IOSR Journal of Applied Chemistry (IOSR-JAC)
e-ISSN: 2278-5736.Volume 8, Issue 11 Ver. I (Nov. 2015), PP 12-20
www.iosrjournals.org

A Semi-empirical based QSAR study of indole*β*- Diketo acid, Diketo acid and Carboxamide Derivativesas potent HIV-1 agent Using Quantum Chemical descriptors

Emmanuel Israel Edache[1*], Ahmed Jibrin Uttu[2], Adebirin Oluwaseye[3], Hassan Samuel[4]and Ajala Abduljelil[1]

[1]Department of Chemistry, Ahmadu Bello University Zaria-Nigeria
[2]Department of Chemistry, Federal University Gashua-Nigeria
[3]Chemistry Advance Lab., Sheba Science and Technology Complex, Abuja-Nigeria
[4]Science Technology, Nigerian Institute of Leather and Science Technology, Samaru-Zaria
Corresponding author: inalegwu334real@yahoo.com, +2348024664850, +2348066776802

Abstract: *In this study, a set of novel synthesized indoleβ- diketo acid, diketo acid and carboxamide derivativeswas investigated by quantitative structure–activity relationship (QSAR) analysis using semi-empirical (PM3) based descriptors. The best molecular descriptors identified were LogP, polar area corresponding to absolute values of electrostatic potential greater than 75 (P-area(75)), Energy (E), Minimum values of electrostatic potential (as mapped onto an electron density surface) (MinEIPot), Polar surface area and Maximum values of electrostatic potential (as mapped onto an electron density surface) (MaxEIPot) that contributed to the anti-HIV activity of the indoleβ- diketo acid, diketo acid and carboxamide derivatives as independent factors. The correlation of these descriptors with their anti-HIV activity increases indicating their importance in studying biological activity. Quantitative structure activity relationship (QSAR) analysis was applied to 37 of the above mentioned derivatives using physicochemical and structural molecular descriptors obtained by the semi-empirical method by employing PM3 basis set. By using the multiple linear regression (MLR) technique several QSAR models have been drown up with the help these calculated descriptors and the anti-HIV activity of indoleβ- diketo acid, diketo acid and carboxamide derivatives. The regression method was used to derive the most significant models as a calibration model for predicting the LogIC50 of this class of compounds. Among the obtained QSAR models presented in the study from the MLR method, statistically the most significant one is the last model with the squared correlation coefficient 0.8932, Q = 3.1854 and F= 27.8644 that could be useful to predict the biological activity of indoleβ- diketo acid, diketo acid and carboxamide derivatives as Potent HIV-1 Drugs.*

***Keywords**: QSAR, β- diketo acid, diketo acid and carboxamide derivatives, MLR, PM3, HIV*

I. Introduction

The HIV epidemic is still a major concern. Infection with the human immunodeficiency virus type-1 (HIV-1) causes increasing destruction of immunity, which finally results in the development of theimmunodeficiency syndrome (AIDS) [1].Up to 19 different drugs have been approved for the treatment of HIV-infected individuals, including 7 nucleoside reverse transcriptase (RT) inhibitors (NRTIs), 1 nucleotide RT inhibitors (NtRTI), 3 non-nucleoside RT inhibitors (NNRTIs), 7 protease inhibitors (PIs) and 1 fusion inhibitor [2]. Virtually every country in the world has seen new infections in 1998, and the epidemic is out of control in many places according to the World Health Organization (WHO) and the Joint United Nations Programme on HIV/AIDS (UNAIDS)[3].Human immunodeficiency virus type1 (HIV–1) Integrase is an enzyme required for viral replication. HIV Integrase catalyzes integration of viral DNA into host genome I two separate but chemically similar reactions known as 3'processing and DNA strand transfer. In 3' processing IN removes a dinucleotide next to conserved cytosine–adenine sequence from each 3'– end of the viral DNA. IN then attaches the processed 3'– end of the viral DNA to the host cell DNA in the strand transfer reaction. As there is no known human counterpart of HIV Integrase, IN is an attractive target for anti–retroviral drug design [4].During the past two decades an increasing number of quantitative structure-activity/property relationship (QSAR/QSPR) models have been studied using theoretical molecular descriptors for predicting biomedical, activity, toxicological and technological properties of chemicals [5]. QSAR/QSPR includes all statistical methods, by which biological activities are related with structural elements, physicochemical properties or fields [6].QSAR studies of anti-HIV activity represent an emerging and exceptionally important topic in the area of computed-aided drug design. Following our interest in this field, our present research aimed to describe the structure-property relationships study on indole *β*- diketo acid, diketo acid and carboxamide derivatives and developed a QSAR model on these compounds with respect to their inhibitory activity (IC_{50}).

