**Journal of Computational Methods in Molecular Design, 2015, 5 (2):66-82**

**Scholars Research Library**
*(http://scholarsresearchlibrary.com/archive.html)*

**ISSN : 2231- 3176**
**CODEN (USA): JCMMDA**

# Predicting the drug abuse activity of some novel 4-Alkoxycarbonyl-1, 5-diaryl-1, 2, 3-triazoles on CB1 cannabinoid receptor using 2D and 3D-QSAR (kNN-MFA) analysis

**Anwar Rafique Shaikh[1]*, Malik Amber[1], Imadulislam[1], Ziyaul Haque[2], Sanjay J. Kshirsagar[1], Kankate R. S.[1] and Ketki Eknath Shelar**

*[1]Bhujbal Knowledge City, MET's Institute of Pharmacy, Adgaon, Nashik*
*[2]AliAllana College of Pharmacy, Jamia Isalamia Isha-atul Uloom's Ali-Allana College of Pharmacy, Akkalkuwa*

---

**ABSTRACT**

*2D-QSAR and 3D-QSAR studies by means of MLR, PLS and PCR were performed on a series of 4-Alkoxycarbonyl-1, 5-diaryl-1, 2, 3-triazole analogues for the treatment of drug abuse disorder using software MDS 4.2 version (VLife Science). This study was performed with 13 compounds (Data set) using random as well as manual data selection methods for the division of the data into training and test set. MLR methodology with stepwise forward-backward variable selection method was used for building the QSAR models. Statistically significant QSAR models were generated. Among them most significant model has squared correlation coefficient ($r^2$) = 0.9895, cross validated correlation coefficient ($q^2$) = 0.9747 and predictive correlation coefficient (pred_$r^2$) = 0.9826 for the treatment of drug abuse disorders. The idea of the present study is the search for novel 4-Alkoxycarbonyl-1, 5-diaryl-1, 2, 3-triazole analogues that would show promise to be useful in the treatment of drug abuse disorders.*

**Keywords**: CB1 cannabinoid receptor, 4-Alkoxycarbonyl-1,5-diaryl-1,2,3-triazoles , 2D QSAR, 3D QSAR kNN-MFA, VLife MDS

---

## 1INTRODUCTION

The wide range of pharmacological effects of cannabinoid and endogenous cannabinoid ligands, are mediated by two subtypes of transmembrane G-protein coupled receptors: CB1 and CB2. [1-4] CB1 receptors are expressed in the central nervous system (CNS) with high density in the cerebellum, hippocampus and striatum. [5] CB1 receptors are found in some peripheral tissues (urinary bladder, testis, and ileum) as well. CB2 receptors are predominantly located in the immune system (tonsils, spleen and immune cells) with very low concentration in the CNS. [6] CB1 agonists have potential therapeutic applications in developing drugs for pain, nausea, glaucoma, stroke, cancer and neurological disorders such as multiple sclerosis and Parkinson's disease. The potential applications of CB1 antagonists include the therapeutic treatment of obesity and related metabolic disorders as well as medications for drug addiction. The role of CB1 receptors in these disease states and disorders, the nature of the receptors active sites, and the molecular interactions between the receptors and ligands are not fully understood and are under intensive investigation. [7-8]

Our interest in developing new effective medications for drug abuse has focused our attention on the CB1 receptors because of their selective presence within CNS and their mediation of the effects of psycho-stimulants on brain

circuitry. Pharmacological studies have shown that CB1 receptor antagonists or inverse agonists have the ability to attenuate the elevation of dopamine levels that occurs with psycho-stimulant use, suggesting their potential applications in the treatment of drug abuse disorders 10-12. [9-11]

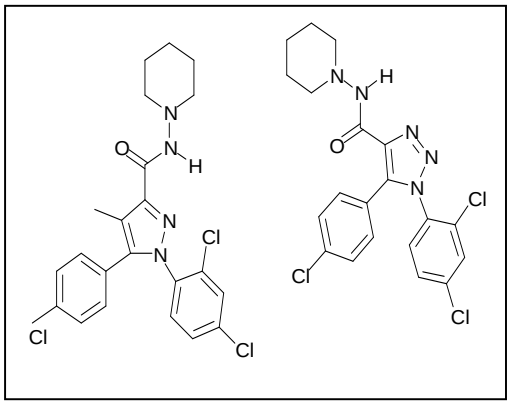
The pyrazole derivative SR141716 (**1**) was the first compound reported to be a potent and selective antagonist for the CB1 receptor. [12] However, further characterization of SR1411716 has shown it to possess an inverse agonist pharmacological profile.[13] This prototypical CB1 receptor antagonist/inverse agonist has been studied as a potential therapeutic for the treatment of obesity, smoking cessation and a variety of other CB1 receptor mediated pathological conditions.[14] However, inverse agonists typically elicit pharmacological responses opposite to agonists and thus are not ideally suited for the treatment of drug addiction due to potential dysphoric side effects associated with these drugs. These include increased nociceptive sensitivity, decreased food intake and body weight, disruption of operant behavior and potential nausea in humans.[15-19] Such side effects undoubtedly would lead to low compliance and relapse among addicts. Moreover, recent reports that describe increased levels of anxiety and depression in patients, along with a higher incidence of suicide (3 to 1 over placebo) in clinical trials with the diet drug Rimonabant (SR141716) strongly disfavors its development as a drug abuse medication.[20]

Based upon the structure of 1,5-diarylpyrazole core template in SR141716 (**1**), bioisosteric replacement has been an important approach to discover new lead compounds of potent CB1 ligands and a variety of bioisosteric analogues of SR14716 have been synthesized and identified as potent ligands[14,21]. In reviewing the literature, the absence of 1,2,3-triazole analogues were significant. To explore this deficiency, our aim was to perform 2D and 3D-QSAR analysis on a series of 1,2,3-triazole analogues of Fig.1, and explore further the binding motifs of CB1 receptor ligands.

The purpose of the present study is to investigate the Physico-chemical parameters responsible for the protective effect of 4-alkoxycarbonyl-1,5-diaryl-1,2,3-triazoles derivatives as a drug abuse medication, explore the correlation between them and to obtain more information for designing novel substituted 4-alkoxycarbonyl-1,5-diaryl-1,2,3-triazoles derivatives with potent protective activity. In the present investigation, three widely used techniques, viz. stepwise forward variable selection method, Genetic algorithm and simulated annealing have been applied for descriptor optimization and multiple linear regression analysis, principal component regression and partial least square has been applied for two and three-dimensional QSAR models development. The generated models provide insight into the influence of various interactive fields on the activity and, thus, can help in designing and forecasting the protecting effect of novel 4-alkoxycarbonyl-1,5-diaryl-1,2,3-triazoles molecules.

A series of 4-alkoxycarbonyl-1,5-diaryl-1,2,3-triazoles analogues which were reported [21] are chosen for QSAR study in order to establish quantitative relationship between physiochemical properties and biological activities of the compounds using MDS software (VlifeScience). [22]

## MATERIALS AND METHODS




**(1)** SR141617A **(2)**

**Figure 1: Parent Chemical structure of 4-alkoxycarbonyl-1,5-diaryl-1,2,3-triazoles derivative**

All molecular modeling studies (2D and 3D) were performed using the Molecular Design Suite (VLife MDS software package, version 3.5; from VLife Sciences, Pune, India), on a HP computer with a Pentium IV processor and a Windows 7 operating system. Structures were sketched using the 2D draw application and converted to 3D structures (Fig.1, Table 1).

