

# Comprehensive Structure Activity Relationship Studies for Angiotensin II Receptor Antagonists as Antihypertensive Agents

Anupama Parate\*<sup>1</sup>, Rajesh Sharma<sup>2</sup>, Subhash Chandra Chaturvedi<sup>3</sup>

<sup>1</sup>School of Pharmacy, Devi Ahilya Vishwavidyalaya, Indore (MP), INDIA

<sup>2</sup>School of Pharmacy, Devi Ahilya Vishwavidyalaya, Indore (MP), INDIA

<sup>3</sup>Aurbindo Institute of Pharmacy, Indore (MP), INDIA

## Address for Correspondence

School of Pharmacy,  
Devi Ahilya  
Vishwavidyalaya,  
Indore (MP), INDIA,  
Tel:+ 919425011060

## E-mail:

[anupama\\_sainy@  
yahoo.com](mailto:anupama_sainy@yahoo.com)

## ABSTRACT

Angiotensin II receptor antagonists (ATIIRA) has become an attractive molecular target for drugs that aim to treat hypertension triggered by renin angiotensin system. To study the relationship between the structure of several ATIIRA we have performed a two dimensional and three-dimensional quantitative structure–activity relationship (QSAR) study of benzimidazole based derivatives. A series of 40 compounds containing 4, 5, 6, 7 substituted benzimidazoles were subjected to comprehensive 2D and 3D advanced kNN-MFA QSAR analysis employing multiple linear regression, partial least square, principle component analysis, advanced kNN molecular field analysis, stepwise forward back method, simulated annealing and genetic algorithm method. The model allowed the identification of relevant structural features required for the interaction with the AT<sub>1</sub> receptor, enabling the prediction of activity of molecules. Some highly predictive 2D and 3D-QSAR models, with significant models with  $r^2 = 0.84$  and  $r^2 = 0.83$  were obtained in 2D analysis and with  $q^2 = 0.77$  and  $q^2 = 0.67$  by advanced kNN MFA method in 3D analysis. These models are in good agreement with the structural characteristics of the potential angiotensin II receptor antagonists and provide some structural insights for the improvement of bioactivities.

**Keywords:** Antihypertensive agents; Angiotensin II receptor antagonists; QSAR; knn-MFA; Substituted benzimidazoles.

## INTRODUCTION

Hypertension is a major risk factor for cerebro-cardiovascular diseases; the renin-angiotensin-system (RAS) plays a

pivotal role in many cardiovascular and renal diseases, including hypertension, heart failure, renal artery stenosis and diabetic &

nondiabetic nephropathies<sup>1</sup>. Angiotensin II is one of the most powerful endogenous vasoconstrictors produced by limited and very specific proteolysis of its precursor protein, angiotensin I in RAS. The effects of angiotensin II include constricting vascular smooth muscle cells directly and thereby producing hypertension when those cells are in small arterioles, angiotensin II increases myocardial contractility, stimulates aldosterone release by the adrenal gland (leading to salt and water retention and exacerbating hypertension), and stimulates catecholamine release from sympathetic nerve endings, which serves to raise blood pressure (BP) even further. Angiotensin II is also involved in cell growth and proliferation, with its greatest impact in human biology and disease in the heart, kidney, and cerebral vessels. The action of Ang II is mediated through selective membrane bound Angiotensin II receptors Type 1 (AT<sub>1</sub>) and Type 2 (AT<sub>2</sub>). These receptors have been identified and belong to the G-protein coupled receptor super family (GPCRs). The AT<sub>1</sub> receptor exists in the blood vessels, liver, kidneys, adrenal cortex, and heart, and cardiovascular effects of AT II are mainly mediated by AT<sub>1</sub> receptor<sup>2,3</sup>. The type 1 (AT<sub>1</sub>) receptor for the octapeptide hormone angiotensin II (Ang II) is a member of the G-protein-coupled receptor super family (GPCRs)<sup>4</sup>. In the last decades several selective antagonists have been designed developed and are used to treat both hypertension and damage associated with the diseases such as atherosclerosis and diabetes<sup>5-15</sup>. Numerous data sets which are reported in the literature were subjected to QSAR analysis in order to design novel angiotensin II receptor antagonists<sup>16-28</sup>. In order to understand the design and key findings of experimental studies of Ries et. al. a 2D and 3D model for QSAR was generated by advanced methods employing Vlife MDS software package,

version 3.0<sup>29</sup> to optimize biological activity and to design novel surrogates.

## EXPERIMENTAL PROTOCOL

### 2D Model Building

The physicochemical descriptors are based on the physicochemical properties of molecule. Another class of descriptors called the Alignment Independent (AI) descriptors around more than 700 descriptors AI descriptors are calculated<sup>30</sup>. For calculation of AI descriptors every atom in the molecule was assigned at least one and at most three attributes. The first attribute is 'T-attribute' to thoroughly characterize the topology of the molecule. The second is the atom type attribute. The atom symbol is used here. The third attribute is assigned to atoms taking part in a double or triple bond. After all atoms have been assigned their respective attributes, selective distance count statistics for all combinations of different attributes are computed. Three significant statistical methods were used while establishing a 2D QSAR relationship between the biological activity and physicochemical parameters namely multiple linear regression (MLR), partial least square analysis (PLS) and principle component analysis (PCA) methods are used to build a QSAR model. The QSAR model can then be used to predict activities for new molecules, for screening a large set of molecules whose activities are not known.

### Selection of dataset

The in vitro AT<sub>1</sub> activity values pIC<sub>50</sub> (nM) of 6- substituted benzimidazoles as shown in Table 1 was used as dependent variable and physicochemical and alignment independent descriptors in the 2D QSAR study while molecular fields as independent variables in the 3D QSAR study [8]. In 2D QSAR analysis three statistical methods (MLR, PLS, PCA) were applied on different combinations of test and training set to yield

six models (Model 1, 2, 3, 4, 5 and 6). The models 1, 2 and 3 were generated by sphere exclusion algorithm with a dissimilarity value of 2.2. The dataset was finally divided in a training set of 30 molecules and a test set of 10 molecules. In Model 1, 2 and 3 the test set comprised of ten molecules BZ1, BZ2, BZ3, BZ4, BZ6, BZ8 BZ11, BZ13, BZ15 and BZ17; remaining molecules are kept in training set. The unicolon statistics for the mentioned test set and training set is shown in Table 2. The other three models 4, 5 and 6 were generated by manual method by dividing the test and training set on the basis of structural diversity. The test set consists of eight molecules BZ1, BZ2, BZ3, BZ8, BZ9, BZ19, BZ32, and BZ33 remaining molecules of the data set were kept in training set. The activity distribution plot for Model 4, 5, and 6 is shown in Figure 1. The QSAR models with pertinent statistical parameters are shown in Table 3.

#### Validation of 2D Models

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

##### Internal Validation

The Internal validation was carried out using leave-one-out (LOO) method. For calculating  $q^2$ , each molecule in the training set was eliminated once and the activity of the eliminated molecule was predicted by using the model developed by the remaining molecules. The  $q^2$  was calculated using the equation which describes the internal stability of a model:

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

where  $y_i$ , and  $\hat{y}_i$  are the actual and predicted activity of the  $i^{\text{th}}$  molecule in the training set,

respectively, and  $y_{\text{mean}}$  is the average activity of all molecules in the training set.

##### External Validation

For external validation, the activity of each molecule in the test set was predicted using the model developed by the training set. The  $\text{pred}_r^2$  value is calculated as follows:

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

where  $y_i$ , and  $\hat{y}_i$  are the actual and predicted activity of the  $i^{\text{th}}$  molecule in the test set, respectively, and  $y_{\text{mean}}$  is the average activity of all molecules in the training set. Both summations are over all molecules in the test set. Thus, the  $\text{pred}_r^2$  value is indicative of the predictive power of the current model for the external test set. Both summations are over all molecules in the test set.

##### Randomization Test

To evaluate the statistical significance of the QSAR model for an actual dataset, one tail hypothesis testing was used. The robustness of the models for training sets was examined by comparing these models to those derived for random datasets. Random sets were generated by rearranging the activities of the molecules in the training set. The statistical model was derived using various randomly rearranged activities (random sets) with the selected descriptors and the corresponding  $q^2$  were calculated. The significance of the models hence obtained was derived based on a calculated Z score. Z score value is calculated by the following formula:

$$\text{Zscore} = \frac{(h - \mu)}{\sigma}$$

where  $h$  is the  $q^2$  value calculated for the actual dataset,  $\mu$  the average  $q^2$ , and  $\sigma$  is its standard deviation calculated for various

iterations using models build by different random datasets. The probability ( $\alpha$ ) of significance of randomization test is derived by comparing Z score value with Z score critical value, if Z score value is less than 4.0, otherwise it is calculated by the formula as given in the literature. For example, a Z score value greater than 3.10 indicates that there is a probability ( $\alpha$ ) of less than 0.001 that the QSAR model constructed for the real dataset is random.

#### Evaluation of the QSAR Models

The developed QSAR models were evaluated using the following statistical measures: number of observations (molecules); k, number of variables; optimum component, number of optimum PLS components in the model;  $r^2$ , coefficient of determination;  $q^2$ , cross-validated  $r^2$  (by leave one-out);  $pred_r^2$ ,  $r^2$  for external test set; Z score, Z score calculated by the randomization test;  $best\_ran\_q^2$ , highest  $q^2$  value in the randomization test;  $best\_ran\_r^2$ , highest  $r^2$  value in the randomization test;  $\alpha$ , statistical significance parameter obtained by the randomization test; SEE, standard error of estimate of the model; SECV, standard error of cross-validation; and SEP, standard error of external test set prediction.

#### 3D Model Building

k-nearest neighbor molecular field analysis (kNN-MFA) requires suitable alignment of given set of molecules calculation of descriptors. This is followed by generation of a common rectangular grid around the molecules where the steric and electrostatic interaction energies are computed at the lattice points of the grid using a methyl probe of charge +1 which act as 3D descriptors. These interaction energy values are considered for relationship generation and utilized as descriptors to decide nearness between molecules. The descriptors that get selected in a given model

are the field points either of steric or electrostatic nature at different locations in a common grid around set of molecules. The extreme of field values of compounds in the cluster of most active compounds decide range of field values which is recommended for new compound design. The kNN-MFA requires suitable alignment of set of molecules. This is followed by generation of a common rectangular grid around the molecules. The steric and electrostatic energies are computed at the lattice points of the grid using methyl probe of charge +1. These interaction energy values at the grid points are considered for relationship generation using kNN method and utilized as descriptors for obtaining distances within this method. An optimal training and test set can be generated for kNN method using sphere exclusion method. This algorithm allows constructing training sets covering all descriptor space areas occupied by representative points. It is expected that the predictive ability of QSAR models generally decreases when the dissimilarity level increases. Once the training and test sets are generated, kNN methodology is applied to descriptors generated over the grid. The methods adopted for 3D QSAR analysis are discussed below:

#### a. kNN Stepwise forward backward molecular field analysis (SWFB)<sup>31</sup>

The kNN technique is a conceptually simple approach to pattern recognition problems. In this method, an unknown pattern is classified according to the majority of the class memberships 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: (i) calculate distances between an unknown object (u) and all the objects in the training set; (ii) select k

objects from the training set most similar to object  $u$ , according to the calculated distances, (iii) classify object  $u$  with the group to which a majority of the  $k$  objects belong. An optimal  $k$  value is selected by the optimization through the classification of a test set of samples or by the leave-one out cross-validation. The variables and optimal  $k$  values are chosen using different variable selection methods as described below.

#### b. Genetic algorithm (GA)<sup>32</sup>

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. Genetic information can however be altered through genetic operations such as mutation and crossover. In genetic algorithm, each “chromosome” is a set of genes, which constitutes a candidate solution to the discrimination problem. A population of “chromosomes” is used. The passage of each “chromosome” to the next generation is determined by its relative fitness, i.e., the closeness of its properties to those desired. Random combinations and/or changes of the transmitted “chromosomes” produce variations in the next generation of “offspring”. Better the fitness (correspondence with desired properties), greater is the chance of that chromosome being selected for transmission. Optimal or near optimal solutions are obtained through evolution over many generations. There are four major components of GA: chromosome generation, fitness assessment, selection, and mutation. This method employs a stochastic variable selection procedure, combined with kNN, to optimize

- (i) The number of nearest neighbors ( $k$ ) and
- (ii) The selection of variables from the original pool as described in simulated annealing.

