e- ISSN: 2394-5532
p- ISSN: 2394-823X
General Impact Factor (GIF): 0.875
International Journal of Applied And Pure Science and Agriculture
www.ijapsa.com

QSAR Studies of Benzamidine derivatives Using Graph Theory And Multilinear Regression Analysis

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

[1]P.G. Department of Chemistry, T.N.B. College, T.M.Bhagalpur University India 812007

[2]Chemical & Metallurgical Laboratory, Diesel Loco shed , S.E. Railway Kharagpur, W.B., India

[3]Department of Mathematics, T.N.B. College, T.M.Bhagalpur University India 812007

ABSTRACT

In the present work in mathematical modeling, graph theory based quantitative structure activity relationship (QSAR) studies were performed on some benzamidine derivatives using statistical work. Graph theory was successfully applied in developing a relationship between chemical structure and biological activity. The regression analysis of the data has shown that the activities of the compound can be modeled excellently in tri-parametric model. 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= 0.9945, R^2 = O.9889, PRESS= 0.2223, R^2_{cv}= 0.9888, S_{PRESS} = 0.1144). The results are discussed critically.

Keywords: QSAR, MFA, Thrombin Inhibitors Activity, MLR, QSPR

I. INTRODUCTION

The extended applications of computational chemistry to biodisciplines play an essential role in molecular property based drug design. In constracting graph theoretical schemes to traditional quantitative structure activity relationship(QSAR) methods one must not be wary of using a complimentry approch. Traditional QSAR is usually based on a large number of empirical parameters. The graph theoritical approch involves(a rather smell set of) structural or graph invarients. In QSAR, one uses statistical methods in order to select critical descriptors and demomonstrate a structure –activity correlation. In graph theory , one manuplates a structure algebraically, using partial order and ranking based on selected standards. Of course, graph theoritical descriptors also yield stucture-property or structure activity correlations[1]. In this work, we have examined the chemically and structurally differerent 21 set of benzamidine derivatives. Some studies indicate that benzamidine based dyes can be metabolized to benzamidine and that human exposure to such dyes is associated with bladder cancer [2].

We seek to uncover correlations of biological activity with molecular structure with Quantative structure activity 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 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 egn. (1) reduces to the simple form of a straight line, $A_s = mx_s + b$. 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].

II. METHODOLOGY:

This methodology used is to transform the chemical structure in to its molecular graph i.e. Two Dimensonal structure. This can be done by depeleting all the Carbon- hydrogen atom as well as hetro atom hydrogen bonds of chemical structure. In the parsent investigation, initially, We have used a set of distance based topological in dices and physico-chemical parameter[4].

Statistical Analysis:

Correlation Analysis:

Correlation analysis[4] 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 was carried out by 'Multi Regress' a programme carried out using stepwise regression methodology5. It was carried out using a computer program, graph pad software, In order to obtain appropriate models; we used the maximum R2 method. In addition we also calculation the quality factor Q, as the ratio of correlation coefficient (R) and the standard error of estimation (Sc) i.e. Q= R/Sc. Finally, the cross-validation method was used to establish the predictive potential of our models. At this stage, It is interesting to comments an adjustable R2 (R_A^2) Coefficients. It takes into accounts of adjustment of R2 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 model.

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 coefficent r^2; which is defined as[5, 6].

$$R^2 = 1-\Sigma (Y_{pred}-Y_{act})^2 \ / \Sigma(Y_{act}-Y_{mean})$$

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

$\Sigma (Y_{pred}-Y_{act})^2$ 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 coefficent 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). 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 values.

Software:

Different types of descriptors were calculated for each molecule in the study table using default settings within cerisu2 and Dragon software. These descriptors include topological, spatial and thermodynamic type. A complete list of descriptors used in the study is given table 1.

III. RESULT AND DISCUSSION:

The descriptors involved in the selected models and their thrombin inhibition activities (pIC_{50}) of 21 Benzamidine Derivatives used in present study are shown in table-1. We have carried out single and step wise multiple regression analysis for modeling of pIC_{50}.

Mono Parametric Regression:

$$0.1887(\pm 0.005538)MR - 6.790$$

R=0.9919, $R^2 = 0.9839$, $R_A^2 = 0.9830$, $R_{CV}^2 = 0.9836$, $S_{PRESS} = 0.1305$

Bi Parametric Regressions:

Successive regression yielded bi parametric models all having better statistics which models discussed below. We still attempt modeling with five correlation parameters.

$$0.1925(\pm 0.07713)\sigma + 0.1544(\pm 0.01459)MR - 7.036$$

R= 0.9940, $R^2 = 0.9880$, $R_A^2 = O.9867$, $R_{CV}^2 = 0.9879$, $S_{PRESS} = 0.1156$

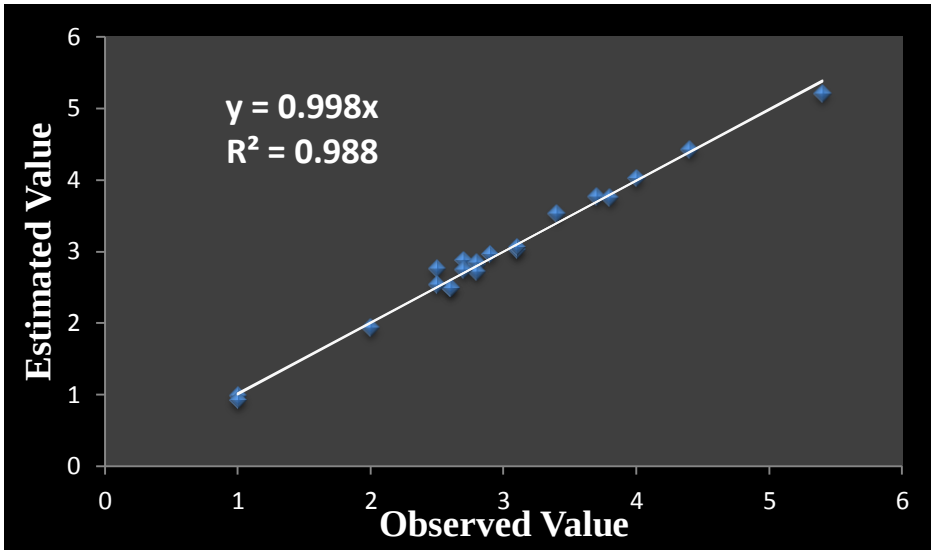
Tri Parametric Regressions:

Finally, stepwise regression resulted into the most significant model containing three parameters. The regression expression has the highest R^2 value and can be expressed as below.

$$0.1934(\pm 0.07632)\sigma + 0.1568(\pm 0.01459)MR - 0.001490(\pm 0.001268)ZM_2 - 7.072$$

R= 0.9945, $R^2 = 0.9889$, $R_A^2 = 0.9870$, $R_{CV}^2 = 0.9888$, $S_{PRESS} = 0.1144$

Fig 2: Plot Of Observed Vs. Estimated Activity (pIC_{50})



IV. CONCLUSION:

On the basis of above observation it leads to the conclusion that the activity pIC_{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, MR and ZM_2 most useful for this purpose.The best produced model is a tri-parametric regression equation with very good statistical fit good predictive power as evident from its R^2 =0.9889, R^2_{cv} =0.9888,S_{PRESS}=0.1144 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 MR and σ descriptors are favour the activity while (-ve)sign of ZM_2 indices indicate that they disfavour the activity.

Table -1 : Benzamidine compounds and there Thrombin inhibitor activities (pIC_{50})

Sl. No.	Compound Name	σ	MR	ZM_2	pIC_{50}
1	4-nitrobenzenecarboximidamide	10.039	50.876	62	2.5
2	3-(hydroxymethyl)benzenecarboximidamide	10.127	49.034	55	2.6
3	2-methylbenzenecarboximidamide	8.1240	41.803	51	1.0
4	3-nitrobenzenecarboximidamide	10.127	49.012	62	2.6
5	3-carbamimidoylbenzoic acid	10.224	51.432	62	2.7
6	3-benzylbenzenecarboximidamide	10.905	54.998	87	3.4
7	benzenecarboximidamide	10.418	51.485	43	2.9
8	3-aminobenzenecarboximidamide	11.877	59.114	50	4.4
9	biphenyl-3-carboximidamide	11.987	55.110	84	3.7
10	3-(dimethylamino)benzenecarboximidamide	10.613	51.822	62	3.1
11	3-methoxybenzenecarboximidamide	10.613	52.012	55	3.1
12	3,4-dimethylbenzenecarboximidamide	10.921	50.123	58	2.8
13	3-bromobenzenecarboximidamide	10.321	50.975	50	2.8
14	3,5-dimethylbenzenecarboximidamide	9.5430	46.187	57	2.0
15	3-(4-carbamimidoylphenyl)-2-oxopropanoic acid	12.112	64.065	77	5.4
16	3-(3-phenoxypropoxy)benzenecarboximidamide	11.293	56.098	103	3.8
17	4-ethoxybenzenecarboximidamide	10.321	50.234	59	2.8
18	4-methoxybenzenecarboximidamide	10.224	50.465	55	2.7
19	3-(3-carbamimidoylphenyl)-2-oxopropanoic acid	8.5710	41.119	77	1.0
20	5-(3-carbamimidoylbenzyl)naphthalene-2-carboximidamide	11.488	57.913	140	4.0
21	4-(hydroxymethyl)benzenecarboximidamide	10.030	49.446	55	2.5

MR=Molar Refractivity, σ = Shape Index(Order 1), ZM_2= Second Zagreb Index, pIC_{50} = Experimental Activity

BIBLIOGRAPHY

1 Balaban A.T. Ed. **1976** *Chemical application of graph theory*, academic press, London

2 Prabhat Kumar*, Preeti Singh and Jitendra Pratap Singh, **2011** Topological and Energy Descriptor Based QSAR Study of Benzamidine Derivatives as Inhibitor of Tripsin International Journal of Research in Pharmaceutical and Biomedical Sciences, Vol. 3 (4), 296-303,

3 Kumar Nandan,Baidyanath Prasad,Sunita Gupta & Kumar Ranjan, **2011** Statistical Significance of Phathalimide Derivatives,*Indian J. of Mathematics and Mathematical Sci.*, , 7, 51-62

4 Kumar Nandan, Baidyanath Sah, Md. Belal Ahmad & Kumar Ranjan, **2013** QSAR Studies on polychlorinated aromatic compounds using topological descriptors, *Int. J. Pharmaceutical Science and Research*, 4(7), 2691-2695

5 Kumar Nandan, Baidyanath Sah, Md. Belal Ahmad & Kumar Ranjan, **2013** Identification of New HIV-1 Protease Inhibitors By Multiple Linear Regression(MLR) and Physico-Chemical Descriptors, *Int. J. Pharmaceutical Science and Research*, 4(10), 3971-3975

6 Kumar Nandan, Sunita Gupta, Baidyanath Sah, Md. Belal Ahmad & Kumar Ranjan, **2014**, Design of 5,6 –Dihydro-2-pyrones Derivatives as HIV-1 Protease Inhibitors Using Molecular Descriptors, 2(3), 684-690

