# Trends in Bioinformatics

ISSN 1994-7941

science alert

ANSInet
an open access publisher
http://ansinet.com

OPEN ACCESS Trends in Bioinformatics

ISSN 1994-7941
DOI: 10.3923/tb.2016.1.13

# Research Article
# Generation of 2D-QSAR Model for Angiogenin Inhibitors: A Ligand-Based Approach for Cancer Drug Design

[1]Joseph Christina Rosy, [2]Saminathan Balamurali, [3]John Asnet Mary, [3]Rajaiah Shenbagarathai and [1]Krishnan Sundar

[1]Department of Biotechnology, Kalasalingam University, Krishnankoil, 626126, Tamilnadu, India
[2]Department of Computer Applications, Kalasalingam University, Krishnankoil, 626126, Tamilnadu, India
[3]PG and Research Department of Zoology and Biotechnology, Lady Doak College, Madurai, 625002, Tamilnadu, India

## Abstract

**Background and Objective:** Angiogenin is a monomeric protein which has been considered as an important factor in angiogenesis. Recent studies on angiogenin proved that it is an ideal drug target for treating cancer and vascular dysfunctions. The present study aimed to develop a Quantitative Structure Activity Relationship (QSAR) model with small molecules of angiogenin inhibitors. **Methodology:** The small molecule inhibitors were divided into training and test sets to build the QSAR model. Multiple Linear Regression (MLR) and Partial Least Square (PLS) methods were used to develop QSAR models. **Results:** In the MLR model, the descriptors generated for the compounds showed multicollinearity and resulted in a mono-parametric equation. The model generated by PLS satisfied both internal and external cross validation parameters. The predicted model showed the positive contribution of ring atoms and donor hydrogen bonds to the activity. **Conclusion:** As these parameters are reported to be crucial for biological activity of drugs, it can be used to do develop effective small molecule drug candidates for angiogenin.

Key words: Angiogenesis, descriptor, multicollinearity, cross validation, regression

Received: August 14, 2015 Accepted: October 26, 2015 Published: June 15, 2016

Citation: Joseph Christina Rosy, Saminathan Balamurali, John Asnet Mary, Rajaiah Shenbagarathai and Krishnan Sundar, 2016. Generation of 2D-QSAR model for angiogenin inhibitors: A ligand-based approach for cancer drug design. Trends Bioinform., 9: 1-13.

Corresponding Author: Krishnan Sundar, Department of Biotechnology, Kalasalingam University Krishnankoil, 626126, Tamilnadu, India
Tel: +91-4563289042 Fax: +91-4563289322



Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

## INTRODUCTION

Angiogenesis is a natural process involved in the formation of new blood vessels which promises greater importance in normal physiological functions. However, angiogenesis has also been known for its devastating effects under pathological conditions such as cancer and vascular dysfunction[1]. Clinically, it has been ascertained that tumor angiogenesis is closely connected to metastasis in many kinds of human cancers[2,3]. Hence, blocking angiogenesis could be an additional therapeutic approach to treat cancer[4]. Advancements in technology has uncovered the role of various angiogenic factors involved in angiogenesis such as TGF$^2$-α, TGF-β, aFGF and bFGF[5,6]. Angiogenin, one among the various factors of angiogenesis, is considered as an ideal drug target for blocking angiogenesis[7]. It is a single chain protein containing 123 amino acids and is a homolog of bovine pancreatic ribonuclease A[8]. It has been reported that it contributes to the ribonucleotlytic cleavage of both 28S and 18S RNA[9] and has distinctive catalytic and cell-binding domains which are essential for its angiogenic property[10,11]. The role of angiogenin in neovascularization has been shown to be interacting with endothelial and smooth muscle cells to induce various cellular responses like cell migration, invasion, proliferation and tubular structure formation. The detailed mechanism of action of angiogenin-induced angiogenesis involves ribonucleolytic activity, cell basement membrane degradation, activation of signal transduction and nuclear translocation[7]. From these observations it is quite clear that the inhibition of angiogenin would be an effective therapeutic strategy for cancer. Angiogenesis inhibitors are becoming a new class of drugs which target one of these several angiogenic factors[12]. Although, angiogenin was reported to be regulated by endogenous ribonuclease inhibitors[13], synthetic inhibitors are needed for the treatment of diseases caused by pathological neovascularization. Search for the inhibitors of angiogenin was started even before two decades. Many angiogenic inhibitors have been reported so far which includes proteins, oligonucleotides, peptides and nucleotides[14,15]. The main drawback with the nucleotide inhibitors is that they had very high $K_i$ values which were greater than or equal to 500 μM under physiological conditions[16]. This had led to search for low molecular weight inhibitors. Shapiro *et al.*[17] have patented many small molecular weight compounds as effective inhibitors of angiogenin.

Quantitative Structure-Activity Relationship (QSAR) is one of the promising areas of research in medicinal chemistry and chemometrics arena. It aims to derive relationship between the structural features or descriptors of the chemical entities and their own biological activity through linear or nonlinear mathematical equation. Thus QSAR studies provide useful information that how the structural features of a chemical or drug molecule influences the biological activity. *In silico* QSAR has become one of the advantageous approaches for bioactivity evaluation as compared to experimental testing[18]. The success of QSAR model depends on the quality of the input data, selection of appropriate descriptors and statistical methods to validate the developed model[19]. In this study, we report the development of a 2D-QSAR model for angiogenin inhibitors which has not been explored so far.

## MATERIALS AND METHODS

**Data set:** The inhibitor compounds of angiogenin and their biological activities ($K_i$) were collected from the literature[17]. They were used for QSAR analysis (Table 1). These compounds showed significant diversity in their structure and activity profiles. The $K_i$ values (given in μM unit) were converted into their molar units and then into their negative logarithmic scale ($pK_i = -\log K_i$). The $K_i$ values collected from the literature were in the range of 3-500 μM. Hence the insignificant compounds with higher $K_i$ values were removed from the data set and the compounds in the activity range of 3-85 μM were chosen for the present study. The data set comprises of 30 compounds of these, 75% of the compounds were assigned to training set and 25% of the compounds were assigned to test set. The grouping of compounds into training (23) and test set (7) was done by sorting out the compounds with their biological activities (that is both set of compounds should span the entire activity range). A univariate statistics for the training and test set compounds was generated to check the correctness of selection criteria which is shown in Table 2. The distribution of compounds within the experimental activity range is shown in Fig. 1.

