IJPSR (2013), Vol. 4, Issue 10 **(Research Article)**

INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES AND RESEARCH

Received on 04 May, 2013; received in revised form, 04 September, 2013; accepted, 26 September, 2013; published 01 October, 2013

IDENTIFICATION OF NEW HIV-1 PROTEASE INHIBITORS BY MULTIPLE LINEAR REGRESSION (MLR) AND PHYSICO-CHEMICAL DESCRIPTORS

Kumar Nandan*[1], Md. Belal Ahmad [1], Kumar Ranjan [3] and Baidyanath Sah [2]

[1]Department of Chemistry [1], Department of Mathematics [2], T.N.B. College, T.M. Bhagalpur University, Bhagalpur- 812 007, Bihar, India
Department of Chemistry, BUIT [3], Barkatullah University, Bhopal, Madhya Pradesh, India

Keywords:

QSAR, MFA, HIV-1 activity, MLR, QSPR

Correspondence to Author:

Kumar Nandan

Research Scholar, Department of Chemistry, T.N.B. College, T.M. Bhagalpur University, Bhagalpur-812 007, Bihar, India

E-mail: kumarnandan.2008@rediffmail.com

ABSTRACT: In the present work in mathematical modeling, quantitative structure activity relationship (QSAR) studies were performed on some 5, 6-dihydro-2-pyrones derivatives using statistical work.Using only 4 topological and physico-chemical molecular descriptors, we have achieved 84.81% correct classification of the compounds with and without its activity.A heurisimatedtic algorithm selects the best multiple linear regression(MLR) equation showed the correlation between the observed values and the estimated values of activity is very good(R=0.9209, R^2 =O.8481, PRESS=0.7312, R^2_{cv} =0.8210, S_{PRESS} =0.2074). The results are discussed in the light of the main factors that influence the inhibitory activity of the HIV-1 protease.

INTRODUCTION: The construction and investigation of Physico-Chemical Descriptor which could be used to describe molecular structures is one of the important directions of mathematical chemistry. Nowadays, scientists routinely work with collection of hundreds of thousands of molecular structures which cannot be efficiently processed without use of diverse sets of QSAR parameters. Modern QSAR science uses a broad range of atomic and molecular properties varying from merely empirical to quantum-chemical. QSAR studies have often been carried out by using regression analysis the biological activities are being modeled using a set of molecular descriptor.

QUICK RESPONSE CODE

DOI: 10.13040/IJPSR.0975-8232.4(10).3971-75

Article can be accessed online on: www.ijpsr.com

DOI link: http://dx.doi.org/10.13040/IJPSR.0975-8232.4(10).3971-75

Such varieties of available descriptors in combination with numerous powerful statistical and machine learning techniques allow creating effective and sophisticated structure-bioactivity relationship.To evalute the substrate-envelope hypothesis, new protease inhibitors were designed based on the 5,6-dihydro-2-pyrones derivatives.The binding affinities of thease inhibitors to wild-type HIV-1 protease were measured as previously described [1-2].

Representative compounds from designed inhibitors were also tested against a panel of three to four drug-resistant protease varients. In QSAR, we seek to uncover correlations of biological activity with molecular structure with Quantitative structure property relationship (QSPR); we extend the same notion to general chemical property predication and just biological activity. In either case, the relationship is most often expressed by a linear equation that related molecular properties, X, Y to the desired activity A_i for compounds i.

$$A_i = mx_i + ny_i + oz_i + b$$

Where m, n and o are the linear slopes that express the correlation of the particular molecular property with the activity of the compound, and b is a constant. If only one molecular property is important, for example molecular volume, then above eqn. reduces to the simple form of a straight line, $A_i = mx_i + b$.

The slopes and the constant are often calculated using multiple linear regression (MLR) which is analogous with regular linear regression when there is just one independent variable. In constructing graph theoretical schemes to traditional QSAR methods the graph theoretical approach involves (a rather small set of) structural or graph invariants. In QSAR, one uses statistical methods in order to select critical descriptors and demonstrate a structure – activity correlation. In graph theory, one manipulates a structure algebraically, using partial order and ranking based on selected standards of course, graph theoretical descriptors also yield structure property or structure activity correlations[3].Although the 22 inhibitors analysed in this study had the same molecular scaffold, their various X_a, X_b and X_c substituents generated a range of inhibitors sizes and shape and a range of affinities for the wild-type protease.