**Table 1**: **Inhibition of [$^3$H]SR141716A at CB1 receptors**

| Comp. No. | Code | Activity | |
|---|---|---|---|
| | | ClogP[b] | (nM) |
| 1. | HS69 | 5.33 | 590 ± 170 |
| 2. | HS53-2 | 4.46 | 6900 ± 1300 |
| 3. | HS57-2 | 5.11 | 1420 ± 266 |
| 4. | HS53-1 | 4.68 | 4400 ± 760 |
| 5. | HS57-1 | 5.32 | 66 ± 7.0 |
| 6. | HS60 | 4.69 | 54%[d] |
| 7. | HS57-3 | 5.68 | 180 ± 27 |
| 8. | HS57-4 | 6.21 | 4.6 ± 0.012 |
| 9. | HS57-5 | 6.70 | NA[e] |
| 10. | HS57-6 | 7.62 | NA[e] |
| 11. | HS57-8 | 6.83 | 11 ± 3.4 |
| 12. | HS57-9 | 6.85 | 97 ± 55 |
| 13. | HS57-7 | 7.23 | 240 ± 79 |

[a] *All compounds were tested as the freebase.;* [b] *See Reference 25.;* [c] *All values are the mean ±SEM of three experiments performed in triplicate.;* [d] *Percent inhibition at 100µM.;* [e] *NA, not available.*

### 2.1 Biological data

The cannabinoid receptor affinity of substituted 4-alkoxycarbonyl-1,5-diaryl-1,2,3-triazoles derivatives on CB1 receptors were taken from the reported work. The total set of compounds was divided into a training set for generating 2D and 3D QSAR models and a test set for validating the quality of the models. Selection of the training set and test set molecules was done on the basis of structural diversity and a wide range of activity such that the test-set molecules represent a range of biological activity similar to that of the training set; thus, the test set is truly representative of the training set. The biological activity values [IC50 (µM)] reported in micro molar units were converted to their molar units and then further to negative logarithmic scale and subsequently used as the dependent variable for the QSAR analysis. The values of $IC_{50}$ along with the structure of the compounds in the series are listed in Table 1.

### 2.2 Molecular modeling for 2D QSAR

In 2D QSAR analysis, significant methods Multiple linear regression, principal component regression and partial least square were applied to generate the 2D-QSAR model. The 2D structures were converted to 3D structures by sending them to MDS software. Each compound was energy minimized and batch optimized by using Merck Molecular Force Field force field and charges followed by Austin Model-1. Hamiltonian method was available in MOPAC module with the convergence criterion 0.001 kcal/mol A˚ fixing Root Mean Square Gradients (RMS) to 0.001 kcal/mol A˚. 2D descriptors (physicochemical and alignment independent) were calculated for the optimized compounds on QSAR plus work sheet. The invariable descriptors (the descriptors that are constant for all the molecules) were removed, as they do not contribute to QSAR. Most stable structure for each compound was generated after energy minimization and used for calculating various physico-chemical descriptors like thermodynamic, steric and electronic. The energy-minimized geometry was used for the calculation of the various 2D descriptors (Individual, Chi, ChiV, Path count, Chi Chain, ChiV Chain, Chain path count, Cluster, Path cluster, Kappa, Element Count, Estate number, Estate contribution, Semi-empirical, Hydrophilic– hydrophobic, Polar surface area and Alignment independent) and was considered as independent variables in the present study.

QSAR analysis was performed after removal of all the invariable columns, as they do not contribute to the QSAR. The optimal test and training data set were generated using the manual as well as random data selection method. Sphere exclusion method was also adopted for division of training and test set. Sphere exclusion method is used for creating training and test set from the data. This is a rational selection method which takes into consideration both biological and chemical space for division of data set. Dissimilarity value provides handle to vary train/test set size. It needs to be adjusted by trial and error until a desired division of train and test set is achieved. As a rule, increase in dissimilarity value will lead to increase in number of molecules in the test set. All 13 molecules were subjected to

regression analysis using Multiple linear regression analysis, as model building methods coupled with stepward forward backward variable selection method. Regression analysis was carried out for treatment of drug abuse disorders and the best model was cross-validated. Best two dimensional QSAR results obtained by multiple linear regression analysis (using random and manual data selection method), Partial Least Squares and Principal Component Regression are obtained by the following Table.2a, 2b and 2c respectively.

**Table 2.1: 2D QSAR model optimization by Multiple linear regression analysis (using Random selection method)**

| **Model No. Parameters** | I | II | **III** | IV |
|---|---|---|---|---|
| **Test set** | 1,2,5,13 | 10,4,7,8 | **11,13,3,9** | 13,3,4,8 |
| **N** | 9 | 9 | **9** | 9 |
| **DOF** | 5 | 6 | **6** | 7 |
| **$r^2$** | 0.9917 | 0.9995 | **0.9871** | 0.9670 |
| **$q^2$** | 0.9696 | 0.9988 | **0.9558** | 0.9459 |
| **F-test** | 199.7050 | 6160.0070 | **230.300** | 204.86 |
| **$r^2$se** | 0.1214 | 0.027 | **0.1398** | 0.2128 |
| **$q^2$se** | 0.2328 | 0.0429 | **0.2591** | 0.2723 |
| **pred_$r^2$** | 0.9976 | 0.9608 | **0.9942** | 0.9261 |
| **pred_$r^2$se** | 0.0628 | 0.2475 | **0.1011** | 0.3146 |

**Table 2.2: 2D QSAR model optimization by Partial least square and Principal component regression methods**

| **Model No. Parameters** | **XI** | **XII** |
|---|---|---|
| **2D-QSAR Method** | **PLS results** | **PCR results** |
| **Test set** | **11,12,6,7** | **13,3,4,8** |
| **Training set** | **1,2,3,4,5,8,9,10,13** | **1,2,5,6,7,9,10,11,12** |
| **Optimum components** | **1** | **1** |
| **N** | **9** | **9** |
| **DOF** | **7** | **7** |
| **$r^2$** | **0.9700** | **0.9670** |
| **$q^2$** | **0.9382** | **0.9459** |
| **F-test** | **226.0004** | **204.86** |
| **$r^2$se** | **0.2107** | **0.2128** |
| **$q^2$se** | **0.3014** | **0.27** |
| **pred_$r^2$** | **0.9388** | **0.9261** |
| **pred_$r^2$se** | **0.2607** | **0.3146** |

### 2.3. Three dimensional (3-D) QSAR studies

In the kNN-MFA method three models were generated for the selected members of training and test sets, and the corresponding best two models are reported herein. VLife Molecular Design Suite 3.5 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. To derive the kNN-MFA descriptor fields, a 3D cubic lattice with grid spacing of 2 A˚ in x, y, and z dimensions were created to encompass the aligned molecules. kNN-MFA descriptors were calculated using $sp^3$ carbon probe atom with a van der Waals radius of 2A˚ and a charge of +1.0 with default cut off energy 30 kcal/mol to generate steric field, electrostatic and hydrophobic fields. The steric, and electrostatic energy values were truncated at a default value of ±30 kcal/mol. The kNN-MFA steric, and electrostatic fields thus generated were scaled by the standard method in the software. The 3D-QSAR studies were performed by kNN-MFA using stepwise forward backward, simulated annealing selection method and genetic algorithm method. The software produced more than 7568 descriptors and prior to model development descriptors having zero values or same values were removed which resulted in more than total 2500 descriptors for all the compounds in separate columns This algorithm allows constructing training sets covering all descriptor space areas occupied by representative points. kNN-MFA with simulated annealing and stepwise variable selection was employed for selection of variables to obtain the QSAR model. The standard leave one- out (LOO) procedure was implemented to calculate cross validated $r^2$ ($q^2$) value, that is 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.

