QSAR Study on a Series of Aryl Carboxylic Acid Amide Derivatives as Potential Inhibitors of Dihydroorotate Dehydrogenase (DHODH)

Vivek K. Vyas* and Manjunath Ghate

Department of Pharmaceutical Chemistry, Institute of Pharmacy, Nirma University, Ahmedabad 382 481, Gujarat, India

Abstract: QSAR study was performed on a series of aryl carboxylic acid amide derivatives (62 analogs) to establish structural features required for human dihydroorotate dehydrogenase (*h*DHODH) inhibition. Statistical significant QSAR models were developed for the prediction of *h*DHODH inhibitory activity by applying MLR analysis ($r^2 = 0.851$ and $q^2 = 0.795$), PCR analysis ($r^2 = 0.713$ and $q^2 = 0.667$) and PLS analysis ($r^2 = 0.848$ and $q^2 = 0.802$). QSAR study emphasized the importance of topological, estate number, hydrophobic and alignment independent descriptors for the prediction of *h*DHODH inhibitory activity. SaasCcount descriptor suggested the presence of carbon atoms in five member aryl ring system. Positive impact of alignment independent descriptors reveals the presence of carbonyl oxygen and chloro group in this series of compounds. DistTopo signifies basic connectivity of atoms in the molecules. High degree of predictability of the proposed QSAR models offers a great potential for the design and development of potent *h*DHODH inhibitors.

Keywords: Dihydroorotate dehydrogenase (DHODH), QSAR, Multiple linear regression (MLR), Principle component regression (PCR), Partial least square (PLS) analysis, aryl carboxylic acid amide derivatives.

INTRODUCTION

Dihydroorotate dehydrogenase (DHODH) [EC 1.3.99.11] [1] is a fourth enzyme of pyrimidine *de-novo* synthesis, which catalyses the conversion of dihydroorotate to orotate, a rate-limiting step in pyrimidine biosynthesis [2]. Pyrimidines are required for the biosynthesis of DNA, RNA, glycoproteins and phospholipids [3]. There are two routes for the biosynthesis of pyrimidines in most of the organisms; salvage pathways and *de novo* synthesis from small metabolites. Rapidly proliferating human cells such as activated T lymphocytes [4] and cancer cells [5] are heavily dependent on *de novo* nucleotide synthesis to meet their growth requirements [6]; therefore *de novo* pyrimidine biosynthesis represents an attractive and selective target for the development of new therapeutics as anticancer, antimalarial and antimicrobial agents. Inhibition of DHODH leads to reduced levels of essential pyrimidine nucleotides. Inhibitors of *h*DHODH have proven efficacy for the treatment of cancer [7,8] and immunological disorders, such as rheumatoid arthritis and multiple sclerosis [9-12]. Leflunomide [13,14] and brequinar [15] are two examples of such compounds. DHODH catalyses the only oxidation/reduction reaction in the pyrimidine nucleotide biosynthesis: Oxidation of dihydroorotate (DHO) to orotate (ORO) and reduction of flavin mononucleotide (FMN) to dihydroflavin mononucleotide (FMNH$_2$) (Fig. **1**), comprise two half reactions of the redox couple. The co-substrate electron acceptor used by DHODH varies in different organisms. In the human enzyme, flavin cofactor is FMN and ubiquinone (CoQ) is second substrate. DHODH enzymes are divided into two family based upon their localization, amino acid sequence, substrate/cofactor dependence and cellular localization [16]. Family-1 enzymes are located in the cytosol, electron acceptors involved in second half reaction of redox reaction are either fumarate or NAD$^+$ whereas family-2 enzymes transfer electrons to ubiquinone (CoQ), to which *h*DHODH belongs [17,18]. Many excellent research publications characterized exponential growth in therapeutic potential of DHODH inhibitors. QSAR study on a series of aryl carboxylic acid amide derivatives quantifies necessary structural and physicochemical requirements for design and development of potent *h*DHODH inhibitors.

MATERIAL AND METHODS

QSAR is the study of the quantitative relationship between the experimental activity of a set of compounds and their physicochemical properties using statistical methods [19,20]. The experimental information associated with biological activity, which is used as dependent variables in building a QSAR model [21,22]. The calculated parameters are various descriptors, which are indicative of molecular structure, and used as independent variable. QSAR is being used to gain insight into the interaction of molecules with macromolecules and macromolecular systems [23,24]. Modification in molecular structures may lead to the development of potent and selective drug candidate. In the present study, all the computational work was performed using Vlife MDS QSAR plus software on a HP computer

*Address correspondence to this author at the Department of Pharmaceutical Chemistry, Institute of Pharmacy, Nirma University, S.G. Highway, Chharodi, Ahmedabad, 382 481, Gujarat, India; Tel: +91 9624931060; Fax: +91 2717 241916; E-mails: vicky_1744@yahoo.com, vivekvyas@nirmauni.ac.in
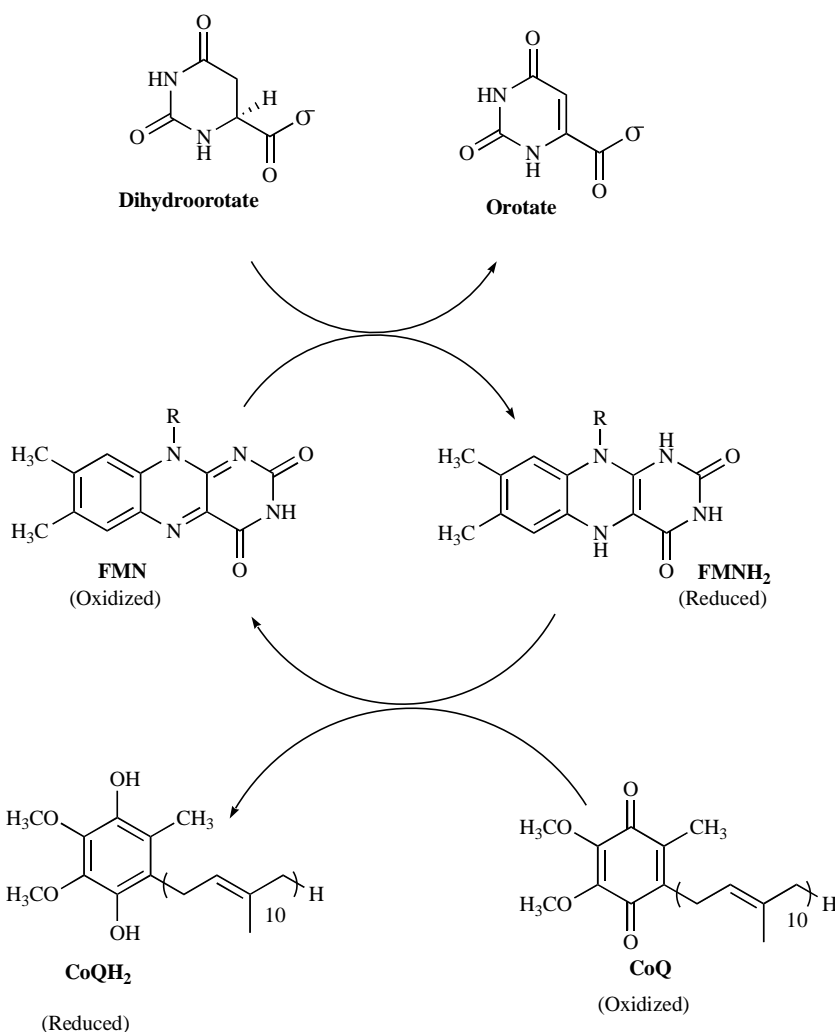
Fig. (1). Reactions catalyzed by DHODH.

with Core2 Duo processor and a window XP operating system.

