Journal of Computational Methods in Molecular Design, 2015, 5 (3):135-149

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

ISSN : 2231- 3176
CODEN (USA): JCMMDA

Investigation of 5,6-dihydro-2-pyrones derivatives as potent anti-HIV agents inhibitors

Emmanuel Israel Edache*, Adamu Uzairu and Stephen Eyije Abechi

Department of Chemistry, Ahmadu Bello University Zaria, Nigeria
Corresponding author: edacheson2004@gmail.com

ABSTRACT

Human immunodeficiency virus protease (HIV-1PR) is an emerging and potential drug target for anti-HIV therapy. This enzyme plays a crucial role in the process of viral maturation and infectivity. Increased resistance in strains of protease inhibitors is a major obstacle to anti-HIV therapy. The aim of the study was to correlate the chemical structure of compounds with experimental data from biological activity anti-HIV-1 protease. In the present study a series of 5,6-dihydro-2-pyrones derivatives (containing 24 compounds) as HIV-1PR inhibitors was subjected to quantitative structure–activity relationship (QSAR) analysis. For building the regression models, genetic function approximation (GFA) were used to predict the HIV-1PR inhibition activity. Based on prediction, the best validation model for anti-HIV inhibition activity with squared correlation coefficient (R^2)= 0.9541, cross validated correlation coefficient (Q^2cv)= 0.7440 and external prediction ability pred_R^2= 0.9266 showed that Sum of E-State descriptors of strength for potential Hydrogen Bonds of path length 6, Minimum E-State descriptors of strength for potential Hydrogen Bonds of path length 6, Overall or summation solute hydrogen bond basicity and Solvation Energy were the positive contributors. Leave one out cross validation and Y-randomization analysis were performed in order to confirm the robustness of the model. The proposed model provides a better understanding on the binding mode pattern of the compounds to the binding site of HIV-1 enzyme. The results of the present study is useful for designing more potent HIV-1PR inhibitors.

Keyword: QSAR analysis, Anti-HIV, Protease inhibitors, Regression analysis, semi-empirical, GFA.

INTRODUCTION

A number of targets for potential chemotherapeutic intervention of the human HIV-1 are provided by the retrovirus life cycle [1]. The HIV-1 protease-mediated transformation from the immature, non-dangerous virion, to the mature, infective virus is a crucial stage in the HIV-1 life cycle [2,3]. HIV-1 protease has therefore become a major target for anti-AIDS drug design, and its inhibition has been shown to extend the length and improve the quality of life of AIDS patients [4]. A large number of inhibitors have been designed, synthesized, and assayed, and several HIV-1 protease inhibitors are now utilized in the treatment of AIDS [5]. The role of the protease in infection spreading is to act as a “molecular scissor”, cleaving inactive viral polyproteins into smaller, functional proteins. In the presence of protease inhibitors, viral particles are unable to mature and are cleared rapidly [6].

Extensive clinical trials have led to the development of the following five HIV-1 PR inhibitors that are approved by the United States Food and Drug Administration (US-FDA): Saquinavir mesylate (SAQ), Ritonavir (RIT), Indinavir

sulfate (IND), Nelfinavir mesylate (NLF), and Amprenavir (APR) [7]. Such drugs particularly are effective in short-term treatments, while resistance limits their long-term efficacy. The widespread use of these chemotherapeutics has resulted in the emergence of drug-resistant mutants that pose a continuing challenge to the design of new active compounds. Many crystallographic [8] and energetic studies [9] about the HIV-1 PR, wild type and mutants, have made the enzyme an attractive target for the computer-aided drug design strategy [10]. The resistance of the HIV-1 PR mutants indulged us to apply very fast and precise techniques, able to predict the biological activity for the new HIV-1 PR inhibitors.

Quantitative structure activity relationships (QSARs), one of the most widely used methods in chemometrics, and molecular docking are two of the supportive methods for drug design and prediction of drug activity [11]. QSAR models are mathematical equations which generate a communication between chemical structures and their biological activities, while molecular docking is done to specify the structural features that are important for interaction with a receptor [12].

In this report, we have performed a QSAR study on 24 compounds of 5,6-dihydro-2-pyrones derivatives [13] which had been synthesized and evaluated as Anti-HIV-1 protease inhibitors. The study is hoped to provide a better understanding of the binding pattern of the compounds and may also be useful for the designing more potent HIV-PR inhibitors.

MATERIALS AND METHODS

Chemical Data

Molecules

24 molecules belonging to 5,6-dihydro-2-pyrones used as anti-HIV-1 protease inhibitors were taken from the literature and used for the present study [13]. These were divided into eighteen (18) training six (6) test sets, respectively. The structures observed, and the biological activities of the training and test sets of these compounds are presented in Fig.1 and Table, respectively. Predictive power of the models obtained was evaluated by a test set of 6 5,6-dihydro-2-pyrones as anti-HIV-1 protease inhibitors molecules with uniformly distributed biological activity. Selection of test set molecules was made by considering the fact that, test set molecules represent range of biological activity similar to training set. The mean of biological activity of training and test set was 0.9789 and 0.6889, respectively. Therefore test set is the true representative of the training set.

Figure 1: General structure of 5,6-dihydro-2-pyrones compounds used in the present study.

