International Journal of Research and Reviews
in Applied Sciences and Engineering (IJRRASE)
Vol 8. No.2 Year 2016 Pp.371-393

available at : www.ijcns.com
ISSN: 2231-0061

QSAR Study on HIF – PHD2 Inhibitors by Monte Carlo Method

S.Sugunakala
Assistant Professor and Head, Department of Bioinformatics
A.V.C.College (Autonomous)
Mannampandal, Mayiladuthurai, Tamilnadu, India
E mail: suguna_murali@hotmail.com

Abstract

The main aim of the present work is to study the quantitative structure activity relationship (QSAR) of structural indicators of a series of 71 pyridine analogues and their inhibitory activity on Hypoxia inducible factor Prolyl Hydroxylase Domain II (HIF – PHD2) usingCORAL (http://www.insilico.eu/coral) software. To develop the QSAR models, the molecular structure by SMILES [simplified molecular input line entry system] and molecular graph by HSG [Hydrogen suppressed Graph]were used. The datasets were divided into sub-training set, the calibration set, and the test set. The quality of predicted models were examined by splitting the dataset into two types and the models were validated by validation set in which the dataset compounds are not included in the model building steps. The statistically significant predicted model can be used to define the structural characteristics responsible for the increased or decreased inhibition activity on HIF – PHD2.

Keywords: QSAR, Inhibitory activity on HIF – PHD2, Monte Carlo method, CORAL software

---

1. Introduction

Hypoxia inducible factor (HIF) plays an important role in hypoxic condition by regulating the various essential genes (Semenza, 2004). It consists of two subunits namely HIF - 1α and 1 β. Regulation of various processes namely erythropoiesis, energy utilization, angiogenesis, cellular proliferation and apoptosis are carried out with HIF - 1α subunit (Brahimi-Horn and Pouysségur, 2009; Chiche, et al., 2010) and the functional level mainly depends on the oxygen concentration i.e. in normoxia, the expression level of HIF - 1α was suppressed where as in hypoxia their expression is increased (Bruick, 2003).

The level of HIF - 1α was regulated by hydroxylation process at 4^{th} position of Pro 402 and Pro 564 which is catalyzed by prolyl hydroxylase domains PHD1, 2 & 3 which lead to their degradation (Appelhoff, et al., 2004). Especially, PHD2 domain is considered to be a key regulator in the hydroxylation process based on oxygen availability (Metzen, et al., 2005).

The up regulation of HIF - 1 α activity is made possible by preventing the hydroxylation through the use of PHD2 inhibitors (Linden, et al., 2003; Tsukiyama, et al., 2006; Nangaku, et al., 2006; Lee, et al., 2010; Rosen, et al., 2010; Rabinowitz, et al., 2010) .

Hence, inhibition of PHD2 domain can be considered as an important therapeutic drug target to treat various hypoxia related diseases namely anemia, myocardial infarction etc., Many PHD2 inhibitors are available in the market. But developing new and effective HIF – PHD2 inhibitors still remains a problem.

To solve this problem, QSAR models can be developed and used (Ojha and Roy, 2011; Ibezim, et al., 2012). The fundamental basis of QSAR analysis involves calculation of molecular descriptors by means of various methods like multiple linear regression (MLR), artificial neural networks, principle component analysis (PCA) etc., (Rasulev, 2008; Furtula, B. and Gutman 2011; Došlic et al., 2011; Afantitis, 2012).

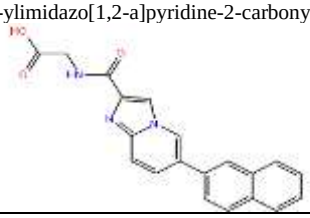
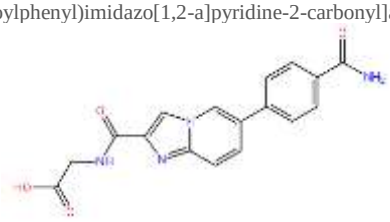
In this study, CORAL software was used to calculate molecular descriptors and to build QSAR models by Monte Carlo method (Toropova, et al., 2011) and to define the molecular structure responsible for HIF – PHD2 inhibition activity.

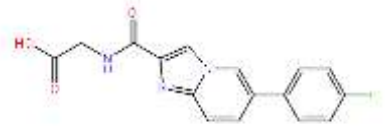
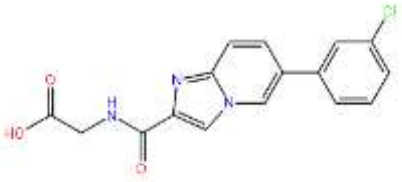
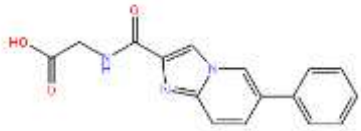
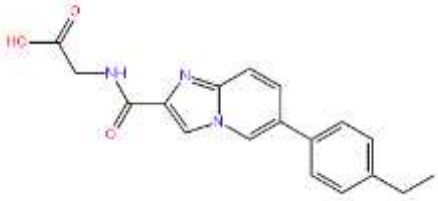
2. Method

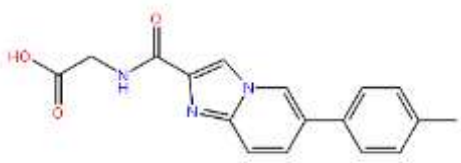
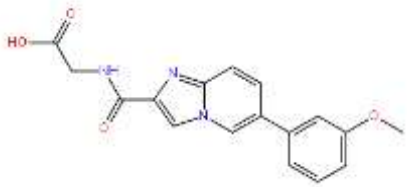
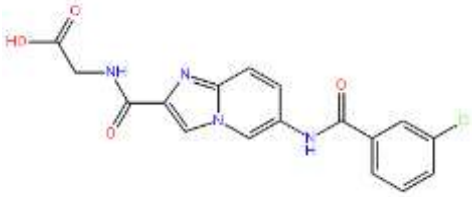
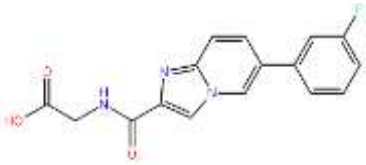
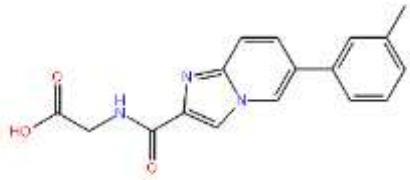
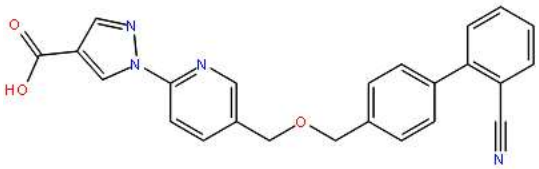
The data set

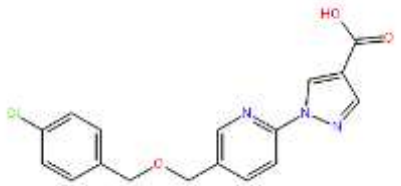
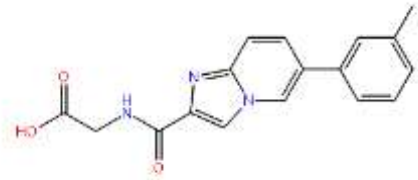
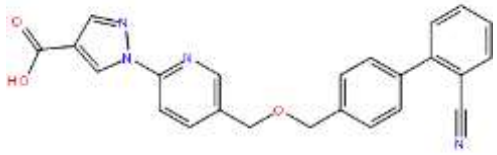
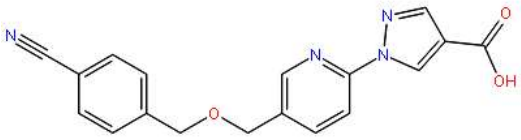
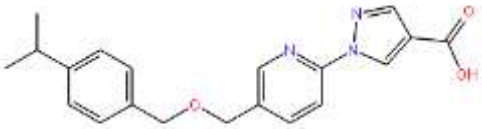
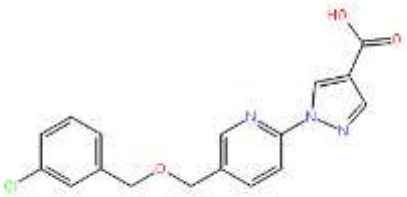
A series of 71 pyridine analogs with HIF –PHD2 inhibition activity (Warshakoon, et al., 2006) has been taken from the Binding database (Table 1). The logarithm of the inhibitory concentration [PIC50 = - log (IC50) nM] has been used as the endpoint. The QSAR models were built up for two random splits.