Data Preparation

A data set of 62 compounds (*h*DHODH inhibitors) was collected from the literatures [25-27]. The experimental IC_{50} value (50% inhibitory concentration of the enzyme) was evaluated by Leban, *et al.* [25-27] in an enzyme assay by using N-terminally truncated recombinant *h*DHODH. The enzyme inhibition data (IC_{50} value) were converted pIC_{50} and subsequently used as the dependent variable for QSAR study (Table **1**). All the compounds were sketched using the 2D draw application and converted to 3D structures. Energy minimization and geometry optimization were conducted using Merck molecular force field (MMFF) as force field and charge, maximum number of cycles were 1000, convergence criteria (RMS gradient) was 0.01 and medium's dielectric constant of 1 by batch energy minimization method [28]. Energy-minimized geometry was used for calculation of descriptors. A total of 208 2D descriptors were calculated, which encoded different aspects of molecular structure and consists of electronic, thermodynamic, spatial and structural descriptors like retention index (chi), atomic valence connectivity index (chiV), path count, chain path count, cluster, path cluster, element count, estate number, molecular weight, molecular refractivity, logP and topological index. Various Baumann alignment-independent (AI) descriptors were also calculated.

Selection of Training and Test Set

Dataset of 62 molecules was divided into training (44) and test (18) set compounds. Selection of training set and test set molecules was done on the basis of Sphere Exclusion (SE) method [29] for MLR, PCR and PLS model with dissimilarity value of 2.5, where dissimilarity value gives sphere exclusion radius. Sphere exclusion is dissimi-larity based compound selection method. Sphere exclusion consists in selecting training and test molecules, whose similarities with each of the other selected molecules are not higher than the defined threshold. Therefore, each training set molecule creates a sphere around itself so that any test set molecules inside the sphere are excluded from selection. Radius of the sphere is an adjustable parameter; increase in dissimilarity value will lead to increase in number of molecules in a test set. Unicolumn statistics of test and training sets (Table **2**) showed accurate selection of test and

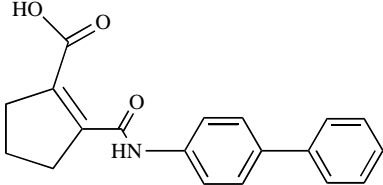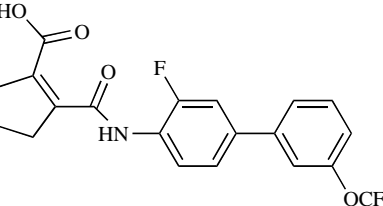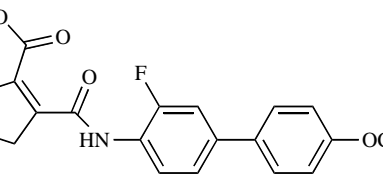**Table 1. Structures and Activity of Aryl Carboxylic Acid Amide Derivatives**
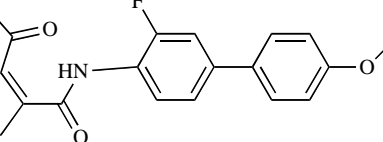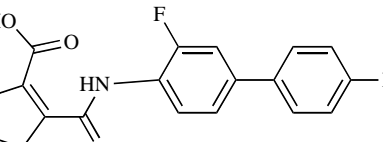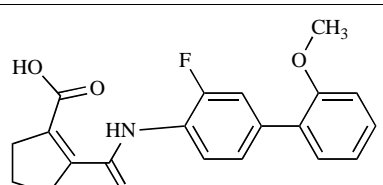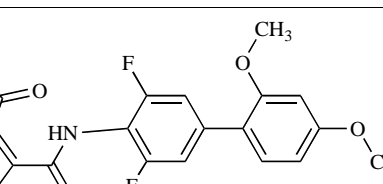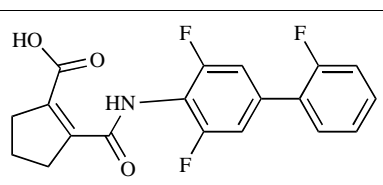
S. no.	Compounds	IC$_{50}$[a]	Observed pIC$_{50}$[b]	Predicted pIC$_{50}$		
				MLR	PCR	PLS
1*		410 nM	6.388	5.813	5.985	5.731
2		280 nM	6.551	6.45	6.65	6.471
3		5700 nM	5.244	6.216	6.524	6.266
4		1700 nM	5.77	6.409	6.54	6.402
5		735 nM	6.134	6.282	6.745	6.262
6		42 nM	7.377	6.893	6.748	6.849
7*		90 nM	7.046	7.487	7.15	7.464
8		45 nM	7.347	7.279	7.224	7.188

Table 1. contd…
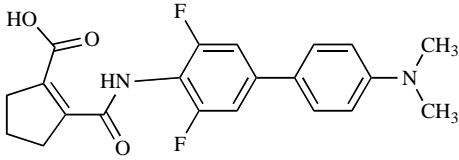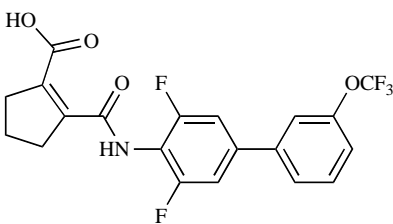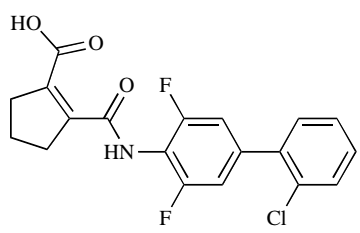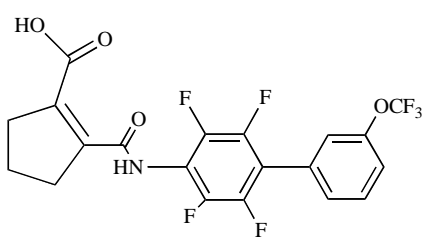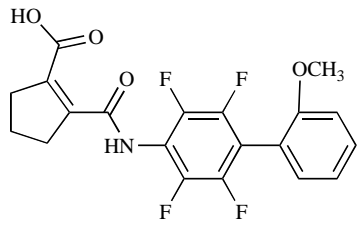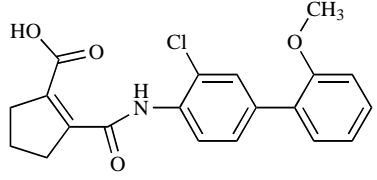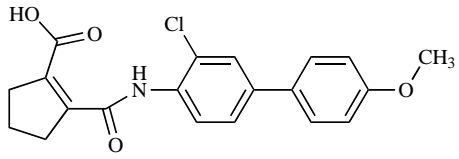
S. no.	Compounds	IC_{50}[a]	Observed pIC_{50}[b]	Predicted pIC_{50}		
				MLR	PCR	PLS
9		610 nM	6.215	6.148	6.646	6.362
10		33 nM	7.482	7.051	7.077	7.069
11		18 nM	7.745	7.67	7.18	7.646
12*		7 nM	8.155	8.13	7.944	8.121
13		8 nM	8.097	8.139	8.075	8.088
14		310 nM	6.509	6.249	6.1	6.202
15*		8400 nM	5.076	5.663	5.391	5.648
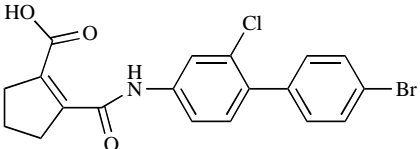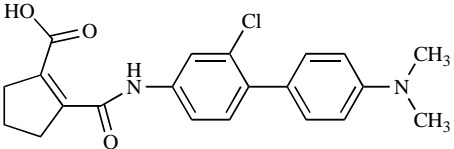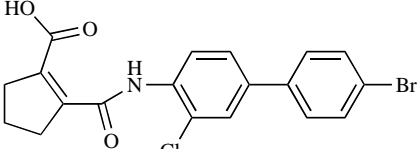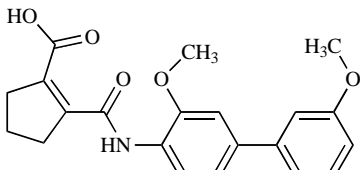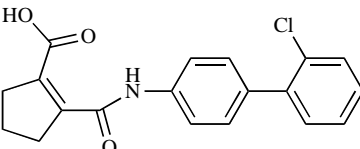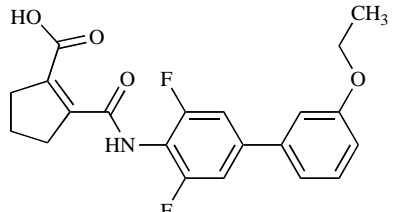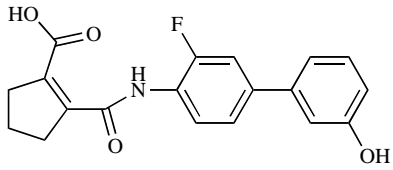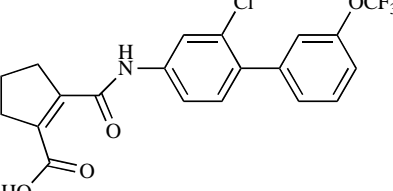
S. no.	Compounds	IC_{50}[a]	Observed pIC_{50}[b]	Predicted pIC_{50}		
16		70 nM	7.155	7.127	7.33	7.12
17		50 nM	7.301	7.185	7.294	7.18
18*		2900 nM	5.538	5.698	5.583	5.68
19		170 nM	6.771	6.044	6.611	6.096
20		225 nM	6.648	6.753	6.467	6.695
21		17 nM	7.767	7.213	7.103	7.224
22*		610 nM	6.215	6.148	7.136	6.126
23		20 nM	7.699	6.560	7.446	6.686