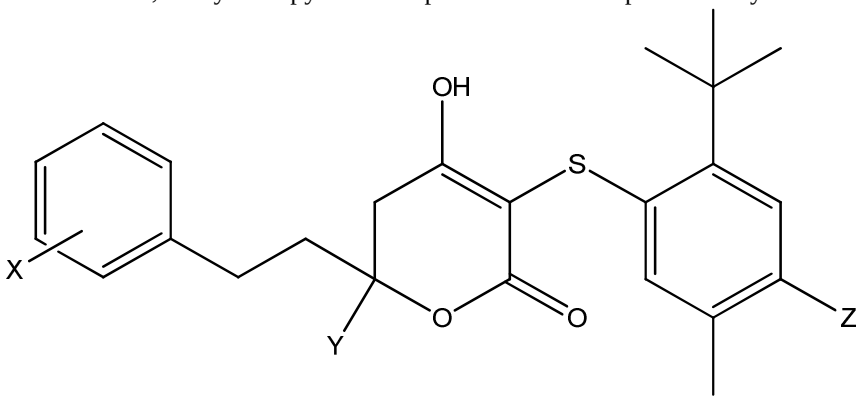

Figure 1. Substituted 5,6-dihydro-2-pyrones compounds

Table 1. Biological activities of training and test set derivatives

Comp No	X	Y	Z	pIC50
1[a]	H	Ph	H	1.5440
2[a]	H	Ph	OH	1.5185
3[a]	H	Ph	$O(CH_2)_2OH$	0.8325
4[a]	H	Ph	CH_2OH	0.8195
5[a]	H	Ph	OCH_3	1.1760
6[a]	4-OH	Ph	H	1.0413
7[a]	4-NH_2	Ph	H	1.3802
8[b]	H	Ph-OH	H	1.6020
9[a]	H	Ph-NH_2	H	1.5051
10[a]	H	$PhO(CH_2)_2OH$	H	1.0792
11[b]	4-OH	Ph	CH_2OH	0.2304
12[a]	3-OH	Ph	CH_2OH	0.3979
13[a]	4-NH_2	Ph	CH_2OH	0.4913
14[a]	3-NH_2	Ph	CH_2OH	0.6020
15[o]	H	$PhO(CH_2)_2OH$	CH_2OH	0.1461
16[a]	H	$PhO(CH_2)_2OH$	$O(CH_2)_2OH$	0.8061
17[b]	H	$PhO(CH_2)_2OH$	OH	0.5682
18[a]	H	$PhO(CH_2)_2OH$	CH_2OCH_3	0.6532
19[a]	4-OH	PhOH	CH_2OH	2.0791
20[b]	4-OH	Cyclohexyl	CH_2OH	0.6127
21[a]	4-OH	Isopropyl	CH_2OH	0.5563
22[a]	4-OH	Methyl	CH_2OH	0.6334
23[a]	4-NH_2	Cyclohexyl	CH_2OH	0.5051
24[b]	4-NH_2	isopropyl	CH_2OH	0.4313

[a]Training set; [b]Test set; [o]Outlier

Biological activity

The logarithm of measured IC_{50} (μM) against anti-HIV-1 protease inhibitors as pIC_{50} ($pIC_{50} = \log 1/IC_{50}$) was used as dependent variable, consequently correlating the data linearly to the independent variable/descriptors.

Molecular Modeling

Software

All molecular modeling studies were carried out using Spartan'14 version 1.1.2 [14] and PaDEL-Descriptor version 2.18 [15] running on DELL INSPIRON Pentium, Dual-core processor window seven (7) operating system. The molecular structures of the compounds in the selected series were drawn in the graphic user interface of the software. Structures were built using 2D application tool and exported in 3D format. All 3D structures were geometrically optimized by minimizing energy. Calculation of the structural electronic and other descriptors of 5,6-dihydro-2-pyrones derivatives was conducted with the semi-empirical Parametric model 3 (PM3) method, which is a better method than the others semi-empirical method. Method of PM3 is repair method of before all like modified intermediate neglect of differential overlap (MNDO) method [16], which can predicts compounds having valence many with the best accuracies [17]. The PM3 method can be used for the analysis of 5,6-dihydro-2-pyrones derivatives, because 5,6-dihydro-2-pyrones derivatives are organic compounds considering atoms of C, H, O and N.

Calculation of descriptors

Different types of descriptors were calculated for each molecule in the study table using default settings within Spartan'14 version 1.1.2 [14] and PaDEL-Descriptor version 2.18 [15]. These descriptors include electronic, spatial, structural, and thermodynamic. Some of the list of descriptors used in the study is given in given Table 3.

Generation of QSAR models

QSAR analysis in computational research is responsible for the generation of models to correlate biological activity and physicochemical properties of a series of compounds. The underlying assumption is that the variations of biological activity within a series can be correlated with changes in measured or computed molecular features of the molecules. In the present study, we have used genetic function approximation (GFA) technique to generate different 1D, 2D, and 3D QSAR models from various descriptors available within Spartan'14 version 1.1.2 [14] and PaDEL version 2.18 [15] modeling software in order to deduce correlation between the structure and biological activity of the present series of molecules. Our approach follows the methodology used previously to generate successful 3D QSAR models for anticancer [18], antimalarial [19], anti-tubercular [20], and anti-fungi agents [21]. GFA technique was used since it generates a population of equations rather than one single equation for correlation between

biological activity and physicochemical properties. GFA developed by Rogers, involved the combination of Friedman's multi variant adaptive regression splines (MARS) algorithm with Holland's genetic algorithm to evolve population of equations that best fit the training set data [22]. This is done as follows:

(i) An initial population of equations is generated by random choice of descriptors. The fitness of each equation is scored by Lack- of- Fit (LOF) measure,

$$LOF = \frac{LSE}{\left\{1-\left[\frac{c+d*p}{m}\right]\right\}^2} \quad (1)$$

Where LSE is least square error, *c* is the number of basic functions in the model, *d* is the smoothing parameter which controls the number of terms in the equations and *p* is the number of features contained in all terms of the models, and *m* is the number of compounds in the training set.

(ii) Pair's form the population of equations are chosen at random and "crossovers" are performed and progeny equations are generated.
(iii) The fitness of each progeny equation is assessed by LOF measure.

If the fitness of new progeny equation is better, then it is preserved. The model with proper balance of all statistical terms will be used to explain variance in the biological activity. A distinctive feature of GFA is that it produces a population of models (e.g. 100), instead of generating a single model, as do most other statistical methods. The range of variations in this population gives added information on the quality fit and importance of the descriptors. By examining these models, additional information can be obtained. For example, the frequency of use of a particular descriptor in the population of equations may indicate how relevant the descriptor is to the prediction of activity. Combination of robust statistical technique GFA coupled with the use of different types of descriptors would result in better prediction of biological activity for anti-HIV-1 protease inhibitors. The number of terms in the equation was fixed to one, two, three and four including constant in the training set. The set of equations generated were evaluated on the following basis: (a) LOF measure; (b) Variable terms in the equations; (c) Cross validated and non-cross validated R^2; (d) Randomized cross validated R^2; (e) Predictive ability of equation. Cross validated R^2 (Q^2cv), Randomized cross validated R^2, were calculated using cross validated test option in the statistical tools supported in material studio version 7.0.

All possible combination of descriptors was considered to find the best regression model. The multi-collinearity among variables was identified using variance inflation factor (VIF) [23]. The VIF for the ith regression coefficient is expressed as:

$$VIF = \frac{1}{1-R_i^2} \quad (2)$$

R_i represents the coefficient produced by regressing the descriptor xi against the other descriptors, x_j $(j \neq i)$ If VIF was greater than 10, it was not considered as a model. The predictive activity of the model is quantitated in terms of R^2 which is defined as:

$$R^2 = 1 - \frac{\Sigma(y_{pred}-y_{actual})^2}{\Sigma(y_{actual}-y_{mean})^2} \quad (3)$$

In this equation, y_{pred}, y_{actual} and y_{mean} are the predicted, actual, and mean values of the target property, respectively. We have used leave-one-out cross-validation to verify the performance of a trained model. If the cross-validation criteria were less than 0.5, then they were not considered as models.

The predictive R^2 was based only molecules not included in the training set and is defined as:

$$R^2_{pred} = 1 - \frac{\Sigma(Y_{pred(Test)}-Y_{Test})^2}{\Sigma(Y_{(Test)}-\bar{Y}_{training})^2} \quad (4)$$

Where,$Y_{pred(Test)}$, and $Y_{(test)}$ indicate predicted, observed and activity values respectively of the test set compounds and $\bar{Y}_{training}$ indicates mean of observed activity values of the training set. For a predictive QSAR model, the value of R^2_{pred} should be more than 0.5 [20, 24-26].

