

A 2D QSAR Analysis of HIF-1 inhibitors (2-phenoxy- N-substituted acetamides) as Anti-Cancer Agents

LR. Jagtap*, BA. Bhairav and RB. Saudagar

Department of Quality Assurance and Techniques, KCT's R.G. Sapkal College of Pharmacy,
Anjaneri - 422213, Tal. Trimbakeshwar, Dist. Nashik, Maharashtra, India.

ABSTRACT

A set of thirty one substituted 2-phenoxy-*N*-phenylacetamide derivatives with HIF-1 inhibitory activities was subjected to 2D and 3D Quantitative Structure Activity Relationship (QSAR) studies using various combinations of descriptors. 2D-QSAR was performed using Multiple Linear Regression (MLR), Principal Component Regression (PCR) and Partial Least Squares Regression (PLS) methods. Among these three methods Multiple Linear Regression (MLR) led to the statistically significant best 2D-QSAR Model-I having correlation coefficient $r^2 = 0.9469$ and cross validated squared correlation coefficient $q^2 = 0.8933$ with external predictive ability of $\text{pred}_r^2 = 0.7128$ with the descriptors like SssNHE-index, slopp, T_O_N_1 and T_2_Cl_1. 3D-QSAR study was performed using the simulated annealing variable selection procedures k-nearest neighbor molecular field analysis approach. 3D-QSAR shows interesting results in terms of internal and external predictability. Molecular field analysis was applied for the generation of steric, hydrophobic and electrostatic descriptors based on aligned structures which shows good correlative and predictive capabilities in terms of $q^2 = 0.9672$ and $\text{pred}_r^2 = 0.8480$. Hence the model proposed in this work provides important structural insight in designing novel derivatives with specific HIF-1 inhibitory activity.

Keywords: 2-phenoxy-*N*-phenylacetamide, 3D-QSAR, Anti-Cancer Agent.

INTRODUCTION

Development of intratumoral hypoxia is a hallmark of rapidly growing solid tumors and of their metastases. Hypoxic tumor cells are resistant to conventional chemotherapy and radiotherapy^[1-4] and consequently the presence of hypoxia in tumors play a negative role in patient prognosis. A key modulator expressed in many cancer cells in response to hypoxic stress is hypoxia-inducible factor-1 (HIF-1), a heterodimeric transcription factor that consists of an HIF-1 α subunit and a HIF-1 β subunit, both of which are members of the basic helix-loop-helix PER/ Arnt/Sim (bHLH-PAS) transcription family. HIF-1 β is also known as the aryl hydrocarbon nuclear translocator (ARNT). HIF-1 activity is dependent on the availability of the HIF-1 α subunit, which is regulated by cellular oxygen levels. At normal oxygen levels, HIF-1 α is degraded via the pVHL-mediated ubiquitin-proteosomal pathway. Under hypoxic conditions, however, HIF-1 α rapidly accumulates in the cell and dimerizes with HIF-1 β , which is constitutively expressed^[5-11]. Binding of HIF-1 with co-activators to a specific DNA sequence within the target gene promoter, called the hypoxia-responsive elements (HRE), leads to transcriptional activation of a variety of genes involved in angiogenesis, glycolysis, growth factor signaling, tumor invasion, and metastasis^[12]. In many cancers, the HIF-1 pathway is not only activated by low oxygen tension but is also induced or amplified by a wide range of growth-promoting stimuli and oncogenic signals.

Clinically, HIF-1 α overexpression has been shown to be a marker of highly aggressive cancers,^[13] and inhibition of HIF-1 production and function significantly reduces tumor growth in animal models^[14]. Accordingly, HIF-1 represents an attractive molecular target for the development of novel anticancer agents^[15-16]. Among the recently identified inhibitors of HIF-1, 2-phenoxy-*N*-phenylacetamide derivatives were found to be very active^[17]. The development of tyrosine HIF-1 inhibitors has therefore become an active area of research in pharmaceutical science. One could not, however, confirm that the compounds synthesized would always possess good inhibitory activity to HIF-1, while experimental assessments of inhibitory activity of these compounds are time-consuming and expensive. Consequently, it is of interest to develop a prediction method for biological activities before the synthesis. Quantitative structure activity relationship (QSAR) searches information relating chemical structure to biological and other activities by developing a QSAR model. Using such an approach one could predict the activities of newly designed compounds before a decision is being made whether these compounds should be really synthesized and tested. With the above facts and in

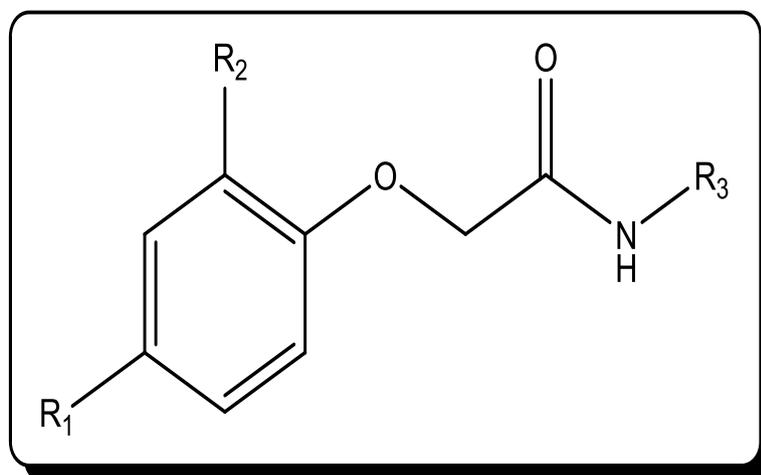
continuation of our research for newer anti-cancer agent ^[18-25] in the present study, we reported 2D and 3D-QSAR studies on a series of HIF1 inhibitors to provide further insight into the key structural features required to design potential drug candidates of this class.

MATERIAL AND METHOD

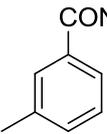
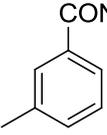
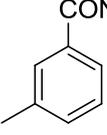
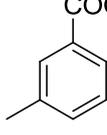
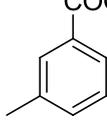
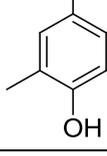
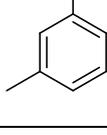
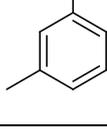
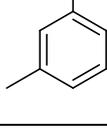
1 Data Set for 2D and 3D-QSAR

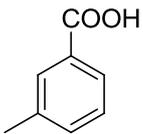
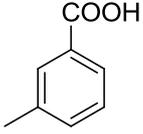
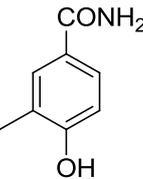
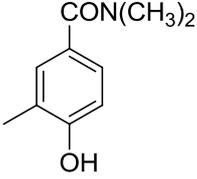
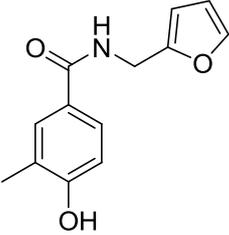
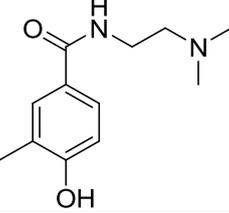
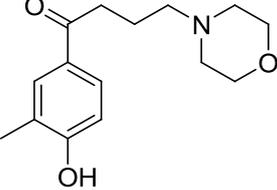
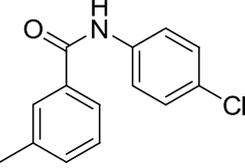
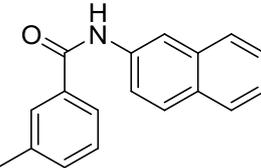
In the present study, a series of substituted 2-phenoxy-*N*-phenylacetamide having HIF-1 inhibitory activity from published result ^[17] was taken. The HIF-1 inhibitory activities (IC_{50}) converted to $-\log(IC_{50})$ were used in 2D and 3D-QSAR. Various 2D-QSAR models were generated for this series using Multiple Linear Regression (MLR), Principal Component Regression (PCR) and Partial Least Squares (PLS) Regression methods and those which come out with promising results are discussed here. QSAR models were generated by a training set of 25 (MLR), 23 (PCR) and 24 (PLS) molecules. Predictive power of the resulting models was evaluated by a test set of 6 (MLR), 8 (PCR) and 7 (PLS) molecules with uniformly distributed biological activities. The test set was selected based on the criteria given by Oprea et al ^[26]. The structures of all the compounds along with their actual and predicted biological activities are presented in Table 1 for 2D.

