Send Orders for Reprints to reprints@benthamscience.net

Letters in Drug Design & Discovery, **2014**, *11*, 000-000

QSAR Study on Diketo Acid and Carboxamide Derivatives as Potent HIV-1 Integrase Inhibitor

Arodola Olayide Adebimpe, Radha Charan Dash and Mahmoud E. S. Soliman*

Discipline of Pharmaceutical Sciences, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

Abstract: Herein, we present a validated predictive QSAR model to provide more insight into the relationship between the molecular properties of diketo acid and carboxamide derivatives as well as HIV-I integrase inhibition. A set of 40 diketo acid and carboxamide derivatives possessing integrase inhibitory activity was subjected to 2D-QSAR using Discovery studio V3.5. The QSAR results presented here were based on a genetic function algorithm (GFA) approach. Logarithmic inverse values of IC_{50} (μM) were taken as the dependent variables, and physicochemical parameters were taken as the independent variable. A suitable set of molecular descriptors was calculated using GFA approach (max 500 generations). Results showed that radius of gyration, Zagreb index, Wiener index and minimized energy are statistically significant with the correlation coefficient value of 0.820 and play an important role in HIV-1 integrase inhibition.

Keywords: Diketo acid, carboxamide, 2D-QSAR, GFA, Integrase inhibitor.

INTRODUCTION

Acquired immunodeficiency syndrome (AIDS), currently regarded as one of the most devastating diseases of the human immune system is caused by the human immunodeficiency virus (HIV) [1-3]. AIDS was first reported in 1981[4, 5], and now has become a global pandemic. The integration of HIV-1 DNA into the host chromosome contains a series of DNA cutting and joining reactions. The first step in the integration process is 3 end processing. In the second step, termed DNA strand transfer, the previously processed viral DNA end is inserted into the target DNA [6, 7]. Thus, the integrase enzyme is crucial for viral replication and represents a potential target for antiretroviral drug design [8-11].

It has been almost forty years since the quantitative structure-activity relationship (QSAR) paradigm first found its way into the practice of pharmaceutical chemistry, toxicology (QSTR) [12], property (QSPR) [13], and eventually most aspects of chemistry. The first application of QSAR is attributed to Hansch *et al.* [14], who developed an equation that related biological activity to certain physicochemical properties of a set of structures [15] and this has become one of the most useful approach to speed up drug design process [16, 17].

QSAR yield power may be attributed to the strength of its initial postulate that activity was a function of structure as described by electronic attributes and steric properties. QSAR attempts to find consistent relationships between the variations in the values of molecular properties and the biological activity for a series of compounds so that it can be used to evaluate new chemical entities [14]. The formulation of thousands of equations using the QSAR methodology attempts a validation of its concepts and its utility in the elucidation of the mechanism of action of drugs at the molecular level and a more complete understanding of physicochemical properties. It is now possible to develop a model for a system as well as compare models from a biological database [18].

There is a series of statistical model analyses that are used to develop a QSAR model, which include multiple linear regression (MLR), principle component analysis (PCA), partial least square (PLS), genetic function algorithm (GFA).

In this study, we describe the application of QSAR models based on GFA approach. GFA is a heuristic search method used for identifying optimal solutions to a problem where the possible solution space is too large to be exhaustively enumerated. GFA has been widely used for feature optimization in QSAR models for variable selection [19-21]. The purpose of variable selection is to select the variables significantly contributing to prediction and to discard other variables by a fitness function. The GFA approach has a number of important advantages, which include: ability to build multiple models rather than a single model; automatic selection of features to be used in its basic functions and to determine the appropriate number of basic functions to be used by testing full-size models rather than incrementally building them; reliable discovery of combinations of basic functions that take advantage of correlations between features; ability to incorporate the lack of fit (LOF) error measure developed by Friedman [22] that resists over fitting and allows user control over the smoothness of fit; use of larger variety of basic functions in construction of its models, preferred model length and useful partitions of the data set, automatic removal of outliers and finally, provision of additional information not available from other statistical standard regression analysis. The GFA has been applied to three published data sets to demonstrate it is an effective tool for doing both QSAR and QSPR [23-25].

*Address correspondence to this author at the Discipline of Pharmaceutical Sciences, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa; Tel: +27 0312607413, Fax: +27 031260 779; E-mail: soliman@ukzn.ac.za

1570-1808/14 $58.00+.00

Although several QSAR studies on HIV integrase inhibitors have been reported [26-32] using MLR, PLS and PCA, the QSAR study on HIV-1 integrase using the GFA method has been lacking in literature. Such an understanding about the GFA method might provide a new starting point for the design of novel inhibitors against HIV-1. The main purpose of this work is to find out how accurate the QSAR analysis predicted the activities of compounds that were already synthesized in comparison to their experimental biological activities. Therefore, a 2-dimensional QSAR model was used to analyze some potential diketo acid and carboxamide-based HIV 1 integrase inhibitors.

The list of the structures of 40 inhibitors employed in this study and their experimental inhibitory concentration (IC_{50}) effective against HIV-1 integrase enzyme was taken from literature [33-35] (Table **1**).

METHODS

The QSAR study that we conducted was performed on 40 molecules of the diketo acid and carboxamide derivatives (Table **1**) which had strand transfer data (IC_{50} - molar concentration of the drug leading to 50% inhibition of enzyme Integrase) that was collected from literature [33-35]. Out of 40 molecules, 30 were used as a training set and 10 molecules as a test set to evaluate the internal degree of predicitivity of the QSAR equation. With the help of CHEM 3D, which has been used in a previous study [31], different 2D structures were drawn (see Table **1**), followed by the conversion to 3D structures of reasonable conformations using Discovery studio v3.5 software, which has been previously used in a publication [36]. A large number of descriptors were also calculated (e.g. ALogP, molecular weight, molar refractivity, dipole moment, heat of formation, Radius of gyration, Wiener index, Zagreb index etc.) (the full list of descriptors is provided in the supplementary material file). Total charge and total formal charge, which are atomistic descriptors, were found to be irrelevant, due to the insignificant values of zero, and were therefore discarded from this study. 2D QSAR analysis was carried out using genetic function approximation (GFA) analysis (with 500 maximum generations) [31].