Like R^2cv the predictive R^2 can assume a negative value reflecting a complete lack of predictive ability of the training set for the molecules included in the test set [27,28].
Q is the quality factor [29,30]. The quality factor Q is used to decide the predictive potential of the models. The quality factor Q is defined as the ratio of correlation coefficient to the standard error of estimation. We found it to be a good parameter to explain the predictive potential of the models proposed by us. The higher the value of Q the better is the predictive potential of the models [29-31].

$$Q = \frac{R}{SE} \quad (5)$$

To check the external predictability of the selected model is r^2m which was proposed by Roy and Paul, 2008 [32] and it was calculated by the following formula:

$$r_m^2 = r^2(1 - \sqrt{r^2 - r_o^2}) \quad (6)$$

Where r^2 is squared correlation coefficient between observed and predicted values and r^2_o is squared correlation coefficient between observed and predicted values with intercept value set to zero. A value of r^2_m is greater than *0.5* may be taken as an indicator of good external predictability.

According to Roy, 2007 [33] R^2_p is used to check the acceptability of the selected model. The parameter R^2_p which penalized the model R^2 for the difference between squared mean correlation coefficient (R^2_r) of the randomized models and squared correlation coefficient (R^2) of the non-randomized model. A value of R^2p should be greater than *0.5* may be taken as an indication of model acceptability and can be calculated by the following formula:

$$R_p^2 = R^2(1 - \sqrt{R^2 - R_r^2}) \quad (7)$$

Both the models have one outlier's compound 15, because its residual values exceeded twice the standard error of estimate. When these outliers have been removed from the data set, we have got highly significant model.

Regression analysis
The regression analysis is done using Material Studio version 7.0 software.

Model Validation
The reliability of the models was indicated by cross-validation experiments quantified with predictive Q^2cv. For leave one out (LOO) cross-validation a data point is removed (left-out) from the set, and the model refitted; the predicted value for that point is then compared to its actual value. This is repeated until each datum has been omitted once; the sum of squares of these deletion residuals can then be used to calculate Q^2, an equivalent statistic to R^2.

$$Q_{cv}^2 = 1 - \frac{\Sigma(Y_{pred} - Y_{exp})^2}{\Sigma(Y_{pred} - \bar{Y}_{av})^2} \quad (8)$$

Where Y_{pred} and Y_{exp} are the predicted and experimental biological activities of the left out compound, respectively, and Y_{av} is the average experimental activity of left-in compounds. In addition to the traditional LOO cross-validation. The Q^2 values can be considered a measure of the predictive power of a model: whereas R^2 can always be increased artificially by adding more parameters or descriptors, Q^2 decreases if a model is over parameterized [27,34] and is therefore a more meaningful summary statistic for predictive models [35].

Y-Randomization Test
The statistical significance of the relationship between the anti-HIV activity and chemical structure descriptors was further demonstrated by randomization procedure. Y-randomization is the most popular and probably the most powerful technique for the validation of a given QSAR model [36, 37]. In this approach, dependent variable vector

(anti-HIV activity in this study) is randomly shuffled and a new QSAR model is built using the original independent variables. The procedure is repeated number of times. If the new QSAR models have lower R^2 and Q^2 values for several trials (100 times in this study), then the given QSAR model is thought to be robust. Therefore, Y-randomization is useful to avoid any chance-comer correlation between dependent variable vector and independent variables. This Y-randomization was tested for model and low values of R^2 and Q^2 were observed (see Table 5).

Estimation of the Predictive Ability of a QSAR Model

According to Golbraikh and Tropsha group [38, 39] a QSAR model is considered predictive, if the following conditions are satisfied:

$R \geq 0.8$ (9)

$R^2 \geq 0.6$ (10)

If cross-validated R^2 $(Q^2) \geq 0.5$ (11)

If R^2 for external test set, R^2pred ≥ 0.6 (12)

Randomized R^2 value should be as low as to R^2 (13)

Randomized Q^2 value should be as low as to Q^2 (14)

$(r^2 - r_o^2)/r^2 < 0.1 \; and \; 0.85 \leq K \leq 1.15$ (15)

$(r^2 - r_o'^2)/r^2 < 0.1 \; and \; 0.85 \leq K' \leq 1.15$ (16)

$r^2_{m(Overall)}$ And R_p^2 are ≥ 0.5 (or at least near 0.5) (17)

If the standard deviation SEE is not much larger than standard deviation of the biological data.

If its F value indicate that overall significance level is better than 95%.

If its confidence interval of all individual regression coefficients proves that they are justified at the 95% significance level. Equation has to be rejected, if the above mentioned statistical measures are not satisfied, the number of the variables in the regression equation is unreasonably large and standard deviation is smaller than error in the biological data [25].

RESULTS AND DISCUSSION

The statistical quality of the developed equations was judged by the parameters such as standard error of estimate (SE), Fisher ratio (F test), Root mean square error of cross validation (RMSECV), Root mean square error of prediction (RMSEP), Quality factor (Q) and predicted correlation coefficient of multiple determination (R^2_{pred}).

Model 1. One-variable mode. Successive regression analysis indicated that one-variable model Sum of E-State descriptors of strength for potential Hydrogen Bonds of path length 6 (SHBint6) as correlating descriptor is the best modelling the pIC50. This model is as follows:

$$pIC_{50} = 0.3400(SHBint6) + 1.0098$$

$$N = 18, R = 0.7735, R^2 = 0.5984, R_a = 0.5733, Q^2_{cv} = 0.2681, LOF = 0.3654, F = 23.8367, SE = 0.3062, Q = 2.5261, SSY = 1.4998, RMSECV = 0.2887, PRESS = 0.6898, RMSEP = 0.3714, R^2_{pred} = 0.5553 \quad (18)$$

One-parametric equation modeled for HIV-1 Protease inhibitory activity and has good correlation between biological activity and descriptor as indicated by R = 0.7735 and explains 57.33% variance in inhibition. Low standard deviation of the model demonstrates accuracy of the model. The model showed overall significance level better than 95%, with the F = 23.8367. The positive coefficient of SHBint6 indicates that the increase in its magnitude will enhance the activity (pIC50). Cross validated value (Q^2 = 0.2681<0.6) reflects the poor predictive power of the model. The experimental and predicted values of activity data, correlation matrix are shown in Table 2 and 4, respectively, the plot of experimental vs. predicted property/ external test plot are shown in Figure 2a and 3a, respectively**.**

Model 2. Two-variable model. When (SHBint8) Sum of E-State descriptors of strength for potential Hydrogen Bonds of path length 8 and (Wlambda3.mass) Directional WHIM, weighted by atomic masses where added together, the model shows significant improvement all the parameters. The improved model is as follows:

$$pIC_{50} = 0.0613(SHBint8) - 0.0621(Wlambda3.mass) + 1.8561$$

$$N = 18, R = 0.8534, R^2 = 0.7284, R_a = 0.6921, Q_{cv}^2 = 0.6071, LOF = 0.2783, F = 20.1095, SE = 0.2601, Q = 3.2810, SSY = 1.0144, RMSECV = 0.2374, PRESS = 1.3689, RMSEP = 0.5232, R_{pred}^2 = 0.1175 \quad (19)$$