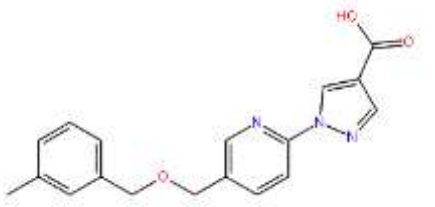
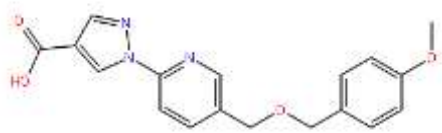
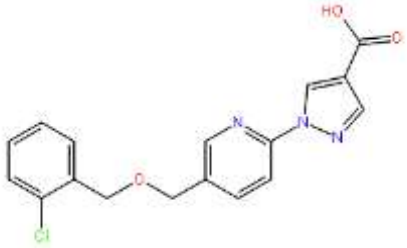
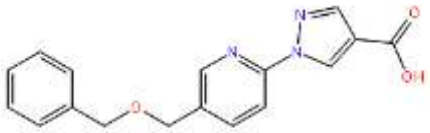
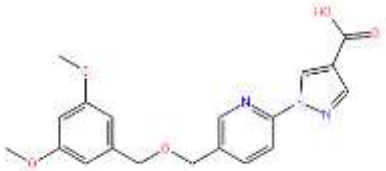
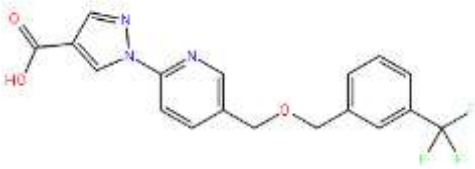
Table 1 – Pubchem CID, IUPAC name, chemical structure, IC_{50}, PIC_{50}

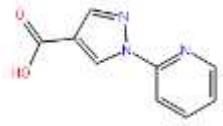
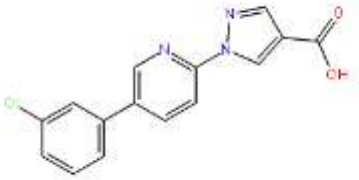
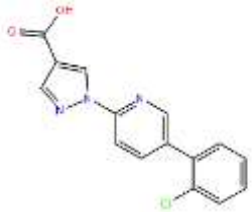
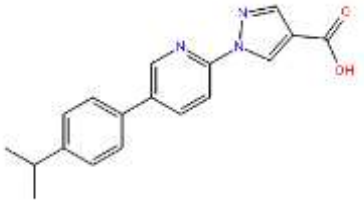
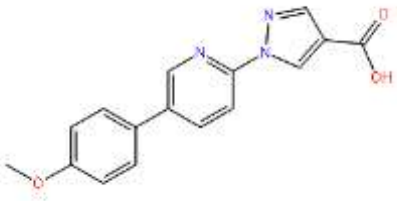
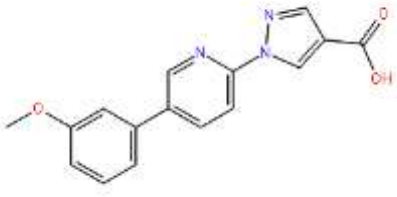
S.NO	Pubchem CID	Structure	Experimental IC_{50} (nM)	Predicted IC_{50} (PIC_{50}) (Mole)
1.	16126774	2-[(6-naphthalen-2-ylimidazo[1,2-a]pyridine-2-carbonyl)amino]acetic acid 	3600	5.44
2.	44419281	2-[[6-(4-carbamoylphenyl)imidazo[1,2-a]pyridine-2-carbonyl]amino]acetic acid 	5900	5.22
3.	44419265	2-[[6-(4-chlorophenyl)imidazo[1,2-a]pyridine-2-carbonyl]amino]acetic acid	6800	5.16

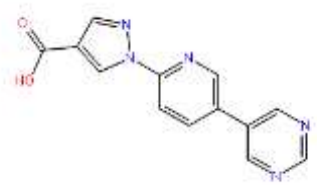
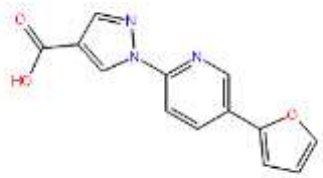
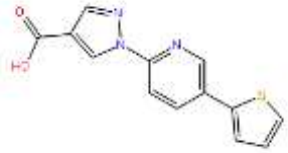
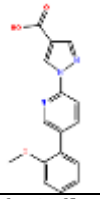
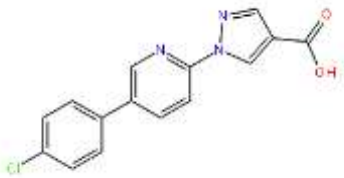
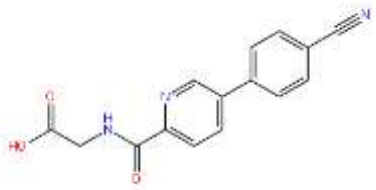
				
4.	44419280	2-[[6-(4-methoxyphenyl)imidazo[1,2-a]pyridine-2-carbonyl]amino]acetic acid 	8100	5.09
5.	44419336	2-[[6-(3-chlorophenyl)imidazo[1,2-a]pyridine-2-carbonyl]amino]acetic acid 	12000	4.92
6.	44419282	2-[[6-(4-fluorophenyl)imidazo[1,2-a]pyridine-2-carbonyl]amino]acetic acid 	14000	4.85
7.	44419264	2-[(6-phenylimidazo[1,2-a]pyridine-2-carbonyl)amino]acetic acid 	14200	4.84
8.	44419270	2-[[6-(4-ethylphenyl)imidazo[1,2-a]pyridine-2-carbonyl]amino]acetic acid 	16000	4.79
9.	44419252	2-[[6-(4-methylphenyl)imidazo[1,2-a]pyridine-2-carbonyl]amino]acetic acid	17000	4.76

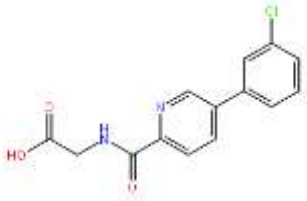
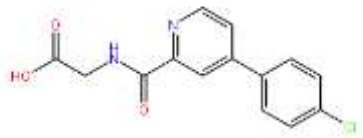
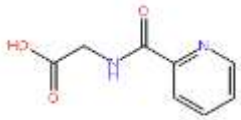
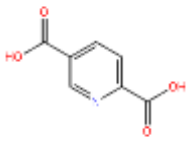
				
