display the contribution of electrostatic potential and steric field. The statistical results showed significant correlation coefficient r^2 (q^2) of 0.9455, r2 for external test set (pred_r^2) 0.8087, coefficient of correlation of predicted data set (pred_r^2se) of 0.5873, degree of freedom 20 and k nearest neighbor of 2. Therefore, this study not only casts light on binding mechanism between *Trypanosoma cruzi* and its inhibitors, but also provides new hints for the design of antitrypanosomal agents with observable structural diversity.

(Received April 11, 2010; accepted May 19, 2010)

Keywords: Chagas'disease, Trypanosoma cruzi, 1,3,4-thiadiazole-2-arylhydrazone (Brazilizone A derivatives), k-Nearest Neighbor Molecular Field Analysis (kNN-MFA).

1. Introduction

Chagas' disease is one of the most important parasitic infections of Latin America, with over 17 million people infected, mainly in endemic areas, and at least 120 million people at risk. It is caused by the hemoflagellate protozoan *Trypanosoma cruzi*, which infected humans through the bite of a triatomine insect vector or blood transfusion. The infective trypomastigote form of the parasite penetrates into mammalian cells and undergoes differentiation into proliferative amastigotes. Rupture of these cells leads to liberation of the parasites and perpetuation of the infection, which after several years can lead to the chronic forms o the disease, cardiac and/or digestive[1]. Currently, this pathology is treated with nitro heterocyclic agents such as nifurtimox and benzenidazole. These two drugs are effective against the circulating form of the parasite (trypomastigotes) during the acute phase of the disease, but not during the chronic stage. Additionally, they produce serious adverse effects including mutagenesis. Megazol is a 5-nitroimidazole derivative with antibacterial and antiparasitic activity particularly against trypanosomes[2]. For this reason, megazol was considered as an alternative lead-compound for the treatment of the Chagas' disease. In spite of its impressive antiprotozoal profile, which was associated with its interference with oxygen metabolism as well as its role as thiol scavenger for

trypanothione, cofactor for trypanothione reductase, megazol development was discontinued due to the toxicity and mutagenicity induced by its use in animals. Try to circumvent this undesired profile several megazol analogues were synthesized. However, none of these derivatives have showed to be more potent than the prototype.

Considering this panorama and trying to circumvent this undesirable profile, several megazol analogues belonging to a new class of 1,3,4-thiadiazole-2-arylhydrazone derivatives have been designed and synthesized as attractive antichagasic drug candidates. Due to the hypothesis that the introduction of a radical scavenger subunit linked to the heterocyclic scaffold of megazol could modulate the production of toxic nitro anion radical species, they could potentially avoid mutagenic properties[3].

One could not, however, confirm that the compounds designed would always possess good inhibitory activity to *Trypanosoma cruzi*, while experimental assessments of inhibitory activity of these compounds are time-consuming and expensive. Consequently, it is of interest to develop a prediction method for biological activities before the synthesis[4]. Quantitative structure-activity relationship (QSAR) searches information relating chemical structure to biological and other activities by developing a QSAR model. Using such an approach one could predict the activities of newly designed compounds before a decision is being made whether these compounds should be really synthesized and tested[5].

Many different approaches to QSAR have been developed over the years. The rapid increase in three-dimensional structural information (3D) of bioorganic molecules, coupled with the development of fast methods for 3D structure alignment (e.g. active analogue approach), has led to the development of 3D structural descriptors and associated 3D QSAR methods. The most popular 3D QSAR methods are comparative molecular field analysis (CoMFA) and comparative molecular similarity analysis (CoMSIA)[6,7]. The CoMFA method involves generation of a common three dimensional lattice around a set of molecules and calculation of the steric and electrostatic interaction energies at the lattice points. The interaction energies are numerically very high when a lattice point is very close to an atom and special care needs to be taken in order to avoid problems arising because of this. The CoMSIA method avoids these problems by using similarity function represented as Gaussian. This information around the molecule is converted into numerical data using the partial least squares (PLS) method that reduces the dimensionality of data by generating components. However, a major disadvantage is that PLS attempts to fit a linear curve among all the points in the data set. Further, the PLS method does not offer scope for improvement in results. It has been observed from several reports that the predictive ability of PLS method is rather poor due to fitting of a linear curve between the available points. In the case of the CoMSIA method, molecular similarity is evaluated and used instead of molecular field, followed by PLS analysis.

Variable selection methods have also been adopted for optimal region selection in 3D QSAR methods and shown to provide improved QSAR models as compared to the original CoMFA technique. For example, GOLPE was developed using chemometric principles, and q^2-GRS was developed on the basis of independent analyses of small areas (or regions) of near molecular space to address the issue of optimal region selection in CoMFA[8,9]. These considerations provide an impetus for the development of fast, generally nonlinear, variable selection methods for performing molecular field analysis. With the above facts and in continuation of our research for newer antitrypanosomal agent[10] in the present study, we report here the development of a new method (kNN-MFA) that adopts a k-nearest neighbor principle for generating relationships of molecular fields with the experimentally reported activity to provide further insight into the key structural features required to design potential drug candidates of this class. This method utilizes the active analogue principle that lies at the foundation of medicinal chemistry.

