Computational modeling studies on anti-HIV-1 non-nucleoside reverse transcriptase inhibition by dihydroalkoxybenzyloxopyrimidines analogues: an electrotopological atomistic approach

Nitin S. Sapre[1*], Tarang Bhati[1], Swagata Gupta[2], Nilanjana Pancholi[1], Urmila Raghuvanshi[1], Divya Dubey[1], Vandana Rajopadhyay[1], Neelima Sapre[3]

[1]Department of Applied Chemistry, SGSITS, Indore, India, [*]Corresponding Author: sukusap@yahoo.com, nsapre@sgsits.ac.in
[2]Department of Chemistry, Bherulal Patidar Post Graduate College, Mhow, Indore, India;
[3]Department of Computer Applications, SV College of Engineering, Indore, India.

ABSTRACT

For the first time we report quantitative structure activity relationship (QSAR) studies based on Kier-Hall Electrotopological State (E-State) Indices for Dihydroalkoxybenzyloxopyrimidines (DABO) derivatives acting as NNRTIs of HIV-1. A dataset of 74 compounds was compiled from published studies and randomly subdivided into training and test sets. To understand the pharmacophoric effect, Kier-Hall Electrotopological State descriptors namely S_{N_1}, S_{N_3}, S_F, S_{Ar}, S_S, S_O, S_{NO_2}, S_{Cl}, S_Y (Y = S-alkyl and NH-alkyl), S_X (X = Me) and biological activity were used as independent and dependent variable respectively. Statistical results were highly encouraging for the training set [multiple linear regression [(MLR): r^2 = 0.961, F = 100.41 and q^2 = 0.926, neural networks (NN): r^2 = 0.966, F = 115.594, degrees of freedom = 40 and k-nearest neighbour (k-NN): r^2 = 0.770, q^2 = 0.757, degrees of freedom = 40]. Results of validation using a test set showed the same trend as training set (NN > MLR > kNN). The above results suggest that of various functional groups present in DABO S_{N_3}, S_O, S_{Cl}, S_{Ar} and S_{NO_2} contribute more significantly towards activity. On the other hand S_{N_1}, S_S, and S_F do not play any role in enhancing the activity. The substitution of S-alkyl and NH-alkyl at C_2 position is essential though it does not contribute much towards the activity. The substitution of methyl group at C_5 position is unfavorable and exhibit negative impact on inhibitory activity. Therefore, it seems reasonable to choose E-State indices as suitable and significant descriptors for exploring the relationship between the pIC_{50} and the pharmacological properties of the compounds.

Keywords: AIDS; HIV-1; NNRTIs; DABOs; QSAR; pIC_{50}; Kier Hall E-State Indices; MLR; NN; k-NN

1. INTRODUCTION

Acquired Immunodeficiency Syndrome (AIDS) is a set of symptoms and infections resulting from the damage of the human immune system caused by the human immunodeficiency virus (HIV) [1]. AIDS cause very serious public health problem and economic burden. Globally, an estimated 33.4 million people are living with HIV with nearly 7500 new infections each day [2]. An anti-HIV agent may exert its activity by inhibiting a variety of steps in the life cycle of the virus [3]. Reverse transcription is a highly choreographed, multistep process in which the plus-strand RNA genome of the HIV-1 is converted into a double-stranded cDNA [4]. HIV-1 reverse transcriptase (RT) is an asymmetric heterodimer composed of two related subunits, p66 (560 amino acids) and p51 (440 amino acids). p66 folds into two domains: polymerase and RNase H. The polymerase domain of p66 is divided into four subdomains fingers (amino acids 1 - 85 and 118 - 155), a palm (amino acids 86 - 117 and 156 - 237) and a thumb (amino acids 238 - 318) domain. [5] The two subunits form a stable dimer which is essential for the enzymatic activity [6]. The RT enzyme is known to be flexible, moving both the domains to accommodate inhibitors and to take on different shapes depending on the inhibitor bound to it [7]. Inhibitors of HIV-1 RT fall under two categories: the nucleoside (NRTIs) and the non-nucleoside reverse transcriptase

inhibitors (NNRTIs). The NRTIs are substrate analogs that act as chain terminators, whereas the NNRTIs are a chemically diverse group of compounds that non-competitively inhibit DNA polymerization [8]. Twenty-five years ago, nucleoside analog 3'-azidothymidine (AZT) was shown to efficiently block the replication of HIV in a cell culture [9]. NNRTIs form a second class of compounds that target HIV-1 RT. [10] NNRTIs bind at an allosteric site about 10 Å away from the polymerase active site. The structural mechanism of NNRTI inhibition involves distortion of the catalytically essential triad of aspartic acid residues [11]. The non-nucleoside site, although contained within the p66 subunit, also has Glu138 from the p51 subunit located at the edge of the inhibitor pocket [12]. Over 40 amino acid substitutions have been identified to be associated with NNRTI resistance, *in vitro* and *in vivo* [13]. Almost twenty-one years ago, the NNRTIs *i.e.*, 1-(2-2-hydroxyethoxy methyl)-6-(phenylthio) thymine (HEPT) [14,15] and Tetrahydro-imidazo[4,5,1-jkj] [1,4] benzodiazepin-2(1H)-one and thione (TIBO) compounds [16,17] have been discovered. [18] Since then three more NNRTIs-nevirapine, delavirdine and efavirenz have been approved for the treatment of HIV-1 and a few more are in clinical development, including rilpivirine, etravirine and dapivirine [19]. Till date various molecular modelling studies have been performed using QSAR [20], QSPR [21], Simple docking [22] and Template docking [23] on some NNRTIs.

Pharmacophore based descriptors are of immense importance in understanding the structure activity relationships of ligand molecules. Dihydroalkoxybenzyloxopyrimidines (DABO) derivatives have been successfully used as HIV-1 reverse transcriptase inhibitors [24] and are useful precursors for deriving novel potent compounds. The first step of rational drug design often consists of performing QSAR modeling, based on congeneric structures of ligands; this procedure has been carried out by several researchers for a class of DABO derivatives [25-28]. Like other NNRTIs, DABOs also assume a "butterfly like" spatial arrangement with the methylene group or moiety serving as a bridge between the two wings of the benzene and pyrimidine rings [27]. 5-isopropyl-2-[(methylthiomethyl)thio]-6-benzylpyrimidin-4-(1H)-one DABO analogue was found to be very potent against HIV-1 Reverse Transcriptase (RT) [28]. Dihydro-alkyl-thio-benzyl-oxopyrimidines (S-DABOs) are common with EBUs (Emivirine, formerly MKC-442) in having a benzyl moiety at position C-6 of the pyrimidine ring while they differ from EBUs in that the alkoxy chain is linked at position C-2 instead of N-1. Sudbeck *et al.* have performed derivatization of DABO by modelling studies of potent 1-[(2-hydroxy ethoxy) methyl]-6-(phenylthio) thymine (HEPT) and by the sub-

stitution of benzyl moiety with phenyl thio moiety at the C-6 position of the pyrimidine ring [29]. A structure-activity relationship (SAR) profile of DABOs together with molecular modeling investigation on their putative binding mode have shown that the presence of a C-2-alkoxy (DABOs), C-2-alkylthio (S-DABOs), or C-2-alkylamino (NH-DABOs) side chain is the structural determinant for the antiviral activity of these derivatives, with the length and size of the C-2 side chain having only modulator effects on potency [30-34]. The 2,6-difluoro substitution at the C-6-benzyl moiety of S-DABOs and NH-DABOs produced favorable π-stacking interactions with the Tyr188 side chain into the non-nucleoside binding site (NNBS), leading to compounds (F2-S-DABOs and F2-NH-DABOs) active in the nanomolar range. [35] The biological testing results clearly indicated that the substitution of halogen at the C5 position of pyrimidine ring could increase the anti-HIV-1 RT activity. The most active compounds showed activity in the lower micromole range with IC_{50} values (IC_{50} 0.18 - 3.03 μM) comparable to nevirapine (IC_{50} 4.12 μM). The docking showed that a new halogen bond was formed between halogen and carbonyl of TYR188 in the HIV-I RT [36].

