Turkish Computational and Theoretical Chemistry

Turkish Comp Theo Chem (TC&TC)

Volume(Issue): 4(2) – Year: 2020 – Pages: 49-58

e-ISSN: 2602-3237

https://doi.org/10.33435/tcandtc.624156

Received: 25.09.2019 ***Accepted***: 14.06.2020 *Research Article*

In Silico study for investigating and predicting the activities of 7-Hydroxy-1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]pyridine-4-carboxylate Derivatives as Potent Anti-HIV Agents

Ahanonu Saviour Ugochukwu [1], **Gideon Adamu Shallangwa, Adamu Uzairu**

Department of Chemistry, Ahmadu Bello University, Zaria- Nigeria

Abstract: In this study a QSAR was carried out on a data set of 7-Hydroxy-1,3- dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]pyridine-4-carboxylate Derivatives to investigate their activities on HIV-1. Genetic Function Algorithm(GFA) and Multi Linear Regression Analysis (MLRA) were used to select the optimum descriptors and to generate the correlation QSAR model that relate their activities against HIV with the molecular structures of the derivatives. After the internal validation, the model was found to have a squared correlation coefficient (R^2) of **0.9334**, adjusted squared correlation coefficient (R^2_{adj}) of **0.9134** and leave one out cross validated coefficient (LOO- Q^2_{cv}) value of **0.8604**. The external validation (R^2_{pred}) set used for confirming the predictive power of the model was **0.8935**. Y randomization value of **0.6463** was used to confirm the robustness of the model**.** The robustness and stability of the model obtained by validation of the test set also confirmed that the model can be used to design and synthesize other 7-Hydroxy-1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]pyridine-4-carboxylate Derivatives with improved Anti- HIV activities.

Keywords: *QSAR,7-Hydroxy-1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]pyridine-4-carboxylate Derivatives, Y Randomization, HIV, MLR.*

1. Introduction

HIV/AIDS epidemic which was reported in the United States in the spring of 1981 is today still a major concern. Human immunodeficiency infection is the chief reason for AIDS [1]. Acquired Immunodeficiency syndrome (AIDS) and human immunodeficiency virus (HIV) infection are global health hazards with huge social, economic and ethical consequences. [2,3]. Human immune virus type 1(HIV-1) decrease the immunity of the body and consequently results in Acquired Immune deficiency syndrome (AIDS). The drugs which were screened and approved for the treatment of HIV can only dramatically slow down the disease's progress and also prevent secondary infections and complications. A person living with HIV has a severe reduction in CD4+T cells which means the person develops a very weak immune system and becomes vulnerable to contracting life threatening infections such as pneumocystis carinii pneumonia which eventually results in AIDS.

Once an HIV particle enters a person's body, it binds to the surface of a target cell (CD4+T cell) by shedding its own viral envelope, allowing the HIV particle to release an HIV ribonucleic acid (RNA) chain into the cell which is then converted into deoxyribonucleic acid (DNA). The HIV DNA enters the cell's nucleus and is copied onto the cell's chromosomes. As HIV infection progress, the CD4 + T cell population declines slowly and the infected individual becomes progressively more susceptible to certain opportunistic infections and neoplasms [4].

QSAR is a mathematical model relating the biological activity measurements of a set of chemical compounds to the variation in their chemical structure. It is used to predict the biological effects of yet untested chemical compounds. The application of quantitative structure Activity Relationship (QSAR) technique to this problem has potential to minimize effort and time required to discover new compounds or to improve current ones in terms of their efficiency

[1] Corresponding Authors

e-mail: favour_saviour@yahoo.com

[5]. QSAR has become inexorably embedded as an essential tool in the pharmaceutical industry. The fundamental assumption of QSAR is that variations in the biological activities of a series of chemicals that target a common mechanism of action are correlated with variation in their structural, physical and chemical properties. It provides a discussion of several qualitative approximations of the structure activity relationship to search the preferred conformations to establish correlations between structural parameters and the various properties of the investigated

macromolecules and improving the conception of new therapeutic drugs [6].

Early QSARs comprised of moderately little number of molecules being utilized to infer a basic direct comparison to foresee the following particle in the arrangement to be combined. Presently, it correlates the response of chemicals (activity/property) with their structural and physicochemical information in the form of numerical quantities, i.e., descriptors which has made people to be very interested.