**Computational data:** The compounds were sketched with Maestro (Maestro, version 9.1) provided by Schrodinger, LLC, NY. They were then converted into their 3D-structures using Ligprep (LigPrep, version 2.4, Schrodinger, NY). Addition of implicit hydrogen atoms, geometry optimization and energy minimization were carried out using MacroModel by applying

Table 1: List of angiogenin inhibitor compounds chosen for the study along with their $K_i$ and $pK_i$ values

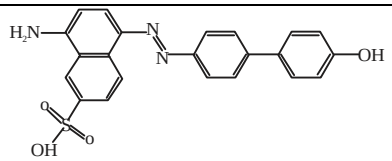
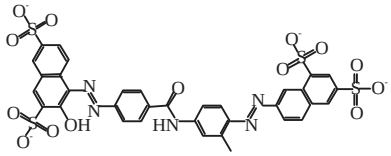
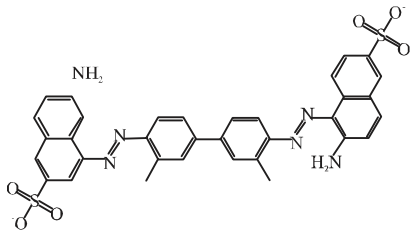
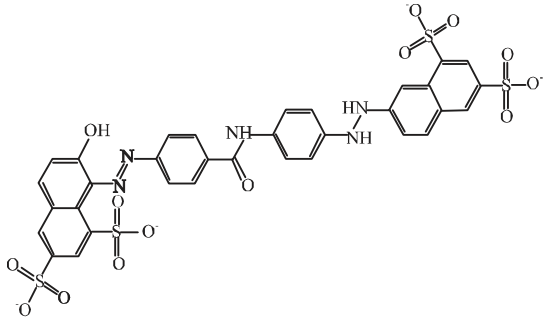
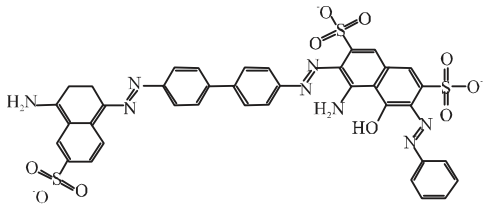
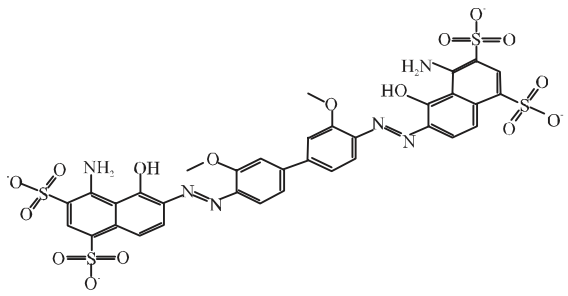
| Compound I.D | Structures | $K_i$ (μM) | $K_i$ (M) | $pK_i$ |
|---|---|---|---|---|
| NCI-65828 |  | 81 | 0.000081 | 4.091515 |
| NCI-65845 |  | 3 | 0.000003 | 5.522879 |
| NCI-242027[a] |  | 5 | 0.000005 | 5.30103 |
| NCI-65841[a] |  | 5 | 0.000005 | 5.30103 |
| NCI-79596 |  | 5 | 0.000005 | 5.30103 |
| NCI-9617 |  | 5 | 0.000005 | 5.30103 |

Table 1: Continue

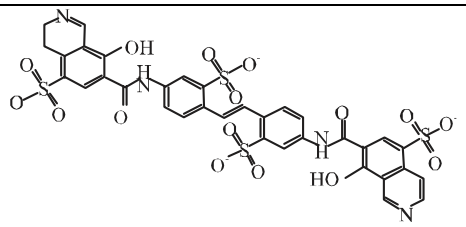
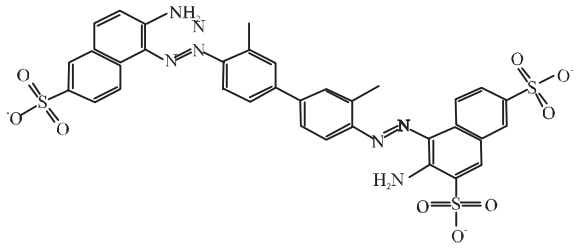
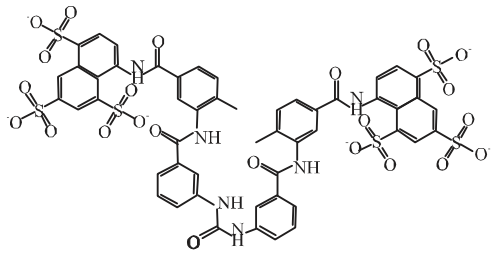
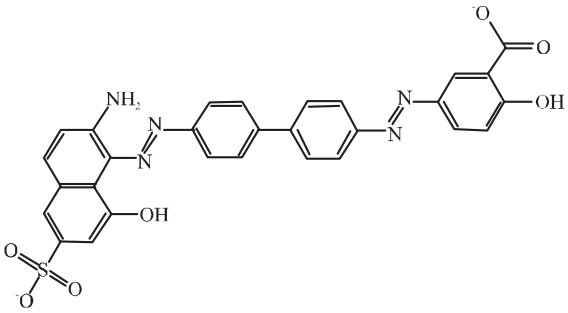
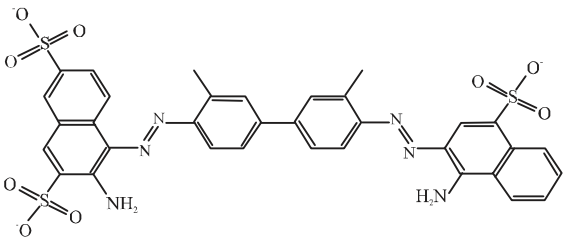
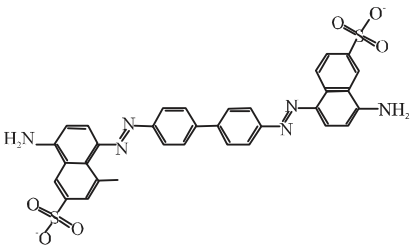
| Compound I.D | Structures | $K_i$ (μM) | $K_i$ (M) | $pK_i$ |
|---|---|---|---|---|
| NCI-665534-P |  | 5 | 0.000005 | 5.30103 |
| NCI-16224 |  | 5.5 | 5.5E-06 | 5.259637 |
| Sigma-Suramin |  | 10 | 0.00001 | 5 |
| NCI-N-73358 |  | 14 | 0.000014 | 4.853872 |
| NCI-7815 |  | 14 | 0.000014 | 4.853872 |
| NCI-45618[a] |  | 15 | 0.000015 | 4.823909 |