2. Computational methods

2.1. Methodology

We hereby report the models, as generated by kNN-MFA in conjunction with stepwise (SW) forward-backward variable selection methods. In the kNN-MFA method, several models were generated for the selected members of training and test sets, and the corresponding best models are reported herein. VLife Molecular Design Suite (VLifeMDS), allows user to choose probe, grid size, and grid interval for the generation of descriptors. The variable selection methods along with the corresponding parameters are allowed to be chosen, and optimum models are generated by maximizing q^2. k-nearest neighbor molecular field analysis (kNN-MFA) requires suitable alignment of given set of molecules. This is followed by generation of a common rectangular grid around the molecules. The steric and electrostatic interaction energies are computed at the lattice points of the grid using a methyl probe of charge +1. These interaction energy values are considered for relationship generation and utilized as descriptors to decide nearness between molecules. The term descriptor is utilized in the following discussion to indicate field values at the lattice points. The optimal training and test sets were generated using the sphere exclusion algorithm[11]. This algorithm allows the construction of training sets covering descriptor space occupied by representative points. Once the training and test sets were generated, kNN methodology was applied to the descriptors generated over the grid.

Nearest Neighbor (kNN) Method

The kNN methodology relies on a simple distance learning approach whereby an unknown member is classified according to the majority of its k-nearest neighbors in the training set. The nearness is measured by an appropriate distance metric (e.g., a molecular similarity measure calculated using field interactions of molecular structures). The standard kNN method is implemented simply as follows: Calculate distances between an unknown object (u) and all the objects in the training set; select k objects from the training set most similar to object u, according to the calculated distances; and classify object u with the group to which the majority of the k objects belongs. An optimal k value is selected by optimization through the classification of a test set of samples or by leave-one out cross-validation[12].

kNN-MFA with Simulated Annealing

Simulated annealing (SA) is the simulation of a physical process, 'annealing', which involves heating the system to a high temperature and then gradually cooling it down to a preset temperature (e.g., room temperature). During this process, the system samples possible configurations distributed according to the Boltzmann distribution so that at equilibrium, low energy states are the most populated.

kNN-MFA with Stepwise (SW) Variable Selection

This method employs a stepwise variable selection procedure combined with kNN to optimize the number of nearest neighbors (k) and the selection of variables from the original pool as described in simulated annealing.

kNN-MFA with Genetic Algorithm

Genetic algorithms (GA) first described by Holland[13] mimic natural evolution and selection. In biological systems, genetic information that determines the individuality of an organism is stored in chromosomes. Chromosomes are replicated and passed onto the next generation with selection criteria depending on fitness.

2. 2. Chemical Data

Twenty nine 1,3,4-thiadiazole-2-arylhydrazone derivatives as *Trypanosoma cruzi*

inhibitors were taken from the literature and used for kNN-MFA analysis[1-3]. The above reported 1,3,4-thiadiazole-2-arylhydrazone derivatives showed wide variation in their structure and potency profiles. kNN-MFA (3D QSAR) models were generated for these derivatives using a training set of 22 molecules. Predictive power of the resulting models was evaluated by a test set of 7 molecules with uniformly distributed biological activities. Selection of test set molecules was made by considering the fact that test set molecules represent structural features similar to compounds in the training set. The structures of all compounds along with their actual and predicted biological activities are shown in Table 1.

2. 3. Biological Activities

The negative logarithm of the measured IC_{50} (μM) against *Trypanosoma cruzi* as pIC_{50} [$pIC50 = -\log (IC_{50} \times 10^{-6})$] was used as dependent variable, thus correlating the data linear to the free energy change. Since some compounds exhibited insignificant/no inhibition, such compounds were excluded from the present study. All the IC_{50} values had been obtained using the trypomastigote form of *T. cruzi* and the assays were performed in Dulbecco's modified Eagle medium. The IC_{50} values of reference compounds were checked to ensure that no difference occurred between different groups. The pIC_{50} values of the molecules under study spanned a wide range from 2 to 6.

2. 4. Data Set

All computational work was performed on Apple workstation (8-chore processor) using Vlife MDS QSAR plus software developed by Vlife Sciences Technologies Pvt Ltd, Pune, India, on windows XP operating system . All the compounds were drawn in Chem DBS using fragment database and then subjected to energy minimization using batch energy minimization method[14].

2. 5. Molecular Modeling and Alignment

Conformational search were carried out by systemic conformational search method and lowest energy conformers were selected. All the compounds were aligned by template based method. The selection of template molecule for alignment was done by considering the following facts: a) the most active compound; b) the lead or commercial compound; c) the compound containing the greatest number of functional group[15,16]. Generally, the low energy conformer of the most active compound is selected as a reference[17]. In the present study, all the compounds were aligned against minimum energy conformation of most active compound no.2 (Fig.1) by using megazol nucleus as template shown in Fig.2

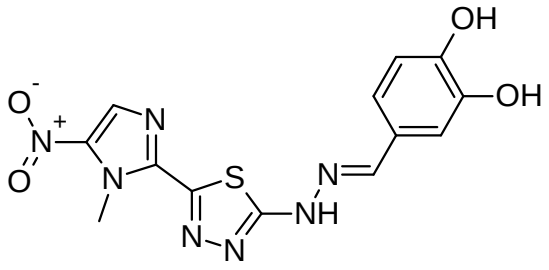


Fig.1. Reference molecule (2) used for alignment by template based alignment.

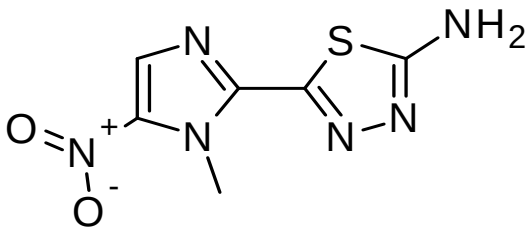


Fig.2. Megazol moiety as a template for alignment.