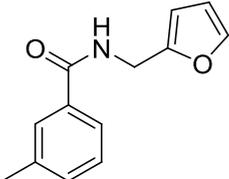
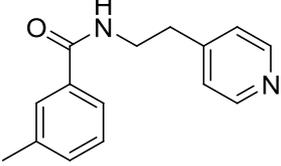
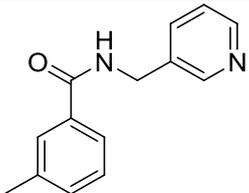
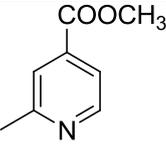
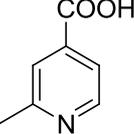
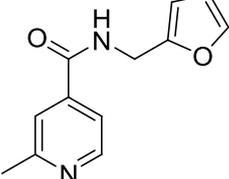
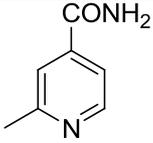
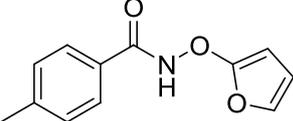
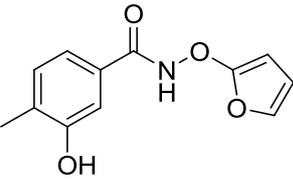
Table 1: Structure, experimental and predicted activity of 2-phenoxy-*N*-substituted acetamide acid analogues used in training and test set for HIF-1 α inhibition against Hep3B Cell lines by Model-I

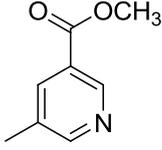


Sr. No.	R ₁	R ₂	R ₃	IC ₅₀ (μM)	pIC ₅₀		Residual
					Exp.	Pre.	
1 ^T	Admantine	-		9.9	5.00	4.98	0.02
2	Admantaie	-		2.6	5.58	5.62	-0.04
3	Admantine	-		4.3	5.36	5.42	-0.06

4	-C(CH ₃) ₃	-		30	4.52	4.49	0.03
5	-CH ₃	-		30	4.52	4.33	0.19
6	-NO ₂	-		30	4.52	4.61	-0.09
7 ^T	Admantine	-		5	5.30	5.42	-0.12
8	-C(CH ₃) ₃	-		29	4.53	4.49	0.04
9	Admantine	-		0.4	6.39	6.44	-0.05
10	F	-		30	4.52	4.43	0.09
11	Cl	-		30	4.52	4.62	-0.01
12	Br	-		30	4.52	4.58	-0.06

13	I	-		7.2	5.14	4.98	0.16
14 ^T	-COCH ₃	-		30	4.52	4.65	-0.13
15	Admantine	-		30	4.52	4.43	0.09
16	Admantine	-		30	4.52	4.14	0.38
17	Admantine	-		12.6	4.89	4.97	-0.08
18	Admantine	-		51.7	4.29	4.49	-0.2
19	Admantine	-		30	4.52	4.61	-0.09
20	Admantine	-		30	4.52	4.45	0.07
21	Admantine	-		30	4.52	4.61	-0.09

22	Admantine	-		14.06	4.85	4.79	0.06
23	Admantine	-		8.2	5.08	5.12	-0.04
24 ^I	Admantine	-		10	5.00	5.19	-0.19
25	Admantine	-		1.2	5.92	5.87	0.05
26	Admantine	-		1.03	6.00	5.90	0.1
27 ^T	Admantine	-		10	5.00	5.19	-0.19
28	Admantine	-		3.1	5.50	5.63	-0.13
29	Cl	Cl		30	4.52	4.45	0.007
30	Cl	Cl		23.2	4.63	4.59	0.04

31	Admantine	-		5.9	5.22	5.39	-0.17
----	-----------	---	---	-----	------	------	-------

Exp. = Experimental activity, Pred. = Predicted activity;
 a = $-\log(\text{IC}_{50} \times 10^{-6})$; T = Test Set

3D-QSAR models were generated using a training set of 25 molecules with varied chemical and biological activities. The sphere exclusion method was used for the selection of molecules in training and test sets. In order to obtain a validated QSAR model for the purpose of meaningful prediction, an available dataset should be divided into the training and test sets. For the prediction statistics to be reliable, the test set must include at least five compounds^[27]. Ideally, the division into the training and test set must satisfy the following three conditions: (i) All representative compound-points of the test set in the multidimensional descriptor space must be close to those of the training set. (ii) All representative points of the training set must be close to those of the test set. (iii) The representative points of the training set must be distributed within the whole area occupied by the entire dataset^[27]. Since some compound exhibited insignificant/no inhibition, such compounds were excluded from the present study

2 Two Dimensional Molecular Modeling

The data set used for the QSAR analyses contains 31 molecules belonging to 2-phenoxy-N-phenylacetamide derivatives as HIF-1 inhibitors. All the structures of the compounds were drawn in 2D-APPL mode of software and exported to 3D model. The modeling analyses, calculations, and visualizations for 2D-QSAR were performed using the V-Life Molecular Design Suite 3.0 (Vlife MDS). The compounds were then subjected to energy minimization and conformational analysis using Monte Carlo conformational search with RMS gradient of 0.001 kcal/mol using MMFF force field. Monte Carlo conformational search method is similar to the RIPS method that generates a new molecular conformation by randomly perturbing the position of each coordinate of each atom in molecule^[28]. Most stable structure for each compound was generated after energy minimization and used for calculating various physico-chemical descriptors. A total of 238 2D descriptors were computed using VLife science Molecular Design Suite Software, which include various physicochemical descriptors; Baumann alignment-independent topological descriptors^[29]. After computation of all descriptors, invariable descriptors that have no variation in their values were removed. 2D-QSAR equations were generated by MLR, PCR and PLS analysis using forward-backward variable selection method. Various 2D descriptors like element counts, molecular weight, molecular refractivity, log P, topological index, electro-topological index, Baumann alignment independent topological descriptors, different quantum chemical descriptors including heat of formation, dipole moment, local charges, and different topological, elemental count including bromine count, fluorine count, Path count and constitutional descriptors for each molecule. For calculation of descriptors every atom in the molecule was assigned at least one and at most three attributes. The first attribute is 'T-attribute' to thoroughly characterize the topology of the molecule. The second attribute is the atom type. The atom symbol is used here. The third attribute is assigned to atoms taking part in a double or triple bond. In 2D-QSAR techniques the cross-validation analysis was performed using leave one out method.

3 Three Dimensional Molecular Modeling

Like many 3D-QSAR methods^[30, 31] k-nearest neighbor molecular field analysis (kNN-MFA) requires suitable alignment of given set of molecules. This is followed by generation of a common rectangular grid around the molecules. The steric and electrostatic interaction energies are computed at the lattice points of the grid using a methyl probe of charge +1. These interaction energy values are considered for relationship generation and utilized as descriptors to decide nearness between molecules. The term descriptor is utilized in the following discussion to indicate field values at the lattice points. The optimal training and test sets were generated using the sphere exclusion algorithm. This algorithm allows the construction of training sets covering descriptor space occupied by representative points. Once the training and test sets were generated, kNN methodology was applied to the descriptors generated over the grid.