### 2.4. Molecular alignment

Molecular alignment was used to visualize the structural diversity in the given set of molecules. This was followed by generation of common rectangular grid around the molecules. The Atom based alignment method was used for alignment by considering the common elements of the series. kNN-MFA method requires suitable alignment of given set of molecules after optimization, alignment was carried out by Atom based alignment method (Fig. 2)

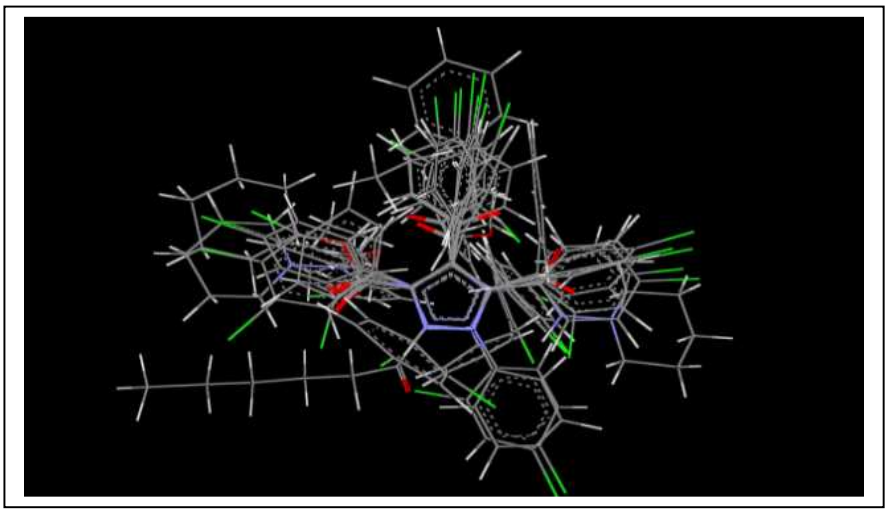


**Figure 2: 3D view of aligned molecules by Atom based type of method of alignment**

### 2.5. k-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). This method employs the kNN classification principle combined with the stepwise selection procedure for optimization of the number of nearest neighbors (k) used to estimate the activity of each compound and optimization of selection of variable from the original pool of all molecular descriptors (steric and electrostatic fields at the lattice points) that are used to calculate similarities between compounds). The descriptors that get selected in a given model are the field points either of steric or of electrostatic nature at particular locations in a common grid around reported set of molecules. For utilizing these descriptors for new ligand design, we consider the field values at different grid points of compounds cluster having most active compound. The extrema of field values of compounds in the cluster of most active compounds decide the range of field values which is preferred and recommended for new compound design.

#### 2.5.1. kNN-MFA Stepwise (SW) forward variable selection:

In stepwise (SW) forward backward variable selection algorithm, the search procedure begins with developing a trial model step by step with a single independent variable and to each step. Independent variables are added one at a time, examining the fit of the model by using the PLS cross-validation procedure. Thus, the model is repeatedly altered from the previous one by adding or removing a predictor variable in accordance with the 'stepping criteria' (in this case F= 4 for inclusion; F=3.99 for exclusion for the forward–backward selection method). The method continues until there is no more significant variable remaining outside the model (Table 3).

**Table 3: Best results of 3D QSAR model obtained by kNN-MFA method**

| Model No. Parameters | XIII | XIV | XV |
|---|---|---|---|
| kNN-MFA method | Stepwise Forward Backward (SWFB) | Simulated Annealing (SA) | Genetic Algorithm (GA) |
| Test set | 11,13,3,9 | 13,3,4,8 | 11,12,6,7 |
| kNN | 2 | 2 | 3 |
| N | 9 | 9 | 9 |
| DOF | 7 | 4 | 5 |
| $q^2$ | 0.7063 | 0.6349 | 0.1414 |
| $q^2$se | 0.5788 | 0.6618 | 1.0508 |
| pred_$r^2$ | -1.4133 | -0.9634 | -2.6121 |
| pred_$r^2$se | 2.0629 | 1.6212 | 2.0034 |
| Descriptors | E_592 | E_1100, E_1179, S_1157, E_892 | S_910, S_323, E_528 |

### 2.6. Development and validation of QSAR models

Models were generated by using significant statistical methods, namely, Multiple linear regression (MLR). The multiple linear regression model and its estimation using ordinary least squares (OLS) is doubtless the most widely used tool. The multiple linear regression model assumes a linear (in parameters) relationship between a dependent variable $y_i$ and a set of explanatory variables $x_i = (x_i0; x_i1; :::; x_iK)$. $X_ik$ is also called an independent variable, a covariate or a regressor. The first regressor $x_i0 = 1$ is a constant unless otherwise specified. Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable *x* is associated with a value of the dependent variable *y*. Here all the calculated descriptors were considered as independent variable and biological activity as dependent variable.

The cross-validation analysis was performed using the leave-one-out method. The following statistical parameters were considered to compare the generated QSAR models: correlation coefficient (r), squared correlation coefficient ($r^2$), predicted $r^2$ (pred_$r^2$), and Fischer's value (F) [24]. To validate the generated QSAR models, the leave-one-out (LOO) method was used, indicated as the value of q2 (cross-validated explained variance), which is a measure of the internal predictive ability of the model. The cross-validated r2 (q2) value was calculated, where yi and are the actual and predicted activities of the $i^{th}$ molecule respectively, and $y_{mean}$ is the average activity of all the molecules in the training set.

Both summations are over all molecules in the training set and hence the predictions were based on the current trial solution, the $q^2$ obtained indicates the predictive power of the current model[25].

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

The predicted $r^2$ (pred_$r^2$) value was calculated, where yi and ^yi 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 model for external test set[25].

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

To evaluate the statistical significance of the QSAR model or 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[25].

## RESULTS AND DISCUSSION

The importance and utility of the new 2D and 3D QSAR method discussed has been established by applying it to known sets of molecules as described above. All the calculated descriptors were considered as independent variable and biological activity as dependent variable. In 2D QSAR analysis, significant methods like Multiple linear regression analysis, Partial Least Square (PLS) and Principal Component Regression (PCR) were applied to generate the model having good $q^2$ and pred_$r^2$ values, one of which was selected having good internal and external predictivity. Selection of training and test set was by Manual data selection and random data selection method. Training and test set were selected if they follow the unicolumn statistics, i.e. maximum of the test is less than maximum of training set and minimum of the test set is greater than of training set, which is prerequisite for further QSAR analysis. This result shows that the test is interpolative i.e., derived from the min-max range of training set. The mean and standard deviation of the training and test set provides insight to the relative difference of mean and point density distribution of the two sets The QSAR models developed by kNN-MFA include both the electrostatic, steric descriptors along with their range to indicate their importance for interaction in molecular field. Models 5,6 and 7 are with 3D QSAR studies. QSAR investigations of the substituted 4-alkoxycarbonyl-1,5-diaryl-1,2,3-triazoles derivatives series resulted in several QSAR equations. Some statistically significant 2D and 3D QSAR models were chosen for discussion.

### 3.1 2D-QSAR model:

**Model III:**

**By Random data selection method;**

$IC_{50}$ = 1.1144(±0.0582) XlogP + 0.1834(±0.0541) $SsCH_3E$-Index – 0.5761

n = 9, Degree of freedom = 6, $r^2$ = 0.9871, $q^2$ = 0.9558, F test = 230.300, $r^2$ se = 0.1398, $q^2$ se = 0.2591, pred_$r^2$ = 0.9942, pred_$r^2$ se = 0.1011

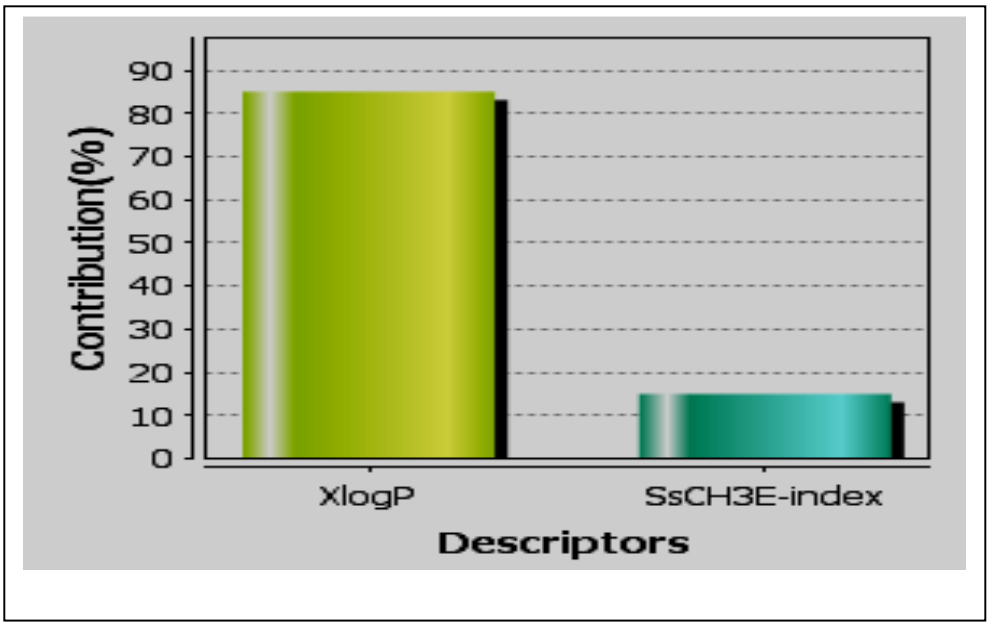



**Figure 3a: Contribution plot of 2D-QSAR Model III**

Among all the significant models the above is the best model generated for anti drug abuse disorders. The equation explains 98.71% ($r^2$ = 0.9871) of the total variance in the training set and has an internal ($q^2$) and external (pred_$r^2$) predictive ability of ~95.58% and ~99.42% respectively. The F-test = 230.300 which is far greater than the F-tabulated value = 3.2850 (http://url.ie/fjj2). This shows the statistical significance of 99.99% of the model, which

means that the probability of failure of the model is 1 in 10,000. All the standard error values are very low (less than 1) hence this model is highly acceptable.