10.	44419297	2-[[6-(3-methoxyphenyl)imidazo[1,2-a]pyridine-2-carbonyl]amino]acetic acid 	19000	4.72
11.	44419302	2-[[6-[(3-chlorobenzoyl)amino]imidazo[1,2-a]pyridine-2-carbonyl]amino]acetic acid 	24000	4.61
12.	44419342	2-[[6-(3-fluorophenyl)imidazo[1,2-a]pyridine-2-carbonyl]amino]acetic acid 	24000	4.61
13.	44419346	2-[[6-(3-methylphenyl)imidazo[1,2-a]pyridine-2-carbonyl]amino]acetic acid 	100000	4.00
14.	16126655	1-[5-[[4-(2-cyanophenyl)phenyl]methoxymethyl]pyridin-2-yl]pyrazole-4-  carboxylic acid	190	6.72
15.	44419318	1-[5-[(4-chlorophenyl)methoxymethyl]pyridin-2-yl]pyrazole-4-carboxylic acid	580	6.23

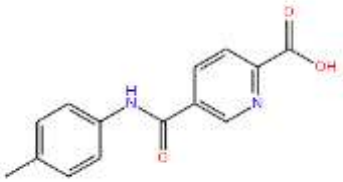
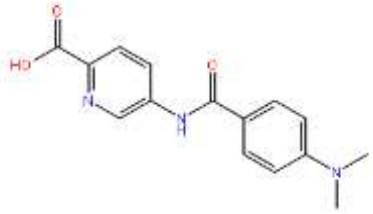
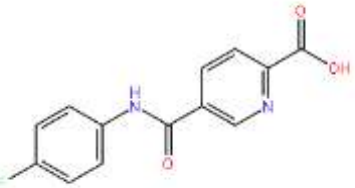
				
16.	6914666	2-[(1-chloro-4-hydroxyisoquinoline-3-carbonyl)amino]acetic acid 	1400	5.85
17.	44419325	1-[5-[(3-methoxyphenyl)methoxymethyl]pyridin-2-yl]pyrazole-4-carboxylic  acid	1600	5.79
18.	44419326	1-[5-[(4-cyanophenyl)methoxymethyl]pyridin-2-yl]pyrazole-4-carboxylic acid 	1600	5.79
19.	44419333	1-[5-[(4-propan-2-ylphenyl)methoxymethyl]pyridin-2-yl]pyrazole-4-carboxylic  acid	1700	5.76
20.	44419319	1-[5-[(3-chlorophenyl)methoxymethyl]pyridin-2-yl]pyrazole-4-carboxylic acid 	1700	5.76
21.	44419329	1-[5-[(3-methylphenyl)methoxymethyl]pyridin-2-yl]pyrazole-4-carboxylic acid	2100	5.67

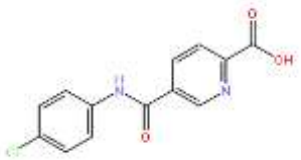
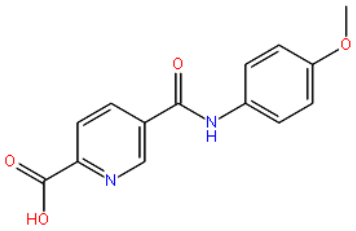
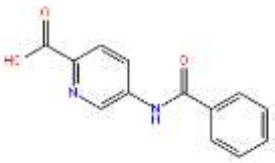
				
22.	44419321	1-[5-[(4-methoxyphenyl)methoxymethyl]pyridin-2-yl]pyrazole-4-carboxylic  acid	2400	5.61
23.	44419320	1-[5-[(2-chlorophenyl)methoxymethyl]pyridin-2-yl]pyrazole-4-carboxylic acid 	2400	5.61
24.	44419317	1-[5-(phenylmethoxymethyl)pyridin-2-yl]pyrazole-4-carboxylic acid 	2700	5.56
25.	44419334	1-[5-[(3,5-dimethoxyphenyl)methoxymethyl]pyridin-2-yl]pyrazole-4-  carboxylic acid	2800	5.55
26.	44419330	1-[5-[[3-(trifluoromethyl)phenyl]methoxymethyl]pyridin-2-yl]pyrazole-4-  carboxylic acid	2900	5.53
27.	8027170	1-pyridin-2-ylpyrazole-4-carboxylic acid	12000	4.92

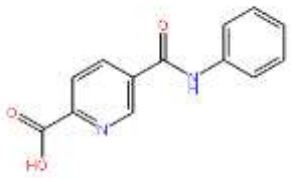
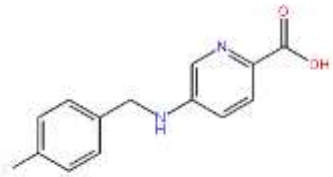
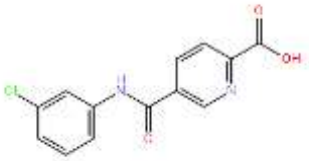
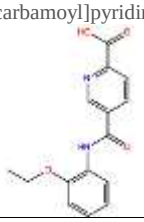
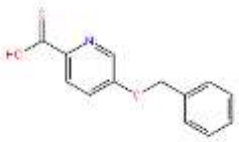
				
28.	44419139	1-[5-(3-chlorophenyl)pyridin-2-yl]pyrazole-4-carboxylic acid 	4900	5.30
29.	44419140	1-[5-(2-chlorophenyl)pyridin-2-yl]pyrazole-4-carboxylic acid 	6500	5.18
30.	44419165	1-[5-(4-propan-2-ylphenyl)pyridin-2-yl]pyrazole-4-carboxylic acid 	6800	5.16
31.	44419160	1-[5-(4-methoxyphenyl)pyridin-2-yl]pyrazole-4-carboxylic acid 	7300	5.13
32.	44419179	1-[5-(3-methoxyphenyl)pyridin-2-yl]pyrazole-4-carboxylic acid 	10000	5.00
33.	44419159	1-(5-pyrimidin-5-ylpyridin-2-yl)pyrazole-4-carboxylic acid	15000	4.82

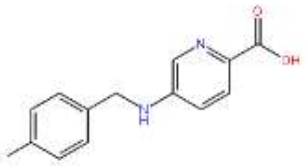
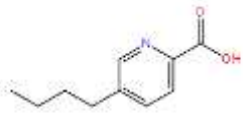
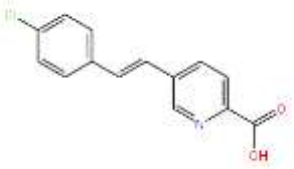
				
34.	44419147	1-[5-(furan-2-yl)pyridin-2-yl]pyrazole-4-carboxylic acid 	15000	4.82
35.	44419131	2-[[5-(4-chlorophenyl)pyridine-2-carbonyl]amino]acetic acid 	15000	4.82
36.	44419154	1-(5-thiophen-2-ylpyridin-2-yl)pyrazole-4-carboxylic acid 	18000	4.74
37.	44419171	1-[5-(2-methoxyphenyl)pyridin-2-yl]pyrazole-4-carboxylic acid 	19000	4.72
38.	44419133	1-[5-(4-chlorophenyl)pyridin-2-yl]pyrazole-4-carboxylic acid 	20000	4.69
39.	44419124	2-[[5-(4-cyanophenyl)pyridine-2-carbonyl]amino]acetic acid 	39000	4.40

40.	44419123	2-[[5-(4-methylphenyl)pyridine-2-carbonyl]amino]acetic acid 	46000	4.33
41.	44419132	2-[[5-(3-chlorophenyl)pyridine-2-carbonyl]amino]acetic acid 	100000	4.00
42.	24891150	2-[[4-(4-chlorophenyl)pyridine-2-carbonyl]amino]acetic acid 	100000	4.00
43.	11788622	2-(pyridine-2-carbonylamino)acetic acid 	100000	4.00
44.	15121190	5-[[4-(dimethylamino)phenyl]carbamoyl]pyridine-2-carboxylic acid 	4100	5.38
45.	15121179	5-[(4-ethoxyphenyl)carbamoyl]pyridine-2-carboxylic acid 	5000	5.30
46.	7493	 pyridine-2,5-dicarboxylic acid	5500	5.25