2. 6. Selection of Training and Test Set

The dataset of 29 molecules was divided into training and test set by Sphere Exclusion (SE) method for model 1, model 2 and model 3 having dissimilarities values of 5.0, 5.3 and 5.1 respectively with pIC_{50} activity field as dependent variable and various 3D descriptors calculated for the compounds as independent variables.

2. 7. Cross-Validation Using Weighted k-Nearest Neighbor

This is done to test the internal stability and predictive ability of the QSAR models. Developed QSAR models were validated by the following procedure:

2. 7.1 Internal Validation

a.) A molecule in the training set was eliminated, and its biological activity was predicted as the weighted average activity of the k most similar molecules (eq.1). The similarities were evaluated as the inverse of Euclidean distances between molecules (eq.2) using only the subset of descriptors corresponding to the current trial solution.

$$wi = \frac{\text{Exp}(-dj)}{\sum \text{Exp}(-dj)}$$

k –Nearest neighbor

$$\hat{y}_i = \sum Wi\, y_i \qquad (1)$$

$$d_{if} = \sum_{k=1}^{vn} (X_{ik} - X_{jk})^2)^{1/2} \qquad (2)$$

b.) Step 1 was repeated until every molecule in the training set has been eliminated and its activity predicted once.

c.) The cross-validated r^2 (q^2) value was calculated using eq. 3, where *yi* and *ŷi* are the actual and predicted activities of the *i*th molecule, respectively, and y_{mean} is the average k-Nearest neighbor activity of all molecules in the training set. Both summations are over all molecules in the training set. Since the calculation of the pair wise molecular similarities, and hence the predictions, were based upon the current trial solution, the q^2 obtained is indicative of the predictive power of the current kNN-MFA model.

$$q^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{mean})^2} \qquad (3)$$

2.7.2 External validation

The predicted r^2 (pred_r^2) value was calculated using eq. 4, where *yi* and *ŷi* are the actual and predicted activities of the *i*th molecule in test set, respectively, and y_{mean} is the average activity of all molecules in the training set. Both summations are over all molecules in the test set. The pred_r^2 value is indicative of the predictive power of the current kNN-MFA model for

external test set.

$$pred\_r^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{mean})^2} \quad (4)$$

Both summations are over all molecules in the test set. Thus, the pred_r^2 value is indicative of the predictive power of the current model for external test set.

2.7.3 Randomization Test

To evaluate the statistical significance of the QSAR model for an actual data set, we have employed a one-tail hypothesis testing. The robustness of the QSAR models for experimental training sets was examined by comparing these models to those derived for random data sets. Random sets were generated by rearranging biological activities of the training set molecules. The significance of the models hence obtained was derived based on calculated Zscore[18].

$$Z\text{score} = \frac{(h - \mu)}{\sigma} \quad (5)$$

Where *h* is the q^2 value calculated for the actual dataset, μ the average q^2, and σ is its standard deviation calculated for various iterations using models build by different random data sets. The probability (α) of significance of randomization test is derived by comparing Zscore value with Zscore critical value, if Zscore value is less than 4.0; otherwise it is calculated by the formula as given in the literature. For example, a Zscore value greater than 3.10 indicates that there is a probability (α) of less than 0.001 that the QSAR model constructed for the real dataset is random. The randomization test suggests that all the developed models have a probability of less than 1% that the model is generated by chance.

3. Experimental

All the twenty nine compounds were built on workstation of molecular modeling software VlifeMDS, which is a product Vlife Sciences Pvt Ltd., India[19]. We hereby report the models, as generated by kNN-MFA in conjunction with stepwise (SW) forward-backward variable selection methods shown in **Table 3**.

In the present kNN-MFA study, (-13.2343 to19.1320) x (-12.0268 to15.04940) x (-11.2513 to 15.4959) A^0grid at the interval of 2.00 was generated around the aligned compounds. The steric and electrostatic interaction energies are computed at the lattice points of the grid using a methyl probe of charge +1 of Gasteiger-Marsili type. These interactions energy values are considered for relationship generation and utilized as descriptors to decide nearness between molecules. The QSAR models were developed using forward-backward variable selection method with pIC_{50} activity field as dependent variable and physico-chemical descriptors as independent variable having cross-correlation limit of 20, 19.2 and 19.5 for model 1, model 2 and model 3 respectively. Selection of test and training set was done by sphere exclusion method having dissimilarity value of 5.0, 5.3 and 5.1 for model 1, model 2 and model 3 respectively. Variance cut off point was 1.0. Numbers of maximum and minimum neighbors were 5 and 2 respectively.

The method described above has been implemented in software, Vlife Molecular Design Suite (VlifeMDS), [19] which allows user to choose probe, grid size, and grid interval for the generation of descriptors. The variable selection methods along with the corresponding parameters are allowed to be chosen, and optimum models are generated by maximizing q^2.

Steps involved in kNN-MFA method

1. Molecules are optimized before alignment optimization is done by MOPAC energy minimization and optimization is necessary process for proper alignment of molecules around template.
2. kNN-MFA method requires suitable alignment of given set of molecules, alignment are template based.
3. This is followed by generation of common rectangular grid around the molecules, the steric and electrostatic interaction energies are computed at the lattice points of the grid using a methyl probe of charge +1.
4. The optimal training and test set were generated using sphere exclusion method.
5. Model was generated by various kNN methods, and models validated internally and externally by leave one out, external validation.
6. Predict the activity of test set of compounds.

Since the final equation are not very useful to represent efficiently the kNN-MFA models, 3D master grid maps of the best models are displayed. They represent area in space where steric and electrostatic field interactions are responsible for the observed variation of the biological activity.