The aim of this research is to develop QSAR model using Genetic Function Algorithm (GFA) for variable selection of descriptors and multiple linear regression (MLR) method for predicting the activity of 7-Hydroxy-1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]pyridine-4-carboxylate Derivatives as potent anti- HIV.

2. Materials and Methods

The five guidelines adopted by the OECD to obtain the validated QSAR model are as follows:

1. A defined end point
2. an unambiguous algorithm
3. a defined domain of applicability
4. appropriate measures of goodness of fit, robustness and predictivity
5. a mechanistic interpretation, if possible

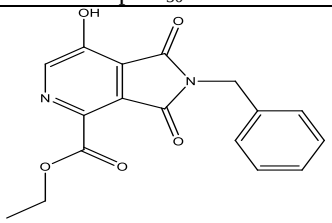
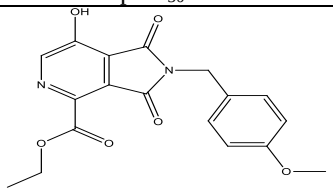
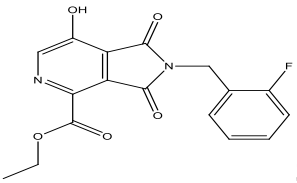
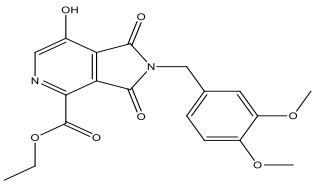
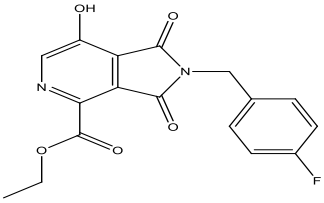
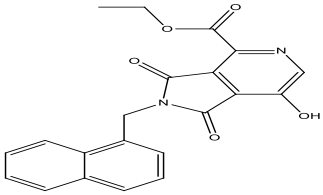
The OECD principles are the best possible outline of the essential points to be addressed while developing reliable and reproducible QSAR models [7].

The data set of 7-Hydroxy-1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]pyridine-4-carboxylate Derivatives as potent anti-HIV that were used in this research was selected from the literature [8].The derivatives containing 20 compounds with well-defined activities selected for QSAR study. Some compounds were excluded because their activities were not well-defined.

The biological activity data in the form EC50 (half –maximal effective concentration) were reported in □M and were converted into pEC50 = (-logEC50) ------- (1).

It represents the molar concentration in mol/L of an agonist that produces half of the maximal possible effect of the agonist. The dataset of 20 compounds were divided into a training set of 14 molecules to generate the QSAR model and a test set of 6 molecules to validate the quality of the generated model. The Table 1 below shows the pEC50 and structures of the compounds used in QSAR.

Table 1. Structures and pEC_{50} values of the lead compounds

S/N	Structures/pEC_{50}	S/N	Structures/pEC_{50}
1.	 4.1981	6.	 4.1600
2.	 3.5976	7.	 3.5392
3.	 4.0511	8.	 4.3159