47.	15121188	5-[(4-methylphenyl)carbamoyl]pyridine-2-carboxylic acid 	8300	5.08
48.	15121186	5-[(2-fluorophenyl)carbamoyl]pyridine-2-carboxylic acid 	11000	4.95
49.	15121200	5-[(4-ethoxybenzoyl)amino]pyridine-2-carboxylic acid 	11200	4.95
50.	15121202	5-[[4-(dimethylamino)benzoyl]amino]pyridine-2-carboxylic acid 	11300	4.94
51.	15121191	5-[(2-methoxyphenyl)carbamoyl]pyridine-2-carboxylic acid 	14800	4.82
52.	15121187	5-[(4-fluorophenyl)carbamoyl]pyridine-2-carboxylic acid 	16100	4.79
53.	15121182	5-[(2-chlorophenyl)carbamoyl]pyridine-2-carboxylic acid	16300	4.78

				
54.	15121198	5-[(4-methylbenzoyl)amino]pyridine-2-carboxylic acid 	16500	4.78
55.	15121194	5-[(2,4-dimethoxyphenyl)carbamoyl]pyridine-2-carboxylic acid 	16600	4.77
56.	15121185	5-[(4-chlorophenyl)carbamoyl]pyridine-2-carboxylic acid 	17400	4.75
57.	15121193	5-[(4-methoxyphenyl)carbamoyl]pyridine-2-carboxylic acid 	21100	4.67
58.	15121199	5-[(4-chlorobenzoyl)amino]pyridine-2-carboxylic acid 	22500	4.64
59.	12640225	5-benzamidopyridine-2-carboxylic acid 	25600	4.59

60.	15121195	5-[(3,4,5-trimethoxyphenyl)carbamoyl]pyridine-2-carboxylic acid 	26500	4.57
61.	15121181	5-(phenylcarbamoyl)pyridine-2-carboxylic acid 	26500	4.57
62.	12640191	5-[(4-fluorophenyl)methylamino]pyridine-2-carboxylic acid 	31000	4.50
63.	15121183	5-[(3-chlorophenyl)carbamoyl]pyridine-2-carboxylic acid 	32200	4.49
64.	15121189	5-[(2-ethoxyphenyl)carbamoyl]pyridine-2-carboxylic acid 	32900	4.48
65.	15121204	5-[(2-fluorobenzoyl)amino]pyridine-2-carboxylic acid 	33800	4.47
66.	18379352	5-phenylmethoxypyridine-2-carboxylic acid 	39000	4.40

67.	15121196	5-[(4-nitrophenyl)carbamoyl]pyridine-2-carboxylic acid 	44000	4.35
68.	15121201	5-[(4-fluorobenzoyl)amino]pyridine-2-carboxylic acid 	51200	4.29
69.	12640191	5-[(4-fluorophenyl)methylamino]pyridine-2-carboxylic acid 	80000	4.09
70.	3442	5-butylpyridine-2-carboxylic acid 	149000	3.82
71.	15121207	5-[(E)-2-(4-chlorophenyl)ethenyl]pyridine-2-carboxylic acid 	150000	3.82

2.2 Splitting

Development of statistically significant QSAR models were based on the placing of dataset compounds into training and test set compounds (Roy et al., 2011; Puzyn, et al., 2012). The following criteria were used while splitting the dataset compounds. (i) In each subset the uniform range of end point is maintained (ii) Random splits and (iii) the splits should not be identical. The percentage deviation in the splits is detailed in Table 2 which revealed that the splits are in enough random.

Table 2 Percentage of identity of splits 1–2

	Set	Split1%	Split2%
Split1	Sub training	100	33.8
	Calibration		4.44
	Test		50
	Validation		0
Split2	Sub training		100
	Calibration		

	Test		
	Validation		

*Identity% = $\frac{N_{ij}}{0.5 * (N_i+N_j)}$ X 100 (Toropov, et al., 2013)

Where,

Ni; j is the number of substances which are distributed into the same set for both i-th split and j-th split (set = sub-training, calibration, test, validation)

Ni is the number of substances which are distributed into the set for i-th split

Nj is the number of substances which are distributed into the set for j-th split

2.3 Optimal descriptors

If the molecular structure was represented by hybrid form i.e. a combination of SMILES and HSG (Hydrogen suppressed graph), statistically significant QSAR models were obtained. The hybrid optimal descriptors were used to build up model for the pIC_{50} of PHD2 inhibitors:

$$_{Hybrid}DCW\ (T; Nepoch) = {}_{SMILES}DCW\ (T; Nepoch) + {}_{HSG}DCW\ (T; Nepoch)$$

Where,

$^{SMILES}DCW\ (T, N_{epoch}) = \sum CW\ (SSSk)$

$^{HSG}DCW\ (T, N_{epoch}) = \sum CW\ (A_K) + \sum CW\ (^0EC_K) + \sum CW\ (^2EC_K)$ …… (2)

T is threshold for definition of rare (noise) molecular features: T = 1 and 2 were examined.

N_{epoch} is the number of the epochs of the Monte Carlo optimization:

N_{epoch} = 150 have been selected;

Where,

SSSk is three-components SMILES attributes,

Ak is chemical element, such as, C, N, O, etc., for HSG,

^{0}ECk is the Morgan extended connectivity of zero order (i.e., the vertex degree) in the HSG,

^2EC_K is the Morgan extended connectivity of second order (i.e., the vertex degree) in the HSG.

CW(x) is the correlation weights for x (some molecular feature which can be extracted from SMILES or from HSG). Table 3 contains an example of the CW(x) calculated with the Monte Carlo method. Table 4 contains an example of the calculation with the correlation weights.

Table 3 Correlation weights for calculation of Hybrid DCW (T; Nepoch) for the case of split 1 (Eq 4)

Structural attribute (SA)	CW(SA)	N_{TRN}	N_{CLB}	N_{TST}
ECO-C…2…:	-0.3468	33	25	10
EC0-C...2...:	-0.3468	33	25	10
EC0-C...2...:	-0.3468	33	25	10
EC0-C...3...:	-1.7853	33	25	10
EC0-C...2...:	-0.3468	33	25	10
EC0-C...3...:	-1.7853	33	25	10
EC0-C...2...:	-0.3468	33	25	10
EC0-C...2...:	-0.3468	33	25	10
EC0-C...3...:	-1.7853	33	25	10
EC0-C...2...:	-0.3468	33	25	10
EC0-C...3...:	-1.7853	33	25	10
EC0-C...2...:	-0.3468	33	25	10
EC0-N...3...:	-1.3057	15	14	8
EC0-C...2...:	-0.3468	33	25	10
EC0-C...3...:	-1.7853	33	25	10
EC0-N...2...:	-3.1436	33	25	10