II. Materials And Methods

2.1. Experimental data set

The experimental inhibitory concentrations (IC_{50}) in micromolar units of selected indoleβ- diketo acid, diketo acid and carboxamide derivatives against HIV integrase inhibitors are extracted from a recent publication[7, 8]. In Table 1 we provide the experimental activities, and for modeling purposes these values are converted into logarithm units ($-\log_{10}IC_{50}$).

2.2Structural calculations

In the first step, the structures of the 37 investigated molecules were pre-optimized using the current geometry included in the Spartan'14 package version 1.1.2 [9]. In the next step, the minimized structures were refined using the semi-empirical PM3 Hamiltonian also implemented in Spartan'14 version 1.1.2 [9]. For geometry optimization. To display the "real" spatial orientation of the substituents of the indoleβ- diketo acid, diketo acid and carboxamide derivatives. (Fig. 1,2,3, 4and 5 atom numbering as per the IUPAC convention).

Figure 1: Compound 1-8

Figure 2: Compound 9-11

Figure 3: Compound 12-15

Figure 4: Compound 16-23

Figure 5: Compound 24-37

Table 1: Indole β- diketo acid, diketo acid and carboxamide derivatives selected with their activities

Compd No	R	R1	R2	X	Log IC$_{50}$
1	H	H	CH_3	2-CO	0.7780
2	H	H	CH_2CH_3	2-CO	0.2040
3*	OCH_2O		CH_2CH_3	2-CO	0.6990
4	H	H	Bn	2-CO	0.0000
5	OCH_2O		Bn	2-CO	0.3010
6	H	H	CH_3	3-CO	0.3010
7	H	H	CH_2CH_3	3-CO	0.4770
8*	H	H	Bn	3-CO	0.0000

Compd No	R	R1	R2	X	Log IC$_{50}$
9*	OCH_2O		CH_3	2-CO	1.6990
10	OCH_2O		CH_2CH_3	2-CO	1.8130
11	OCH_2O		CH_3	3-CO	1.7780

Compd No	R1	R2	R3	IC50
12	4’-Cl	-	-	0.000
13	3’-F	-	-	0.602
14	-	4-OCH_3	-	0.824
15*	-	3-OCH_3	-	0.854

Compd No.	R$_1$	R$_2$	R$_3$	LogIC$_{50}$
16*	4-F	-	-	1.000
17	H	-	-	0.638
18	2-Cl	-	-	0.432
19	4-Cl	-	-	0.420
20	4-F, 3-Cl	-	-	1.398
21*	4-F	CN	-	1.699
22	4-F	Br	-	1.523
23	4-F	I	-	1.699

Compd No.	R$_1$	R$_2$	R$_3$	LogIC$_{50}$
24*	$NHCOCH_3$	CH_3	4-fluorotoluene	2.155
25	NH-SO_2-CH_3	CH_3	4-fluorotoluene	2.097
26	NHCO-N(CH3)$_2$	CH_3	4-fluorotoluene	1.745
27*	NHSO2-N(CH3)$_2$	CH_3	4-fluorotoluene	1.921
28	NHCOCO-N(CH3)$_2$	CH_3	4-fluorotoluene	2.000
29	NHCOCO-OCH_3	CH_3	4-fluorotoluene	1.824
30	NHCOCO-OH	CH_3	4-fluorotoluene	2.398
31	N(CH3)COCO-N(CH3)$_2$	CH_3	4-fluorotoluene	1.824
32	NHCO-pyridazine	CH_3	4-fluorotoluene	1.824
33	NHCO-pyrimidine	CH_3	4-fluorotoluene	2.155
34	NHCO-oxazole	CH_3	4-fluorotoluene	2.155
35*	NHCO-thiazole	CH_3	4-fluorotoluene	2.097
36	NHCO-1H imidazole	CH_3	4-fluorotoluene	2.222
37*	NHCO-1,3,4-oxadiazole	CH_3	4-fluorotoluene	1.8224

*Test set

2.3.Statistical Analyses

2.3.1. Multiple Linear Regression (MLR)

Multiple linear regression analysis of molecular descriptors was carried out using the Microsoft Excel for Windows. Multiple linear regression (MLR) is a method used to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to the observed was employed to correlate the binding affinity and molecular descriptors [10]. This method has been widely applied in many QSAR studies, and has upheld to be a useful linear regression method to build QSAR models that may explore forthright the properties of the chemical structure in combination with its ability of inducing a pharmacological response [11]. The advantage of MLR is its simple method and easily interpretable mathematical expression.