Table 1: Continue

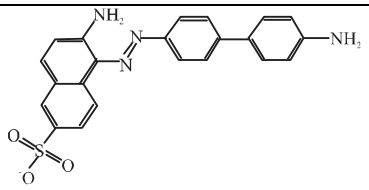
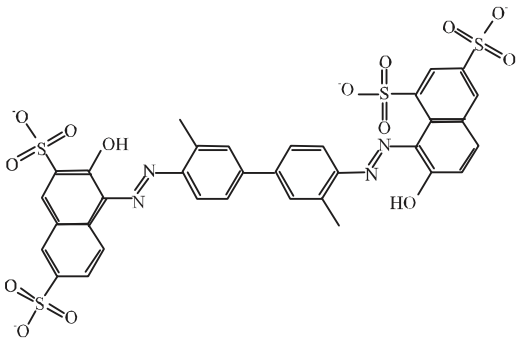
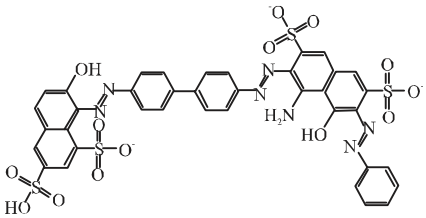
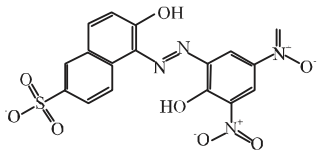
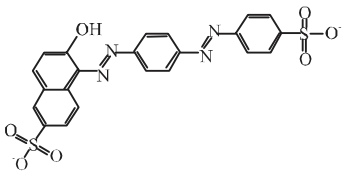
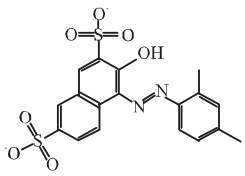
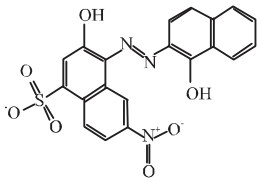
| Compound I.D | Structures | $K_i$ (μM) | $K_i$ (M) | $pK_i$ |
|---|---|---|---|---|
| NEW[a] |  | 20 | 0.00002 | 4.69897 |
| NCI-65568 |  | 23 | 0.000023 | 4.638272 |
| NCI-79741 |  | 23 | 0.000023 | 4.638272 |
| NCI-65820 |  | 25 | 0.000025 | 4.60206 |
| NCI-58047 |  | 36 | 0.000036 | 4.443697 |
| Sigma-xylidene |  | 49 | 0.000049 | 4.309804 |
| Sigma-eriochrome[a] |  | 50 | 0.00005 | 4.30103 |

Table 1: Continue

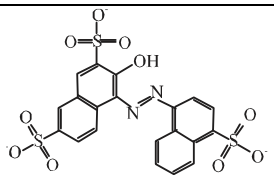
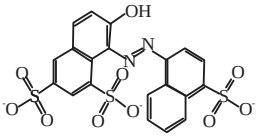
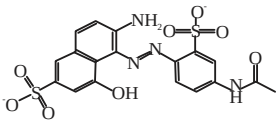
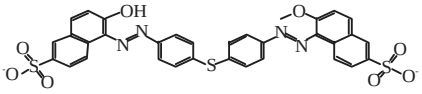
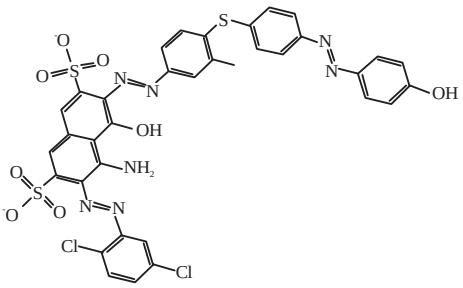
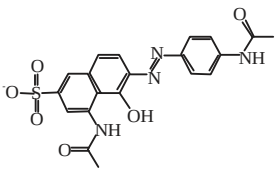
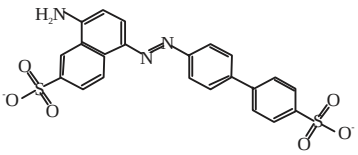
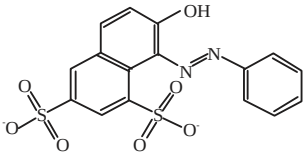
| Compound I.D | Structures | $K_i$ (μM) | $K_i$ (M) | $pK_i$ |
|---|---|---|---|---|
| Sigma-Amaranth[a] |  | 60 | 0.00006 | 4.221849 |
| Sigma-newcoccine |  | 69 | 0.000069 | 4.161151 |
| Sigma-acidred |  | 70 | 0.00007 | 4.154902 |
| Sigma-acidviolet |  | 71 | 0.000071 | 4.148742 |
| NCI-75661 |  | 76 | 0.000076 | 4.119186 |
| NCI-73416 |  | 77 | 0.000077 | 4.113509 |
| NCI-724225[a] |  | 81 | 0.000081 | 4.091515 |
| Sigma-orange G |  | 83 | 0.000083 | 4.080922 |