#### c. Simulated annealing (SA)<sup>33</sup>

The idea of simulated annealing is to simulate a physical process called annealing in which a system is heated to a high temperature and then is gradually lowered to a preset temperature value (e.g. room temperature). During this process the system samples possible configurations according to Boltzmann distribution. At equilibrium low energy states will be mostly populated. The first implementation of the SA procedure was described by Metropolis et al., followed by the development of a more generalized mathematical optimization protocol. The implementation of SA is as follows:

- (i) Generate a trial solution to the underlying optimization problem; i.e. a kNN-MFA model is built based on a random selection of descriptors, a trial hypothetical pharmacophore (HP).
- (ii) Calculate the value of the fitness function, which characterizes the quality of the trial solution to the underlying problem i.e. the  $q^2$  value for a QSAR model built using only the HP descriptors ( $q^2_{curr}$ ).
- (iii) Perturb (i.e. slightly modify) the trial solution to obtain a new solution i.e. change a fraction of the current HP descriptors to other randomly selected descriptors and build a new kNN-MFA model for the new trial HP.
- (iv) Calculate the new value of the fitness function ( $q^2_{new}$ ) for the new trial solution.
- (v) Apply the optimization criteria. If  $q^2_{curr} < q^2_{new}$  the new solution is accepted and used to replace the current trial solution, if  $q^2_{curr} > q^2_{new}$ , the new solution is accepted only if the following Metropolis criterion is satisfied i.e. where  $rnd$  is a random number uniformly distributed between 0 and 1 and  $T$  is a parameter analogous to the temperature in Boltzmann distribution law,  $rnd < e^{- (q^2_{curr} - q^2_{new}) / T}$
- (vi) Steps iii-v is repeated until the termination condition is satisfied. The temperature-lowering scheme and the termination condition used in this work have

been adapted from Sun et al. Thus, every time when a new solution is accepted or when a preset number of successive steps of generating trial solutions (100 steps) do not lead to a better result, the temperature is lowered by 10% (the default initial temperature is 1000). The calculations are terminated when either the current temperature of simulations is lowered to the value of  $(T) 10^{-6}$  or the ratio between the current temperature and the temperature corresponding to the best solution found is equal to  $10^{-6}$ .

#### Selection of dataset

A set of 40 compounds used in 2D QSAR were now subjected to 3D QSAR analysis by advance kNN molecular field analysis. The 3D QSAR analysis resulted in six models for different combinations of test and training sets. Sphere exclusion algorithm method was opted for division of test and training set. Three models (Model 7, 8, 9) were obtained when dissimilarity value was set to 9.0. The set of compounds BZ2, BZ3, BZ5, BZ6, BZ7, BZ31, BZ33 and BZ37 - comprised of test set and the remaining molecules were in training set where the value of dissimilarity value was set to 9.0. The other three models (Model 10, 11, 12) were obtained when dissimilarity value was set to 9.5. The compounds BZ1, BZ2, BZ3, BZ8, BZ9, BZ19, BZ32 and BZ33 comprised of test set and the remaining molecules were in the training set when the dissimilarity value was set to 9.5. The Unicolumn statistics for training set and test set for 3D study is reported in Table 5. Three advanced kNN molecular field analysis methods (SWFB, SA and GA) were employed in 3D QSAR analysis as shown in Table 6.

#### Validation of 3D Models

This is done to test the internal stability and predictive ability of the QSAR

models. Developed QSAR models were validated by the following procedures:

#### Cross-Validation Using Weighted k-Nearest Neighbor

The standard leave-one-out procedure was implemented and can be summarized as follows:

(1) 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.

$$w_i = \frac{\exp(-d_j)}{\sum_{\text{k-nearest neighbors}} \exp(-d_j)}$$

$$\hat{y}_i = \sum w_i y_i \quad (1)$$

$$d_{ij} = \left[ \sum_{k=1}^{n} (X_{ik} - X_{jk})^2 \right]^{1/2} \quad (2)$$

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

(3) The cross-validated  $r^2$  ( $q^2$ ) value was calculated using eq. 3, where  $y_i$  and  $\hat{y}_i$  are the actual and predicted activities of the  $i^{\text{th}}$  molecule, respectively, and  $y_{\text{mean}}$  is the average 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.

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

#### External Validation

The following procedure was applied for external validation.

(1) Predict the biological activity of a molecule in the test set as the weighted average activity of the  $k$  most similar molecules in the training set (eq. 1). The similarities were evaluated as the inverse of Euclidean distances between molecules (eq. 2) as calculated using the descriptors determined by the current model.

(2) Step 1 was repeated for every molecule in the test set.

(3) The predicted  $r^2$  ( $\text{pred}_r^2$ ) value was calculated using eq. 4, where  $y_i$  and  $\hat{y}_i$  are the actual and predicted activities of the  $i^{\text{th}}$  molecule in test set, respectively, and  $y_{\text{mean}}$  is the average activity of all molecules in the training set. Both summations are over all molecules in the test set. The  $\text{pred}_r^2$  value is indicative of the predictive power of the current kNN-MFA model for external test set.

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

#### Randomization Test

To evaluate the statistical significance of the QSAR model for an actual data set, one-tail hypothesis testing was employed. 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 Z-score.

#### Evaluation of the QSAR Models

The QSAR models were evaluated using following statistical measures:  $n$ , number of observations (molecules);  $V_n$ , number of descriptors;  $k$ , number of nearest neighbors;  $q^2$ , cross-validated  $r^2$  (by the leave-one-out method),  $\text{pred}_r^2$ , predicted  $r^2$  for the external test set; Z-score, the Z score calculated by  $q^2$  in the randomization test;

best\_ran\_ $q^2$ , the highest  $q^2$  value in the randomization test, and  $R$ , the statistical significance parameter obtained by the randomization test.

## RESULTS & DISCUSSION

### 2D QSAR approach

The statistically significant model (Model 1) generated by multiple regression analysis method with  $r^2 = 0.8350$  as coefficient of determination was considered. Model 1 explains 84% of variance in the observed activity values. The model showed an internal predictive power  $q^2=0.6147$  of 61% and predictivity for external test set  $\text{pred}_r^2 = 0.5958$  about 60%. Model 2 was generated by partial least square analysis with four significant parameters with  $r^2 = 0.8349$  as coefficient of determination. It is capable of explaining 83% of variance in the observed activity values. The model showed an internal predictive power  $q^2=0.6137$  of 61% and predictivity for external test set  $\text{pred}_r^2 = 0.5907$  about 59%. Model 3 a bi-parametric model generated with principle component analysis with same distribution of test and training set. It confirmed that two alignment independent descriptors T\_N\_N\_3 and T\_2\_C\_4 play pivotal role in determining activity.

The reported models Model 4, 5, and 6 were found to be more significant in terms of statistics then the models obtained by SE algorithm. The statistical four parametric Model 4 with multiple regression analysis method with coefficient of determination  $r^2 = 0.8393$  is capable of explaining 84% of variance in the observed activity values. The contribution chart for Model 4 for the four descriptors discussed is shown in Figure 2. The model showed an internal predictive power ( $q^2=0.7436$ ) of 74% and predictivity for external test set ( $\text{pred}_r^2 = 0.7826$ ) about 78%. The graph for experimental vs.

predicted activity for Model 4 is shown in Figure 3.

Model 5 is a four parametric model generated with partial least square analysis method with coefficient of determination  $r^2 = 0.8317$  is capable of explaining 83% of variance in the observed activity values. The model showed an internal predictive power ( $q^2=0.7420$ ) of 74% and predictivity for external test set ( $\text{pred}_r^2 = 0.7578$ ) about 76%. The contribution chart and graph for experimental versus predicted activity for Model 5 is shown in Figure 4 and Figure 5 respectively. Model 6 is generated using principle component analysis with same distribution of test and training set. The model confirmed the role of three descriptors which contributed in determining the biological activity. The contribution chart and graph for experimental vs. predicted activity is shown in Figure 6 and Figure 7 respectively.

The statistically significant Model 4 has shown a positive correlation with  $\text{chi3Cluster}$ ,  $T\_T\_N\_7$  and a negative correlation with  $T\_N\_O\_5$  and  $\text{HDonor}$  count. As a positively contribution descriptor (41%),  $T\_T\_N\_7$  is an alignment independent descriptor influencing the activity. This descriptor means the count of fragments having any heavy atom and a nitrogen atom separated by seven bond distance. The descriptor indicates that any substituent at 6<sup>th</sup> position of benzimidazole should possess nitrogen atom. In the same model, descriptor  $\text{chi3Cluster}$  is directly proportional to activity, indicating the importance of fused rings or branched molecules are important in determining biological activity. This descriptor signifies simple 3<sup>rd</sup> order cluster  $\text{chi}$  index in a compound. The descriptors  $T\_N\_O\_5$  was inversely proportional to the activity. This is the count of number of oxygen atoms (single, double or triple bonded) separated from oxygen atom by 5 bond distance in a molecule (N-C-C-C-C-O). In the same model H Donor count that is the

number of hydrogen atoms is inversely proportional to activity and thus adding hydrogen atoms are not conducive for the activity. The proposed models provide significant interpretations due to the fact that they provide an insight into role of various moieties featuring interactions in determining activity, secondly both internal and external validation procedures were used for the validation of the model significance, lastly both the models have improved correlation coefficient  $r^2$  and cross validation  $q^2$  values along with a significant predictive ability ( $\text{pred}_r^2$ ). The data for predicted activity for all the 2D models is reported in Table 4.

### 3D kNN Molecular field approach

In 3D QSAR analysis the descriptors that get selected in a given model were the field points either of steric or electrostatic nature at particular locations in a common grid around reported set of molecules. For utilizing these descriptors for new ligand design, the field values at different grid points of compounds cluster having most active compound were considered (Figure 8). The extrema of field values of compounds in the cluster of most active compounds decide range of field values which is preferred and recommended for new compound design. Activity distribution plot for 3D QSAR Model 7, 8, 9 is shown in Figure 9. In the analysis, Model 7 the triparametric model obtained by SWFB as shown in Figure 10 indicates that three steric interactions are contributing to activity. The descriptor  $S\_566$  (21.7998, 30.0000) has shown positive range inferring that analogs with large aromatic and aliphatic substituents will result in better biological activity at 6<sup>th</sup> position. However the descriptor  $S\_1321$  (-0.3085 to -0.1401) suggests bulky groups can be detrimental to activity at 4<sup>th</sup> position of benzimidazole. The descriptor  $S\_1958$  (30.0000 to 30.0000) exhibiting positive range implying that bulkier groups like  $-C_2H_5$ ,  $C_3$ ,  $7H_7$ ,  $C_4H_9$



might result in improved biological activity. The graph of experimental vs. predicted for Model 7 is shown in Figure 11. The Model 8 was not found to be significant in terms of statistics and thus cannot be exploited in designing of novel analogs. Model 9 obtained from GA was found to have analogous contribution of descriptors as discussed in Model 10 in detail except for steric interaction which is exhibiting positive contribution as shown in Figure 12, S\_1088 (2.7412 to 4.0075) where bulky aromatic and aliphatic systems are more favorable for activity at 6<sup>th</sup> position. The graph for experimental vs. predicted activity is shown in Figure 13. In the study the kNN-MFA Model 10, 11, 12 obtained by using the three variable selection methods, stepwise forward backward variable selection (SWFB), simulated annealing (SA), genetic algorithm (GA) shows steric and electrostatic interactions. In Model 10 three descriptors in kNN SWFB method plays major role in determining activity. The activity distribution plot for Model 10, 11, 12 is given in Figure 14. The model implies the significant role of the electrostatic field interaction for structure activity relationship.

Statistically kNN-SWFB is better as compared to SA and GA with respect to both the internal ( $q^2 = 0.77$ ) and external ( $\text{pred}_r^2 = 0.75$ ) model validation and correctly predicts ~ 77% and ~ 75% for the training and test set respectively in Model 10, 11, 12. It uses two electrostatic field descriptors E\_929, E\_1786 and one steric descriptor S\_1319 with its nearest neighbor ( $k=5$ ) to evaluate the activity of new molecule. Model 10 for kNN SWFB algorithm the plot of the kNN-MFA shows the relative position and ranges of the electrostatic and steric fields in the model provide guidelines for new molecule design as shown in Figure 15. The descriptor E\_929 (-10.0000, -4.1791) suggest that electronegative groups are preferred in the region (6<sup>th</sup> position of benzimidazole) where hetero-aromatic moiety with electronegative

groups or atoms like (electronegative atoms like N, N in benzimidazole, imidazole, imidazo pyridazine, N, O in benzoxazoles, N, S in benzthiazoles) might result in conducive activity. However, the descriptor S\_1319 (-0.1381 to -0.1292) in the region (4<sup>th</sup> position of benzimidazole) suggest that less bulkier groups imparting less steric hindrance are favored to give appreciable biological activity. Unsubstituted or less bulky groups like H, CH<sub>3</sub> at the benzimidazole are preferred at this position. The electrostatic descriptor E\_1786 (-10.0000 to 0.2402) at the terminal biphenyl moiety again suggest that the electronegative atoms at the proximal phenyl moiety like carboxylic and tetrazole moiety will always be conducive to activity. This triparametric model can assist in designing of novel analogs with improved biological activity. The graph of experimental versus predicted activity is shown in Figure 16.