In the equations, the figures in the parentheses are the standard errors of the regression coefficients, N is the total number of compounds in the data set, $N_{training}$ is the number of compound in the training set, $N_{test}$is the number of compound in the test set, R is the correlation coefficient, R^2 is the determination coefficient, Q^2 is the leave many out(LOO) cross validated, F is the significance test (F-test), *RMSECV* is the root mean square error of cross validation(training set), *RMSEP* is the root mean square error of prediction(external validation set) Se is the standard error of estimate represents standard deviation which is measured by the error mean square, which expresses the variation of the residuals or the variation about the regression line. Therefore, standard deviation is an absolute measure of quality of fit and should have low value for the regression to be significant. PRESS is the predictive residual sum of the squares, R^2_{pred} is the correlation coefficient of multiple determination (external validation set). F-test values are for all equation statistically significant at 95% level probability.

R^2, Q^2, *RMSECV*, *Q*, and *RMSEP* of a model can be obtained from:

$$R^2 = 1 - \frac{\Sigma(Y_{obs} - Y_{cal})^2}{\Sigma(Y_{obs} - \bar{Y})^2} \underline{\hspace{8cm}} (1)$$

R^2 is a measure of explained variance. Each additional X variable added to a model increases R^2. R^2 is a relative measure of fit by the regression equation. Correspondingly, it represents the part of the variation in the observed data that is explained by the regression.

Calculation of Q^2 (cross-validated R^2) is called as internal validation.

$$Q^2 = 1 - \frac{\Sigma\left(Y_{obs} - Y_{pred}\right)^2}{\Sigma(Y_{obs} - \bar{Y})^2} \underline{\hspace{6cm}} (2)$$

$$RMSECV = \sqrt{\Sigma \frac{\left(Y_{obs} - Y_{pred}\right)^2}{N}} \underline{\hspace{5cm}} (3)$$

Where, Y_{obs} , Y_{pred} and *N* indicate observed, predicted activity values and number of samples in the training set respectively and $\bar{Y}$ indicates mean activity value. A model is considered acceptable when the value of Q^2 exceeds 0.5. External validation or predictability of the models are performed by calculating predictive R^2(R^2_{pred}).

$$R^2_{pred} = 1 - \frac{\Sigma\left(Y_{pred\ (Test)} - Y_{Test}\right)^2}{\Sigma\left(Y_{(Test)} - \bar{Y}_{training}\right)^2} \underline{\hspace{5cm}} (4)$$

$$RMSEP = \sqrt{\Sigma \frac{\left(Y_{pred\ (Test)} - Y_{Test}\right)^2}{M}} \underline{\hspace{5cm}} (5)$$

Where,$Y_{pred\,(Test)}$, $Y_{(test)}$ and *M* indicate predicted, observed activity values and number of samples respectively of the test set compounds and $\bar{Y}_{training}$ indicates mean of observed activity values of the training set. For a predictive QSAR model, the value of R^2_{pred} should be more than 0.5 [6, 12, 13].

However, this is not a sufficient condition to guarantee that the model is really predictive. It is also recommended to check:

1) the slope K or K' of the linear regression lines between the observed activity and the predicted activity in the external validation, where the slopes should be 0.85 ≤ K≤1.15 or 0.85≤K'≤1.15 and

2) the absolute values of the difference between the coefficients of multiple determination, R_o^2 and $R_o^{'2}$ smaller than 0.3 [14].

Q is the quality factor [15, 16]. The quality factor Q is used to decide the predictive potential of the models. The quality factor Q is defined as the ratio of correlation coefficient to the standard error of estimation. We found it to be a good parameter to explain the predictive potential of the models proposed by us. The higher the value of Q the better is the predictive potential of the models [15-17].

$$Q = \frac{R}{SE} \underline{\hspace{8cm}} (6)$$

2.4Y-Randomization Test

The Y-randomization test is a widely used technique that displays the robustness of a QSAR model, being a measure of the model over[18]. The biological activity is randomly shuffled and a new QSAR model is developed using the same descriptors. The procedure is repeated several times and a new QSAR model are developed. The obtained MLR model (after 500 randomizations) must have low R^2 and Q^2_{cv}values. If the opposite happens then an acceptable QSAR model cannot be obtained for the specific modeling and data [19].

