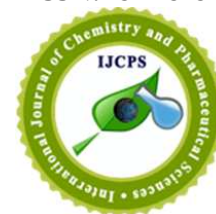




Research Article

International Journal of Chemistry and Pharmaceutical Sciences

www.pharmaresearchlibrary.com/ijcps



Design of 5,6-Dihydro-2-Pyrones Derivatives As HIV-1 Protease Inhibitors Using Molecular Descriptors

Kumar Nandan*¹, Sunita Gupta², Md. Belal Ahmad³, Kumar Ranjan⁴, Baidyanath Sah⁵

¹University Department of Chemistry, T. M. Bhagalpur University India 812007

²University Department of Chemistry, A.P.S. University, Rewa (M.P), India 486003

³Department of Chemistry, T.N.B. College, T. M. Bhagalpur University India 812007

⁴C & M laboratory, Diesel loco shed S.E. Railway Kharagpur (W.B) India

⁵Department of Mathematics, T.N.B. College, T. M. Bhagalpur University India 812007

Abstract

In the present work in mathematical modelling, 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 84.72% correct classification of the compounds with and without its activity. A heuristic 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 ($R=0.9204$, $R^2=0.8472$, $PRESS=0.7357$, $R_{cv}^2=0.8197$, $S_{PRESS}=0.2080$). The results are discussed in the light of the main factors that influence the inhibitory activity of the HIV-1 protease.

Keywords: QSAR, MFA, HIV-1 Activity

Contents

1. Introduction	684
2. Experimental	686
3. Results and discussion	687
4. Conclusion	690
5. Acknowledgement	690
6. References	690

*Corresponding author

Kumar Nandan

E-mail: kumarnandan.2008@rediffmail.com

Manuscript ID: IJCPS2007

Published Online 27 March 2014



PAPER-QR CODE

© 2013, IJCPS All Rights Reserved

1. Introduction

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 which the biological activities are being modeled using a set of molecular descriptor. Such varieties of available descriptors in combination with numerous powerful statistical and machine learning techniques allow creating effective and sophisticated structure-bioactivity relationship to evaluate the substrate-envelope hypothesis, new protease inhibitors were designed based on the 5,6-dihydro-2-pyrones derivatives. Due largely to the emergence of multi-drug-resistant HIV strains, the development of new HIV protease inhibitors remains a high priority for the pharmaceutical industry. Toward this end, we previously identified a 4-

hydroxy-5, 6-dihydropyrone lead compound (CI-1029, 1) which possesses excellent activity against the protease enzyme, good antiviral efficacy in cellular assays, and promising bioavailability in several animal species. The crystal structures of a number of representative 5, 6-dihydro-4-hydroxy-2-pyrones complexes with the HIV protease were also determined to provide better understanding of the interaction between the enzyme and these inhibitors to aid the structure-based drug design effort. The binding affinities of these inhibitors to wild-type HIV-1 protease were measured as previously described¹. Representative compounds from designed inhibitors were also tested against a panel of three to four drug-resistant protease variants. 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 prediction and just biological activity. In either case, the relationship is most often expressed by a linear equation that related molecular properties, X, Yto 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, than 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 correlations². 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. Several graph theoretical invariants have been generalized so that they produce structure dependent descriptors³⁻⁷. In such ageneralization, most biological activities are dominated by molecular size, which is well characterized by most of the physicochemical properties Ideally, the activities and properties are connected by some known mathematical function, f : biological activity = f [structure(in present study topological & physicochemical descriptors are used as the structural parameters.)].

The authors have developed a QSAR models to predict protease inhibitors of 5, 6-dihydro-2-pyrones derivatives. The negative logarithm of IC_{50} ($\log IC_{50}$) was used as the biological activity in QSAR studies. In the present work we have taken 22 compounds (a set of 24 compounds) and their anti-HIV activity from the reported work⁷ (Vijay K Agrawal et al 2006). Often in an automated selection of descriptors a descriptor will be discarded because it is highly correlated with another descriptor already selected. But what is important is not whether two descriptors parallel one another, i.e. duplicate much of the same structural information but whether they in those parts that are important for structure-property-activity correlations. If they differ in the domain, which is important for the property/ activity considered, both descriptors should be retained. If they differ in parts that are not relevant for the correlation of considered property/ activity then one of them can be discarded. Hence, the residual of the correlation between two descriptors (compound no. 2 & 15) should be examined and kept or discarded depending on how well it can improve the correlation based on already selected descriptors. Alternately, one should replace the set of descriptors used by descriptors that can be extracted from them through the orthogonalization procedure that has been introduced in regression analysis⁷ (see p. 164). The results obtained are better (discussed in the Result and Discussion part of this study) than those of a previous QSAR study performed by (Vijay K Agrawal et al 2006) in their review work on 5, 6-dihydro-2-pyrones derivatives parameters in QSAR.