Calculation of pIC_{50}

Reported IC_{50} (μM) for strand transfer values were manually converted into -log IC_{50} (pIC_{50}) using the formula given below. The term 'pIC_{50}' is a scale for expressing IC_{50} value exponentially, which normalizes the actual activity using negative logarithmic function.

$$pIC_{50} = -\log IC_{50}$$

Calculation of Molecular Descriptors

The molecular descriptors were calculated for the data set using QSAR properties and utility of Discovery studio v3.5. Topological (surface area) and constitutional descriptors (AlogP, Dipole, Molecular weight, Energy, Radius of gyration, Wiener index and Zagreb index etc.) were computed (Table **2**).

CONCLUSION

In this study, we screened 26 preselected descriptors for 40 compounds using GFA method. GFA was then used to generate three different 2D-QSAR models to determine the degree of predicitivity of these diketo acid and carboxamide derivatives as HIV-1 integrase inhibitors. A QSAR model was generated for integrase activity. As all the descriptors were not important for specific model generation, in order to select the optimal set of descriptors, we used systematic variable selection leave one out (LOO) method in a stepwise forward manner for the selection of descriptors. Three best QSAR equations models generated for this study using the GFA approach and LOO method are shown in Table **3**.

The statistical quality of the generated models was determined by the parameters like correlation co-efficient (R^2), cross-validated squared correlation co-efficient (Q^2), LOF, which is the relative measure of quality of fit, p-value which represents the variance of calculated and observed activity, and chance statistics assuring that the results are not merely based on chance correlations. Best models were selected on the basis of their statistical significance.

Equation 1: The QSAR result for integrase inhibition produced highly predictive model that has excellent R^2 = 0.820. Furthermore, the developed model showed Q^2 = 0.558 and LOF value 0.193 indicating the model is well validated and explained. The four best descriptors selected on the basis of importance in biological activity were Zagreb index (Z), Wiener index (W), Radius of gyration (R) and minimized energy (M).

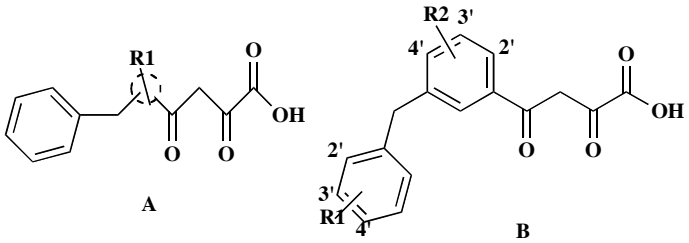
From the QSAR study, the statistically significant equation derived was,

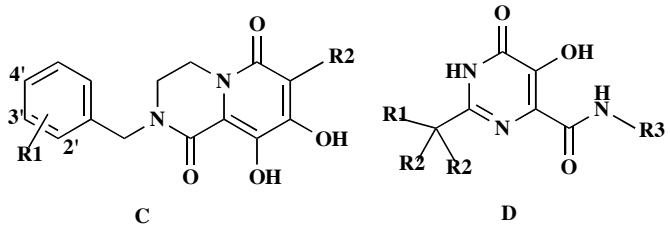
$$pIC_{50} = -11.65 - 0.0024W + 0.089Z + 0.019M + 1.187R$$

N= number of training compounds =30, R^2 = 0.820; Q^2 = 0.558; LOF = 0.193.

From the equation it was observed that, the radius of gyration and minimized energy and Zagreb index are positively correlated with the biological activity, however radius of gyration had the most contribution towards the biological activity. The radius of gyration of a molecule describes its dimensions and is calculated as the root mean square distance between its center of gravity and its ends. The value of R in compounds **30, 34, 35** and **36** having amide residue attached to a cyclic aromatic ring is high. These compounds showed good activity. On the contrary the value of R in the compounds **13, 17** and **19** having a halogenated group is low and resulted into lower biological activity. The second parameter Z contributed relatively lower in the above QSAR study but positively correlated with biological activity. The Z is a topological descriptor based on the vertex degree of heavy atoms. It also interprets the activity of compounds **18** and **20** having heavy atoms compared to compounds **9** and **13.** The W is the sum of the chemical bonds existing between all pairs of heavy atoms in the molecule. The value of W in the compounds **1, 3,** and **5** having a heavy atom attached to one of the aromatic rings is less. This resulted in lower biological activity. On the other hand, the high value of W in the compounds **30, 34** and **36** is due to the fact that

Table 1. Structures and Biological Activity of Training and Test Set Studied in this Report.





Comp	Core	R1	R2	R3	$IC_{50}(\mu M)$	$*pIC_{50}(\mu M)$
1	A	Pyrrole	4'-F	-	0.17	0.770
2	A	O-xylene	-	-	5.67	-0.754
3	A	1,2-(CH_3)-1*H*-pyrrole	-	-	0.22	0.658
4[a]	A	2,3-(CH_3) thiopene	-	-	0.18	0.745
5	A	2,4-(CH_3) thiopene	-	-	0.16	0.796
6	A	1,3-(CH_3)-1*H*- pyrrole	-	-	0.5	0.301
7	A	2,5-(CH_3) thiopene	-	-	0.5	0.301
8[a]	B	4'-Cl	-	-	1.0	0.000
9	B	3'-F	-	-	0.25	0.602
10	B	-	4'-OCH_3	-	0.15	0.824
11	B	-	3'-OCH_3	-	0.14	0.854
12[a]	C	4'-F	-	-	0.10	1.000
13	C	H	-	-	0.23	0.638
14	C	2'-Cl	-	-	0.37	0.432
15	C	3'-Cl	-	-	0.04	1.398
16[a]	C	4'-Cl	-	-	0.38	0.420
17	C	4'-F, 3'-Cl	-	-	0.04	1.398
18	C	4'-F	CN	-	0.02	1.699
19	C	4'-F	Br	-	0.03	1.523
20[a]	C	4'-F	I	-	0.02	1.699
21	D	$N(CH_3)_3$	tetrahydro-2*H*-pyran	4-fluorotoluene	0.002	2.699
22	D	$NHCOCH_3$	CH_3	4-fluorotoluene	0.007	2.155