In this QSAR model, the positive coefficient of XlogP [it is a prediction of logP based on the atom type method. It is a constitutional descriptor and is determined by the Kellog method] on the biological activity indicates that higher the value leads to the better activity, while lower the value indicates less biological activity. The positive coefficient of $SsCH_3E$-Index [it is an electrotopological state indices for total no. of CH3- groups connected with single bonds] on the biological activity indicates that higher the value leads to the better activity, while lower the value indicates less biological activity. In addition, the randomization test shows confidence of 99.9999 (Alpha Rand Pred R^2 = 0.0000) that the generated model is not random and hence chosen as the QSAR model.

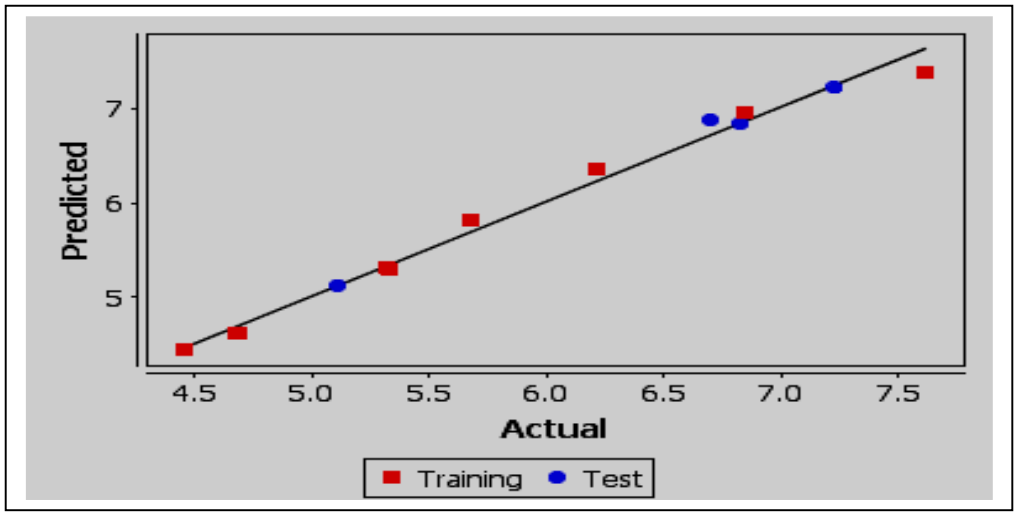



**Figure 3b: Fitness plot showing actual activity versus predicted activity of Model III**

**Model VIII:**
**By Manual data selection method;**
$IC_{50}$= 1.1627(±0.0506) XlogP + 0.3031(±0.0908) $SsCH_3$count – 0.8332n = 9, Degree of freedom = 6, $r^2$ = 0.9895, $q^2$ = 0.9747, F test = 281.5701, $r^2$ se = 0.1345, $q^2$ se = 0.2082, pred_$r^2$ = 0.9826, pred_$r^2$ se = 0.1390

Among all the significant models the above are the best models generated for anti drug abuse disorders. The equation explains 98.95% ($r^2$ = 0.9895) of the total variance in the training set and has an internal ($q^2$) and external (pred_$r^2$) predictive ability of ~97% and ~98% respectively. The F-test = 281.5701 which is far greater than the F-tabulated value = 3.2850 (http://url.ie/fjj2). This shows the statistical significance of 99.99% of the model, which means that the probability of failure of the model is 1 in 10,000. All the standard error values are very low (less than 1) hence this model is highly acceptable.

In this QSAR model, the positive coefficient of XlogP [it is a prediction of logP based on the atom type method. It is a constitutional descriptor and is determined by the Kellog method] on the biological activity indicates that higher the value leads to the better activity, while lower the value indicates less biological activity. The positive coefficient of $SsCH_3$count [This descriptor signifies the no. of CH3-groups connected with single bond] on the biological activity indicates that higher the value leads to the better activity, while lower the value indicates less biological activity.

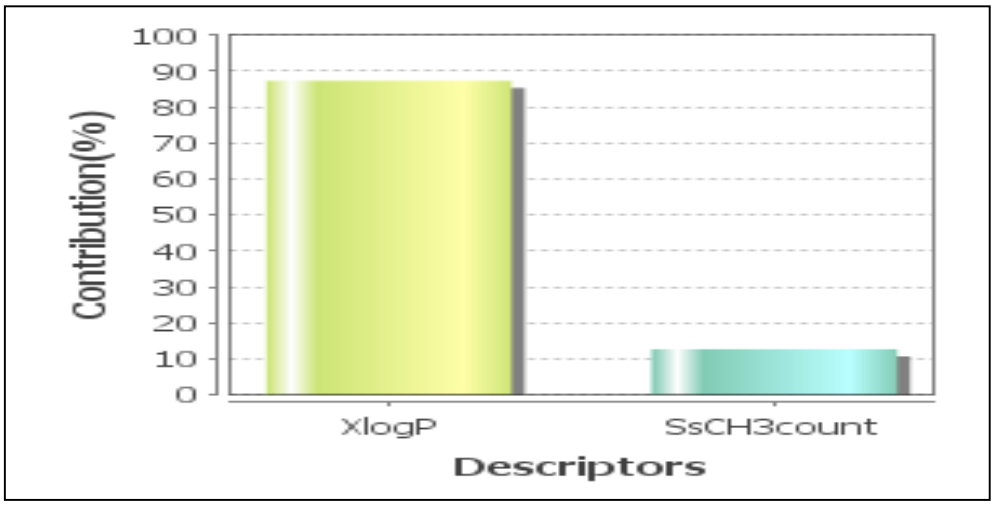



**Figure 4a: Contribution plot of 2D-QSAR Model VIII**

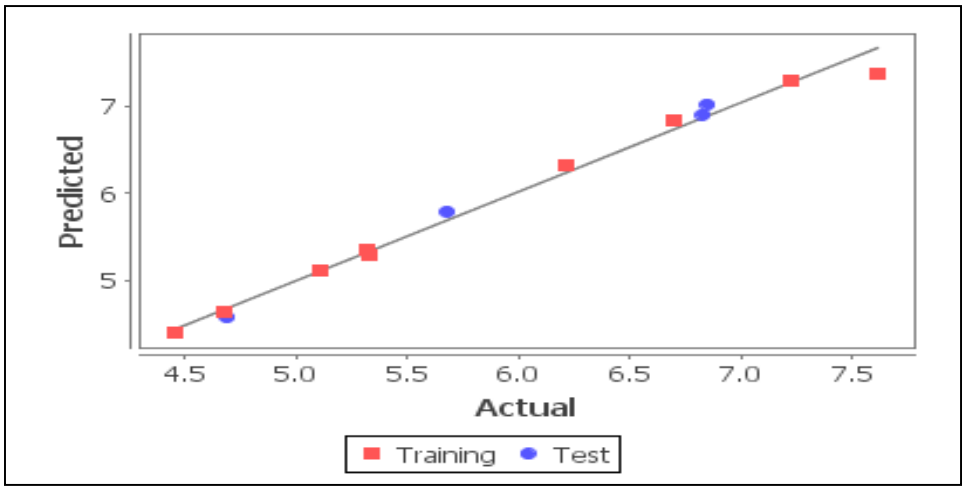



**Figure 4b: Fitness plot showing actual activity versus predicted activity of Model VIII**

**Model XI:**
**By partial least square method;**
$IC_{50}$= 1.1847 XlogP – 0.7883
n = 9, Degree of freedom = 7, $r^2$ = 0.9700, $q^2$ = 0.9382, F test = 226.0004, $r^2$ se = 0.2107, $q^2$ se = 0.3014, pred_$r^2$ = 0.9388, pred_$r^2$ se = 0.2607

Among all the significant models the above are the best models generated for anti drug abuse disorders. The equation explains 97% ($r^2$ = 0.9700) of the total variance in the training set and has an internal ($q^2$) and external (pred_$r^2$) predictive ability of ~93.82% and ~93.88% respectively. The F-test = 226.0004 which is far greater than the F-tabulated value = 3.2850 (http://url.ie/fjj2). This shows the statistical significance of 99.99% of the model, which means that the probability of failure of the model is 1 in 10,000. All the standard error values are very low (less than 1) hence this model is highly acceptable.