Model 2 is a two parametric equation modeled and was considered to be a poor predictor of activity because of its poor validation (R^2_{pred} = 0.1175<0.6) although the coefficient of determination (R^2 = 0.7284) is higher as compared with model 1. Further, the descriptors (SHBint8)) Sum of E-State descriptors of strength for potential Hydrogen Bonds of path length 8 and (Wlambda3.mass) Directional WHIM, weighted by atomic mass are found to be highly correlated to each other as shown in the correlation matrix (Table 4). The experimental and predicted values of activity data is shown in Table 2, the plot of experimental vs. predicted property and external test set are shown in Figure 2b and 3b, respectively**.** The correlation coefficient between SHBint8 and Wlambda3.mass is 0.853 which is outside the acceptable range (< 0.8). The above model indicates that decrease in Wlambda3.mass and increase in SHBint8 will improve the activity (pIC_{50}) values. However, if the descriptor SHBint8 and Wlambda3.mass is replaced by minHBint6 and MLFER_BH is added to model 1, Model 3 is derived.

$$pIC_{50} = 0.5912(SHBint6) - 0.4315(minHBint6) - 1.0201(MLFR\_BH) + 2.7221$$

$$N = 18, R = 0.9417, R^2 = 0.8869, R_a = 0.8627, Q_{cv}^2 = 0.6210, LOF = 0.1314, F = 36.6071, SE = 0.1767, Q = 5.3294, SSY = 0.4222, RMSECV = 0.1532, PRESS = 0.1657, RMSEP = 0.1820, R_{pred}^2 = 0.8932 \quad (20)$$

Model 3. When (minHBint6) Minimum E-State descriptors of strength for potential Hydrogen Bonds of path length 6 and (MLFR_BH) Overall or summation solute hydrogen bond basicity is added to model 1, improves the quality of prediction as indicated by much better statistical parameters in Table 6. (R^2 = 0.8869, Q^2 = 0.6210). A plot of experimental *vs* predicted property and the external test set plot are shown in Figure 2c and 3c**,** respectively**.** The negative coefficient of minHBint6 and MLFER_BH revealed that the molecular flexibility of 5,6-dihydro-2-pyrones is detrimental to the activity.

Model 4

Addition Solvation energy (Solvation E) to the above three-parametric model yielded a four-parametric model. A drastic improvement in variance is observed.

$$pIC_{50} = 0.6363(SHBint6) - 0.5368(minSHBint6) - 1.9624(MLFER\_BH) - 0.0111(Solvation\ E.) + 3.5827$$

$$N = 18, R = 0.9768, R^2 = 0.9541, R_a = 0.9400, Q_{cv}^2 = 0.7440, LOF = 0.0610, F = 67.6020, SE = 0.1148, Q = 8.5087, SSY = 0.1713, RMSECV = 0.0976, PRESS = 0.1139, RMSEP = 0.1509, R_{pred}^2 = 0.9266 \quad (21)$$

Model 4 is the best model since it shows best correlation coefficient R = 0.9768 and explains 95.41% variance in inhibition. Further, smaller standard error of estimate, higher *F* and leave-one-out (LOO) cross validated value (Q2 = 0.744>0.6) demonstrates satisfactory predictive ability of the model. The statistical parameters of Model 4 are shown in Table 6**.** This model indicates that Solvation energy are highly correlated to the activity. The negative sign of the coefficient of minSHBint6, MLFER_BH and Solvation E., indicates that inhibitory activity increases as the descriptors decreases. A closed look at model 4 reveals that Solvation energy play dominant role in exhibiting the activity. They belong to 2D and 3D autocorrelation category. The brief description of the descriptors is given in Table 3. The experimental and predicted values of activity data is shown in Table 2**.** The plot of experimental vs. predicted property and external test set plot are shown in Figure 2d and 3d**,** respectively**.**

Cross validation

The cross validation analysis was performed using leave one out (LOO) method [27, 34] in which one compound is removed from the data set and the activity is correlated using the rest of the data set. The cross-validated R^2 in each case was found to be very close to the value of R^2 for the entire data set and hence these models can be termed as statistically significant.

Cross validation provides the values of PRESS, SSY and R^2cv and PSE/RMSEP from which we can test the predictive power of the proposed model. It is argued that PRESS, is a good estimate of the real predictive error of the model and if it is smaller than SSY the model predicts better than chance and can be considered statistically significant. Also, if the PRESS value is transformed into a dimension-less term by relating it to the initial sum of squares, we obtain R^2cv, i.e., the complement to the traces on of unexplained variance over the total variance. The PRESS and R^2cv have good properties. Still, for practical purposes of end users the use of the square root of PRESS/N, which is called the predictive square error (PSE/RMSEP), is more directly related to the uncertainty of the predictions. The PSE/RMSEP values also support our results. The calculated cross-validated parameters confirm the validity of the models. All the requirements for an ideal model have been fulfilled by model number 4, which made it the best model. R^2_a takes into account the adjustment of R^2. R^2 is a measure of the percentage explained variation in the dependent variable that takes into account the relationship between the number of cases and the number of independent variables in the regression model, whereas R^2 will always increase when an independent variable is added. R^2_a will decrease if the added variable does not reduce the unexplained variable enough to offset the loss of decrease of freedom.

Comparison with other QSAR Studies

Agrawal and coworkers [13] proposed QSAR-based multiple regression analysis method for anti-HIV-1 protease inhibitors activity of 24 5,6-dihydro-2-pyrones derivatives. They developed QSAR-based models on the entire data set of 22 compounds and found that the best model involves 4 correlating descriptors with statistical quality given by R = 0.9513 and 0.9519, F-value of 40.470 and 40.994 respectively. It is interesting to compare our results with the results of Agrawal and coworkers. Our model is with four correlating parameters having the R = 0.9768, R^2 = 0.9541 and F-value of 67.602 in the training set and R^2_{pred} = 0.9266 in the test set. GFA technique is better than the previously reported one by Agrawal et al.; in addition to that we also applied other statistical parameters which are better than the Agrawal and coworkers.

CONCLUSION

This studies obtained a multivariate QSAR model for a set of 5,6-dihydro-2-pyrones that have the capacity of inhibiting HIV-1 protease inhibitors. The LOO cross validation method, the Y-randomization technique and the external validation indicated that the model is significant, robust and has good internal and external predictability. From the results it is concluded that 5,6-dihydro-2-pyrones as anti-HIV-1 protease inhibitors derivatives can be modeled using a four-parametric model which contains variety of molecular descriptors including Sum of E-State descriptors of strength for potential Hydrogen Bonds of path length 6, Minimum E-State descriptors of strength for potential Hydrogen Bonds of path length 6, Overall or summation solute hydrogen bond basicity and Solvation Energy. The results obtained is useful for pharmaceutical as well as medicinal chemists to synthesis new drugs having still better anti-HIV-1 potential.