III. Results And Discussion

In the present study authors tried to develop best QSAR model to explain the correlation between the physicochemical parameters and HIV integrase inhibitory activity of indoleβ-diketo acid, diketo acid and carboxamide derivatives. After regression analysis on the Excel software, the best equation received for 3' processing inhibitory activity was-

4-variables,

Model-1:

$$pIC_{50} = -0.4107(\pm 0.3258) - 0.1423(\pm 0.0476) LogP - 0.1441(\pm 0.0653) Dipole\ M. - 0.0014(\pm 0.0004) Energy + 0.0104(\pm 0.0022) P - Area(75)$$

$$N = 37, N_{training} = 27, N_{test} = 10, R = 0.9218, R^2 = 0.8498, R_{adj} = 0.8225, SE = 0.3354, F = 31.1188, PRESS = 1.4038, RMSECV = 0.3028, RMSEP = 0.3748, Q = 2.7484, R^2_{pred} = 0.7100,$$

$$K = 0.7673\ and\ K' = 0.9904, where\ 0.85 \leq K\ or\ K' \leq 1.15, /R_o^2 - R_0'^2/ = 0.05 < 0.3, \frac{(R^2 - R_o^2)}{R^2} = 0.1056 > 0.1,$$

Model-2:

$$pIC_{50} = -0.6112(\pm 0.3384) - 0.1715(\pm 0.0453) LogP - 0.1228(\pm 0.0643) Dipole\ M. - 0.0016(\pm 0.0004) Energy + 0.0150(\pm 0.0032) Acc. P - Area(75)$$

$$N = 37, N_{training} = 27, N_{test} = 10, R = 0.9201, R^2 = 0.8465, R_{adj} = 0.8186, SE = 0.3391, F = 30.3345, PRESS = 1.2874, RMSECV = 0.3061, RMSEP = 0.3588, Q = 2.7134, R^2_{pred} = 0.7340, K = 0.7496\ and\ K' = 1.0024, where\ 0.85 \leq K\ or\ K' \leq 1.15, /R_o^2 - R_0'^2/ = 0.0174 < 0.3, \frac{(R^2 - R_o^2)}{R^2} = 0.1123 > 0.1$$

Model-3:

$$pIC_{50} = -0.4266(\pm 0.3381) - 0.1421(\pm 0.0484) LogP - 0.1400(\pm 0.0665) Dipole\ M. - 0.0014(\pm 0.0004) Energy_{(aq)} + 0.0097(\pm 0.0022) P - Area(75)$$

$$N = 37, N_{training} = 27, N_{test} = 10, R = 0.9192, R^2 = 0.8449, R_{adj} = 0.8167, 0.8167, SE = 0.3408, F = 29.9615, PRESS = 1.3553, RMSECV = 0.3077, RMSEP = 0.3681, Q = 2.697, R^2_{pred} = 0.7200, K = 0.7628\ and\ K' = 0.9882, where\ 0.85 \leq K\ or\ K' \leq 1.15, /R_o^2 - R_0'^2/ = 0.0338 < 0.3, \frac{(R^2 - R_o^2)}{R^2} = 0.1056 > 0.1$$

5-variables

Model-4:

$$pIC_{50} = -0.0396(\pm 0.3593) - 0.1453(\pm \pm 0.0442) LogP - 0.1788(\pm 0.0629) Dipole\ M. - 0.0016(\pm 0.0004) Energy_{(aq)} - 0.0131(\pm 0.0056) PSA + 0.0166(\pm 0.0036) P - Area(75)$$

$$N = 37, N_{training} = 27, N_{test} = 10, R = 0.9364, R^2 = 0.8768, R_{adj} = 0.8474, SE = 0.3110, F = 29.8815, PRESS = 1.3487, RMSECV = 0.2742, RMSEP = 0.3672, Q = 3.0109, R^2_{pred} = 0.7214,$$

$$K = 0.8031\ and\ K' = 0.9953, where\ 0.85 \leq K\ or\ K' \leq 1.15, /R_o^2 - R_0'^2/ = 0.0779 < 0.3, \frac{(R^2 - R_o^2)}{R^2} = 00884 < 0.1,$$

Model-5:

$$pIC_{50} = -0.0335(\pm 0.3498) - 0.1464(\pm 0.0442) LogP - 0.1808(\pm 0.0629) Dipole\ M. - 0.0016(0.0004) Energy - 0.0119(\pm 0.0056) PSA + 0.0168(\pm 0.0036) P - Area(75)$$