- **3.1 Building k-nearest-neighbor (kNN-MFA) model**

The kNN methodology relies on a simple distance learning approach whereby an unknown member is classified according to the majority of its k-nearest neighbors in the training set. The nearness is measured by an appropriate distance metric (e.g., a molecular similarity measure calculated using field interactions of molecular structures). The standard kNN method is implemented simply as follows: (i) Calculate distances between an unknown object (u) and all the objects in the training set; select k objects from the training set most similar to object u, according to the calculated distances; and classify object u with the group to which the majority of the k objects belongs. (ii) 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.

- **3.2. kNN-MFA with Simulated Annealing method**

Simulated annealing^[32] is the simulation of a physical process, 'annealing', which involves heating the system to a high temperature and then gradually cooling it down to a preset temperature (e.g., room temperature). During this process, the system samples possible configurations distributed according to the Boltzmann distribution so that at equilibrium, low energy states are the most populated. As the temperature falls so the lower energy states become more probable. In a variable selection procedure a move typically consists of adding or deleting a single variable from the set under consideration. The concept is based on the manner in which liquids freeze or metals recrystallize in the process of annealing. In Simulated annealing the process starts from an initial state of the very high temperature and introduces perturbations, or random moves, that create a new state. The difference in fitness between the initial and perturbed states defines whether the move is accepted. If the fitness function is better, leading to an improved model (in terms of fitness function like cross validated q^2) then the new subset is accepted for the next iteration. In the SA variant the temperature of the system is gradually reduced, hence increasing the chance of finding the globally optimal solution. The maximum and minimum temperature in this analysis were set as 100K and 0.01K, respectively with cross correlation limit 0.4 and the temperature was decreased by 5 units with 5 iteration at that particular temperature.

- **3.3 Alignment of Molecules**

The most critical input for the 3D-QSAR modeling is the alignment of the molecules. Structures of all compounds were sketched using molecular builder of MDS 3.5 and each structure was subjected to energy minimization up to a gradient of 0.01 kcal/mol using MMFF-94 force field^[33]. Conformational search of each energy minimized structure was performed using stochastic approach. Stochastic conformational search method is similar to the RIPS method, which generate new molecular conformation by randomly perturbing the position of each coordinate of each atom in the molecule followed by the energy minimization. Energy minimized and geometry optimized structure of molecules were aligned by the template-based method^[34] using VLife MDS 3.5 software where a template structure is defined and used as a basis for alignment of a set of molecules, and a reference molecule is chosen on which the other molecules of the dataset get aligned considering the chosen template. The template structure, i.e. 2-phenoxy-N-phenylacetamide was used for alignment by considering the common elements of the series as shown in Fig. 4. The reference molecule is chosen in such a way that it is the most active among the series of molecules considered. The superimposition of all molecules based on minimizing root mean square deviation (RMSD) is shown in Fig. 4.

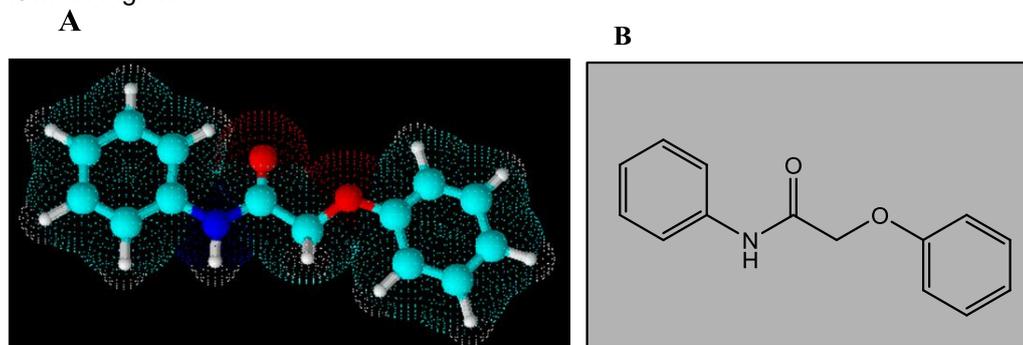


Fig. 4: 2-phenoxy-N-phenylacetamide template used for alignment of molecules in 3D-QSAR

A) 3D view B) 2D view

• 3.4 Computation of steric and electrostatic fields

The aligned biologically active conformations of 2-phenoxy-*N*-phenylacetamide are used for the calculation of molecular fields. Molecular fields are the steric and electrostatic interaction energies which are used to formulate a relationship between steric and electrostatic properties together with the biological activities of compounds. Each conformation is taken in turn, and the molecular fields around it are calculated. This is done by generating three-dimensional rectangular grids around the molecule and calculating the interaction energy between the molecule and probe group placed at each grid point. Steric and electrostatic fields are computed at each grid point considering MMFF charges^[34]. Methyl probe of charge +1 with 10.0 kcal/mole electrostatic and 30.0 kcal/mole steric cutoff were used for fields generation. A value of 1.0 is assigned to the distance-dependent dielectric constant. Steric and electrostatic field descriptors were calculated using Lennard-Jones and Coulomb potentials^[30, 34]. In the present study, molecular field analysis coupled with stepwise forward-backward variable was applied to obtain a 3D-QSAR model based on steric and electrostatic descriptors. The calculated steric and electrostatic field descriptors were used as independent variables and pIC₅₀ values were used as dependent variables in the present study^[35] to derive the 3D-QSAR models using MDS software.

4 Cross-Validation

The developed QSAR models are evaluated using the following statistical measures: *n*, (the number of compounds in regression); *k*, (number of variables); DF, (degree of freedom); *r*² (the squared correlation coefficient), F test (Fischer's value) for statistical significance, *q*² (cross-validated correlation coefficient) and *pred_r*², (*r*² for external test set). The regression coefficient *r*² is a relative measure of fit by the regression equation. However, a QSAR model is considered to be predictive, if the following conditions are satisfied: *r*² > 0.6, *q*² > 0.6 and *pred_r*² > 0.5^[36]. Internal validation is carried out using 'leave-one-out' (LOO) method^[29]. The cross-validated coefficient, *q*², is calculated using the following equation: The cross-validated *r*² (*q*²) value was calculated using where *y_i* and *ŷ_i* are the actual and predicted activities of the *i*th molecule, respectively and *y_{mean}* is the average activity of all molecules in the training set. Both summations are over all molecules in the training set. Since the calculation of the pair wise molecular similarities, and hence the predictions, were based upon the current trial solution, the *q*² obtained is indicative of the predictive power of the current 2D and 3D-model.

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

External validation is also carried out in the present study. The external predictive power of the model is assessed by predicting pIC₅₀ value of the eight test set molecules, which are not included in the QSAR model development. The predictive ability of the selected model is also confirmed by *pred_r*² or rCVext². The *pred_r*² value is indicative of the predictive power of the current 2D and 3D model for external test set.

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

Where *y_i* and *ŷ_i* are the actual and predicted activity of the *i*th molecule in the test set, respectively, and *y_{mean}* is the average activity of all molecules in the training set. The significance of the models hence obtained is derived based on a calculated Z score^[31]. A Z score value is calculated by the following formula:

$$Zscore = \frac{(h - \mu)}{\sigma} \quad \dots\dots 3$$

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

- Result and Discussion
- 2D-QSAR modeling and its validation

A series of total thirty one compounds for which absolute IC₅₀ values reported was used for correlating chemical composition (structure) with their HIF-1 inhibitory activity. 2D-QSAR (MLR, PCR and PLS) and 3D-QSAR (kNN MFA-Simulated Annealing) models were generated. The best QSAR model was selected on the basis of value of Statistical parameters like *r*² (square of correlation coefficient for training set of compounds), *q*² (cross-validated *r*²), and *pred_r*² (predictive *r*² for the test

set of compounds). The generated QSAR models were evolved by repeating the MLR, PCR and PLS and kNN–MFA methods to check the accuracy and precision of both the methods. The frequency of use of a particular descriptor in the population of equations indicated the relevant contributions of the descriptors. Statistical results generated by both 2D and 3D-QSAR analysis showed that both QSAR model have good internal as well as external predictability as shown in Table 2 and descriptor contributing to the present study is shown in Table 3.