The authors have developed a QSAR models to predict protease inhibitors of 5, 6-dihydro-2-pyrones derivatives. The negative logarithm of IC_{50} ($logIC_{50}$) was used as the biological activity in QSAR studies [4].

MATERIAL AND METHODS

Methodology: This methodology used is to transform the chemical structure in to its molecular graph. This can be done by depleting all the Carbon- hydrogen atom as well as hetro atom hydrogen bonds of chemical structure [5-6]. In the present investigation, initially, we have used a set of distance based topological indices and physico-chemical parameter.

Molecular Descriptor: The physico-chemical volume parameter Vol. and logRB is the sum of branching indices in MFA-qsar equation specify the regions of different compounds in the training set, leading to either an increase or decrease in activities.

Indicator Parameter: These are dummy parameters that are some times used to obtain better (i.e. statistically more significant) QSAR models in multivariate regression analysis. In the present study we have used two such dummy parameters (indicator parameter) IP_1 and IP_2. The indicator parameter, IP_1, is equal to one unit if OH is present at X_a otherwise zero. If OH is present at X_c the indicator parameter is IP_2 and and is equal to one otherwise zero

Correlation Analysis: Correlation analysis of biological activity, topological indices and physicochemical parameter was carried out- Inter-Correlated parameter were eliminated stepwise depending on their individual correlation with the biological activity. All possible combinations of parameters were considered for multiple regression analysis.

Regression Analysis: Multiple regression analysis [7-8] a programmed carried out by 'Multi Regress' using stepwise regression methodology carried out. It was carried out using a computer program, graph pad and NCSS software, In order to obtain appropriate models; we used the maximum R^2 Method. In addition we also calculate the quality factor [9] Q, as the ratio of correlation coefficient (R) and the standard error of estimation (Se) i.e. Q= R/Se. Finally, the cross-validation method was used to establish the predictive potential of our models.

Cross-validation: A "cross-validated R^2_{cv} "may then be defined completely analogously to the definition of the conventional R^2_{cv}, as;

$$R^2_{cv} = \frac{SSY - PRESS}{SSY}$$

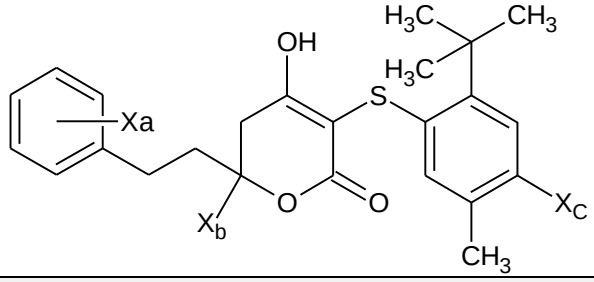
Where PRESS is the standard errors of the cross-validated predictions and SSY is the sum of squared deviations of each biological property value from their mean and PRESS, or predictive sum of squares, is the sum, over all compounds, of the squared differences between the actual and "predicted" biological property values [10].

Software: All molecular modeling studies were carried out using HYPERCHEM (version 7.5) and DRAGON software. The structures of molecules were drawn using Chemsketch software.

NCSS Inc. is a leading worldwide provider of predictive analytics software and solutions.

RESULT AND DISCUSSION: The basic 5, 6-dihydro-2-pyrones derivatives pharmacophore used in the present studies is shown in **table 1**.