The triparametric Model 12 obtained by GA is in parity with Model 7 obtained from SWFB with dissimilarity value 9.0. It indicates that three steric interactions at various points of substitution play major role in determining biological activity implying the major role of steric field interaction for illustrating structure activity relationship as shown in Figure 17. The external validation is not considerably high in this mathematical model. The descriptor S\_929 (30.0000 to 30.0000) exhibiting positive range analogously suggest as in S\_566 that analogs with bulky aromatic and aliphatic substituents will enhance biological activity at 6<sup>th</sup> position in benzimidazole. However the descriptor S\_932 (-0.0044 to -0.0041), suggests that the groups at 4<sup>th</sup> position should not possess bulky groups as they can be detrimental to activity. The steric parameter is in full compliance with the steric interaction obtained in the previous SWFB model. The descriptor S\_1945 (-0.4802 to -0.4698) exhibiting negative range implying that less

bulkier groups might result in improved biological activity. The graph of experimental vs. predicted is shown in Figure 18. The data for predicted activity for all 3D the models is reported in Table 4.

## CONCLUSION

A QSAR study on 40 derivatives, Angiotensin antagonists for AT1 receptor acting as antihypertensive agents is described in this research. The model validation was done by creating different combination of test and training set by SE algorithm and manual method of selection. Amongst all the six models obtained in 2D QSAR analysis, Model 4 obtained by MLR was statistically significant in 6 substituted benzimidazoles. The model was investigated for reliability and stability by using statistical analysis criteria. The resultant Models 4 and Model 5 ( $r^2 = 0.84$  and  $r^2 = 0.83$ ) suggest importance of hydrogen bonding, electrostatic interaction, and the shape/size of the molecules in determining the activities in 2D QSAR analysis. Since PLS approach utilizes simultaneously information of available dependent response (i.e., activity) and independent variables (molecular descriptors) while building the QSAR model, it offers a significant advantage to compare behavior of common subset descriptors toward individual activity. Similarly In 3D QSAR analysis, advanced kNN MFA Model 10 Step Wise Forward Back Ward kNN Method ( $q^2 = 0.77$ ) yielded a triparametric model with contributing descriptors E\_929, S\_1319 and E\_1786 and Model 12 ( $q^2 = 0.67$ ) provided confidence in robustness of descriptors toward data distribution. The developed models have consistent contribution (percentage) of descriptors toward activities and statistical significance with different training and test sets (Manual method and SE) which could be immensely useful in exploring and understanding the structural requirements for the design of novel

angiotensin II receptor antagonists as antihypertensive agents. QSAR model validation becomes an essential part in the development of a statistically valid and predictive model, because the real utility of a QSAR model was to design and predict accurately the modeled properties of the newly synthesized compounds as antihypertensive agents.

## ACKNOWLEDGEMENT

The author Anupama Parate is also thankful to University Grants Commission, New Delhi for providing financial assistance for the project. The author also wishes to express gratitude to V-life Science Technologies Pvt. Ltd for providing the software for the study, and Head, School of Pharmacy, Devi Ahilya Vishwavidyalaya for providing facilities to carry out the work. The authors also thank the anonymous reviewers whose valuable comments and suggestions greatly helped in improving the manuscript.

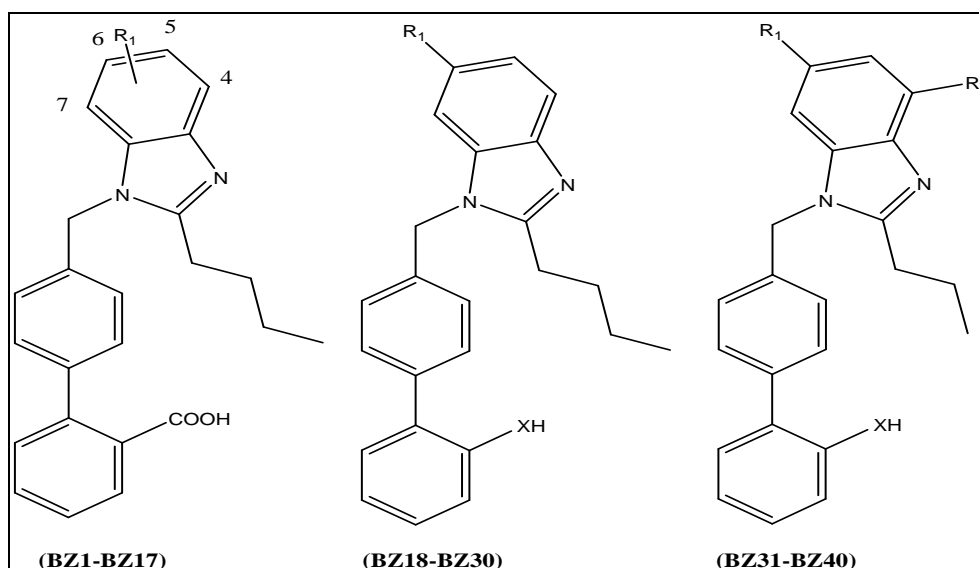
## REFERENCES

1. Le MT, Pugsley MK, Vauquelin G, Liefde IV (2007) British Journal of Pharmacology 151:952–962.
2. Burnier M, Brunner HR (2000) Angiotensin II receptor antagonists. Lancet 355:637-45.
3. Gasparo M de, Catt KJ, Inagami T, Wright JW, Unger T (2000) International Union of Pharmacology, XXIII The Angiotensin II Receptors. Pharmacol Rev 52:415-72.
4. Miura S, Saku K, Karnik SS (2003) Molecular analysis of the structure and function of the angiotensin II type 1 receptor. Hypertens Res 26:937–943.
5. Wong PC, Price Jr WA, Chiu AT, Duncia JV, Carini DJ, Wexler RR, Johnson AL, Timmermans PB (1996) Nonpeptide angiotensin II receptor antagonists XI Pharmacology of EXP3174: an active metabolite of DuP 753, an orally active antihypertensive agent. J Pharmacol Exp Ther 255:211-217.

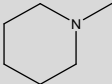
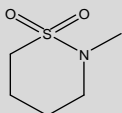
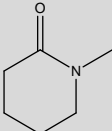
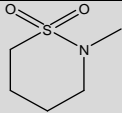
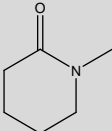
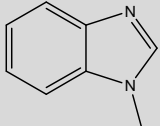
6. Buhl Mayer P, Furet P, Criscione L, Gaspro M de, Schimidilin TS, R Lattman (1994) Valsartan a Potent Orally Active Angiotensin II Antagonist Developed from a structurally new Amino Acid Series. *Biorg Med Chem Lett* 4:29-34.
7. Bernhart CA, Perreaut PM, Ferrari BP, Muneaux YA, Assens JL, Clement J, Haudricourt F, Muneaux CF, Taillades JE, Vignal MA (1993) A new series of imidazolones: highly specific and potent nonpeptide AT<sub>1</sub> angiotensin II receptor antagonists. *J Med Chem* 36:3371-80.
8. Ries UJ, Mihm G, Narr B, Hasselbach KM, Wittneben H, Entzeroth M, Meel JC, W Wiene, NH Huel (1993) 6-Substituted benzimidazoles as new nonpeptide angiotensin II receptor antagonists: Synthesis, biological activity, and structure-activity relationships. *J Med Chem* 36:4040-51.
9. Yanagisawa H, Amemiya Y, Kanazaki T, Shimoji Y, Fujimoto K, Kitahara Y, Sada T, Mizuno M, Ikeda M, Miyamoto S, Furukawa Y, Koike H (1996) Nonpeptide angiotensin II receptor antagonists: synthesis, biological activities, and structure-activity relationships of imidazole-5- carboxylic acids bearing alkyl, alkenyl, and hydroxyalkyl substituents at the 4- position and their related compounds. *J Med Chem* 39:323-38.
10. Ellingboe JW, Antane M, Nguyen TT, Collini MD, Antane S, Bender R, Hartupee D, White V, McCallum J, Park CH (1994) Pyrido[2,3-d]pyrimidine angiotensin II antagonists. *J Med Chem* 37:542-50.
11. Keenan RM, Weinstock J, Finkelstein JA, Franz RG, Gaitanopoulos DE, Girard GR, Hill DT, Morgan TM, Samanen JM, Peishoff CE (1993) Potent non-peptide angiotensin II receptor antagonists 2 1-(Carboxybenzyl)imidazole-5-acrylic acids. *J Med Chem* 36:1880-1892.
12. Middlemiss D, Drew GM, Ross BC, Robertson MJ, Scopes DIC, Dowle MD, Akers J, Cardwell K, Clark KL, Coote S, Idred CD, Hamblett J, Hilditch A, Hirst GC, Jack T, Motana J, Panchal TA, Paton JMS, Shah P, GT Stuart (1991) Benzobromofurans: A new class of potent non peptide antagonists of angiotensin II. *Bioorg Med Chem Lett*, 1:711-716.
13. Judd DB, Dowle MD, Middlemiss D, Scopes DI, Ross BC, Jack TI, Pass M, Tranquillini E, Hobson JE, Panchal TA (1994) Bromobenzofuran-based non-peptide antagonists of angiotensin II: GR138950, a potent antihypertensive agent with high oral bioavailability. *J Med Chem* 37:3108-3120.
14. K Kubo, Y Kohara, Y Yoshimura, Y Inada, Y Shibouta, Y Furukawa, T Kato, K Nishikawa, and T Naka (1993) Nonpeptide angiotensin II receptor antagonists Synthesis and biological activity of potential prodrugs of benzimidazole -7- carboxylic acids. *J Med Chem*, 36:2343-2349.
15. A Salimbeni, R Canevotti, F Paleari, D Poma, S Caliarì, F Fici, R Cirillo, A R Renzetti, A Subissi, L Belvisi (1995) N-3- substituted pyrimidinones as potent, orally active, AT<sub>1</sub> selective angiotensin II receptor antagonists. *J Med Chem* 38:4806-4820.
16. A Parate and SC Chaturvedi (2010) Structural insights for 3H-1, -2, -4 triazolinones as angiotensin II receptor antagonists using QSAR techniques *Med Chem Res* 19:375–391.
17. PZ Yan, BC Chen, H Shi, HJ Jian, LW Hai, LS Guo, and QY Ru (2007) QSAR study of angiotensin II antagonists using robust boosting partial least squares regression. *Ana Chima Acta*, 593:68-74.
18. N Balasubramanian, A Dhake, and VK Mourya (2007) QSAR Studies of 4,5-dihydro-4-oxo- 3H-imidazo[4,5-c]pyridines as potent angiotensin II receptor antagonists by MLR and NLR analysis. *Arkivoc* 1:189-204.
19. Pandya T, Chaturvedi SC (2005) Structure-activity relationship study of some triazolinones based compounds with antagonistic balanced activity on angiotensin II receptor subtypes AT<sub>1</sub> and AT<sub>2</sub> A three-dimensional quantitative structure-activity relationship investigation. *Arzneimittel-Forschung*, 55:265-70.
20. Datar PA, Desai PV, Coutinho EC (2004) A 3D-QSAR of Angiotensin II (AT<sub>1</sub>) Receptor Antagonists Based on Receptor Surface Analysis. *J Chem Inf Mod* 44:210-220.