In this QSAR model, the positive coefficient of XlogP [it is a prediction of logP based on the atom type method. It is a constitutional descriptor and is determined by the Kellog method] on the biological activity indicates that higher the value leads to the better activity, while lower the value indicates less biological activity.

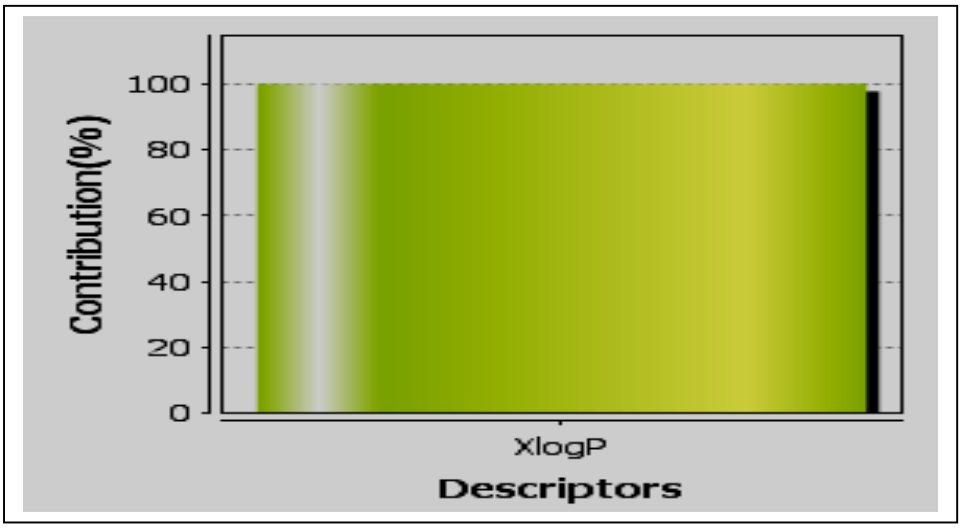



**Figure 5a: Contribution plot of 2D QSAR Model XI**

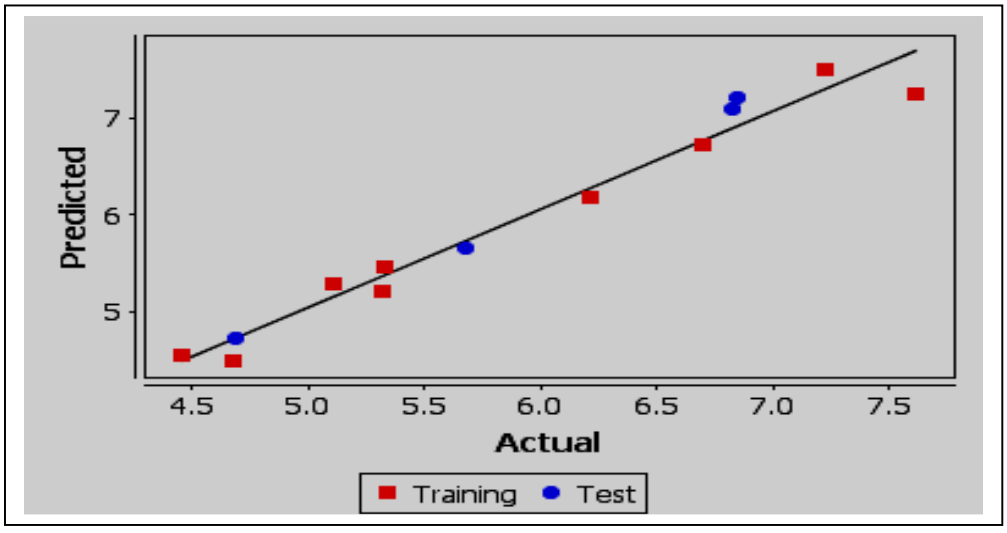



**Figure 5b: Fitness plot showing actual activity versus predicted activity of Model XI**

**Model XII:**
**By principal component regression method;**
$IC_{50}$= 1.1739 SlogP – 0.4720
n = 9, Degree of freedom = 7, $r^2$ = 0.9670, $q^2$ = 0.9459, F test = 204.86, $r^2$ se = 0.2128, $q^2$ se = 0.27, pred_$r^2$ = 0.9261, pred_$r^2$ se = 0.3146

Among all the significant models the above are the best models generated for anti drug abuse disorders. The equation explains 96.70% ($r^2$ = 0.9670) of the total variance in the training set and has an internal ($q^2$) and external (pred_$r^2$) predictive ability of ~94.59% and ~92.61% respectively. The F-test = 204.86 which is far greater than the F-tabulated value = 3.2850 (http://url.ie/fjj2). This shows the statistical significance of 99.99% of the model, which means that the probability of failure of the model is 1 in 10,000. All the standard error values are very low (less than 1) hence this model is highly acceptable.

In this QSAR model, the positive coefficient of SlogP [it is the log of Octanol/water partition coefficient, including implicit hydrogens. This is an atomic contribution model that calculates logP from the given structure, i.e. correct

protonation state (washed structures). Results may vary from logP (o/w) descriptor] on the biological activity indicates that higher the value leads to the better activity, while lower the value indicates less biological activity.