TABLE 1: STRUCTURE OF COMPOUNDS MOLECULAR DESCRIPTORS AND THEIR ACTIVITY



Compound no.	logIC$_{50}$	X_a	X_b	X_c	Vol.	logRB	IP_1	IP_2
1	1.5440	H	Ph	H	1308.94	906.8839	0	0
2*	1.5185	H	Ph	OH	1318.64	971.9099	0	1
3	0.8325	H	Ph	$O(CH_2)_2OH$	1462.89	1199.427	0	1
4	0.8195	H	Ph	CH_2OH	1372.46	1042.593	0	1
5	1.1760	H	Ph	OCH3	1380.08	1042.593	0	0
6	1.0413	4-OH	Ph	H	1338.85	977.6093	1	0
7	1.3802	4-NH_2	Ph	H	1352.84	977.6093	0	0
8	1.6020	H	$PhNH_2$	H	1324.63	973.9668	0	0
9	1.5051	H	PhOH	H	1342.22	973.9668	0	0
10	1.0792	H	$PhO(CH_2)_2OH$	H	1461.12	1205.855	0	0
11	0.2304	4-OH	Ph	CH_2OH	1393.27	1118.924	1	1
12	0.3979	3-OH	Ph	CH_2OH	1383.22	1115.836	1	1
13	0.4913	4-NH_2	Ph	CH_2OH	1404.57	1118.924	0	1
14	0.6020	3-NH_2	Ph	CH_2OH	1411.29	1115.836	0	1
15*	0.1461	H	$PhO(CH_2)_2OH$	CH_2OH	1525.90	1363.722	0	1
16	0.8061	H	$PhO(CH_2)_2OH$	$O(CH_2)_2OH$	1603.48	1543.661	0	1
17	0.5682	H	$PhO(CH_2)_2OH$	OH	1474.29	1281.834	0	1
18	0.6532	H	$PhO(CH_2)_2OH$	OH	1579.53	1451.090	0	0
19	2.0791	4-OH	PhOH	CH_2OH	1411.67	11193.84	1	1
20	0.6127	4-OH	Cyclohexyl	CH_2OH	1437.05	1118.924	1	1
21	0.5563	4-OH	Isopropyl	CH_2OH	1336.10	921.4896	1	1
22	0.6334	4-OH	methyl	CH_2OH	1251.48	806.3530	1	1
23	0.5051	4-NH_2	Cyclohexyl	CH_2OH	1453.02	1118.924	0	1
24	0.4313	4-NH_2	Isopropyl	CH_2OH	1357.76	921.4896	0	1

*Delete compounds

The numbers accompanying descriptors in the equation represent their positions in three-dimensional MFA grid **(fig**. 1). We have carried out stepwise multiple regression analysis for modeling of compound no. 20.

Final equation of tetraparametric regression analysis;

$$-0.001792(\pm 0.0006116)Volume + 0.0001665(\pm 2.248E-05)logRB - 0.2791(\pm 0.1141)IP_1 - 0.5626(\pm 0.1007)IP_2 + 3.589$$