Table 1. contd…

S. no.	Compounds	$IC_{50}{}^{a}$	Observed $pIC_{50}{}^{b}$	Predicted pIC_{50}		
24*		150 nM	6.824	6.81	6.857	6.773
25		290 nM	6.538	6.55	5.806	6.467
26		370 nM	6.432	6.295	6.86	6.233
27*		390 nM	6.409	6.333	5.855	6.262
28		840 nM	6.076	6.626	6.57	6.669
29		80 nM	7.097	6.924	7.281	6.93
30*		180 nM	6.745	6.615	6.4	6.689
31		435 nM	6.362	6.335	6.035	5.856
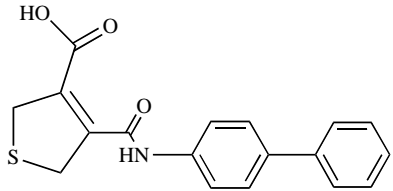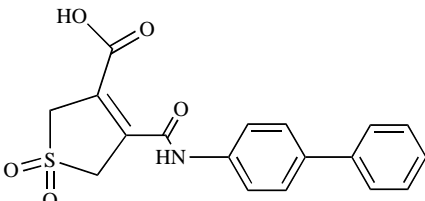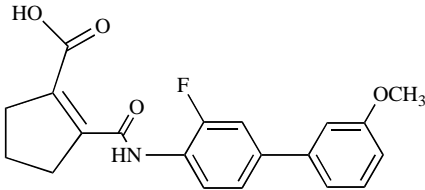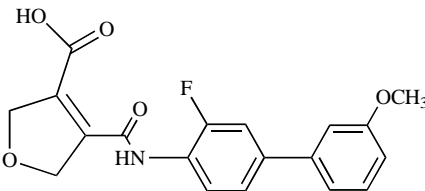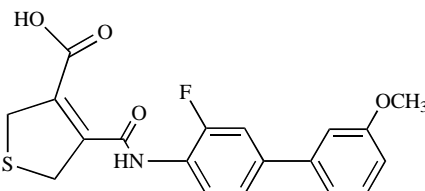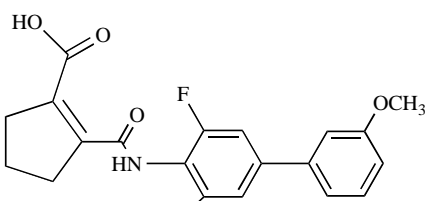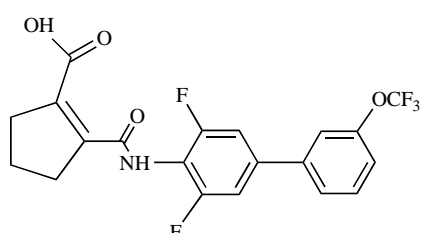
S. no.	Compounds	IC$_{50}$[a]	Observed pIC$_{50}$[b]	Predicted pIC$_{50}$		
32*		0.667 μM	6.176	5.894	5.923	5.896
33		3.8 μM	5.421	5.57	5.585	5.649
34*		0.134 μM	6.873	6.651	6.672	6.614
35		0.36 μM	6.444	6.415	6.624	6.331
36		0.131 μM	6.883	6.897	6.615	6.922
37		0.011 μM	7.959	7.895	7.022	6.896
38*		0.033 μM	7.482	7.321	6.99	7.313

Table 1. contd…
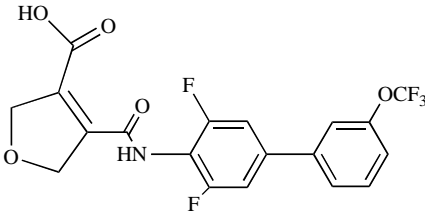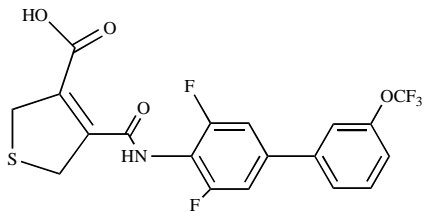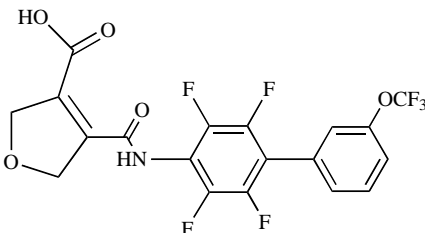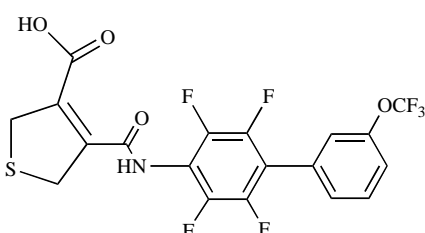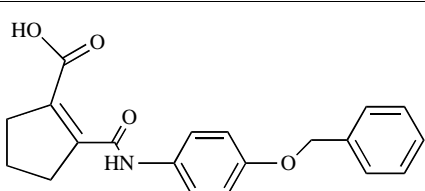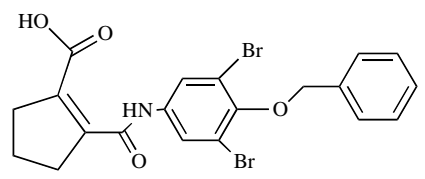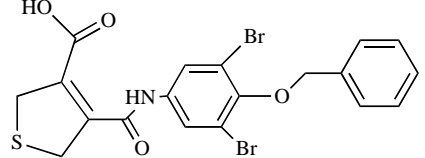
S. no.	Compounds	$IC_{50}{}^a$	Observed $pIC_{50}{}^b$	Predicted pIC_{50}		
39		0.205 µM	6.689	6.872	6.942	6.843
40		0.015 µM	7.824	7.71	7.08	7.744
41		0.058 µM	7.237	7.493	7.68	7.509
42*		0.004 µM	8.398	8.335	7.865	8.395
43		2 µM	5.699	5.253	5.861	5.28
44		0.127 µM	6.897	6.837	6.625	6.91
45		0.173 µM	6.762	7.051	6.77	7.17

Table 1. contd…

S. no.	Compounds	IC$_{50}$[a]	Observed pIC$_{50}$[b]	Predicted pIC$_{50}$		
46		0.105 μM	6.979	7.714	7.664	7.792
47*		0.011 μM	7.959	7.959	7.958	7.859
48		0.041 μM	7.388	6.583	7.432	6.655
49		0.11 μM	6.959	7.049	7.458	7.156
50*		0.68 μM	6.168	6.219	6.431	6.238
51		1.46 μM	5.836	6.245	6.442	6.259
52		303 nM	6.519	6.485	6.796	6.423
53		44 nM	7.357	7.604	7.58	7.559