4. Results and discussion

Training set of 22 and test set of 7 1,3,4-thiadiazole-2-arylhydrazone derivatives having different substitution were employed. Following statistical measure was used to correlate biological activity and molecular descriptors: n = number of molecules, Vn = number of descriptors, k = number of nearest neighbor, df = degree of freedom, r^2= coefficient of determination, q^2 = cross validated r^2 (by the leave-one out method), pred_r^2 = r^2 for external test set, pred_r^2se = coefficient of correlation of predicted data set, Z score = the Z score calculated by q^2 in the randomization test, best_ran_q^2 = the highest q^2value in the randomization test and α = the statistical significance parameter obtained by the randomization test.

Selecting training and test set by spherical exclusion method, Unicolumn statics shows that the max of the test is less than max of train set and the min of the test set is greater than of train set shown in **Table 2**, which is prerequisite analysis for further QSAR study. The above result shows that the test is interpolative i.e. derived within the min-max range of the train set. The mean and standard deviation of the train and test provides insight to the relative difference of mean and point density distribution of the two sets. In this case the mean in the test set higher than the train set shows the presence of relatively more active molecules as compared to the inactive ones. Also the similar standard deviation in both set indicates that the spread in both the set with their respective mean is comparable.

The activity distribution graph shows the comparison between the activity of training and test set. It can be observed from Hierarchical Graph that the test set molecule activities lie within the range of training set, shown in **Fig.3.**

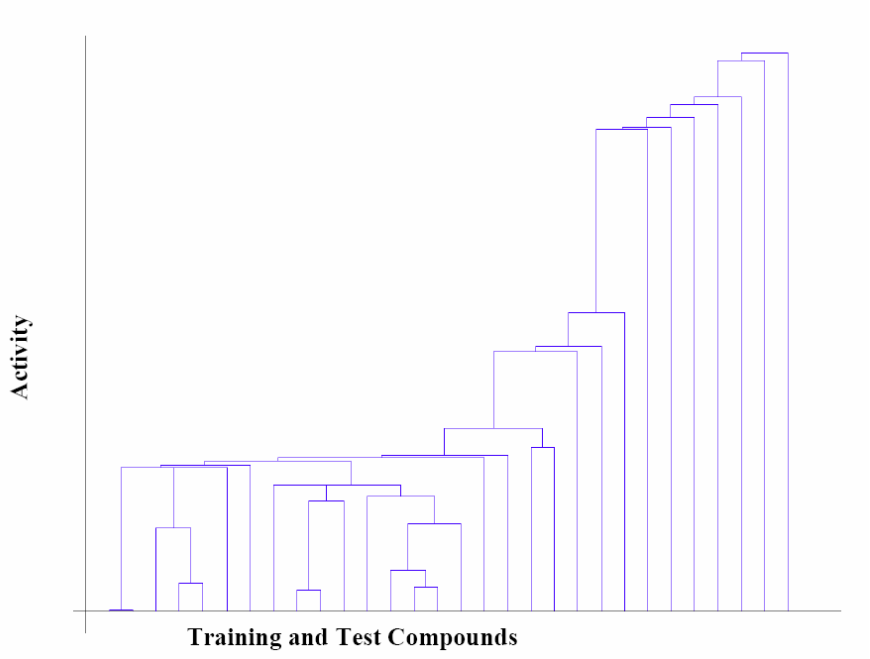
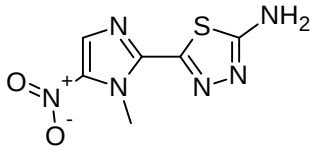
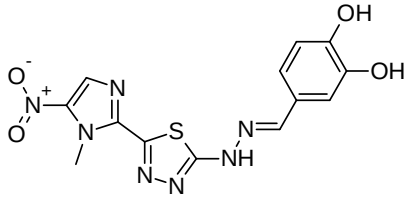
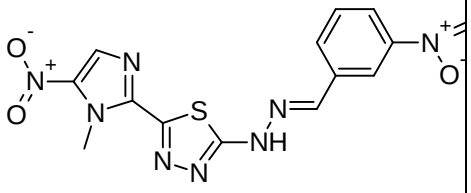
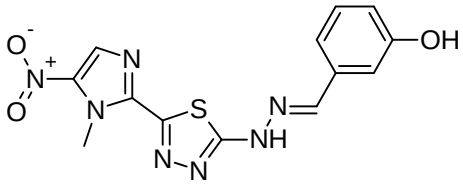
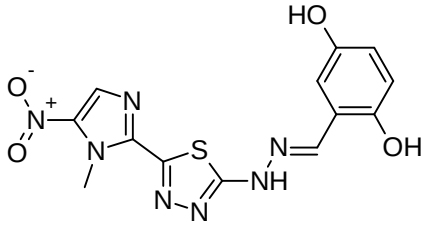
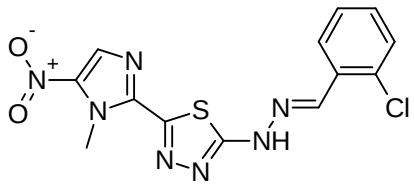
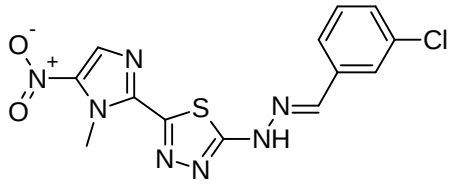
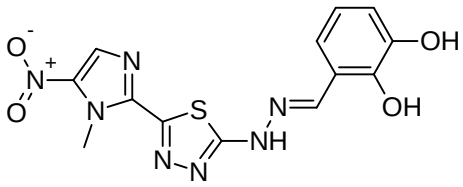