Table 2: Statistical parameters of 2D and 3D-QSAR

QSAR APPROACH	2D-QSAR			3D-QSAR
	MLR	PCR	PLSR	kNN MFA Model
Parameters				
N	25	23	24	25
k	-	-	-	2
Df	21	18	20	-
r^2	0.9469	0.8471	0.9179	-
q^2	0.8933	0.6954	0.8317	0.9672
F test	62.440	27.6923	59.6450	-
r^2 se	0.5376	0.2311	0.2541	-
q^2 se	0.4628	0.3261	0.3307	-
pred_ r^2	0.7128	0.6932	0.7541	0.8480
pred_ r^2 se	0.6386	0.5386	0.6155	-
α _ran_ r^2	0.0000	0.00000	0.0000	-
α _ran_ q^2	0.0001	0.0001	0.0010	0.00000
α _ran_pred_ r^2	0.0000	0.0000	0.0000	-
best_ran_ r^2	0.7444	0.6518	0.6834	-
best_ran_ q^2	0.5423	0.3989	0.4533	0.51099
Zscore_ran_ r^2	5.1702	5.6195	4.9604	-
Z score_ran_ q^2	4.2915	4.2591	4.9112	8.3245
				S_815 (30, 0000 30, 0000) E_1166 (-10, 0000 -10, 0000) H_1547 (0.3970 0.4039) S_848 (-0.1636 -0.0958)

N, number of molecules in training set, MLR = Multiple Linear Regression, PCR = Principal Component Regression, PLS = Partial Least Squares, N = number of molecules of training set, Df =degree of freedom, r^2 = coefficient of determination, q^2 = cross validated r^2 , pred_ r^2 = r^2 for external test set, Zscore = the Zscore calculated by q^2 in the randomization test, best_ran_ q^2 = the highest q^2 value in the randomization test and α _ran_ q^2 = the statistical significance parameter obtained by the randomization test,

Table 3: Molecular descriptors contributing in the present 2D-QSAR study

DESCRIPTOR	DESCRIPTION
Individual	
H-acceptor count	This descriptor signifies number of hydrogen bond acceptor atoms.
Rotatable bond count	This descriptor signifies number of rotatable bonds.
slogp	This descriptor signifies log of the octanol/water partition coefficient This property is an atomic contribution model that calculates logP from the given structure; i.e., the correct protonation state
ChiV ChiV5 Estate Number	This descriptor signifies atomic valence connectivity index
SssCH2count Electrotopological descriptor	This descriptor defines the total number of -CH ₂ group connected with two single bonds

SssOE-index	Electrotopological state indices for number of oxygen atom connected with two single bonds.
SssNHE-index	Electrotopological state indices for number of -NH group connected with two single bonds.
Alignment Independent Descriptors	
T_N_O_3	This is the count of number of nitrogen atoms separated from Oxygen atoms by three bond distance.
T_2_N_1	This is the count of number of double bonded atoms separated from nitrogen atoms by one bond distance.
T_O_N_1	This is the count of number of oxygen atoms separated from nitrogen atoms by one bond distance.
T_2_Cl_1	This is the count of number of double bonded atoms separated from chlorine atoms by one bond distance.

$$pIC_{50} = 15.8435(\text{SssNHE-index}) + 2.1497(\text{slogP}) - 6.2531(\text{T}_O_N_1) + 0.5003(\text{T}_2\text{Cl}_1) - 9.5332 \dots\dots\dots 4$$

2D QSAR Model-I developed has a correlation coefficient (r^2) of 0.9469, significant cross validated correlation coefficient (q^2) of 0.8933, F test of 62.440, F-statistics proves it to be statistically significant as the calculated Fischer value (F) exceeds the Tabulated F value and low Standard Error of Estimate. r^2 for external test set (pred_r^2) 0.7128 and degree of freedom 18. The model is validated by $\alpha_{\text{ran}_r^2} = 0.0000$, $\alpha_{\text{ran}_q^2} = 0.0001$, $\alpha_{\text{ran}_{\text{pred}_r^2}} = 0.0000$, $\text{best}_{\text{ran}_r^2} = 0.7444$, $\text{best}_{\text{ran}_q^2} = 0.54239$, $\text{Zscore}_{\text{ran}_r^2} = 5.1702$ and $\text{Z score}_{\text{ran}_q^2} = 4.2915$. The randomization test suggests that the developed model have a probability of less than 1% that the model is generated by chance. The activity distribution plot and descriptor contribution plot for 2D-QSAR Model-I is shown in Fig. 1 and 2.

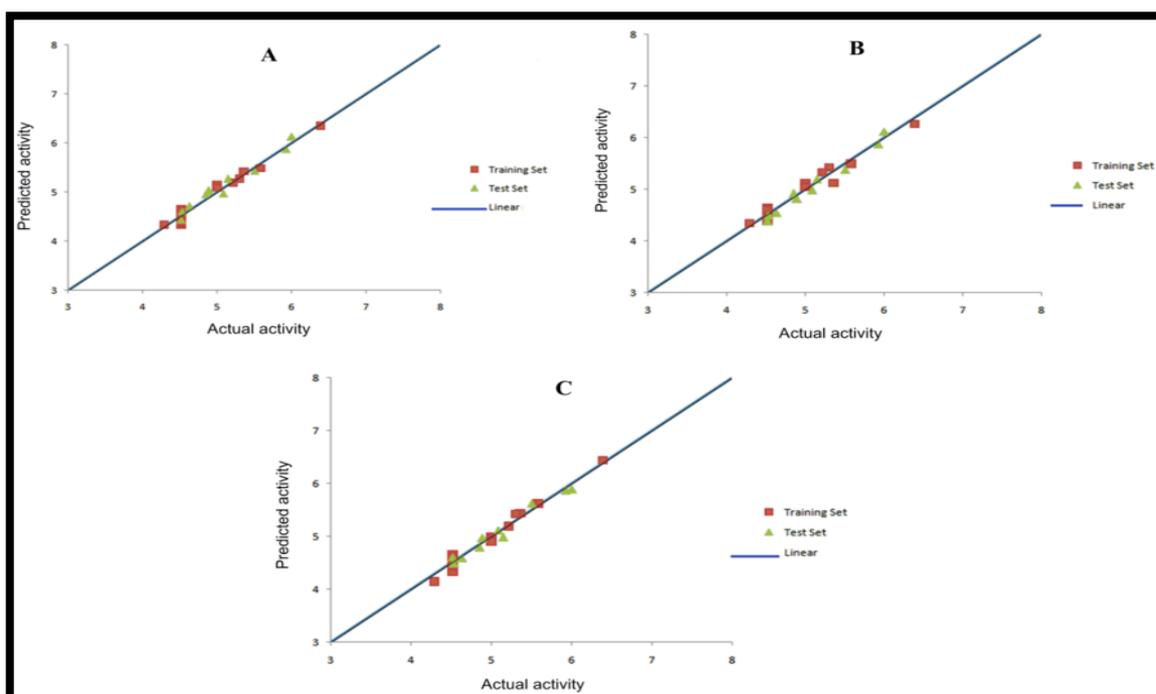


Fig. 1: Graph of Actual vs. Predicted activities for training and test set molecules by Multiple Linear Regression model (A), Principal Component Regression (B) and Partial Least Square Regression (C).