Table 1. contd…
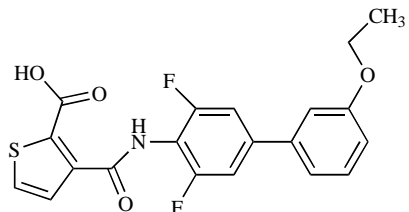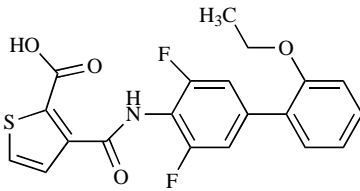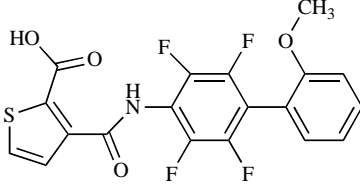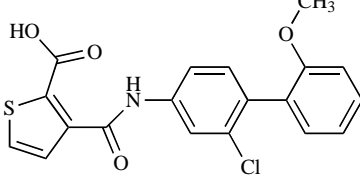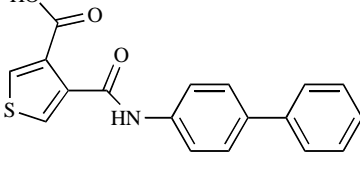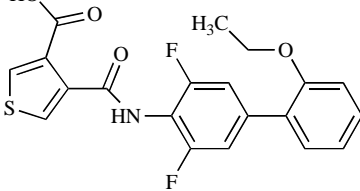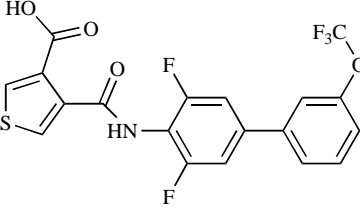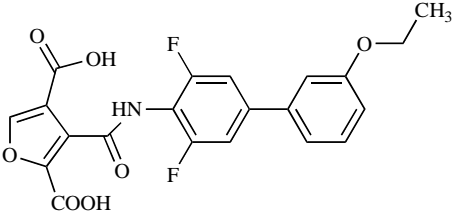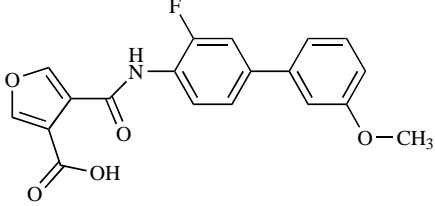
S. no.	Compounds	IC$_{50}$[a]	Observed pIC$_{50}$[b]	Predicted pIC$_{50}$		
54*		3 nM	8.523	8.395	8.132	8.356
55		9 nM	8.046	8.52	8.227	8.461
56		1 nM	9	9.228	9.183	9.142
57		12 nM	7.921	8.515	8.223	8.474
58*		1000 nM	6	6.424	6.75	6.38
59		10 nM	8	7.915	7.825	7.954
60*		10 nM	8	7.928	7.938	7.965

Table 1. contd…

S. no.	Compounds	IC$_{50}$[a]	Observed pIC$_{50}$[b]	Predicted pIC$_{50}$		
61		16 nM	7.796	7.69	7.963	7.692
62		340 nM	6.469	6.469	6.828	5.775

[a]Experimental *h*DHODH inhibitory activity (IC$_{50}$ μM/nM)
[b] -log (IC$_{50}$) nM/μM
*Test set

training sets, as maximum of the training set was more than that of test set and the minimum of training set was less than or equal to that of test set.

Table 2. Unicolumn Statistics of Training and Test Set Compounds

Set	Average	Max	Min	Sum
Training	6.9567	9.0000	5.0757	306.0953
Test	6.9038	8.0969	5.4202	124.2684

Statistical Computation

QSAR models were developed using MLR, PCR and PLS regression methods coupled with forward-backward variable selection method. Statistical measures were used for the evaluation of QSAR models; the number of compounds in regression *n*, regression coefficient r^2, number of descriptors in a model *k*, F-test (Fisher test value) for statistical significance *F*, cross validated correlation coefficient q^2, predictive squared correlation coefficients *pred_r^2*, coefficient of correlation of predicted data set *pred_r^{2}se* and standard error of estimation r^2 *se* and q^2 *se*.

Multiple Linear Regression (MLR) Analysis

MLR is a method used to model linear relationship between a dependent variable Y (*h*DHODH inhibitory activity) and independent variables X (2D descriptors). MLR is based on least squares: the model is fit such that sum-of-squares of differences of observed and predicted value is minimized. MLR estimates values of regression coefficients (r^2) by applying least squares curve fitting method. The model creates a relationship in the form of a straight line (linear) that best approximates all the individual data points. In regression analysis, conditional mean of dependant variable (pIC$_{50}$) Y depends on (descriptors) X. MLR analysis

extends this idea to include more than one independent variable. Regression equation takes the form $Y = b1*x1 + b2*x2 + b3*x3 + c$ where, Y is dependent variable, 'b's are regression coefficients for corresponding 'x's (independent variable), 'c' is a regression constant or intercept [30,31].

Principal Component Regression (PCR) Method

PCR is a data compression method based on the correlation among dependent and independent variables. PCR provides a method for finding structure in data sets. Its aim is to group correlated variables, replacing the original descriptors by new set called principal components. These principal components uncorrelated and are built as a simple linear combination of original variables. It rotates the data into a new set of axes such that first few axes reflect most of the variations within the data. First principal component (PC$_1$) is defined in the direction of maximum variance of the whole data set. Second principal component (PC$_2$) is the direction that describes the maximum variance in orthogonal subspace to PC$_1$. Subsequent components are taken orthogonal to those previously chosen and describe maximum of remaining variance, by plotting the data on new set of axes, it can spot major underlying structures automatically. Value of each point, when rotated to a given axis, is called the principal component value. PCA selects a new set of axes for the data. These are selected in decreasing order of variance within the data. Purpose of principal component PCR is the estimation of values of a dependent variable on the basis of selected principal components (PCs) of independent variables [32].

Partial Least Squares (PLS) Regression Method

PLS analysis is a popular regression technique, can be used to relate one or more dependent variable (Y) to several independent (X) variables. PLS relates a matrix Y of dependent variables to a matrix X of molecular structure

descriptors. PLS is useful in situations where the number of independent variables exceeds the number of observation, when X data contain colinearties or when N is less than 5 M, whares N is number of compound and M is number of dependant variable. PLS creates orthogonal components by using existing correlations between independent variables and corresponding outputs while also keeping most of the variance of independent variables. The aim of PLS regression is to predict the activity (Y) from X and to describe their common structure [33]. PLS is probably the least restrictive of various multivariate extensions of MLR model. PLS is a method for constructing predictive models when factors are many and highly collinear. QSAR models were generated using pIC_{50} values as dependent variable and various descriptors values as independent variables. The cross-correlation limit was set at 0.5, number of variables in final equation at six in MLR and five in PCR and PLS. Term selection criteria was set as r^2 and F-test, 'in' at 4 'out' at 3.99. Variance cutoff was set at 0, scaling to autoscaling and number of random iterations to 10.

Validation of QSAR Models

The definitive validity of the model is examined by mean of external validation (q^2), which evaluates how well the equation generalizes [34]. The training set was used to derive an adjustment model, which was used to predict the activity of a test set. The predictive power of the equations was validated using predictive squared correlation coefficients r^2_{pred}. Statistical significance of these models was further supported by a plot of observed versus predicted activity of training and test set compounds, which provides an idea that how fit model was trained and how well it predicts activity of test set (Fig. **2**). Nearness of observed to predicted activity reported in Table **1** also adds to this fact. Contribution charts for all the significant models are presented in Fig. (**3**), which gives percentage contribution of descriptors used in deriving the QSAR models.

RESULTS AND DISCUSSION

Generation of QSAR Models

QSAR study of aryl carboxylic acid amide derivatives resulted in several models. Statistically significant QSAR models were selected for discussion.

