

RSC Advances



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. This *Accepted Manuscript* will be replaced by the edited, formatted and paginated article as soon as this is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

In-silico de-novo Design of Novel NNRTIs: A Bio-Molecular Modelling Approach

Nilanjana Jain (Pancholi)^a, Swagata Gupta^b, Neelima Sapre^c and Nitin S. Sapre^{a,*}

Received (in XXX, XXX) Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXX 20XX

DOI: 10.1039/b000000x

ABSTRACT Six novel Non-Nucleoside Reverse Transcriptase Inhibitors exhibiting high efficacy are designed using *in-silico* mathematical modelling techniques and the results are validated using docking technique. An *in-silico* assessment of interaction potential and structural requirements of 5-Alkyl-2-alkylamino-6-(2,6-difluorophenylalkyl)-3,4-dihydropyrimidin-4(3H)-one (DABO) analogues in the non-nucleoside inhibitor binding pocket is also performed. Efficient use of 3D-Pharmacophoric (S_{ALL}, HD_{ALL}, HA_{ALL} and R_{ALL}) and 3D-averaged alignment (ClogP and dipole moment) descriptors is made in this study. The chemometric analyses, using Support Vector Machine, Back Propagation Neural Network and Multiple Linear Regression, are performed. The relative potentials of these chemometric methods is also assessed and the results, SVM ($r = 0.939$, $MSE = 0.071$, $q^2 = 0.876$), BPNN ($r = 0.923$, $MSE = 0.104$, $q^2 = 0.818$) and MLR ($r = 0.912$, $MSE = 0.096$, $q^2 = 0.832$), indicates that SVM describes the relationship between the descriptors and inhibitory activity in a better manner. The results also suggest that there is a non-linear relationship between the descriptors and inhibitory activity. The study further suggests that isopropyl/propenyl groups as R and R', oxobutyl group as X and di or tri-substitution as R" are the best suited substituents for exhibiting better inhibitory activity.

INTRODUCTION

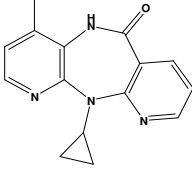
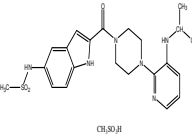
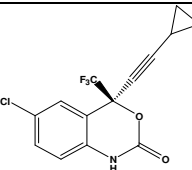
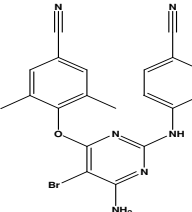
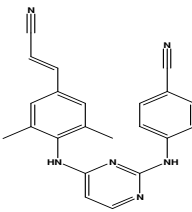
Extensive research is going on to develop a cure for Human immunodeficiency virus type-1 (HIV-1) infection, one of the fatal viruses.¹⁻⁴ The non-nucleoside reverse transcriptase inhibitors (NNRTIs) are prominent members of current combinatorial drug therapy against HIV-1 infection which exhibit significant potency and are relatively less toxic.⁵⁻⁷ However, the rapid manifestation of these into drug-resistant viral strains has relegated the therapeutic efficacy of the inhibitors.⁸⁻¹⁰ Recent advances in ligand-based and structure-based drug design (LBDD and SBDD) approaches, coupled with virtual screening, are robust tools for design of newer compounds acting against HIV-1 infection.¹¹⁻¹⁵

The reverse transcriptase (RT) enzyme is essential for the conversion of genetic RNA into DNA and thus plays a significant role in drug discovery pipeline to combat HIV-1 infection.^{16,17}

The RT enzyme is a heterodimer made of p66 and p51 subunits, where each subunit contains thumb, palm and finger domains. The p66 subunit contains the functional active site that binds the nucleic acid template primer to the nucleotide triphosphate.^{18,19} NNRTIs are highly potent, moderately toxic than the nucleoside reverse transcriptase inhibitors (NRTIs) and do not require cellular activation to inhibit the HIV-1 RT.²⁰ These are non-competitive inhibitors that bind to a hydrophobic "pocket" in the p66 subunit of HIV-1 RT, approximately 10 Å away from the polymerase binding site.²¹ X-ray crystallographic studies have revealed the prominent interactions of the inhibitor within the

non-nucleoside inhibitor binding pocket (NNIBP) of the protein and have facilitated design of more effective inhibitors.²² The NNIBP does not exist in the unliganded RT and is produced by binding of the ligand with the side chains of aromatic (including Y181 and Y188) and hydrophobic amino-acid residues of the viral protein.^{23,24} NNRTI resistance mutations influence the binding of the inhibitors to their binding pocket either by changing the size, shape and polarity of the NNIBP or affecting the entry of NNRTIs into this pocket.²⁵ Among the various FDA-approved NNRTIs, Nevirapine (Viramune/Viramune XR) is a highly effective inhibitor which has emerged as a key drug for the prevention of vertical transmission. Recent studies have suggested that nevirapine is more effective in crossing the blood-brain barrier.²⁶ Another NNRTI, Delavirdine, which is bulkier in size than nevirapine, creates better interactions with RT, viz. hydrogen bond interactions with K103 and hydrophobic interactions with P236.²⁷ Efavirenz (Sustiva, Stocrin, EFV, DMP-266) is a potent NNRTI that binds to HIV-1 RT at a site distinct from the polymerase catalytic site which has also been found to be effective when combined with either nevirapine, nelfinavir or indinavir.^{28,29} Etravirine (ETR/TMC125), a second-generation NNRTI, has exhibited an enhanced barrier to resistance and is found to be extremely effective in achieving the viral suppression as well as improving the immunity in treatment-experienced HIV-infected patients.³⁰ Among the newly discovered NNRTIs, rilpivirine is an antiretroviral exhibiting better bioavailability, easier formulation and administration compared to etravirine³¹⁻³² and the FDA approved NNRTIs are given in Table 1.

Table 1 : Chronological approval status of NNRTIs

S.No	Name	Structure	Approval Status
1.	Nevirapine(BI-RG-587)/ Viramune/Viramune XR 11-cyclopropyl-5,11-dihydro-4-methyl-6H-dipyrido [3,2-b:2',3'-e][1,4] diazepin-6-one		Approved in 1996/ Extended Release 2011
2.	Delavirdine /DLV/Rescriptor N-[2-((4-[3-(propan-2-ylamino)pyridin-2-yl]piperazin-1-yl)carbonyl)-1H-indol-5-yl]methanesulfonamide		Approved in 1997
3.	Efavirenz(DMP 266)/Stocrin™/Sustiva™ (S)-6-chloro-(cyclopropylethynyl)-1,4-dihydro-4-(trifluoromethyl)-2H-3,1-benzoxazin-2-one.		Approved in 1998
4.	R165335/ Etravirine/ TMC125/ Intelligence™ 4-[[6-amino-5-bromo-2-[(4-cyanophenyl)amino]-4-pyrimidinyl]oxy]-3,5-dimethylbenzonitrile		Approved in 2008
5.	R278474/ rilpivirine/TMC 278/ Edurant 4-[[4-[(1E)-2-cyanoethenyl]-2,6-dimethylphenyl]amino]-2-pyrimidinyl]amino]benzo-nitrile		Approved in 2011

<http://aidsinfo.nih.gov/education-materials/fact-sheets/21/58/fda-approved-hiv-medicines>, updated 28, November 2014.

ATP analogs are also reported to be used for inhibition of HIV-transcription.³³ The latest entrant, undergoing clinical trials, in the armamentarium of anti-retrovirals is doravirine (MK-1439) which has robust antiviral activity and better tolerability.³⁴ The 5-Alkyl-2-alkylamino-6-(2, 6-difluorophenylalkyl)-3, 4-dihydropyrimidin-4(3H)-one (DABO) analogs effectively inhibit

the replication of variety of HIV-1 strains at the reverse transcriptase step.³⁵ Efforts are going on to make structural changes so that the DABO analogs demonstrate enhanced potency. Derivatives, belonging to rationally designed broad spectrum NNRTIs, such as Dihydroalkoxy benzyloxypyrimidines (O-DABOs), Dihydroalkylthiobenzyloxypyrimidines (S-DABOs), Dihydroalkylamino difluorobenzyloxypyrimidines (NH-DABOs), N,N-disubstituted amino(2,6-difluorophenylalkyl)-3,4-dihydropyrimidin-4(3H)-ones (F2-N,N-DABOs), Dihydroalkyl thio naphthyl methyloxypyrimidines (DATNOS), are synthesized.³⁶⁻⁴²

The common structure of DABO consists of a pyrimidinone ring and a di-fluoro substituted aromatic ring attached through a substituted CH bridging group. The three substituents R and X are attached to the pyrimidinone ring while R' is attached to the bridging CH group. Earlier studies on DABO are based on various SAR analyses, where the structural requirements for enhancing the biological activity have been quantitatively analyzed.⁴³⁻⁴⁶ The 3-dimensional RT complex-DABO crystal structure analysis has provided newer dimensions to interpretation of the drug-receptor interaction profile and this has certainly aided in substantiating the SAR analyses for enhanced structural refinement.⁴⁷⁻⁵²

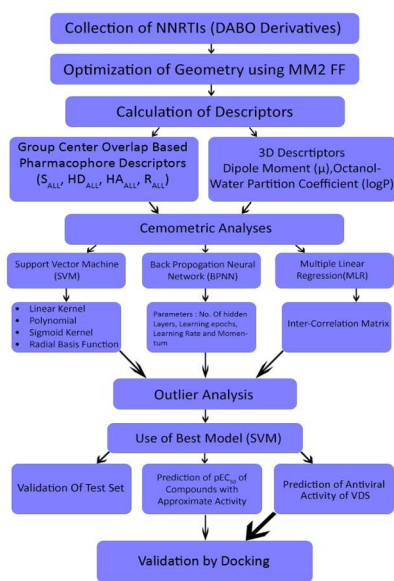
The present study deals with design of novel NNRTIs by performing chemometric analyses of two important types of descriptors namely (a) 3D pharmacophoric [dipole moment, S_{ALL} , HD_{ALL} , R_{ALL} and HA_{ALL}] and (b) 2D average alignment [octanol/water partition coefficient (ClogP)] descriptors in understanding the interactions potential of 5-Alkyl-2-alkylamino-6-(2, 6-difluorophenylalkyl)-3, 4-dihydropyrimidin-4(3H)-one (DABO) analogs in the NNIBP.⁵³ It is reported that docking templates can be very useful if 3D structural conformation of a ligand is evident.⁵⁴ These templates basically accounts for vital contribution of the pharmacophores in the structural alignment of the ligand within the NNIBP. These pharmacophore based templates assist in extracting requisite information based on similarity overlap for each individual template point. During template alignment procedure, ligand flexibility is also considered. The poses similar to reference ligand are searched by the docking engine guided by pharmacophoric template groups. The validation procedure focuses on atomic overlap of each template group centre with the concerned ligand.