Table 1. contd…

Comp	Core	R1	R2	R3	$IC_{50}(\mu M)$	$*pIC_{50}(\mu M)$
23	D	$NH\text{-}SO_2\text{-}CH_3$	CH_3	4-fluorotoluene	0.008	2.097
24[a]	D	$NHCO\text{-}N(CH_3)_2$	CH_3	4-fluorotoluene	0.018	1.745
25	D	$NHSO_2\text{-}N(CH_3)_2$	CH_3	4-fluorotoluene	0.012	1.921
26	D	$NHCOCO\text{-}N(CH_3)_2$	CH_3	4-fluorotoluene	0.01	2.000
27	D	$NHCOCO\text{-}OCH_3$	CH_3	4-fluorotoluene	0.015	1.824
28[a]	D	NHCOCO-OH	CH_3	4-fluorotoluene	0.004	2.398
29	D	$N(CH_3)COCO\text{-}N(CH_3)_2$	CH_3	4-fluorotoluene	0.015	1.824
30	D	NHCOCO-1,4-(CH_3) morpholine	CH_3	4-fluorotoluene	0.02	1.699
31	D	NHCOCO-1,4-(CH_3) piperazine	CH_3	4-fluorotoluene	0.026	1.585
32[a]	D	$NHCOCO\text{-}N(CH_3)_2$	CH_3	1'-ethyl-2',3'-(OCH_3)	0.021	1.678
33	D	$NHCOCO\text{-}N(CH_3)_2$	CH_3	1'-ethyl-3'-Cl-4'-F benzene	0.009	2.046
34	D	NHCO-pyridine	CH_3	4-fluorotoluene	0.02	1.699
35	D	NHCO-pyridazine	CH_3	4-fluorotoluene	0.015	1.824
36[a]	D	NHCO-pyrimidine	CH_3	4-fluorotoluene	0.007	2.155
37	D	NHCO-oxazole	CH_3	4-fluorotoluene	0.007	2.155
38	D	NHCO-thiazole	CH_3	4-fluorotoluene	0.008	2.097
39	D	NHCO-1*H* Imidazole	CH_3	4-fluorotoluene	0.006	2.222
40[a]	D	NHCO-1,3,4- oxadiazole	CH_3	4-fluorotoluene	0.015	1.824

IC_{50} = Biological activity (μM) $^*pIC_{50}$ [Experimental pIC_{50} value] = -log IC_{50}

^{*a} = Compounds were used in test set

Table 2. List of descriptors used in this study.

Abbreviations	Definition
M	Minimized energy. Gives the energy after a fast minimization procedure using clean force field
W	Wiener index. It is the sum of the chemical bonds existing between all pairs of heavy atoms in the molecule) [37]
Z	Zagreb index. It is the sum of the squares of vertex valences
R	Radius of gyration. It is the measures of the size of an object, a surface, or an ensemble of points. It is the root mean square distance of the object's parts from either its center of gravity or a given axis
Ms	Molecular_3D_SASA. It calculates the total solvent accessible surface area for each molecule using a 3D method.[37].

Table 3. The 3 different equations derived from the QSAR model.

	Equation	R^2	Q^2	LOF	P-value
1	Y= -11.65 − 0.0024929W + 0.088809Z + 0.01936M + 1.1879R	0.820	0.558	0.193	5.174e-09
2	Y= -12.896 − 0.0028585W + 0.077907Z + 0.020068M + 0.015681Ms	0.812	0.470	0.202	9.270e-09

Table 3. contd…

	Equation	R^2	Q^2	LOF	P-value
3	Y= -9.6736 − 0.0020098W + 0.078883Z + 0.89779R	0.790	0.620	0.190	5.641e-09

Y: pIC_{50}, set of descriptors (W, Z, M, R, Ms,) have been explained in table 2, R^2: correlation coefficient, Q^2: cross-validated R squared, LOF: Lack of fit, P-value: significance level. [31]

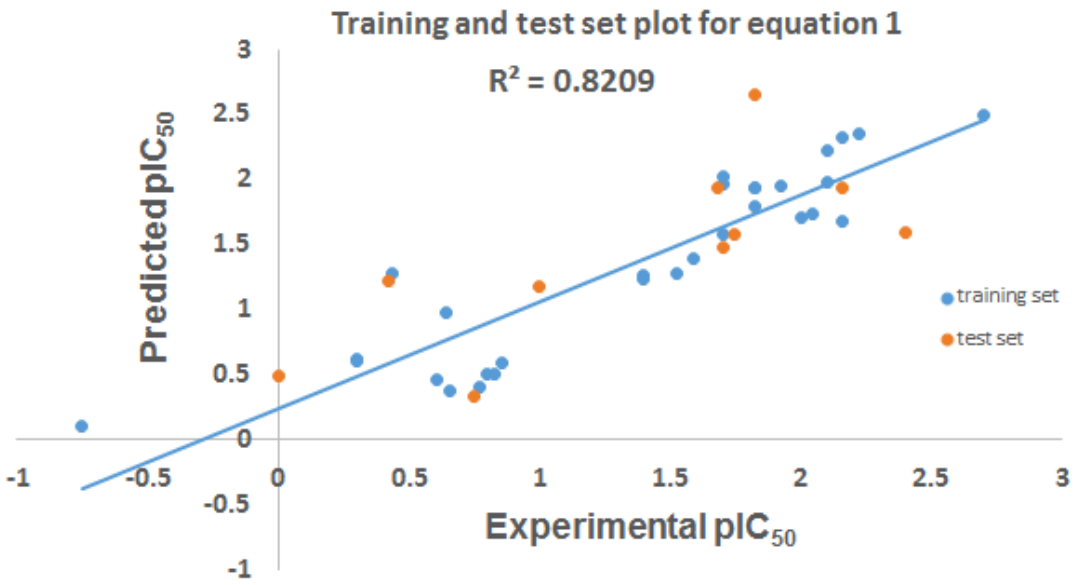


Fig. (1). Predicted pIC_{50} against experimental pIC_{50} for equation 1.