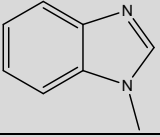
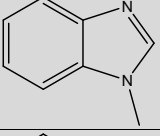
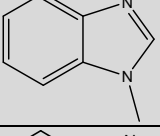
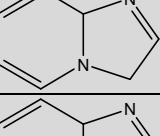
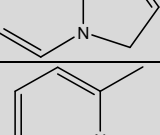
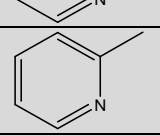
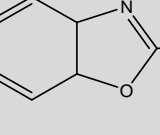
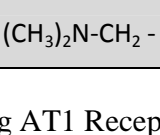
21. Pandya T, Chaturvedi SC (2004) QSAR study of a series of 2, 3, 6-substituted quinazolinones as AT1 selective angiotensin II receptor antagonists. *Ind J Chem*, 43:2440-2445.
22. Datar PA, Coutinho EC, Srivastava S (2004) Comparative Study of 3D-QSAR Techniques on Angiotensin II Receptor (AT<sub>1</sub>) Antagonists. *Lett in Drug Design Disc* 1:115-120.
23. Pandya T, Pandey SK, Tiwari M, Chaturvedi SC, Saxena AK (2001) 3-D QSAR studies of triazolinones based balanced AT1/AT2 receptor antagonists. *Bioorg Med Chem*, 9:291-300
24. Kurup A, Garg R, Carini DJ, Hansch C (2001) Comparative QSAR: angiotensin II antagonists. *Chem Rev*, 101:2727.
25. Yoo SE, Kim SK, Lee SH, Yi KY, Lee DW (1999) A comparative molecular field analysis and molecular modeling studies on pyridyl imidazole type of angiotensin II antagonists. *Bioorg Med Chem* 7:2971-2976.
26. Belvisi L, Bravi G, Catalano G, Mabiliab M, Salimbeni A, Scolastico C (1996) A 3D QSAR CoMFA study of non-peptide angiotensin II receptor antagonists. *J Comp Aid Mol Des*, 10:567-582.
27. Jain A, Chaturvedi SC (2008) Rationalization of Physicochemical Property of Some Substituted Benzimidazole Bearing Acidic Heterocyclic Towards Angiotensin II Antagonist: A QSAR Approach. *Asian J Biochem* 5:330-336.
28. Parate A, Chaturvedi SC, Predicting 3H- 1, 2, 4 Triazolinones as Angiotensin II Receptor Antagonists: 2D & 3D QSAR By kNN -Molecular Field Analysis Approach *Med Chem Res* DOI: 101007/s00044-011-9622-4.
29. Vlife MDS software package, version 30, supplied by Vlife science technologies Pvt Ltd, 1, Akshay 50, Anand Park, Aundh, Pune, India 411007.
30. Baumann K (2000) An Alignment-Independent Versatile Structure Descriptor for QSAR and QSPR Based on the Distribution of Molecular Features. *J Chem Inf Comput Sci* 42:26-35.
31. Ajmani S, Jhadav K, Kulkarni SA (2006) Three-Dimensional QSAR Using the k-Nearest Neighbor Method and Its Interpretation. *J Chem Inf Mod* 46:24–31.
32. Rogers D, Hopfinger AJ (1994) Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. *J Chem Inf Comp Sci* 34:854-866.
33. Shen M, Tiran AL, Xiao Y, Golbraikh A, Kohn H, Tropsha H (2002) QSAR Analysis of Functionalized Amino Acid Anticonvulsant Agents Using k-Nearest Neighbor and Simulated Annealing-PLS Methods. *J Med Chem* 45:2811-2823.

**Table 1. AT<sub>1</sub> Receptor Binding of Substituted Benzimidazoles at the Phenylene Ring Substituted with Alkylamino and Acylamino Residues in and Substituted with Nitrogen-Containing Heterocycles**



C. No.	R <sub>1</sub>	R <sub>2</sub>	XH	IC <sub>50</sub> nM <sup>a</sup>	Actual Log (1/IC <sub>50</sub> nM)
BZ1	H	-	-	400	6.397940
BZ2	4-CH <sub>3</sub>	-	-	1200	5.920819
BZ3	5-CH <sub>3</sub>	-	-	1200	5.920819
BZ4	6-CH <sub>3</sub>	-	-	850	6.070581
BZ5	7-CH <sub>3</sub>	-	-	480	6.318759
BZ6	4-NH <sub>2</sub>	-	-	1700	5.769551
BZ7	5-NH <sub>2</sub>	-	-	820	6.086186
BZ8	6-NH <sub>2</sub>	-	-	540	6.267606
BZ9	7-NH <sub>2</sub>	-	-	1060	5.974694
BZ10	4-NHCOCH <sub>3</sub>	-	-	5700	5.244125
BZ11	5-NHCOCH <sub>3</sub>	-	-	460	6.337242
BZ12	6-NHCOCH <sub>3</sub>	-	-	180	6.744727
BZ13	7-NHCOCH <sub>3</sub>	-	-	1800	5.744727

<b>BZ14</b>	4-NHCONHC <sub>6</sub> H <sub>11</sub>	-	-	29300	<b>4.533132</b>
<b>BZ15</b>	5-NHCONHC <sub>6</sub> H <sub>11</sub>	-	-	800	<b>6.096910</b>
<b>BZ16</b>	6-NHCONHC <sub>6</sub> H <sub>11</sub>	-	-	26	<b>7.585027</b>
<b>BZ17</b>	7-NHCONHC <sub>6</sub> H <sub>11</sub>	-	-	160	<b>6.795880</b>
<b>BZ18</b>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>4</sub> NH	-	-COOH	390	<b>6.408935</b>
<b>BZ19</b>			-COOH	160	<b>6.795880</b>
<b>BZ20</b>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub> CONH-		-COOH	86	<b>7.065502</b>
<b>BZ21</b>	(CH <sub>3</sub> ) <sub>2</sub> NCONH-		-COOH	24	<b>7.619789</b>
<b>BZ22</b>	CH <sub>3</sub> C <sub>6</sub> H <sub>11</sub> NHCO N-		-COOH	26	<b>7.585027</b>
<b>BZ23</b>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub> SO <sub>2</sub> N- CH <sub>3</sub>		-COOH	33	<b>7.481486</b>
<b>BZ24</b>			-COOH	34	<b>7.468521</b>
<b>BZ25</b>			-COOH	81	<b>7.091515</b>
<b>BZ26</b>	C <sub>6</sub> H <sub>11</sub> NHCONH		-tetrazole	21	<b>7.677781</b>
<b>BZ27</b>	CH <sub>3</sub> C <sub>6</sub> H <sub>11</sub> NHCON -		-tetrazole	10	<b>8.000000</b>
<b>BZ28</b>	(CH <sub>3</sub> ) <sub>2</sub> NCONH-		-tetrazole	8	<b>8.096910</b>
<b>BZ29</b>			-tetrazole	3	<b>8.522879</b>
<b>BZ30</b>			-tetrazole	4	<b>8.397940</b>
<b>BZ31</b>		CH <sub>3</sub>	-COOH	3	<b>8.522879</b>

<b>BZ32</b>		H	-COOH	3	<b>8.522879</b>
<b>BZ33</b>		CH <sub>3</sub>	tetrazole	13	<b>7.886057</b>
<b>BZ34</b>		H	tetrazole	5	<b>8.301030</b>
<b>BZ35</b>		CH <sub>3</sub>	-COOH	4	<b>8.397940</b>
<b>BZ36</b>		CH <sub>3</sub>	tetrazole	3	<b>8.522879</b>
<b>BZ37</b>		CH <sub>3</sub>	tetrazole	5	<b>8.301030</b>
<b>BZ38</b>		H	tetrazole	11	<b>7.958607</b>
<b>BZ39</b>		H	tetrazole	240	<b>6.619789</b>
<b>BZ40</b>	(CH <sub>3</sub> ) <sub>2</sub> N-CH <sub>2</sub> -	CH <sub>3</sub>	-COOH	158	<b>6.801343</b>

a *In vitro* binding AT1 Receptor Binding of Substituted Benzimidazoles.

**Table 2. Unicolumn statistics of the training and test sets for Model 1, 2, 3**

Name Column	Average	Max	Min	Std. Dev	Sum
<b>For Model 1, 2, 3</b>					
<b>Column<sup>a</sup></b>	7.3511	8.5229	4.5331	1.0327	220.5332
<b>Column<sup>b</sup></b>	6.1322	6.7959	5.7447	0.3245	61.3221
<b>For Model 4, 5, 6</b>					
<b>Column<sup>c</sup></b>	7.1303	8.5229	4.5331	1.0647	228.1686
<b>Column<sup>d</sup></b>	6.7108	8.5229	5.9208	0.9824	53.6867

a Unicolumn statistics for 2D model generated by sphere exclusion method with dissimilarity value 2.2 for training set where n=30

b Unicolumn statistics for 2D model generated by sphere exclusion method with dissimilarity value 2.2 for test set.

c Unicolumn statistics for the 2D model generated by manual method for training set

d Unicolumn statistics for 2D model generated by manual method for test set.

**Table 3. Statistical Analysis of Multiple Regression, Partial Least Square and Principle Component analysis**

Parameter	2D – Multiple Regression Analysis <sup>a</sup> Model 1	2D – Partial Least Square Analysis <sup>b</sup>	2D – Principle Component Analysis <sup>c</sup> Model 3	2D – Multiple Regression Analysis <sup>d</sup> Model 4	2D – Partial Least Square Analysis <sup>e</sup> Model 5	2D – Principle Component Analysis <sup>f</sup> Model 6
<b>Descriptor_1 (Coefficient)</b>	T_N_N_3 -1.8105	T_N_N_3 -1.8209	T_N_N_3 -1.7740	T_T_N_7 0.1438	T_T_N_7 0.1317	T_T_N_7 (0.1512)
<b>Descriptor_2 (Coefficient)</b>	T_2_C_4 0.0797	T_2_C_4 0.0795	T_2_C_4 0.0623	T_N_O_5 -1.1280	T_N_O_5 -1.2551	T_N_O_5 (-1.1383)
<b>Descriptor_3 (Coefficient)</b>	SaaOcount -1.5170	Slogp -0.3994	-	H-Donor Count	H-Donor Count	H-Donor Count
<b>Descriptor_4 (Coefficient)</b>	Slogp -0.3946	SaaOcount -1.4864	-	chi3Cluster 0.7855	chi3Cluster 0.6336	-
<b>Constant</b>	4.3229	4.3688	3.0049	4.8400	5.3782	
<b>Degree of freedom</b>	25	26	28	27	30	
<b>Optimum Components</b>	-	3	1	-	1	2



n (training/test)	30	30	30	32	32	32
$r^2$	0.8350	0.8349	0.7137	0.8393	0.8317	0.8140
$q^2$	0.6147	0.6137	0.6205	0.7436	0.7420	0.7496
F test	31.6186	43.8298	69.7855	35.2605	148.2101	63.4565
$r^2_{se}$	0.4519	0.4432	0.5624	0.4573	0.4440	0.4747
$q^2_{se}$	0.6904	0.6779	0.6474	0.5777	0.5497	0.5509
pred_ $r^2$	0.5958	0.5907	0.5994	0.7826	0.7578	0.7562
pred_ $r^2_{se}$	0.8425	0.8478	0.8388	0.5036	0.5315	0.5333
ZScore R <sup>2</sup>	6.48286	6.63543	17.60610	8.03616	10.16299	10.69093
ZScore Q <sup>2</sup>	5.68628	4.95703	12.18563	7.24712	7.67120	9.61975
Best Rand R <sup>2</sup>	0.55856	0.53391	0.14779	0.42231	0.39401	0.27600
Best Rand Q <sup>2</sup>	0.33302	0.39398	0.03964	0.20066	0.13737	0.13197
Alpha Rand R <sup>2</sup>	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Alpha Rand Q <sup>2</sup>	0.00000	0.00001	0.00000	0.00000	0.00000	0.00000
Z Score Pred R <sup>2</sup>	1.94044	1.92461	1.89192	1.40344	1.36147	1.72687
best Rand Pred R <sup>2</sup>	0.57464	0.79754	0.53357	0.76818	0.81602	0.67705
alpha Rand Pred R <sup>2</sup>	0.05000	0.05000	0.05000	0.10000	0.10000	0.05000

- a Statistical data for 2D model generated by multiple regression method with sphere exclusion method with dissimilarity value 2.2 for training set in Model 1
- b Statistical data for 2D model generated by partial least square method with sphere exclusion method with dissimilarity value 2.2 for training set in Model 2
- c Statistical data for 2D model generated by principle component method with sphere exclusion method with dissimilarity value 2.2 for training set in Model 3
- d Statistical data for 2D model generated by multiple regression method with manual method in Model 4
- e Statistical data for 2D model generated by partial least square method with manual method in Model 5
- f Statistical data for 2D model generated by principle component method with manual method in Model 6