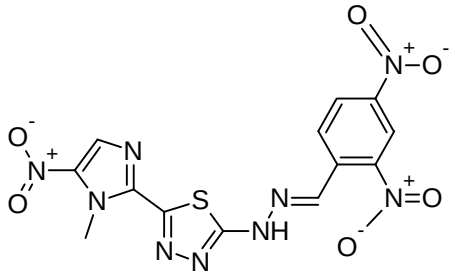
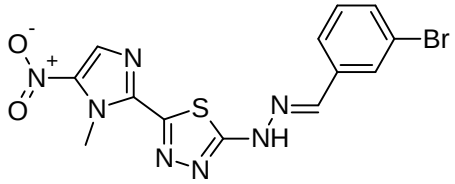
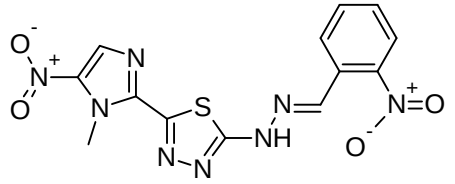
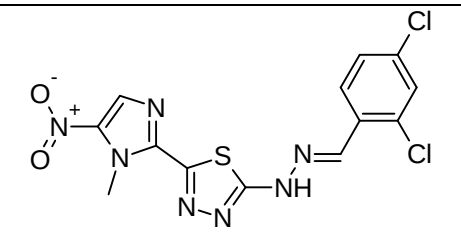
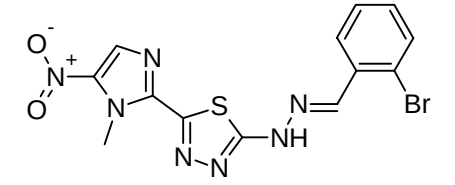
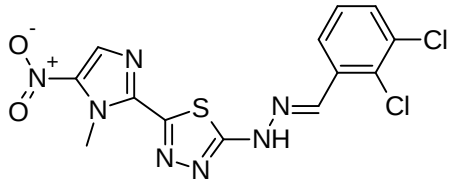
Fig 3. Hierarchical Graph Showing Uniform Distribution of Training and Test Set

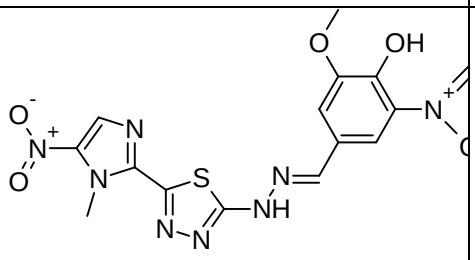
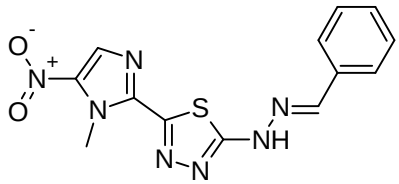
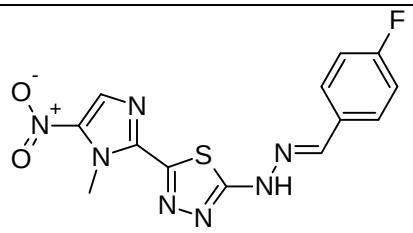
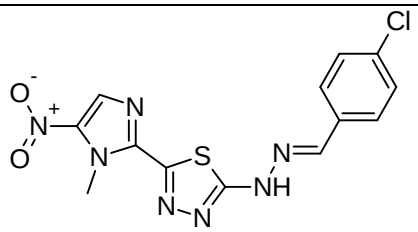
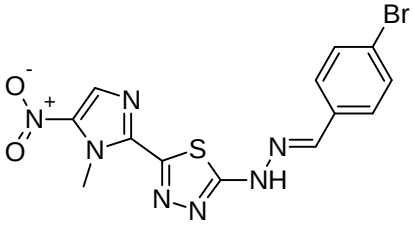
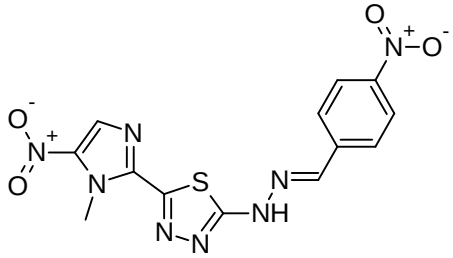
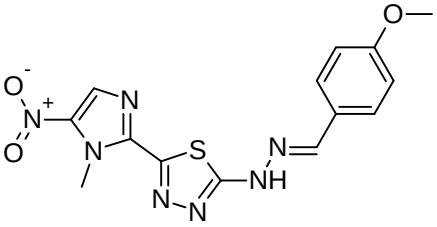
The observed and predicted pIC50 along with residual values for model 1 are shown in **Table 1.** The plot of observed vs. predicted activity is shown in **Fig.4.** From the plot it can be seen that kNN-MFA model is able to predict the activity of training set quite well (all points are close to regression line) as well as external.