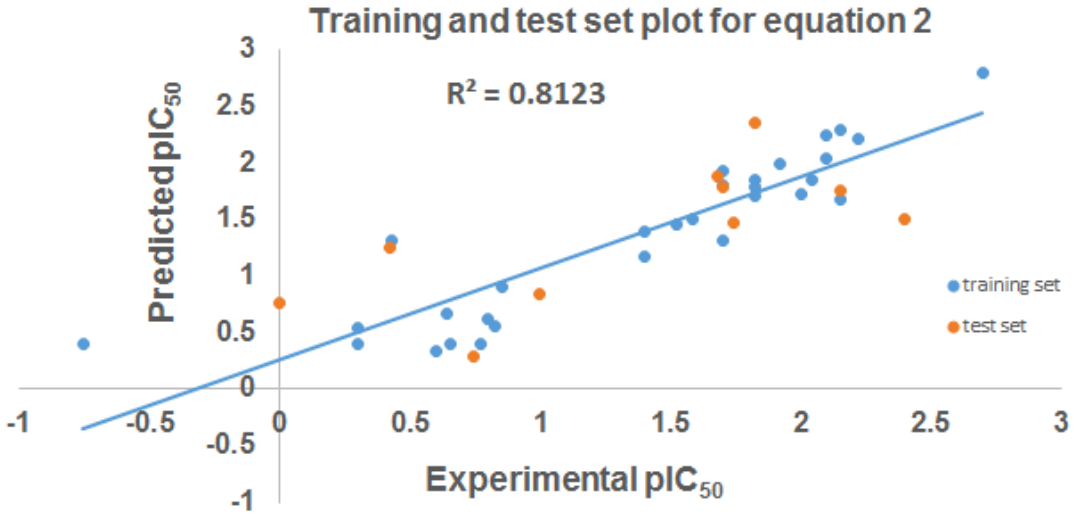


Fig. (2). Predicted pIC_{50} against experimental pIC_{50} for equation 2.

these compounds have one or more amide group attached to one or more cyclic aromatic ring, which gives room for an addition or substitution reaction and thus an increased biological activity.

Equation 2: The QSAR result for integrase inhibition produced a good predictive model that also has $R^2 = 0.812$. Furthermore the developed model showed $Q^2 = 0.470$ and LOF value 0.202 indicating the model is also well validated and explained. The four best descriptors selected on the basis of importance in biological activity were Wiener index (W), Zagreb index (Z), minimized energy (M) and total solvent accessible surface area (Ms). From the QSAR study, the statistically significant equation derived was:

pIC_{50} = -12.896 − 0.0028585W + 0.077907Z + 0.020068M+ 0.015681Ms

N= number of training compounds =30, $R^2 = 0.812$; $Q^2 = 0.470$; LOF = 0.202.

From the equation it was observed that, the total solvent accessible surface area and minimized energy and Zagreb

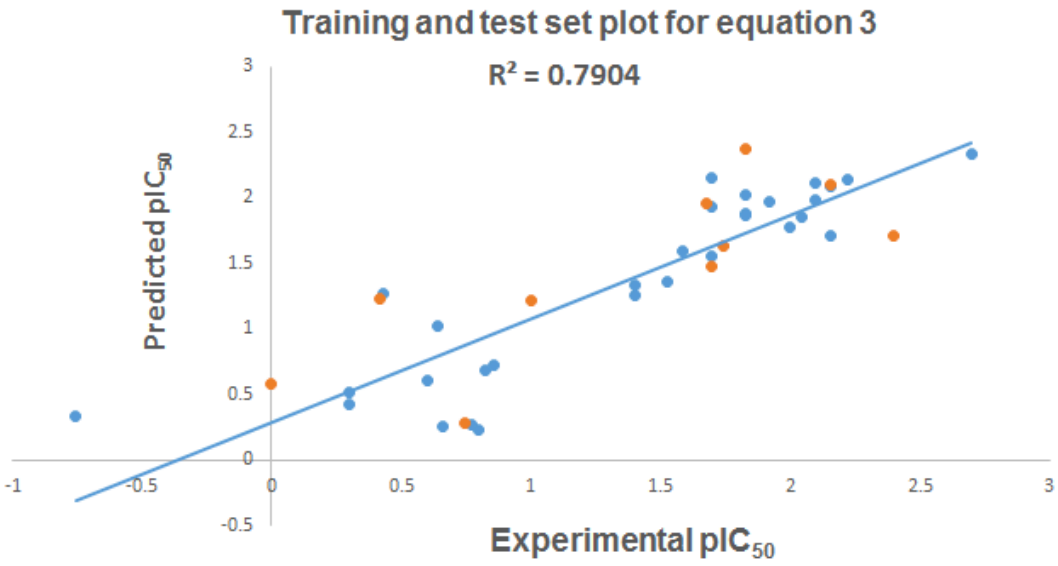


Fig. (3). Predicted pIC_{50} against experimental pIC_{50} for equation 3.

index positively correlated negatively with biological activity, however Zagreb index contributed mostly towards the biological activity. The Zagreb index of a molecule describes the sum of squares of vertex valences. The value of Z in compounds **21, 30, 31** and **40** having amide residue attached to a cyclic aromatic ring is high. These compounds showed good activity. On the contrary, the value of Z in the compounds **4, 5** and **7** having a sulfide group attached to an aromatic cyclic ring is less and resulted into lower biological activity. The second parameter M contributed relatively low in the above QSAR but positively correlated with biological activity. The M is a topological descriptor that best describe the energy of a system after a fast minimization procedure and it interprets the high activity of compounds **26** and **29** respectively.

Equation 3: The QSAR result for integrase inhibition produced a good predictive model that also has good R^2 = 0.790. Furthermore the developed model showed Q^2 = 0.620 and LOF value 0.190 indicating the model is also well validated and explained. The three best descriptors selected on the basis of importance in biological activity were Wiener index (W), Zagreb index (Z), Radius of gyration (R). From the QSAR study, the statistically significant equation derived was:

$$pIC_{50} = -9.6736 - 0.0020098W + 0.078883Z + 0.89779R$$

N= number of training compounds =30, R^2 = 0.790; Q^2 = 0.620; LOF = 0.190.

From the equation it was observed that the Zagreb index (Z) and Radius of gyration (R) positively correlated with biological activity, however R contributed mostly towards the biological activity. The value of R in compounds **30, 34, 35** and **36** having at least an amide residue attached to a cyclic aromatic ring is high. These compounds showed good activity. On the contrary, the value of R in compounds **13, 17** and **19** having halogenated group is lower and this resulted in decreased biological activity. The second descriptor Z gave the least contribution in the above equation but positively correlated with the biological activity. Z also interprets the activity of compounds **18** and **20** having heavy atoms compared to compounds **9** and **13,** which shows lower biological activity.

Table **4** shows the experimental pIC_{50} and the predicted pIC_{50} using the GFA approach for the training set. This shows how the GFA method predicted the pIC_{50}.