Table 1: Continue

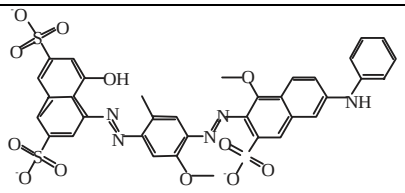
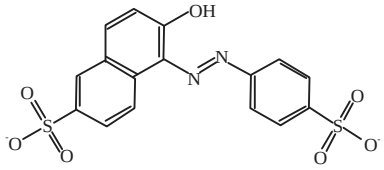
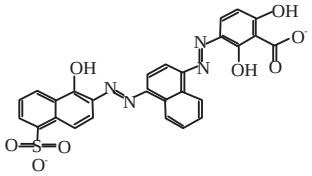
| Compound I.D | Structures | $K_i$ (μM) | $K_i$ (M) | $pK_i$ |
|---|---|---|---|---|
| NCI-47755 |  | 84 | 0.000084 | 4.075721 |
| Sigma-sunset |  | 85 | 0.000085 | 4.070581 |
| NCI-47735 |  | 85 | 0.000085 | 4.070581 |

Table 2: Univariate statistics of the training and test set compounds

| Data set | Mean | Median | Mode | Range | Max. | Min. | Variance | SD | RMSD |
|---|---|---|---|---|---|---|---|---|---|
| Training | 4.5700 | 4.4437 | 5.3010 | 1.4523 | 5.5229 | 4.0706 | 0.2530 | 0.5030 | 0.4919 |
| Test | 4.6770 | 4.6990 | 5.3010 | 1.2095 | 5.3010 | 4.0915 | 0.2487 | 0.4987 | 0.4617 |

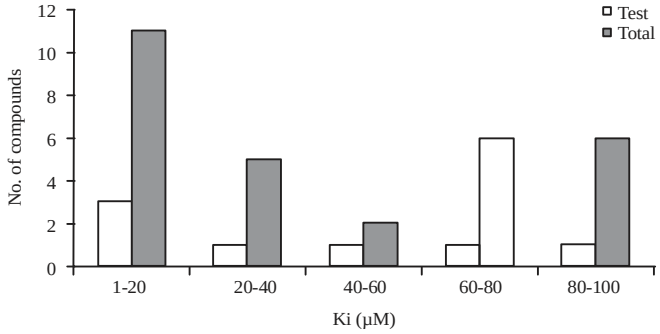



Fig. 1: Distribution of compounds within the activity range of 1-90 μM which explains the significant diversity in the data set

OPLS-2005 all atoms forcefield. Descriptors for QSAR model were generated from the program Qikprop (QikProp, version 3.3, Schrodinger, NY). The important descriptors are listed in Table 3. Two regression approaches, Partial Least Square (PLS) regression and Multiple Linear Regression (MLR), were used to build the QSAR model. MINITAB (MINITAB, statistical software of Minitab Inc., USA) was used for PLS method. Strike (Strike, version 1.9, Schrodinger, NY) was used for MLR. All these programs were implemented and executed on a machine with Core 2 duo 2.8GHz processor and Windows 7 operating system.

**2D-QSAR:** The 2D-QSAR models were developed from the dataset using the methods MLR and PLS. The MLR is used to find the linear relationship between a dependent and a set of independent variables. In the present study, 50 different structural descriptors were taken into consideration. Eighteen of them were found to possess constant values which were removed from further analysis. The remaining 32 descriptors were chosen for MLR studies. The descriptors were chosen based on their inter correlation coefficient. The PLS or Partial Least Square regression or Projection on Latent Structures is a method that combines features from Principal Component

Table 3: Descriptors used for building the QSAR model

| Descriptors | Description |
|---|---|
| #rotor | Number of non-trivial (not CX3), non-hindered (not alkene, amide, small ring) rotatable bonds |
| donor HB | Estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer |
| QPlog Kp | Predicted skin permeability, log Kp |
| EA (eV) | PM3 calculated electron affinity |
| #ring atoms | Number of atoms in a ring |

Analysis (PCA) and MLR. The PLS model finds new variables or latent variables or components which are linear combinations of the original variables. The PLS is especially useful in cases where the data set contains highly inter-correlated descriptors (Multicollinearity) and in cases where the number of descriptors exceeds the number of observations[20]. The optimum number of PLS components (latent variables) for the study was determined based on leave one out cross validation approach. The model refinement procedure applies Predicted Residual Error Sum of Squares (PRESS) of the cross validation to select the optimum number of PLS components[21]. The number of factors that produced the least PRESS was selected as optimum value. The same 32 descriptors, as used in MLR, were selected for the PLS studies. The descriptors with negligible/low regression coefficients were sequentially removed from further process until reliable statistical measures were obtained. The PRESS was calculated based on the following expression:

$$PRESS = \Sigma (Y_{i,obs}-Y_{i,pred})^2$$

In the above expression, $Y_{i,obs}$ means observed activity of "i"th compound in the training set and $Y_{i,pred}$ means leave-one-out cross validated activity of the same "i"th compound in the training set.

**Statistical measures:** The statistical reliability of the model was adjudged by the following parameters; squared correlation coefficient ($R^2$), Adjusted squared correlation coefficient ($R^2_a$), Standard Deviation (SD), Predicted residual Error Sum of Squares (PRESS), Fisher's value for statistical significance (F) and Significance level of variance ratio (P).