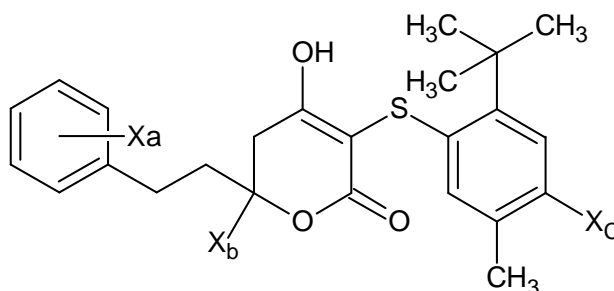


Figure 1. Parent structure of 5, 6-dihydro-2-pyrones derivatives used in present study

2. Materials and methods

Biological Activity:

Biological Activity expressed as $\log IC_{50}$, was taken from the literature⁷. In the present work we have taken 22 compounds and their anti-HIV activity from the reported work (Vijay K Agrawal et al 2006). Many of these compounds inhibited wild type HIV-1 with IC_{50} values between 0.001 μM and 0.005 μM in MT-4 cells. One of these 5, 6-dihydro-2-pyrones derivatives showed good anti-HIV activity (0.02 μM , in clinical trial) with low cytotoxicity for MT4 cells high structural diversity and a sufficient range of the biological activity in the selected series of 5, 6-dihydro-2-pyrones derivatives. It insists as to select these series of compounds for our QSAR studies.

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. In the present investigation, initially, we have used a set of distance based topological indices and physico-chemical parameter⁹.

Molecular Descriptor:

The physico-chemical parachor parameter σ' and is the σ'' 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.

Physicochemical parameter:

In present study molar refractivity (MR), molar volume (MV), parachor (Pr), index of refraction (η), surface tension (ST), density (D), shown in Table 3 are tested and calculated from the computer software acdlabs (Chem sketch 5.0)⁹.

Indicator Parameter:

The indicator parameters^{10, 11} (variables) take on only two values, usually zero and one. The two values signify that the observation falls in one of two possible categories. The numerical values of the dummy variables are not intended to reflect a quantitative ordering of categories, but only serve to identify category or class membership. Therefore, they show the significance of a particular group or a substituent in a given series of drug. They account for the abrupt increase or decrease of a given pharmacological activity at any specific site in the drug molecule. If the coefficient of indicator parameter carries a negative sign in the regression expression, this makes it very clear that the compound having this particular group at a particular position has considerably Lower potency. 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) σ''' and σ'''' . The indicator parameter, σ''' , 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 σ'''' and is equal to one otherwise zero

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.

Correlation Analysis:

Correlation analysis of biological activity, topological indices and physicochemical parameter was carried out- Inter-Related parameter were eliminated stepwise depending on their individual correlation with the biological activity. All possible combinations of parameters were considered for multiple regression analysis^{12, 13}.

Regression Analysis:

Multiple regression analysis a programmed carried out by 'Multi Regress' using stepwise regression methodology was carried out using a computer program, graph pad and NCSS software¹⁴, In order to obtain appropriate models; we used the maximum R^2 Method. In addition we also calculation the quality factor¹⁵ 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. 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 decreases, 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 model¹⁶ (See p. 3974).

Cross-validation:

The predictive power of the equations were validated by leave-one-out(LOO) cross-validation equation. Predicted residual sum of square (PRESS), cross-validated correlation coefficient, and standard deviation error of prediction (SSY) were considered for the validation of these equations. The results from cross-validated analysis were expressed as the cross-validated squared correlation coefficient r^2 ; which is defined as:

$$R^2 = 1 - \frac{\sum (Y_{\text{pred}} - Y_{\text{act}})^2}{\sum (Y_{\text{act}} - Y_{\text{mean}})^2}$$

Where Y_{pred} , Y_{act} , and Y_{mean} are predicted, actual and mean value of target property respectively.