Various statistical modeling techniques have been widely used in quantitative structure-activity relationship (QSAR) studies, such as multiple linear regression (MLR), [37] partial least squares (PLS), [37] artificial neural networks (ANN) [38]. Using such an approach, one could predict the activities of newly designed compounds before a decision is being made whether these compounds should be really synthesized and tested. Thus, for the first time a QSAR study based on Kier-Hall Electrotopological State (E-State) [39-41] indices is performed on DABO derivatives to analyze the pharmacophoric effect on their inhibitory activities. Recently, a comparative molecular field analysis (CoMFA) study was performed in order to derive three-dimensional quantitative structure activity relationship (3D-QSAR) models of DABO analogues [28]. The rationale of the present work is to explore the structure activity relationship of E-State indices of DABO compounds and the biological activity in a quantitative manner using linear (MLR) as well as non-linear (NN and k-NN) techniques.

2. MATERIALS AND METHODS

A series of 74 compounds of DABO derivatives was taken from the literature [28] for the QSAR study. All the 74 compounds were sketched using V-life MDS software [41]. The inhibitory activity pIC_{50} was taken as the biological activity variable (dependent). The functional groups namely S_{N_1}, S_{N3}, S_O, S_X, S_Y, S_{Ar}, S_S, S_F, S_{Cl}, and S_{NO_2} were considered as descriptor variable

(independent) in the QSAR analysis. The Kier-Hall Electrotopological State (E-State) indices of such functional groups were calculated using E-Calc version 1.1 [42]. Multiple linear regression (MLR), neural network (NN) regression and k-Nearest Neighbor (k-NN) were calculated with V-Life MDS software.

2.1. Kier-Hall Electrotopological State (E-State) Indices [38]

Electrotopological State (E-State) indices are widely used in QSAR modeling, including AIDS related research [21,43,44]. The large amount of variables in E-State Indices represent the structural characteristics of molecules such as information about their non-covalent interactions. It encodes electronic and topological information in a single number for each skeletal atom in the hydrogen suppressed graph of a molecule. The E-State index value (S) of an atom "I" is given by the sum of the intrinsic state value (I) and perturbations (ΔI) by the fields of all other atoms in the chemical graph, including substituent atoms. Thus, S is defined by the following set of equations:

$$S = I + \Delta I$$

$$I = \left(\left(2/N \right)^2 \delta^v + 1 \right) \Big/ \delta$$

$$\delta^v = \sigma + \pi + n - h$$

$$\delta = \sigma - h$$

$$\Delta I_i = \Sigma \left(I_i - I_j \right) r_{ij}^2$$

where, N = Principal quantum number (*i.e.*, row in which the atom occurs in the periodic table)

σ = the number of sigma electron contributed by the atom.

Π = the number of pi electron contributed by the atom.

h = the number of hydrogen atoms attached to the atom

n = the number of lone pair of e^- on the atom

r = the number of atoms in the shortest graph path connecting atom "I" and other atom "j" in the chemical graph of the whole molecule.

The resulting E-State of any atom is a numerical value depicting the accessibility of that atom to interaction across space with some reference atom (s) or group (s).

2.2. Statistical Analyses

2.2.1. Multiple Linear Regression (MLR)

Multiple linear regression (MLR) is a method used to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to the observed was employed to correlate the bind-

ing affinity and molecular descriptors. This method has been widely applied in many QSAR studies, and has proven to be a useful linear regression method to build QSAR models that may explore straightforward the properties of the chemical structure in combination with its ability of inducing a pharmacological response [45]. The advantage of MLR is its simple form and easily interpretable mathematical expression.

2.2.2. Neural Network (NN)

Intriguing approaches using machine learning methods have been studied in the field of QSAR modeling [46]. The concept of neural network was first introduced in 1943 by McCulloch and Pitts [47]. Interest in neural networks was slow until the 1980s when new computer architecture and learning algorithms began to appear. The use of neural networks (NN) in all fields has since grown substantially. In 1988, Hoskins *et al*. reported the first use of process control in chemistry [48]. NNs are well suited for modelling complex relationships between the variables and non-linearity in the relationships is suspected.

2.2.3. Kernel Nearest Neighbor (k-NN)

k-NN [49,50] is a simple decision scheme that requires practically no training and is asymptotically optimal, *i.e.*, with increase in training data it converges to the optimal prediction error. For a given compound in the descriptor space, the method analyzes its k-nearest neighboring compounds from the training set and predicts the activity class that is most highly represented among these neighbors.

3. RESULTS AND DISCUSSION

Table 1 records a series of 74 DABO derivatives along with the position of substituents (X, Ar and Y), where X represents substituent such as methyl attached at C-5 position, Ar represents the aryl groups substituted with the terminal methyl group which is attached at C-6 position of the pyrimidinone ring and Y represents sulphur substituted moieties attached at C-2 position of the basic nucleus. It also records the inhibitory concentration pIC_{50} (where IC_{50} is the effective concentration of a compound required to activate 50% protection of MT-4 cell against the cytopathic effect of HIV-1) and the calculated E-State values (N_1, N_3, O, X, Y, Ar, S, F, Cl, and NO_2) of various atoms or groups in the DABO analogues.

The correlation matrix for the correlation of E-State Indices with pIC_{50} for the training set of DABO derivatives is shown in **Table 2**. A perusal of the correlation matrix indicates that in univariate correlation S_{N_1} has the highest correlation potential while S_X has the lowest one.