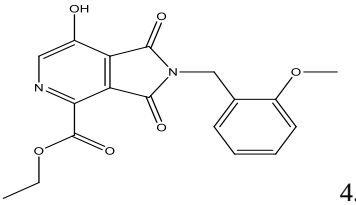
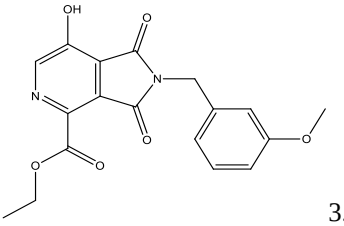
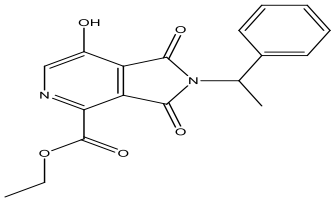
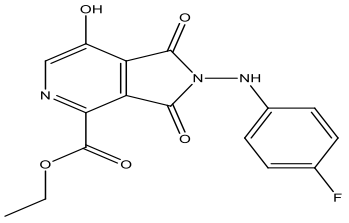
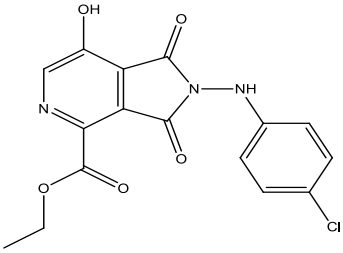
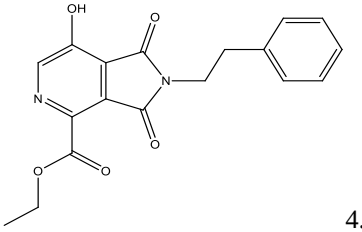
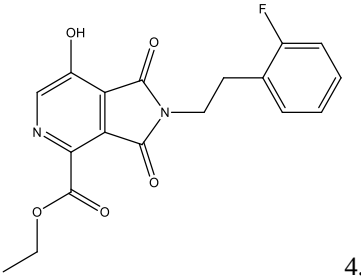
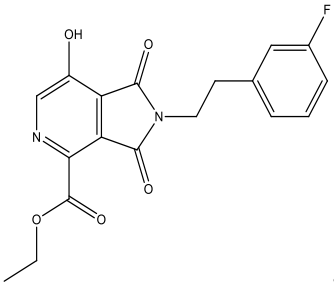
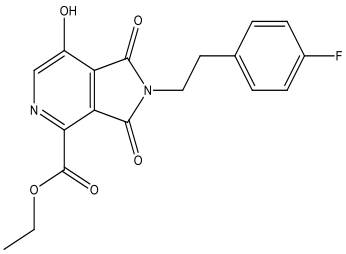
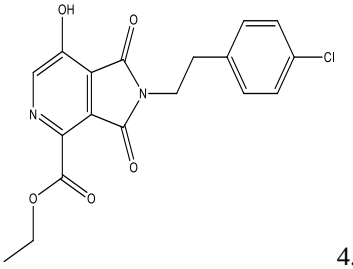
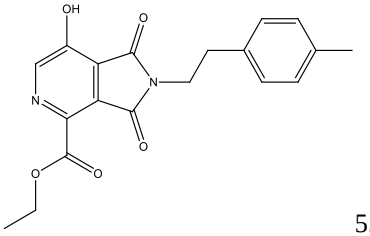
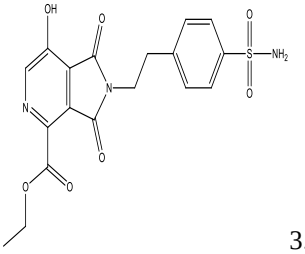
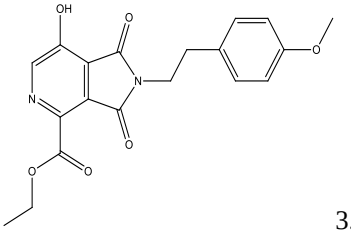
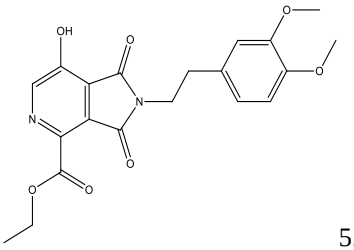
4.		4.2397	
5.		3.6713	
9.		4.1625	
10.		4.4866	
11.		4.8359	
12.		4.2525	
13.		4.3484	
14.		4.2818	
15.		5.7825	
16.		4.8193	
17.		5.6308	
18.		3.8651	
19.		3.7854	
20.		5.0386	

Table 2. Biological activities of training set compounds

Compound number	X	R^1	R^2	$logEC_{50}$
2	CH_2	2-FPh	OEt	3.5976
3	CH_2	4-FPh	OEt	4.0511
5	CH_2	3-OMePh	OEt	3.6713
8	CH_2	Naphthalene-1-yl	OEt	4.3159
9	$CH(cH_3)$	Ph	OEt	4.1625
11	NH	4-ClPh	OEt	4.8359
12	CH_2CH_2	Ph	OEt	4.2525
13	CH_2CH_2	2-FPh	OEt	4.3484
14	CH_2CH_2	3-FPh	OEt	4.2818
15	CH_2CH_2	4-FPh	OEt	5.7825
17	CH_2CH_2	4-MePh	OEt	5.6308
18	CH_2CH_2	$4\text{-}SO_2NH_2$	OEt	3.8651
19	CH_2CH_2	4-OMePh	OEt	3.7854
20	CH_2CH_2	3,4-OMePh	OEt	5.0386