Model -1 (MLR)

pIC_{50} = -0.1868 (T_C_O_7) -0.1617 (MHHD) -0.4014 (T_C_N_3) +0.4123 (chiV1) + 0.5853 (SaasCcount) +1.8345 (DistTopo) +5.1433

where, n = $44_{training}$ and 18_{test}, DF = 37, r^2 = 0.851, q^2 = 0.795, F test = 35.278, r^2 se = 0.375, q^2 se = 0.441, pred_r^2 = 0.684, pred_r^2se = 0.351

Model -2 (PCR)

pIC_{50} = +0.3964 (SaasCcount) -1.1575 (T_N_Cl_3) +0.5060 (T_O_Cl_6) -0.0982 (T_C_O_7) -0.0732 (MHHD) +5.8788

where, n = $44_{training}$ and 18_{test} , DF = 38, r^2 = 0.713, q^2 = 0.667, F test = 23.603, r^2 se = 0.373, q^2 se = 0.309, pred_r^2 = 0.757, pred_r^2se = 0.359

Model -3 (PLS)

pIC_{50} = +0.6026 (SaasCcount) -0.1743 (T_C_O_7) -0.1378 (MHHD) -0.3943 (T_C_N_3) +1.6959 DistTopo) +0.4691 (chiV1) + 5.5180

where, n = $44_{training}$ and 18_{test} , DF = 41, r^2 = 0.848, q^2 = 0.802, F test = 114.342, r^2 se = 0.361, q^2 se = 0.311, pred_r^2 = 0.733, pred_r^2se = 0.428

Model-2 (PCR)

Model-3 (PLS)
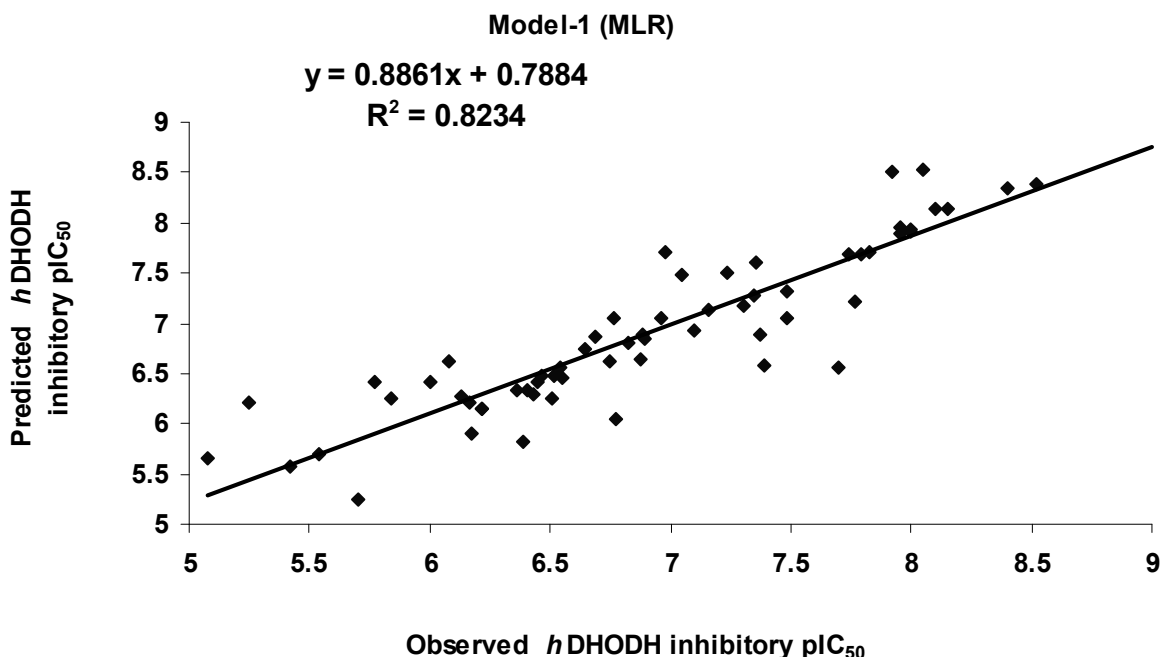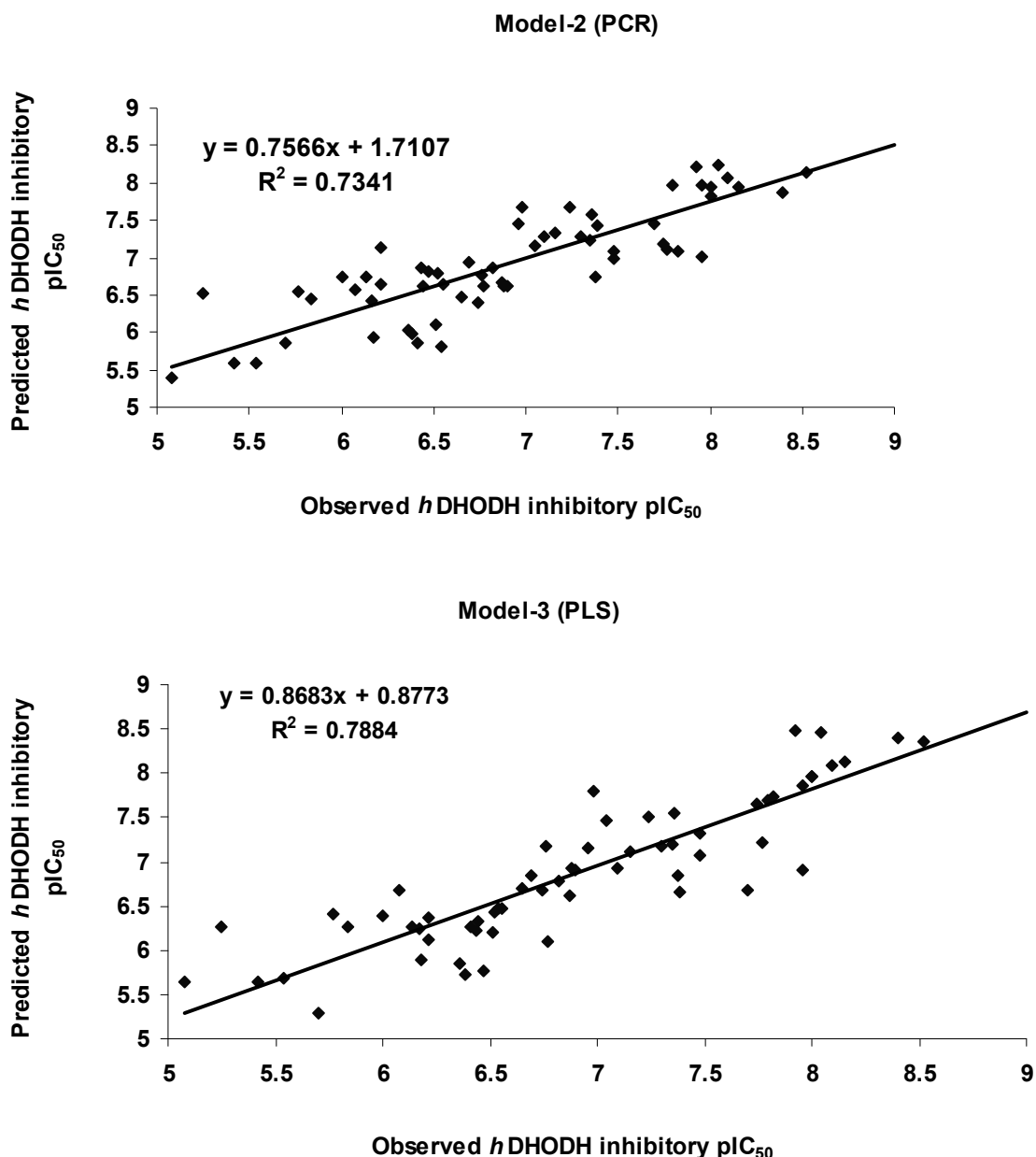
Fig. (2). Graphs of observed vs. predicted hDHODH inhibitory pIC$_{50}$ using model-1, 2 and 3.

In above QSAR models, r^2 is a correlation coefficient that multiply by one hundred to give explained variance in biological activity. Predictive ability of generated QSAR models was evaluated by q^2 employing leave-one-out method. F value reflects ratio of variance explained by models and variance due to error in regression. High F value indicates that model is statistically significant. Low standard error (SE) of estimation indicted by r^2 se and q^2 se, suggested that models are statistically significant. Predictive ability of QSAR model was also confirmed by external validation of test set compounds denoted by pred_r^2. Among these three QSAR models PLS has come out with very good results as compare to other two models. Results of PLS analysis showed very good predictive ability as indicted by r^2, pred_r^2, and q^2 values. Statistical parameters of significant QSAR models are shown in Table **3**.