The Gaussian function is used to evaluate an atom, matching a group definition based on its distance from the group centres. The four template groups incorporated in the present work are steric (S_{ALL}), hydrogen donor (HD_{ALL}), hydrogen acceptor (HA_{ALL}) and ring (R_{ALL}). Also, two extremely relevant descriptors, octanol/water partition coefficient (ClogP) (an 'alignment-average' descriptor), and dipole moment (μ), a 3D parameter, are included for the sake of better assessment of anti-viral activity in the non-nucleoside inhibitor binding pocket (NNIBP).⁵⁵ Optimization of physicochemical properties such as lipophilicity and increase ligand-lipophilicity efficiency (LLE) is reported by Mowbray et al.⁵⁶ These descriptors are selected to characterize the molecular architecture for their capability to correlate diverse biochemical phenomena of the concerned molecular series. The obtained relevant information is assessed chemometrically to

correlate the structural features with the anti-viral activity of the DABO derivatives.⁵⁷

The chemometric models have established themselves soundly in quantifying the correlation between selected molecular aspects and their impact on the biological response.⁵⁸⁻⁶¹ The development of these rational SAR models focus on necessary chemical features leading to better pharmaco-toxico-kinetic profile of a lead candidate curtailing irrelevant experimental determinations.^{62,63} These statistical validation methodologies form the basis of SAR studies, though there are still some limitations.⁶⁴ The quantitative relationships between the molecular entity and its physicochemical and biological properties appear to be rather more complex and nonlinearity also prevails in many cases, thus extracting pivotal information from a SAR model is an aspect that needs to be focused very seriously.^{65,66}

In the present study, non-linear (BPNN [back propagation neural network] and SVM [support vector machine]) and Linear (MLR [multiple linear regression]) techniques are used for chemometric assessment of the training dataset. The results are validated using a test set. Finally, an external dataset (compounds with approximate inhibitory activity) and an *in-silico* designed virtual dataset of molecules is assessed and reported with their predicted activities. External and structural validation of the designed *in-silico* ligands of VDS is performed using docking technique. The complete work performed is presented in scheme 1.



Scheme 1: A scheme presenting the flow of the work.

MATERIALS AND METHODS

MOLECULAR STRUCTURAL DATA SET

The structures of 53-DABO analogues are drawn using ChemDraw Ultra 7.0⁶⁷ and optimized using MM2 force field. The pharmacophoric 3D descriptors based on group centre overlap are calculated using Molegro Virtual Docker (MVD)⁶⁸, while ClogP and dipole moment are calculated using Sybyl -X 2.0 suite

software.⁶⁹

METHODOLOGY

CONFORMATIONAL ANALYSIS: Conformational search for the all the DABO derivatives (36-Training, 9-Test, 1-outlier, 7-Appx., 6-VDS) has been performed using Sybyl-X 2.0 suite software⁶⁹. Random search method is used for performing conformational analysis as which is suggested for complex cases. The randomization is involved with the Cartesian as well as internal coordinates. The method locates energy minima of a molecule by randomly adjusting the selected bonds and minimizing the energy of the resulting geometry. It involves making random torsion changes to selected bonds, followed by minimization. The parameters used for performing random conformational sampling are: Maximum Cycles: 200, Energy cut-off: 10Kcal/mol., RMS threshold 0.10, Convergence threshold: 0.05, Maximum hits: 6, Force Field: Tripos (using default setting), Charges: Gasteiger-Marsili and checking the symmetry of conformer. The details of structure of compounds, conformers and energy are given separately (in supplementary information).

GENERATION OF 3D DESCRIPTORS: 3D descriptors (hydrogen donor, hydrogen acceptor, steric and ring) based on group centre overlap are generated using the known 3D conformation of highest active ligand. A template represents a collection of specific chemical features associated with an atom (in a molecule) crucial for binding interactions. The molecules can thus be aligned with the template. Based on the similarity overlap, crucial information can be deduced.

The Gaussian formula, given in Eq. 1, is used to determine the amount of overlap for the specific group centre:

$$e = \omega * \exp\left(\frac{-d^2}{r_0^2}\right) \quad (1)$$

where, d is the distance from the position of the atom to the centre in the group, ω is a weight factor for the template group and r_0 is a distance parameter.^{70,54}

Dipole Moment (μ). Presence of polarity and its magnitude in a molecule are crucial parameters in determination of specific binding interactions within the binding pocket of an enzyme. Dipole moment is a 3D electrostatic descriptor represented by the vector μ and reported in Debye units (D).^{71,72}

$$\mu = -\sum_{i=1}^{occ} \int \phi_i \hat{r} \phi_i dv + \sum_{a=1}^M Z_a \bar{R}_a \quad (2)$$

where, ϕ_i is molecular orbitals, \hat{r} is electron position operator, Z_a is a-th atomic nuclear charge and \bar{R}_a is position vector of a-th atomic nucleus.

Octanol/water Partition Coefficient (ClogP). Octanol-water partition coefficient is the ratio of equilibrium concentration of solute in a non-polar solvent to the concentration of same solute in a polar solvent. The logarithm of partition coefficient, logP, has been accounted as hydrophobic parameter in “extra-thermodynamic” Hammett methodology. Generally 1-octanol is suggested in the non-polar phase. The logP (Oct/water) from structure based on a substitution method is calculated as given in the equation Eq. 2.^{73,74}

$$\pi_x = \log P(C_6H_5X) - \log P(C_6H_5) \quad (3)$$

Rekker used a reductionist approach, to derive the constants for carbon, hydrogen and for polar fragments.⁷⁵ The fragment values (f) and interaction factors (F) had to be identified and evaluated by the relation reported by Hansch.⁷⁶

$$\log P = \sum_{i=1,N} a_i f_i + \sum_{j=1,N} b_j f_j \quad (4)$$

Chou and Jurs used a constructionist approach to calculate logP, considering the hydrophobic portions of solutes as “hydrocarbon-like” and defining these carbon and hydrogen fragment values as being truly constant and is reported as ClogP.⁷⁷

Chemometric validation Methods. Till date a variety of chemometric methods have been developed and used to handle multivariate data analyses, on which lies the onus of reliable QSAR interpretations.⁷⁸⁻⁸⁰ The robustness of a derived model is extremely important while validating the utility of descriptors in deducing the pharmacophoric qualities of the inhibitors. Three techniques namely SVM, BPNN and MLR are used in the present study. A brief account of these methods is presented herewith.

Support Vector Machine (SVM). SVM is based on the structural risk minimization (SRM) principle which is least sensitive to data over fitting.⁸¹ This method can be applied to linear as well as nonlinear classification and trained faster.⁸² SVM has been successful in correlating various quantitative structure activity/property relationships in the areas of computer-aided drug design methods.⁸³⁻⁸⁷ It is a supervised learning method and support vectors are used with suitable kernel functions.⁸⁸⁻⁹⁰ For the present study v-support vector regression and ε-support vector regression based on LIBSVM are considered and in each case linear, polynomial, sigmoid, and radial basis functions are used.

Back Propagation Neural Network (BPNN). Neural networks resemble human brain neuron network and can handle complex and non-linear data and thus extract the hidden relationships between the dependent and independent variables.⁹¹ Rumelhart et al.⁹² developed the back-propagation neural network (BPNN) as a solution to the problem of training multi-layer perceptrons.

The molecular descriptors are encoded in the form of input neurons, which are multiplied by weight of each neuron. A sigmoid non-linear transfer function is then applied and a suitable bias is applied for shifting the transfer function to either side. These are then multiplied to the output weights, transformed and interpreted.⁹³ The residual error in this supervised learning method i.e. the difference between the experimental quantity and network's predicted quantity is evaluated. This calculated error is allowed to propagate backward through the network and the weights are accordingly adjusted so as to observe the same input pattern and minimize the residual error. This pattern is repeated till a relationship or no relationship is derived.⁹⁴⁻⁹⁷ The most crucial characteristic of a neural network setup is deciding the number of neurons within the hidden layer and is decided by the ratio rho (ρ). Neural network models are very efficient in handling and extracting non-linear relationships. The ratio is maintained within the range 1.0 - 3.0 to get sensible results.^{98,99}

Multiple linear Regressions (MLR). Multiple linear regression

(MLR) is a method where the values of the regression coefficients (b_n 's) are evaluated using least squares curve fitting method.¹⁰⁰⁻¹⁰¹

$$y = b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n + c$$

Where, 'y' is the dependent variable, ' $x_1, x_2 \dots x_n$ ' are the independent variables, ' $b_1, b_2 \dots b_n$ ' are the regression coefficients and 'c' is the constant.

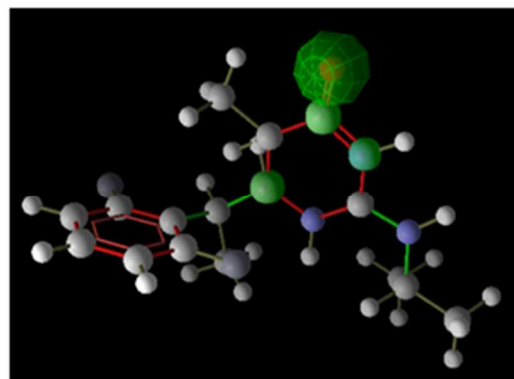
This is the most widely used owing to its fast and easy interpretability. However, for complex systems such as a biological system, the linear combination of descriptor information can often lead to a model with limited accuracy, simply due to the assumption of linearity in the data.

RESULTS AND DISCUSSION

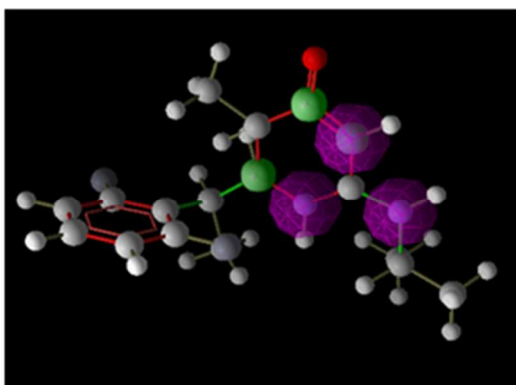
The main aim of the present study is to design novel NNRTIs of HIV-1. 3D pharmacophoric [dipole moment, S_{ALL} , HD_{ALL} , R_{ALL} and HA_{ALL}] and (b) 2D average alignment [octanol/water partition coefficient (ClogP) descriptors are used for extracting their relationship with anti-viral activity using SVM, ANN and MLR techniques. The inhibitory activity of virtually designed compounds is predicted using the derived relationship. The compounds which exhibited activity higher than the highest active reference compound (Cpd. No. 26, $pEC_{50} = 8.30$) are extracted from this virtual dataset.