Table 3. Biological activities of test set compounds

Compound number	X	R^1	R^2	$logEC_{50}$
1	CH_2	Ph	OEt	4.1981
4	CH_2	2-OMePh	OEt	4.2397
6	CH_2	4-OMePh	OEt	4.1600
7	CH_2	3,4-OMePh	OEt	3.5392
10	NH	4-FPh	OEt	4.4866
16	CH_2CH_2	4-ClPh	OEt	4.8193

2.1. Molecular modelling and generation of molecular descriptors

Molecular structures of the dataset compounds were drawn using chemdraw ultra version 12.0.2 software to create the three dimensional structure. These compounds were optimised using density function Theory (DFT) with basis (B3LYP 6-31G*) after energy minimization. The optimised structures were transferred to PaDEL- Descriptor - software to calculate various physiochemical parameters like thermodynamic, steric and electronic descriptors. PaDEL-Descriptor is a freely available open source software to calculate chemical descriptors and fingerprints [9].

PaDEL-Descriptor is the best choice because it has a user-friendly interface and can run all major platforms, which makes it easy for modellers to calculate descriptors during their model development [9].

2.2. Descriptors transformation

In QSAR, bias has to be over comed because model is usually biased toward descriptor with high positive and negative values [10]. To overcome this the molecular descriptor of the training set data were transformed by normalization [11] using the equation below

$$X^n = \frac{x - x_{max}}{x_{max} - x_{min}} \quad \text{..... (2)}$$

In the equation above, X^n is the normalised descriptors, X_{max} is the maximum value in a descriptor column, X_{min} is the minimum value in the descriptor column and X is the original descriptor.

2.3. Data pretreament

The data from PaDEL-Descriptor were transferred to data pretreatment software from DTC Lab for pretreatment in order to remove redundant and unwanted descriptors. All descriptors column with constant column or near constant values were deleted to remove the redundant descriptors. Only one descriptor among those showing high mutual intercorrelation should be retained [12].

The principle of elimination of redundant descriptors was based on absolute correlation limit between them which was set to: 0.8000, 0.8500, 0.9000, 0.9500, 0.9700, 0.9900, 0.9950, 0.9970, 0.9990, 0.9999, and 1.0000 [13].

2.4. Data Division

After data pretreament, the pretreated data was transferred to data division software also from DTC Lab to divide it into training and test set by using Kennard and Stone's Algorithm (Kennard and Stone). Selection of training and test set was carried out in such a manner that compounds of the test set resembled compounds of the training set in multidimensional descriptor space and all representative compounds of training set resembled compounds of test set [14]. Thus, a test set was a true demonstrative of a training set. This was achieved by randomly setting aside test compounds with distributed biological data [15].

2.5. Selection of best descriptors

Genetic function Algorithm (GFA) incorporated in material studio software version 8.0 which is based on the principle of Darwinian evolution [16] was used to select combination of descriptors that best correlate the structure of the compounds with their respective activities [10]. Ga by Johnson Holland and their applications in chemistry date back to the 1970s. The most common use of GA in **in** silico materials or drug design has been for feature selection to alleviate the 'curse of dimensionality' problem alluded to above by reducing the large pool of features to a smaller set that can be easily correlated with the molecular property or biological activity of interest. Multi linear regression (MLR) was used was used to generate predictive models by using small number of descriptors. **GA**s have been shown to generate accurate and robust QSAR.

2.6. Model construction

The best descriptors combination selected by GFA was obtained for both training and test set from the descriptor pool [10]. Their anti HIV activities were placed at the last column in their respective spread sheets. Only the training set descriptors and their activities were imported into the material studio software version 8.0 to generate the model and to validate the internal validation parameters such as R^2, R^2_{adj}, Q^2_{cv}, F-test and Y-randomization and test set to validate the built model.

Three descriptors **minssO, PPSA-3** and **RDF135v** were used to validate the model.