Table 4. Predicted activity for Model 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Comp. No.	Actual log (1/IC50 nM)	Predicted MLR Model 1	Predicted PLS Model 2	Predicted PCA Model 3	Predicted MLR Model 4	Predicted PLS Model 5	Predicted PCA Model 6	Predicted kNN_SWFB Model 7	Predicted kNN_SA Model 8	Predicted kNN_GA Model 9	Predicted kNN_SWFB Model 10	Predicted kNN_SA Model 11	Predicted kNN_GA Model 12
BZ1	6.397940	6.579481**	6.585741**	6.553244**	6.047914**	6.271575**	6.350986**	6.598000	6.358720	6.837460	6.272630**	6.000250**	5.779570**
BZ2	5.920819	6.617161**	6.621524**	6.677747**	6.209044**	6.401543**	6.350986**	8.397940**	7.650820**	8.100240**	6.272800**	5.998710**	5.779570**
BZ3	5.920819	6.617161**	6.621524**	6.677747**	6.274682**	6.454487**	6.350986**	6.809260**	7.895740**	8.475870**	6.331180**	6.090670**	5.940080**
BZ4	6.070581	6.537476**	6.542047**	6.615495**	6.274682	6.454487	6.350986	6.707150	8.183920	7.097130	6.378800	5.704690	6.412020
BZ5	6.318759	6.696846	6.701001	6.739998	6.213453	6.405099	6.350986	6.223940**	6.673300**	5.976100**	6.236440	5.594190	5.186820
BZ6	5.769551	4.933819**	4.931743**	4.779264**	6.174345	6.292642	6.409962	6.380160**	6.411000**	5.954760**	6.258410	5.703090	5.375780
BZ7	6.086186	6.744329	6.752593	6.553244	6.096148	6.213860	6.258771	6.223950**	6.412460**	6.791340**	6.192440	5.640110	6.527330
BZ8	6.267606	6.744329**	6.752593**	6.553244**	6.239983**	6.345586**	6.409962**	6.120240	6.005170	6.901420	6.989830**	6.258020**	8.335680**
BZ9	5.974694	4.933819	4.931743	4.779264	6.178754**	6.296198**	6.409962**	6.179490	5.756210	5.806870	6.181510**	6.256810**	5.779230**
BZ10	5.244125	5.502551	5.496795	5.339526	5.312843	5.252497	5.271652	6.093910	5.669580	5.702620	4.533130	5.808660	5.555040
BZ11	6.337242	7.233376**	7.238168**	7.051255**	6.638098	6.682416	6.561153	4.533130	5.902940	5.816650	6.103700	5.590730	6.538280
BZ12	6.744727	7.233376	7.238168	7.051255	6.781933	6.814142	6.712344	6.399130	5.990910	6.812460	6.002820	6.605980	6.198680
BZ13	5.744727	5.422866**	5.417318**	5.277275**	5.317252	5.256053	5.271652	6.527200	6.366420	5.914420	6.269140	6.181740	5.529360
BZ14	4.533132	5.315591	5.300496	5.713034	5.632233	5.460622	5.633010	6.224320	6.569440	5.670100	5.244130	5.950780	5.787220
BZ15	6.096910	7.046416**	7.041870**	7.424762**	6.669818	6.627089	6.620129	5.244130	6.043770	6.043250	6.443490	6.494580	6.414430
BZ16	7.585027	7.046416	7.041870	7.424762	6.957488	6.890541	6.922511	6.393350	6.429790	7.103960	7.400280	8.072900	7.434230
BZ17	6.795880	5.235906**	5.221020**	5.650782**	5.636642	5.464178	5.633010	7.962600	7.879600	7.571330	6.172950	7.622320	6.265090
BZ18	6.408935	6.345782	6.343311	6.864501	6.461234	6.555464	6.712344	6.948450	6.360110	7.024770	6.118380	5.589970	6.315220
BZ19	6.795880	6.905914	6.905547	7.113506	7.172773**	7.273140**	7.258132**	6.090610	6.805230	5.663370	7.652940**	5.667130**	7.247630**
BZ20	7.065502	7.010678	7.009229	7.238008	6.831838	6.870103	6.863535	6.382450	6.561960	5.977650	7.169820	8.072900	7.627090
BZ21	7.619789	7.180071	7.184215	7.051255	7.627414	7.574651	7.468299	8.136070	7.805940	7.643200	7.820720	8.021430	7.921230
BZ22	7.585027	7.560014	7.555826	7.736019	7.072329	7.064084	6.863535	7.723420	8.015970	6.139420	7.034680	6.782070	7.060200
BZ23	7.481486	7.376205	7.376848	7.362511	7.316608	7.404866	7.409323	6.780250	7.481490	7.481490	7.284610	7.488520	7.210960
BZ24	7.468521	7.632637	7.634043	7.487014	7.422096	7.505661	7.560514	7.384830	7.585030	7.585030	8.005180	8.055600	7.724650
BZ25	7.091515	7.491125	7.491987	7.424762	7.477738	7.534834	7.409323	7.645220	7.865030	7.531640	7.572430	8.131660	7.642480

BZ26	7.677781	7.569578	7.566685	7.673768	7.689573	7.575292	7.829657	7.724510	7.802010	7.686180	7.944070	8.072900	7.902300
BZ27	8.000000	7.799046	7.795411	7.860521	8.235919	8.144014	8.224254	7.686060	8.004150	7.999540	7.319220	7.940020	7.803120
BZ28	8.096910	7.703233	7.709030	7.300260	8.359500	8.259403	8.375445	7.648160	7.799610	7.677430	7.493870	7.917880	7.769520
BZ29	8.522879	8.155799	8.158859	7.736019	8.154182	8.190412	8.467660	7.622700	7.776260	6.385790	8.223280	7.837810	7.948550
BZ30	8.397940	8.014287	8.016802	7.673768	8.209824	8.219586	8.316469	6.223950	5.924710	5.727190	8.331150	7.859370	8.337600
BZ31	8.522879	8.150179	8.143750	8.234029	8.130387	8.092680	7.711705	8.522880**	7.540110**	8.182080**	8.392740	7.903880	7.826840
BZ32	8.522879	8.032814	8.028490	8.047275	7.969256**	7.962711**	7.711705**	7.977080	6.444720	7.349920	7.956400**	7.868210**	8.097760**
BZ33	7.886057	8.673341	8.668565	8.483034	8.862472**	8.777431**	8.618851**	7.649270**	7.834110**	8.047840**	7.691870**	7.663110**	7.338320**
BZ34	8.301030	8.555976	8.553305	8.296280	8.701342	8.647463	8.618851	7.977080	6.740490	7.529970	7.732580	7.948250	7.856160
BZ35	8.397940	8.066840	8.061753	8.109527	8.027259	8.009496	7.711705	7.724640	7.877710	8.193500	8.519510	7.928870	7.737880
BZ36	8.522879	8.590002	8.586568	8.358532	8.759344	8.694247	8.618851	8.518510	7.539350	8.268330	8.240410	7.903880	7.627290
BZ37	8.301030	8.309550	8.308592	8.047275	7.922154	7.956134	8.014087	8.393350**	7.515090**	8.209940**	7.730010	7.644730	7.681920
BZ38	7.958607	8.192185	8.193332	7.860521	7.761023	7.826165	8.014087	7.360040	6.728810	7.080710	8.522650	7.715400	7.814100
BZ39	6.619789	6.619800	6.647341	8.047275	7.038738	6.914021	7.026968	7.286520	7.696770	7.607720	7.917670	8.284500	8.259440
BZ40	6.801343	7.125448	7.131283	6.926752	7.265432	7.316463	6.955750	6.159550	6.044880	5.718830	6.205960	6.947000	7.547810

\*\* Compounds in test set.

Table 5. Unicolumn statistics of the training and test sets for Model 7, 8, 9

Model 7, 8, 9					
Name Column	Average	Max	Min	Std. Dev.	Sum
Column <sup>a</sup>	7.0978	8.5229	4.5331	1.0296	227.1292
Column <sup>b</sup>	6.8408	8.5229	5.7696	1.1794	54.7261
Model 10, 11, 12					
Column <sup>c</sup>	7.1303	8.5229	4.5331	1.0647	228.1686
Column <sup>d</sup>	6.7108	8.5229	5.9208	0.9824	53.6867

*a* Unicolumn statistics for the 3D model generated by sphere exclusion method with dissimilarity value 9.0 for training set

*b* Unicolumn statistics for the 3D model generated by sphere exclusion method with dissimilarity value 9.0 for test set.

*c* Unicolumn statistics for 3D model generated by sphere exclusion method with dissimilarity value 9.5 for training set.

*d* Unicolumn statistics for 3D model generated by sphere exclusion method with dissimilarity value 9.5 for test set.

**Table 6. Statistical Analysis of kNN Stepwise forward backward variable selection Simulated Annealing and Genetic Algorithm models**

Parameter	3D kNN- MFA (Stepwise Forward Backward)	3D kNN- MFA (Simulated Annealing)	3D kNN- MFA (Genetic Algorithm GA) Model 9	3D kNN- MFA (Stepwise Forward Backward)	3D kNN- MFA (Simulated Annealing SA ) Model 11	3D kNN- MFA (Genetic Algorithm GA) Model 12
Descriptor_1 (coefficient)	S_566 21.7998 30.0000	S_750 -2.3935 - 0.2561	E_592 -0.4456 - 0.0181	E_929 -10.0000 - 4.1791	S_566 -0.0004 - 0.0003	S_929 30.0000 30.0000
Descriptor_2 (coefficient)	S_1321 -0.3085 - 0.1401	-	S_1088 2.7412 4.0075	S_1319 -0.1381 - 0.1292		S_932 -0.0044 - 0.0041
Descriptor_3 (coefficient)	S_1958 30.0000 30.0000	-	E_1789 -0.6973 - 0.3994	E_1786 -10.0000 0.2402		S_1945 -0.4802 - 0.4698
knn method	5	-	5	5	5	3
Degree of freedom	28	27	27	28	27	27
n (training/test)	32	32	32	32	32	32
$q^2$	0.8144	0.5681	0.4990	0.7754	0.5930	0.6744
$q^2_{se}$	0.4436	0.6766	0.7287	0.5046	0.6792	0.6075
pred_r <sup>2</sup>	0.8152	0.1079	0.6414	0.7592	0.7519	0.3382
pred_r <sup>2</sup> <sub>se</sub>	0.5206	1.1438	0.7252	0.5299	0.5379	0.8785