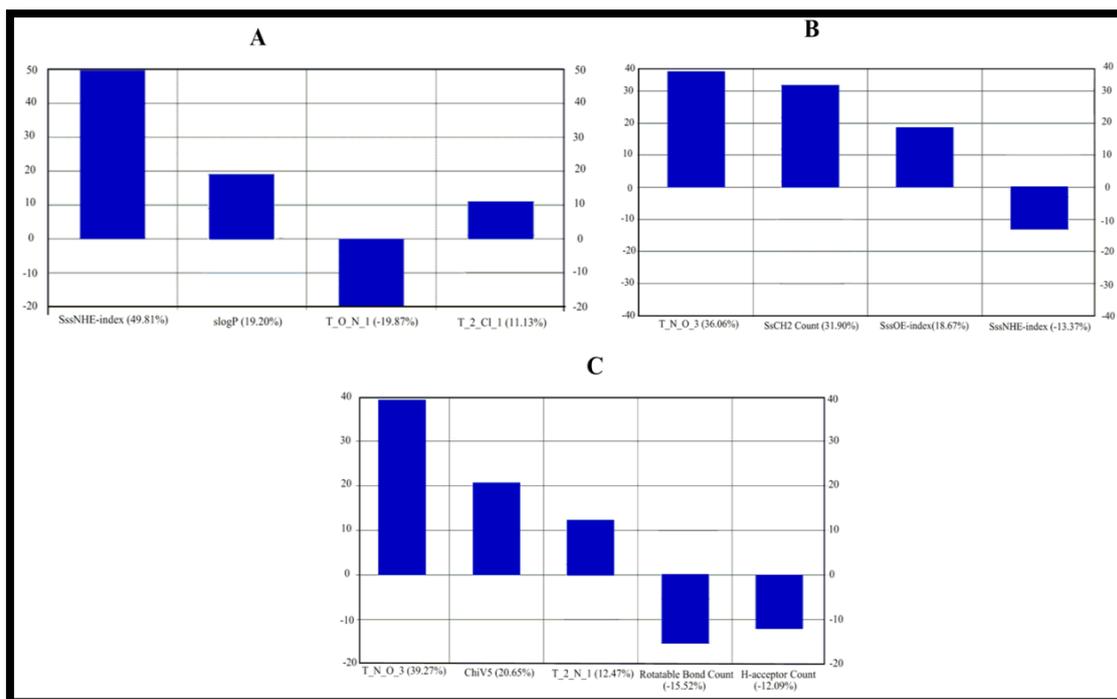


Fig. 2: Plot of percentage contribution of each descriptor in developed Multiple Linear Regression model (A), Principal Component Regression (B) and Partial Least Square Regression (C) model explaining variation in the activity

The statistically significant Model-I shows a positive correlation with SssNHE-index which is an electro topological state index for number of NH group connected with two single bonds. The positive correlation of the descriptor in the model indicates that presence of secondary amine in the basic skeleton 2-phenoxy-*N*-phenylacetamide is favorable for the activity (Compound 1-31). The positive correlation of slog P and alignment independent descriptor (T_2_Cl_1) reveals that electron withdrawing group such as Cl, F, Br which enhances the lipophilicity of compound at R1 favors the HIF-1 inhibitory activity (Compound 11, 12 and similar analogues) **Fig. 3**.

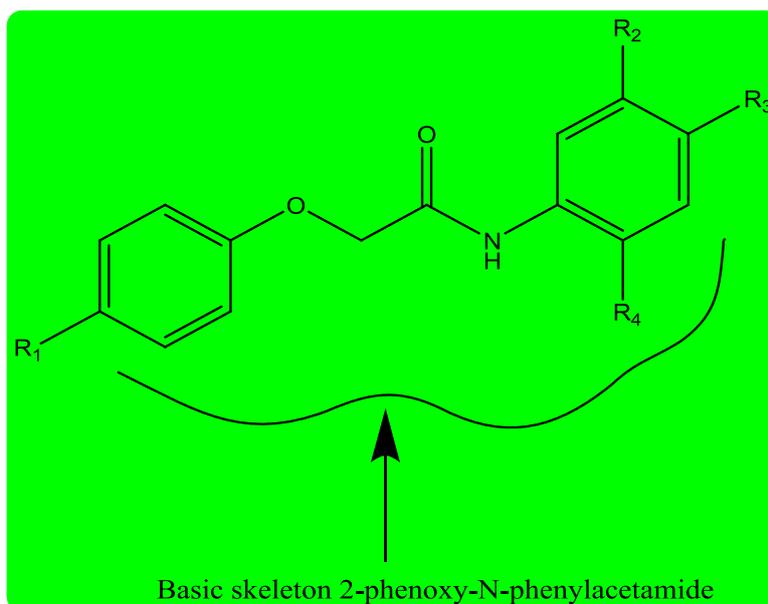


Fig. 3: Basic skeleton of 2-phenoxy-*N*-phenylacetamide with different substituent as R₁, R₂, R₃ and R₄

Inverse relationship of the alignment independent descriptor T_O_N_1 (This is the count of number of oxygen atom separated from nitrogen atom by one bond distance) shows that the distance between "oxy" group and carboxamide nitrogen should be more than one bond in the basic skeleton 2-phenoxy-*N*-phenylacetamide (Compound 1-31)

- 2D-QSAR PCR Model-II

$$pIC_{50} = 0.2520 (T_N_O_3) + 0.1325 (SsCH2count) + 0.0599 (SssOE-index) - 0.3027(SssNHEindex) + 5.5244 \dots\dots 5$$

The training (23 compounds) and test set [8 compounds (compound 2, 8, 10, 14, 18, 21, 30 and 31)] were divided by spherical exclusion method. Electrotopological state index descriptor SssNHE index is common in Model-I and Model-II and the definition of remaining descriptor which contributes to the HIF-1 inhibitory activity is mentioned here. The activity distribution plot and descriptor contribution plot for 2D-QSAR Model-II is shown in Fig. 1 and 2. Equation (5) shows the positive contribution of T_N_O_3 (This is the count of number of nitrogen atom separated from oxygen atom by three bond distance) shows that the three bond distance between "oxy" group and carboxamide nitrogen in the basic skeleton 2-phenoxy-*N*-phenylacetamide is favorable for the activity and any increase or decrease in bond distance may inversely affect over the biological activity (Compound 1-31). The next important estate number descriptor SsCH2count is the total number of CH₂ group connected with two single bond, reveals that bulkier admantine ring at R1 in which CH₂ group connected with two single bond is favorable for HIF-1 inhibitory activity (Compound 1,2,3 and similar analogues). Electrotopological state descriptor SssOE-index is the number of oxygen atoms connected with two single bonds reveals that "oxy" group in the basic skeleton 2-phenoxy-*N*-phenylacetamide is favorable for the activity (Compound 1-31) Fig.3

- 2D QSAR PLSR Model-III

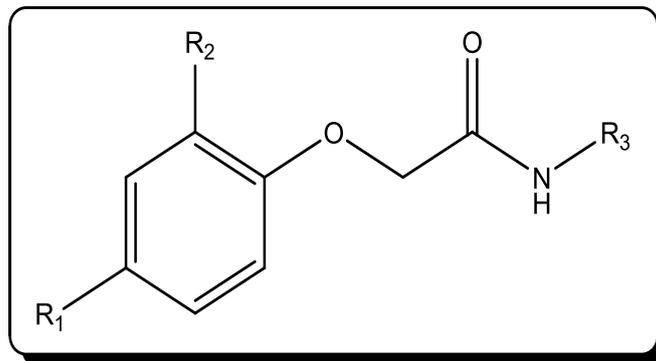
$$pIC_{50} = 0.3211(T_N_O_3) + 0.0994(ChiV5) + 0.1172(T_2_N_1) - 0.0625 (Rotatable bondcount) + 0.1240(H-acceptor count) + 4.0781 \dots\dots 6$$

It is simple to interpret 2D-QSAR PLSR equation where each descriptor's contribution can be seen by the magnitude and sign of its regression coefficient. The activity distribution plot and descriptor contribution plot for 2D-QSAR Model-III is shown in Fig. 1 and 2. The training (24 compounds) and test set [7 compounds (compound 5, 12, 16, 19, 22, 28 and 30)] were divided by spherical exclusion method. Alignment independent descriptor (T_N_O_3) is common in Model-II and Model-III and the definition of remaining descriptor which contributes to the HIF-1 inhibitory activity in Model-III is discussed in this section. The developed PLSR Model -III indicated that the direct correlation of alignment independent descriptor T_2_N_1 (This is the count of number of double bonded atom separated from nitrogen atom by one bond distance) explains that CONH₂ (carboxamide group) in which nitrogen is separated from double bonded carbonyl group by one bond distance in the basic skeleton of 2-phenoxy-*N*-phenylacetamide is favorable for the activity (Compound 1-31) as shown in Fig. 3. Inverse relation of rotatable bond count descriptor indicates that rigid bond is essential for HIF-1 inhibitory activity. Chiv5 signifies direct relationship of the atomic valence connectivity index with HIF-1 inhibitory activity. Positive correlation of H-acceptor count descriptor shows that hetero atoms such as O, N etc which are capable of forming H-bonding are favorable for drug receptor interaction.