$\Sigma (Y_{\text{pred}} - Y_{\text{act}})^2$ is the predicted Residual Error Sum of Squares (PRESS). PRESS is an important cross-validation parameter as it is a good approximation of the real predictive error of the equations. The statistical parameters considered to compare and select the generated QSAR equations were correlation coefficient (r), standard deviation (s), sequentia Fischer (F) test, Cross-validated correlation coefficient r^2 . A data point is considered as an outlier if it has a large magnitude (when the residual value exceeds twice the standard error of estimate of the equation)¹⁷ (see p. 232).

A "cross-validated R_{cv}^2 " may then be defined completely analogously to the definition of the conventional R_{cv}^2 , as

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

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 values¹⁸ (See p.106).

3. Results and Discussion

The most straightforward way to perform QSAR analysis is to divide the set of molecules into training set and test set. We obtain the QSAR equations using the training set and then apply them on the test set. Such an analysis clearly gauges the reliability of QSAR equations. However, if leave-one-out methodology is used and then cross-validation is done, there is no need for such division into training and test sets. The basic 5, 6-dihydro-2-pyrones derivatives⁷ pharacophare used in the present studies is shown in figure 1.

The numbers accompanying descriptors in the equation represent their positions in three-dimensional MFA grid (fig 2). We have carried out stepwise multiple regression analysis for modelling of compound no 20.

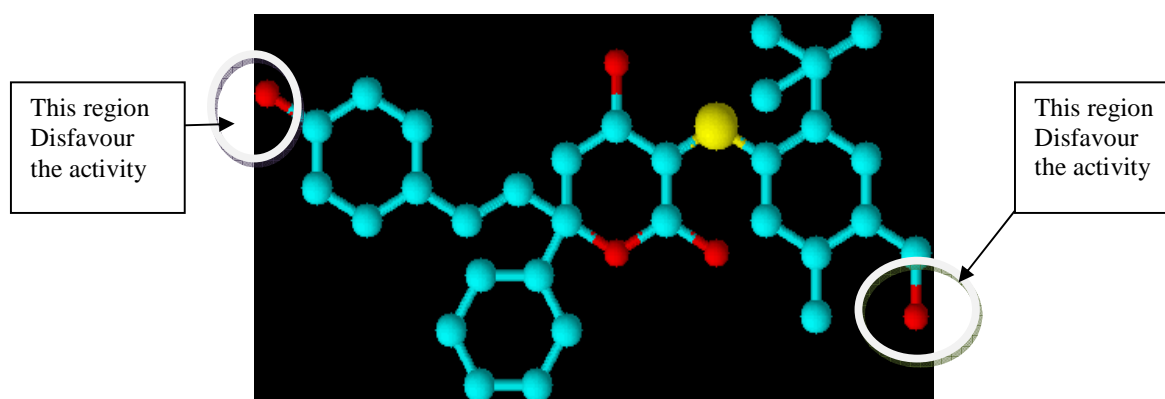


Figure 2: Alignment of the Used In Training Set

Final equation of tetraparametric regression analysis

$$-0.002190(\pm 0.0007543)\sigma' + 0.0001676(\pm 2.263E - 05)\sigma'' - 0.2799(\pm 0.1147)\sigma''' - 0.5614(\pm 0.1012)\sigma'''' + 3.571$$

$$N=22 \quad R=0.9204 \quad R^2=0.8472 \quad F=23.5597 \quad R_A^2=0.8112 \quad Q=4.425$$

$$PRESS=0.7357, \quad SSY=4.0786, \quad R_{cv}^2=0.8197, \quad S_{PRESS}=0.2080, \quad PSE=0.1528$$

PRESS – predicted residual sum of squares,

SSY – sum of the squares of regression value,

R_{cv}^2 – cross-validation correlation coefficient

S_{PRESS} – uncertainty of prediction

PSE – predictive square error

σ' = Parachor

σ'' = Sum of Branching Index

σ''' = OH at X_a position

σ'''' = OH at X_c position

To be a reasonable QSAR model, PRESS/SSY should be the smaller than 0.4, and the value of this ratio smaller than 0.1 indicates an excellent model.