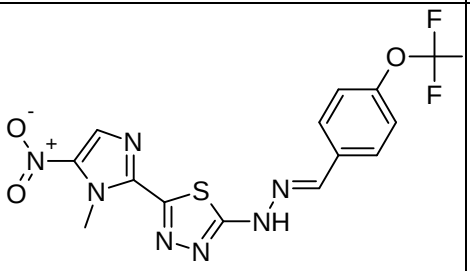
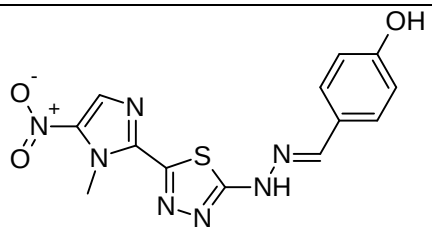
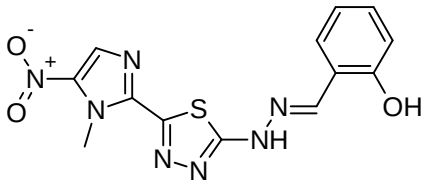
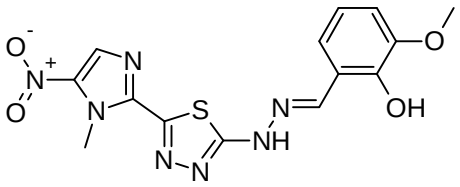
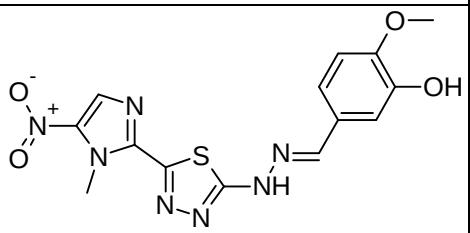
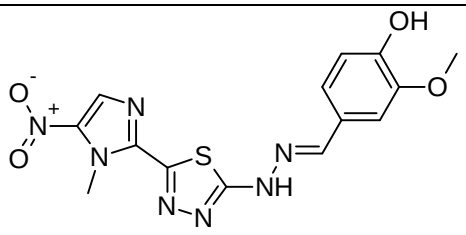
Table 1. Structure, Experimental and Predicted Activity of 1,3,4-thiadiazole-2-arylhydrazone used in training and test set by model-1. IC_{50} = Compound concentration in micro mole that led to 50% lysis of the parasite, pIC_{50} = -Log ($IC_{50} \times 10^{-6}$): Training and Test data set developed using model 1 T = Test set molecule.

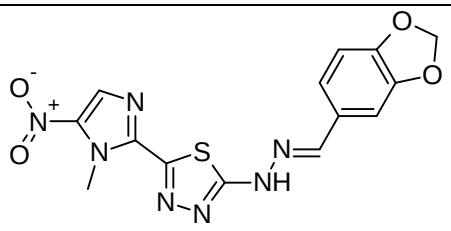
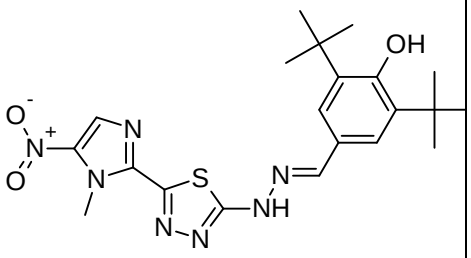
S.NO	Molecules	IC_{50} (µm)	pIC_{50}		Residual
			Experimental activity	Predicted activity	
1^T		9.9	5.004	4.422	0.582
2^T		5.3	5.275	5.488	-0.213

S.NO	Molecules	IC_{50} (μm)	pIC_{50}		Residual
3		48.2	4.317	3.942	0.375
4		73.9	4.131	3.908	0.223
5		140.0	3.854	3.963	-0.109
6		140.3	3.853	3.535	0.318
7		186.9	3.728	3.795	-0.067
8		309.3	3.509	3.589	-0.008

S.NO	Molecules	IC_{50} (μm)	pIC_{50}		Residual
9^T		330.2	3.481	3.838	-0.357
10		727.6	3.138	3.841	-0.703
11		726.2	3.139	3.249	-0.110
12^T		793.5	3.101	3.160	-0.059
13		900.0	3.046	3.213	-0.167
14		1208.8	2.918	2.958	-0.040

S.NO	Molecules	IC_{50} (μm)	pIC_{50}		Residual
15^{T}		19.0	4.721	3.519	1.202
16		200	3.699	3.302	0.397
17		200	3.699	4.007	-0.308
18		38.4	4.415	4.422	-0.007
19		17.0	4.769	4.125	0.644
20		22.0	4.657	4.102	0.555
21		200	3.699	3.840	-0.141

S.NO	Molecules	IC_{50} (μm)	pIC_{50}		Residual
22		200	3.699	3.655	0.044
23^{T}		11.6	4.935	4.698	0.237
24		54.2	4.266	4.267	-0.001
25		200	3.699	3.900	-0.201
26		50.5	4.296	3.858	0.438
27		200	3.699	4.267	-0.568

S.NO	Molecules	IC_{50} (µm)	pIC_{50}		Residual
28^{T}		200	3.699	3.781	-0.082
29		33.2	4.478	3.863	0.615

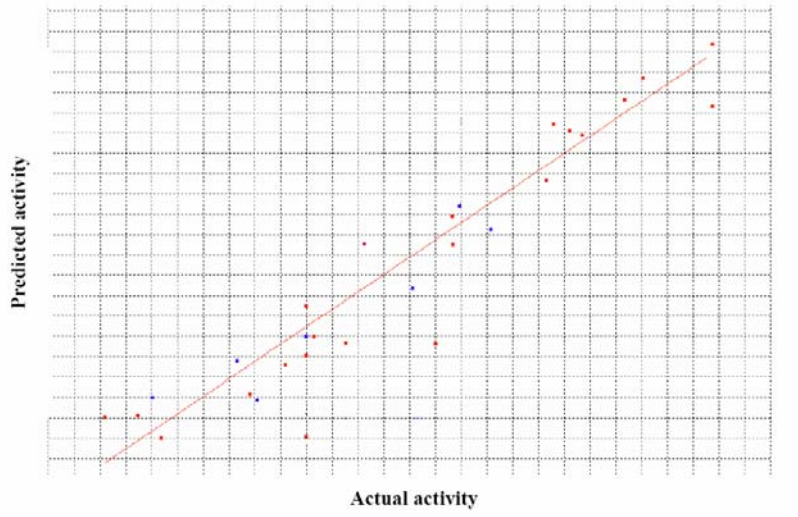


Fig 4. Graph of Actual vs. Predicted activities for training and test set molecules from the kNN-MFA model 1, A) Training set (Red dots) B) Test set (Blue dots)

Table 2. Unicolumn Statics of Training and Test Sets.