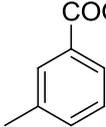
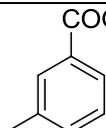
- **3D-QSAR modeling and its validation**

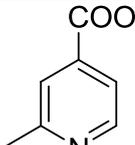
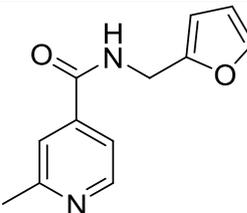
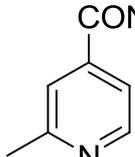
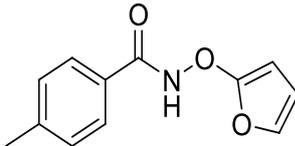
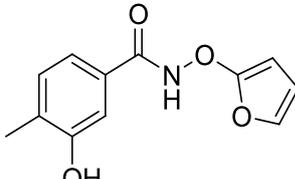
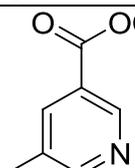
In the present study, kNN-MFA model is developed coupled with simulated annealing variable selection method to develop 3D-QSAR models of 2-phenoxy-*N*-phenylacetamide derivatives based on steric and electrostatic fields. The total data set was divided into training and test sets using the sphere exclusion algorithm for diversity of the sampling procedure. Compounds marked with (T) in Table 4 were selected as test-set molecules.

Table 4: Structure, experimental and predicted activity of 2-phenoxy-*N*-substituted acetamide analogues used in training and test set for HIF-1 α inhibition against Hep3B Cell lines by Model-IV



Sr. No.	R ₁	R ₂	R ₃	IC ₅₀ (μM)	pIC ₅₀		Residual
					Exp.	Pre.	
1	Admantine	-		9.9	5.00	5.01	-0.01
2	Admantaie	-		2.6	5.58	5.61	-0.03
3 ¹	Admantine	-		4.3	5.36	5.42	-0.06
4	-C(CH ₃) ₃	-		30	4.52	4.59	-0.07
5	-CH ₃	-		30	4.52	4.47	0.05
6	-NO ₂	-		30	4.52	4.49	0.03

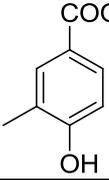
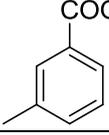
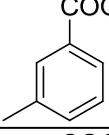
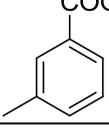
7	Admantine	-		5	5.30	5.28	0.02
8	-C(CH ₃) ₃	-		29	4.53	4.47	0.06

Sr. No.	R ₁	R ₂	R ₃	IC ₅₀ (μM)	pIC ₅₀		Residual
					Exp.	Pre.	
26	Admantine	-		1.03	6.00	6.09	-0.09
27 [†]	Admantine	-		10	5.00	5.09	-0.09
28	Admantine	-		3.1	5.50	5.61	-0.11
29 [†]	Cl	Cl		30	4.52	4.61	-0.09
30	Cl	Cl		23.2	4.63	4.7	-0.07
31	Admantine	-		5.9	5.22	5.19	0.03

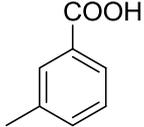
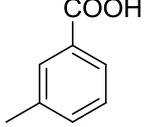
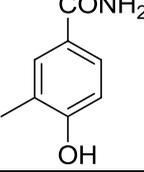
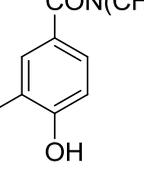
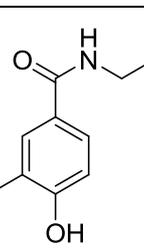
Exp. = Experimental activity, Pred. = Predicted activity;
a = $-\log(\text{IC}_{50} \times 10^{-6})$; T = Test Set

Model IV- kNN MFA Simulated Annealing variable selection

$\text{pIC}_{50} = \text{S}_{815} (30, 0000 \ 30, 0000) - \text{E}_{1166} (-10, 0000 \ -10, 0000) + \text{H}_{1547} (0.39700.4039) - \text{S}_{848} (-0.1636-0.0958) \dots\dots 7$

Sr. No.	R ₁	R ₂	R ₃	IC ₅₀ (μM)	pIC ₅₀		Residual
					Exp.	Pre.	
9	Admantine	-	 <chem>CC(=O)Oc1ccc(O)cc1</chem>	0.4	6.39	6.35	0.04
10 ¹	F	-	 <chem>CC(=O)Oc1ccc(F)cc1</chem>	30	4.52	4.59	-0.07
11	Cl		 <chem>CC(=O)Oc1ccc(Cl)cc1</chem>	30	4.52	4.60	-0.08
12	Br	-	 <chem>CC(=O)Oc1ccc(Br)cc1</chem>	30	4.52	4.48	0.04

For 3D-QSAR a kNN-MFA of substituted 2-phenoxy-*N*-phenylacetamide with reported activities against the HIF-1 receptor was performed. The Simulated annealing variable selection method resulted in several statistically significant models, of which the corresponding best Model-IV is reported herein. The model selection criterion is the value of q^2 , the internal predictive ability of the model, and that of pred_r^2 , the ability of the model to predict the activity of external test set. For activity against HIF-1 receptor, Model-IV was found to be statistically most significant, especially with respect to the internal predictive ability ($q^2 = 0.9672$) of the model. A data set of compounds containing seven molecules was selected as the test set from the original data of thirty one compounds for the validation experiments. The actual verses predicted graph by kNN-MFA is given in Fig. 6

13	I	-		7.2	5.14	5.09	0.05
14	-COCH ₃	-		30	4.52	4.59	-0.07
15	Admantine	-		30	4.52	4.60	-0.08
16	Admantine	-		30	4.52	4.48	0.04
17 ¹	Admantine	-		12.6	4.89	4.95	-0.06