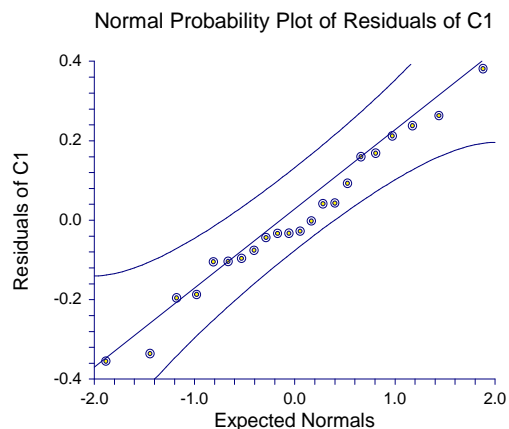
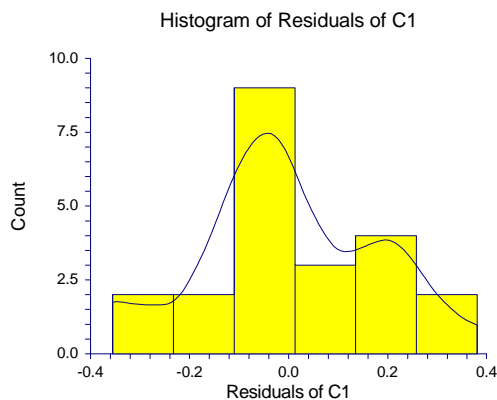
It is interesting to record that the value of S_{PRESS} in the present case is very close to that of standard error of estimation, Se . Hence, both these parameters carry the same meaning and S_{PRESS} does not give any additional information regarding the uncertainty of prediction. In view of this, we have calculated predictive square error, PSE,

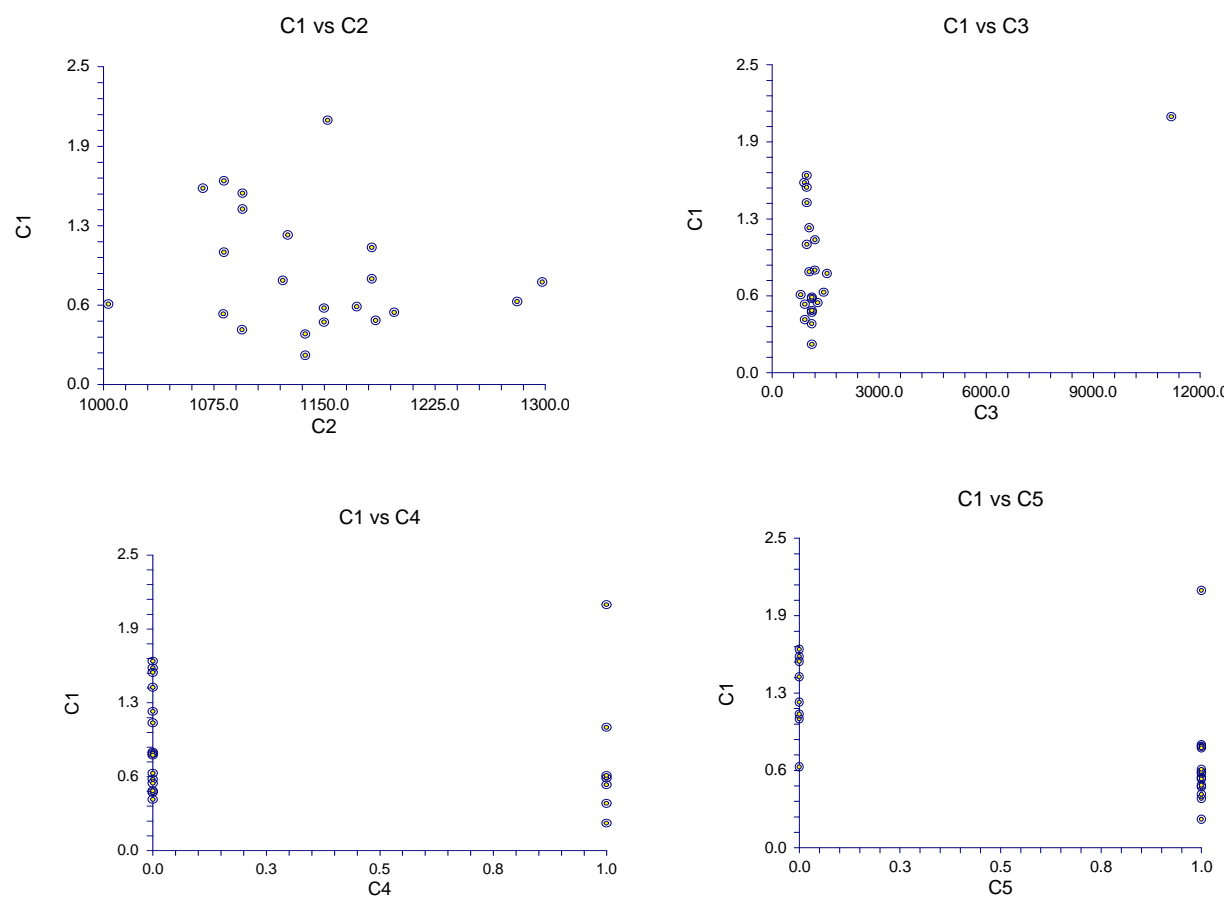
as it seems to be most directly related to the uncertainty of prediction. The value of PSE is found smallest for model indicating that this model has the highest predictive potential⁸. Further support in out favors in obtained by estimating IC₅₀ and compares the same with observed IC₅₀ value. Such a comparison is demonstrated in table 1. We observed that the estimated value is very close to the observed values.