$N = 37, N_{training} = 27, N_{test} = 10, R = 0.9363, R^2 = 0.8767, R_{adj} =, SE = 0.3111, F = 29.8571, PRESS = 1.3047, RMSECV = 0.2743, RMSEP = 0.3612, Q = 3.010, R^2_{pred} = 0.7305, \quad K = 0.8098\ and\ K' = 0.9826, where\ 0.85 \leq K\ or\ K' \leq 1.15, /R^2_o - R'^2_0/ = 0.0652 < 0.3, \frac{(R^2 - R^2_o)}{R^2} = 0.0924 < 0.1$

6-variable
Model-6:

$$pIC_{50} = 2.3804(\pm 0.9980) - 0.3449(\pm 0.0620)LogP + 0.0188(\pm 0.0036)P - Area(75) - 0.0018(\pm 0.0004)Energy + 0.0127(\pm 0.0037)MinEIPot - 0.0287(\pm 0.0082)PSA + 0.0084(\pm 0.0040)MaxEIPot$$

$N = 27, N_{training} = 27, N_{test} = 10, R = 0.9451, R^2 = 0.8932, R_{adj} =, SE = 0.2967, F = 27.8644, PRESS = 0.9168, RMSECV = 0.2554, RMSEP = 0.3028, Q = 3.5043, R^2_{pred} = 0.8106,$

$$K = 0.85\ and\ K' = 0.9838, where\ 0.85 \leq K\ or\ K' \leq 1.15, /R^2_o - R'^2_0/ = 0.0225 < 0.3, \frac{(R^2 - R^2_o)}{R^2} = 0.0673 < 0.1,$$

Where IC_{50} is the molar concentration of the drugleading to 50% inhibition of enzyme Integrase, LogP = Partition coefficient, DM. = Dipole moment, E = Energy, P-Area(75)=Polar area corresponding to absolute values of the electrostatic potential greater than 75, Acc. P-Area(75) = Accessible polar area corresponding to absolute values of the electrostatic potential greater than 75. PSA= Polar surface area, MinEIPot = Minimum values of the electrostatic potential (as mapped onto an electron density surface), E(aq) = Aqueous energy, In the above equations $N_{training}$ is the number of compounds used to derive the model and values in parentheses are the 95% confidence limit of respective coefficient, R = correlation coefficient, SE= Standard error of estimation F = F–ratio between variances of calculated and observed value, R^2squared correlation coefficient, Q^2_{cv} = cross validated squared correlation coefficient, RMSECV= Root mean square error of cross-validation (training set) , RMSEP = Root mean square error of prediction (external validation set), Q = quality factor. We extended our study for eight parametric correlations as they are permitted for a data set of 37 compounds in accordance with the lower limit of rule of thumb. The calculated and predicted (LOO) activities of the compounds by the above models are shown in Table 2.

Model–1 shows good correlation coefficient (*R*) of 0.9218 between descriptors (LogP, DM, E and P-Area (75))explains 84.98% variance in biological activity. This model also indicates statistical significance >99.9% with F values F = 31.1188. Cross validated squared correlation coefficient of this model was 0.7784, which shows the good internal prediction power of this model (Fig. 6). The results of the validation steps show that model 1 cannot be classified as a good model, since according to the criteria used, it suffers the last criteria $\frac{(R^2 - R^2_o)}{R^2} = 0.1056$, which is greater than 1.

Model–4 shows good correlation coefficient (*R*) of 0.9364 between descriptors (LogP, DM, $E_{(aq)}$, PSA and P-Area(75)) and HIV integrase integration inhibitory activity. Squared correlation coefficient (R^2) of 0.8768 explains 87.68% variance in biological activity. This model also indicates statistical significance >99.9% with F values F = 29.8815. Cross validated squared correlation coefficient of this model was 0.8014, which shows the good internal prediction power of this model (Fig. 7), and cannot be classified as a good model, since the linear regression line between the predicted activity and the observed activity K is less than 0.85.