Table 1. Table represents the various substituents with the anti-HIV-1 activity (pIC_{50}) and E-State indices of Dihydroalkoxybenzy-loxopyrimidines (DABO) derivatives.
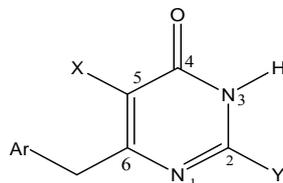
S.No.[*]	X	Y	Ar	pIC_{50}[**]	S_{N_1}	S_{N_3}	S_O	S_X	S_Y	S_{Ar}	S_S	S_F	S_{Cl}	S_{NO_2}
1	H	1-naphtyl	S-Cyclopentyl	4.31	4.71	2.89	12.03	0.00	7.35	18.32	1.73	0.00	0.00	0.00
2	Me	1-naphtyl	S-Cyclopentyl	4.35	4.81	2.96	12.38	1.87	7.35	18.39	1.74	0.00	0.00	0.00
3	H	2-naphtyl	S-sec-Bu	4.83	4.60	2.83	11.88	0.00	7.38	18.28	1.62	0.00	0.00	0.00
4	H	4-F-Ph	S-sec-Bu	4.83	4.45	2.74	11.66	0.00	7.16	19.78	1.56	12.86	0.00	0.00
5	H	4-Cl-Ph	S-sec-Bu	5.02	4.51	2.78	11.70	0.00	7.31	15.26	1.60	5.86	0.00	0.00
6	H	3-Me-Ph	S-tert-Bu	5.09	4.55	2.80	11.75	0.00	7.91	12.74	1.58	0.00	0.00	0.00
7	Me	3-Me-Ph	S-sec-Bu	5.27	4.65	2.88	12.08	1.84	7.41	12.83	1.63	0.00	0.00	0.00
8	Me	Ph	S-Me	5.31	4.45	2.75	11.70	1.81	3.35	11.22	1.45	0.00	0.00	0.00
9	Me	3-Me-Ph	S-tert-Bu	5.34	4.65	2.87	12.10	1.84	7.93	12.80	1.59	0.00	0.00	0.00
10	H	3-Cl-Ph	S-sec-Bu	5.42	4.51	2.78	11.70	0.00	5.18	15.33	1.59	0.00	5.97	0.00
11	Me	3-Me-Ph	S-Cyclopentyl	5.47	4.73	2.94	12.18	1.86	7.40	12.93	1.74	0.00	0.00	0.00
12	Me	3-F-Ph	S-sec-Bu	5.52	4.53	2.80	12.01	1.75	7.13	20.19	1.55	13.25	0.00	0.00
13	H	2,6-di-Cl-Ph	S-Me	5.52	4.42	2.71	11.81	1.75	3.25	19.59	1.39	0.00	6.14	0.00
14	H	Ph	S-Cyclohexyl	5.52	4.63	2.88	11.83	0.00	8.75	11.34	1.74	0.00	0.00	0.00
15	H	Ph	S-Cyclopentyl	5.55	4.60	2.86	11.77	0.00	7.40	11.33	1.72	0.00	0.00	0.00
16	H	3-Me-Ph	S-Cyclohexyl	5.59	4.66	2.89	11.90	0.00	8.74	12.89	1.74	0.00	0.00	0.00
17	Me	3-Me-Ph	S-Me	5.60	4.48	2.76	11.77	1.82	5.18	12.76	1.45	0.00	0.00	0.00
18	H	4-NO₂-Ph	S-sec-Bu	5.62	4.45	2.73	11.72	0.00	7.05	28.05	1.53	0.00	0.00	0.00
19	Me	3-Me-Ph	S-Cyclohexyl	5.66	4.76	2.96	12.25	1.87	8.74	12.95	1.75	0.00	0.00	0.00
20	Me	Ph	S-tert-Bu	5.72	4.62	2.86	12.03	1.83	7.92	11.27	1.58	0.00	0.00	0.00
21	H	2,6-di-Cl-Ph	S-iso-Pr	5.89	4.44	2.72	11.68	0.00	5.95	19.58	1.51	0.00	6.14	0.00
22	Me	2,6-di-Cl-Ph	S-iso-Pr	5.94	4.54	2.79	12.03	1.76	5.96	19.72	1.52	0.00	6.19	0.00
23	Me	2,6-di-Cl-Ph	S-n-Bu	5.94	4.57	2.82	12.08	1.77	6.85	19.80	1.57	0.00	6.21	0.00
24	Me	2,6-di-Cl-Ph	S-tert-Bu	5.96	4.59	2.82	12.13	1.76	7.70	19.76	1.52	0.00	6.21	0.00
25	Me	2-Cl-Ph	S-sec-Bu	6.10	4.61	2.85	12.06	1.80	7.29	15.54	1.60	0.00	6.19	0.00
26	Me	3-NO₂-Ph	S-sec-Bu	6.10	4.53	2.79	12.08	1.72	7.00	28.53	1.52	0.00	0.00	21.31
27	H	3-NO₂-Ph	S-sec-Bu	6.22	4.43	2.72	11.73	0.00	6.99	28.33	1.52	0.00	0.00	21.17
28	H	2,6-di-Cl-Ph	S-tert-Bu	6.22	4.49	2.75	11.78	0.00	7.68	19.61	1.52	0.00	6.16	0.00
29	H	2,6-di-Cl-Ph	S-n-Bu	6.30	4.47	2.75	11.73	0.00	6.84	19.65	1.56	0.00	6.16	0.00
30	H	2,6-di-Cl-Ph	S-Cyclopentyl	6.40	4.57	2.82	11.87	0.00	7.12	19.78	1.67	0.00	6.19	0.00
31	H	2,6-di-F-Ph	S-tert-Bu	6.70	4.27	2.63	11.66	0.00	7.23	29.63	1.39	13.65	0.00	0.00
32	H	2,6-di-F-Ph	S-sec-Bu	7.00	4.27	2.64	11.65	0.00	6.73	29.65	1.44	13.64	0.00	0.00
33	Me	2,6-di-F-Ph	S-sec-Bu	7.00	4.37	2.71	11.99	1.61	6.74	29.95	1.45	13.78	0.00	0.00
34	H	2,6-di-F-Ph	S-Cyclohexyl	7.05	4.38	2.72	11.82	0.00	7.88	2.96	1.56	13.75	0.00	0.00

35	Me	2,6-di-F-Ph	S-tert-Bu	7.05	4.27	2.63	11.66	0.00	7.23	29.63	1.39	13.65	0.00	0.00
36	H	2,6-di-F-Ph	S-Cyclopentyl	7.10	4.35	2.70	11.75	0.00	6.62	29.85	1.54	13.71	0.00	0.00
37	Me	2,6-di-F-Ph	S-Cyclopentyl	7.10	4.45	2.77	12.10	1.63	6.62	30.16	1.55	13.84	0.00	0.00
38	H	2,6-di-F-Ph	NH-Cyclopentyl	7.15	4.28	2.62	11.72	0.00	7.86	29.76	0.00	13.70	0.00	0.00
39	Me	2,6-di-F-Ph	NH-Cyclopentyl	7.52	4.37	2.69	12.07	1.61	7.88	30.06	0.00	13.83	0.00	0.00
40	Me	1-naphtyl	S-sec-Bu	4.35	4.73	2.90	12.28	1.85	7.39	18.27	1.63	0.00	0.00	0.00
41	H	2-naphtyl	S-Cyclohexyl	4.48	4.74	2.91	12.10	0.00	8.69	18.35	1.74	0.00	0.00	0.00
42	H	Ph	S-sec-Bu	5.27	4.52	2.80	11.66	0.00	7.39	11.24	1.62	0.00	0.00	0.00
43	Me	Ph	S-Cyclopentyl	5.47	4.70	2.93	12.12	1.86	7.40	11.38	1.73	0.00	0.00	0.00
44	H	3-Me-Ph	S-Cyclopentyl	5.59	4.63	2.87	11.83	0.00	7.40	12.87	1.73	0.00	0.00	0.00
45	Me	Ph	S-iso-Pr	5.60	4.62	2.86	12.01	1.84	7.40	11.30	1.63	0.00	0.00	0.00
46	Me	3-Cl-Ph	S-sec-Bu	5.74	4.61	2.85	12.05	1.81	7.31	15.42	1.60	0.00	6.01	0.00
47	H	2-NO_2-Ph	S-sec-Bu	6.22	4.40	2.70	11.75	0.00	6.91	28.76	1.49	0.00	0.00	21.67
48	H	2,6-di-Cl-Ph	S-Cyclohexyl	6.40	4.60	2.84	11.94	0.00	8.44	19.83	1.68	0.00	6.21	0.00
49	Me	2,6-di-F-Ph	S-Me	6.70	4.20	2.59	11.68	1.59	3.04	29.53	1.27	13.59	0.00	0.00
50	Me	2,6-di-F-Ph	S-n-Bu	7.05	4.36	2.70	11.96	1.61	6.44	29.94	1.45	13.76	0.00	0.00
51	Me	2,6-di-F-Ph	S-Cyclohexyl	7.15	4.48	2.79	12.17	1.63	7.88	30.26	1.57	13.88	0.00	0.00
52	Me	4-Cl-Ph	S-sec-Bu	4.77	4.61	2.85	12.05	1.81	7.32	15.34	1.61	0.00	5.90	0.00
53	Me	2,6-di-Cl-Ph	S-Cyclohexyl	5.31	4.70	2.91	12.29	1.79	8.44	19.98	1.69	0.00	6.26	0.00
54	Me	2,6-di-Cl-Ph	S-sec-Bu	6.92	4.59	2.83	12.12	1.77	7.18	19.79	1.57	0.00	6.22	0.00
55	H	3-F-Ph	S-sec-Bu	5.92	4.43	2.73	11.66	0.00	7.12	20.04	1.55	13.15	0.00	0.00
56	Me	2-naphtyl	S-sec-Bu	4.23	4.70	2.89	12.22	1.85	7.39	18.35	1.63	0.00	0.00	0.00
57	Me	4-F-Ph	S-sec-Bu	4.59	4.55	2.81	12.01	1.77	7.18	19.91	1.57	12.94	0.00	0.00
58	H	1-naphtyl	S-sec-Bu	4.79	4.63	2.83	11.93	0.00	7.38	18.19	1.62	0.00	0.00	0.00
59	H	Ph	S-tert-Bu	5.07	4.52	2.79	11.68	0.00	7.90	11.21	1.57	0.00	0.00	0.00
60	Me	Ph	S-sec-Bu	5.32	4.62	2.86	12.01	1.84	7.40	11.30	1.63	0.00	0.00	0.00
61	Me	Ph	S-Cyclohexyl	5.37	4.73	2.95	12.18	1.86	8.75	11.40	1.75	0.00	0.00	0.00
62	Me	4-NO_2-Ph	S-sec-Bu	5.44	4.55	2.80	12.07	1.74	7.07	28.21	1.54	0.00	0.00	20.95
63	H	2-Cl-Ph	S-sec-Bu	5.49	4.51	2.78	11.71	0.00	7.27	15.44	1.59	0.00	6.14	0.00
64	H	3-Me-Ph	S-iso-Pr	5.54	4.51	2.77	11.64	0.00	6.14	12.74	1.57	0.00	0.00	0.00
65	Me	3-Me-Ph	S-iso-Pr	5.60	4.61	2.84	11.99	1.83	6.16	12.80	1.58	0.00	0.00	0.00
66	Me	2,6-di-Cl-Ph	S-Cyclopentyl	5.80	4.66	2.89	12.22	1.79	8.91	19.93	1.68	0.00	6.25	0.00
67	H	2,6-di-F-Ph	S-Me	6.10	4.10	2.52	11.33	0.00	3.02	29.23	1.26	13.46	0.00	0.00
68	Me	2-F-Ph	S-sec-Bu	6.10	4.50	2.78	12.00	1.72	7.07	20.63	1.54	13.75	0.00	0.00
69	H	2-F-Ph	S-sec-Bu	6.22	4.40	2.72	11.65	0.00	7.06	20.44	1.53	13.62	0.00	0.00
70	H	2,6-di-F-Ph	S-n-Bu	6.70	4.26	2.63	11.61	0.00	6.43	29.63	1.44	13.63	0.00	0.00
71	H	2,6-di-Cl-Ph	S-sec-Bu	6.70	4.49	2.76	11.77	0.00	7.16	19.64	1.56	0.00	6.16	0.00
72	H	2,6-di-F-Ph	S-iso-Pr	7.30	4.23	2.60	11.56	0.00	5.57	29.52	1.39	13.59	0.00	0.00
73	H	3-Me-Ph	S-sec-Bu	5.62	4.55	2.81	11.73	0.00	7.39	12.77	1.62	0.00	0.00	0.00
74	Me	2,6-di-F-Ph	S-iso-Pr	7.30	4.33	2.67	11.91	1.60	5.59	29.82	1.40	13.72	0.00	0.00