Table 2. Experimental and Predicted values of activity data

Name	LogIC50	Eq1: predicted values	Eq1: residual values	Eq2: predicted values	Eq2: residual values	Eq3: predicted values	Eq3: residual values	Eq4: predicted values	Eq4: residual values
1	1.544	1.00981	0.53419	1.538427	0.005573	1.45518	0.08882	1.586565	-0.04257
2	1.5185	1.00981	0.50869	1.69202	-0.17352	1.20016	0.31834	1.283918	0.234582
3	0.8325	1.00981	-0.17731	0.508488	0.324012	0.984922	-0.15242	0.89199	-0.05949
4	0.8195	0.685429	0.134071	0.744298	0.075202	0.867186	-0.04769	0.850133	-0.03063
5	1.176	1.00981	0.16619	0.823853	0.352147	1.290947	-0.11495	1.284637	-0.10864
6	1.0413	1.00981	0.03149	1.060881	-0.01958	1.153236	-0.11194	1.218004	-0.1767
7	1.3802	1.00981	0.37039	1.032644	0.347556	1.185878	0.194322	1.258113	0.122087
9	1.5051	2.014505	-0.50941	1.654045	-0.14895	1.657961	-0.15286	1.54014	-0.03504
10	1.0792	1.00981	0.06939	0.737553	0.341647	0.937999	0.141201	0.931614	0.147586
12	0.3979	0.679356	-0.28146	0.693084	-0.29518	0.562388	-0.16449	0.486468	-0.08857
13	0.4913	0.683539	-0.19224	0.861803	-0.3705	0.596996	-0.1057	0.522129	-0.03083
14	0.602	0.68311	-0.08111	0.696135	-0.09414	0.596794	0.005206	0.553253	0.048747
16	0.8061	1.00981	-0.20371	0.928993	-0.12289	0.467741	0.338359	0.852914	-0.04681
18	0.6532	1.00981	-0.35661	0.631527	0.021673	0.726841	-0.07364	0.577896	0.075304
19	2.0791	1.720489	0.358611	1.759877	0.319223	2.0791	0	2.0791	0
21	0.5563	0.685845	-0.12955	0.693528	-0.13723	0.596039	-0.03974	0.512763	0.043537
22	0.6334	0.682861	-0.04946	0.724226	-0.09083	0.612999	0.020401	0.601704	0.031696
23	0.5051	0.697272	-0.19217	0.839318	-0.33422	0.648333	-0.14323	0.589359	-0.08426
					Test Set				
8	1.602	2.0703	-0.4683	1.743631	-0.14163	1.651309	-0.04931	1.492929	0.109071
11	0.2304	0.680289	-0.44989	0.906479	-0.67608	0.56283	-0.33243	0.502392	-0.27199
17	0.5682	1.0098	-0.4416	1.447847	-0.87965	0.68292	-0.11472	0.538205	0.029995
20	0.6127	0.69409	-0.08139	0.852429	-0.23973	0.614197	-0.0015	0.500979	0.111721
24	0.4313	0.689029	-0.25773	0.677249	-0.24595	0.630182	-0.19888	0.552359	-0.12106
15	0.1461	OUTLIER							

Table 3. List of some descriptors used in this studies

Descriptor	Description	Class
ALogP	Ghose-Crippen LogKow	2D
AMR	Molar refractivity	2D
apol	Sum of the atomic polarizabilities (including implicit hydrogens)	2D
nH	Number of hydrogen atoms	2D
nN	Number of nitrogen atoms	2D
nO	Number of oxygen atoms	2D
ATSc1	ATS autocorrelation descriptor, weighted by charges	2D
ATSc3	ATS autocorrelation descriptor, weighted by charges	2D
ATSc5	ATS autocorrelation descriptor, weighted by charges	2D
ATSm1	ATS autocorrelation descriptor, weighted by scaled atomic mass	2D
ATSm5	ATS autocorrelation descriptor, weighted by scaled atomic mass	2D
BCUTw-1l	nhigh lowest atom weighted BCUTS	2D
BCUTc-1l	nhigh lowest partial charge weighted BCUTS	2D
BCUTc-1h	nlow highest partial charge weighted BCUTS	2D
BCUTp-1l	nhigh lowest polarizability weighted BCUTS	2D
BCUTp-1h	nlow highest polarizability weighted BCUTS	2D
SHBint6	Sum of E-State descriptors of strength for potential Hydrogen Bonds of path length 6	2D
SHBint8	Sum of E-State descriptors of strength for potential Hydrogen Bonds of path length 8	2D
minHBint6	Minimum E-State descriptors of strength for potential Hydrogen Bonds of path length 6	2D
MLFER_BH	Overall or summation solute hydrogen bond basicity	2D
Wlambda3.mass	Directional WHIM, weighted by atomic masses	3D
P-Area(75)	Polar area corresponding to absolute values of the electrostatic potential greater than 75	3D
Acc.P-Area(75)	Accessible Polar area corresponding to absolute values of the electrostatic potential greater than 75	3D
MaxEIPot	Maximum values of the electrostatics potential	3D
LogP	Partition coefficient	3D
Solvation.E	Solvation Energy	3D
EHOMO	Highest occupied molecular orbital energy	3D
Dipole M	Dipole Moment	3D
ELUMO	Lowest unoccupied molecular orbital energy	3D
Ovality	Ovality	2D

Table 4. Correlation matrix showing correlation among various physic-chemical parameters and inhibitory activity

	LogIC50	SHBint6	SHBint8	minHBint6	Wlambda3.mass	MLFER_BH	Solvation.E
LogIC50	1						
SHBint6	0.7735	1					
SHBint8	0.6947	0.7115	1				
minHBint6	0.4273	0.7442	0.3082	1			
Wlambda3.mass	-0.492	-0.224	0.0046	-0.3122	1		
MLFER_BH	-0.353	-0.113	0.0749	-0.4726	0.5711	1	
Solvation.E	0.2172	0.0175	-0.039	0.23717	-0.2523	-0.8553	1

Table 5. Y-randomization tested for the model

Model	R	R^2_{yrand}	Q^2_{yrand}
1	0.2018	0.0608	-0.2517
2	0.3514	0.1478	-0.3400
3	0.3988	0.1806	-0.5794
4	0.4355	0.2097	-1.7048

Figure 2a. Correlation between the predicted pIC_{50} and the Observed PIC_{50} by Eq (1)