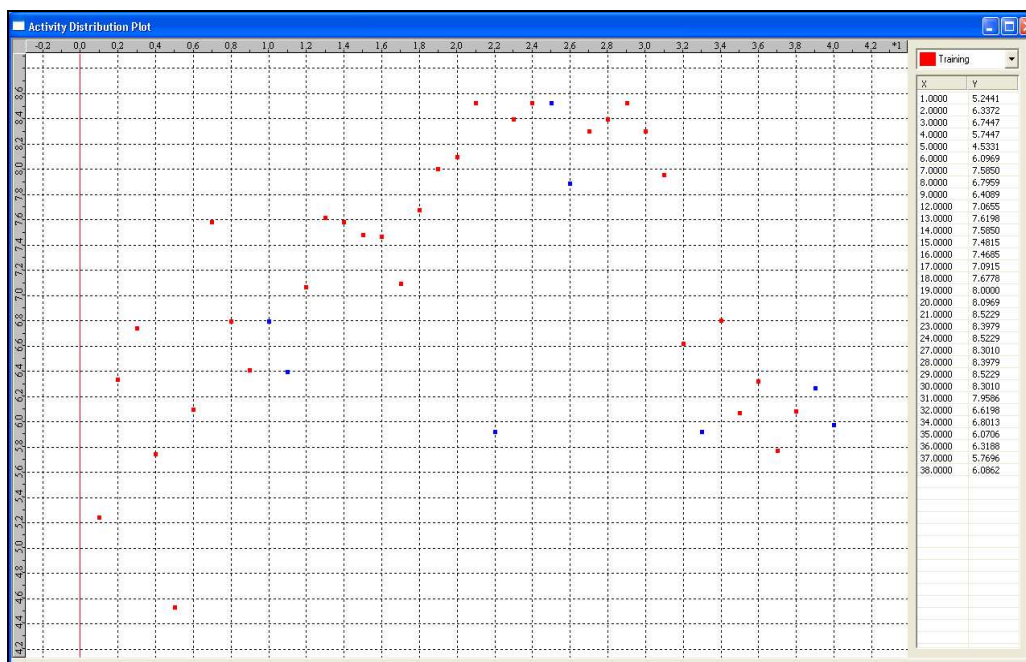


Figure.1. Activity distribution plot for 2D QSAR models

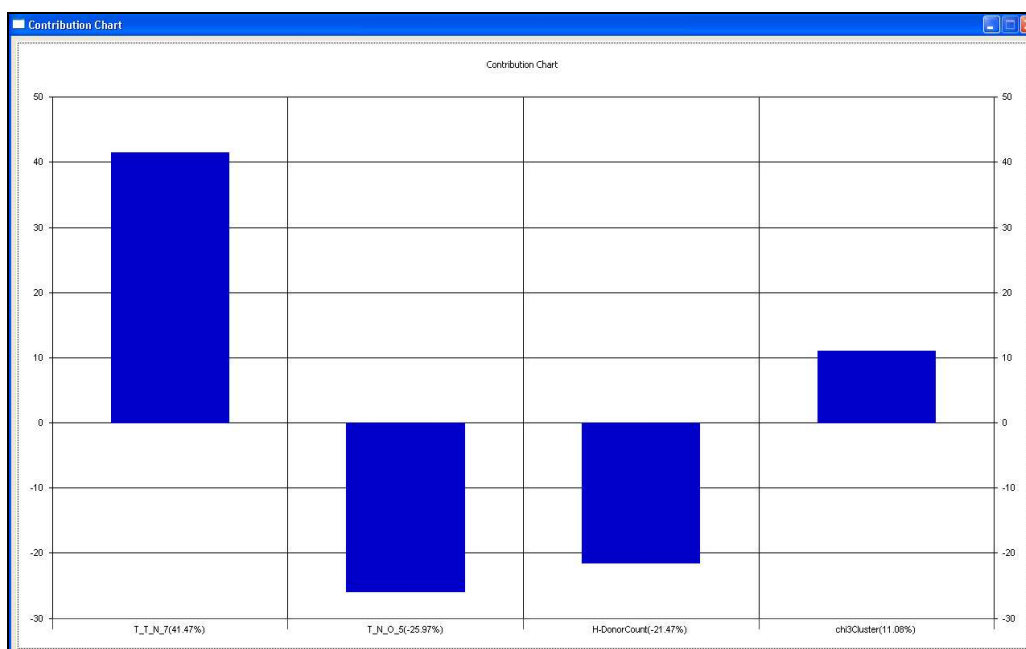


Figure.2. Contribution Chart for Model 4

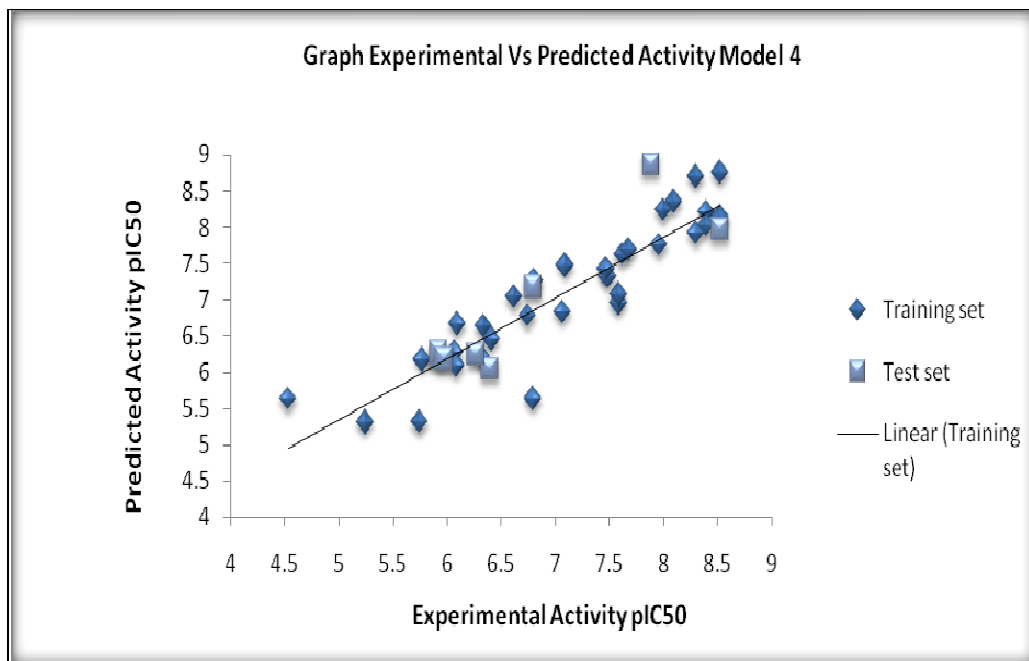


Figure.3. Graph experimental Vs predicted activity for Model 4

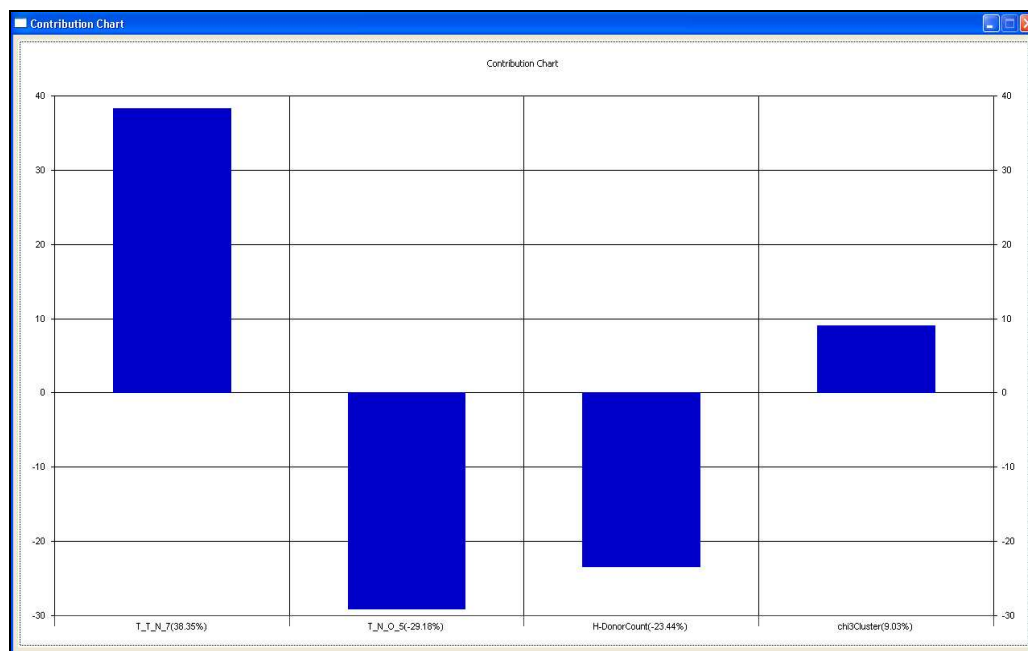


Figure.4. Contribution Chart for Model 5

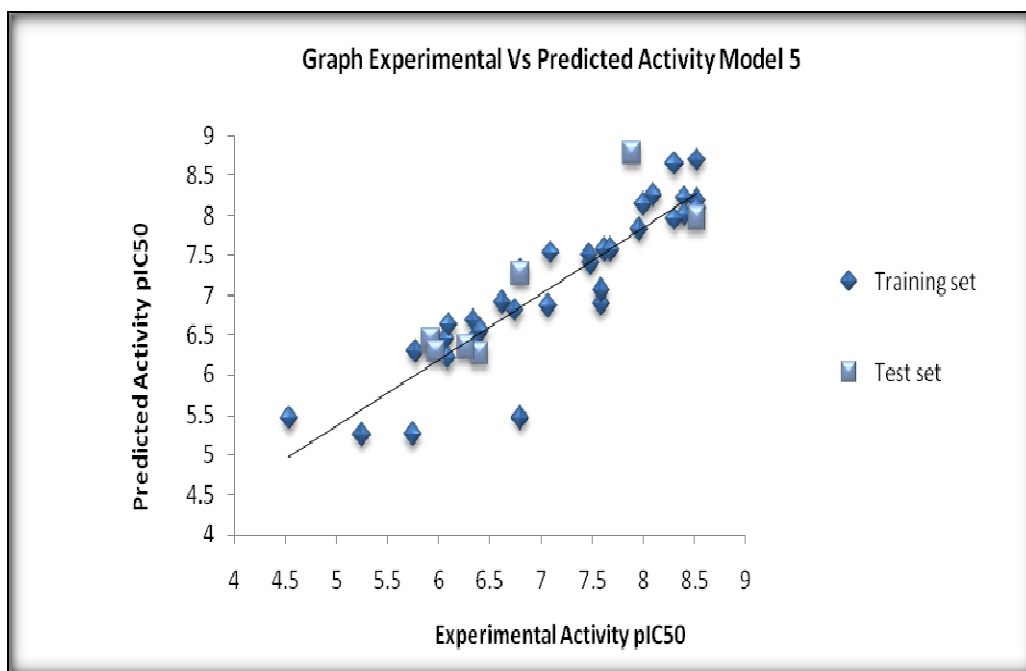


Figure.5. Graph experimental Vs predicted activity for Model 5

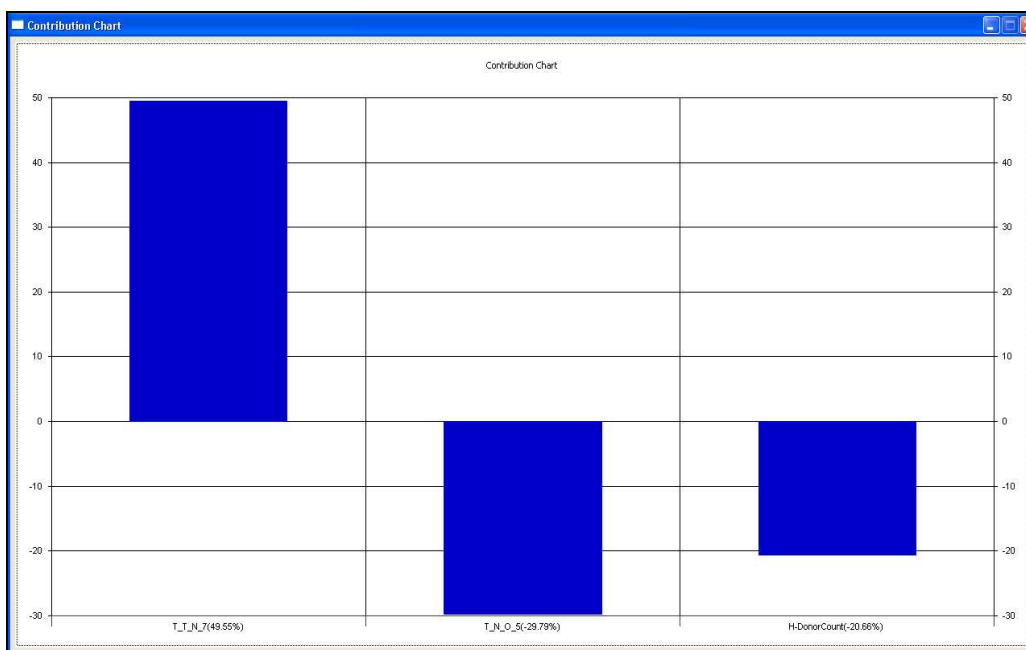


Figure.6. Contribution Chart for Model 6

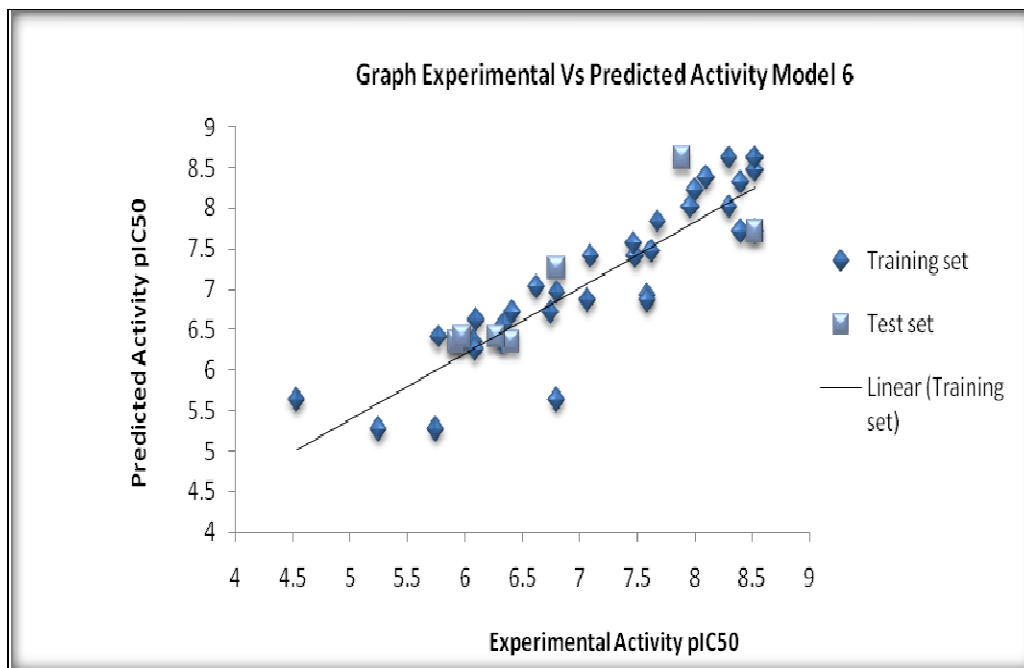