[*]Compounds no. (1-55) corresponds to training set and underlined numbers (56-74) corresponds to test set. [**]$pIC_{50} = -\log IC_{50}$ (where IC_{50} is the effective concentration of a compound required to activate 50% protection of MT-4 cell against the cytopathic effect of HIV-1)

Table 2. Correlation matrix between the E-State Indices and biological activity of DABO derivatives.

	pIC_{50}	S_{N_1}	S_{N_3}	S_O	S_X	S_Y	S_{Ar}	S_S	S_F	S_{Cl}	S_{NO_2}
pIC_{50}	1										
S_{N_1}	0.54	1									
S_{N_3}	0.46	0.98	1								
S_O	0.05	0.46	0.48	1							
S_X	0.00	0.07	0.08	0.49	1						
S_Y	0.00	0.20	0.22	0.12	0.05	1					
S_{Ar}	0.35	0.41	0.42	0.01	0.00	0.03	1				
S_S	0.25	0.29	0.31	0.02	0.00	0.00	0.20	1			
S_F	0.44	0.50	0.43	0.04	0.00	0.00	0.34	0.22	1		
S_{Cl}	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.10	1	
S_{NO_2}	0.00	0.01	0.02	0.00	0.00	0.00	0.09	0.00	0.00	0.01	1

The order of univariate correlations of E-State values for various substituents attached to the parent compound with inhibitory activity follow the order:

$$S_{N_1} > S_{N_3} > S_F > S_{Ar} > S_S > S_O > S_{NO_2} > S_{Cl} = S_Y > S_X$$

The QSAR models were generated using MLR, NN and k-NN techniques to study the antiviral activity of Dihydroalkoxybenzyloxopyrimidines derivatives. The dataset of 74 compounds was divided into a training set of 55 compounds for developing the MLR, NN and k-NN models and a prediction test set of 19 compounds were used to determine the external predictivity of the QSAR model. Both the training and the test sets were divided manually according to a representative range of biological activities and structural variations. Similar training and test sets were used to perform MLR, NN and k-NN methods.

3.1. Multiple Linear Regression (MLR)

The statistical analysis was performed by multiple linear regression method using back-propagation. The best model is presented by the **Eq.1** for the training set of 55 compounds.

$$pIC_{50} = -32.263 * S_{N_1} + 35.330 * S_{N_3} + 3.359 * SO$$
$$-0.168 * S_X + 0.189 * S_Y + 0.001 * S_{Ar} - 0.271 * S_S$$
$$-0.0754 * S_F + 0.0500 * S_{CL} + 0.000 * S_{NO_2} - 12.597 \quad (1)$$

$$(n = 51; \ r^2 = 0.961; \ F = 100.114; \ q^2 = 0.926)$$

In **Eq.1** a high positive coefficient for S_{N_3} indicates that the presence of nitrogen at 3rd position in the compound enhances the activity while a high negative coefficient for S_{N_1} indicates the presence of nitrogen at 1st position will decrease the antiviral activity to a great extent. It has been deduced earlier that the 3-NH function enhances anti-HIV activity by donation of a hydrogen bond to the carbonyl oxygen of Lys101 in RT. [51] This result coincide with the results reported by Guang Fu Sun et al that 3-NH of DABO was essential for the inhibition of HIV RT. [52] As far as functional groups such as oxygen, nitro and sulphur are concerned, O and NO_2 are the ones contributing towards enhancing the activity whilst presence of sulphur is exhibiting a retarding effect on the biological activity. Although, the coefficients for halogen groups namely F and Cl are not very significant (low coefficient) for activity, but substitution of fluorine group at benzyl group is less beneficial than chlorine moiety. This has been also observed in the work of Elena Petricci *et al.* [53] but their previous finding also highlighted the importance of 2, 6-difluorobenzyl substituent at position 6 for optimal activity. [54] The substitution of S-alkyl and NH-alkyl at C-2 position is essential, though it does not contribute much towards the activity. Similarly, presence of Ar exhibit negligible effect on inhibitory activity with its low positive coefficient value. On the other hand substitution of methyl group at C-5 position is unfavorable and exhibits a negative impact on the inhibitory activity. An earlier work also suggest that the DABOs are active when an H atom is present at position 5 of the pyrimidine ring of DABO derivatives [25].