The ligand selected as reference for template alignment in the docking wizard is compound no. 26 with highest activity ($pEC_{50} = 8.30$) in the dataset. Four template groups are taken for comparing the reference ligand with rest of the ligands of the dataset. These are classified into (i) Steric: consisting of all 22 non-hydrogen atoms and are used for shape matching only, (ii) Hydrogen Donor: consisting of 3 hydrogen donor functional groups, (iii) Hydrogen Acceptor: consisting of 1 hydrogen bond acceptor group and (iv) Ring: constituting 12 atoms which are part of rings. Any atom closer than 2.0 Å to the existing centre of a template group is accepted as equal to that centre and the optimal match corresponds to value of 1.0.

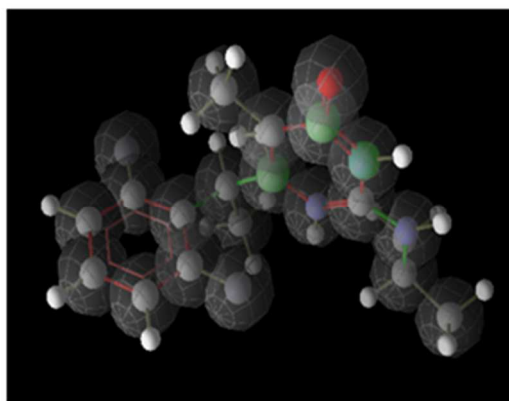
Figure 1:A-D shows all the template groups derived from compound number 26 ($pEC_{50} = 8.30$) and the contribution of each specific atom to a template group is also evident from this figure.



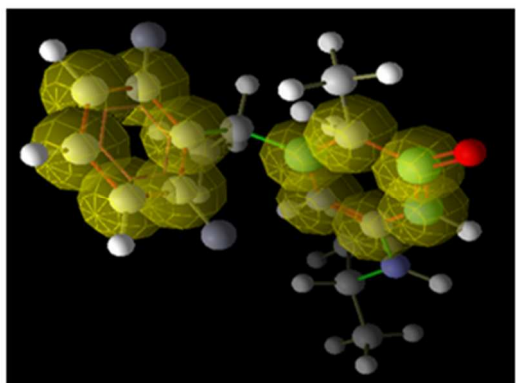
(A) Hydrogen Acceptor (HA) group



(B) Hydrogen Donor (HD) group



(D) Steric (S) group



(C) Ring (R) group

5

Fig. 1: Figure 1-A-D Template groups derived from the reference ligand [Compound no. 26, ($pEC_{50}=8.3010$)] of DABO derivatives.

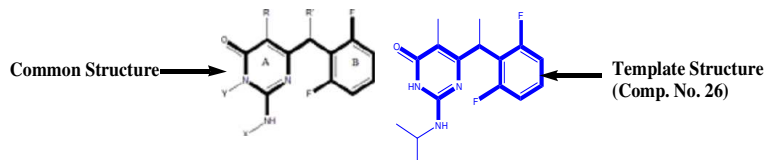
10 A training set of 36 ligands is considered for the 3D similarity-based alignment using template groups and is presented in table 2. The table 2 presents substituents (R, R', X and Y) on the parent structure, experimental and calculated anti-HIV-1 activities (effective concentration pEC_{50}), all the descriptors
 15 namely, octanol/water partition coefficient (ClogP), dipole moment (μ) in debye units, four 3D descriptors based on group centre overlap based approach [steric (S_{ALL}), hydrogen donor (HD_{ALL}), hydrogen acceptor(HA_{ALL}), ring (R_{ALL})]. Table 2 also presents the compounds of the test set and outlier. The structure
 20 shown in blue colour is the highest active ligand used as a template for estimation of 3D pharmacophoric descriptors.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Table 2 Substituents (R, R', X and Y) on the parent structure, Anti-HIV-1 activity (Effective concentration pEC₅₀), octanol/water partition coefficient (ClogP), dipole moment (μ) in debye units, steric (S_{ALL}), hydrogen donor (HD_{ALL}), hydrogen acceptor (HA_{ALL}), ring (R_{ALL}) and calculated pEC₅₀ using MLR, BPNN and SVM of DABO analogues.



S.No.	R	R'	X	Y	Exp-pEC ₅₀	ClogP	S _{ALL}	HD _{ALL}	HA _{ALL}	R _{ALL}	μ	Calculated pEC ₅₀		
												MLR	BPNN	SVM
1	H	H	Iso-Propyl	H	6.398	2.562	0.699	0.758	0.050	0.718	2.768	6.284	6.063	5.400
2	H	H	sec-Butyl	H	6.886	3.091	0.731	0.715	0.161	0.726	3.348	6.712	6.573	5.760
3	H	H	Methoxyethyl	H	6.097	1.754	0.657	0.510	0.045	0.592	4.692	5.881	5.873	5.090
4	H	H	Methylthioethyl	H	6.523	2.472	0.617	0.429	0.035	0.546	4.191	6.027	6.009	5.187
5	H	H	Cyclohexyl	H	6.854	3.755	0.743	0.879	0.378	0.803	3.342	7.339	7.190	7.384
6	H	H	Phenyl	H	6.538	3.703	0.711	0.667	0.337	0.770	3.060	6.762	6.655	5.750
7	Me	H	n-Propyl	H	7.699	3.231	0.846	0.890	0.713	0.868	2.680	7.409	7.317	7.404
8	Me	H	Iso-Propyl	H	7.523	3.011	0.733	0.907	0.215	0.750	3.269	7.058	6.936	7.159
9	Me	H	n-Butyl	H	7.699	3.760	0.669	0.763	0.167	0.658	3.259	7.200	7.057	7.308
10	Me	H	sec-Butyl	H	7.222	3.540	0.706	0.858	0.157	0.657	3.262	7.448	7.285	7.565
11	Me	H	Cyclohexyl	H	7.523	4.204	0.753	0.920	0.384	0.767	3.261	7.848	7.588	7.890
12	Me	H	Phenyl	H	7.222	4.152	0.742	0.748	0.449	0.780	3.000	7.344	7.192	7.311
13	Me	H	2,6-F ₂ -Phenyl	H	6.854	4.488	0.474	0.068	0.490	0.326	1.672	6.619	6.632	5.798
14	Me	H	2-Cl-Phenyl	H	6.886	4.902	0.440	0.064	0.385	0.308	1.799	6.699	6.683	5.929
15	Me	H	3-Cl-Phenyl	H	6.699	4.902	0.463	0.054	0.448	0.322	2.749	6.878	6.772	5.910
16	Me	H	4-Cl-Phenyl	H	7.000	4.902	0.462	0.079	0.487	0.325	3.442	7.038	6.835	5.993
17	Me	H	4-Me-Phenyl	H	6.745	4.651	0.424	0.108	0.497	0.338	2.733	6.691	6.662	5.739
18	Me	H	Phenylmethyl	H	6.301	3.621	0.650	0.524	0.106	0.584	2.699	6.684	6.566	5.733
19	H	Me	Methyl	H	6.959	2.123	0.867	0.940	0.987	0.952	2.591	6.893	6.823	7.044
20	H	Me	n-Propyl	H	7.699	3.181	0.829	0.848	0.745	0.800	2.585	7.511	7.423	7.444
21	H	Me	Iso-Propyl	H	7.523	2.961	0.946	0.964	0.901	0.967	2.550	7.557	7.453	7.723
22	H	Me	n-Butyl	H	7.523	3.710	0.697	0.692	0.477	0.564	2.578	7.663	7.527	7.517
23	H	Me	sec-Butyl	H	7.398	3.490	0.870	0.847	0.718	0.808	2.581	7.775	7.615	7.713
24	Me	Me	Methyl	H	7.523	2.572	0.861	0.892	0.888	0.865	2.694	7.282	7.232	7.277
25	Me	Me	n-Propyl	H	8.222	3.630	0.910	0.936	0.875	0.896	2.684	7.986	7.772	3.035
26	Me	Me	Iso-Propyl	H	8.301	3.410	1.000	1.000	1.000	1.000	2.651	8.041	7.813	3.286
27	Me	Me	n-Butyl	H	8.000	4.159	0.782	0.811	0.646	0.687	2.676	8.146	7.847	3.004
28	Me	Me	Methylthioethyl	H	6.854	3.320	0.735	0.768	0.582	0.636	3.686	7.713	7.517	7.491
29	Me	Me	Cyclohexyl	H	7.585	4.603	0.834	0.685	0.618	0.856	2.708	7.576	7.369	7.557
30	Me	Me	Phenyl	H	7.620	4.551	0.807	0.694	0.503	0.766	2.404	7.705	7.496	7.672
31	H	H	-	Me	4.260	1.149	0.635	0.531	0.123	0.674	2.803	5.017	4.869	4.954
32	H	Me	-	Me	5.620	1.548	0.741	0.484	0.905	0.784	2.092	5.605	5.387	3.640
33	H	H	Cyclopentyl	H	7.046	3.196	0.743	0.870	0.331	0.805	3.356	6.984	6.870	7.033
34	Me	H	Cyclopentyl	H	7.699	3.645	0.758	0.923	0.349	0.775	3.268	7.525	7.359	7.572
35	H	Me	Cyclopentyl	H	7.523	3.595	0.557	0.773	0.232	0.506	3.348	7.386	7.271	7.502
36	Me	Me	Cyclopentyl	H	7.699	4.044	0.601	0.696	0.116	0.509	2.748	7.433	7.291	7.580
^α 37	H	H	Ethyl	H	6.097	2.253	0.692	0.776	0.081	0.738	2.770	6.081	5.830	5.178
^α 38	H	H	n-Propyl	H	6.959	2.782	0.708	0.630	0.106	0.718	2.768	6.181	5.991	5.216
^α 39	H	H	Cyclopropyl	H	5.499	2.078	0.732	0.909	0.287	0.778	3.293	6.520	6.366	5.549
^α 40	H	H	n-Butyl	H	7.000	3.311	0.640	0.471	0.039	0.570	3.328	6.445	6.341	5.536
^α 41	Me	H	Methyl	H	6.398	2.173	0.794	0.818	0.711	0.804	2.689	6.752	6.658	5.655
^α 42	Me	H	Methylthioethyl	H	7.796	2.921	0.669	0.752	0.192	0.665	3.799	6.799	6.694	5.882
^α 43	H	Me	Methylthioethyl	H	7.886	2.871	0.651	0.651	0.423	0.514	3.584	7.242	7.153	7.070
^α 44	H	Me	Cyclohexyl	H	7.658	4.154	0.765	0.640	0.501	0.794	2.616	7.120	6.998	7.077
^α 45	H	H	H	H	3.967	0.912	0.570	0.770	0.016	0.554	2.962	5.622	5.328	5.757
^β 46	H	H	Methyl	H	5.824	1.724	0.416	0.089	0.134	0.367	2.729	4.692	4.942	4.582

^α pEC₅₀ = -logEC₅₀ (where EC₅₀ is the effective concentration of a compound required to activate 50% protection of MT-4 cell against the cytopathic effect of HIV-1). The data points ^α represent the test set and ^β as Outliers.