Unicolumn statics	Average	Max	Min	Std. Deviation
For Training Set	3.9642	5.2750	2.9180	0.7310
For Test Set	4.1763	4.4780	3.7280	0.3101

During the kNN-MFA investigation, dissimilarity value for the selection of training and test by spherical exclusion method of range 5.000 to 6.500 were investigated. The dissimilarity value of 5.000 produced a significant result as compare to the 5.100 and 5.300 shown in the **Table 3.** Further increases in resolution have produced decrease in model quality. From the **Table 3** it

was observed that the results were less sensitive to resolution of dissimilarity value.

Table 3. Stastical Results of kNN-MFA method.

Parameters	Model 1 (Dissimilarity value = 5.0)	Model 2 (Dissimilarity value = 5.3)	Model 3 (Dissimilarity value = 5.1)
n	22	19	20
k	2	2	2
q^2	0.9455	0.7095	0.8197
pred_r^2	0.8087	0.3569	0.4038
pred_r^2se	0.5873	0.6985	0.6850
Z score	4.4032	4.90615	4.2120
best_ran_q^2	0.72332	0.39887	0.49614
α_ran_q^2	0.00001	0.00002	0.0001
Descriptors	E_634 -1.0241 -0.8484	E_745 -0.3133 -0.0011	E_1635 0.1810 0.2870
	E_500 -0.7053 -0.4139	S_533 -0.2987 0.2386	S_2195 -0.0142 -0.0100
	E_1140 0.6002 1.2771	E_1262 0.0797 0.1640	E_2682 0.0891 0.1154
Vn	S_1060 -0.0202 0.0313 04	03	03

n, number of observations (molecules); Vn, number of descriptors; k, number of nearest neighbors; q^2, cross-validated r^2 (by the leave-one out method); pred_r^2, predicted r^2 for the external test set; Zscore, the Zscore calculated by q^2 in the randomization test; best_ran_q^2, the highest q^2 value in the randomization test and α _ran_q^2, the statistical significance parameter obtained by the randomization test.

It is known that the CoMFA method provides significant value in terms of a new molecule design, when contours of the PLS coefficients are visualized for the set of molecules. Similarly, the kNN-MFA models provide direction for the design of new molecules in a rather convenient way. The points which contribute to the kNN-MFA model 1 are displayed in **Fig. 5**. The range of property values for the chosen points may aid in the design of new potent molecules **(Fig. 5)**.

The range is based on the variation of the field values at the chosen points using the most active molecule and its nearest neighbor set.

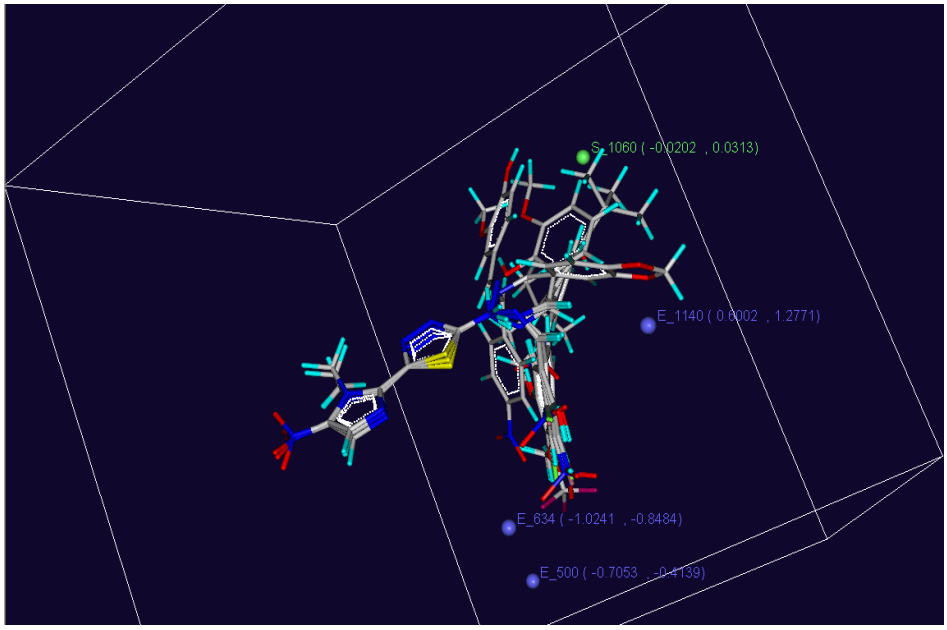


Fig. 5. 3D-alignement of molecule with the important steric and electrostatic point Contributing to the model with range of values shown in parenthesis

The q^2, pred_r^2, Vn and k value of kNN-MFA with model 1, 2 and 3 were (0.9455, 0.8087, 04/2) (0.7095, 0.3569, 03/2) and (0.8197, 0.4038, 03/2) respectively. Among these three methods, model 1 have better q2 (0.9455) and pred_r2 (0.8087) than other two models, model 1 correctly predicts activity 94.55% and 80.87% for the training and test set respectively. It uses 1 steric and 3 electronic descriptors with 2 k nearest neighbor to evaluate activity of new molecule. The model is validated by $\alpha\_ran\_q^2$ = 0.00001, $best\_ran\_q^2$ = 0.72332, and $Zscore\_ran\_q^2$ = 4.4032 .The randomization test suggests that the developed model have a probability of less than 1% that the model is generated by chance.