3.2. Neural Networks (NN)

An NN model was developed for a training set of 55 DABO derivatives. V-life MDS software has been used to perform NN analysis. A typical artificial neural network (ANN) consists of an input layer, a hidden layer and an output layer (neurons). Each neuron in any layer is fully connected with the neuron of a succeeding layer. There are neither connections between the neurons

within a layer nor any direct connections between those of the input and output layers. A convergence criterion implies that the rms error value between the predicted and observed values is lower than the convergence criteria. Epoch means number or iterations in neural network runs. Pruning limit indicates the limit of variance of weights. Following parameters were set for the NN regression: max training epoch was set to 100,000 with pruning limit 0.3, two hidden neurons were set in the hidden layer. The convergence criteria were set to 0.01 and the variance cut off was 0.01 without scaling the data.

3.3. k-Nearest Neighbour (k-NN)

The k-NN model was built using V-life Molecular Design Software. The k-NN regression analysis was performed using following parameters: data were pre-processed using auto scaling option from the software, number of max neighbour = 5, number of min neighbour = 2 and variance cut-off = 0.01. The prediction ability method used was distance based weighted average.

The corresponding observed and predicted values of the pIC_{50} values of 55 compounds studied by MLR, NN and k-NN in this work are given in **Table 3**. The predictive model building abilities of all the three methods were analyzed and compared. Also, the quality of correlation is demonstrated by their respective residual values *i.e.* the difference between observed and calculated pIC_{50} values. The respective residual values thus obtained are also given in **Table 3**.

The models predictive power was judged based on various statistical parameters namely correlation coefficient (r^2), Fischer statistical or fitness value (F) and cross validation coefficient (q^2). All these statistical parameters were computed as defined in the V-Life MDS software.

A plot of observed and calculated pIC_{50} values of NN, MLR and k-NN for the training set of DABO derivatives is presented in **Figure 1**.
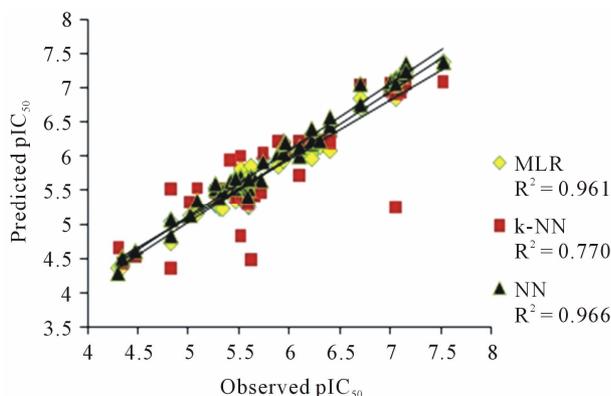
Figure 1. Graph between observed and calculated pIC_{50} values for training set.

3.4. Uni-column Statistics

The min and max values in both train and test set should be compared in a way that
- The maximum of the test should be less than or equal to maximum of training set.
- The minimum of the test set should be greater than or equal to minimum of the training set.

The above observation show that the test set is interpolative *i.e.* derived within the min-max range of the training set. The mean and standard deviation of the training and test set provides insight to the relative difference of mean and point density distribution (along mean) of the test set. In this case the mean in the training set slightly higher than the test set shows the presence of relatively more active molecules as compared to the inactive ones. Also a relatively higher standard deviation in training set indicated that training set has widely distributed activity of the molecules as compared to the test set.

Uni-column statistics for training set:

Name	Average	Max	Min	StdDev	Sum
pIC_{50}	5.888	7.52	4.31	0.819	300.30

Uni-column statistics for test set:

Name	Average	Max	Min	StdDev	Sum
pIC_{50}	5.747	7.30	4.23	0.832	132.00

A test set of 19 compounds are reported in **Table 4** and is used to verify the validity of the QSAR model of training set. **Table 4** also represents the predicted and observed values obtained by the statistical methods (MLR, NN and k-NN) showing a degree of relatedness.

A plot of observed and calculated pIC_{50} values of MLR, k-NN and NN for the test set of DABO derivatives is presented in **Figure 2**.

Methods namely MLR, NN and k-NN regression are widely used methods for building QSAR model. It is simple to interpret a regression model, in which contribution of each descriptor could be observed by the magnitude of its regression coefficient. In case of MLR, high value of squared regression coefficient $r^2 = 0.961$ was observed. The results obtained from NN were the most significant and gave a very high value of squared regression coefficient $r^2 = 0.966$. Though, the results of k-NN were significant with squared regression coefficient $r^2 = 0.770$, but were inferior amongst all the three methods. The results of all the three methods could be cross validated with the squared regression coefficient between observed activity and predicted activity values of the test set compounds. For the best linear model, *i.e.*, MLR ($r^2 =$

Table 3. Observed and calculated anti-HIV-1 activity (pIC_{50}) of dihydroalkoxybenzyloxopyrimidines (DABO) derivatives (training set).

S.No.	pIC_{50} Obs.	pIC_{50} Calc MLR	pIC_{50} Calc k-NN	pIC_{50} Calc NN	Residual MLR	Residual k-NN	Residual NN
1	4.31	4.36	4.66	4.29	0.05	0.35	0.02
2	4.35	4.42	4.42	4.53	0.07	0.07	–0.18
3	4.83	4.73	4.36	4.83	–0.10	–0.47	0.00
4	4.83	5.04	5.52	5.07	0.21	0.69	–0.24
5	5.02	5.13	5.33	5.13	0.11	0.31	–0.11
6	5.09	5.16	5.54	5.35	0.07	0.45	–0.26
7	5.27	5.28	5.47	5.52	0.01	0.20	–0.25
8	5.31	5.23	5.53	5.38	–0.08	0.22	–0.07
9	5.34	5.23	5.48	5.49	–0.11	0.14	–0.15
10	5.42	5.50	5.94	5.63	0.08	0.52	–0.21
11	5.47	5.34	5.39	5.53	–0.13	–0.08	–0.06
12	5.52	5.36	4.83	5.59	–0.16	–0.69	–0.07
13	5.52	5.76	6.01	5.71	0.24	0.49	–0.19
14	5.52	5.82	5.43	5.70	0.30	–0.09	–0.18
15	5.55	5.61	5.36	5.61	0.06	–0.19	–0.06
16	5.59	5.49	5.38	5.57	–0.10	–0.21	0.02
17	5.60	5.25	5.41	5.53	–0.35	–0.19	0.08
18	5.62	5.86	4.47	5.71	0.24	–1.15	–0.09
19	5.66	5.55	5.43	5.65	–0.11	–0.23	0.01
20	5.72	5.55	5.48	5.65	–0.17	–0.24	0.07
21	5.89	5.86	6.22	6.04	–0.03	0.33	–0.15
22	5.94	5.94	6.04	6.05	0.00	0.10	–0.11
23	5.94	6.22	6.08	6.18	0.28	0.14	–0.24
24	5.96	6.06	6.11	6.20	0.10	0.15	–0.24
25	6.10	6.02	5.73	6.00	–0.08	–0.37	0.10
26	6.10	6.14	6.22	6.14	0.04	0.12	–0.04
27	6.22	6.05	6.19	6.18	–0.17	–0.03	0.04
28	6.22	5.97	6.28	6.23	–0.25	0.06	–0.01
29	6.30	6.14	6.21	6.23	–0.16	–0.09	0.07
30	6.40	6.09	6.20	6.44	–0.31	–0.20	–0.04
31	6.70	6.85	7.05	7.05	0.15	0.35	–0.35
32	7.00	6.91	6.97	7.05	–0.09	–0.03	–0.05
33	7.00	7.01	7.08	7.08	0.01	0.08	–0.08
34	7.05	7.11	5.25	7.14	0.06	–1.80	–0.09
35	7.05	6.85	6.91	7.05	–0.20	–0.14	0.00
36	7.10	6.95	6.95	7.17	–0.15	–0.16	–0.07
37	7.10	7.04	7.05	7.19	–0.06	–0.05	–0.09
38	7.15	7.29	7.06	7.35	0.14	–0.09	–0.20