**Cross validation:** Cross validation is a practical and reliable method of testing the significance of the model[22]. Internal cross validation of the model was carried out with Leave-One-Out (LOO) approach in which every single compound from the training set is left out once and the activity of that compound is predicted using the model generated with the remaining compounds. The internal cross validation parameter is stated by Leave-One-Out cross validation coefficient ($q^2$) which is calculated using the formula:

$$q^2 = 1-(\Sigma(y_i-y_i')^2/\Sigma (y_i-y_{mean})^2)$$

where, $Y_i$ and $Y_i'$ are observed and leave-one-out predicted activities, respectively, of the "i"th compound in the training set and $Y_{mean}$ is mean value of the training set compounds[20]. Internal cross validation is never adequate to validate the predictive ability of the model. So it is necessary to validate the model with the test set compounds which are not included in the QSAR model development. In external cross validation, values of each of the test set compounds were predicted by applying their descriptor values in the model generated with training set compounds. External cross validation is expressed in terms of external cross validation coefficient, $r^{2pred}$ which can be calculated using the formula:

$$r^{2pred} = 1 - (\Sigma(Y_{i(Test)}-Y'_{i(Test)})^2/\Sigma(Y_{i(Test)}-Y_{mean})^2)$$

where $Y_{i(Test)}$ and $Y'_{i(Test)}$ are actual and predicted activities of the $i^{th}$ compound in the test set, respectively and $Y_{mean}$ is mean value of the training set compounds[23].

## RESULTS AND DISCUSSION

**Multiple linear regression:** Based on the inter-correlation coefficients of the descriptors, highly correlated descriptors were removed from the study. According to the rule of thumb in MLR (ratio of sample size to the number of descriptors should be greater than or equal to 5), a tetra-parametric model can be expected with the current training set of 23 compounds. The existence of multicollinearity among the descriptors in the present study resulted in a mono-parametric model as described below.

$$pIC_{50} = 3.5727(\pm 0.1620)-0.1756\ (\pm 0.0264)\ QPlogBB \quad (1)$$

$$n = 23,\ R^2 = 0.6777,\ R^2_a = 0.6147,\ F = 44.1,\ p = 0.000001,$$
$$q^2 = 0.6240,\ r^{2Pred} = 0.2138$$

Equation 1 indicates that the model obtained with MLR showed good squared correlation coefficient ($R^2$) value and good internal predictive power ($q^2$) but lack in external predictive power ($r^{2Pred}$). The removal of two outliers,

compounds 7 and 15, from the data set did not improve the external predictivity of the model. The scatter plot which is plotted between observed and predicted $pK_i$ values for training set and test set are shown in the Fig. 2a and b,respectively. To overcome the issues like, multicollinearity among the descriptors and poor external predictivity, PLS method was preferred for building a 2D-QSAR model.

**Partial least square regression:** The same training set, as used in MLR, was used to build the PLS model. The PLS regression was initially started with 32 descriptors. The descriptors with negligible regression coefficients were removed from the study until there was no improvement in $q^2$. The number of optimum components and descriptors for PLS model was found to be 5 (optimized by leave one out cross validation). The scatter plot which is plotted between observed and predicted values for training set and test set are shown in the Fig. 3a and b, respectively. The values of descriptors obtained for PLS analysis is shown in Table 4. The Table 5 represents the observed and predicted values for both MLR and PLS models. The following model equation was obtained by PLS regression analysis:

$$pIC_{50} = 3.02371\text{-}0.12701(\#rotor)\text{-}0.37289\ (EA(eV))+0.03648\ (donorHB)+0.05865\ (\#ringatoms)\text{-}0.29521\ (QPlogKp) \quad (2)$$

$$n = 23,\ R^2 = 0.7846,\ R^2_a = 0.6624,\ F = 12.38\ (p = 0.000005),\ q^2 = 0.638184,\ r^2_{pred} = 0.65008$$

The above equation is based on five PLS components and five descriptors. The equation could explain 66.2% of the variance and predict 63.8% of the variance. The coefficients of donor HB and #ring atoms contribute positively to the biological activity of the model equation whereas the other coefficients like #rotor, EA(eV) and QPlogKp negatively contribute to the activity. It is quite evident from Table 4 that the activity and the number of donor hydrogen bonds correlate well together. Hydrogen bonding is considered as one of the most important interactions that take place in any protein-ligand complex formation[24,25]. Moreover, Hydrogen bonds are observed with a variety of strengths and geometries in the active sites of protein-ligand complexes[26].

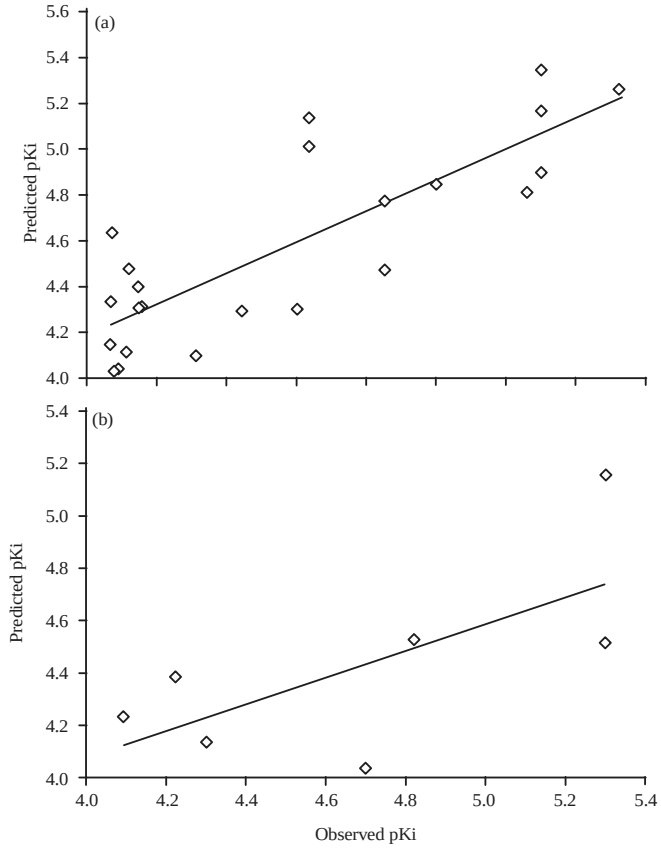



Fig. 2(a-b): MLR analysis demonstrating the correlation between observed and predicted $pK_i$ values for the (a) Training set and (b) Test set