EC0-C...3...:	-1.7853	33	25	10
EC0-C...2...:	-0.3468	33	25	10
EC0-C...2...:	-0.3468	33	25	10
EC0-C...3...:	-1.7853	33	25	10
EC0-O...1...:	1.5020	33	25	10
EC0-N...2...:	-3.1436	33	25	10
EC0-C...2...:	-0.3468	33	25	10
EC0-C...3...:	-1.7853	33	25	10
EC0-O...1...:	1.5020	33	25	10
EC0-O...1...:	1.5020	33	25	10
EC2-C...9...:	2.6074	17	14	6
EC2-C...9...:	2.6074	17	14	6
EC2-C...11..:	-1.0254	32	21	9
EC2-C...18..:	0.6983	11	6	6
EC2-C...14..:	0.4531	17	13	8
EC2-C...18..:	0.6983	11	6	6
EC2-C...12..:	-0.4728	32	25	10
EC2-C...12..:	-0.4728	32	25	10
EC2-C...17..:	-0.1240	18	17	6
EC2-C...11..:	-1.0254	32	21	9
EC2-C...18..:	0.6983	11	6	6
EC2-C...14..:	0.4531	17	13	8
EC2-N...19..:	1.8369	5	2	5
EC2-C...14..:	0.4531	17	13	8
EC2-C...18..:	0.6983	11	6	6
EC2-N...14..:	0.1925	5	3	5
EC2-C...18..:	0.6983	11	6	6
EC2-C...12..:	-0.4728	32	25	10
EC2-C...12..:	-0.4728	32	25	10
EC2-C...15..:	1.6991	29	20	10
EC2-O...6...:	0.5313	20	14	6
EC2-N...11..:	-0.1270	9	6	6
EC2-C...9...:	2.6074	17	14	6
EC2-C...11..:	-1.0254	32	21	9
EC2-O...4...:	0.8428	8	6	5
EC2-O...4...:	0.8428	8	6	5
C...1...=...:	3.5585	33	25	10
C...=...1...:	2.7832	33	25	10
C...C...=...:	0.1044	31	24	10
C...C...=...:	0.1044	31	24	10
C...=...C...:	-0.4267	32	25	10
=...C...2...:	0.1210	31	24	10
C...2...C...:	0.0	1	1	0
=...C...2...:	0.1210	31	24	10
C...=...C...:	-0.4267	32	25	10

=...C...(...:	-0.6044	31	24	10
C...(...C...:	-0.1797	28	22	6
=...C...(...:	-0.6044	31	24	10
C...=...C...:	-0.4267	32	25	10
C...C...=...:	0.1044	31	24	10
C...C...2...:	1.7773	6	4	3
C...2...=...:	1.7832	31	23	10
C...=...2...:	2.3175	31	23	10
=...C...1...:	9.6230	33	25	10
C...1...(...:	2.2540	23	16	6
C...(...1...:	2.1240	16	11	4
3...C...(...:	0.0	14	0	0
C...3...=...:	0.0	1	1	0
C...=...3...:	0.0	1	1	0
N...C...=...:	3.1260	32	24	10
C...N...4...:	0.0	1	0	0
N...4...C...:	0.0	1	1	0
=...C...4...:	0.0	1	1	0
C...=...C...:	-0.4267	32	25	10
=...C...(...:	-0.6044	31	24	10
N...(...C...:	-2.9816	7	5	6
=...N...(...:	-2.2782	5	3	5
N...=...C...:	0.0284	32	24	10
=...C...4...:	0.0	1	1	0
C...4...C...:	0.0	1	0	0
=...C...4...:	0.0	1	1	0
C...=...C...:	-0.4267	32	25	10
=...C...3...:	-2.4375	14	12	8
C...3...(...:	0.0	1	1	0
C...(...3...:	1.1699	10	9	3
(...C...(...:	-3.3790	33	25	10
C...(...=...:	-0.6845	33	25	10
O...=...(...:	2.1553	33	25	10
=...O...(...:	3.8477	33	25	10
O...(...N...:	1.1280	12	9	5
C...N...(...:	1.3770	18	11	6
N...C...C...:	2.0206	8	6	5
C...C...(...:	0.6054	18	16	9
C...(...=...:	-0.6845	33	25	10
O...=...(...:	2.1553	33	25	10
=...O...(...:	3.8477	33	25	10
O...(...O...:	-4.9960	33	25	10

* NTRN, NCLB, and NTST are the numbers of the x in the sub training, calibration, and test set

Table 4 - calculation for hybrid DCW (1, 9) = 18.53975

Molecular features	Correlation Weight, CW(x)
Features extracted from SMILES	
(...C...(...	-3.37900
(...O...(...:	0.84875
1...C...(...:	-0.62300
2...C...(...:	1.15425
2...C...1...:	0.73137
3...N...(...:	0.71775
=...C...(...:	-0.60438
=...C...1...:	9.62300
=...C...2...:	0.12100:

=...C...3...:	-2.43750
=...N...(...:	-2.27825
=...N...3...:	2.69250
=...O...(...:	3.84775
C...(...1...:	2.12400
C...(...2...:	1.94331
C...(...3...:	1.16988:
C...(...=...:	-0.68450:
C...(...C...:	-0.17969:
C...1...(...:	2.25400:
C...1...=...:	3.55850:
C...1...C...:	3.14644:

C...2...(...:	5.15325:
C...2...=...:	1.78325:
C...3...C...:	-1.32813:
C...=...(...:	0.99800:
C...=...1...:	2.78325:
C...=...2...:	2.31750:
C...=...C...:	-0.42669:
C...C...(...:	0.60538:
C...C...1...:	5.22375:
C...C...2...:	1.77725:
C...C...=...:	0.10438:
C...N...(...:	1.37700:

C...N...1...:	2.69050:
C...N...3...:	1.31250:
C...N...=...:	2.21575:
C...O...(...:	-1.31050:
C...O...C...:	0.65344:
Features extracted from HSG	
EC0-C...1...:	2.96975:
EC0-C...2...:	-0.34675:
EC0-C...3...:	-1.78525:
EC0-F...1...:	3.72275:
EC0-Cl..1...:	0.24419:
EC0-N...2...:	-3.14363:
EC0-N...3...:	-1.30569
EC0-O...1...:	1.50200:
EC0-O...2...:	0.40425:
EC2-C...10..:	-2.29787:
EC2-C...11..:	-1.02544:
EC2-C...12..:	-0.47275:
EC2-C...13..:	0.24700:
EC2-C...14..:	0.45313:
EC2-C...15..:	1.69913:
EC2-C...16..:	-1.93250:

EC2-C...17..:	-0.12400:
EC2-C...18..:	0.69831:
EC2-C...3...:	7.99900:
EC2-C...4...:	1.53425:
EC2-C...5...:	1.62500
EC2-C...7...:	-4.12600:
EC2-C...8...:	2.72275:
EC2-C...9...:	2.60738:
EC2-F...5...:	3.43750:
EC2-Cl..5...:	3.17488:
EC2-N...11..:	-0.12700:
EC2-N...12..:	2.08975:
EC2-N...13..:	-1.00400:
EC2-N...14..:	0.19250:
EC2-N...18..:	-1.00300:
EC2-N...19..:	1.83694:
EC2-O...10..:	-2.08494:
EC2-O...4...:	0.84275:
EC2-O...5...:	0.70894:
EC2-O...6...:	0.53125:
EC2-O...8...:	2.55850:
EC2-O...9...:	-2.14463:

N...(...1...:	-1.00000:
N...(...2...:	1.68550:
N...(...C...:	-2.98156:
N...1...C...:	1.53025:
N...3...(...:	1.28025:
N...3...C...:	1.68950:
N...=...C...:	0.02844:
N...C...(...:	2.03625:
N...C...2...:	2.94150:
N...C...=...:	3.12600:
N...C...C...:	2.02062:
O...(...C...:	-2.52725:
O...(...F...:	1.25300:
O...(...Cl..:	2.71294:
O...(...N...:	1.12800:
O...(...O...:	-4.99600:
O...=...(...:	2.15525:
O...C...(...:	0.12600:
O...C...1...:	-0.02825:
O...C...C...:	-3.12100:

Threshold=2;
Number of SMILES Attributes (SA) =160;
Number of active SA=96
In CORAL software, the QSAR models were built up by calculating the CW(x) which provides the maximum of target function (TF) (Toropov, et al., 2008; 2011).
TF = R + R' - |R _ R' | X Const
where R and R' are correlation coefficients between pIC_{50} and Hybrid DCW(T; Nepoch); Const is an empirical constant (In this case it is 0.1). Several runs of the Monte Carlo optimization was carried out to define preferable threshold and Nepoch with best quality of QSAR model followed by validation of the model performed with external validation dataset.