39	7.52	7.39	7.10	7.37	–0.13	–0.42	0.15
40	4.35	4.37	4.42	4.50	0.02	0.07	–0.15
41	4.48	4.57	4.51	4.61	0.09	0.03	–0.13
42	5.27	5.54	5.50	5.60	0.28	0.23	–0.33
43	5.47	5.67	5.55	5.68	0.20	0.08	–0.21
44	5.59	5.28	5.30	5.41	–0.31	–0.29	0.18
45	5.60	5.61	5.51	5.67	0.01	–0.09	–0.07
46	5.74	5.98	6.05	5.92	0.24	0.31	–0.18
47	6.22	6.34	6.19	6.40	0.12	–0.03	–0.18
48	6.40	6.30	6.22	6.57	–0.10	–0.18	–0.17
49	6.70	6.71	7.05	6.76	0.01	0.35	–0.06
50	7.05	7.13	7.06	7.06	0.08	0.01	–0.01
51	7.15	7.24	7.13	7.24	0.09	–0.02	–0.09
52[a]	4.77	5.95	5.93	5.88	1.18	1.16	–1.11
53[a]	5.31	6.37	5.94	6.47	1.06	0.63	–1.16
54[a]	6.92	6.11	5.94	6.20	–0.81	–0.98	0.72
55[a]	5.92	5.29	6.37	5.45	–0.63	0.45	0.47

[a]Data points not included in the equation.

Table 4. Observed and calculated anti-HIV-1 activity (pIC50) of dihydroalkoxybenzyloxopyrimidines (DABO) derivatives (test set).

S.No.	pIC_{50} Obs.	pIC_{50} Calc MLR	pIC_{50} Calc k-NN	pIC_{50} Calc NN	Residual MLR	Residual k-NN	Residual NN
56	4.23	4.79	5.14	5.07	0.57	0.91	–0.84
57	4.59	5.12	6.63	5.25	0.53	2.04	–0.67
58	4.79	4.31	5.15	4.27	–0.48	0.36	0.52
59	5.07	5.48	5.38	5.56	0.41	0.31	–0.50
60	5.32	5.61	5.49	5.66	0.29	0.17	–0.35
61	5.37	5.88	5.45	5.73	0.51	0.08	–0.37
62	5.44	5.96	5.77	6.03	0.52	0.33	–0.59
63	5.49	5.95	6.06	6.02	0.46	0.57	–0.54
64	5.54	5.04	5.30	5.19	–0.50	–0.24	0.35
65	5.60	5.11	5.48	5.37	–0.49	–0.12	0.23
66	5.80	6.50	5.94	6.26	0.70	0.14	–0.46
67	6.10	6.61	6.91	6.66	0.52	0.81	–0.57
68	6.10	5.79	6.66	6.20	–0.31	0.56	–0.11
69	6.22	5.72	6.42	6.12	–0.50	0.20	0.09
70	6.70	7.03	6.66	7.02	0.33	-0.04	–0.32
71	6.70	6.03	6.08	6.23	–0.67	-0.62	0.47
72	7.30	6.76	6.65	6.94	–0.54	-0.65	0.36
73	5.62	5.22	5.27	5.39	–0.40	-0.35	0.23
74	7.30	6.86	7.01	6.99	–0.44	-0.29	0.30
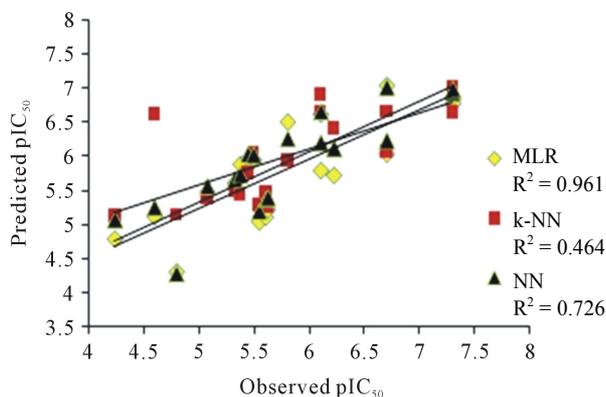
Figure 2. Graph between observed and calculated pIC$_{50}$ values for test set.

0.645) was lower in comparison to the best non-linear model developed using NN ($r^2 = 0.726$) but that of k-NN non-linear model gave very low value of correlation coefficient ($r^2 = 0.464$). Also the degree of relatedness could be visualized by the difference between the observed and predicted values of pIC$_{50}$ *i.e.* from the residual values. Thus, the above results account for some non-linear relationship between the pIC$_{50}$ and E-State Indices.

4. CONCLUSIONS

From the above study related to Dihydroalkoxybenzyloxopyrimidines derivatives following conclusions can be drawn: The 2D-QSAR studies indicated that the anti-HIV-1 activity of DABO derivatives was strongly dependent on the nature of the substituents at N$_1$, N$_3$, C$_2$, C$_4$, C$_5$, and C$_6$ of the pyrimidine ring. Three models developed here viz. MLR, NN and k-NN has different level of robustness for the complete set of compounds. Among the three methods, NN was found to be more successful than the linear model MLR, reflecting that the relationship between descriptors and antiviral activity of DABO derivatives being nonlinear. Though, MLR regression is a simple but powerful method to obtain a subset of significant input variables but it does not account for non-linear relationships while NN plays a significant role depicting non-linearity in building structure activity relationships. Also the degrees of freedom were found to be 40 less than the data points *i.e.* 51 which proves that the model fits the general trend. Our studies have proven the feasibility of the combination of the E-State indices and the NN method to build successful QSAR models.

The most relevant structural conclusions of this study are the following: 1) replacement of pyrimidine nucleus with pyridine nucleus enhance the activity; 2) the presence of bulkiness at the C$_2$ position would lead to higher activity compounds; 3) presence of oxygen group at C$_4$ position in the parent compound is favourable for the activity; 4) DABOs are active with hydrogen in the fifth position comparable with compounds having methyl group.

5. ACKNOWLEDGEMENTS

REFERENCES

[1] Lam, S.N., Acharya, P., Wyatt, R., Kwong, P.D. and Bewley, C.A. (2008) Tyrosine-sulfate isosteres of CCR5 N-terminus as tools for studying HIV-1 entry. *Bioorganic Medicinal Chemistry*, **16**, 10113-10120. doi:10.1016/j.bmc.2008.10.005

[2] AIDS Epidemic Update (2008). http://data.unaids.org/pub/Report/2009/JC1796_Outlook _en.pdf

[3] Adamson, C.S. and Freed, E.O. (2010) Novel approaches to inhibiting HIV-1 replication. *Antiviral Research*, **85**, 119-141. doi:10.1016/j.antiviral.2009.09.009

[4] Sarafianos, S.G., Marchand, B., Das, K., Himmel, D.M., Parniak, M.A., Hughes, S.H. and Arnold, E. (2009) Structure and function of HIV-1 reverse transcriptase: Molecular mechanisms of polymerization and inhibition. *Journal of Molecular Biology*, **385**, 693-713. doi:10.1016/j.jmb.2008.10.071

[5] Sluis-Cremer, N. and Tachedjian, G. (2002) Modulation of the oligomeric structures of HIV-1 retroviral enzymes by synthetic peptides and small molecules. *European Journal of Biochemistry*, **269**, 5103-5111. doi:10.1046/j.1432-1033.2002.03216.x

[6] Sahlberg, C. and Zhou, X.-X. (2008) Development of non-nucleoside reverse transcriptase inhibitors for anti-HIV therapy. *Anti-Infective Agents in Medicinal Chemistry*, **7**, 101-117.