Figure.7. Graph experimental Vs predicted activity for Model 6

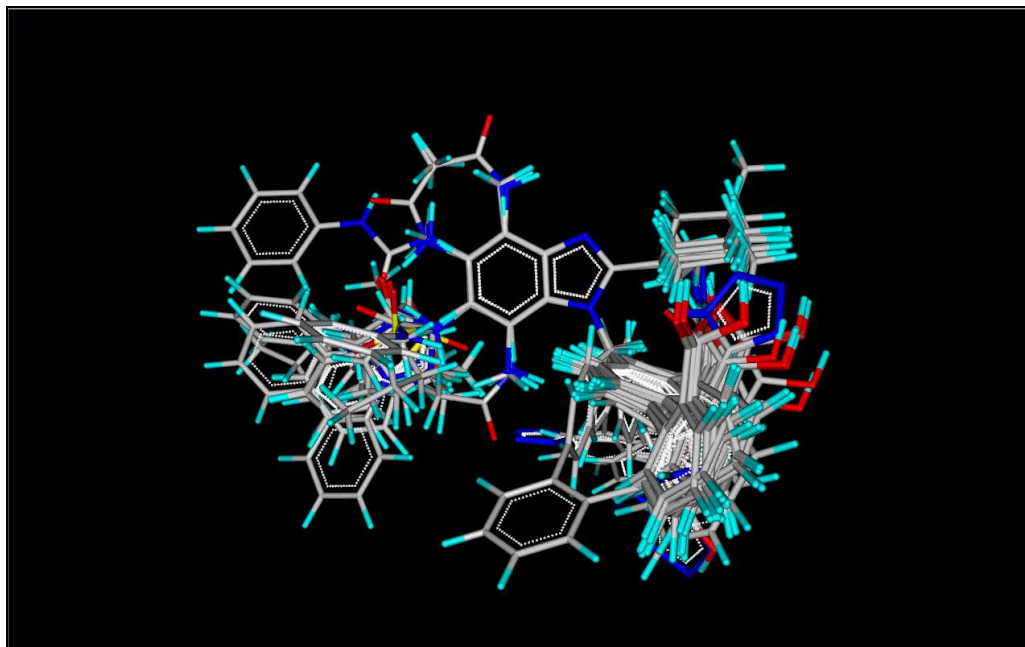


Figure.8. Template based alignment for 3D QSAR study



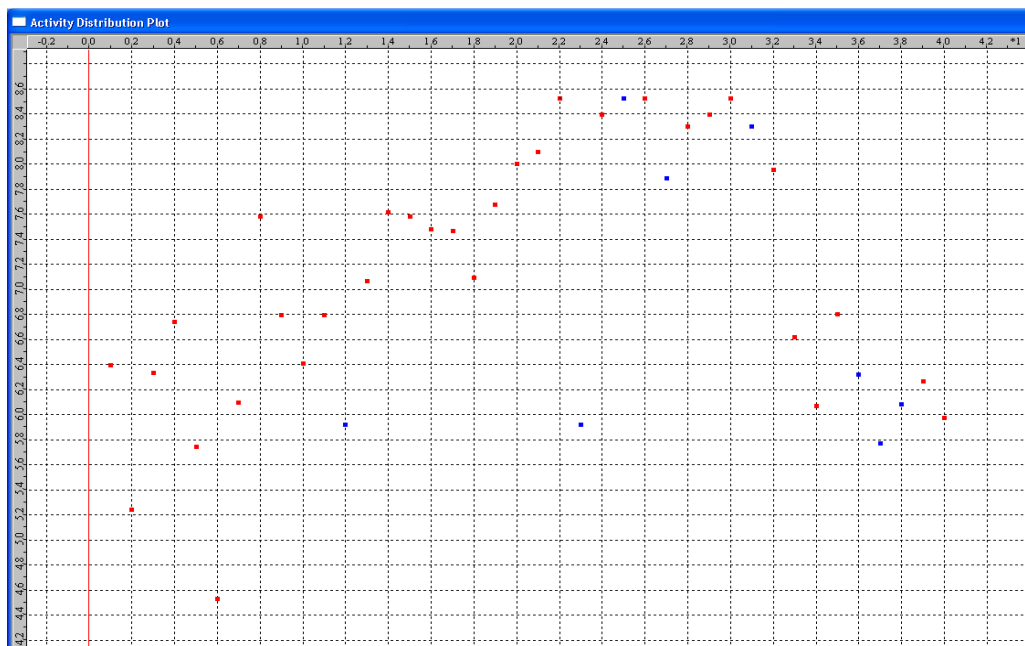


Figure.9. Activity distribution plot for 3D QSAR Model 7, 8, 9

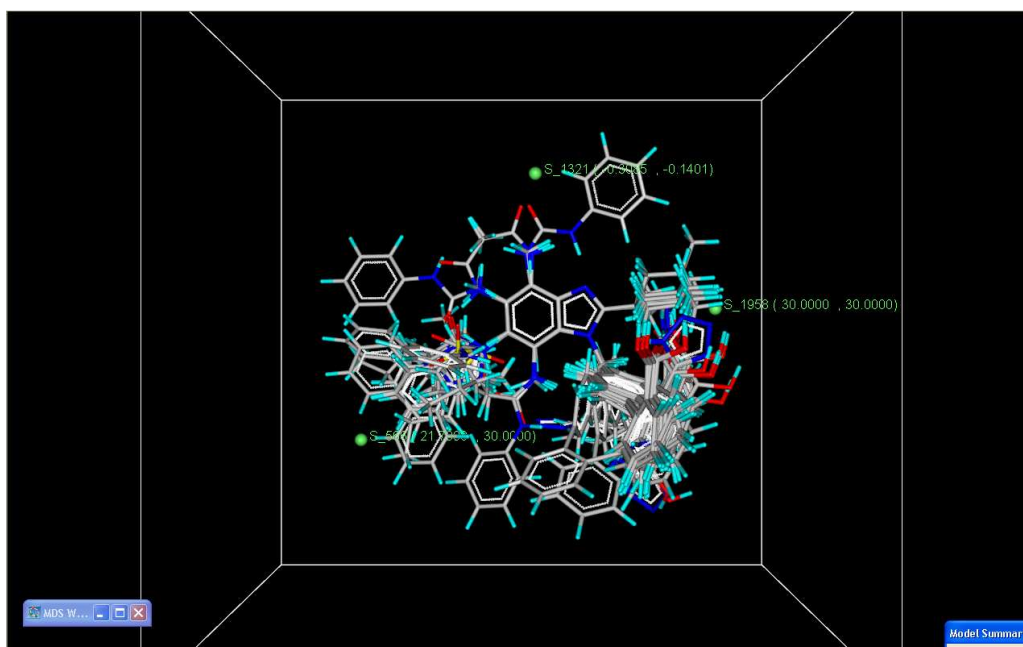


Figure.10. Plot for Model 7 for Stepwise forward backward kNN MFA analysis

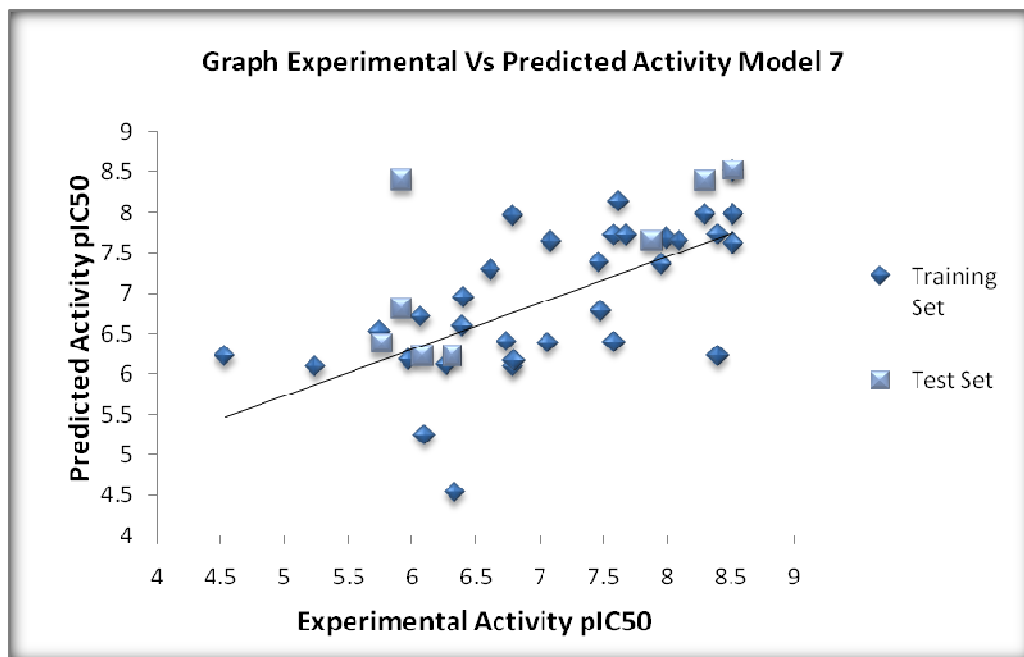


Figure.11. Graph experimental Vs predicted activity for Model 7

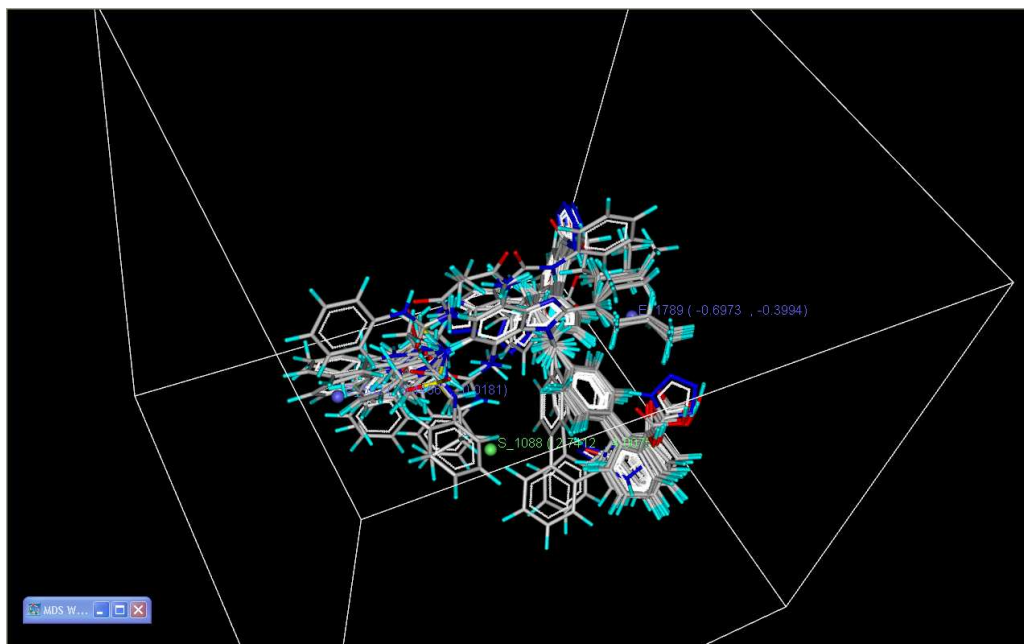


Figure.12. Plot for Model 9 for Genetic Algorithm kNN MFA analysis

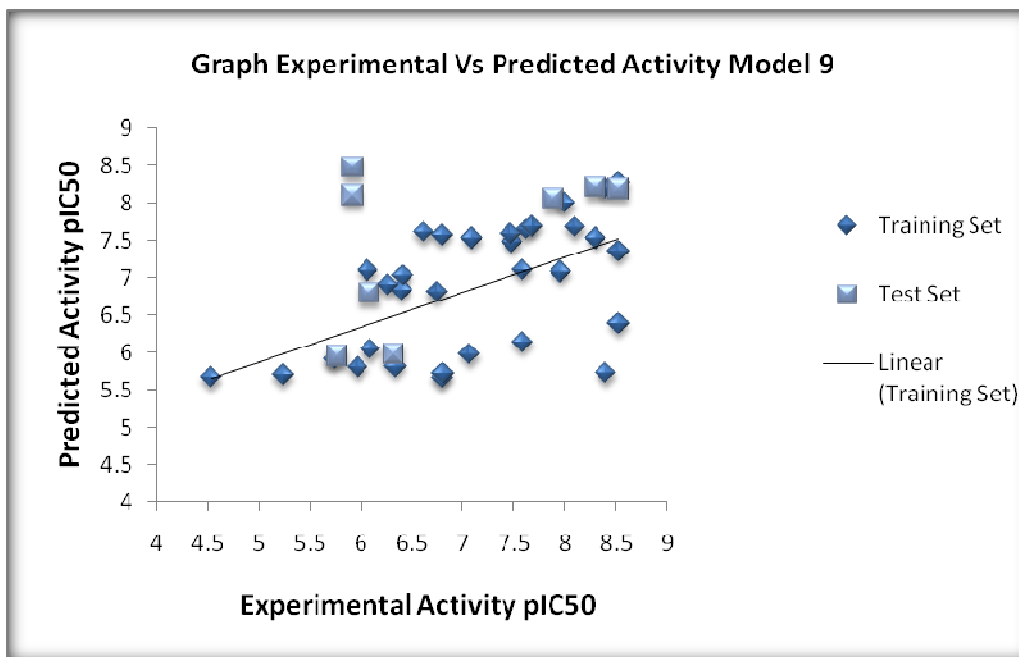