3. Results and Discussion

The best one variable models statistical quality are given in Table 5 and are represented by the following formula

$$pIC_{50} = C0 + C1 \times {}^{Hybrid}DCW(T, N_{epoch}) \ldots\ldots (3)$$

Table 5 Statistical characteristics of one-variable models for HIF – PHD2 inhibition activity (PIC_{50}) with the threshold T = 0, 1 & 2and the number of epochs of the Monte Carlo optimization Nepoch = 150

Table 5 Statistical characteristics of one variable model

Split-1	Run	T	Nact	Sub training						Calibration						Test					
				N	r2	q2	s	Mae	F	N	r2	q2	s	Mae	F	N	r2	q2	s	Mae	F
	1	0	160	33	0.9822	0.9804	0.071	0.042	1707	25	0.9370	0.9281	0.423	0.334	342	10	0.6512	0.4226	0.423	0.334	15
	2	0		33	0.9820	0.9803	0.071	0.042	1694	25	0.9372	0.9283	0.403	0.292	343	10	0.6287	0.4015	0.424	0.305	14
	3	0		33	0.9821	0.9803	0.071	0.043	1705	25	0.9370	0.9281	0.396	0.287	342	10	0.6446	0.4229	0.474	0.359	15
	1	1	127	33	0.9820	0.9802	0.071	0.042	1688	25	0.9373	0.9284	0.410	0.298	344	10	0.4880	0.1931	0.608	0.298	8

	2	1		33	0.9822	0.9804	0.071	0.042	1713	25	0.9370	0.9280	0.401	0.290	342	10	0.6331	0.4276	0.543	0.410	14
	3	1		33	0.9819	0.9801	0.070	0.042	1682	25	0.9373	0.9284	0.406	0.0295	344	10	0.4758	0.1932	0.636	0.453	7
	Average																				
	1	2	96	33	0.9831	0.9813	0.069	0.044	1798	25	0.9351	0.9256	0.468	0.349	331	10	0.7899	0.6806	0.313	0.249	30
	2	**2**		**33**	**0.9827**	**0.9809**	**0.070**	**0.045**	**1758**	**25**	**0.9354**	**0.9260**	**0.473**	**0.353**	**333**	**10**	**0.8185**	**0.7138**	**0.297**	**0.244**	**36**
	3	2		33	0.9828	0.9810	0.070	0.044	1767	25	0.9355	0.9261	0.469	0.350	334	10	0.7296	0.5856	0.432	0.321	22
Split-2																					
	1	0	155	34	0.9528	0.9485	0.134	0.083	646	24	0.9947	0.9938	0.338	0.255	4125	10	0.8008	0.6855	0.263	0.202	32
	2	0		34	0.9532	0.9490	0.134	0.081	652	24	0.9957	0.9949	0.336	0.251	5094	10	0.7394	0.4916	0.326	0.224	23
	3	0		34	0.9518	0.9474	0.136	0.083	632	24	0.9956	0.9948	0.343	0.255	4983	10	0.6938	0.5057	0.278	0.219	18
	1	1	137	34	0.9528	0.9486	0.134	0.080	646	24	0.9957	0.9949	0.343	0.254	5084	10	0.6139	0.2857	0.318	0.232	13
	2	1		34	0.9508	0.9464	0.137	0.082	619	24	0.9962	0.9955	0.326	0.242	5754	10	0.8223	0.7349	0.236	0.186	37
	3	**1**		**34**	**0.9537**	**0.9495**	**0.133**	**0.081**	**659**	**24**	**0.9954**	**0.9946**	**0.347**	**0.257**	**4744**	**10**	**0.8828**	**0.8314**	**0.187**	**0.139**	**60**
	Average																				
	1	2	104	34	0.9437	0.9382	0.147	0.098	536	24	0.9967	0.9960	0.333	0.248	6732	10	0.5245	0.2021	0.347	0.249	9
	2	2		34	0.9440	0.9386	0.146	0.098	540	24	0.9958	0.9951	0.335	0.246	5267	10	0.5153	0.1687	0.356	0.262	9
	3	2		34	0.9479	0.9429	0.141	0.095	582	24	0.9969	0.9963	0.337	0.249	7173	10	0.5629	0.2829	0.354	0.270	10

Best models are indicated by bold
* Nact = number of molecular features
n = number of substances in the set,
r^2 = square correlation coefficient between experimental and calculated values of PIC50,
q^2 = cross-validated squared correlation coefficient,
s = standard error of estimation and F = Fischer F-ratio.
In both the splits, best statistical quality was obtained at threshold equals to 2 & 1 and the average values of the N_{epoch} are 12 and 27 for split 1 & split 2 respectively. Further, the PIC_{50} models for Split1 and 2 are as follows

Split 1

PIC_{50} = 2.3687(± 0.0181) + 0.0621 (± 0.0006) * DCW (2, 8) …… (4)
n = 33, R^2 = 0.9827, q^2 = 0.9809, s = 0.070 (sub training set)
n = 25, R^2 = 0.9354, q^2 = 0.9260, s = 0.343(Calibration set)
n = 10, R^2 = 0.8185, q^2 = 0.7138, s = 0.297, R_m^2 = 0.7154(should be larger than 0.5) (Test set)

Split 2

PIC_{50} = 3.0005 (± 0.0057) + 0.0491 (± 0.0003) * DCW (1, 9) ……. (5)

n = 34, R^2 = 0.9537, q^2 = 0.9495, s = 0.133, F = **659** (sub training set)
n = 24, R^2 = 0.9954, q^2 = 0.9946, s = 0.347 (Calibration set)
n = 10, R^2 = 0.8828, q^2 = 0.8314, s = 0.187, R_m^2 = 0.6730(should be larger than 0.5) (Test set)
n = 10,

Where,
n= Number of compounds in the dataset,
r = correlation coefficient,
q^2 = leave-one-out cross validated correlation coefficient
F = Fischer F-ratio
R_m^2 = Predictive ability of the model
The robustness of the predicted models were further validated by using Y-scrambling (Mitra, et al., 2010) and the obtained values are given in Table 6

Table 6 Y-randomization results for test set

Probe of Y-scrambling	Split 1,Eq. 4 R_r^2	Split 2, Eq. 5, R_r^2
R^2	0.8185	0.8828
1	0.0781	0.0782
2	0.1119	0.0754
3	0.2572	0.0222
4	0.1146	0.1688
5	0.1919	0.1444
6	0.0918	0.0088
7	0.0093	0.994
8	0.0048	0.1214
9	0.0052	0.0012
10	0.0691	0.3721
Randaomized R $[{}^{C}R_p^2=R*\sqrt{(R^2-Rr^2)}]$	0.7421	0.7170

From the obtained correlation weights of Monte Carlo method of optimization in each split, it is possible to detect the structural features responsible for the increased or decreased inhibition activity of compounds against HIF - PHD2. Structural features having positive correlation values indicate that the feature may enhance the inhibition activity as well as increased end point. Similarly, the features having negative correlation values may represent the decreased inhibition activity. Based on this knowledge, we concluded that the presence of nitrogen is the promoter of the inhibition activity and vice versa and the presence of side chain atoms may decrease the inhibition activity (Table -7). The obtained predicted inhibitory values, experimental values and the residuals are also given in Table 8. The obtained results reveals that the predicted models can be effectively used to compute inhibitory activity of novel pyridine based molecule against HIF – PHD2 domains and may be used as a drug to treat various disorders of hypoxic conditions.