Correlation Analyses. Various regression models (nonlinear and

linear) are generated to evaluate the behaviour of the descriptors. The uni-variate linear correlation matrix for the correlation of all the descriptors with pEC_{50} and their individual impact coefficient (IC) for DABO analogues is presented in table 3.

Table 3. The uni-variate correlation (R , R^2) and impact (IC) coefficients of 3D and 2D descriptors with pEC_{50} and linear equation for DABO analogues.(Training set)

Descriptor	R	R^2	Impact Coefficient(IC)	Equation
ClogP	0.502	0.252	0.413	$0.413 * \text{clogP} + 5.664$
S_{ALL}	0.426	0.182	2.290	$2.290 * S_{ALL} + 5.469$
HD_{ALL}	0.466	0.217	1.264	$1.264 * HD_{ALL} + 6.250$
HA_{ALL}	0.388	0.151	1.031	$1.031 * HA_{ALL} + 6.631$
R_{ALL}	0.293	0.085	1.178	$1.178 * R_{ALL} + 6.296$
μ	0.107	0.011	-0.141	$-0.141 * \mu + 7.515$

Perusal of the coefficients of the descriptors suggests that ClogP exhibit highest correlation potential while the dipole moment shows the lowest one under linearity conditions. Considering the uni-variate relationship of the descriptors with the antiviral

activity of the training set of the DABO analogues, following order of correlation (R^2) is observed.

$$\text{ClogP} > HD_{ALL} > S_{ALL} > HA_{ALL} > R_{ALL} > \mu$$

While the order of impact (IC) of descriptor follows the following order:

$$S_{ALL} > HD_{ALL} > R_{ALL} > HA_{ALL} > \text{ClogP} > \mu$$

The result of uni-variate linear correlation analyses shows that though ClogP has highest R^2 (0.252) indicating relatively high linear relatedness yet its impact (coefficient of the descriptor =0.413) on the biological response is not the highest. S_{ALL} showed a low potential i.e. a less linear relationship with the biological response ($R^2=0.182$) yet it has highest impact (coefficient of the descriptor =2.290) on the antiviral activity.

Table 4 presents detailed analyses of non-linear and linear chemometric methods used in the present investigation. In table 4, 'k' is the no. of descriptors, ' r^2 ' is the correlation coefficient, ' q^2 ' is cross validated ' r^2 ' from the (LOO) and N-CV procedure, rho (ρ) is the Spearman rank correlation coefficient, MSE is the mean squared error and PRESS is the predictive sum of squares.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Table 4. Comparative analyses of models build by multiple linear regressions (MLR), Back Propagation Neural Network (BPNN) and Support Vector Machine (SVM) techniques for Training set.

S.No	MLR	Model	K	R	r ²	r ² adj	spearman(rho)	PRESS	MSE	q ²
1	MLR	MS	6	0.912	0.832	0.798	0.813	--	0.096	0.832
2		LOO	6	0.850	0.723	0.666	0.773	5.791	0.161	0.719
3		NCV (N=10)	6	0.859	0.739	0.684	0.789	5.465	0.152	0.735
1	BPNN	MS	6	0.923	0.852	--	0.827	--	0.104	0.818
2		LOO	6	0.833	0.693	--	0.749	6.540	0.182	0.683
3		NCV (N=10)	6	0.799	0.639	--	0.727	8.263	0.230	0.600
1	SVM (Epsilon-Radial) 36SV : RBFK	MS	6	0.939	0.883	-	0.849	--	0.071	0.876
2		LOO	6	0.895	0.802	-	0.789	-	0.117	0.796
3		NCV (N=10)	6	0.899	0.809	-	0.785	-	0.112	0.805
1	SVM (Epsilon-Polynomial) 34SV :PK	MS	6	0.332	0.111	--	0.676	--	0.587	-0.023
2		LOO	6	-0.643	0.414	--	-0.674	-	0.589	-0.027
3		NCV (N=10)	6	-0.390	0.157	--	-0.340	-	0.666	-0.161
1	SVM (Epsilon-Sigmoid) 36SV : SK	MS	6	0.908	0.824	-	0.803	--	0.103	0.820
2		LOO	6	0.846	0.716	-	0.751	-	0.170	0.704
3		NCV (N=10)	6	0.818	0.669	-	0.731	-	0.199	0.653
1	SVM (Epsilon-Linear) 36SV : LK	MS	6	0.906	0.820	--	0.800	--	0.105	0.817
2		LOO	6	0.832	0.692	--	0.739	-	0.184	0.678
3		NCV (N=10)	6	0.813	0.660	--	0.710	-	0.203	0.646
1	SVM (Nu-Radial) 22SV : RBFK	MS	6	0.930	0.864	-	0.820	--	0.080	0.861
2		LOO	6	0.872	0.760	-	0.768	-	0.140	0.757
3		NCV (N=10)	6	0.878	0.771	-	0.772	-	0.132	0.770
1	SVM (Nu -Polynomial) 18SV : PK	MS	6	0.406	0.165	--	0.689	--	0.583	-0.016
2		LOO	6	-0.804	0.646	--	-0.791	-	0.615	-0.072
3		NCV (N=10)	6	-0.423	0.179	--	-0.372	-	0.609	-0.062
1	SVM (Nu -Sigmoid) 22SV : SK	MS	6	0.910	0.828	-	0.818	--	0.099	0.828
2		LOO	6	0.884	0.782	-	0.793	-	0.125	0.781
3		NCV (N=10)	6	0.851	0.724	-	0.768	-	0.160	0.721
1	SVM (Nu -Linear) 22SV : LK	MS	6	0.911	0.829	--	0.814	--	0.098	0.829
2		LOO	6	0.876	0.767	--	0.788	-	0.134	0.766
3		NCV (N=10)	6	0.850	0.723	--	0.761	-	0.161	0.719

^a MS = Manual Selection, RBFK = Radial Basis Function Kernel, PK = Polynomial Kernel, SK = Sigmoid Kernel, LK = Linear Kernel

Support Vector Machine (SVM) Analyses. In the present study, ϵ -support vector regression and ν -support vector regression with variable kernels [linear (SVM-LK), polynomial (SVM-PK), sigmoid (SVM-SK), and radial basis function (SVM-RBFK)] are considered and eight models are generated using a random seed 3386003317. Optimal parameter settings are fine-tuned and accordingly good results are obtained. The following parameters, Cost: 100000, Gamma=0.00038, Epsilon (ϵ): 0.001/ Nu (ν): 0.5, Termination criterion tolerance: 0.001 are chosen for performing the analyses. Table 4 also presents all the statistical details of the eight SVM models. It is observed that the radial basis function kernel performs best, followed by sigmoid and linear kernels in ϵ and ν techniques both. In all the cases the correlation coefficients are comparable. In either case the polynomial kernels perform worst. The SVM regression models (ϵ -RBFK and ν -RBFK) are

better (higher correlation coefficients and lower mean squared errors) than the MLR and BPNN regression models. The better results obtained using SVM method can be attributed to non-linearity among the various parameters and also signifies the robustness of the models derived.

Back propagation Neural Network (BPNN) Analyses. In the back-propagation method considered for training the neural network same random seed (3386003317) as in the case of SVM is used. For training the network, following parameters: max training epoch = 733, learning rate = 0.33, output learning rate = 0.44, momentum = 0.20, Neural architecture = (71, 3H, 10) and initial weight (\pm) = 0.20 are considered. The best BPNN model ($r=0.923$), points towards significant non-linearity. The relevance scores of the respective 3D descriptors, as suggested by the best model relating biological activity using BPNN method is presented in Figure 2.

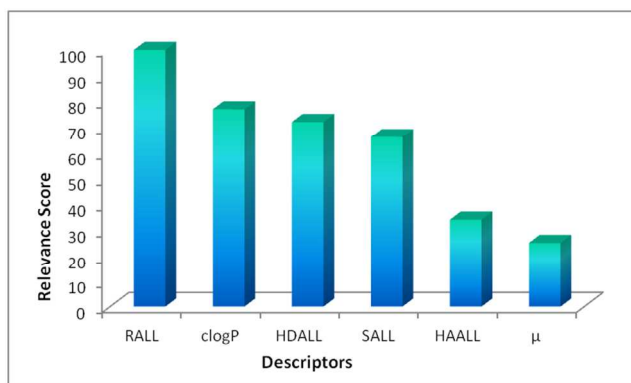


Figure 2. The relevance scores of the 3D descriptors of DABO derivatives as obtained from the BPNN regression analysis.

($n = 36$ $r=0.923$, $r^2 = 0.852$ Spearman (rho) = 0.827 MSE = 0.104)

Of these descriptors, R_{ALL} has the most prominent enhancing effect on the anti-viral activity, followed by the lipophilicity (ClogP). HD_{ALL} shows a better effect on antiviral activity than HA_{ALL} . Leave-one-out (LOO) and N-cross validated (N-CV) methods are used to validate the results. Dipole moment (μ), a 3D electrostatic descriptor, exhibits lowest impact on the antiviral activity.