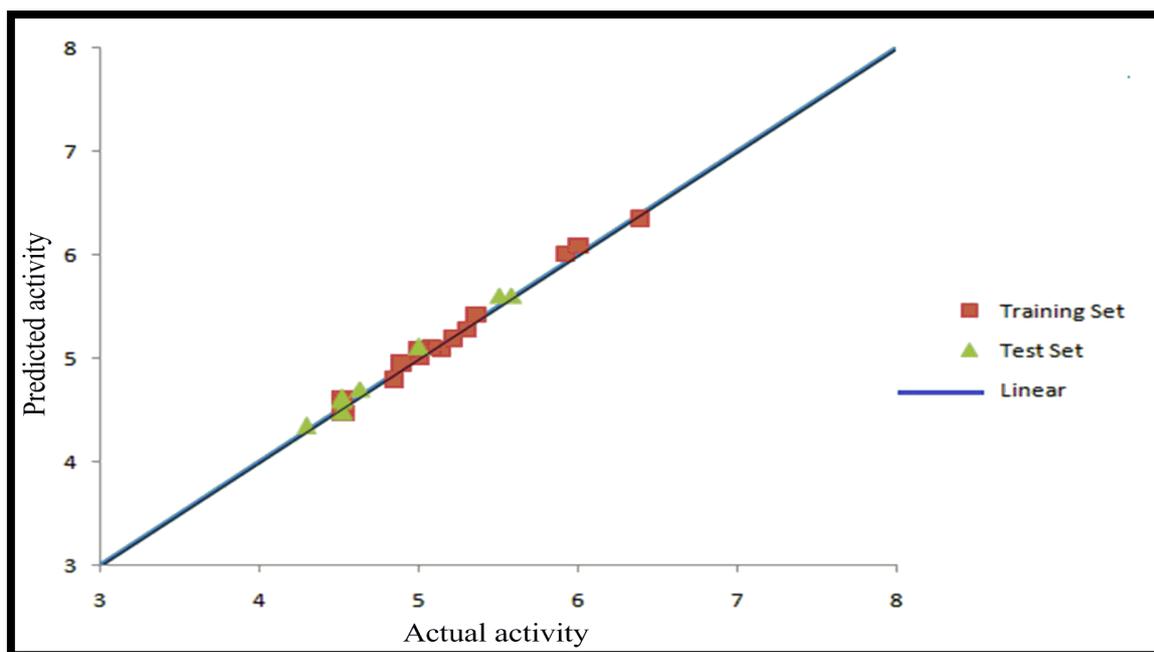


Fig. 6: Graph of Actual vs. Predicted activities for training and test set molecules by kNN-MFA model

and statistical data is shown in Table 2. The value of pred_r^2 was obtained for the test set and gave better results, with a value of 0.8480, which means 84% predictive power for the external test set. Thus, our model displays good predictivity in regular cross validation. 3D-QSAR based alignment shows a q^2 (cross validated r^2) with four descriptors namely S_815, E_1166, H_1547 and S_848 imply that interactions at the selected points are indeed significant for the structure-activity relationship. Steric, electrostatic and hydrophobic field energy of interactions between probe (CH₃) and compounds at their corresponding spatial grid points of 815, 1166, 1547 and 848. In the present study, negative range of electrostatic field descriptor E_1166 (-10, 0000 -10, 0000) indicates that the electronegative atom such as "oxy" group in 2-phenoxy-*N*-phenylacetamide skeleton is essential for the activity (Compounds 1-31) Fig.7.

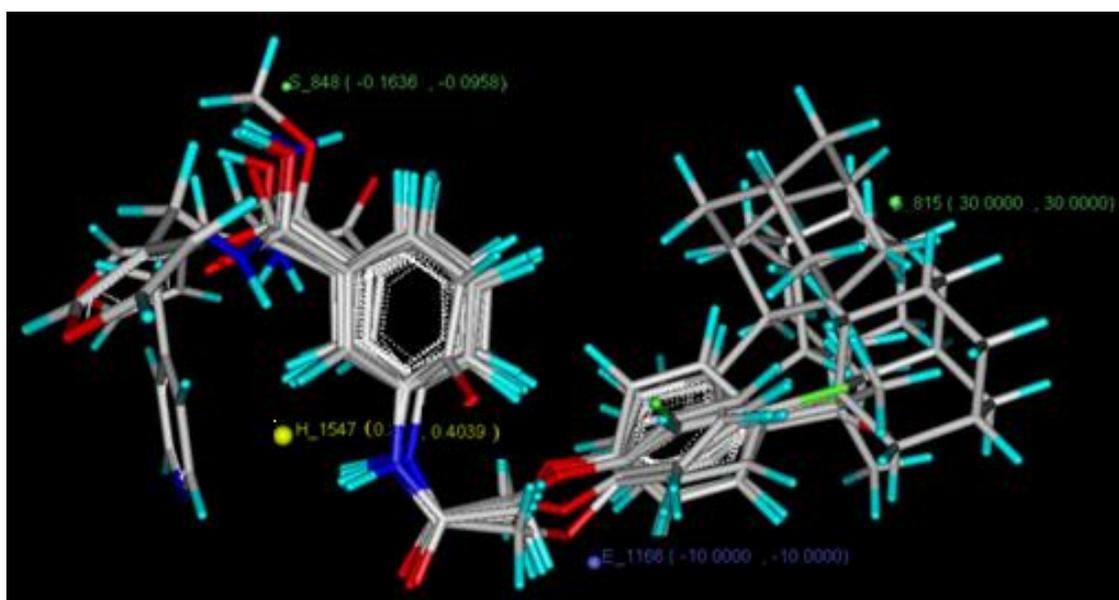


Fig. 7: Contribution plot steric and electrostatic field of interactions by kNN-MFA SA models
 Negative range S_848 (-0.1636 -0.0958) indicates that less bulky groups are preferred at R₂, R₃ and R₄ position of template. The positive description of steric descriptor S_815 (30, 0000 30, 0000)

indicates that presence of bulkier group at R1 favors the activity such as adamantane ring (Compound 1, 2, 3 and similar analogues). Positive contribution of H_1547 (0.3970 0.4039) indicates that groups which forms hydrogen bonding with receptor is favorable for the activity such as secondary amino group which is present almost in all the compounds.

CONCLUSION

The present work reveals how the HIF-1 inhibitory activities of substituted 2-phenoxy-*N*-phenylacetamide derivatives can be treated statistically to uncover the molecular characteristics which are essential for specific HIF-1 inhibitory activity.

In the present 2D-QSAR investigation, all proposed QSAR models were statistically significant. However, Model-1 by Multiple Linear Regression analysis could be considered as best one in terms of excellent internal and external predictive abilities. According to Model-1 (MLR) anticancer activity of 2-phenoxy-*N*-phenylacetamide derivatives was influenced by SssNHE-index, slogp, T_O_N_1 and T_2_Cl_1 descriptors, which help in understanding the effect of substituent at different position of 2-phenoxy-*N*-phenylacetamide derivatives.

The result obtained from 2D-QSAR study suggests that presence of secondary amine in the basic skeleton 2-phenoxy-*N*-phenylacetamide and electron withdrawing group such as Cl, F, Br which enhances the lipophilicity of compound at R1 is favorable for the HIF-1 inhibitory activity. Electron withdrawing groups enhances the lipophilicity of compounds and favors the HIF-1 α inhibition. It also suggests that distance between "oxy" group and carboxamide nitrogen should be more than one bond in the basic skeleton 2-phenoxy-*N*-phenylacetamide.

The 3D results reveal that electronegative atom such as "oxy" group in 2-phenoxy-*N*-phenylacetamide skeleton is essential for the activity and less bulky groups are preferred at R₂, R₃ and R₄ position of template. Presence of bulkier group at R1 favors the activity such as adamantane ring. 3D-QSAR studies also explain that groups forming hydrogen bonding with receptor is favorable for the activity such as secondary amino group which is present almost in all the compounds.