Table **5** also shows how well the pIC_{50} was predicted using the GFA approach for the test set. From the residual values (Figs. **4**, **5** and **6**), it can be clearly seen that the lower residual values show that there is a minimal difference between the experimental value and the predicted value of the biological activity of this test set.

The histograms of the three different QSAR models are represented below.

In Fig. (**7**), the Y-axis represents the different molecular descriptors used in this study as shown on the right side of the graph. On the other hand, the X-axis represents the number of generations we could generate for each of these molecular descriptors.

According to Fig. (**7**), at each step, the GFA uses the current population to create the children that make up the next generation. The algorithm selects a group of individuals in the current population, called parents, who contribute their genes—the entries of their vectors—to their children. The algorithm usually selects individuals that have better fitness values as parents. User can specify the function that the algorithm uses to select the parents. The GFA creates three types of children for the next generation: Elite children, Crossover children, and Mutation children. In our QSAR study, the algorithm stops when the number of generations reaches the value of 500 Generations.

In this present study, QSAR models have been developed based on molecular, structural, physicochemical, 2D and 3D

Table 4. Experimental pIC_{50} and GFA predicted pIC_{50} for training set.

Cmpounds	pIC_{50}	Predicted[1]	Residual[1]	Predicted[2]	Residual[2]	Predicted[3]	Residual[3]
1	0.77	0.409	0.361	0.393	0.377	0.274	0.496
2	-0.754	0.105	-0.859	0.407	-1.161	0.335	-1.089
3	0.658	0.377	0.281	0.397	0.261	0.261	0.397
5	0.796	0.498	0.298	0.618	0.178	0.228	0.568
6	0.301	0.616	-0.315	0.536	-0.235	0.422	-0.121
7	0.301	0.608	-0.307	0.398	-0.097	0.512	-0.211
9	0.602	0.463	0.139	0.330	0.272	0.602	0.000
10	0.824	0.505	0.319	0.563	0.261	0.692	0.132
11	0.854	0.591	0.263	0.900	-0.046	0.725	0.129
13	0.638	0.971	-0.333	0.676	-0.038	1.017	-0.379
14	0.432	1.280	-0.848	1.316	-0.884	1.276	-0.844
15	1.398	1.239	0.159	1.166	0.232	1.260	0.138
17	1.398	1.267	0.131	1.401	-0.003	1.340	0.058
18	1.699	1.580	0.119	1.311	0.388	1.559	0.139
19	1.523	1.276	0.247	1.464	0.059	1.362	0.160
21	2.699	2.495	0.204	2.796	-0.097	2.334	0.365
22	2.155	1.681	0.474	1.672	0.483	1.713	0.442
23	2.097	1.973	0.124	2.034	0.063	1.989	0.108
25	1.921	1.957	-0.036	1.998	-0.077	1.975	-0.054
26	2.000	1.704	0.296	1.724	0.276	1.777	0.223
27	1.824	1.797	0.027	1.707	0.117	1.867	-0.043
29	1.824	1.943	-0.119	1.851	-0.027	1.883	-0.059
30	1.699	1.970	-0.271	1.926	-0.227	1.929	-0.230
31	1.585	1.391	0.194	1.499	0.086	1.594	-0.009
33	2.046	1.739	0.307	1.845	0.201	1.860	0.186
34	1.699	2.020	-0.321	1.809	-0.110	2.154	-0.455
35	1.824	1.931	-0.107	1.787	0.037	2.017	-0.193
37	2.155	2.325	-0.170	2.302	-0.147	2.090	0.065
38	2.097	2.221	-0.124	2.243	-0.146	2.109	-0.012
39	2.222	2.357	-0.135	2.219	0.002	2.133	0.089

Table 5. Experimental pIC_{50} and GFA predicted pIC_{50} for test set.

Cmpounds	pIC_{50}	Predicted[1]	Residual[1]	Predicted[2]	Residual[2]	Predicted[3]	Residual[3]
4	0.745	0.326	0.419	0.287	0.458	0.282	0.463
8	0.000	0.485	-0.485	0.761	-0.761	0.587	-0.587
12	1.000	1.178	-0.178	0.836	0.164	1.215	-0.215

Table 5. contd….

Cmpounds	pIC_{50}	Predicted[1]	Residual[1]	Predicted[2]	Residual[2]	Predicted[3]	Residual[3]
16	0.420	1.212	-0.792	1.259	-0.839	1.233	-0.813
20	1.699	1.482	0.217	1.784	-0.085	1.473	0.226
24	1.745	1.580	0.165	1.471	0.274	1.634	0.111
28	2.398	1.594	0.804	1.500	0.898	1.706	0.692
32	1.678	1.937	-0.260	1.877	-0.199	1.961	-0.283
36	2.155	1.936	0.219	1.765	0.390	2.096	0.059
40	1.824	2.656	-0.832	2.360	-0.536	2.371	-0.547

Predicted[1]: predicted value for equation 1, Residual[1]: residual value for equation 1

Predicted[2]: predicted value for equation 2, Residual[2]: residual value for equation 2

Predicted[3]: predicted value for equation 3, Residual[3]: residual value for equation 3