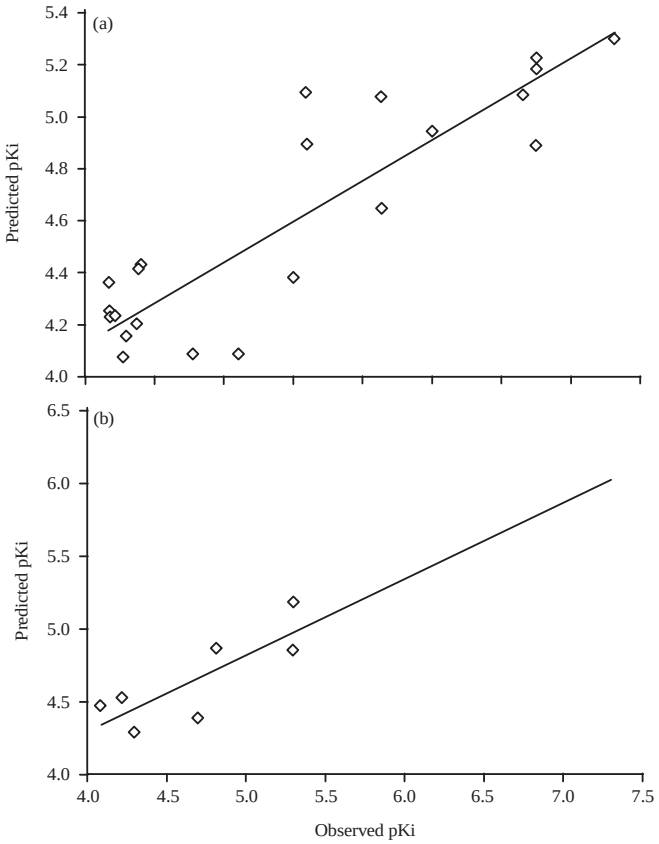



Fig. 3(a-b): PLS analysis demonstrating the correlation between observed and predicted $pK_i$ values for the (a) Training set and (b) Test set

Table 4: Values of important descriptors for training and test set compounds

| Compounds | #rotor | donor HB | QPlogKp | EA (eV) | #ring atoms |
|---|---|---|---|---|---|
| **Training set** | | | | | |
| 1 | 7 | 3.5 | -4.022 | 1.368 | 22 |
| 2 | 15 | 5 | -9.877 | 2.137 | 32 |
| 5 | 16 | 7 | -6.829 | 1.627 | 38 |
| 6 | 19 | 8 | -10.513 | 1.777 | 32 |
| 7 | 17 | 6 | -9.984 | 1.953 | 32 |
| 8 | 13 | 5.5 | -7.678 | 1.694 | 32 |
| 9 | 22 | 12 | -9.595 | 3.054 | 44 |
| 10 | 11 | 4.5 | -5.848 | 1.386 | 28 |
| 11 | 13 | 5.5 | -7.516 | 1.607 | 32 |
| 14 | 15 | 5 | -9.111 | 2.085 | 32 |
| 15 | 18 | 7.5 | -8.347 | 2.171 | 38 |
| 16 | 8 | 3 | -7.491 | 2.383 | 16 |
| 17 | 9 | 3 | -5.379 | 2.093 | 22 |
| 18 | 7 | 2 | -5.504 | 1.833 | 16 |
| 21 | 9 | 4 | -6.927 | 2.194 | 20 |
| 22 | 9 | 5.5 | -6.926 | 1.748 | 16 |
| 23 | 12 | 3 | -4.528 | 1.665 | 32 |
| 24 | 15 | 5.5 | -4.856 | 1.576 | 34 |
| 25 | 7 | 4 | -5.001 | 1.656 | 16 |
| 27 | 7 | 3 | -4.915 | 1.951 | 16 |
| 28 | 15 | 5 | -6.380 | 2.234 | 32 |
| 29 | 7 | 3 | -5.790 | 1.715 | 16 |
| 30 | 10 | 3 | -5.584 | 1.812 | 26 |

Table 4: Continue

| Compound | #rotor | donor HB | QPlogKp | EA (eV) | #ring atoms |
|---|---|---|---|---|---|
| **Test set** | | | | | |
| 3 | 11 | 5 | -5.946 | 1.589 | 32 |
| 4 | 16 | 8 | -9.488 | 2.103 | 32 |
| 12 | 11 | 5 | -5.837 | 1.485 | 32 |
| 13 | 7 | 4 | -4.014 | 0.986 | 22 |
| 19 | 7 | 3 | -5.439 | 1.981 | 20 |
| 20 | 9 | 3 | -7.362 | 2.172 | 20 |
| 26 | 8 | 3.5 | -5.529 | 1.59 | 22 |

Table 5: Observed and predicted values of MLR and PLS

| | Predicted $pK_i$ | | |
|---|---|---|---|
| Compounds | $pK_i$ | Model-1 (MLR) | Model-2 (PLS) |
| **Training set** | | | |
| 1 | 4.09151 | 4.03846 | 4.22986 |
| 2 | 5.52288 | 5.25402 | 5.2967 |
| 5 | 5.30103 | 4.89116 | 4.88494 |
| 6 | 5.30103 | 5.3371 | 5.22009 |
| 7 | 5.30103 | 5.16445 | 5.17936 |
| 8 | 5.25964 | 4.80493 | 5.08499 |
| 9 | 5.00000 | 4.84005 | 4.94161 |
| 10 | 4.85387 | 4.47052 | 4.64255 |
| 11 | 4.85387 | 4.76506 | 5.06961 |
| 14 | 4.63827 | 5.00497 | 5.08996 |
| 15 | 4.63827 | 5.12669 | 4.89443 |
| 16 | 4.60206 | 4.30666 | 4.37829 |
| 17 | 4.4437 | 4.28892 | 4.08785 |
| 18 | 4.3098 | 4.09923 | 4.08734 |
| 21 | 4.16115 | 4.30543 | 4.42634 |
| 22 | 4.1549 | 4.30209 | 4.41247 |
| 23 | 4.14874 | 4.39623 | 4.20172 |
| 24 | 4.11919 | 4.47245 | 4.15921 |
| 25 | 4.11351 | 4.10801 | 4.0778 |
| 27 | 4.08092 | 4.02705 | 3.90594 |
| 28 | 4.07572 | 4.63298 | 4.2282 |
| 29 | 4.07058 | 4.14419 | 4.25224 |
| 30 | 4.07058 | 4.3316 | 4.36075 |
| **Test set** | | | |
| 3 | 5.30103 | 4.51355 | 4.84863 |
| 4 | 5.30103 | 5.15338 | 5.17697 |
| 12 | 4.82391 | 4.52356 | 4.85523 |
| 13 | 4.69897 | 4.04215 | 4.38818 |
| 19 | 4.30103 | 4.13524 | 4.28404 |
| 20 | 4.22185 | 4.38253 | 4.52648 |
| 26 | 4.09152 | 4.23043 | 4.46495 |