Model–6 shows good correlation coefficient (*R*) of 0.9451 between descriptors (LogP, P-Area (75), E, MinEIPot PSA and MaxEIPot) and HIV integrase integration inhibitory activity. Squared correlation coefficient (R^2) of 0.8932explains 89.32% variance in biological activity. This model also indicates statistical significance >99.9% with F values F = 27.8644. Cross validated squared correlation coefficient of this model was 0.8042, which shows the good internal prediction power of this model (Fig. 8). Therefore, the results of validations steps show that the model can be classified as a good model, since according to the criteria used, it has good internal quality, it is robust, it does not suffer from chance correlation at random, and it shows a good capacity of external predictions.
The predictive ability of model – 1, 4 and 6 was also confirmed by external validation (model – 2, 3 and 5 respectively). The R^2_{pred} value of the selected model is greater than the prescribed value (R^2_{pred}>0.5). The QSAR model for training set of 3’ processing inhibitionactivity using model –6: To further evaluate the significance of the developed model, it needs to undergo a stability test. For this, standard error of estimate and root mean squares are used. The values of standard error (*SEE*), root mean square error cross validation (*RMSECV*), and root mean squares error prediction (*RMSEP*) in this model are 0.2967, 0.2743 and 0.3612, respectively, which

further adds to the statistical significance of the developed model. In addition, the low values of *SEE*, *RMSE* and *RMSEP* indicate that the developed QSAR model is stable for predicting unknown compounds in the test set. The high and positive value of quality factor (*Q*) for this QSAR's model suggest its high predictive power and lack of over fitting.

Table 2. The calculated and predicted (LOO) activities of the compounds by the above models.

Observation	Predicted LogIC50	Residuals	Predicted LogIC50	Residuals	Predicted LogIC50	Residuals
IN01	0.4065	0.37154	0.271858	0.506142	0.481003	0.296997
IN02	0.3925	-0.1885	0.242167	-0.03817	0.337824	-0.13382
IN03*	1.1949	-0.4959	1.0046	-0.3056	1.2685	-0.5696
IN04	0.356	-0.356	0.243665	-0.24366	0.167002	-0.167
IN05	0.9062	-0.6052	0.685643	-0.38464	0.801561	-0.50056
IN06	0.5505	-0.2495	0.632063	-0.33106	0.720515	-0.41951
IN07	0.4441	0.0329	0.476553	0.000447	0.46276	0.01424
1N08*	0.0064	-0.0064	-0.1159	0.1156	-0.3034	0.3034
IN09*	1.3274	0.3716	1.4398	0.2592	1.792	-0.093
IN10	1.2919	0.52112	1.376575	0.436425	1.588633	0.224367
IN11	1.5068	0.27124	1.763859	0.014141	1.68343	0.09457
IN12	0.3173	-0.3173	0.37661	-0.37661	0.181454	-0.18145
IN13	0.4781	0.12393	0.579746	0.022254	0.701157	-0.09916
IN14	0.3602	0.46376	0.265339	0.558661	0.146102	0.677898
IN15*	0.3258	0.5281	0.3231	0.5309	0.5803	0.2737
IN16*	1.2476	-0.2476	1.3656	-0.3656	1.3903	-0.3903
IN17	0.649	-0.011	0.596745	0.041255	0.406269	0.231731
IN18	0.6597	-0.2277	0.608783	-0.17678	0.547389	-0.11539
IN19	0.8296	-0.4096	0.8373	-0.4173	0.685139	-0.26514
IN20	1.3523	0.04572	1.513279	-0.11528	1.634337	-0.23634
IN21*	1.0581	0.6409	1.1593	0.5397	1.8415	-0.1425
IN22	1.2998	0.22315	1.513251	0.009749	1.522865	0.000135
IN23	1.1955	0.50345	1.408262	0.290738	1.355348	0.343652
IN24*	1.9708	0.1842	2.0209	0.1314	2.0169	0.1381
IN25	2.3444	-0.2474	2.418666	-0.32167	2.213856	-0.11686
IN26	1.6202	0.12483	1.558166	0.186834	1.588876	0.156124
IN27*	1.7696	0.1514	1.5562	0.3648	2.2207	-0.2997
IN28	2.2597	-0.2597	2.164327	-0.16433	2.0459	-0.0459
IN29	2.2494	-0.4254	2.082152	-0.25815	2.097345	-0.27334
IN30	2.5221	-0.1241	2.305139	0.092861	2.252666	0.145334
IN31	1.6251	0.19894	1.734992	0.089008	1.902217	-0.07822
IN32	1.6482	0.17576	1.617584	0.206416	1.499226	0.324774
IN33	2.1296	0.02543	2.197025	-0.04202	2.206225	-0.05122
IN34	2.1782	-0.0232	2.199737	-0.04474	2.250837	-0.09584
IN35*	1.676	0.421	1.6583	0.4387	1.7373	0.3597
IN36	1.8592	0.36278	1.762515	0.459485	1.952065	0.269935
IN37*	1.6379	0.1862	1.4714	0.3526	1.9012	-0.0772