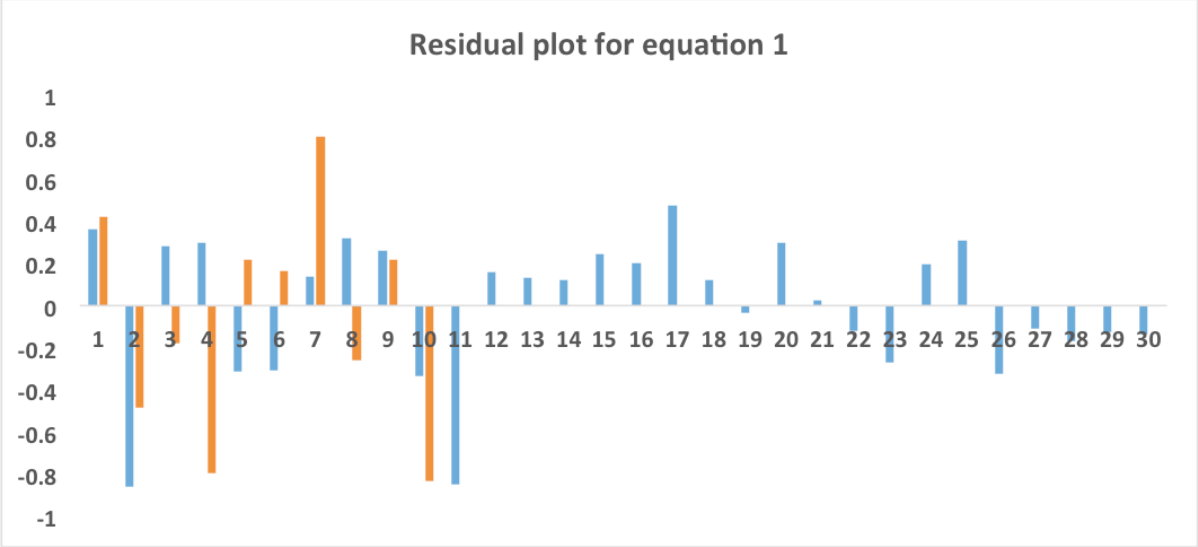


Fig. (4). Histogram of residual values obtained from QSAR model for equation 1.

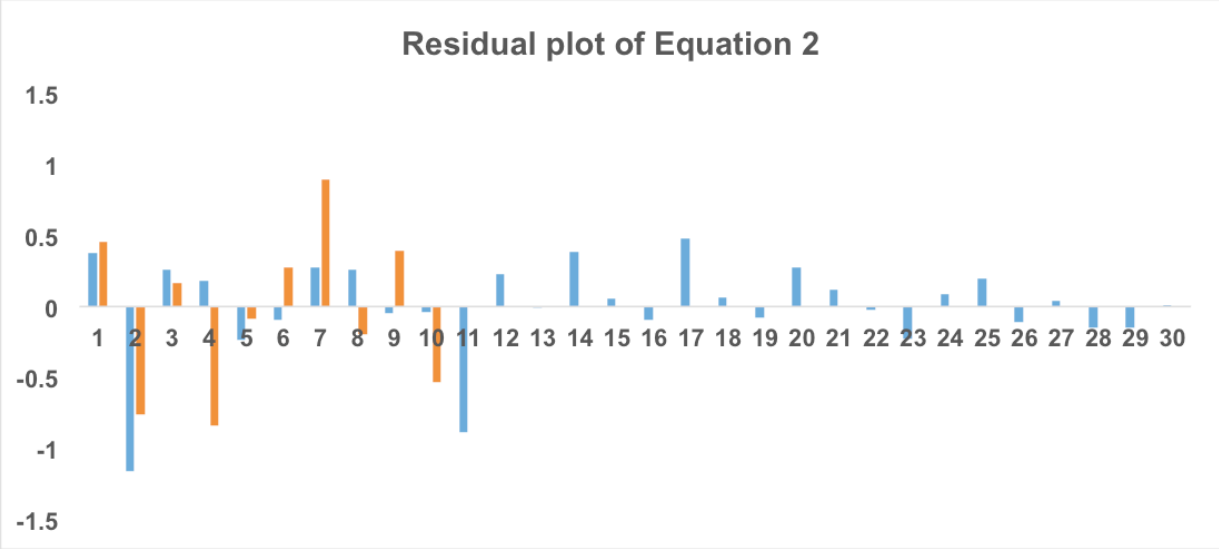


Fig. (5). Histogram of residual values obtained from QSAR model for equation 2.

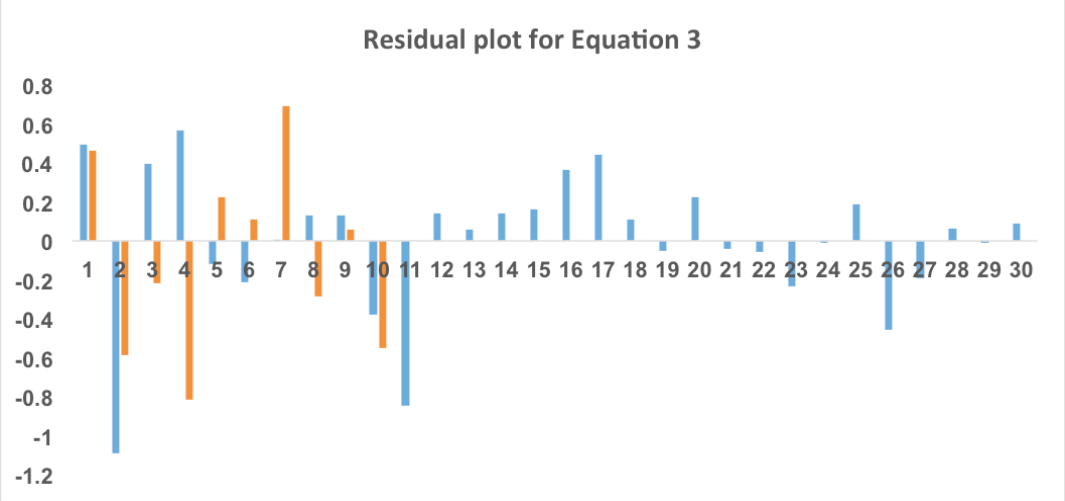


Fig. (6). Histogram plot of residual values obtained from QSAR model for equation 3.

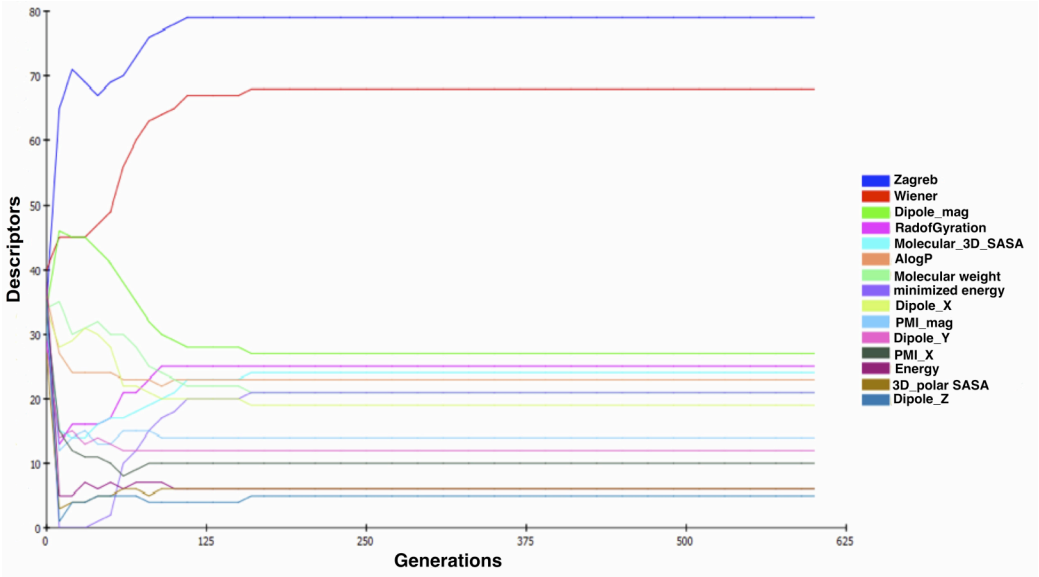


Fig. (7). The graph of the variable usage against generation number.