Multiple Linear Regression (MLR) Analyses. MLR analyses are also performed on data with the same random seed (3386003317) as that in SVM and BPNN. Of the various MLR analyses, backward elimination method gave best model (MLR: $r = 0.912$) and the multivariate relationship of descriptors with antiviral activity is presented herewith in equation 5:

$$pEC_{50} = +0.519(\pm 0.631)C \log P + 3.908(\pm 0.728)S_{ALL} + 2.320(\pm 0.854)HD_{ALL} + 0.789(\pm 0.297)HA_{ALL} - 4.099(\pm 1.018)R_{ALL} + 0.126(\pm 0.096)\mu + 3.018 \quad (5)$$

($n = 36$, $r=0.912$ $r^2 = 0.832$ $r_{adj}^2 = 0.798$ Spearman(rho) = 0.813 MSE = 0.096)

The high positive coefficients of S_{ALL} and HD_{ALL} show that they impart an enhancing effect while a high negative coefficient of R_{ALL} shows that it imparts an adverse effect on anti HIV-1 activity. ClogP, HA_{ALL} and Dipole moment (μ) have comparatively less enhancing effect.

Figure-3 presents graph between experimental and calculated activity (pEC_{50}) derived from the best of MLR, BPNN and SVM models for the training set. A good fit is observed between observed and calculated activity reconfirming the robustness of the methods used.

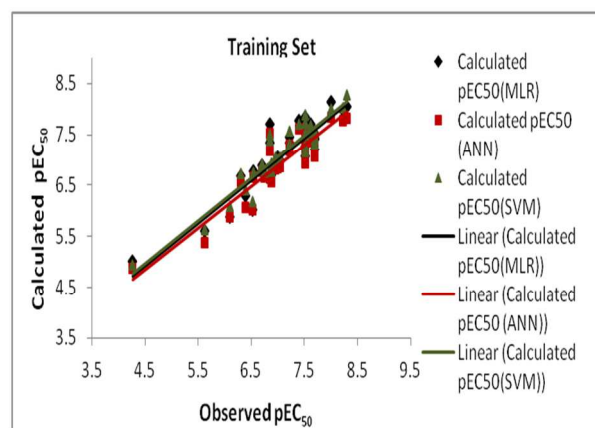


Figure 3. A graph of comparative analyses of observed and calculated pEC_{50} (MLR, BPNN and SVM) values for the training set of DABO derivatives.

Compound no. 46 is observed as an outlier. It has unsubstituted R, bridged R' and methyl moiety present at X, and has a moderate value of antiviral activity.

Cross-validation. A test-set of DABO derivatives with antiviral activity spanning the complete range is considered for external validation of the aforementioned chemometric regression models.

A comparison of the validated results is shown in Figure 4 suggests good agreement between the experimental and estimated activity. The results of regression analyses using all the methods for the test set are found to be highly encouraging and are given herewith: SVM: $r = 0.839$, $r^2 = 0.703$, Spearman (rho) = 0.767, BPNN: $r = 0.822$, $r^2 = 0.676$, Spearman (rho) = 0.800 and MLR: $r = 0.816$, $r^2 = 0.666$, Spearman (rho) = 0.800. In case of training as well as the test set the results of the SVM validation are the best.

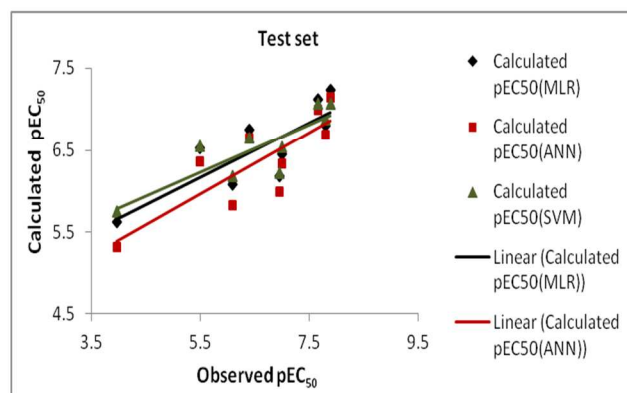


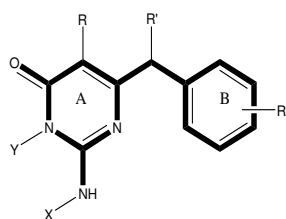
Figure 4. A graph of comparative analyses of observed and calculated pEC_{50} (MLR, BPNN and SVM) values for the test set of DABO derivatives.

It is observed that the steric attributes are important in elucidating the interaction mechanisms in the case of DABO derivatives and the present findings establish the role of steric interactions very well. The hydrogen donor attribute behaves better than the hydrogen acceptor attribute. It is again well evident that the hydrogen donor ability of 3-NH is crucial to the antiviral activity of the DABO derivatives.¹⁰²

Estimation of Anti-Viral Activity for Approximate dataset and Virtual Dataset (VDS). The SVM (ϵ -RBFK), BPNN and

MLR methods are used for prediction of the antiviral activities of the previously synthesized DABOs (whose antiviral activities are reported in approximate terms). Table 5 records the details of such 7 compounds (Comp. Nos. 47-53) and their predicted antiviral activities. It is observed that in certain instances, there is a wide gap between the reported and the predicted activity values. A virtual dataset of 150 compounds, with structure similarity with DABO derivatives, is created and their anti-viral activities are predicted using the best derived SVM (ϵ -RBFK) model and for comparison sake the activities predicted using BPNN and MLR are shown. Care is taken to obtain molecules with not only better activities but also with favourable substitution(s). Finally, six compounds (comp nos. 54-59) with antiviral activity higher than **Table 5**. Descriptors, anti-HIV-1 activity (Approximate) and predicted anti-viral activity (SVM, BPNN and MLR) for the compounds with approximate activity and the virtual data set.

8.25 are extracted, which could be earmarked for synthesis and subsequent development as lead or probable drug derivatives. The results suggest that substitution of di-fluoro groups at 2 and 6 positions in ring 'B' by Chloro and/or Bromo and tri-substitution at 2,4 and 6 position by Br/Cl/F, smaller modification in R and X, yielded better inhibitory activity (pEC_{50}). Therefore, compounds designed *in-silico* may reduce the time frame in the search for better drug like candidates. Table-4 records the six structures which are virtually generated using the DABO like template. The template used for deriving the virtual data set is also given in the Table 5. Table 5 also records all the descriptors and predicted effective concentration (pEC_{50}) for the virtual data set.



30

Comp. No.	R	R'	X	Y	R''	pEC_{50} (Exp)	ClogP	S_{ALL}	HD_{AL} _L	HA_{AL} _L	R_{ALL}	μ	pEC_{50} (Predicted) SVM	pEC_{50} (Predicted) BPNN	pEC_{50} (Predicted) MLR
47 ^a	Me	H	1-Naphthyl	H	-	>3.699	5.326	0.559	0.450	0.373	0.419	3.599	7.971	7.790	8.042
48 ^a	Me	Me	H	Me	-	>4.032	1.997	0.718	0.383	0.731	0.658	2.196	5.783	6.001	5.906
49 ^a	H	H	CN	H	-	>3.699	0.807	0.624	0.819	0.025	0.626	1.218	5.515	5.057	5.387
50 ^a	Me	H	CN	H	-	>3.699	1.256	0.813	0.816	0.757	0.818	2.718	6.218	6.466	6.328
51 ^a	H	H	NO ₂	H	-	>3.699	-2.228	0.664	0.777	0.185	0.638	3.031	3.942	4.251	4.174
52 ^a	Me	H	NO ₂	H	-	>3.699	1.779	0.837	0.811	0.745	0.796	5.221	6.973	7.194	7.079
53 ^a	H	H	COCH ₂ CH ₃	H	-	>3.699	2.201	0.718	0.871	0.061	0.730	3.998	6.790	6.706	6.548
54 ^b	i-propenyl	Me	2-methylbutyl	H	2,6-di-Cl	-	6.370	0.543	0.187	0.010	0.257	1.528	8.452	7.628	8.026
55 ^b	i-propenyl	H	2-methylbutyl	H	2-Br,4-F,6-Cl	-	6.270	0.505	0.116	0.002	0.226	1.104	8.288	7.462	7.730
56 ^b	i-propenyl	i-propyl	i-pentyl	H	2,6-di-Cl	-	6.370	0.473	0.148	0.005	0.226	1.612	8.307	7.460	7.796
57 ^b	i-propyl	i-propyl	2-Oxobutyl	H	2,4,6-tri-Cl	-	6.300	0.505	0.251	0.013	0.245	1.655	8.508	7.659	8.058
58 ^b	i-propyl	i-propyl	2-Oxobutyl	H	2,6-di-Br	-	5.880	0.553	0.283	0.013	0.279	1.912	8.328	7.620	7.997
59 ^b	i-propenyl	i-propyl	2-Oxobutyl	H	2,6-di-Br,4-F	-	5.770	0.565	0.316	0.010	0.280	2.526	8.402	7.655	8.132

^acompounds with activity reported in approximate terms, ^bvirtually designed compounds.

Figure 5 presents the graphical comparison of predicted values of pEC_{50} estimated using SVM technique for the compounds with approximate activity and that of virtual dataset.

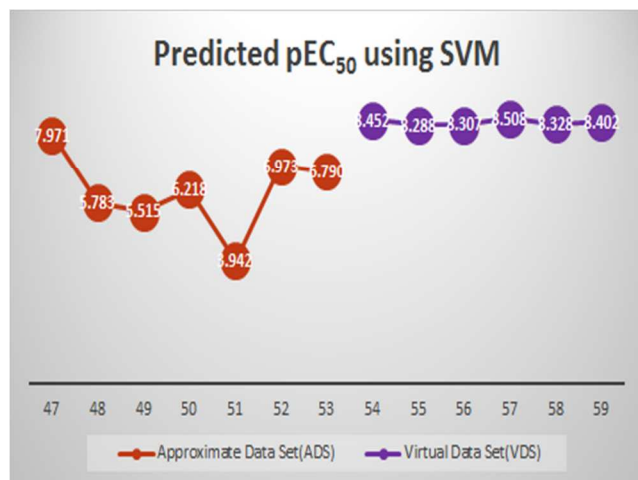


Figure 5. A graph depicting the estimated pEC_{50} values for compounds with approximate activity and the VDS of DABO analogues

The structures of the virtually designed compounds, exhibiting pEC_{50} values > 8.25 , that are identified as lead/drug molecules from the virtual data set are presented in Figure 6.