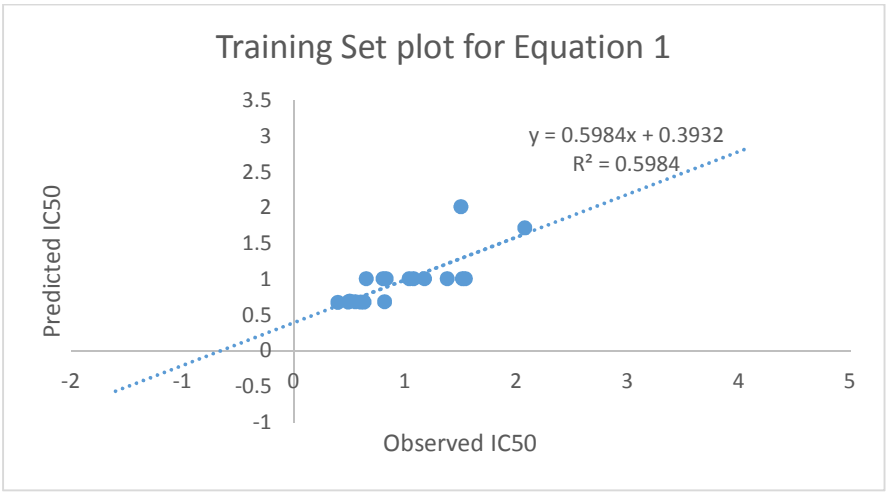

Figure 2. Correlation between the predicted pIC_{50} and the Observed PIC_{50} by Eq (2)

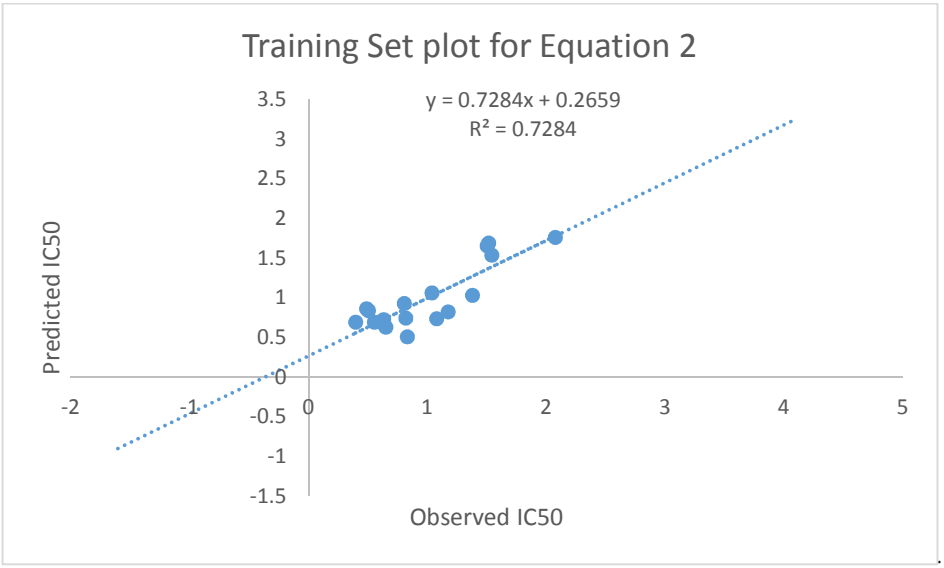

Figure 2c. Correlation between the predicted pIC_{50} and Observed pIC_{50} by Eq (3)

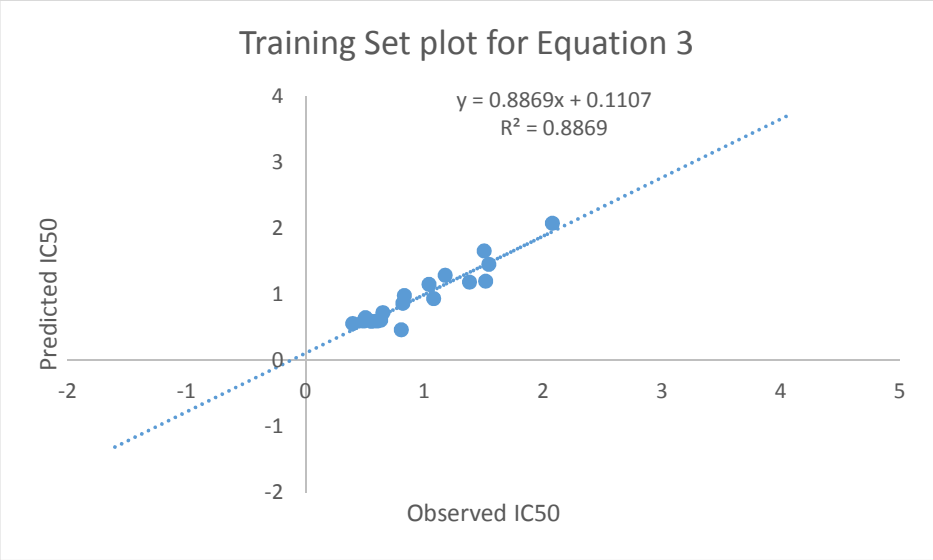

Figure 2d. Correlation between the predicted pIC_{50} and the Observed PIC_{50} by Eq (4)

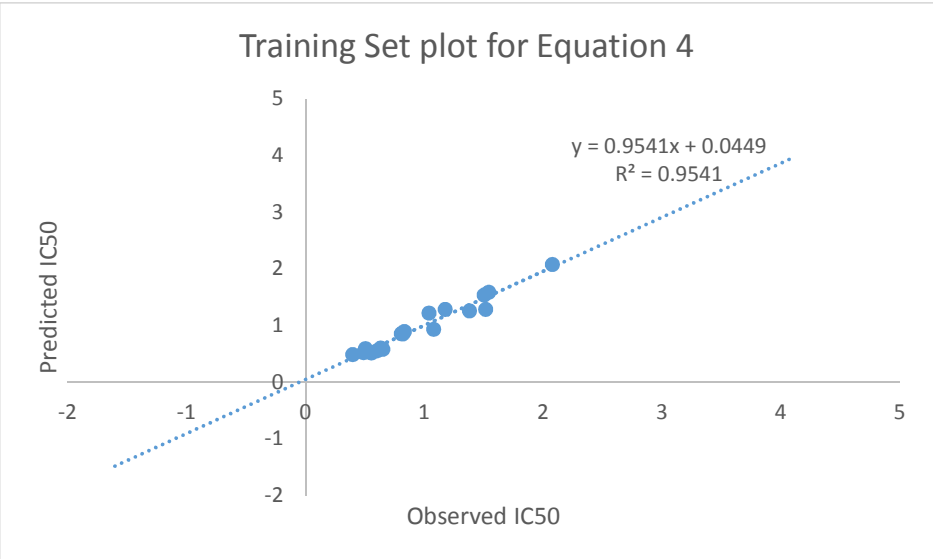

Figure 3a. Correlation (external set) between the predicted pIC_{50} and Observed pIC_{50} by Eq (1)