Table 4. Descriptor Name, Type, Meaning and Class

Descriptor Name	Type	Meaning	Descriptor class
minssO	2D	Minimum atom type E-state:-O-	Electrotopogical State Atom type descriptor
PPSA-3	3D	Charge weighted partial positive surface area	CPSA descriptor
RDF135v	3D	Radial distribution function-135/weighted by relative Vander Waals volumes	RDF descriptors

2.7. Validation of QSAR model

2.7.1. Internal validation

The QSAR were developed by GFA and MLR methods and evaluated using the following statistical parameters such as standard errors of regression coefficient, R^2(squared correlation coefficient, R^2_{adj}(adjusted squared correlation coefficient, Q^2(leave one out cross validated coefficient, F-test , Y-randomization etc.

The most common internal method of validating the model is least square fitting, R^2 (squared correlation coefficient) for the comparison between the predicted and the experimental activities. An improved method of determining R^2 is the robust straight line fit. The difference between R^2 and R^2_{adj} value is less than 0.3 indicates that the number of descriptors involved in the QSAR model is acceptable. The number of descriptors is not acceptable if the difference is more than 0.3. Also for good predictability, the difference between R^2 and Q^2 value should not exceed 0.3 [17].

Cross validation process repeats the regression many times usually each molecule is left out once (only) in turn, and the R is computed using the predicted values of the missing molecule. It can also be more than one molecule (leave many out, LMO) is left at a time, CV is used to determine how large a model can be used for a given data set. CV is especially useful if the training set used to create the model is small (≤ 20 compounds)

Equations of internal validation parameters are as follows:

$R^2 = 1 - \frac{\sum(Y_{cal} - Y_{obs})^2}{\sum(Y_{obs} - \bar{Y})^2}$ (3)

$Q^2 = 1 - \frac{\sum(Y_{obs} - Y_{predicted})^2}{\sum(Y_{obs} - \bar{Y})^2}$ (4)

From equations 3 and 4, Y_{obs}, Y_{cal} and $\bar{Y}$ are the observed activity, the calculated activity and the mean observed activity of the samples in the training set, respectively.

R_2 is adjusted for the number of explanatory variables in the model.

It is defined as: $R^2_{adj} = \frac{R^2 - P(n-1)}{n-p+1}$ (5)

P in equation 5 is the number of independent variables in the model.

PRESS = $\sum_{i=1}^{n}(y_{pred,i} - y)^2$ (6)

PRESS means the Predictive residual sum of the squares.

Y is the data value(s) not used to construct the cross validation model.

2.7.2. Y – Randomization

The predictive power of the equation is poor when the observations are not sufficiently independent of each other. To test this, y – randomization of dependent variables is carried out. This process ensures that the model is not due to chance and is strong. Coefficient of determination (cR_p^2) for y – randomization should be greater than 0.5.

$cR_p^2 = R \times [R^2 - (R_r)^2]^2$ ..... (7)

R is the coefficient of determination for y – randomization and R_r is the average R of random models. In Y- randomization procedure, the set of activity values are reassigned randomly to different molecules and repeating the entire modelling procedure. After several repetitions, if the model prediction is comparable to the original equation, the set of compounds observations is not sufficient to support the model. The aim of this method is to test for the validity of the original QSAR model and to ensure that the selected model is appropriate.

In Y –randomization, a number generator is used to allocate the integers between 1 and N to sequence of N numbers. In each cycle, the resulting arrangement of random integers is employed in order to reorder the y- data – leaving the x data intact and then the full data analysis is carried out on these scrambled data. Every run will yield estimates of R^2 and Q^2 which are recorded. Each case of the scrambled data gives much lower R^2 and Q^2 than the original data to show that the model is strong and not due to chance [18].

Y – Randomization is important if there are small numbers of compounds in the training set [11].

2.7.3. External validation

The best method of validating a model is an external method, such as evaluating the QSAR model on a test set compounds [17]. These are statistical methodologies used to ensure the model is sound and unbiased ("good model").

To estimate the predictive power of a QSAR model, Golbraikh and Tropsha recommended the use of the following statistical characteristics of the test set which are as follows: R^2_{pred} , , $\frac{r^2 - r_o^2}{r^2}$, $\frac{r^2 - r_o'^2}{r^2}$

R^2_{pred} is the coefficient of determination between the predicted and observed activities, r_0^2 is predicted vs observed activities and $r_0'^2$ is observed vs predicted activities, k and k' are slopes of the regression lines through the origin [11].