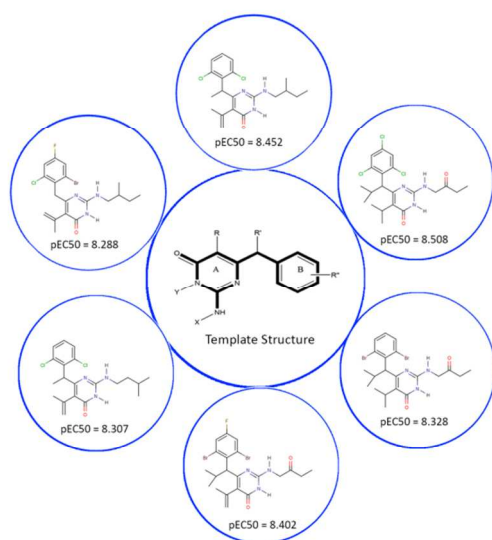
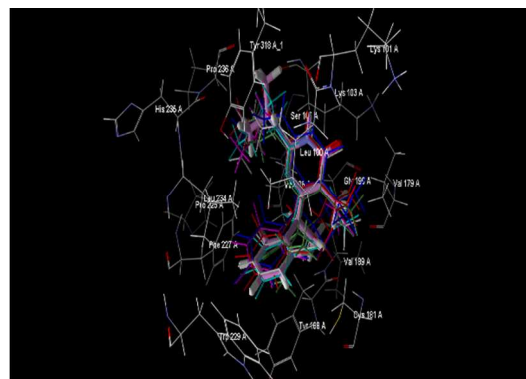
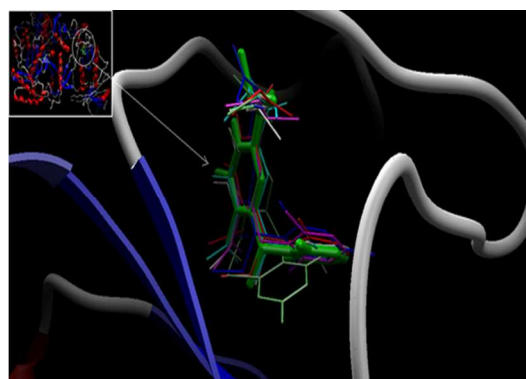


Figure 6. Structures of lead compounds (Virtual Data Set) with their predicted pEC_{50} .

Validation of results for VDS by Docking and Virtual Screening. The docking protocol involving receptor-ligand crystal structure (1JLA)¹⁰³ is highly satisfactory (RMSD = 0.37 Å) to run the docking simulations of the virtual compounds. Figure 7-a shows the snapshot of the overall superimposition of Compound no. 26 (highest active ligand with pEC_{50} = 8.301) along with all the VDS compounds while Figure 7-b shows the secondary view. From the figure 7a it can be observed that the compound no. 26 and all the VDS compounds are closely surrounded by amino acids GLY190, LYS101, VAL106, TYR318, PHE227, LEU234, CYS181, TYR188, VAL189, VAL175, LYS103, SER105, PRO236, PRO225, HIS235, and TRP229.



7-a



7-b

Figure 7 a. Superimposed structures of Comp. No. 26 and all VDS compounds (Comp. Nos. 54-59) showing interactions with the surrounding amino acids, **7 b.** Secondary view of Superimposed structures of Comp. No. 26 and all VDS compounds (Comp. Nos. 54-59).

Table 6 presents the respective distances (in Å) from the nearest amino acid residues of binding pocket for the TNK-651, highest active ligand (26), compound no. 47 and virtually designed compounds (54-59). From table-5, it is clear that compound no.26 is better bound as compared to TNK-651 in NNIBP, due to closer interaction with the surrounding amino acid residues. The compound No. 47 (with highest predicted activity amongst compounds whose activity is reported in approximate terms) and compound no. 26 exhibit similar interactions based on their respective distance from the referred amino acids. The comparison made above suggest that all compounds of VDS show more proximity with the surrounding amino acids and thus resulting in better interactions. Thus, the order of proximity and thereby possibilities of closer interaction in NNIBP as follows: VDS Compounds $> 26 \cong 47 > TNK-651$. Among all the compounds of VDS, it is observed that the compound no. 57 (also whose predicted activity is the highest, 8.508) is relatively in more near vicinity than the other VDS compounds and thus better binding can be interpreted in term of more favourable interactions. Compared to compound no. 26, five of the VDS compounds have shown better interactions with the amino acids while compound no.55 whose activity though a little less has shown nearly similar binding.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

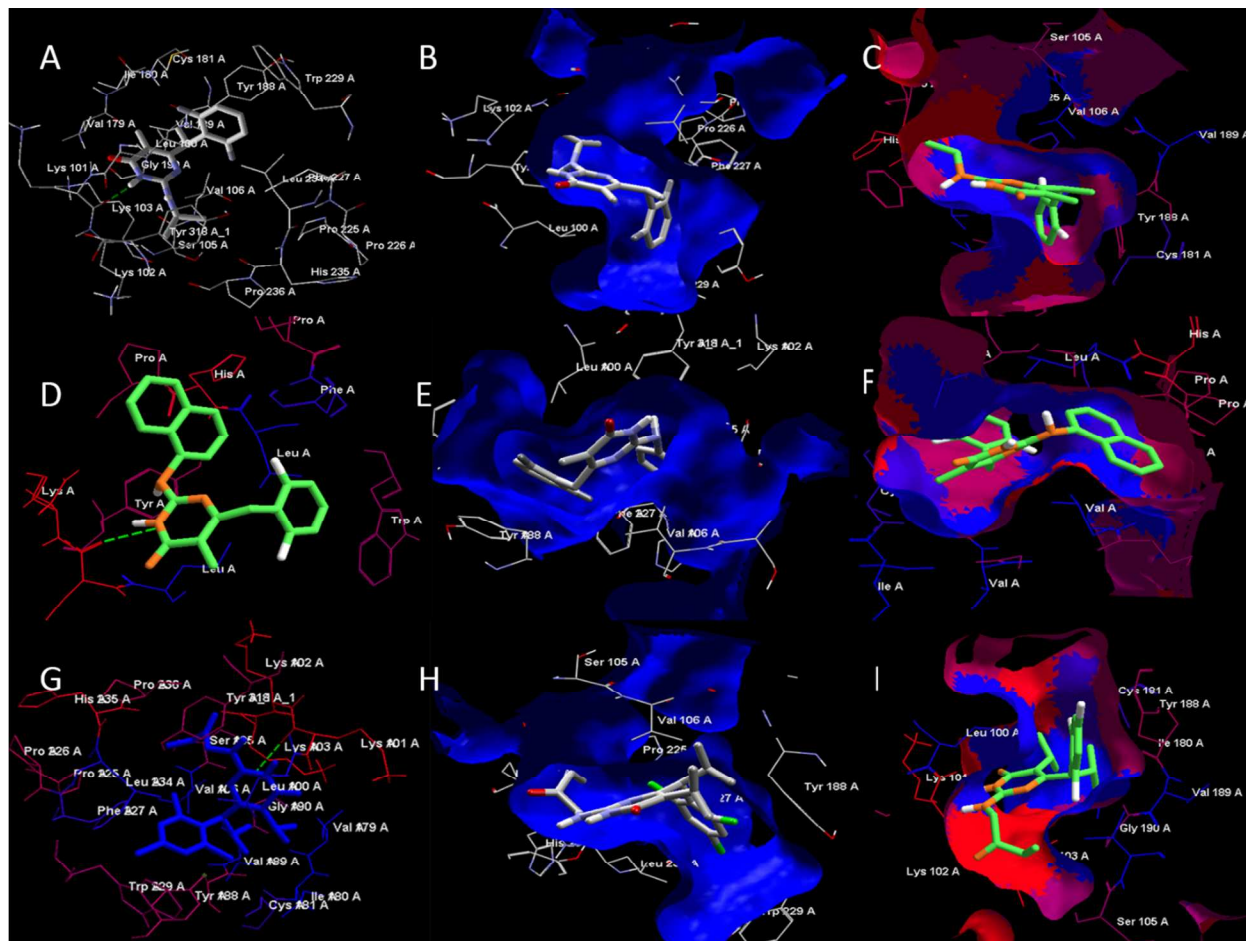
Table 6. Distances (in Å) of TNK-651, highest active ligand (26), compound 47* and virtually designed compounds (54-59) from the nearest amino acid residues.

Compounds → Amino Acid Residues	TNK_651	26	47	54	55	56	57	58	59
LYS101	2.606	1.626	1.997	1.619	1.643	1.643	1.466	1.449	1.879
LYS103	4.310	3.868	3.768	4.195	4.096	3.817	3.578	4.169	3.558
LEU234	3.576	4.236	2.930	4.626	3.701	4.256	3.252	3.299	3.300
PRO236	3.610	3.590	4.718	3.538	3.581	4.094	3.630	3.596	3.753
TRP229	5.026	4.747	3.812	4.367	4.945	4.229	3.321	4.101	4.750
TYR188	4.156	3.700	4.054	2.594	3.327	2.489	3.568	3.032	3.298
TYR318	2.839	3.302	3.297	3.636	1.618	3.379	3.613	1.787	3.701

*The compounds whose activity is reported in approximate terms.

The interactions of the docked pose of compound nos. 26, 47 and 57 with the 1JLA molecule are presented in figures 8 (a-i).

Electrostatic contour maps (8-b, e, h) indicate the favourable polar interactions between the protein and the ligand.

**Figure 8 (a-i).** Hydrogen bond, Electrostatic and hydrophobic interactions of Comp. 26 (A-C), Comp. no. 47 (D-F) and Comp. 57 (G-I) with surrounding amino acid residues.

The following observations are that the compounds with bulkier substitution at R'', isopropyl and isopropylene at R and R' and oxo-butyl group as X show better activity. The blue contour

enclosing most of the molecular region represent positive interactions favouring nucleophilic assortments. The hydrogen bond interaction observed for virtual compound no. 57 is similar to that of compound no. 26. Isopropyl group (R and R'), and

chlorine attached on ring B are inserted into blue pocket exhibiting favourable hydrophobic interactions. The oxo-butyl group inserted into the red pocket shows favourable hydrophilic interactions. The ring B also shows π - π stacking behaviour with TYR188 suggesting the stability of the ligand-protein complex.

CONCLUSIONS

The present paper reports six novel virtually designed DABO derivatives. The new compounds have been obtained by combination of peculiar chemical features of well-known family of pyrimidinone- containing NNRTIs (DABOs) and showed a discrete, characteristic structure-activity relationship profile.