Furthermore, it is reported that hydrogen bonds stabilize the ligand at the target site and increases binding affinity and drug efficacy[27]. The next foremost descriptor is #ringatoms. The number of atoms present in the ring correlates positively with the activity. Rotor or number of rotatable bonds is a measure of molecular flexibility (i.e., more the number of rotatable bonds more the molecule is flexible). The negative regression coefficient of rotor in our model indicates that the increase in number of rotatable bonds or increase in flexibility of compounds might reduce the activity. The EA(eV) means electron affinity calculated by PM3 semi empirical methods. Electron affinity is another descriptor which is a measure of change in energy when a neutral atom is attracted towards an electron to form a negative ion. Oxygen, sulphur and nitrogen atoms hold the key for electron affinity values in the compounds used in the present study. As the regression coefficient of EA (eV) contributes negatively to the model, we deduce that the compounds with more electron affinity values will decrease the activity.

The reliability or predictability of the model is further assured by the external validation with test set compounds ($r^{2pred}$). The model resulted in this study is said to be predictive

since it meets the following conditions $R^2$ >0.6, $q^2$>0.6 and $r^{2pred}$>0.5 as reported in the literature[28].

## DISCUSSION

The present study involves 2D-QSAR model development with angiogenin inhibitors. Two regression methods namely, Multiple Linear Regression (MLR) and Partial Least Square Regression were used in the present study to build the QSAR model. This is in accordance with the study by Sahu *et al.*[29] who derived 2D-QSAR models based on MLR for predicting the anti-malarial activity of compounds. Similarly, MLR was used to derive QSAR models to predict biological activities of Aurora-A kinase inhibitors[30] and to predict anti-leishmanial activity of diaryl sulfides[31]. The descriptors or independent variables generated for the model development showed multicollinearity. Multicollinearity is high multiple correlations between subsets of the variables, which leads to removal of variables from the model[32]. This made the task difficult for building the QSAR model with MLR. The model established with MLR had ended up with a mono-parametric equation as well as poor external predictability. This led us to go for another well known regression approach called PLS which handle the data set that shows multicollinearity with numerous descriptors[33]. The PLS was widely used by many researchers to derive QSAR models with predictive capability[34-36]. In this study, the QSAR model developed with PLS exhibited good internal as well as external predictability. The model also indicates the importance of donor hydrogen bonds and ring atoms for the biological activity. The predictivity of the model was tested using various statistical parameters. The model developed in this study can act as a predictive model as the $q^2$ value of the current model is in par with the $q^2$ of models derived by Kovalishyn *et al.*[37]. The model developed in this study can be still enriched with more number of data sets in order to increase the predictivity of the model. As the reported number of effective small molecule inhibitors to angiogenin, albeit their vast potential, has been found to be very low, it is thus indispensable to optimize the existing structures of small molecule inhibitors and to screen for them from various compound libraries for designing an effective lead against angiogenin.

## ACKNOWLEDGMENT

The work is supported by Bioinformatics Infrastructure Facility (No. BT/BI/25/001/2006) Bioinformatics division, Department of Biotechnology, Ministry of Science and Technology, Government of India, New Delhi.

## REFERENCES

1. Carmeliet, P. and R.K. Jain, 2000. Angiogenesis in cancer and other diseases. Nature, 407: 249-257.
2. Blood, C.H. and B.R. Zetter, 1990. Tumor interactions with the vasculature: Angiogenesis and tumor metastasis. Biochimica Biophysica Acta (BBA)-Rev. Cancer, 1032: 89-118.
3. Macchiarini, P., G. Fontanini, F. Squartini, C.A. Angeletti and M.J. Hardin, 1992. Relation of neovascularisation to metastasis of non-small-cell lung cancer. Lancet, 340: 145-146.
4. Folkman, J., 1995. Biomarkers of aging: do we know what to look for? Mol. Med., 1: 120-122.
5. Cross, M. and T.M. Dexter, 1991. Growth factors in development, transformation and tumorigenesis. Cell, 64: 271-280.
6. Okamura, K., A. Morimoto, R. Hamanaka, M. Ono, K. Kohno, Y. Uchida and M. Kuwano, 1992. A model system for tumor angiogenesis: Involvement of transforming growth factor-α in tube formation of human microvascular endothelial cells induced by esophageal cancer cells. Biochem. Biophys. Res. Commun., 186: 1471-1479.
7. Gao, X. and Z. Xu, 2008. Mechanisms of action of angiogenin. Acta Biochim. Biophys., 40: 619-624.
8. Fett, J.W., D.J. Strydom, R.R. Lobb, E.M. Alderman, J.L. Bethune, J.F. Riordan and B.L. Vallee, 1985. Isolation and characterization of angiogenin, an angiogenic protein from human carcinoma cells. Biochemistry, 24: 5480-5486.
9. Shapiro, R., J.F. Riordan and B.L. Vallee, 1986. Characteristic ribonucleolytic activity of human angiogenin. Biochemistry, 25: 3521-3532.
10. Hallahan, T.W., R. Shapiro and B.L. Vallee, 1991. Dual site model for the organogenic activity of angiogenin. Proc. Natl. Acad. Sci. USA, 88: 2222-2226.
11. Hallahan, T.W., R. Shapiro, D.J. Strydom and B.L. Vallee, 1992. Importance of asparagine-61 and asparagine-109 to the angiogenic activity of human angiogenin. Biochemistry, 31: 8022-8029.
12. Kerbel, R. and J. Folkman, 2002. Clinical translation of angiogenesis inhibitors. Nat. Rev. Cancer, 2: 727-739.
13. Shapiro, R. and B.L. Vallee, 1987. Human placental ribonuclease inhibitor abolishes both angiogenic and ribonucleolytic activities of angiogenin. Proc. Natl. Acad. Sci. USA., 84: 2238-2241.
14. Russo, A., K.R. Acharya and R. Shapiro, 2001. Small molecule inhibitors of RNase A and related enzymes. Methods Enzymol, 341: 629-648.
15. Russo, N., K.R. Acharya, B.L. Vallee and R. Shapiro, 1996. A combined kinetic and modeling study of the catalytic center subsites of human angiogenin. Proc. Natl. Acad. Sci. USA., 93: 804-808.