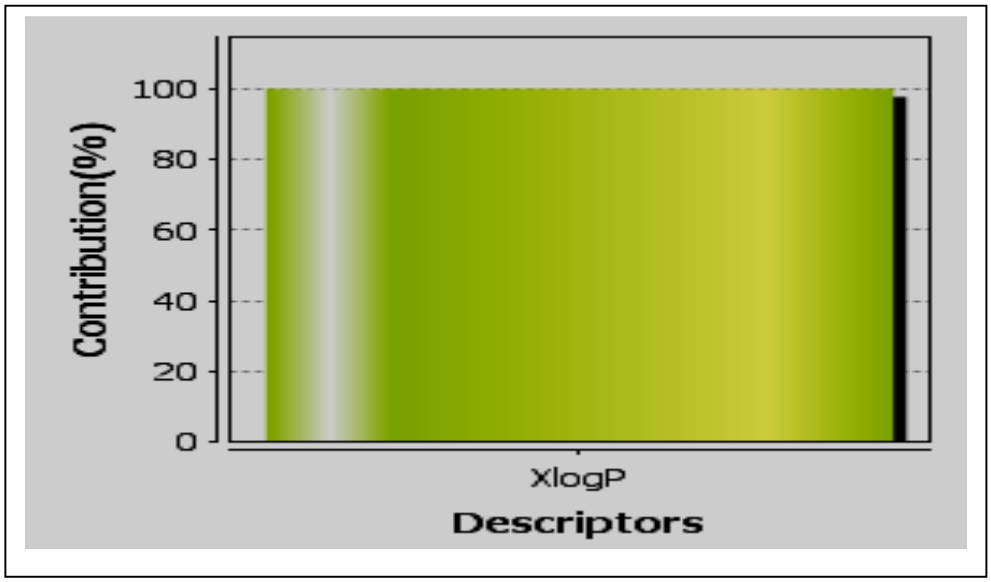



**Figure 6a: Contribution plot showing 2D QSAR of Model XII**

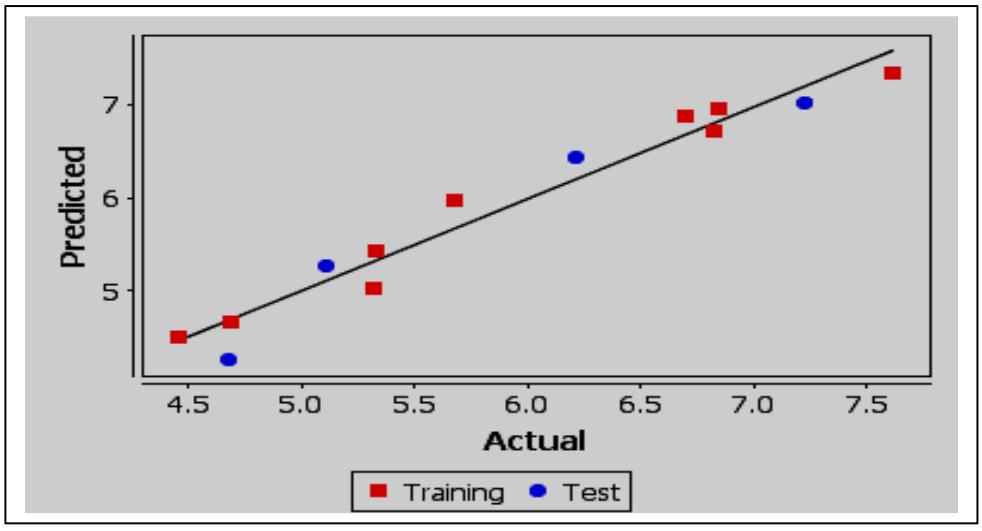



**Figure 6b: Fitness plot showing actual activity versus predicted activity of Model XII**

**3.2 3D-QSAR model:**
**Model XIII:**
**By stepwise forward backward method;**
$IC_{50}$= E_592 (-0.1075, -0.0342)
N= 9, Degree of freedom = 7, $q^2$ = 0.7063, $q^2$se = 0.5788, pred_$r^2$ =-1.4133, and pred_$r^2$ se = 2.0629

According to the **kNN-MFA methodology** $Ic_{50}$ is a function of independent variables, steric and electrostatic fields. Values of it prove the equation statistically to be significant. It also shows the predictive power of the model as 70.63%. It showed steric and electrostatic field energy of interactions between probe (CH) and compounds at their

corresponding spatial grid points of 592. Numbers nearest neighbor's k of 2 were observed with this model i.e. two values are proved statistically significant. It is observed from the **Fig 6** that the negative coefficient of E_592 suggested that electronegative substituent may be favorable on the position of triazole ring for better activity.

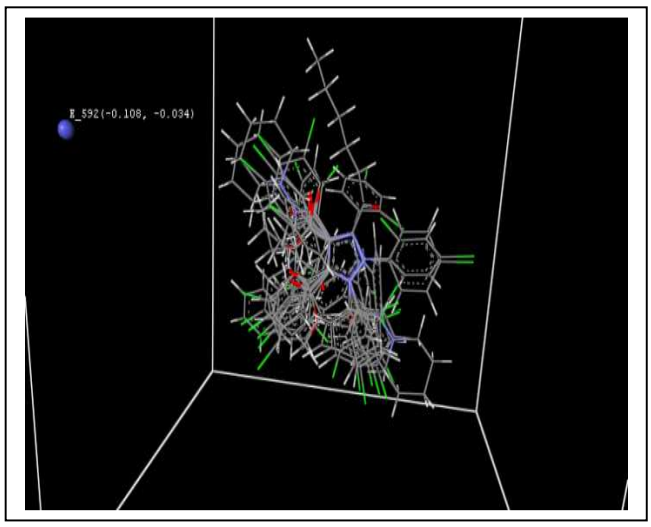



**Figure 7a: Contribution plot of 3D-QSAR Model XIII**

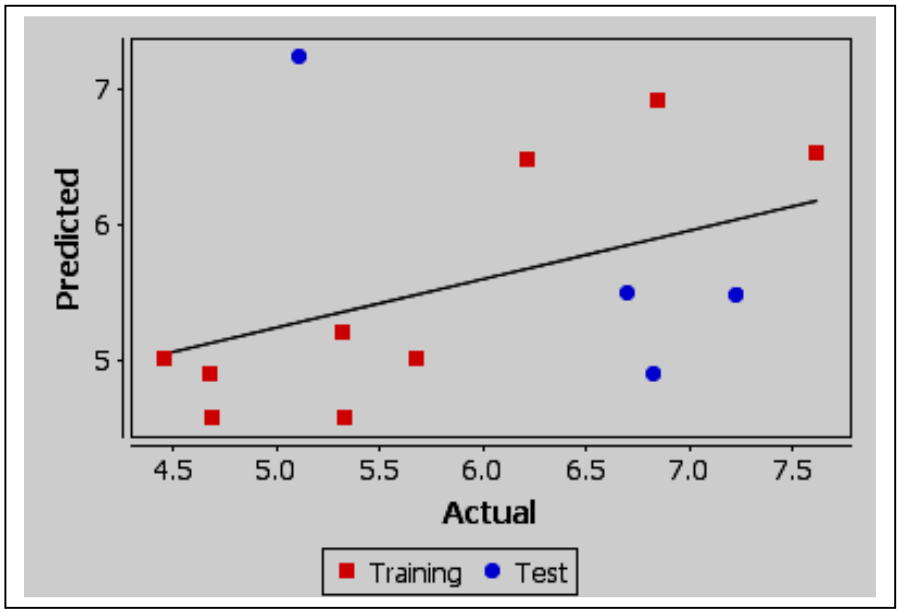



**Figure 7b: Fitness plot showing actual activity versus predicted activity of Model XIII**

**Model XIV:**

**By simulated annealing method;**

$IC_{50}$ = E_1100 (0.0243, 0.1525) + E_1179 (0.3555, 0.8597) + S_1157 (30.0000, 30.0000) – E_892 (-0.0378, 0.0799)
N= 9, Degree of freedom = 4, $q^2$ = 0.6349, $q^2$se = 0.6618, pred_$r^2$ =-0.9634, and pred_$r^2$ se = 1.6212

According to the **kNN-MFA methodology** $IC_{50}$ is a function of independent variables, steric and electrostatic fields. Values of it prove the equation statistically to be significant. It also shows the predictive power of the model as 63.49%. It showed steric and electrostatic field energy of interactions between probe (CH) and compounds at their

corresponding spatial grid points of 1100, 1179, 1157, and 892. Numbers nearest neighbor's k of 2 were observed with this model i.e. two values are proved statistically significant. It is observed from the Fig.6 that the positive coefficient of E_1100 and E_1179 suggested that electropositive substituent may be favorable on the position of triazole ring for better activity. Even the steric factor S_1157 is positive which indicates the favorability of bulky groups on the triazole ring to increase the activity. The negative coefficient of E_892 indicates the addition of electronegative atom also at the position is responsible for the increase in biological activity.