N=22 R =0.0.9209 R^2 =0.8481 F =23.7298 R_A^2 =0.8124 Q =4.44

PRESS=0.7312, SSY=4.0831, R_{cv}^2=0.8210, S_{PRESS}=0.2074

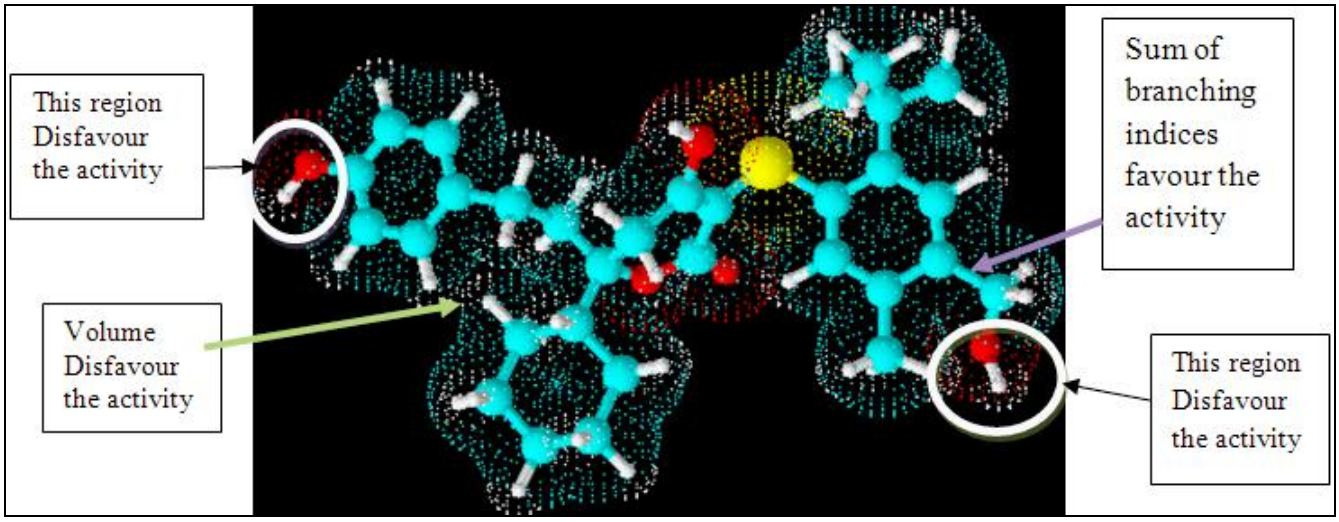


FIG. 1: ALIGNMENT OF THE USED IN TRAINING SET

In order to confirm the above-mentioned finding, we have estimated Q-value and observed that it is highest for model. At this stage, It is interesting to comments an adjustable R^2 (R_A^2) Coefficients. It takes into accounts of adjustment of R^2 therefore If a variable is added that does not contribute its fair share, the R_A^2 will actually decline. If R_A^2 always increases then an independent variable is added. On other side R_A^2 will decrease, this means the added variable does not reduce the unexplained variation enough to offset the loss of degrees of freedom. In our case, R_A^2 value increases with increasing number of parameters. This indicates that the new parameters have a fair share in the proposed model. Further support is out favors in obtained by estimating IC_{50} and compares the same with observed IC_{50} value. Such a comparison is demonstrated in **table 2**. We observed that the estimated value is very close to the observed values.

TABLE 2: COMPARISONS OF OBSERVED AND ESTIMATED IC_{50}

Compound No	logIC_{50}	Predicted	Residuals
1	1.5440	1.394551151	0.149448849
2	0.8325	0.604708811	0.227791189
3	0.8195	0.740664069	0.078835931
4	1.1760	1.289650177	-0.113650177
5	1.0413	1.073664599	-0.032364599
6	1.3802	1.327649537	0.052550463
7	1.6020	1.377601429	0.224398571
8	1.5051	1.346076349	0.159023651
9	1.0792	1.171594170	-0.09239417
10	0.2304	0.437020087	-0.206620087
11	0.3979	0.454517667	-0.056617667
12	0.4913	0.695826086	-0.204526086
13	0.6020	0.683268203	-0.081268203
14	0.8061	0.410060457	0.396039543
15	0.5682	0.597999352	-0.029799352
16	0.6532	1.000212587	-0.347012587
17	2.0791	2.081643626	-0.002543626
18	0.6127	0.358556869	0.254143131
19	0.5563	0.506605762	0.049694238
20	0.6334	0.639091389	-0.005691389
21	0.5051	0.608993219	-0.103893219
22	0.4313	0.746844403	-0.315544403

The most active molecule no. 20 was used for MFA model. A common substructure-based alignment was adopted in the present study, which attempted to align molecules to the template molecule on a common backbone. Finally, we have plotted a graph between observed and calculated value, which yielded predictive correlation co-efficient (**fig. 2**).

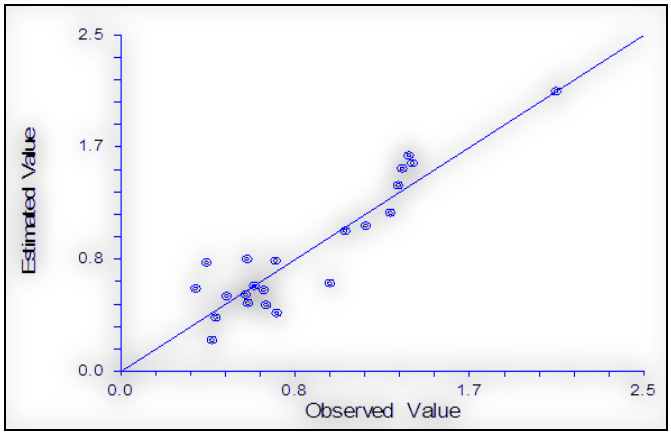


FIG. 2: PLOT OF OBSERVED VS. ESTIMATED ACTIVITY IC_{50}

CONCLUSION: On the basis of above observation it leads to the conclusion that the activity logIC_{50} of the present set of compounds can be successfully modeled using molecular descriptors. It was also observed that out of the molecular descriptors used logRB, volume, IP_1 and IP_2 are most useful for this purpose. The best produced model is a tetra-parametric regression equation with very good statistical fit for good predictive power as evident from its R^2 =0.8481, R^2_{cv} =0.8210,S_{PRESS}=0.2074 values.