4. Conclusion

Statistical QSAR models for HIF – PHD2 domain inhibition activity for pyridine analogs were developed by using CORAL software (Table – 5). The developed models explain the structural feature promoters for the increased or decreased HIF – PHD2 domain inhibition activity (Table 7). These features may be effectively used to predict HIF – PHD2 inhibition

activity for a newly developed pyridine analogs (Table -8; Fig 1). The predicted model in two splits were validated with external dataset compounds also (Table 9)

Table 7 - Stability of the correlation weights for molecular structures extracted from SMILES

Molecular features, x	Correlation weights, CW(x) Run 1
C...1...=...:	5.3780
C...C...=...:	1.0136
C...2...C...:	0.0
=...C...(...:	0.7773
C...=...C...:	-0.3095
C...2...=...:	2.6290
C...1...(...:	2.7540
C...3...=...:	0.0
C...N...4...:	0.0
C...=...C...:	- 0.3095
=...C...(...:	0.7773
=...N...(...:	0.0575
N...=...C...:	1.0020
C...4...C...:	0.0
C...=...C...:	- 0.3095
=...C...3...:	0.5615
(...C...(...:	2.9980
C...(...=...:	0.4990
O...=...(...:	0.5000
=...O...(...:	1.7460
O...(...N...:	2.7480
C...N...(...:	1.1240
N...C...C...:	2.7540
C...C...(...:	1.9874
C...(...=...:	0.4990
O...=...(...:	0.5000
=...O...(...:	1.7460
O...(...O...:	- 3.0040

Table 8 – Observed and Residuals

Experimental PIC50, predicted PIC50

S. NO	Pubchem CID	Observed C_{50} (PIC_{50})	Predicted PIC50	Residual
1.	+ 16126774	5.44	5.554	0.114
2.	# 44419281	5.22	5.044	-0.176
3.	+ 44419265	5.16	4.938	-0.222
4.	# 44419336	4.92	4.737	-0.183
5.	+ 44419282	4.85	4.858	0.008
6.	# 44419264	4.84	4.433	-0.407
7.	- 44419270	4.79	4.44	-0.350
8.	+ 44419252	4.76	4.773	0.013
9.	# 44419297	4.72	4.597	-0.123
10.	# 44419302	4.61	4.75	0.140
11.	- 44419342	4.61	4.54	-0.070
12.	+ 44419346	4.00	4.259	0.259
13.	- 16126655	6.72	6.766	0.046
14.	# 44419318	6.23	5.686	-0.544
15.	- 6914666	5.85	5.953	0.103
16.	# 44419325	5.79	5.95	0.160
17.	- 4441932	5.79	5.866	0.076
	6			
18.	+ 44419333	5.76	5.672	-0.088
19.	- 44419319	5.76	5.978	0.218
20.	+ 44419329	5.67	5.499	-0.171
21.	+ 44419321	5.61	5.667	0.057
22.	# 44419320	5.61	5.695	0.085
23.	- 44419317	5.56	5.596	0.036
24.	+ 44419334	5.55	5.601	0.051
25.	- 44419330	5.53	5.64	0.110
26.	+ 44419139	4.92	4.54	-0.380
27.	- 44419140	5.30	5.126	-0.174
28.	- 44419165	5.18	5.191	0.011
29.	+ 44419160	5.16	5.02	-0.140
30.	+ 44419179	5.13	4.919	-0.211
31.	- 44419159	5.00	4.505	-0.495
32.	- 44419147	4.82	4.82	0.00
33.	- 44419131	4.82	4.266	-0.554
34.	+ 44419154	4.82	4.884	0.064

35.	+ 44419171	4.74	4.938	0.198
36.	- 44419133	4.72	5.16	0.440
37.	+ 44419124	4.69	4.821	0.131
38.	- 44419123	4.40	4.161	-0.239
39.	- 44419132	4.33	4.385	0.055
40.	+ 24891150	4.00	3.914	-0.086
41.	+ 11788622	4.00	3.913	-0.087
42.	- 15121190	4.00	3.828	-0.172
43.	+ 15121179	5.38	5.296	-0.084
44.	- 7493	5.30	4.889	-0.411
45.	+ 15121188	5.25	5.093	-0.157
46.	- 15121186	5.08	4.82	-0.260
47.	+ 15121200	4.95	4.943	-0.007
48.	+ 15121202	4.95	4.945	-0.005
49.	- 15121191	4.94	4.674	-0.266
50.	- 15121187	4.82	4.645	-0.175
51.	+ 15121182	4.79	4.812	0.022
52.	- 15121198	4.78	4.65	-0.130
53.	+ 15121194	4.78	4.734	-0.046
54.	+ 15121185	4.77	4.715	-0.055
55.	+ 15121193	4.75	5.017	0.267
56.	- 15121199	4.67	4.62	-0.050
57.	+ 12640225	4.64	4.36	-0.280
58.	#	4.59	4.957	0.367
	15121195			
59.	+ 15121181	4.57	4.866	0.296
60.	+ 12640191	4.57	4.301	-0.269
61.	+ 15121189	4.50	4.499	-0.001
62.	- 15121204	5.30	5.171	-0.129
63.	+18379352	4.48	4.485	0.005
64.	- 15121196	4.47	4.502	0.032
65.	+ 15121201	4.40	4.62	0.220
66.	# 12640191	4.35	4.491	0.141
67.	+ 3442	4.29	4.524	0.234
68.	+ 15121207	4.09	4.262	0.172

Table 9- Validation of the predicted model using external dataset compounds

S.No	Compound ID	Observed	Predicted	Split	Residuals
1.	44419280	3.910	4.912, 4.121	1 & 2	-1.00 & -0.21
2.	15121183	4.510	4.60	2	-0.09
3.	8027170	4.080	3.920	1	0.16

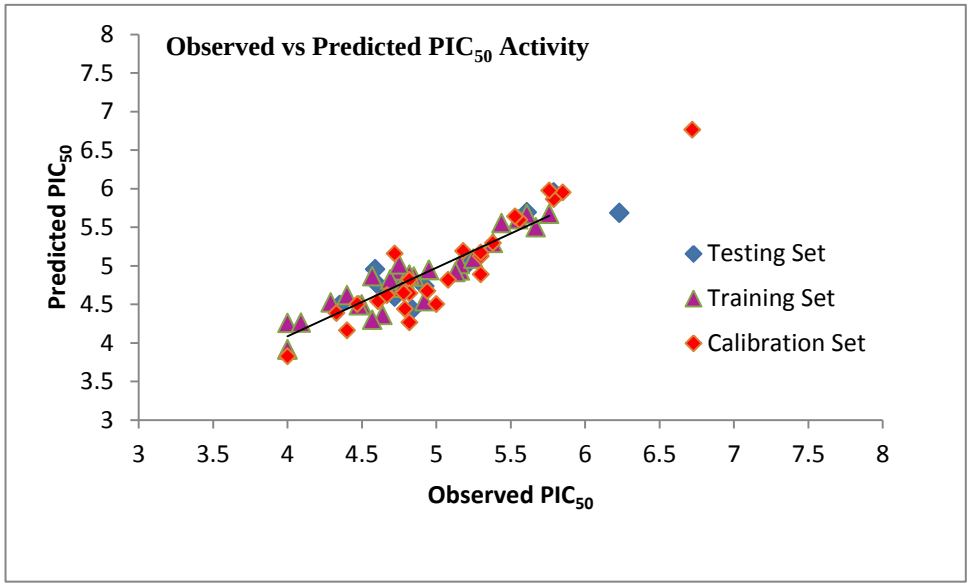


Figure 1 Plot of observed Vs Predicted PIC_{50}

References

[1] Afantitis, A., Melagraki, G., Koutentis, P.A., Sarimveis, H. and Kollias, G., 2011. "Ligand-based virtual screening procedure for the prediction and the identification of novel β-amyloid aggregation inhibitors using Kohonen maps and Counter propagation Artificial Neural Networks". European journal of medicinal chemistry, 46(2), pp.497-508.