[7] Jäger, J. Smerdon, S.J., Wang, J. Boisvert, D.C. and Steitz, T.A. (1994) Comparison of three different crystal forms shows HIV-1 reverse transcriptase displays an internal swivel motion. *Structure*, **2**, 869-876.

[8] Kopp, E.B., Miglietta, J.J., Shrutkowski, A.G., Shih, C.K., Grob, P.M. and Skoog, M.T. (1991) Steady state kinetics and inhibition of HIV-1 reverse transcriptase by a non-nucleoside dipyridodiazepinone, BI-RG-587, using a heteropolymeric template. *Nucleic Acids Research*, **19**, 3035-3039. doi:10.1093/nar/19.11.3035

[9] Cihlar, T. and Ray, A.S. (2010) Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Research*, **85**, 39-58. doi:10.1016/j.antiviral.2009.09.014

[10] Jochmans, D., Vingerhoets, J., Arnoult, E., Geeraert, L. and Guillemont, J. (2009) Antiviral research: Strategies in antiviral drug discovery. Chapter 3, human immunodeficiency virus type 1 non-nucleoside reverse transcriptase inhibitors. ASM Press, Washington, 33-50.

[11] Spence, R.A., Kati, W.M., Anderson, K.S. and Johnson,

K.A. (1995) Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Science*, **267**, 988-993. doi:10.1126/science.7532321

[12] Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A. and Steitz, T.A. (1992) Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science*, **256**, 1783-1790. doi:10.1126/science.1377403

[13] Tambuyzer, L., Azijn, H., Rimsky, L.T., Vingerhoets, J., Lecocq, P., Kraus, G., Picchio, G. and De Bethune, M.P. (2009) Compilation and prevalence of mutations associated with resistance to non-nucleoside reverse transcriptase inhibitors. *Antiviral Therapy*, **14**, 103-109.

[14] Baba, M., Tanaka, H., De Clercq, E., Pauwels, R., Balzarini, J., Schols, D., Nakashima, H., Perno, C.F., Walker, R. T. and Miyasaka, T. (1989) Highly specific inhibition of human immunodeficiency virus type 1 by a novel 6-substituted acyclouridine derivative. *Biochemical and Biophysical Research Communities*, **165**, 1375-1381. doi:10.1016/0006-291X(89)92756-3

[15] Miyasaka, T., Tanaka, H., Baba, M., Hayakawa, H., Walker, R.T., Balzarini, J. and De Clercq, E. (1989) A novel lead for specific anti-HIV-1 agents: 1-[(2-hydroxy ethoxy)methyl]-6-(phenylthio)thymine. *Journal of Medicinal Chemistry*, **32**, 2507-2509. doi:10.1021/jm00132a002

[16] Debyser, Z., Pauwels, R., Andries, K., Desmyter, J., Kukla, M., Janssen, P.A. and De Clercq, E. (1991) An antiviral target on reverse transcriptase of human immunodeficiency virus type 1 revealed by tetrahydroimidazo-[4,5,1-jk][1,4]benzodiazepin-2(1H)-one and -thione derivatives. *Proceedings of National Academy of Sciences USA*, **88**, 1451-1455. doi:10.1073/pnas.88.4.1451

[17] Pauwels, R., Andries, K., Desmyter, J., Schols, D., Kukla, M.J., Breslin, H.J., Raeymaeckers, A., Van Gelder, J., Woestenborghs, R., Heykants, J., Schellekens, K., Janssen, M.A.C., De Clercq E., Janssen, P.A.J. (1990) Potent and selective inhibition of HIV-1 replication in vitro by a novel series of TIBO derivatives. *Nature*, **343**, 470-474. doi:10.1038/343470a0

[18] De Bethune, M.P. (2010) Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: A review of the last 20 years (1989-2009). *Antiviral Research*, **85**, 75-90. doi:10.1016/j.antiviral.2009.09.008

[19] Dickinson, L., Khoo, S. and Back, D. (2010) Pharmacokinetics and drug-drug interactions of anti-retrovirals: An update. *Antiviral Research*, **85**, 176-189. doi:10.1016/j.antiviral.2009.07.017

[20] Sapre, N.S., Gupta, S., Pancholi, N. and Sapre, N. (2009) A group center overlap based approach for "3D QSAR" studies on TIBO derivatives. *Journal of Computational Chemistry*, **30**, 922-933. doi:10.1002/jcc.21114

[21] Sapre, N.S., Pancholi, N., Gupta, S., Sikarwar, A. and Sapre, N. (2007) Molecular Modelling studies on 2-Amino 6-aryl-Sulphonylbenzonitriles as non-nucleoside reverse transcriptase inhibitors of HIV-1: A QSPR approach. *Journal of Chemical Sciences*, **119**, 625-630. doi:10.1007/s12039-007-0078-5

[22] Sapre, N.S., Gupta, S., Pancholi, N. and Sapre, N. (2008) Molecular docking studies on tetrahydroimidazo-[4,5,1-jk]

[1,4]-benzodiazepinone (TIBO) derivatives as HIV-1 NNRT inhibitors. *Journal of Computer Aided Molecular Design*, **22**, 69-80. doi:10.1007/s10822-007-9161-8

[23] Sapre, N.S., Gupta, S. and Sapre, N. (2008) Assessing ligand efficiencies using template-based molecular docking and tabu-clustering on tetrahydroimidazo-[4,5,1-jk] [1,4]-benzodiazepin-2(1H)-one-thione (TIBO) derivatives as HIV-1RT inhibitors. *Journal of Chemical Science*, **120**, 395-404. doi:10.1007/s12039-008-0063-7

[24] Botta, M., Artico, M., Massa, S., Gambacorta, A., Marongiu, M.E., Pani, A. and La Colla, P. (1992) Synthesis, antimicrobial and antiviral activities of isotrimethoprim and some related derivatives. *European Journal of Medicinal Chemistry*, **27**, 251-257. doi:10.1016/0223-5234(92)90009-P

[25] Costi, R., Di Santo, R., Artico, M., Massa, S., Lavecchia, A., Marceddu, T., Sanna, L., La Colla, P. and Marongiu, M.E. (2000) Structure-activity relationship studies on potential non-nucleoside DABO-like inhibitors of HIV-1 reverse transcriptase. *Antiviral Chemistry and Chemotherapy*, **11**, 117-133.

[26] Sbardella, G., Mai, A., Artico, M., Chimenti, P., Massa, S., Loddo, R., Marongiu, M.E., La Colla, P. and Pani, A. (2001) Structure-activity relationship studies on new DABOS: Effect of substitutions at pyrimidine C-5 and C-6 positions on anti-HIV-1 activity. *Antiviral Chemistry and Chemotherapy*, **12**, 37-50.

[27] Yin, L.-Q., Yu, S.-W., Yao, L.-F., He, Y.-P. and Xie, X.-G., (2008) QSAR studies on 6-(1-naphthylmethyl) substituted S-DABO derivatives as novel non-nucleoside HIV-1 reverse transcriptase inhibitors. *Chinese Journal of Structural Chemistry*, **27**, 1214-1222.

[28] De Brito, M.A., Rodrigues, C.R., Cirino, J.J., De Alencastro, R.B., Castro, H.C. and Albuquerque, M.G. (2008) 3D-QSAR CoMFA of a series of DABO derivatives as HIV-1 reverse transcriptase non-nucleoside inhibitors. *Journal of Chemical Information and Modelling*, **48**, 1706-1715. doi:10.1021/ci8001217

[29] Sudbeck, E.A., Mao, C., Vig, R., Venkatachalam, T.K., Tuel-Ahlgren, L. and Uckun, F.M. (1998) Structure-based design of novel dihydroalkoxybenzyloxopyrimidine derivatives as potent non-nucleoside inhibitors of the human immunodeficiency virus reverse transcriptase. *Antimicrobial Agents and Chemotherapy*, **42**, 3225-3233.