The highest value of R^2 and R^2_{cv} and lowest value of S_{PRESS} gave further support to our finding.The MFA equation suggested that (-ve)sign of IP_1 , $IP_2$and volume descriptors are disfavour the activity while (+ve)sign of logRB indices indicate that they favoured activity.Our results open very interesting perspectives regarding 5,6-dihydro-2-pyrones derivatives with protease inhibitors.

ACKNOWLEDGEMENTS: One the author (Kumar Nandan) is highly obliged and thankful Dr. Sunita Gupta, Dept. of Chemistry, A.P.S. University Rewa *India* for introducing him to the fascinating field of Chemical Topology, Graph Theory and statistical work.

REFERENCE:

1. V.K. Agrawal, J. Singh, K.C. Mishra, P.V. Khadikar and Y.A. Jaliwala a QSAR studies on the use of 5, 6-dihydro-2-pyrones as HIV-1 protease inhibitors, *Arkivok*, 2006, 162-177.
2. Abhinav Prasoon Mishra, 2D-QSAR study of 2,5-dihydropyrazolo [4,3-c] quinoline-3-one a novel series of PDE-4-inhibitors, *Int. J. Pharmaceutical and Biomedical Sci.* 2012; 3(1): 105-109.
3. P.V. Khadikar,Anjali Shrivastava,V.K. Agrawal and S. Shrivastava, Topological designing of 4-piperazinylquinazolines as Antagonists of PDGFR tyrosine Kinase family,*Bioorganic & Medicinal Chemistry Letters*, 2003; 13: 3009-3014.
4. Shalini Singh, QSAR Studies on the inhibition of the human carbonic anhydrase cytosile isozyme HCA V^{III+} , *indian J. Chem. Soc*., 2013; 90: 245-252.
5. M.Mladenovic,N.Vukovic,S.Sukdolac and S.Solujic, Design of Novel 4-Hydroxy-Chromene-2-one Derivatives as Antimicrobial agent, *Molecules*, 2010; 15: 4294-4308.
6. Neuwoehner, T. Zilberman, K. Fenner, and B.I. Escher, QSAR-analysis and mixture toxicity as diagnostic tools: Influence of degradation on the toxicity and mode of action of diuron in algae and daphnids, *Aquat. Toxicol*. 2010; 97: 58–67.
7. Kumar Nandan,Baidyanath Prasad,Sunita Gupta & Kumar Ranjan,Statistical Significance of Phathalimide Derivatives,*Indian J. of Mathematics and Mathematical Sci.*, 2011; 7: 51-62.
8. P.V. Khadikar,S. Karmarkar,V.K. Agrawal and S. Shrivastava,Use of distance based topological indices in modeling antihypertensive activity:Case of 2-aryl-imino-imidazolidines, *Indian J. Chem*., 2003; 42A: 1426-1435.
9. Pogliani L, *Amino acids*, 1994; 6: 141.
10. Erol Eroglu, H. Turkmen, S.Guler, S. Palaz and Oral Oltulu, A DFT-based QSARs study of acetazolamide/ sulfanilamide derivatives with carbonic anhydrase(CA-11) isozyme inhibitory activity, *Int. J. Mol. Sci*., 2007; 8: 145-155.

How to cite this article:
Nandan K, Ahmad MB, Ranjan K and Sah B: Identification of new HIV-1 Protease inhibitors by multiple linear regression (MLR) and Physico-chemical descriptors. *Int J Pharm Sci Res* 2013; 4(10): 3971-75. doi: 10.13040/IJPSR. 0975-8232.4(10).3971-75