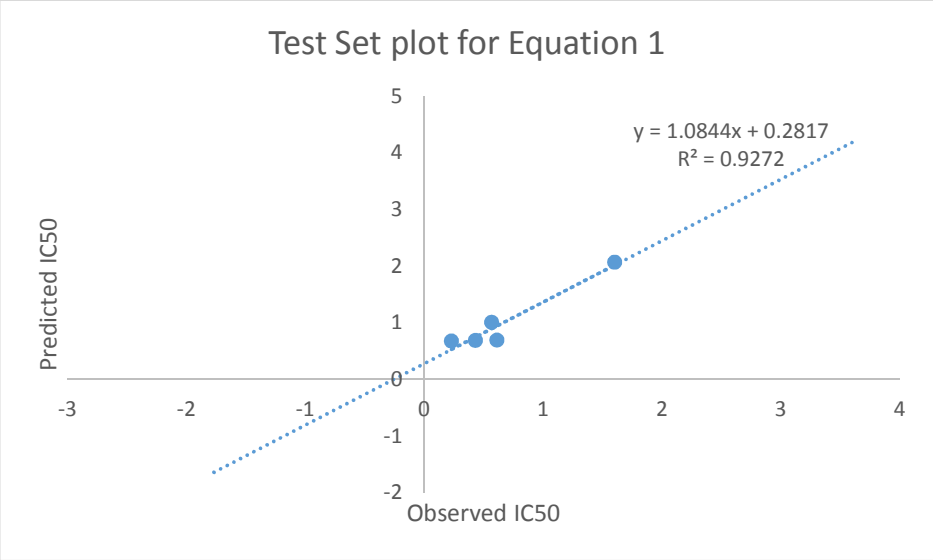

Figure 3b. Correlation (external set) between the predicted pIC_{50} and Observed pIC_{50} by Eq (2)

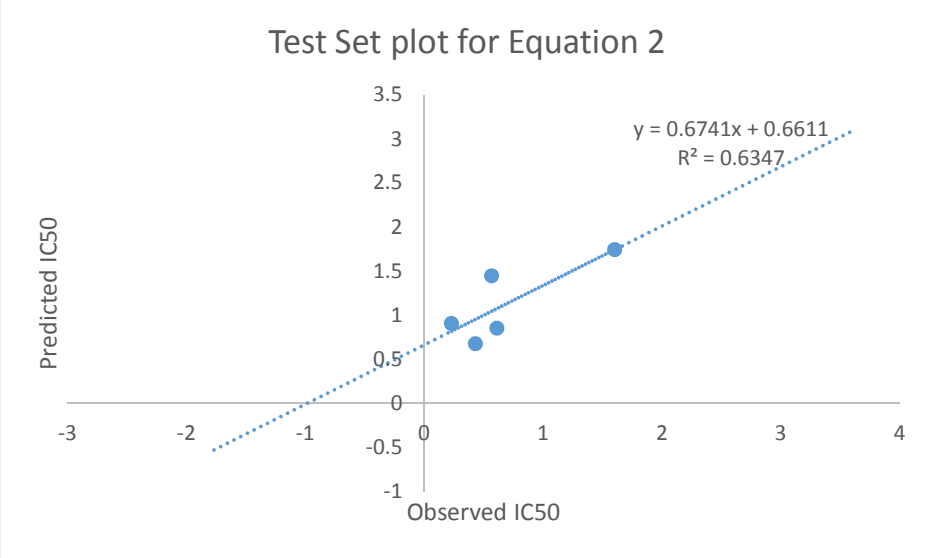

Figure 3c. Correlation (external set) between the predicted pIC_{50} and Observed pIC_{50} by Eq (3)

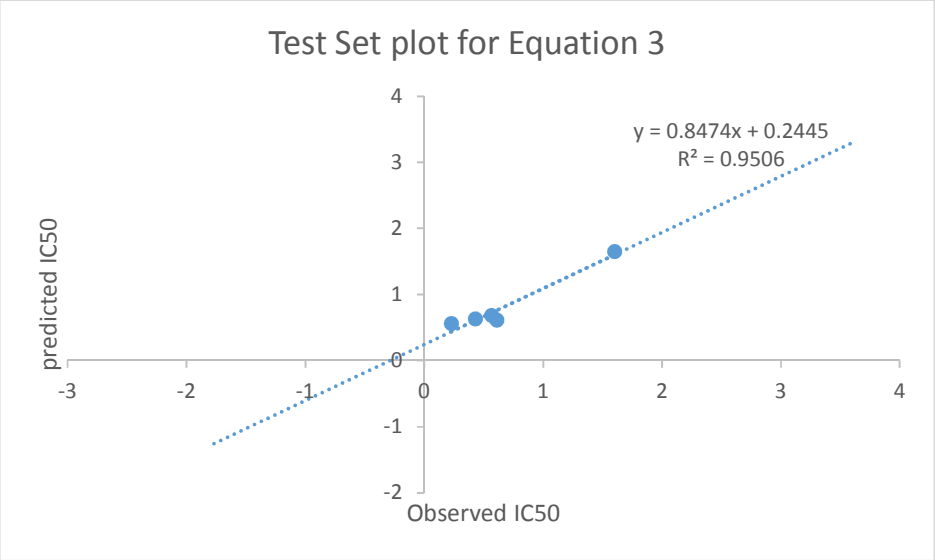

Figure 3d. Correlation (external set) between the predicted pIC_{50} and Observed pIC_{50} by Eq (4)

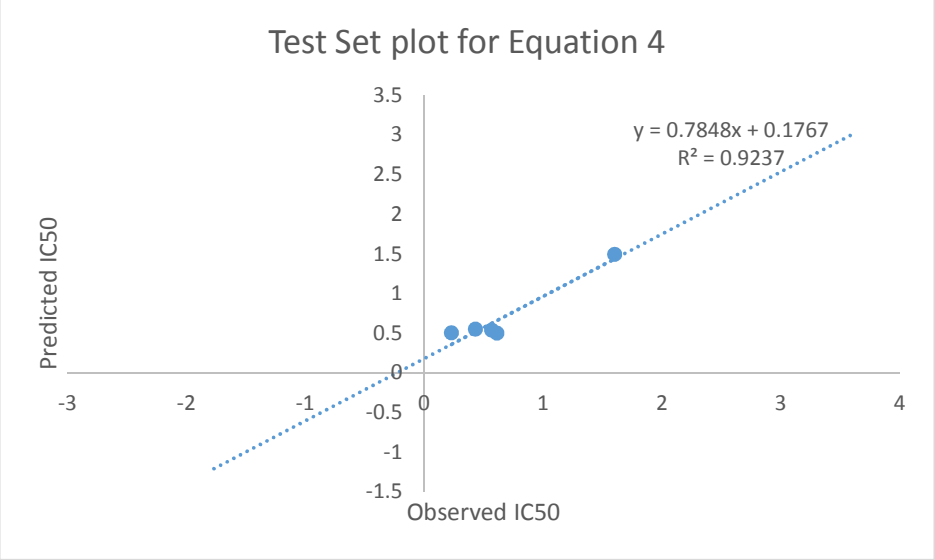

Table 6. Predicted values of the test set (external cross-validation) and results of statistical parameters