[30] Mai, A., Sbardella, G., Artico, M., Ragno, R., Massa, S., Novellino, E., Greco, G., Lavecchia, A., Musiu, C., La Colla, M., Murgioni, C., La Colla, P. and Loddo, R., (2001) Structure-based design, synthesis, and biological evaluation of conformationally restricted novel 2-alkyl thio-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkyl-pyrimidin-4 (3H)-ones as non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Journal of Medicinal Chemistry*, **44**, 2544-2554. doi:10.1021/jm010853h

[31] Tramontano, E., Marongiu, M.E., de Montis, A., Loi, A. G., Artico, M., Massa, S., Mai, A. and La Colla, P. (1994) Characterization of the anti-HIV-1 activity of 3,4-dihydro-2-alkoxy-6-benzyl-4-oxopyrimidines (DABOs), new non-nucleo side reverse transcriptase inhibitors. *New Microbiologica*, **17**, 269-279.

[32] Mai, A., Artico, M. Ragno, R., Sbardella, G., Massa, S., Musiu, C., Mura, M., Marturana, F., Cadeddu, A., Maga, G., and La Colla, P. (2005) 5-Alkyl-2-alkyl amino-6-

(2,6-difluorophenylalkyl)-3,4-di hydro pyrimidin-4(3H)-ones, a new series of potent, broad-spectrum non-nucleoside reverse transcriptase inhibitors belonging to the DABO family. *Bioorganic Medicinal Chemistry*, **13**, 2065-2077. doi:10.1016/j.bmc.2005.01.005

[33] Mai, A., Artico, M., Sbardella, G., Massa, S., Loi, A. G., Tramontano, E., Scano, P. and La Colla, P. (1995) Synthesis and anti-HIV-1 activity of thio analogues of Dihydroalkoxybenzyloxopyrimidines. *Journal of Medicinal Chemistry*, **38**, 3258-3263. doi:10.1021/jm00017a010

[34] Mai, A., Artico, M., Sbardella, G., Massa, S., Novellino, E., Greco, G., Loi, A.G., Tramontano, E., Marongiu, M.E. and La Colla, P. (1999) 5-Alkyl-2-(alkylthio)-6-(2,6-di halophenylmethyl)-3,4-dihydropyrimidin-4(3H)-ones: Novel potent and selective dihydro-alkoxy-benzyl-oxopyrimidine derivatives. *Journal of Medicinal Chemistry*, **42**, 619-627. doi:10.1021/jm980260f

[35] Qin, H., Liu, C., Guo, Y., Wang, R., Zhang, J., Ma, L., Zhang, Z., Wang, X., Cui, Y. and Liu, J. (2010) Synthesis and biological evaluation of novel C5 halogen-func- tion- alized S-DABO as potent HIV-1 non-nucleoside reverse transcriptase inhibitors. *Bioorganic Medicinal Chemistry*, **18**, 3231-3237. doi:10.1016/j.bmc.2010.03.025

[36] Deeb, O., Hemmateenejad, B., Jaber, A., Garduno-Juarez, R. and Miri, R. (2007) Effect of the electronic and physicochemical parameters on the carcinogenesis activity of some sulfa drugs using QSAR analysis based on genetic-MLR and genetic-PLS. *Chemosphere*, **67**, 2122-2130. doi:10.1016/j.chemosphere.2006.12.098

[37] Shen, Q., Shi, W.M., Yang, X.P. and Ye, B.X. (2006) Particle swarm algorithm trained neural network for QSAR studies of inhibitors of platelet-derived growth factor receptor phosphorylation. *European Journal of Pharmaceutical Science*, **28**, 369-376. doi:10.1016/j.ejps.2006.04.001

[38] Kier, L.B. and Hall, L.H. (1990) An electrotopological-state index for atoms in molecules. *Pharmaceutical Research*, **7**, 801-807. doi:10.1023/A:1015952613760

[39] Hall, L.H. and Kier, L.B., (1995) Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information. *Journal of Chemical. Information and Computer Science*, **35**, 1039-1045.

[40] Kier, L.B. and Hall, L.H., (1999) Molecular structure description: The electrotopological state. Academic Press, San Diego.

[41] VLife Sciences Technologies: MDS 3.5 (2010) Molecular Design Suite, Pvt Ltd. Pune, India.

[42] E-Calc version (1999) **1.1**.

[43] Sapre, N.S., Pancholi, N. and Gupta, S. (2008) Computational modeling of substitution effect on HIV-1 non-nucleoside reverse transcriptase inhibitors with Kier-Hall electrotopological state (E-State) indices. *Internet Electronic Journal of Molecular Design*, **7**, 55-67.

[44] Sapre, N.S., Pancholi, N., Gupta, S. and Sapre, N. (2008) Computational modeling of tetrahydroimidazo-[4,5,1-jk][1,4]-benzodiazepinone derivatives: An atomistic drug design approach using Kier-Hall electrotopological state (E-State) indices. *Journal of Computational Chemistry*, **29**, 1699-1706. doi:10.1002/jcc.20931

[45] Wang, H.Y., Li, Y., Ding, J., Wang, Y. and Chang, Y.Q. (2008) Prediction of binding affinity for estrogen receptor alpha modulators using statistical learning approaches. *Journal of Molecular Diversity*, **12**, 93-102.

[46] Ragno, R., Mai, A., Sbardella, G., Artico, M., Massa, S., Musiu, C., Mura, M., Marturana, F., Cadeddu, A. and La Colla, P. (2004) Computer-aided design, synthesis, and anti-HIV-1 activity *in vitro* of 2-alkylamino-6-[1-(2,6-di fluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as novel potent non-nucleoside reverse transcriptase inhibitors, also active against the Y181C variant. *Journal of Medicinal Chemistry*, **47**, 928-934. doi:10.1021/jm0309856

[47] Sun, G.F., Kuang, Y.Y., Chen, F.E., De Clercq, E., Balzarini, J. and Pannecouque, C. (2005) Non-nucleoside HIV reverse transcriptase inhibitors, Part 6[1]: Synthesis and anti-HIV activity of novel 2-[(arylcarbonylmethyl) thio]-6-arylthio DABO analogues. *Archive der Pharmazie*, **338**, 457-461. doi:10.1002/ardp.200400961

[48] Petricci, E., Mugnaini, C., Radi, M., Togninelli, A., Bernardini, C., Manetti, F., Parlato, M.C., Renzulli, M.L., Alongi, M., Falciani, C., Corelli, F. and Botta, M., (2006) Towards new methodologies for the synthesis of biologically interesting 6-substituted pyrimidines and 4(3H)-pyrimidinones. *Arkivoc*, **7**, 452-478.

[49] Petricci, E., Botta, M., Corelli, F. and Mugnaini, C., (2002) An improved synthesis of solid-supported reagents (SSRs) for selective acylation of amines by microwave irradiation. *Tetrahedron Letter*, **43**, 6507-6509. doi:10.1016/S0040-4039(02)01478-8

[50] Guha, R. and Jurs, P.C. (2005) Interpreting computational neural network QSAR models: A measure of descriptor importance. *Journal of Chemical Information and Modelling*, **45**, 800-806. doi:10.1021/ci050022a

[51] McCulloch, W.S. and Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, **5**, 115-133.

[52] Hoskins, J.C. and Himmelbau, D.M. (1988) Artificial neural network models of knowledge representation in chemical engineering. *Computers and Chemical Engineering*, **12**, 881-890. doi:10.1016/0098-1354(88)87015-7

[53] Cover, T. and Hart, P. (1967) Nearest neighbour pattern classification. *IEEE Transactions Information Theory*, **13**, 21-27. doi:10.1109/TIT.1967.1053964

[54] Hand, D., Mannila, H. and Smyth, P. (2001) Principles of data mining. The MIT Press, Cambridge.