REFERENCES

1. Hockel, M.; Vaupel, P. Tumor Hypoxia: Definitions and current clinical biological, and molecular aspects. *J. Natl. Cancer Inst.*, 2001, 93, 266-276.
2. Bos, R.; Van der Groep, P.; Greijer, A. E.; Shvarts, A.; Meijer, S.; Pinedo, H. M.; Semenza, G. L.; Van Diest, P. J.; Van der wall, E. Levels of Hypoxia inducible factor-1 α independently predict prognosis in patients with lymph node breast carcinoma. *Cancer.*, 2003, 97, 1573-1581.
3. Brown, J. M.; Wilson, W. R. Exploiting tumour hypoxia in cancer treatment. *Nat. Rev. Cancer.*, 2004, 4, 437-447.
4. Moulder, J. E.; Rockwell, S. Tumor hypoxia: its impact on cancer therapy. *Cancer Metastasis Rev.*, 1987, 5, 313-341.
5. Wang, G. L.; Jiang, B.; Rue, E. A.; Semenza, G. L. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O₂ tension. *Proc. Natl. Acad. Sci. U.S.A.*, 1995, 92, 5510-5514.
6. Michel, G.; Minet, E.; Ernest, I.; Roland, I.; Durant, F.; Remacle, J.; Michiels, C. A. A model for the complex between the hypoxia-inducible factor-1 (HIF-1) and its consensus DNA sequence. *J. Biomol. Struct. Dyn.*, 2000, 18, 169-180.
7. Zagorska, A.; Dulak, J. Novel variable selection quantitative structure-property relationship approach based on the k-nearest-neighbor principle. *Acta. Biochim. Pol.*, 2004, 51, 563-585.
8. Poellinger, L.; Johnson, R. S. HIF-1 and hypoxic response: the plot thickens. *Curr. Opin. Genet. Dev.*, 2004, 14, 81-85.
9. Jones, M. K.; Szabo, I. L.; Kawanaka, H.; Husain, S. S.; Tarnawski, A. S. Von Hippel-Lindau tumor suppressor and HIF-1 α : new targets of NSAID inhibition of hypoxia-induced angiogenesis. *FASEB J.*, 2002, 16, 264-266.
10. Maxwell, P. H.; Dachs, G. U.; Gleadle, J. M.; Nicholls, L. G.; Harrisa, A. L.; Stratford, I. J.; Hankinson, O.; Pugh, C. W.; Ratcliffe, P. J. Hypoxia-inducible factor-1 modulates gene expression in solid tumors and influences both angiogenesis and tumor growth. *Proc. Natl. Acad. Sci. U.S.A.*, 1997, 94, 8104-8109.
11. Semenza, G. L. Regulation of mammalian O₂ homeostasis by hypoxia-inducible factor 1. *Annu. Rev. Cell Dev. Biol.*, 1999, 15, 551-578.
12. Semenza, G. L. HIF-1 and tumor progression: pathophysiology and therapeutics. *Trends Mol. Med.*, 2002, 8, S62-S67.

13. Birner, P.; Schindl, M.; Obermair, A.; Plank, C.; Breitenecker, G.; Oberhuber, G. Expression of Hypoxia-inducible factor 1 α in epithelial ovarian tumors: its impact on prognosis and on response to chemotherapy. *Cancer Res.*, 2000, 60, 4693-4696.
14. H. Zhong, A. M. De Marzo, E. Laughner, M. Lim, D. A. Hilton, D. Zagzag, P. Buechler, W. B. Isaacs, G. L. Semenza, J. W. Simons, *Cancer Res.* 1999, 59, 5830-5835.
15. Giaccia, A.; Siim, B. G.; Johnson, R. S. HIF-1 as a target for drug development. *Nat. Rev. Drug Discovery.*, 2003, 2, 803-811.
16. G. L. Semenza, Targeting HIF-1 for cancer therapy. *Nat. Rev. Cancer.*, 2003, 3, 721-732.
17. Lee, K.; Lee, J. H.; Boovanahalli, S. K.; Jin, Y.; Lee, M.; Jin, X.; Kim, J. H.; Hong, Y. S.; Lee, J. J. (Aryloxyacetyl amino)benzoic acid analogues: a new class of hypoxia-inducible factor-1 inhibitors. *J. Med. Chem.*, 2007, 50, 1675-1684.
18. Noolvi, M. N.; Patel, H. M.; Bhardwaj, V.; Chauhan, A. Synthesis and in vitro antitumor activity of substituted quinazoline and quinaxoline derivatives: search for anticancer agent. *Eur. J. Med. Chem.*, 2011, 46, 2327-2346
19. Noolvi, M. N.; Patel, H. M. 2D-QSAR studies on a series of quinazoline derivatives as tyrosine kinase (EGFR) inhibitor: an approach to design anticancer agents. *Lett. Drug Des. Discov.*, 2011, 7, 556-586.
20. Noolvi, M. N.; Patel, H. M.; Bhardwaj, B. An approach to design anticancer agents by 2D-QSAR studies on a series of quinazoline analogues as tyrosine kinase (erbB-2) inhibitors. *Med. Chem.*, 2011, 7, 200-212
21. Noolvi, M. N.; Patel, H. M. Synthesis, method optimization, anticancer activity of 2,3,7-trisubstituted quinazoline derivatives and targeting EGFR-tyrosine kinase by rational approach. *Arab. J. Chem.*, 2011, article in press.
22. Noolvi, M. N.; Patel, H. M.; Agrawal, S.; Zambre, A.; Badiger, A. Synthesis, antimicrobial and cytotoxic activity of novel 4 azetidone derivatives of 1H-benzimidazole. *Arab. J. Chem.*, 2011, article in press
23. Noolvi, M. N.; Patel, H. M.; Bhardwaj, B. 2D-QSAR studies on a series of 4-anilino quinazoline derivatives as tyrosine kinase (EGFR) inhibitor: an approach to design anti cancer agents. *Digest. J. Nanomat. and Biostr.*, 2011, 5, 387 – 401
24. Manjula, S. N.; Noolvi, M. N.; Parihar, K.V.; Manohara Reddy, S. A.; Ramani, V.; Gadad, A. K.; Singh, G.; Kutty, N. G.; Rao, M. Synthesis and anti-tumour activity of optically active thiourea and their 2-aminobenzothiazole derivatives: A novel class of anticancer agents. *Eur. J. Med. Chem.*, 2009, 44, 2923-2929.
25. Badiger, A. M.; Noolvi, M. N.; Nayak, P. V. QSAR study of Benzthiazole derivatives as p56 lck inhibitors. *Lett. Drug Des. Discov.*, 2006, 3, 550-560.
26. Opera, T. I., Waller, C. L.; Marshall, G. R. Three-dimensional quantitative structure– activity relationship of human immunodeficiency virus (I) protease inhibitors. 2. Predictive power using limited exploration of alternate binding modes. *J. Med. Chem.*, 1994, 37, 2206.
27. Golbraikh, A.; Tropsha, A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *J. Comput. Aided. Mol. Des.*, 2002, 16, 357-369.
28. Baysal, C.; Meirovitch, H. New conformational search method based on local torsional deformations for cyclic molecules, loops in proteins and dense polymer systems. *J. Chem. Phys.*, 1996, 105, 7868.
29. Baumann, K. An alignment-independent versatile structure descriptor for QSAR and QSPR based on the distribution of molecular features. *J. Chem. Inf. Comput. Sci.*, 2002, 42, 26–35.
30. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA): Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.*, 1988, 110, 5959-5967.
31. Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (COMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.*, 1994, 37, 24- 30
32. Zheng, W.; Tropsha, A. Novel Variable Selection Quantitative Structure-Property Relationship Approach Based on the k-Nearest- Neighbor Principle. *J. Chem. Inf. Comput. Sci.*, 2000, 40, 185-194.
33. Halgren T. A. Merck molecular force field II MMF94 vander Waals and electrostatic parameters for intermolecular interactions. *J. Comp. Chem.*, 1996, 17, 520-522.
34. Ajmani, S.; Jadhav, K.; Kulkarni, S. A. Three-dimensional QSAR using the k-nearest neighbor method and its interpretation. *J. Chem. Inf. Model.*, 2006, 46, 24–31.

35. Gehlhar, D. K.; Verkhivker, G. M.; Rejto, P.; Sherman, C. J.; Fogel, D. B.; Fogel, L.J.; Freer, S. T. "Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming". *Chem. Biol.*, 1995, 2, 317–324
36. Afantitis, A.; Melagraki, G.; Sarimveis, H.; Igglessi, M.O.; Kolliasnovel, G. A. A novel QSAR model for predicting the inhibition of CXCR3 receptor by 4-N-arryl-[1,4] diazepane. *Eur. J. Med.Chem.*, 2009, 44, 877-884.