The coefficient of determination for the test set $R^2_{predicted}$ was calculated using the equation below

$R^2_{predicted} = \frac{\sum(Ypred_{test} - Yexp_{test})^2}{\sum(Yexp_{test} - \bar{Y}_t)^2}$ ..... (8)

Criteria proposed by Golbraikh and Tropsha on a set of parameters for determining the external predictability of QSAR models are as follows [12]:

$R^2_{pred} > 0.6$, $\frac{r^2 - r_o^2}{r^2} < 0.1$, $\frac{r^2 - r_o'^2}{r^2} < 0.1$ and $0.85 \leq k < 1.15$ or $0.85 \leq k' < 1.1$

According to Golbraikh and Tropsha, models are considered satisfactory, if all the above conditions are satisfied.

$r^2_{m(loo)} = (r^2 \times (1 - \sqrt{r^2 - r^2_o})$ ..... (9)

$r'^2_{m(loo)} = (r^2 \times (1 - \sqrt{r^2 - r'^2_o})$ ..... (10)

$\Delta r^2_{m(loo)} = | r^2_m - r'^2_m |$ .....(11)

3. Results and Discussion

3.1. QSAR model generated and its validation parameters

pEC_{50} = 12.8744(**minsso)** - 0.4625**(PPSA-3)** + 0.6350**(RDF135v)** – 45.3480

$N_{training}$ = 14, N_{test} = 6, Friedman LOF = o.1820, R^2 = 0.9334, R^2_{adj} = 0.9134, Q^2_{cv} = 0.8604, F-value = 46.7, $R^2_{predicted}$ = 0.8935

Table 5. Experimental, predicted and residual values of the selected model

Experimental	Predicted	Residual
3.5976	3.931347	-0.333747
4.0511	3.743864	0.307236
3.6713	3.615363	0.055937
4.3159	4.180467	0.135433
4.8359	4.693442	0.142458
4.2818	4.256597	0.025203
5.7825	5.73704	0.04546
5.6308	5.481209	0.149591
4.3484	4.356375	-0.007975
3.8651	3.72197	0.14313
3.7854	3.925758	-0.140358
4.1625	4.343918	-0.181418
4.2525	4.450161	-0.197661
5.0386	5.181888	-0.143288

Table 6. Internal and external validation parameters for the QSAR generated

Parameter	Threshold	Modal score	Comment	Reference
Internal				
R^2	$R^2 > 0.6$	0.9334	Passed	[Tropsha 2010]
R^2_{adj}	$R^2_{adj} > 0.6$	0.9134	Passed	
Q^2	$Q^2 > 0.6$	0.8604	Passed	
$F_{(4,15)}$	> 2.09	46.7	Passed	
Random model				
$\overline{R_r}$	< 0.5	0.3609	Passed	[Torpsha 2010]
$\overline{R_r^2}$	< 0.5	0.1437	Passed	
$\overline{Q_r^2}$	< 0.5	-0.4099	Passed	
cR_p^2	> 0.6	0.6463	Passed	[Roy 2007]
External validation				
$R^2_{predicted}$	$R^2_{predicted} > 0.6$	0.8935	Passed	
$\frac{r^2 - r_o^2}{r^2}$	< 0.1	0.03205	Passed	[Golbraikh and Tropsha 2002]
$\frac{r^2 - r_o'^2}{r^2}$	< 0.1	0.0328	Passed	
r^2_m	> 0.5	0.7715	Passed	
$\Delta r^2_{m(loo)}$	< 0.2	0.0019	Passed	

The aim of this research is to develop a QSAR model that could relate the structure of 7-Hydroxy-1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]pyridine-4-carboxylate Derivatives with their biological activities against HIV-1. Experimental, predicted and residual values of 7-Hydroxy-1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]pyridine-4-carboxylate Derivatives were presented in Table 5. The lower residual values between experimental and predicted values showed that the model has a good predicted power.

GFA method employed in this study led to the selection of three descriptors which were used to build the model for calculating the predicted activities against HIV-1. The combination of Minimum atom type E-state:-O-, Charge weighted partial positive surface area and Radial distribution function-135/weighted by relative Vander Waals volumes descriptors increased R2 obtained to

acceptable value which was an indication that the model generated was robust.

Pearson correlation of the three descriptors used in the QSAR model was reported in **Table 7** which shows that the correlation coefficient between each descriptor is very low. It means that there is no significant inter -correlation among the descriptors used in building the model [10].