*Test set

Figure 6: Plot of predicted IC50 vs observed IC50 of training set

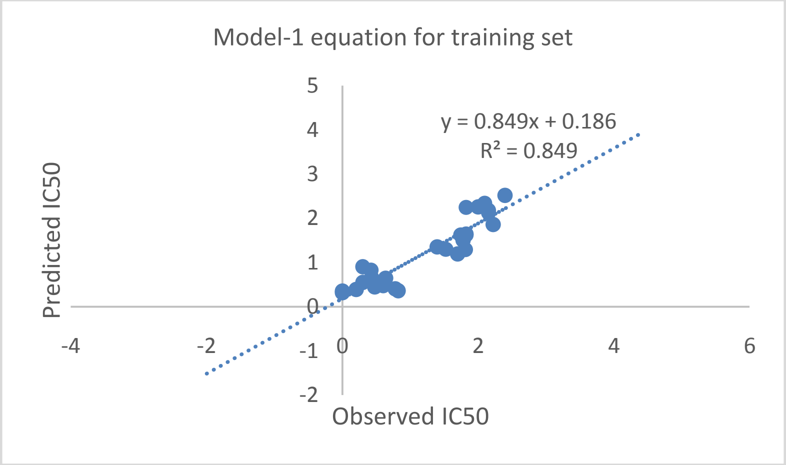


Figure 7: Plot of predicted IC50 vs observed IC50 of training set

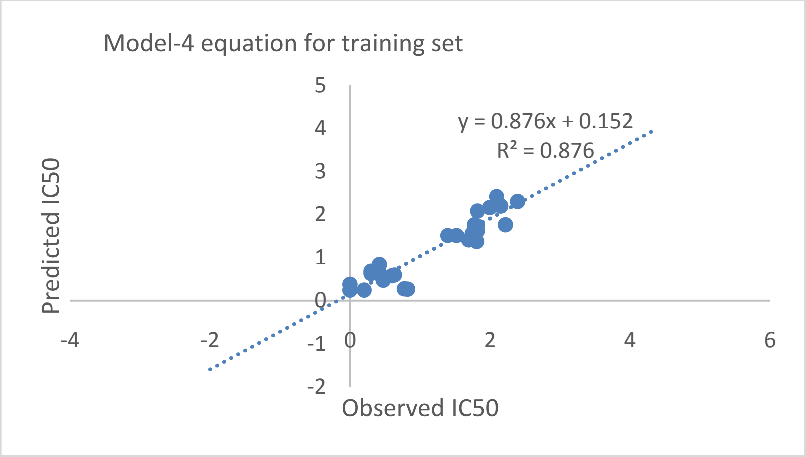
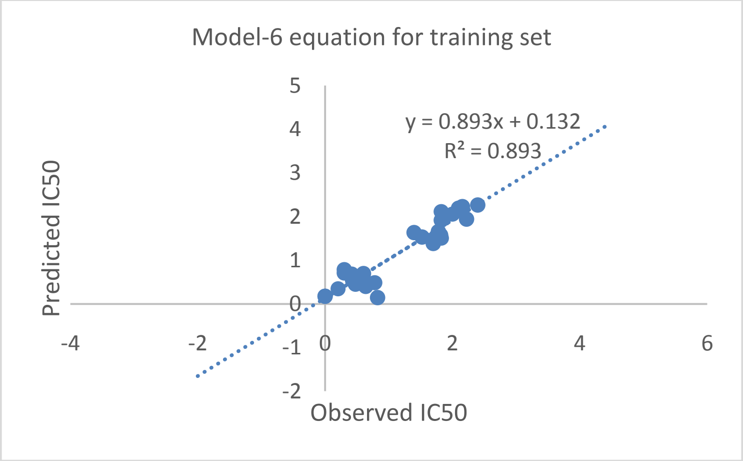


Figure 8: Plot of predicted IC50 vs observed IC50 of training set



IV. Conclusions

In the study, a QSAR model for the activity indoleβ- diketo acid, diketo acid and carboxamide derivatives was successfully developed based on various semi-empirical PM3 descriptors. Significant regression equations were obtained by MLR for 37 indoleβ- diketo acid, diketo acid and carboxamide compounds according to their anti-HIV activity. The QSAR model indicates that the quantum chemical descriptors such as polar surface area, Log P, Energy, minimum and maximum values of electrostatic potential (as mapped onto an electron density surface) play an important role for the anti-HIV activity and pEC50 of indoleβ- diketo acid, diketo acid and carboxamide derivatives. The results of the present study may be useful on the designing of more potent β- diketo acid, diketo acid and carboxamide analogues as anti-HIV agents.