[2] Appelhoff, R.J., Tian, Y.M., Raval, R.R., Turley, H., Harris, A.L., Pugh, C.W., Ratcliffe, P.J. and Gleadle, J.M., 2004. "Differential function of the prolyl hydroxylases PHD1, PHD2, and PHD3 in the regulation of hypoxia-inducible factor". Journal of Biological Chemistry, 279(37), pp.38458-38465.

[3] Brahimi-Horn, M.C. and Pouysségur, J., 2009. "HIF at a glance". Journal of cell science, 122(8), pp.1055-1057.

[4] Bruick, R.K., 2003. "Oxygen sensing in the hypoxic response pathway: regulation of the hypoxia-inducible transcription factor". Genes & development, 17(21), pp.2614-2623.

[5] Chiche, J., Brahimi-Horn, M.C. and Pouysségur, J., 2010. "Tumour hypoxia induces a metabolic shift causing acidosis: a common feature in cancer". Journal of cellular and molecular medicine, 14(4), pp.771-794.

[6] Došlic, T., Furtula, B., Graovac, A., Gutman, I., Moradi, S. and Yarahmadi, Z., 2011. "On vertex–degree–based molecular structure descriptors". MATCH Commun. Math. Comput. Chem, 66, pp.613-626.

[7] Furtula, B. and Gutman, I., 2011. "Relation between second and third geometric–arithmetic indices of trees". Journal of Chemometrics, 25(2), pp.87-91.

[8] Ibezim, E., Duchowicz, P.R., Ortiz, E.V. and Castro, E.A., 2012. "QSAR on aryl-piperazine derivatives with activity on malaria", Chemometrics and Intelligent Laboratory Systems, 110(1), pp.81-88.

[9] Lee, M.J., Kim, J.W. and Yang, E.G., 2010. "Hinokitiol activates the hypoxia-inducible factor (HIF) pathway through inhibition of HIF hydroxylases". Biochemical and biophysical research communications, 396(2), pp.370-375.

[10] Linden, T., Katschinski, D.M., Eckhardt, K., Scheid, A., Pagel, H. and Wenger, R.H., 2003. "The antimycotic ciclopirox olamine induces HIF-1α stability, VEGF expression, and angiogenesis". The FASEB Journal, 17(6), pp.761-763.

[11] Metzen, E., Stiehl, D.P., Doege, K., Marxsen, J.H., Hellwig-Bürgel, T. and Jelkmann, W., 2005. "Regulation of the prolyl hydroxylase domain protein 2 (phd2/egln-1) gene: identification of a functional hypoxia-responsive element". Biochemical Journal, 387(3), pp.711-717.

[12] Mitra, I., Saha, A. and Roy, K., 2010. "Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants". Molecular Simulation, 36(13), pp.1067-1079.

[13] Nangaku, N., Kojima, I., Tanaka, T., Ohse, T., Kato, H. and Fujita, F., 2006. "Novel drugs and the response to hypoxia: HIF stabilizers and prolyl hydroxylase". Recent patents on cardiovascular drug discovery, 1(2), pp.129-139.

[14] Ojha, P.K. and Roy, K., 2011. "Comparative QSARs for antimalarial endochins: importance of descriptor-thinning and noise reduction prior to feature selection". Chemometrics and Intelligent Laboratory Systems, 109(2), pp.146-161.

[15] Puzyn, T., Mostrag-Szlichtyng, A., Gajewicz, A., Skrzyński, M. and Worth, A.P., 2011. "Investigating the influence of data splitting on the predictive ability of QSAR/QSPR models". Structural Chemistry, 22(4), pp.795-804.

[16] Rabinowitz, M.H., Barrett, T.D., Rosen, M.D. and Venkatesan, H., 2010. "Inhibitors of HIF prolyl hydroxylases". Annual Reports in Medicinal Chemistry, 45, pp.123-139.

[17] Rasulev, B.F., Toropov, A.A., Hamme, A.T. and Leszczynski, J., 2008. "Multiple linear regression analysis and optimal descriptors: predicting the cholesteryl ester transfer protein inhibition activity". QSAR & Combinatorial Science, 27(5), pp.595-606.

[18] Rosen, M.D., Venkatesan, H., Peltier, H.M., Bembenek, S.D., Kanelakis, K.C., Zhao, L.X., Leonard, B.E., Hocutt, F.M., Wu, X., Palomino, H.L. and Brondstetter, T.I., 2010. "Benzimidazole-2-pyrazole HIF prolyl 4-hydroxylase inhibitors as oral erythropoietin secretagogues". ACS medicinal chemistry letters, 1(9), pp.526-529.

[19] Roy, K. and Mitra, I., 2012. "On the use of the metric rm 2 as an effective tool for validation of QSAR models in computational drug design and predictive toxicology". Mini reviews in medicinal chemistry, 12(6), pp.491-504.

[20] Semenza, G.L., 2004. "Hydroxylation of HIF-1: oxygen sensing at the molecular level". Physiology, 19(4), pp.176-182.

[21] Toropov, A.A., Rasulev, B.F. and Leszczynski, J., 2008. "QSAR modeling of acute toxicity by balance of correlations". Bioorganic & medicinal chemistry, 16(11), pp.5999-6008.

[22] Toropov, A.A., Toropova, A.P., Benfenati, E., Gini, G. and Fanelli, R., 2013. "The definition of the molecular structure for potential anti-malaria agents by the Monte Carlo method". Structural Chemistry, 24(4), pp.1369-1381.

[23] Toropov, A.A., Toropova, A.P., Martyanov, S.E., Benfenati, E., Gini, G., Leszczynska, D. and Leszczynski, J., 2011. "Comparison of SMILES and molecular graphs as the representation of the molecular structure for QSAR analysis for mutagenic potential of polyaromatic amines". Chemometrics and Intelligent Laboratory Systems, 109(1), pp.94-100.

[24] Toropova, A.P., Toropov, A.A., Benfenati, E., Gini, G., Leszczynska, D. and Leszczynski, J., 2011. "CORAL: quantitative structure–activity relationship models for estimating toxicity of organic compounds in rats". Journal of computational chemistry, 32(12), pp.2727-2733.

[25] Toropova, A.P., Toropov, A.A., Benfenati, E., Gini, G., Leszczynska, D. and Leszczynski, J., 2011. "QSAR modeling of anxiolytic activity taking into account the presence of keto-and enol-tautomers by balance of correlations with ideal slopes". Central European Journal of Chemistry, 9(5), pp.846-854.

[26] Tsukiyama, F., Nakai, Y., Yoshida, M., Tokuhara, T., Hirota, K., Sakai, A., Hayashi, H. and Katsumata, T., 2006. "Gallate, the component of HIF-inducing catechins, inhibits HIF prolyl hydroxylase". Biochemical and biophysical research communications, 351(1), pp.234-239.

[27] Warshakoon, N.C., Wu, S., Boyer, A., Kawamoto, R., Sheville, J., Renock, S., Xu, K., Pokross, M., Evdokimov, A.G., Walter, R. and Mekel, M., 2006. "A novel series of imidazo [1, 2-a] pyridine derivatives as HIF-1α prolyl hydroxylase inhibitors". Bioorganic & medicinal chemistry letters, 16(21),pp.5598-5601