Interpretaion of QSAR Models

Descriptors used in generation of QSAR models are shown in Fig. (**3**) and detailed description is given in Table **4**. MLR (Model-1) and PLS (Model-3) indicated positive contribution of chiV1, DistTopo, SaasCcount and negative contribution of T_C_O_7, most hydrophobic hydrophilic distance (MHHD), T_C_N_3, where as PCR model indicated positive contribution of T_O_Cl_6 along with SaasCcount and negative contribution of T_N_Cl_3 and MHHD.

ChiV1 is a topological descriptor, which signifies atomic valence connectivity index (order 1). This is calculated as the sum of 1/sqrt (vivj) over all bonds between heavy atoms i and j where i < j [35]. The valence degree (also called valence connectivity index), defined by Randic [36] in his

Model-1 (MLR)

Model-2 (PCR)

Model-3 (PLS)

Fig. (3). Contribution charts of the QSAR models.

Table 3. Statistical Parameters Used in QSAR Models

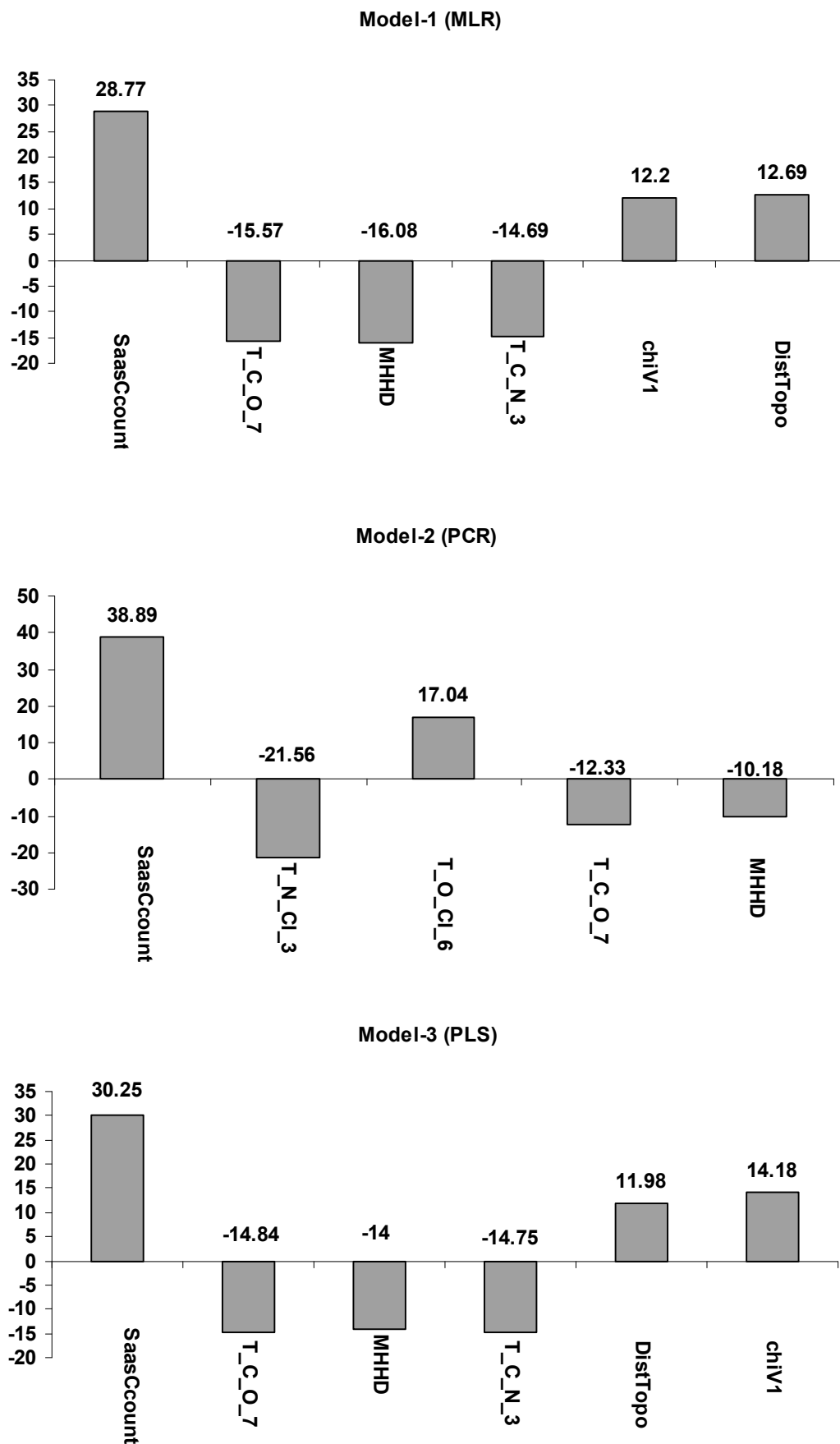| S. No.	Statistical Parameters	Model-1 (MLR)	Model-2 (PCR)	Model-3 (PLS)
1	n	$44_{Training}$ and 18_{Test}	$44_{Training}$ and 18_{Test}	$44_{Training}$ and 18_{Test}
2	DF	37	38	41
3	r^2	0.851	0.713	0.848
4	q^2	0.795	0.667	0.802
5	F test	35.278	23.603	114.342
6	r^2 se	0.375	0.373	0.361
7	q^2 se	0.441	0.309	0.311
8	pred_r^2	0.684	0.757	0.733
9	pred_r^2se	0.351	0.359	0.428

MLR = Multiple Linear Regression, PCR = Principal Component Regression, PLS = Partial Least Squares, n = number of molecules, DF =degree of freedom, r^2 =coefficient of determination, q^2 = cross validated r^2, pred_r^2 = r^2 for test set, pred_r^2se = coefficient of correlation of predicted data set.
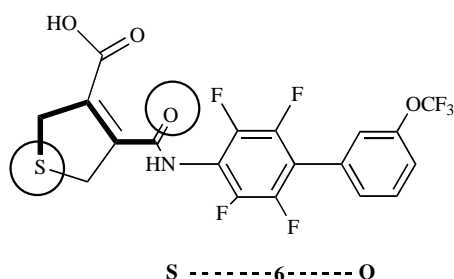

Fig. (4). Example of DistTopo descriptor: two atom types, carbonyl oxygen and aromatic sulfur, are separated by the shortest chemical graph path of six.

branching indices, is used to describe electronic environment and bonding configuration of each non-hydrogen atom (heavy atom) in a molecule for example carbon valence connectivity index takes into account only bonds between carbon atoms. ChiV1 reveals molecular connectivity for all heavy atoms and their bonding configuration in the molecules.

SaasCcount is an estate contribution descriptor, which defines total number of carbon connected with one single bond along with two aromatic bonds. SaasCcount plays most important role in determining *h*DHODH inhibitory activity. It suggested the importance of number of carbon atoms in aryl ring for good *h*DHODH inhibitory activity.

DistTopo is a distance based topological descriptors. Topological indices are numerical values associated with chemical constitutions for the purpose of correlating chemical structure with activity (*h*DHODH inhibition). Distance based topological descriptors are defined by their atom types and topological distance. DistTopo defined by two atom types and shortest path separation (or graph distance) between the atoms (substructure). Graph distance is defined as smallest number of atoms along the path connecting two atoms in a molecular structure. DistTopo descriptors can be generated for any molecular structure. An example of a DistTopo descriptor is shown in Fig. (**4**). Distance between two atoms is measured and then assigned

into one or two distance bins. Distance based topology index quantify chemical structures by translation of chemical structures into entity where some parts (atoms or atomic cores, atom pairs, branching, bonded or not bonded atoms) are involved in a binary relationship. Such a binary relationship depicted the basic connectivity of atoms in molecules.