Reference

[1] I. Daniela, C. Luminita, F. Simona, and M. Mircea, A quantitative structure–activity relationships study for the anti-HIV-1 activities of 1-[(2-hydroxyethoxy)methyl]-6--(phenylthio)thymine derivatives using the multiple linear regression and partial least squares methodologies. *Journal of the Serbian Chemical Society 78 (4)*, 2013, 495–506.

[2] B. Jan, S. Miguel, E. De Clecq., et al. Pyridine N-oxide derivatives: unusual anti-HIV compounds with multiple mechanisms of antiviral action. *Journal of Antimicrobial Chemotherapy 55,* 2005, 135-138.

[3] O.S Pedersen and E.B Pedersen, Non-nucleoside reverse transcriptase inhibitors: the NNRTI boom. *Antiviral Chemistry & Chemotherapy 10*, 1999, 285–314.

[4] K.K Sahu, V. Ravichandran, K.J Prateek, S. Simant, et al. QSAR Analysis of Chicoric Acid Derivatives as HIV–1 Integrase Inhibitors. *Acta Chim. Slov.***,** *55,* 2008, 138–145.

[5] R. Hadanu and Syamsudin. Quantitative Structure-Activity Relationship Analysis of Antimalarial Compound of Mangostin Derivatives Using Regression Linear Approach. *Asian Journal of Chemistry 25 (11),* 2013, 6136-6140.

[6] R. Veerasamy, H. Rajak, A. S. Jain, C.P. Sivadasan, Varghese and R.K. Agrawal. Validation of QSAR models-Strategies and importance. *International journal of drug design and discovery 2(3),* 2011, pp 511-519.

[7] M. Sechi, Design and synthesis of novel indole a-diketo acid derivatives as HIV-1 integrase inhibitors. *J Med Chem., 47,* 2004, 5298–5310

[8] A.O. Adebimpe, R.C. Dash and M.E.S. Soliman, QSAR Study on Diketo Acid and Carboxamide Derivatives as Potent HIV-1 Integrase Inhibitor. *Letters in Drug design & Discovery, volume 11, No. 5*, 2014, 000-000

[9] Wavefunction, Inc. Spantan’14, version 1.1.2, Irvin, California, USA, 2014.

[10] N.S. Sapre, T. Bhati, S. Gupta, N. Pancholi, U. Raghuvanshi, D. Dubey, V. Rajopadhyay, N. Sapre, Computational modeling studies on anti-1 non-nucleoside reverse transcriptase inhibition by dihydroalkoxybenzyloxopyrimidines analogues: an electropological atomistic approach. *Journal of Biophysical Chemistry, 2(3),* 2011, 361-372.

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

[12] R. Veerasamy, S. Ravichandran, A. Jain, H. Rajak, R.K. Agrawal, QSAR studies on novel anti-HIV agents using FA-MLR, FA-PLS and PCRA techniques. *Digest Journal of Nanomaterials and Biostructures, vol.4,* 2009, 823-834.

[13] Doreswany and C.M. Vastrad, Predictive comparative QSAR analysis of sulfathiazole analogues as mycobacterium tuberculosis H37RV. *Journal of advanced bioinformatics applications and research, 3(3),* 2012, 379-390.

[14] L.F. Motta and W.P. Almeida, Quantitative structure-activity relationship (QSAR) of a series of Ketone derivatives as anti-candida albicans. *International Journal of Drug Discovery, vol. 3(2),* 2011, 100-117.

[15] L. Pogliani, “Structure property relationships of amino acids and some dipeptides,” *Amino Acids, vol. 6(2),* 1994, 141–153.

[16] L. Pogliani, “Modeling with special descriptors derived from a medium-sized set of connectivity indices,” *Journal of PhysicalChemistry, vol. 100(46),* 1996, 18065–18077.

[17] S. Chaterjee, A.S. Hadi, B. Price,*Regression analysis by examples*, 3rd Ed. Wiley; New York, 2000

[18] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being Earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb.Sci.; vol, 22,* 2003, 69-77.

[19] A. Antreas, M. Georgia, S. Haralambos, A.K.Panayiotis, M. John and I.M Olga, A novel QSAR model for evaluating and predcting the inhibition activity of dipeptidyl aspartyl fluomethylketones. *Journal of QSAR and Combinational Science, vol. 25, No. 10,* 2006, 928-935.