Table 7. Pearson's correlation Coefficient

	Name	*minssO*	*PPSA-3*	*RDF135v*
Name	1			
minssO	0.314770069	1		
PPSA-3	0.602073729	0.343128	1	
RDF135v	0.817419352	0.146382	0.81500406	1

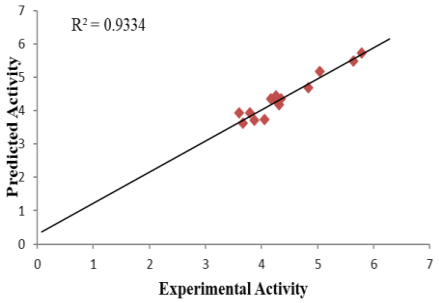


Figure 1. plot of predicted activity against experimental activity of training set

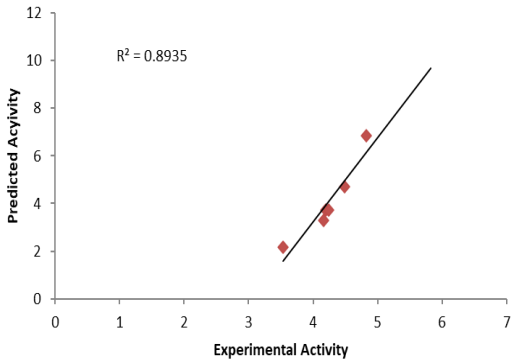


Figure 2: plot of predicted activity against experimental activity of test set

The plot in figure 2 above with $R^2_{predicted} > 0.6$ means that the model was robust.

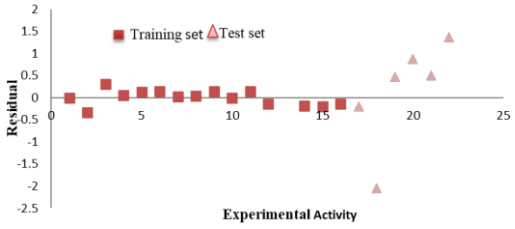


Figure 3: Plot of Residual values versus Experimental Activity

Y-randomization was reported in the Table 8 below. The low values of R^2 and Q^2 for ten trials assured that the developed model was robust, reliable and stable while the high value of $cR^2_p > 0.6$ [19] shows that the QSAR model is strong and not inferred by chance [20].

Table 8. Y- randomization table

Model	R	R^2	Q^2
Original	0.845398	0.714697	0.563412
Random 1	0.356772	0.127286	-0.85834
Random 2	0.400789	0.160632	-0.41882
Random 3	0.292718	0.085684	-0.32289
Random 4	0.352125	0.123992	-0.51328
Random 5	0.194254	0.037734	-0.58644
Random 6	0.476354	0.226913	-0.22621
Random 7	0.414817	0.172073	-0.19033
Random 8	0.287337	0.082563	-0.47896
Random 9	0.22556	0.050877	-0.44226
Random 10	0.607866	0.369501	-0.0622
Random Models Parameters			
Average r :	0.360859		
Average r^2 :	0.143726		
Average Q^2 :	-0.40997		
cRp^2 :	0.646316		

4. Conclusion

In this study, SW-MLR was used to develop linear QSAR model for the prediction of anti-HIV effective activity of 7-Hydroxy-1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]pyridine-4-carboxylate Derivatives. The built model displayed good correlation between the structure and activity of the studied compounds. The model was validated using the following parameters R^2, R^2_{adj}, Q^2_{cv}, y-randomization for internal validation and $R^2_{predicted}$ for external validation. The built model has a good internal and external predictive power. The descriptors minssO, PPSA-3 and RDF135v in the

built model were used to determine the activity of the compounds to functioning effective anti-HIV inhibitors. The robustness and stability of the QSAR model generated have been established by internal and external validation assessment. The result obtained by these validation tests implies that the model can be used to design new 7-Hydroxy-1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]pyridine-4-carboxylate Derivatives with improved anti-HIV activity. The knowledge gained in this piece of work can be used to design more potent 7-Hydroxy-1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]pyridine-4-carboxylate Derivatives as anti HIV agents.

Acknowledgement

The authors wish to thank God for His infinite mercy which has made them to produce this work. They also wish to thank majority of people for their kind advice and encouragement.