Baumann's [37] alignment independent (AI) topological descriptors T_C_O_7 and T_C_N_3 contributed negatively to MLR and PLS model, where as T_O_Cl_6 contributed positively to PCR model. Alignment independent descriptors can be generated by considering topology of a molecule, atom type and bond. For calculation of alignment independent descriptors every atom in the molecule was assigned at least one and at most three attributes. First attribute is 'T-attribute' to thoroughly characterize topology of a molecule. Second attribute is atom type, atom symbol is used here. Third attribute is assigned to atoms taking part in a double or triple bond. After the all atoms have been assigned their respective attributes, selective distance count statistics for all combinations of different attributes are computed. A selective distance count statistic 'XY2' (e.g. 'TOPO2N3) counts all the fragments between start atom with attribute 'X' (e.g. '2'double bonded atom) and end atom with attribute 'Y' (e.g. 'N') separated by graph distance 3. Graph distance can be defined as the smallest number of atoms along the path connecting two atoms in molecular structure. In this study, to calculate AI descriptors, we used following attributes: 2 (double bonded atom), 3 (triple bonded atom), C, N, O, S, H, F, Cl and Br the distance range of 0 to 7. Description of all the AI descriptors used in gene-ration of QSAR models are given in Table **4**. Positive contribution of T_O_Cl_6 reveals the importance of presence of oxygen atom separated by six bond distance from chlorine atom. The positive impact is augmented by the presence of carbonyl oxygen and chlorine atom in biphenyl ring system.

Most hydrophobic hydrophilic distance (MHHD) signifies distance between most hydrophobic and hydrophilic point on van der Walls (vdW) surface. MHHD (SlogpK) is a

Table 4. Molecular Descriptors Used in QSAR Study

Descriptor	Description
Atomic valence connectivity index (chiv)	
Sub class (chiV1)	ChiV1 signifies atomic valence connectivity index
Estate number	
SaasCcount	SaasCcount defines the total number of carbon connected with one single bond along with two aromatic bonds
Distance based topological	
DistTopo	This descriptor signifies distance based topological index
Hydrophobicity SlogpK	
Most hydrophobic hydrophilic distance	This descriptor signifies distance between most hydrophobic and hydrophilic point on the vdW surface
Alignment independent descriptors	
T_C_O_7	T_C_O_7 is a count of number of carbon atoms (single, double or triple bonded) separated from any oxygen atom (single or double bonded) by six bond distance, e.g., C_C_C_C_C_C_C_O
T_C_N_3	T_C_N_3 is a count of number of carbon atoms (single, double or triple bonded) separated from any nitrogen atom (single or double bonded) by three bond distance, e.g., C_C_C_N
T_O_Cl_6	T_O_Cl_6 is a count of number of oxygen atoms (single, double or triple bonded) separated from any chlorine atom (single or double bonded) by six bond distance e.g., O_C_C_C_C_C_C_Cl
T_N_Cl_3	T_N_Cl_3 is a count of number of nitrogen atoms (single, double or triple bonded) separated from any chlorine atom (single or double bonded) by three bond distance e.g., N_C_C_C_Cl

thermodynamic descriptor to characterize the hydrobicity/hydrophilicity of a molecule. Logp (octanol/water partition coefficient) is an important property in describing the affinity of compounds in terms of their partitioning biological membranes. SlogP estimates logP by summing the contribution of atom-weighted solvent accessible surface areas (SASA) and correction factors. Negative correlation of MHHD suggested that compound with less protonation state can increases *h*DHODH inhibitory activity.

Aryl Carboxylic Acid Amide-*h*DHODH Interaction (Hypothetical Model)

Aryl carboxylic acid amide derivatives act by binding to ubiquinone (CoQ) binding site of *h*DHODH. DHODH catalyses oxidation/reduction reaction in pyrimidine nucleotide biosynthesis pathway: oxidation of DHO to ORO and reduction of FMN to $FMNH_2$ (Fig. 1), co-substrate electron acceptor used by *h*DHODH is ubiquinone (CoQ). The ability of one or more functional group of these compounds to interact with one or more subsites of CoQ binding site of *h*DHODH serve as potential inhibitors. Based on QSAR study discussed above some of the important pharmacophoric groups are identified. The carboxyl group at pentacyclic ring system can form ionic interactions with binding site. Biphenyl or biaryl ring system can fits to hydrophobic cavity of binding pocket. Substituted biphenyl (biaryl tail) ring substituted with chloro group in second aromatic ring lead toward better inhibitory activity. At amide linker, electronegativity of oxygen and the presence of N-H dipoles allow amides to form H-bond with the binding site of enzyme. Cyclopentene and other pentacyclic aromatic

heterocycles can occupy hydrophobic binding pocket. Based on these observations, a proposed hypothetical model to explain inhibitor-enzyme interactions is depicted in Fig. (**5**).
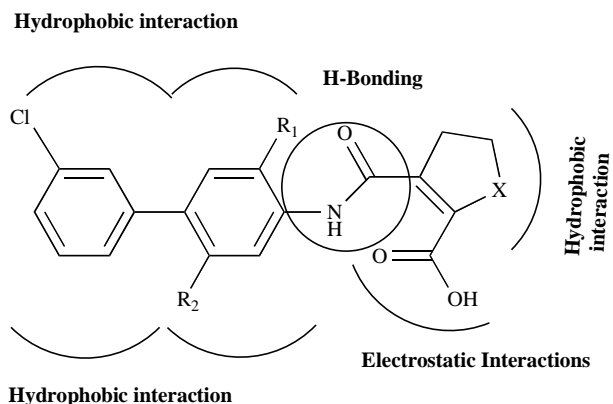
Fig. (5). Proposed hypothetical model of aryl carboxylic acid amide bound to ubiquinone (CoQ) binding site of *h*DHODH.

CONCLUSION

In continuation of prediction of *h*DHODH inhibitory activity, mathematical characterization of physicochemical properties and correlation with the *h*DHODH inhibitory activity has led to the significant QSAR models. The overall degree of prediction was found to be around 84% in case of MLR and PLS. Among the three QSAR models (MLR, PCR and PLS), results of PLS analysis showed significant predictive power and reliability as compare to other two methods. QSAR study on aryl carboxylic acid amide

derivatives quantified necessary structural and physico-chemical requirements for inhibition of *h*DHODH. QSAR study revealed the importance of chiV1, SaasCcount and DistTopo descriptors (present in every models). SaasCcount suggested the importance of carbon atoms in aryl ring system; DistTopo signifies basic connectivity of atoms in the molecules. Results of this QSAR study provide us a new approach to design *h*DHODH inhibitors and highlighted *h*DHODH as a target for chemotherapeutics.