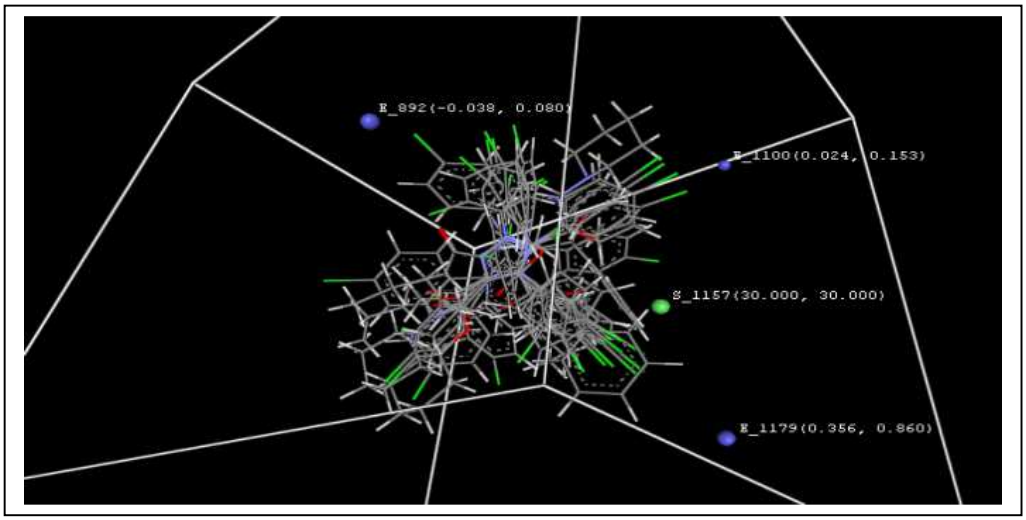



**Figure 8a: Contribution plot of 3D-QSAR Model XIV**

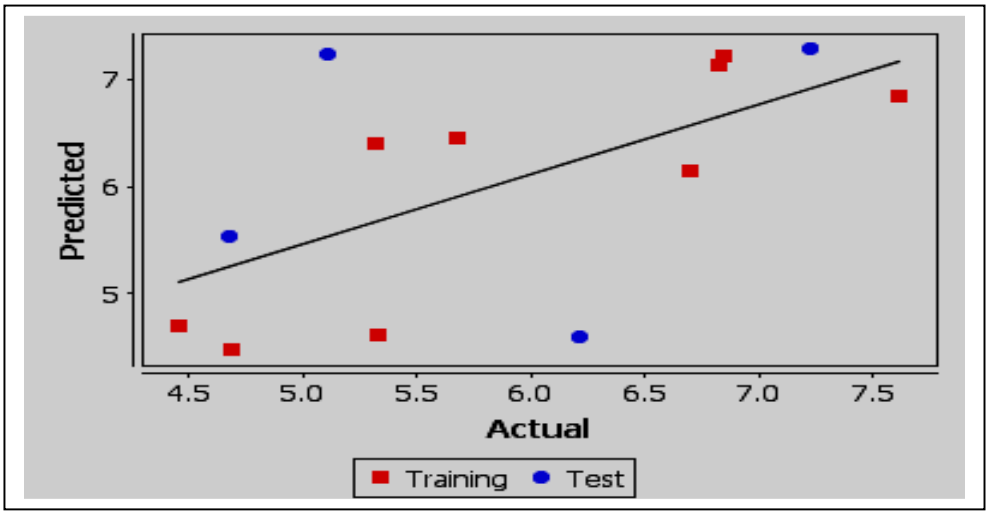



**Figure 8b: Fitness plot showing actual activity versus predicted activity of Model XIV**

**Model XV:**

**By genetic algorithm method;**

$IC_{50}$= - S_910 (-0.1861, -0.0416) – S_323 (-0.0298, -0.0027) + E_528 (0.1386, 2.0222)

N= 9, Degree of freedom = 5, $q^2$ = 0.1414, $q^2$se = 1.0508, pred_$r^2$ =-2.6121 and pred_$r^2$ se = 2.0034

According to the **kNN-MFA methodology** $IC_{50}$ is a function of independent variables, steric and electrostatic fields. Values of it prove the equation statistically to be significant. It also shows the predictive power of the model as 14.14%. It showed steric and electrostatic field energy of interactions between probe (CH) and compounds at their corresponding spatial grid points of 910, 323 and 528. Numbers nearest neighbor's k of 3 were observed with this

model i.e. three values are proved statistically significant. It is observed from the Fig.6 that the negative coefficient of S_910 and S_323 suggested that the bulky groups are not favorable for the activity. While the positive coefficient of E_528 indicates the electropositive substitution is necessary for better biological activity.

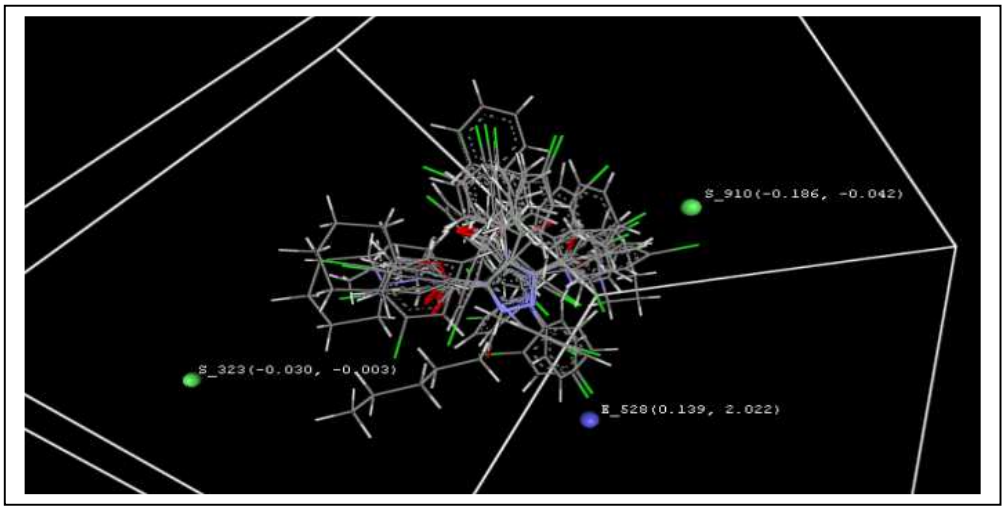



**Figure 9a: Contribution plot of 3D-QSAR Model XV**

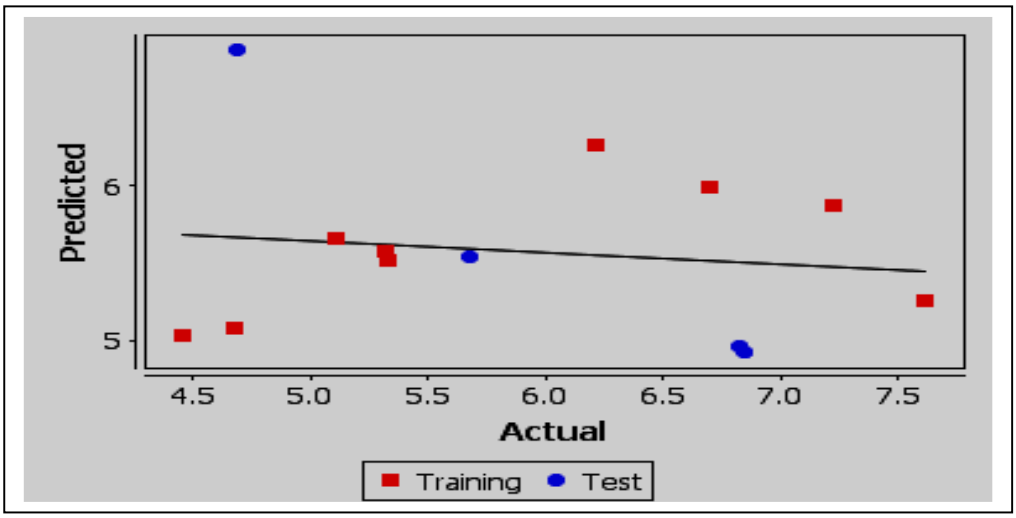



**Figure 9b: Fitness plot showing actual activity versus predicted activity of Model XV**

## CONCLUSION

In the present investigation, all proposed QSAR models were statistically significant, thus, from above QSAR investigations it could be concluded that 2D/3D descriptors properties of substituted 4-Alkoxycarbonyl-1, 5-diaryl-1, 2, 3-triazoles derivatives are mainly involved in treatment of drug abuse disorders. The good correlation between experimental and predicted biological activity for compounds in the test set further highlights the reliability of the constructed QSAR model. The requirements for the more potent biological activity are explored with 2D, 3D and group based QSAR studies. The 2D technique indicates the importance of XlogP, SssCH3-Eindex, SssCH3-count, and SlogP of the compounds on the activity. The 3D QSAR analysis makes it possible to relate chemical structures of ligands and their binding affinity with respect to different bio targets by using the kNN-MFA techniques. Thus it provides a direct view of factors expressed in terms of molecular fields (electrostatic, steric) affecting the binding affinity. This in turn could give the reasonably good prediction of binding affinity. The location and range of

function values at the field points selected by the models provide clues for the design of new molecules. Hence, this method is expected to provide a good alternative for the drug design.