properties that were obtained from various softwares. The Discovery studio result suggests that the Radius of gyration, Zagreb index, Wiener index and minimized energy are statistically significant with a correlation coefficient value of 0.8209, which is highly significant. These descriptors have played an important role in identifying some promising compounds that possess HIV-1 inhibitory properties such as compound **18, 20, 30, 34, 35** and **36**. The synthesis of the compounds considered in this study was done in literature [33-35] but it was validated using 2D-QSAR model. This model holds good predictive performance with Q^2 values ranging from 0.47 to 0.62 that was calculated using LOO method. In conclusion, this model can be used to predict more potent drugs that possess HIV-1 integrase inhibitory property.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

The author acknowledges the computational facility provided by CHPC, Cape Town (http://www.chpc.ac.za) and the School of Health Sciences, UKZN, for financial support.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publishers Web site along with the published article.

REFERENCES

[1] Mitsuya, H.; Guo, H.; Megson, M.; Trainer, C.; Reitz jr, M.Broder, S. Transformation and cytopathogenic effect in an immune human t-cell clone infected by HTLV-I, *Journal of Science*, **1984**, *223*, 1293 - 1296.

[2] Kent A. Sepkowitz, M. D. AIDS - The first 20 years, *N. Engl. J. Med.*, **2001**, *344*, 1764 - 1772.

[3] Weiss, R. A. How does HIV cause AIDS?, *journal of Science*, **1993**, *260*, 1273 - 1279.

[4] Gottlieb, M. S.; Schroff, R.; Schanker, H. M.; Weisman, J. D.; Fan, P. T.; Wolf, R. A.Saxon, A. Pneumocystis-carinii pneumonia and mucosal candidiasis in previously healthy homosexual men - evidence of a new acquired cellular immunodeficiency, *N. Engl. J. Med.*, **1981**, *305*, 1425-1431.

[5] Masur, H.; Michelis, M. A.; Greene, J. B.; Onorato, I.; Vandestouwe, R. A.; Holzman, R. S.; Wormser, G.; Brettman, L.; Lange, M.; Murray, H. W.Cunninghamrundles, S. An outbreak of community-acquired pneumocystis-carinii pneumonia - initial manifestation of cellular immune dysfunction, *N. Engl. J. Med.*, **1981**, *305*, 1431-1438.

[6] Craigie, R. HIV integrase, a brief overview from chemistry to therapeutics, *J. Biol. Chem.*, **2001**, *276*, 23213-23216.

[7] Di Santo, R.; Costi, R.; Artico, M.; Tramontano, E.; La Colla, P.Pani, A. HIV-1 integrase inhibitors that block HIV-1 replication in infected cells. Planning synthetic derivatives from natural products, *Pure Appl. Chem.*, **2003**, *75*, 195-206.

[8] Barreca, M. L.; Ferro, S.; Rao, A.; De Luca, L.; Zappala, M.; Monforte, A. M.; Debyser, Z.; Witvrouw, M.Chimirri, A. Pharmacophore-based design of HIV-1 integrase strand-transfer inhibitors, *J. Med. Chem.*, **2005**, *48*, 7084-7088.

[9] d'Angelo, J.; Mouscadet, J. F.; Desmaële, D.; Zouhiri, F.Leh, H. HIV-1 integrase: the next target for AIDS therapy?, *Pathol. Biol. (Paris)*. **2001**, *49*, 237-246.

[10] Grobler, J. A.; Stillmock, K.; Hu, B. H.; Witmer, M.; Felock, P.; Espeseth, A. S.; Wolfe, A.; Egbertson, M.; Bourgeois, M.; Melamed, J.; Wai, J. S.; Young, S.; Vacca, J.Hazuda, D. J. Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes, *Proc. Natl. Acad. Sci. U. S. A.*, **2002**, *99*, 6661-6666.

[11] Zhuang, L. C.; Wai, J. S.; Embrey, M. W.; Fisher, T. E.; Egbertson, M. S.; Payne, L. S.; Guare, J. P.; Vacca, J. P.; Hazuda, D. J.; Felock, P. J.; Wolfe, A. L.; Stillmock, K. A.; Witmer, M. V.; Moyer, G.; Schleif, W. A.; Gabryelski, L. J.; Leonard, Y. M.; Lynch, J. J.; Michelson, S. R.Young, S. D. Design and synthesis of 8-hydroxy- 1,6 naphthyridines as novel inhibitors of HIV-1 integrase in vitro and in infected cells, *J. Med. Chem.*, **2003**, *46*, 453-456.

[12] Aboul-Kassim; Tareka, T.Simoneit, B., R. T. . QSAR/QSPR and multicomponent joint toxic effect modeling of organic pollutants at aqueous-solid phase interfaces, *Springer Handbook of Environmental Chemistry*, **2001**, *5E*, 243-314.

[13] Das, R. N.Roy, K. Advances in QSPR/QSTR models of ionic liquids for the design of greener solvents of the future, *Mol. Divers.*, **2013**, *17*, 151-196.

[14] Hansch, C.; Maloney, P. P.Fujita, T. Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients, *Nature*, **1962**, *194*, 178.

[15] Yang, G.-F.Huang, X. Development of quantitative structure-activity relationships and its application in rational drug design, *Curr. Pharm. Des.*, **2006**, *12*, 4601-4611.

[16] Hemmateenejad, B.; Miri, R.; Akhond, M.Shamsipur, M. Quantitative structure-activity relationship study of recently synthesized 1,4-dihydropyridine calcium channel antagonists. Application of the Hansch analysis method, *Arch. Pharm.*, **2002**, *335*, 472-480.