CONFLICTS OF INTEREST

The author(s) confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1] Liu, S.; Neidhardt, E.A.; Grossman, T.H.; Ocain, T.; Clardy, J. Structures of human dihydroorotate dehydrogenase in complex with antiproliferative agents. *Protein Struct. Anal. Folding Des.*, **2000**, *8*, 25-33.
[2] Vyas, V.K.; Ghate, M. Recent developments in the medicinal chemistry and therapeutic potential of dihydroorotate dehydrogenase (DHODH) Inhibitors. *Mini-Rev. Med. Chem.,* **2011**, *11*, 1039-1055.
[3] Loeffler, M.; Zameitat, E. Pyrimidine biosynthesis. *Encycl. Biol. Chem.,* **2004**, *3*, 600-605.
[4] Evans, D.R.; Guy, H.I. Mammalian pyrimidine biosynthesis: fresh insights into an ancient pathway. *J. Biol. Chem.,* **2004**, *279*, 33035-33038.
[5] Fairbanks, L.D.; Bofill, M.; Ruckemann, K.; Simmonds, H.A. Importance of ribonucleotide availability to proliferating T-lymphocytes from healthy humans. *J. Biol. Chem.,* **1995**, *270*, 29682-29689.
[6] Cutolo, M.; Sulli, A.; Ghiorzo, P.; Pizzorni, C.; Craviotto, C.; Villaggio, B. Anti-inflammatory effects of leflunomide on cultured synovial macrophages from patients with rheumatoid arthritis. *Ann. Rheum. Dis.*, **2003**, *62*, 297–302.
[7] Shawver, L.K.; Schwartz, D.P.; Mann, E.; Chen, H.; Tsai, J.; Chu, L.; Taylorson, L.; Longhi, M.; Meredith, S.; Germain, L.; Jacobs, J.S; Tang, C.; Ullrich, A.; Berens, M.E.; Hersh, E.; McMahon, G.; Hirth K.P.; Powell, T.J. Inhibition of platelet-derived growth factor-mediated signal transduction and tumor growth by N-[4-(trifluoromethyl)-phenyl]-5-methylisoxazole- 4-carboxamide, *Clin. Cancer Res.*, **1997**, *3*, 1167-1177.
[8] Baumann, P.; Mandl, W.S.; Volkl, A.; Adam, C.; Bumeder, I.; Oduncu, F.; Schmidmaier, R. Dihydroorotate dehydrogenase inhibitor A771726 (leflunomide) induces apoptosis and diminishes proliferation of multiple myeloma cells. *Mol. Cancer Ther.*, **2009**, *8*, 366-375.
[9] Chen, S.F.; Ruben, R.L.; Dexter, D.L. Mechanism of action of the novel anticancer agent 6-fluoro-2-(2'-fluoro-1,1'-biphenyl-4-yl)-3-methyl-4-quinolinecarboxylic acid sodium salt (NSC 368390): inhibition of *de novo* pyrimidine nucleotide biosynthesis. *Cancer Res.*, **1986** *46*, 5014-5019.
[10] Herrmann, M.L.; Schleyerbach, R.; Kirschbaum, B.J. Leflunomide: an immunomodulatory drug for the treatment of rheumatoid arthritis and other autoimmune diseases. *Immunopharmacology*, **2004**, *47*, 273-289.
[11] Merrill, J.E.; Hanak, S.; Pu, S.F.; Liang, J.; Dang, C.; Iglesias, B.D.; Harvey, B.; Zhu, B.; McMonagle, S.K. Teriflunomide reduces behavioral, electrophysiological, and histopathological deficits in the Dark Agouti rat model of experimental autoimmune encephalomyelitis. *J. Neurol.*, **2009**, *256*, 89-103.
[12] Copeland, R. A.; Marcinkeviciene, J.; Haque, T. S.; Kopcho, L. M.; Jiang, W.; Wang, K.; Ecret, L. D.; Sizemore, C.; Amsler, K. A.; Foster, L.; Tadesse, S.; Combs, A. P.; Stern, A. M.; Trainor, G. L.;

Slee, A.; Rogers, M. J.; Hobbs, F. *Helicobacter pylori*-selective antibacterial based on inhibition of pyrimidine biosynthesis. *J. Biol. Chem.,* **2000**, *275*, 33373-33378.
[13] McRobert, L.; McConkey, G.A. RNA interference (RNAi) inhibits growth of Plasmodium falciparum. *Mol. Biochem. Parasitol.,* **2002**, *119*, 273–278.
[14] Fox, R. I.; Herrmann, M. L.; Frangou, C. G.; Wahl, G. M.; Morris, R. E.; Strand, V.; Kirschbaum, B. J. Mechanism of action for leflunomide in rheumatoid arthritis. *Clin. Immunol.,* **1999**, *93*, 198-208.
[15] Rozman, B. Clinical experience with leflunomide in rheumatoid arthritis. *J. Rheumatol Suppl.,* **1998**, *53*, 27-31.
[16] Nara, T.; Hshimoto, T.; Aoki, T. Evolutionary implications of the mosaic pyrimidine-biosynthetic pathway in eukaryotes. *Gene,* **2000**, *257*, 209-222.
[17] Bjornberg, O.; Rowland, P.; Larsen, S.; Jensen, K.F. Active site of dihydroorotate dehydrogenase A from *Lactococcus lactis* investigated by chemical modification and mutagenesis. *Biochemistry,* **1997**, *36*, 16197–1620.
[18] Bjornberg, O.; Gruner, A.C.; Roepstorff, P.; Jensen, K.F. The activity of *Escherichia coli* dihydroorotate dehydrogenase is dependent on a conserved loop identified by sequence homology, mutagenesis, and limited proteolysis. *Biochemistry,* **1999**, *38*, 2899–2908.
[19] Vyas, V.K.; Jain, A.; Ghate, M.; Maliwal, D. QSAR modeling of some substituted alkylidenepyridazin-3-one as a non-cAMP-based antiplatelet agent. *Med. Chem. Res.,* **2011**, *20*, 355–363.
[20] Kaushik, S.; Gupta, S.P.; Sharma, P. K.; Anwar, Z. A QSAR study on some series of HIV-1 integrase inhibitors. *Med. Chem.*, **2011**, *7(6)*, 553-560.
[21] Santos, R.N.; Guido R.V.C.; Oliva, G.; Dias L.C.; Andricopulo A.D. Quantitative structure-activity studies on a series of migrastatin analogs as inhibitors of cancer cell metastasis. *Med. Chem.,* **2011**, *7(3)*, 155-164.
[22] Vyas, V.K.; Joshi, G.; Namdeo B.; Gupta, A. Exploring structure indenture of some aminopyrazolopyridine ureas as potent VEGFR/PDGFR multitargeted kinase inhibitors: A QSAR approach. *Ind. J. Chem.,* **2011**, *50B*, 858-867.
[23] Goodarzi, M.; Freitas, M.P. MIA-QSAR coupled to different regression methods for the modeling of antimalarial activities of 2-aziridinyl and 2,3-bis-(aziridinyl)-1,4-naphtoquinonyl sulfate and acylate derivatives. *Med. Chem.*, **2011**, *7(6)*, 645-654.
[24] Noolvi, M.N.; Patel, H.M.; Bhardwaj, V. A comparative QSAR analysis of quinazoline analogues as tyrosine kinase (erbB-2) inhibitors. *Med. Chem.*, **2011**, *7(3)*, 200-212.
[25] Leban, J.; Saeb, W.; Garcia, G.; Baumgartner R.; Kramer, B. Discovery of a novel series of DHODH inhibitors by a docking procedure and QSAR refinement. *Bioorg. Med. Chem. Lett.*, **2004**, *14*, 55–58.
[26] Leban, J.; Kralik, M.; Mies, J.; Gassen, M.; Tentschert K.; Baumgartner R. SAR, species specificity, and cellular activity of cyclopentene dicarboxylic acid amides as DHODH inhibitors. *Bioorg. Med. Chem. Lett.*, **2005**, *15*, 4854–4857.
[27] Leban, J.; Kralik, M.; Mies, J.; Baumgartner, R.; Gassen M.; Tasler S. Biphenyl-4-ylcarbamoyl thiophene carboxylic acids as potent DHODH inhibitors. *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 267–270.
[28] Vyas, V.K.; Ghate, M. 2D and 3D QSAR study on amino nicotinic acid and isonicotinic acid derivatives as potential inhibitors of dihydroorotate dehydrogenase (DHODH). *Med. Chem. Res.,* DOI 10.1007/s00044-011-9837-4.
[29] Hudson, B.D.; Hyde, R.M.; Ra E.; Wood, J. Parameter based methods for compounds selection from chemical databases. *Quant. Struct. Act. Relat.,* **1996**, *15*, 285-289.
[30] Kubyani, H. Variable Selection in QSAR Studies. I. An evolutionary algorithm. *Quant. Struct. Act. Relat.,* **1994**, *13*, 285-294.
[31] Croux C.; Joossens, K. Influence of observations on the misclassification probability in quadratic discriminant analysis. *J. Multivar. Anal.,* **2005**, *96*, 348-403.
[32] Huberty, C.J. Applied Discriminant Analysis, Willey and Sons, New York, **1994**.
[33] J. Devillers, Neural Networks in QSAR and Drug Design, Academic Press, London, **1996**.
[34] Doucet, J.P.; Barbault F.; Xia, H.; Panaye, A. Fan, B. Nonlinear SVM approaches to QSPR/QSAR studies and drug design. *Curr. Comput.-Aided Drug Des.,* **2007**, *3*, 263-289.

[35]　Kier, L.B.; Hall, L.H. A differential molecular connectivity index. *Quant. Struct. Act. Relat.,* **1991**, *10*, 34.

[36]　Randic, M. Characterization of molecular branching. *J. Am. Chem. Soc.,* **1975**, *97*, 6609- 6615.

[37]　Balaban, A.T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399-404.