The chemometric analyses for better understanding of interactive capability of DABO derivatives within the NNBP have yielded surprising results. The biological response of the antiviral compounds is enhanced mainly by the presence of steric attributes, hydrogen donor capabilities and lipophilicity and bulkier groups at R, R' and X are suggested for better biological response. On the basis of results obtained from VDS it can be concluded that compounds with 2,6, di-chloro and/or di-bromo and tri-halo-substitution on ring-B and minor modifications of R and X are better performer. The better performance of SVM and BPNN models over MLR model is suggestive of the fact that there invariably exists some degree of non-linear relationship between the anti-viral activity and descriptors used. Also, among the chemometric tools, SVM certainly has better applicability as well as interpretability in terms of varied kernel settings leading to search spanning from linear to nonlinear (radial, sigmoid, polynomial) relationships. The statistical results of the training set are validated and complimented by the test set. A significant point observed while comparing all the VDS structures concludes that chain length of X affects activity but the chain length for X of should be five bonds or less to get enhanced activity and presence of O-atom is still more beneficial. Thus, better activity of compound no. 57 can be attributed to the presence of oxo-butyl group and exhibit favourable polar interactions with PRO236, TYR318 and VAL106 residues in the binding pocket. The relative orientation of the aromatic moieties of the ligand with respect to the orientation of the aromatic moieties of the receptor and their involvement in stacking type interactions guide the polar interactions. The predicted antiviral activities of synthesized DABOs (whose activities are reported in approximate terms) show a wide gap. The *in-silico* generated six virtual compounds showed a high biological response and thus they can be used as precursors or lead compounds and can be synthesized and tested.

Acknowledgements: The authors (Nilanjana Jain (Pancholi)) and Swagata Gupta thank DST, New Delhi for grant of Woman Scientist Fellowship (WOS-A) and Principal, GBLPPG College MHOW respectively.

Notes and references

^a Dr. Nitin S. Sapre, Professor, Department of Applied Chemistry, SGSITS, Indore Phone: +919826607444, Email: sukasap@yahoo.com.

^a Dr. Nilanjana Jain(Pancholi), Department of Applied Chemistry, SGSITS, Indore Phone: +919826607444, Email: sukasap@yahoo.com.

^bDr. Swagata Gupta, ^bDepartment of Chemistry, Govt. BLPPG College, MHOW

^cNeelima Sapre, Department of Mathematics and Computational Sc., SGSITS, Indore

- 1 B.M. Mathers, L. Degenhardt, C. Bucello, J. Lemon, L. Wiessing, M. Hickman, *Bull World Health Organ.* 2013, **91(2)**, 102-123.
- 2 L. Genovese, M. Nebuloni, M. Alfano, *Front Immunol.* 2013, **4(86)**, doi: 10.3389/fimmu.2013.00086. eCollection 2013.
- 3 H. Okatch, B. Ngwenya, K.M. Raletamo, K. Andrae-Marobela, *Anal. Chim. Acta.* 2012, **7(30)**, 42-48.
- 4 V. Repunte-Canonigo, C. Lefebvre, O. George, T. Kawamura, M. Morales, G. Koob, A. Califano, E. Masliah, P.P. Sanna, *Mol Neurodegener.* 2014, **9(1)**, 26.
- 5 P. Zhan, X. Chen, D. Li, Z. Fang, E. De Clercq, X. Liu, *Med. Res. Rev.* 2013, **33**, E1-72.
- 6 L. Schneider, N. Ktorza, S. Fourati, L. Assoumou, E. Courbon, F. Caby, C. Blanc, M. Tindel, R. Agher, A.G. Marcelin, V. Calvez, G. Peytavin, C. Katlama, *HIV Clin. Trials*, 2012, **13(5)**, 284-288.
- 7 C. Reynolds, C.B. de Koning, S.C. Pelly, W.A. van Otterlo, M.L. Bode, *Chem. Soc. Rev.* 2012, **41(13)**, 4657-4670.
- 8 R. Paredes, B. Clotet, *Antiviral Res.* 2010, **85**, 245-265.
- 9 M.P. de Béthune, *Antiviral Res.* 2010, **85(1)**, 75-90.
- 10 R. Gupta, A. Hill, A.W. Sawyer, D. Pillay, *Clin. Infect. Dis.* 2008, **47(5)**, 712-722.
- 11 T. Homma, *Nihon Rinsho*, 2012, **70**, 326-330.
- 12 A.M. Almerico, M. Tutone, A. Lauria, *J. Comput. Aided Mol. Des.* 2008, **22**, 287-297.
- 13 (13) N. Jain, S. Gupta, N. Sapre, N.S. Sapre, *Mol. Biosys.* 2014, **10(2)**, 313-325.
- 14 N.S. Sapre, N. Jain(Pancholi), S. Gupta, N. Sapre *RSC Adv.* 2013, **3**, 10442-10451.
- 15 N.S. Sapre, N. Pancholi, S. Gupta, N. Sapre, *J. Comp. Chem.* 2008, **29(11)**, 1699-1706.
- 16 S.E. Galembeck, F.M. Bickelhaupt, Fonseca, C. Guerra, E. Galembeck, *J. Mol. Model.* 2014, **20(7)**, 2332. doi: 10.1007/s00894-014-2332-3.
- 17 H. Huang, R. Chopra, G.L. Verdine, S.C. Harrison, *Science*, 1998, **282**, 1669-1675.
- 18 M.T. Christen, L. Menon, N.S. Myshakina, J. Ahn, M.A. Parniak, R. Ishima, *Chem. Biol. Drug Des.* 2012, **80(5)**, 706-716.
- 19 Y. Hsiou, J. Ding, K. Das, A.D. Clark Jr., S.H. Hughes, E. Arnold, *Structure* 1996, **4**, 853-860.
- 20 A.L. Hopkins, J. Ren, R.M. Esnouf, B.E. Willcox, E.Y. Jones, C. Ross, T. Miyasaka, R.T. Walker, H. Tanaka, D.K. Stammers, D.I. Stuart, *J. Med. Chem.* 1996, **39**, 1589-1600.
- 21 J. Ding, K. Das, C. Tantillo, W. Zhang, A.D. Clark, Jr, S. Jessen, X. Lu, Y. Hsiou, A. Jacobo-Molina, K. Andries, R. Pauwels, H. Moereels, L. Koymans, P.A.J. Janssen, R.H. Smith, Jr, M. Kroeger Koepke, C.J. Michejda, S.H. Hughes, E. Arnold, *Structure*, 1995, **3**, 365-379.
- 22 J. Ren, D.K. Stammers, *Virus Res.* 2008, **134(1-2)**, 157-170.
- 23 G. Maga, M. Radi, M.A. Gerard, M. Botta, E. Ennifar, *Viruses* 2010, **2(4)**, 880-899.
- 24 N.S. Sapre, S. Gupta, N. Pancholi, N. Sapre, *J. Comput.-Aided Mol. Des.* 2008, **22**, 69-80.
- 25 K. Das, A.D. Clark, P.J. Lewi, J. Heeres, M.R. De Jonge, L.M. Koymans, H.M. Vinkers, F. Daeyaert, D.W. Ludovici, M.J. Kukla, B. De Corte, R.W. Kavash, C.Y. Ho, H. Ye, M.A. Lichtenstein, K. Andries, R. Pauwels, M.P. De Béthune, P.L. Boyer, P. Clark, S.H. Hughes, P.A. Janssen, E. Arnold, *J. Med. Chem.* 2004, **47**, 2550-2560.
- 26 E. De Clercq, *Chem. Biodiv.* 2004, **1**, 44-64.
- 27 G. Meng, Y. Liu, A. Zheng, F. Chen, W. Chen, E. De Clercq, C. Pannecouque, J. Balzarini, *Eur. J. Med. Chem.* 2014, **82**, 600-611.
- 28 Valerie Braz, A.L. Holladay, D. M. Barkley, *Biochemistry*, 2010, **49(3)**, 601-610.
- 29 S. Brück, S. Witte, J. Brust, D. Schuster, F. Mosthaf, M. Procaccianti, J.A. Rump, H. Klinker, D. Petzold, M. Hartmann, *Eur. J. Med. Res.* 2008, **13(7)**, 343-348.