	K	K'	$/r_o^2 - r_o'^2/$	$\frac{r^2 - r_o^2}{r^2}$	$\frac{r^2 - r_o'^2}{r^2}$	$r^2_{m(test)}$	R_p^2	$r^2_{m(overall)}$
Threshold values	$\mathbf{0.85 \leq K \leq 1.15}$	$\mathbf{0.85 \leq K' \leq 1.15}$	< 0.3	< 0.1	< 0.1	$\geq \mathbf{0.5}$	$\geq \mathbf{0.5}$	$\geq \mathbf{0.5}$
Model 1	**0.7092**	**1.3613**	**0.5502**	**0.0963**	**0.0370**	**0.65**	**0.5776**	**0.335**
Model 2	**0.6494**	**1.3241**	**0.8**	**0.108**	**1.3723**	**0.080**	**0.6638**	**0.296**
Model 3	**0.8896**	**1.0878**	**0.0620**	**0.1**	**0.053**	**0.631**	**0.8035**	**0.737**
Model 4	**0.9599**	**1.0078**	**0.0435**	**0.074**	**0.027**	**0.685**	**0.8541**	**0.863**

Acknowledgments
Authors would like to acknowledge the anonymous referees for their useful comments that helped to improve the quality of the manuscript. We also greatly acknowledge David Ebuka Arthur for sending valuable information for the development of this work.

REFERENCES

[1]JS. Berns; N Kasbekar, Clin. *J. Amer. Soc. Neph*., **2006**, 1, 117–129.
[2]JM Coffin. Molecular biology of HIV. In Evolution of HIV. Edited by Crandall KA. Baltimore, MD: Johns Hopkins University Press; **1999**:3–39
[3]J Amborzia; JA. Levy, 3d ed, ed. KK Holmes; PF Sparling; PA Mardh; SM Lemon; WE Stamm; P Piot; JN Wasserheit, New York: McGraw-Hill, **1998**, 251–58.
[4]RA Katz; AM Skalka, *Ann. Rev. Biochem*., **1994**, 63, pp 133–173.
[5]P Reddy; J Ross, *Formulary*, **1999**, 34, 567–577.
[6]R Mishra; PS Mishra; Shuaib. *World J. of Pharm. & Pharm. Sci*. **2013**, 2(6), 6312-6324.
[7]PJ Ala; EE Huston; RM Klabe; PK Jadhav; PY Lam; CH Chang, *Biochem.*, **1998**, 37, 15042–15049.
[8]A Wlodawer; J Vondrasek, *Ann. Rev. of Biophys.s and Biomol. Struct*., **1998**, 27, 249–84.
[9]S Bihani; A Das; V Prashar; JL Ferrer; MV Hosur, *Proteins Struct. Funct. & Bioinfor*., **2009**, 74, pp 594–602.
[10]G Klebe. *J. Mol. Med*., **2000**, 78(5), 69-81.
[11]C Hansch; A Kurup; R Garg; H Gao. *Chem. Rev*., **2001**, 101, 619–672.
[12]A Maryam; F Ramezani; M Elyasi; HS Aliabadi; M Amanlou, *DARU J. Pharm. Sci*., **2015**, 23, 29.
[13]VK Agrawal; J Singh; KC Mishra; PV Khadikar; YA Jaliwala, *J. ARKIVOC*; **2006**, 2, 162-177.
[14]Wavefunction. Inc. Spartan'14, version 1.1.2, **2013**, Irvine, California, USA.
[15]Yap Chun Wei. Inc. PaDEL-Descriptor, version 2.18, **2011**. A software to calculate molecular descriptors and fingerprints. http://padel.nus.edu.sg
[16]MJS Dewar; EG Zoebisch; EF Healy and JJP Stewart. *J. Am. Chem. Soc*., **1985**, 107, 3902-9.
[17]GF Fadhil; HA Radhy; A Perjessy; M Samalíkova; E Kolehmainen; WMF Fabian; K Laihia; Z Sustekova, *Molecules;* **2002**, 7, 833-839.
[18]R Yadav; S Nandi, *J. Comput. Methods Mol. Des*., **2014**, 4 (3): 92-105.
[19]R Hadanu; Syamsudin, *Asian J. Chem.*, **2013**, 25(11), 6136-614.
[20]Doreswany; CM. Vastrad, *J. Advanced Bioinfor. Appl. & Res*., **2012**, 3(3), 379-390.
[21]LF Motta; WP Almeida, *Int. J. Drug Dis*., **2011**, 3(2), 100-117.
[22]D. Rogers; A. J. Hopfinger, *J. Chem. Infor. Comp. Sci*., **1994**, 4, 854-866.
[23]RH Myers. Classical and modern regression application. 2nd edition. Duxbury press. CA. **1990**.
[24]R. Veerasamy; H. Rajak; A. Jain; S Sivadasan; CP Varghese; R.K. Agrawal, In. *J. Drug Des. & Discov*., **2011**, 2(3), 511-519.
[25]R. Veerasamy; S Ravichandran; A Jain; H Rajak; R.K. Agrawal, *Digest J. Nano. & Biostruct.*, **2009**, 4, 823-834.
[26]P Gramatica, *J. QSAR Comb. Sci*., **2007**, 5, 694-701.
[27]RD III. Crammer; JD Bunce; DE Patterson, *Quant. Struct. Act. Relat*.; 1998, 7, pp 18-25.
[28]C L Waller; TI Opera; A Giollitti; GR Marshal, *J. Med. Chem.*; **1993**. 36, 4152-4160.
[29]L. Pogliani, *Amino Acids*, **1994**, 6(2), 141–153.
[30]L. Pogliani, *J. Phys. Chem*., **1996**, 100(46), 18065–18077.
[31]S Chaterjee; AS Hadi ; B Price, Regression analysis by examples, 3rd Ed. Wiley; New York. **2000**.
[32]K Roy; S Paul, *QSAR Comb. Sci*., **2008**, 28, 406-425.
[33]K Roy, *Expert Opin. Drug Discov*.; **2007,** 2, pp 1567-1577.
[34]BL Podlgar; DM Ferguson. *Drug Des. Discov*. **2000**, 7 (1), 4–12.
[35]M Fernandez; J Caballero, *Bioorg. Med. Chem*., **2006**, 14, 280–294.

[36]VH Masand; RD Jawarkar; KN Patil; DT Mahajan; TB Hadda; GH Kurhade, *Der Pharm. Lett.*, **2010**, 2(6):350–357.
[37]VH Masand; RD Jawarkar; KN Patil; GM Nazerruddin; SO Bajaj, *Org Chem Indian J*. **2010**, 6(1):30–38.
[38]A Golbraikh; A. Tropsha. *J. Comp. Aided. Mol. Design*, **2002**, 16, 357-366.
[39]PP Roy; S Paul; I Mitra; K Roy, *Molecules*, **2009,** 14; pp 1660-1701.