Compound No.	logIC ₅₀	Predicted	Residuals
1	1.5440	1.384435836	0.159564164
2	0.8325	0.620831012	0.211668988
3	0.8195	0.727053362	0.092446638
4	1.1760	1.281024797	-0.105024797
5	1.0413	1.085246623	-0.043946623
6	1.3802	1.337372737	0.042827263
7	1.6020	1.364577423	0.237422577
8	1.5051	1.336762316	0.168337684
9	1.0792	1.183326233	-0.104126233
10	0.2304	0.426394368	-0.195994368
11	0.3979	0.425876872	-0.027976872
12	0.4913	0.678520482	-0.187220482
13	0.6020	0.678002986	-0.076002986
14	0.8061	0.425116527	0.380983473
15	0.5682	0.601350488	-0.033150488
16	0.6532	1.008254036	-0.355054036
17	2.0791	2.081486551	-0.002386551
18	0.6127	0.349738563	0.262961437
19	0.5563	0.515299992	0.041000008
20	0.6334	0.667057030	-0.033657030
21	0.5051	0.601645660	-0.096545660
22	0.4313	0.767426106	-0.336126106

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. The QSAR model obtained exhibits strong dependencies on directional component to molecular descriptors with output values (Figure 3). Finally, we have plotted a graph between observed and calculated value, which yielded predictive correlation co-efficient (Figure 4).

Plots Section:





C1= Experimental Observed Activity

C2= Parachor(σ')

C3= Sum of Branching Index(σ'')

C4= OH at X_a position(σ''')

C5= OH at X_c position(σ'''')

Figure 3. (Continued) log IC_{50} vs (a) Residuals Norms (b) Expected Norms (c) Parachor (d) Sum of Branching Index (e) OH at X_a position (f) OH at X_c position