Figure.13. Graph experimental Vs predicted activity for Model 9

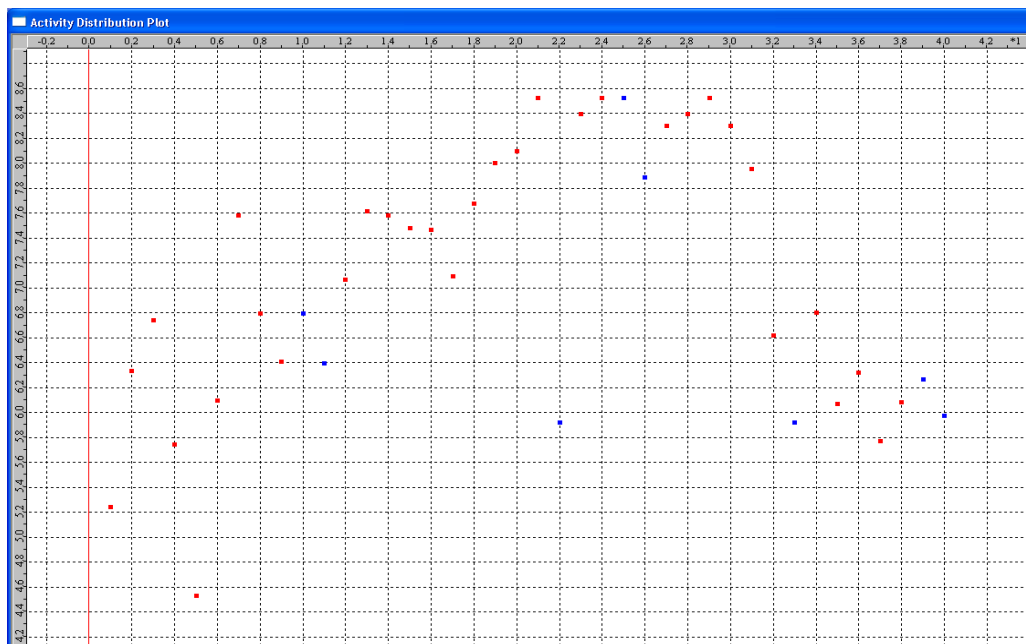


Figure.14. Activity Distribution plot for 3D QSAR study for Model 10, 11, 12.

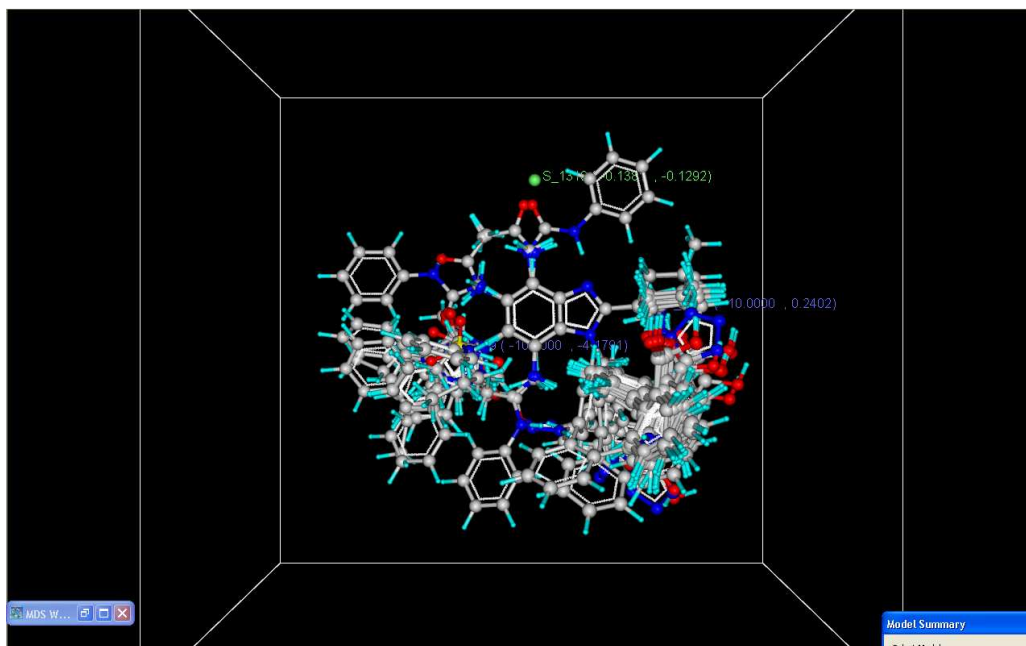


Figure.15. Plot for Model 10 for Stepwise forward backward kNN MFA analysis

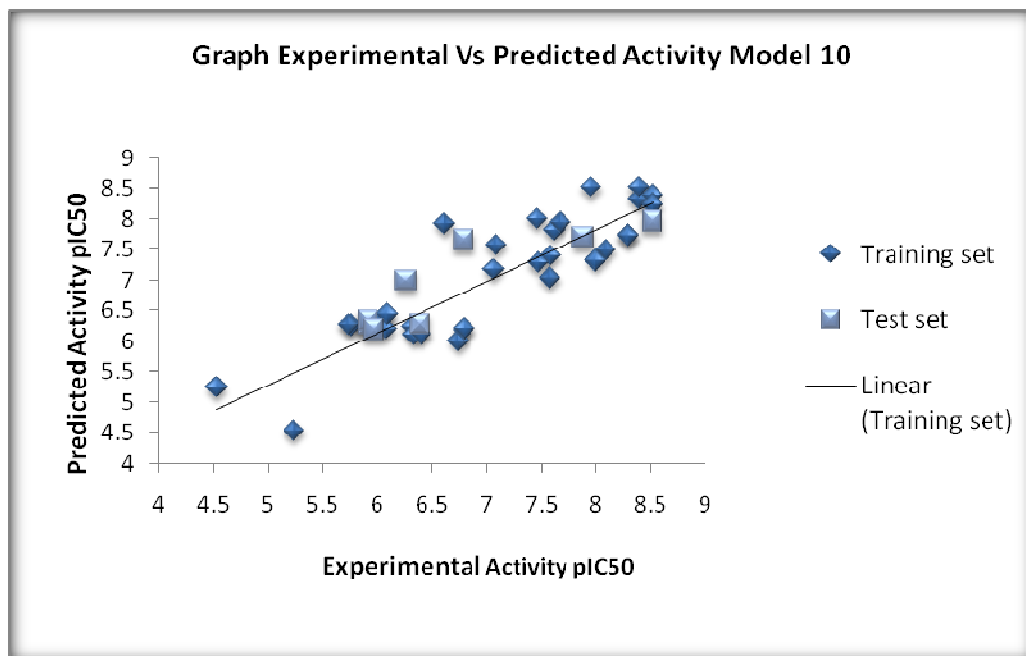


Figure.16. Graph experimental Vs predicted activity for Model 10

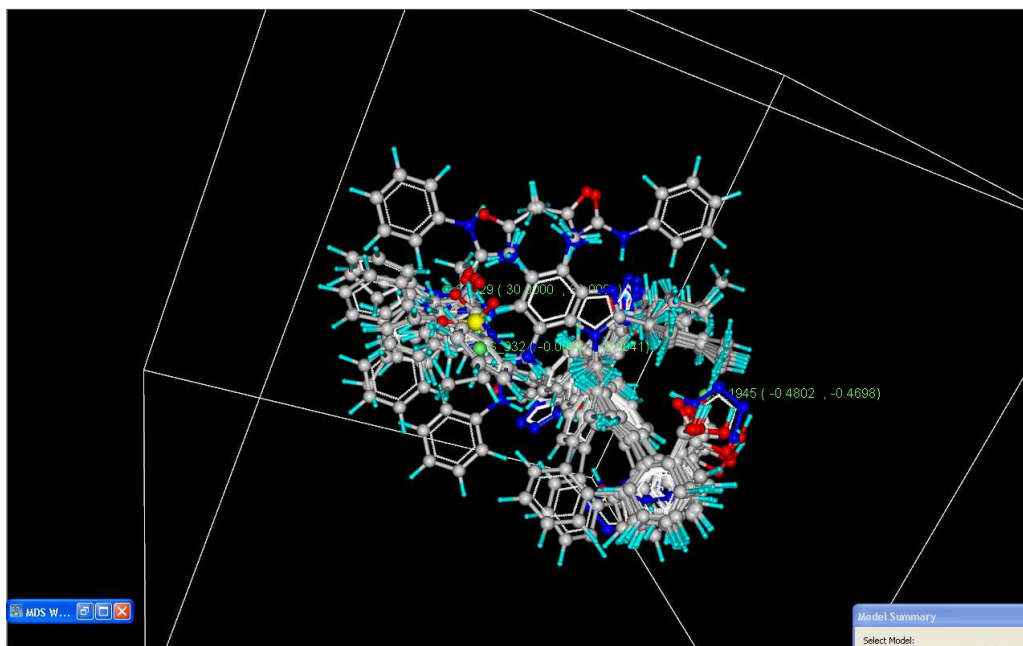


Figure.17. Plot for Model 12 for Genetic Algorithm kNN MFA analysis

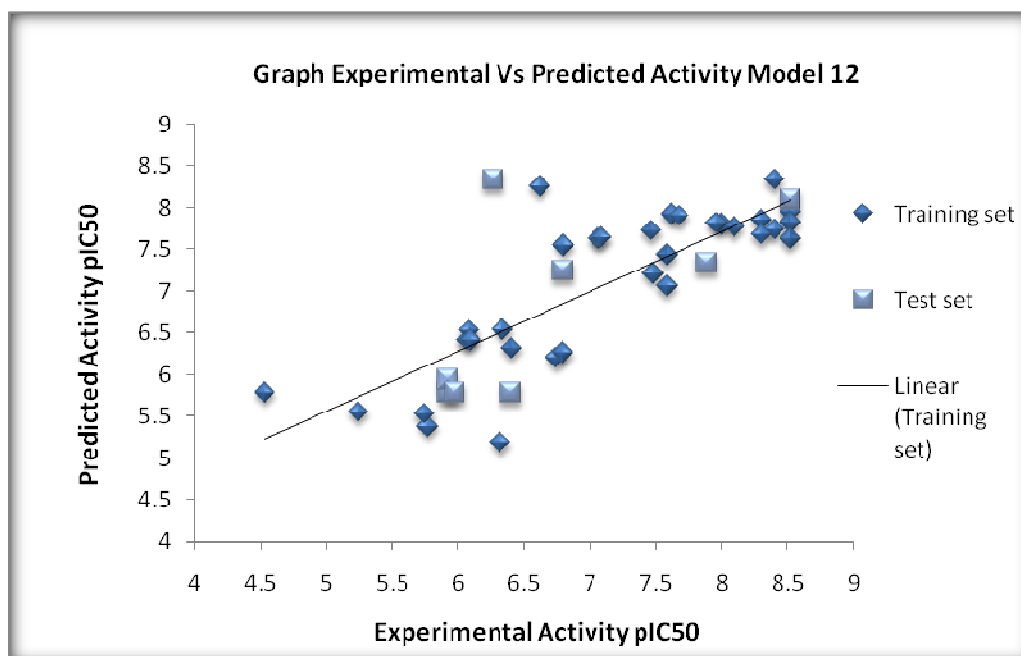


Figure.18. Graph experimental Vs predicted activity for Model 12