The 3D-QSAR model showed that electrostatic effects dominantly determine the binding affinities and these QSAR models developed in this study would be useful for the development of new drugs as a medicament for the drug abuse disorder. The 2D-QSAR studies revealed that alignment independent descriptors were the major contributing descriptors. The descriptor values obtained in this study helped in quantification of the structural features of 4-alkoxycarbonyl-1, 5-diaryl-1, 2, 3-triazole derivative.

After successful QSAR studies, attempts were made to predict the activities of the newly designed analogues of these reported compounds. we have designed 47 compounds among which 11 compounds are showing higher activity than the reported analogues. In future, we can synthesize these compounds using the selected scheme and confirm their activity by carrying out *in vivo* evaluation.

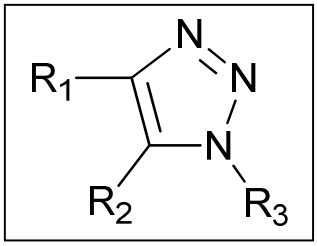


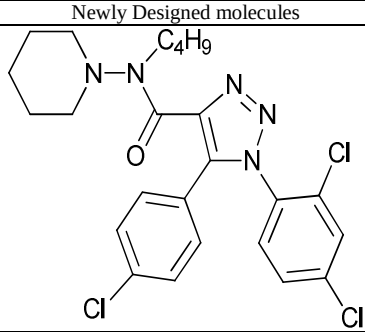

**Figure 10: Parent structure of the triazole derivatives**

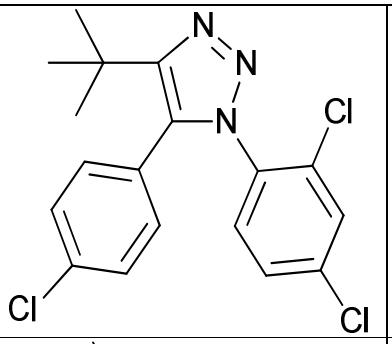
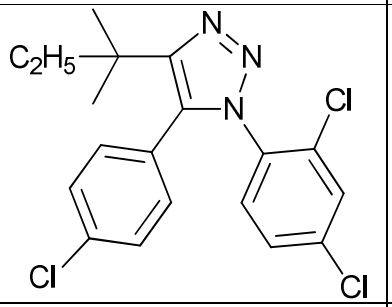
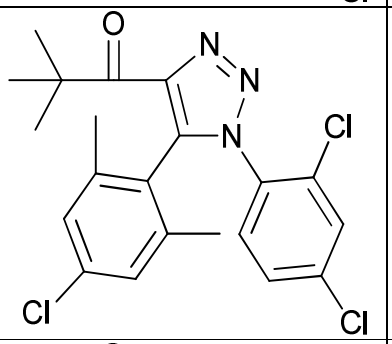
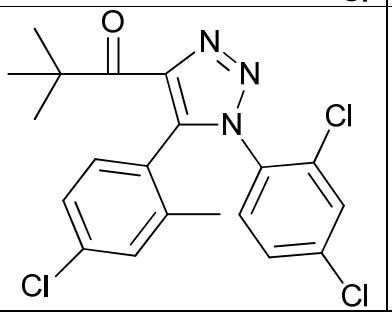
**Newly Designed Molecules**

After successful QSAR studies, attempts were made to predict the activities of the newly designed analogues of these reported compounds. we have designed several compounds among which 5 compounds are showing higher activity than the reported analogues. In future, we can synthesize these compounds using the selected scheme and confirm their activity by carrying out *in vivo* evaluation.

**Table 4: Newly designed triazole derivative with their predicted biological activity**

**These compounds have more biological activity than the reported.**

| Sr. No. | Newly Designed molecules | Activity |
|---|---|---|
| 1 |  | 7.1025 |

| | | |
|---|---|---|
| 2 |  | 8.08166 |
| 3 |  | 8.57019 |
| 4 |  | 8.71542 |
| 5 |  | 8.1691 |

**Acknowledgment**
The author wishes to express gratitude to V-life Science Technologies Pvt. Ltd. for providing the software for the study. Also the authors are thankful to the trustee Bhujbal Knowledge City for providing the necessary facilities to carry out the research work.

**REFERENCES**

[1] W A Devane, F A Dysarz., M R Johnson., L S Melvin., Howlettm, A C *J Mol Pharmacol*,**1988**, *34*, 605.
[2] L A Matsuda, S J Lolait, M J Brownstein, A C Young, Bonner, T I *J Nature*, **1990**, *346*, 561.
[3] S Munro, Thomas K. L, Abu-Shaar, M *J Nature*, **1993**, 365, 61.
[4] C M Gerard, C Mollereau, G Vassart, Parmentier, M *Biochem J*, **1991**, *279*, 129.

[5] M Herkenham, A B Lynn, M D Little, M R Johnson, L S Melvin, Rice, K C *Proc Natl Acad Sci USA*, **1990**, *87*, 1932.
[6] M Herkenham, In: *Cannabinoid Receptors*, ed New York, Academic Press, **1995**, p 145.
[7] V DiMarzo, M Bifulco, De Petrocellis, L *Nat Rev Drug Discov J*, **2004**, *3*, 771.
[8] O M H Salo, J R Savinainen, T Parkkari, T Nevalainen, M Lahtela-Kakkonen, J Gynther, J T Laitinen, T Järvinen, A J Poso, *Med Chem J*, **2006**, *49*, 554.
[9] P M Beardsley, B F Thomas, *Behav Pharmacol J*, **2005**, *16*, 275.
[10] B Le Foll, S R Goldberg, *J Pharmacol Exp Ther*, **2005**, *312*, 875.
[11] S A Tucci, J C Halford, J A Harrold, Kirkham, T C *J Curr Med Chem*, **2006**, *13*, 2669.
[12] M Rinaldi-Carmona, F Barth, M Heaulme, D Shire, B Calandra, C Congy, S Martinez, J Maruani, G Neliat, D Caput, P Ferrara, P Soubrie, J C Brelière, Le Fur, G. *FEBS Lett J*, **1994**, *350*, 240.
[13] R G Pertwee, *Life Sci J*, **2005**, *76*, 1307.
[14] F *J* Barth, *Annu Rep Med Chem*, **2005**, *40*, 103.
[15] G Tanda, Goldberg S R *Psychopharmacology*, **2003**,*169*, 115.
[16] P J Mclaughlin, K Winston, L Swezey, A Wisniecki, J Aberman, Tardif, D J, Betz A.J, K Ishiwari, A Makriyannis, Salamone J D, *J Behav Pharmacol*, **2003,** 1583.
[17] T U Jarbe, N V DiPatrizio, A Makriyannis, Pharmacol, *Biochem Behav J*, **2003**, *75*, 809.
[18] J A Vivian, S Kishioka, E R Butelman, J Broadbear, K O Lee, Woods J H, *J Pharmacol Exp Ther*, **1998**, *286*, 697.
[19] P J Winsauer, P Lambert, J M Moerschbaecher, *J Behav Pharmacol*, **1999**, *10*, 497.
[20] R Christensen, P K Kristensen, E M Bartels, H Biddal, Astrup A *Lancet* **2007**, *370*, 1671.
[21] Hong Shu, Sari Izenwasser, Dean Wade, D Stevens Edwin , and L Trudell Mark , Design, synthesis, and biological activities of novel 4-Alkoxycarbonyl-1,5-diaryl-1,2,3-triazoles derivatives, *Bioorganic and Medicinal Chemistry Letters*, **2009**, 19, 891.
[22] Molecular Design Suite (MDS 4.2) (2013) VLife Sciences Technologies, Pvt Ltd Pune, India, www.vlifesciences.com
[23] S Boltan, *Pharmaceutical statistics- Practical & Clinical applications*, Informa healthcare, New York, 5th edn, (2009) 65, 119.