[17] Lill, M. A. Multi-dimensional QSAR in drug discovery, *Drug Discovery Today*, **2007**, *12*, 1013-1017.

[18] Hansch, C.; Kurup, A.; Garg, R.Gao, H. Chem-bioinformatics and QSAR: A review of QSAR lacking positive hydrophobic terms, *Chem. Rev.*, **2001**, *101*, 619-672.

[19] Cho, S. J.Hermsmeier, M. A. Genetic algorithm guided selection: Variable selection and subset selection, *J. Chem. Inf. Comput. Sci.*, **2002**, *42*, 927-936.

[20] Fernandez, M.; Caballero, J.; Fernandez, L.Sarai, A. Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM), *Mol. Divers.*, **2011**, *15*, 269-289.

[21] Hasegawa, K.; Miyashita, Y.Funatsu, K. GA strategy for variable selection in QSAR studies: GA-based PLS analysis of calcium channel antagonists, *J. Chem. Inf. Comput. Sci.*, **1997**, *37*, 306-310.

[22] Friedman, J. H. Multivariate adaptive regression splines, *Annals of Statistics*, **1991**, *19*, 1-67.

[23] Selwood, D. L.; Livingstone, D. J.; Comley, J. C. W.; Odowd, A. B.; Hudson, A. T.; Jackson, P.; Jandu, K. S.; Rose, V. S.Stables, J. N. Structure-activity-relationships of antifilarial antimycin analogs - a multivariate pattern-recognition study, *J. Med. Chem.*, **1990**, *33*, 136-142.

[24] Cardozo, M. G.; Iimura, Y.; Sugimoto, H.; Yamanishi, Y.Hopfinger, A. J. QSAR analyses of the substituted indanone and benzylpiperidine rings of a series of indanone benzylpiperidine inhibitors of acetylcholinesterase, *J. Med. Chem.*, **1992**, *35*, 584-589.

[25] Koehler, M. G.Hopfinger, A. J. Molecular modelling of polymers: 5. Inclusion of intermolecular energetics in estimating glass and crystal-melt transition temperatures, *Polymer*, **1989**, *30*, 116-126.

[26] Almerico, A. M.; Tutone, M.; Ippolito, M.Lauria, A. Molecular modelling and QSAR in the discovery of HIV-1 integrase inhibitors, *Curr Comput Aided Drug Des*, **2007**, *3*, 214-233.

[27] de Melo, E. B.Castro Ferreira, M. M. Multivariate QSAR study of 4,5-dihydroxypyrimidine carboxamides as HIV-1 integrase inhibitors, *Eur. J. Med. Chem.*, **2009**, *44*, 3577-3583.

[28] Dessalew, N. Investigation of the structural requirement for inhibiting HIV integrase: QSAR study, *Acta Pharmaceutica*, **2009**, *59*, 31-43.

[29] Gupta, P.; Roy, N.Garg, P. Docking-based 3D-QSAR study of HIV-1 integrase inhibitors, *Eur. J. Med. Chem.*, **2009**, *44*, 4276-4287.

[30] Leonard, J. T.Roy, K. Exploring molecular shape analysis of styrylquinoline derivatives as HIV-1 integrase inhibitors, *Eur. J. Med. Chem.*, **2008**, *43*, 81-92.

[31] Srivastav, V. K.Tiwari, M. QSAR and docking studies of coumarin derivatives as potent HIV-1 integrase inhibitors, *Arabian Journal of Chemistry*.

[32] Sharma, H.; Cheng, X.Buolamwini, J. K. Homology Model-Guided 3D-QSAR Studies of HIV-1 Integrase Inhibitors, *Journal of Chemical Information and Modeling*, **2012**, *52*, 515-544.

[33] Summa, V.; Petrocchi, A.; Bonelli, F.; Crescenzi, B.; Donghi, M.; Ferrara, M.; Fiore, F.; Gardelli, C.; Paz, O. G.; Hazuda, D. J.; Jones, P.; Kinzel, O.; Laufer, R.; Monteagudo, E.; Muraglia, E.; Nizi, E.; Orvieto, F.; Pace, P.; Pescatore, G.; Scarpelli, R.; Stillmock, K.; Witmer, M. V.Rowley, M. Discovery of Raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection, *J. Med. Chem.*, **2008**, *51*, 5843-5855.

[34] Wai, J. S.; Egbertson, M. S.; Payne, L. S.; Fisher, T. E.; Embrey, M. W.; Tran, L. O.; Melamed, J. Y.; Langford, H. M.; Guare, J. P.; Zhuang, L. G.; Grey, V. E.; Vacca, J. P.; Holloway, M. K.; Naylor-Olsen, A. M.; Hazuda, D. J.; Felock, P. J.; Wolfe, A. L.; Stillmock, K. A.; Schleif, W. A.; Gabryelski, L. J.Young, S. D. 4-aryl-2,4-dioxobutanoic acid inhibitors of HIV-1 integrase and viral replication in cells, *J. Med. Chem.*, **2000**, *43*, 4923-4926.

[35] Wai, J. S.; Kim, B.; Fisher, T. E.; Zhuang, L.; Embrey, M. W.; Williams, P. D.; Staas, D. D.; Culberson, C.; Lyle, T. A.; Vacca, J. P.; Hazuda, D. J.; Felock, P. J.; Schleif, W. A.; Gabryelski, L. J.; Jin, L.; Chen, I. W.; Ellis, J. D.; Mallai, R.Young, S. D. Dihydroxypyridopyrazine-1,6-dione HIV-1 integrase inhibitors, *Bioorg. Med. Chem. Lett.*, **2007**, *17*, 5595-5599.

[36] Ouyang, L.; He, G.; Huang, W.; Song, X.; Wu, F.Xiang, M. Combined Structure-Based Pharmacophore and 3D-QSAR Studies on Phenylalanine Series Compounds as TPH1 Inhibitors, *International Journal of Molecular Sciences*, **2012**, *13*, 5348-5363.

[37] Karelson, M.; Lobanov, V. S.Katritzky, A. R. Quantum-chemical descriptors in QSAR/QSPR studies, *Chem. Rev.*, **1996**, *96*, 1027-1043.

Received: August 21, 2013 Revised: November 22, 2013 Accepted: December 16, 2013