16. Leonidas, D.D., G.B. Chavali, A.M. Jardine, S. Li, R. Shapiro and K.R. Acharya, 2001. Binding of phosphate and pyrophosphate ions at the active site of human angiogenin as revealed by X-ray crystallography. Protein Sci., 10: 1669-1676.
17. Shapiro, R., J.L. Jenkins, R.Y.T. Kao and G.J. Latham, 2004. Small-molecule inhibitors of angiogenin and rnases and *in vivo* and *in vitro* methods of using same. Publication No. WO 2004076640 A2, September 10, 2004. https://www.google.com/patents/WO2004076640A2?cl=en22
18. Winkler, D.A., 2002. The role of Quantitative Structure-Activity Relationships (QSAR) in biomolecular discovery. Brief. Bioinform., 3: 73-86.
19. Roy, P.P. and K. Roy, 2008. On some aspects of variable selection for partial least squares regression models. QSAR Comb. Sci., 27: 302-313.
20. Abdi, H., 2010. Partial least squares regression and Projection on Latent Structure regression (PLS regression). Wiley Interdisciplin. Rev.: Comput. Stat., 2: 97-106.
21. Wold, S., L. Eriksson and S. Clementi, 1995. Statistical Validation of QSAR Results. In: Chemometric Methods in Molecular Design, Van de Waterbeemd, H. (Ed.). Chapter 5, VCH Publishers, Weinheim, Germany, ISBN-13: 9783527300440, pp: 309-318.
22. Hasegawa, K., Y. Miyashita, S.I. Sasaki, H. Sonoki and H. Shigyou, 1992. Quantitative structure-activity relationship study of antiarrhythmic phenylpyridines using multivariate partial least squares modelling. Chem. Int. Lab. Syst., 16: 69-75.
23. Fan, Y., L.M. Shi, K.W. Kohn, Y. Pommier, and J.N. Weinstein, 2001. Quantitative structure-antitumor activity relationships of camptothecin analogues: Cluster analysis and genetic algorithm-based studies. J. Med. Chem., 44: 3254-3263.
24. Panigrahi, S.K., 2008. Strong and weak hydrogen bonds in protein-ligand complexes of kinases: A comparative study. Amino Acids, 34: 617-633.
25. Jeffrey, G.A. and W. Saenger, 1991. Hydrogen Bonding in Biological Structures. Springer-Verlag, New York, USA., ISBN: 978-3-642-85135-3, Pages: 569.
26. Sarkhel, S. and G.R. Desiraju, 2004. N-H...O, O-H...O and C-H...O hydrogen bonds in protein-ligand complexes: Strong and weak interactions in molecular recognition. Proteins Struct. Funct. Bioinf., 54: 247-259.
27. Patil, R., S. Das, A. Stanley, L. Yadav, A. Sudhakar and A.K. Varma, 2010. Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. PLoS ONE, Vol. 5. 10.1371/journal.pone.0012029
28. Golbraikh, A. and A. Tropsha, 2002. Beware of $q^2$! J. Mol. Graphics. Model., 20: 269-276.
29. Sahu, N.K., M.C. Sharma, V.K. Mourya and D.V. Kohli, 2014. QSAR studies of some side chain modified 7-chloro-4-aminoquinolines as antimalarial agents. Arab. J. Chem., 7: 701-707.
30. Yan, A., Y. Chong, L. Wang, X. Hu and K. Wang, 2011. Prediction of biological activity of Aurora-A kinase inhibitors by multilinear regression analysis and support vector machine. Bioorg. Med. Chem. Lett., 21: 2238-2243.
31. Goodarzi, M., E.F.F. da Cunha, M.P. Freitas and T.C. Ramalho, 2010. QSAR and docking studies of novel antileishmanial diaryl sulfides and sulfonamides. Eur. J. Med. Chem., 45: 4879-4889.
32. Van Damme, S. and P. Bultinck, 2007. A new computer program for QSAR-analysis: ARTE-QSAR. J. Comput. Chem., 28: 1924-1928.
33. Wold, S., M. Sjostrom and L. Erikson, 2001. PLS-regression: A basic tool of chemometrics. Chemometrics Intell. Lab. Syst., 58: 109-130.
34. Hemmateenejad, B., R. Miri, M. Akhond and M. Shamsipur, 2002. QSAR study of the calcium channel antagonist activity of some recently synthesized dihydropyridine derivatives. An application of genetic algorithm for variable selection in MLR and PLS methods. Chemometrics Int. Lab. Syst., 64: 91-99.
35. Hernandez, N., R. Kiralj, M.M.C. Ferreira and I. Talavera, 2009. Critical comparative analysis, validation and interpretation of SVM and PLS regression models in a QSAR study on HIV-1 protease inhibitors. Chemometrics Intell. Lab. Syst., 98: 65-77.
36. Luco, J.M. and F.H. Ferretti, 1997. QSAR based on multiple linear regression and PLS methods for the anti-HIV activity of a large group of HEPT derivatives. J. Chem. Inform. Comput. Sci., 37: 392-401.
37. Kovalishyn, V., J. Aires-de-Sousa, C. Ventura, R.E. Leitao and F. Martins, 2011. QSAR modeling of antitubercular activity of diverse organic compounds. Chemometrics Intell. Lab. Syst., 107: 69-74.