References

[1] I.E. Emmanuel,U. Adamu, E. A. Stephen, Quantitative structure and activity relationship modelling study of anti-HIV-1 RT inhibitors: Genetic function approximation and density functional theory method. Journal computational methods in molecular Designs, 5 (2015) 61-76.

[2] Di Santo, Inhibiting the HIV integration process: past, present and future. J med chem 51 (2014) 539-566.

[3] D.S. Ruelas , W.C. Greene , An integrated overview of HIV-1 latency. Cell, Elsevier Inc. 155 (2013) 519-529.

[4] S. Wang, P. Hortz, M. Schechter, L. Rong, Modelling the slow CD4 + T cell Decline in HIV-infected individuals, PLOS Computational Biology 11 (2015).

[5] A.E. Shola, M.O. Idris, S. Tukur, A.U Saviour, G.A. Shallangwa, A. Uzairu, Theoretical modelling for investigating some active compounds as potent inhibitors against lung cancer. Journal of Engineering and exact science 5(2019) 0125-0136.

[6] F. Soualmia, S. Belaidi, N. Tchouar, T. Lanez, Review of computational studies applied in new macrolide antibiotics, Journal of fundamental and applied sciences, 12 (2020) 392-415.

[7] J.S. Jaworska, M. Comber, C.Auer , C.J.Van Leeuwen, Summary of a workshop on regulatory acceptance of QSARs for human health and environmental endpoints, Environmental Health Perspectives, 111 (2003) 1358–1360.

[8] G. Liu, R. Luo, X. ZHANG, Y. ZHOU, J. LI, Y. ZHENG, H. Liu, Synthesis and Evaluation of Anti-HIV Activities of Novel 7-Hydroxy-1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]pyridine-4-carboxylate Derivatives, 4 (2014) 573-580.

[9] Y. He, C. Y. Liew, N. Sharma, S. K. Woo, Y. T. Chau, C. W. Yap, PaDEL-Descriptor: Open source software for PD-PK-T prediction, Journal of computational chemistry, 34 (2013) 604-610.

[10] A.E.Shola, S. Uba, A. Uzairu, In Silico Study for investigating and predicting the activities of 1, 2, 4-Tirazole Derivatives as potent Anti-Tubercular agents. The Journal of Engineering and Exact Science 4 (2018) 0246-0254.

[11] Alexander Tropsha. Best practices for QSAR model Development, validation and exploitation, Molecular Informatics. Inf. 29 (2010) 476-488.

[12] A. Golbraikh; A Tropsha, Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection, J Comput Aided Mol Des., 20 (2002) 269–276.

[13] A. Racz, D. Bajusz, K. Heberger, Intercorrelation limits in molecular descriptors preselection for QSAR/QSPR, Molecular informatics 38 (2019) 1-6.

[14] K. Roy, J.T. Leonard, On selection of training and test sets for the development of predictive QSAR models. QSAR and Combinatorial Science 25 (2006) 235-251.

[15] M. Patel, N. Malle Shappa, S. Poonam, J. Varun, B. Sumit, L. Sandeep, A. Vikrant, D. Saurabh, B. Varun, QSAR studies as strategic approach in drug discovery. Med Chem 23 (2014) 4991-5007.

[16] J.H. Holland, Adaptation in natural and artificial system, University of Michigan Press. (1975).

[17] V. Ravinchandran, R. Harish, J. Abhishek, S. Shalini, P.V. Christapher, K.A. Ram,

Validation of QSAR models-strategies and importance, International Journal of Drug Design and Discovery, 2 (2011) 511-519.

[18] A.Tropsha, P. Grammatica, V.K. Gombar, The importance of being Earnest: Validation is the Absolute essential for successful Application and interpretation of QSAR models, QSAR and Combinatorial Science 22 (2003) 69-77.

[19] K. Roy, Some aspects of validation of predictive quantitative structure-activity relationship models, Expert Opinion On Drug Discovery 2 (2007) 1567-1577.

[20] A.S. Ugochukwu, G. A. Shallangwa, A. Uzairu, Quantitative Structure and Activity Relationship of 3a, 6a – Dihydro-1H-pyrrolo [3,4-c]pyrazole-4,6-dione Derivatives, Turkish Computational and Theoretical Chemistry. 4 (2020) 32-39.