30. J.D. Croxtall, *Drugs*, 2012, **72(6)**, 847-869.
31. A.C. Achhra, M.A. Boyd, M.G. Law, G.V. Matthews, A.D. Kelleher, D.A. Cooper, *PLoS One*, 2014, **9(6)**, e99530.
32. J. Vingerhoets, L. Rimsky, V. Van Eygen, S. Nijs, S. Vanveggel, K. Boven, G. Picchio, *Antivir. Ther.* 2013, **18(2)**, 253-256.
33. Narayanan, G. Sampey, R. Van Duyne, I. Guendel, K. Kehn-Hall, J. Roman, R. Currer, H. Galons, N. Oumata, B. Joseph, L. Meijer, M. Caputi, S. Nekhai, F. Kashanchi, *Virology*, 2012, **432(1)**, 219-231.
34. M.T. Lai, M. Feng, J.P. Falguyret, P. Tawa, M. Witmer, D. DiStefano, Y. Li, J. Burch, N. Sachs, M. Lu, E. Cauchon, L.C. Campeau, J. Grobler, Y. Yan, Y. Ducharme, B. Côté, E. Asante-Appiah, D.J. Hazuda, M.D. Miller, *Antimicrob. Agents Chemother.* 2014, **58(3)**, 1652-1663.
35. M. Botta, M. Artico, S. Massa, A. Gambacorta, M.E. Marongiu, A. Pani, P. La Colla, *Euro. J. Med. Chem.* 1992, **27(3)**, 251-257.
36. Y. He, F. Chen, X. Yu, Y. Wang, E. De Clercq, J. Balzarini, C. Pannecouque, *Bioorg. Chem.* 2004, **32(6)**, 536-548.
37. Y. He, F. Chen, G. Sun, Y. Wang, E. De Clercq, J. Balzarini, C. Pannecouque, *Bioorg. Med. Chem. Lett.* 2004, **14(12)**, 3173-3176.
38. M. Yu, E. Fan, J. Wu, X. Liu, *Curr. Med. Chem.* 2011, **18(16)**, 2376-2385.
39. S. Yang, F.E. Chen, E. De Clercq, *Curr. Med. Chem.* 2012, **19(2)**, 152-162.
40. R. Ragno, A. Mai, G. Sbardella, M. Artico, S. Massa, C. Musiu, M. Mura, F. Marturana, A. Cadeddu, P. La Colla, *J. Med. Chem.* 2004, **47(4)**, 928-934.
41. Y. Wang, F. Chen, E. Clercq, J. Balzarini, C. Pannecouque, *Euro. J. Med. Chem.* 2009, **44(3)**, 1016-1023.
42. M. Radi, M. Pagano, L. Franchi, D. Castagnolo, S. Schenone, G. Casaluce, C. Zamperini, E. Dreassi, G. Maga, A. Samuele, E. Gonzalo, B. Clotet, J.A. Esté, M. Botta, *ChemMedChem.* 2012, **7(5)**, 883-896.
43. M.A. de Brito, C.R. Rodrigues, J.J. Cirino, R.B. de Alencastro, H.C. Castro, M.G. Albuquerque, *J. Chem. Inf. Model.* 2008, **48(8)**, 1706-1715.
44. M.A. de Brito, C.R. Rodrigues, J.J. Cirino, J.Q. Araújo, T. Honório, L.M. Cabral, R.B. de Alencastro, H.C. Castro, M.G. Albuquerque, *Molecules*, 2012, **17(7)**, 7666-7694.
45. N.S. Sapre, T. Bhati, S. Gupta, N. Pancholi, U. Raghuvanshi, D. Dubey, V. Rajopadhyaya, N. Sapre, *J. Biophys. Chem.* 2011, **2(3)**, 361-372.
46. R. Costi, R. Di Santo, M. Artico, S. Massa, A. Lavecchia, T. Marceddu, L. Sanna, P. La Colla, M.E. Marongiu, *Antivir. Chem. Chemother.* 2000, **11(2)**, 117-133.
47. Y. Mao, Y. Li, M. Hao, S. Zhang, C. Ai, *J. Mol. Model.* 2012, **18(5)**, 2185-2198.
48. D. Rotili, M. Tarantino, M.B. Artico, E. Nawrozkij, B. Gonzalez-Ortega, A. Clotet, J. Samuele, A. Esté, G. Maga, A. Mai, *J. Med. Chem.* 2011, **54(8)**, 3091-3096.
49. M. Yu, Z. Li, S. Liu, E. Fan, C. Pannecouque, E. De Clercq, X. Liu, *ChemMedChem.* 2011, **6(5)**, 826-833.
50. M. Radi, C. Falciani, L. Contemori, E. Petricci, G. Maga, A. Samuele, S. Zanolli, M. Terrazas, M. Castria, A. Togninelli, J.A. Esté, I. Clotet-Codina, M. Armand-Ugón, M. Botta, *ChemMedChem.* 2008, **3(4)**, 573-593.
51. Y.P. He, J. Long, S.S. Zhang, C. Li, C.C. Lai, C.S. Zhang, D.X. Li, D.H. Zhang, H. Wang, Q.Q. Cai, Y.T. Zheng, *Bioorg. Med. Chem. Lett.* 2011, **21(2)**, 694-697.
52. M. Radi, L. Angeli, L. Franchi, L. Contemori, G. Maga, A. Samuele, S. Zanolli, M. Armand-Ugon, E. Gonzalez, A. Llano, J.A. Esté, M. Botta, *Bioorg. Med. Chem. Lett.* 2008, **18(21)**, 5777-5780.
53. Mai, M. Artico, R. Ragno, G. Sbardella, S. Massa, C. Musiu, M. Mura, F. Marturana, A. Cadeddu, G. Maga, P. La Colla, *Bioorg. Med. Chem.* 2005, **13**, 2065-2077.
54. N.S. Sapre, S. Gupta, N. Pancholi, N. Sapre, *J. Comp. Chem.* 2009, **30(6)**, 922-933.
55. R.D. Cramer, B. Wendt, *J. Comput. Aided- Mol. Des.* 2007, **21**, 23-32.
56. C.E. Mowbray, R. Corbau, M. Hawes, L.H. Jones, J.E. Mills, M. Perros, M.D. Selby, P.A. Stupple, R. Webster, A. Wood, *Bioorg. Med. Chem. Lett.* 2009, **19(19)**, 5603-5606.
57. S. Gritsch, S. Guccione, R. Hoffmann, A. Cambria, G. Raciti, T. Langer, *J. Enzyme Inhib.* 2001, **16(3)**, 199-215.
58. H. Van de Waterbeemd, *Drug Des. Discov.* 1993, **9(3-4)**, 277-285.
59. Y. Brito-Sánchez, J.A. Castillo-Garit, H. Le-Thi-Thu, Y. González-Madariaga, F. Torrens, Y. Marrero-Ponce, J.E. Rodríguez-Borges, *SAR QSAR Environ. Res.* 2013, **24(3)**, 235-251.
60. Y. Zhou, J. Jiang, W. Lin, H. Zou, H. Wu, G. Shen, R. Yu, *Euro. J. Pharma. Sci.* 2006, **28**, 344-353.
61. P. Liu, W. Long, *Int. J. Mol. Sci.* 2009, **10(5)**, 1978-1998.
62. J. Li, S. Li, B. Lei, H. Liu, X. Yao, M. Liu, P. Gramatica, *J. Comput. Chem.* 2010, **31(5)**, 973-985.
63. F. Gharagheizi, B. Tirandazi, R. Barzin, *Ind. Eng. Chem. Res.* 2009, **48**, 1678-1682.
64. K. Roy, P.P. Roy, *Eur. J. Med. Chem.* 2009, **44(7)**, 2913-2922.
65. X.J. Yao, A. Panaye, J.P. Doucet, R.S. Zhang, H.F. Chen, M.C. Liu, Z.D. Hu, B.T. Fan, *J. Chem. Inf. Comput. Sci.* 2004, **44(4)**, 1257-1266.
66. H. Tanaka, M. Baba, M. Ubasawa, H. Takashima, K. Sekiya, I. Nitta, S. Shigeta, R.T. Walker, E. De Clercq, T. Miyasaka, *J. Med. Chem.* 1991, **34(4)**, 1394-1399.
67. ChemDraw Ultra 7.0.0 (www.cambridgesoft.com)
68. Molegro Virtual Docker, V. 6.0.0, 2013. (www.molegro.com)
69. Sybyl-X 2.1 suite, 2013
70. Tavlarakis, R.H. Zhou, *Mol. Simul.* 2009, **35**, 1224-1241.
71. K. Osmialowski, J. Halkiewicz, A. Radecki, R.J. Kalisz, *J. Chromatogr.* 1985, **346**, 53-60.
72. P.W. Atkins, Quanta, A Handbook of Concepts (2nd ed.). New York: Oxford University Press, 1991.
73. T. Fujita, J. Iwasa, C. Hansch, *J. Am. Chem. Soc.* 1964, **86**, 5175-5180.
74. R. Smith, C. Hansch, M. Ames, *J. Pharm. Sci.* 1975, **64(4)**, 599-606.
75. R. Mannhold, R. Rekker, *Perspectives in Drug Discovery and Design*, 2000, **18(1)**, 1-18.
76. C. Hansch, A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley Interscience New York, 1979.
77. J. Chou, P. Jurs, *J. Chem. Inf. Comput. Sci.* 1979, **19**, 172-178.
78. G.N. Elliott, H. Worgan, D. Broadhurst, J. Draper, J. Scullion, *Soil. Biol. Biochem.* 2007, **39**, 2888-2896.
79. K. Roy, J.T. Leonard, *Indian J. Chem. Sect. A-Inorg. Bio-Inorg. Phys. Theor. Anal. Chem.* 2006, **45**, 126-137.
80. Niazi, S. Jameh-Bozorgi, D. Nori-Shargh, *J. Hazard. Mater.* 2008, **151**, 603-609.
81. E. Pourbasheer, S. Riahi, M.R. Ganjali, P. Norouzi, *Eur. J. Med. Chem.* 2009, **44**, 5023-5028.
82. C.C. Chang, C.J. Lin, *Neural Computation*, 2002, **14(8)**, 1959-1977.
83. C. Cortes, V. Vapnik, *Machine learning*, 1995, **20**, 273-297.
84. G. Liang, Z. Li, *J. Mol. Graph. Model.* 2007, **26**, 269-281.
85. R. Darnag, E.L. Mostapha Mazouz, A. Schmitzer, D. Villemain, A. Jarid, D. Cherqaoui, *Euro. J. Med. Chem.* 2010, **45**, 1590-1597.
86. Y. Cong, X. Yang, W. Lv, Y. Xue, *J. Mol. Graph. Model.* 2009, **28**, 236-244.
87. Z. Shi, X.H. Ma, C. Qin, J. Jia, Y.Y. Jiang, C.Y. Tan, Y.Z. Chen, *J. Mol. Graph. Model.* 2012, **32**, 49-66.
88. H. Golmohammadi, Z. Dashtbozorgi, W.E. Acree, *Euro. J. Pharma. Sci.*, 2012, **47**, 421-429.
89. R.K. Prasoona, A. Jyoti, Y. Mukesh, S. Nishant, N.S. Anuraj, J. Shobha, *Interdiscip. Sci.* 2013, **5(1)**, 45-52.

90. N. Segata, E. blanzieri, *J. mach. Learn. Res.* 2010, **11**, 1883-1926.
91. Y. Li, Y. Qin, X. Chen, W. Li, *PLoS One*, 2013, **8(9)**, e73186.
- 5 92. D. Rumelhart, G. Hinton, R. Williams, *Nature*, 1986, **323**, 533-536.
93. V. Akman, P. Blackburn, *J. Logic, Lang. and Inform.* 2000, **9**, 391-395.
94. C. Wang, L. Li, L. Wang, Z. Ping, M.T. Flory, G. Wang, Y. Xi, W. Li, *Diabetes Res. Clin. Pract.* 2013, **100(1)**, 111-118.
- 10 95. M. Szaleniec, R. Tadeusiewicz, M. Witko, *Neurocomputing*, 2008, **72**, 241-256.
96. S. So, G. Richards, *J. Med. Chem.* 1992, **35**, 3201-3207.
97. V. Maniezzo, *IEEE Trans. Neural Networks*, 1994, **5**, 39-53.
98. T.A. Andrea, H. Kalayeh, *J. Med. Chem.* 1991, **34**, 2824-2836.
- 15 99. V.W. Porto, D.B. Fogel, L.J. Fogel, *IEEE Expert.* 1995, **10**, 16-22.
100. D.L. Selwood, D.J. Livingstone, J.C.W. Comley, A.B. O'Dowd, A.T. Hudson, P. Jackson, K.S. Jandu, V.S. Rose, J.N. Stables, *J. Med. Chem.* 1990, **33(1)**, 136-142.
- 20 101. L. Eriksson, E. Johansson, M. Muller, Wold, S. *J. Chemometrics*, 2000, **14**, 599-616.
102. N. S. Sapre, S. Gupta, N. Pancholi, A. Sikarwar, N. Sapre, *Acta. Chim. Slov.* 2007, **54**, 797-804.
- 25 103. <http://www.rcsb.org/pdb/explore.do?structureId=1jla>