The kNN-MFA models obtained by using all the three dissimilarity values showed that electrostatic and steric interactions play major role in determining biological activity. S_1060 in model 1, S_533 in model 2 and S_2195 in model 3 are steric field descriptors similarly E_634, E_500, E_1140 in model 1, E_745, E_1262 in model 2 and E_1635, E_2682 in model 3 are electrostatic field descriptors.

Negative value in electrostatic field descriptors indicates that negative electronic potential is required to increase activity and more electronegative substituents group is preferred in that position, positive range indicates that group that imparting positive electrostatic potential is favorable for activity so less electronegative group is preferred in that region. Similarly negative range in steric descriptors indicates that negative steric potential is favorable for activity and less bulky substituents group is preferred in that region, positive value of steric descriptors reveals that positive steric potential is favorable for increase in activity and more bulky group is preferred in that region.

5. Conclusions

In conclusion, the model developed to predict the structural features of 1,3,4-thiadiazole-2-arylhydrazone to inhibit *Trypanosoma cruzi*, reveals useful information about the structural features requirement for the molecule. In all three optimized models, Model 1 is giving very significant results. The master grid obtained for the various kNN-MFA models show that negative value in electrostatic field descriptors indicates the negative electronic potential is required to increase activity and more electronegative substituents group is preferred in that

position, positive range indicates that the group which imparts positive electrostatic potential is favorable for activity so less electronegative group is preferred in that region. Negative range in steric descriptors indicates that negative steric potential is favorable for activity and less bulky substituents group is preferred in that region. Positive value of steric descriptors reveals that positive steric potential is favorable for increase in activity and more bulky group is preferred in that region. On the basis of the spatial arrangement of the various shapes, electrostatic and steric potential contributions model proposed in this work is useful in describing QSAR of 1,3,4-thiadiazole-2-arylhydrazone derivatives as *Trypanosoma cruzi* inhibitor and can be employed to design new derivatives of 1,3,4-thiadiazole-2-arylhydrazone with specific inhibitory activity.

Acknowledgement

The authors would like to thanks Sardar Sangat Singh Longia, Secretary, ASBASJSM College of Pharmacy, Bela (Ropar) for providing the necessary facilities and Vlife Science Pune for providing the software

Reference

[1] S. A. Carvalho, E. F. da Silva, R. M. Santa-Rita, S. L. de Castro, C. A. M. Fraga, Bioorg Med Chem Lett, **14**, 5967 (2004)
[2] J. D. Maya, S. Bollo, L. J. Nunez-Vergara, J. A. Squella, Y. Repetto, A. Morello, J. Perie, G. Chauviere, Biochem Pharrmacol, **65**, 999-1006 (2003)
[3] S. A. Carvalho, F. A. S. Lopes, K. Salomao, N. C. Romeiro, S. M. S. V. Wardell, S. L. de Castro, E. F. da Silva, C. A. M. Fraga, Bioorg Med Chem, **16**, 413 (2008)
[4] A. J. Bridges, Chem Rev, **101**, 2541 (2001)
[5] A. M. Badiger, M. N. Noolvi, P. V. Nayak, Lett Drug Des Discovery, **3**, 550 (2006)
[6] R. D. Cramer, D. E. Patterson, J. D. Bunce, J Am Chem Soc, **110**, 5959 (1988)
[7] G. Klebe, U. Abraham, T. Mietzner, J Med Chem, **37**, 24 (1994)
[8] M. Baroni, G. Costantino, G. Cruciani, D. Riganelli, R. Valigi, S. Clementi, Quant Struct-Act Relat, **12**, 9 (1993)
[9] S. J. Cho, A. Tropsha, J Med Chem, **38**, 1060 (1995)
[10] S. N. Manjula, M. N. Noolvi, K. V. Parihar, S. A. Manohara Reddy, V. Ramani, A. K. Gadad, G. Singh, N. G. Kutty, M. Rao, Eur J Med Chem, 1-7 (2009)
[11] A. Golbraikh, A. Tropsha, J Chem Inf Comput Sci, **43**, 144 (2003)
[12] M. A. Sharaf, D. L. Illman, B. R. Kowalski, Chemometrics, Wiley, New York, (1986)
[13] J. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, (1975).
[14] W. Zheng, A. Tropsha, J Chem Inf Comput Sci, **40**, 185 (2000)
[15] A. Agarwal, E.W. Taylor, J Computat Chem, **14**, 237 (1993)
[16] N. Baurin, E. Vangrevelinghe, L. M. Allory, et al, J Med Chem, 43, 1109 (2000)
[17] M. Xu, A. Zhang, S. Han, L. Wang, Chemosphere, **48**, 707 (2002)
[18] N. Gilbert, Statistics, W. B. Saunders Co, Philadelphia, PA (1976)
[19] VLifeMDS2.0; Molecular Design Suite, Vlife Sciences Technologies Pvt. Ltd., Pune, India (2004), www.vlifesciences.com