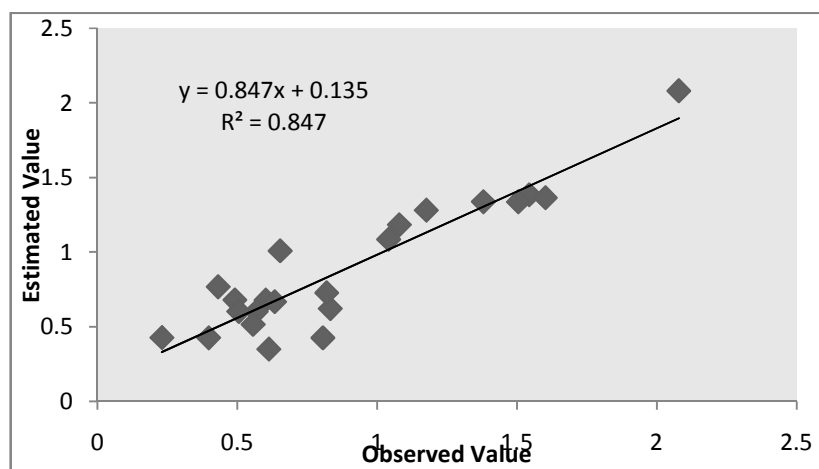


Fig 4: Plot of Observed Vs. Estimated Activity IC_{50}

4. Conclusion

On the basis of above observation it leads to the conclusion that the activity $\log IC_{50}$ of the present set of compounds can be successfully modeled using molecular descriptors. It was also observed that out of the molecular descriptors used, σ'' , σ' , σ''' and σ'''' most useful for this purpose. The predictive power of model was estimated with bootstrapping method and leave-one-out cross validation method. It was observed from the selected models that biological activity of 5, 6-dihydro-2-pyrones derivatives is governed by physicochemical properties of the molecules. The best produced model is a tetra-parametric regression equation with very good statistical fit for good predictive power as evident from its $R^2 = 0.9204$, $R_{CV}^2 = 0.8197$, $S_{PRESS} = 0.2080$ values. The highest value of R^2 and R_{CV}^2 and lowest value of S_{PRESS} gave further support to our finding. The MFA equation suggested that (-ve) sign of σ''' , σ'''' and σ' descriptors are disfavoured the activity while (+ve) sign of σ'' indices indicate that they favoured activity. Our results open very interesting perspectives regarding 5,6-dihydro-2-pyrones derivatives with protease inhibitors.

5. Acknowledgement

Authors are grateful to principal T.N.B. College Bhagalpur TMBU for providing facilities to carry out proposed work. Authors are also thankful to HOD, P.G. Department of chemistry, T.M. Bhagalpur University for their help and encouragement.

6. Reference

1. Susan W Hovorka, Michael J Hageman, Christian Schöneich, Oxidative degradation of a sulfonamide-containing 5,6-dihydro-4-hydroxy-2-pyrone in aqueous/organic cosolvent mixtures. *Pharmaceutical Research*, 2002; 19(4):538-45
2. Kumar Nandan, Baidyanath Prasad, Sunita Gupta & Kumar Ranjan, Statistical Significance of Phthalimide Derivatives, *Indian J. of Mathematics and Mathematical Sci.*, 2011, 7, 51-62
3. Randić, M.: *Quantum Biol. Symp. Int. J. Quantum Chem*, 1988, 15, 201.
4. Randić, M.; Jerman-Blazić, B.; Trinajstić, N. *Computer Chem.* 1990, 14, 237.
5. Randić, M. *J. Chem. Inf. Comput. Sci.* 1995, 35, 373.
6. Kier, L. B.; Hall, L. H. *J. Pharm. Sci.* 1981, 70, 583.
7. 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
8. Shalini Singh, QSAR Studies on the inhibition of the human carbonic anhydrase cytosolic isozyme HCA VI[†], *Indian J. Chem. Soc.*, 2013, 90, 245-252
9. Computer software Chemsketch 5.0 (from ACD labs)
10. Chatterjee S, Hadi S and Price B 2000 *Regression analysis by examples 3rd edn* (New York: Wiley)
11. Trinajstić N 1992 *Chemical graph theory 2nd edn* (Boca Raton, FL: CRC Press)
12. Kumar Nandan, Baidyanath Sah, Md. Belal Ahmad & Kumar Ranjan, QSAR Studies on polychlorinated aromatic compounds using topological descriptors, *Int. J. Pharmaceutical Science and Research* 2013, 4(7), 2691-2695
13. 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
14. 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
15. Pogliani L, *Amino acids*, 1994, 6, 141
16. Kumar Nandan, Baidyanath Sah, Md. Belal Ahmad & Kumar Ranjan, Identification of New HIV-1 Protease Inhibitors By Multiple Linear Regression (MLR) and Physico-Chemical Descriptors, *Int. J. Pharmaceutical Science and Research* 2013, 4(10), 3971-3975
17. Blessy Pothen, Vineet Singh, Surendra Kumar & Meena Tiwari, Structural optimization of new class of selective carbonic anhydrase, *Indian Journal of Chemistry*, 2010, Vol. 49B, pp. 224-233
18. Abhinav Prasoorn 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