The Journal of Engineering and Exact Sciences – jCEC, Vol. 06 N. 01 (2020)
journal homepage: https://periodicos.ufv.br/ojs/jcec
doi: 10.18540/jcecvl6iss1pp0008-0014
OPEN ACCESS – ISSN: 2527-1075

IN SILICO ELUCIDATION OF SOME QUINOLINE DERIVATIVES WITH POTENT ANTI-BREAST CANCER ACTIVITIES

M. O. IDRIS[1], E. S. ABECHI[1], G. A. SHALLANGWA[1] and A. UZAIRU[1]
[1] Ahmadu Bello University, Department of Chemistry, Zaria City, Kaduna, Nigeria

ARTICLE INFO

Article history:
Received 2019-07-10
Accepted 2020-01-03
Available online 2020-02-15

keywords
QSAR model
Model validations
Breast cancer
Quinoline derivatives

ABSTRACT

The toxicity and high resistance to the commercially sold breast-cancer drugs have become more alarming and the demand to produce new and less toxic breast-cancer drugs arises. In silico studies was carried out on some quinoline derivatives to investigate their reported activities against breast cancer and thereby generate a model with a better activity against breast cancer. The chemical structures of the compounds were optimized using Spartan software at Density Functional Theory (DFT) level, utilizing the B3LYP/ 6-31G basis set. Four QSAR models were generated using Multi-Linear Regression (MLR) and Genetic Function Approximation (GFA) method. Equation one was chosen as the best model based on the validation parameters. The validation parameters was found to be statistically significant with square correlation coefficient (R^2) of 0.9853, adjusted square correlation coefficient ($R_{adj}{}^2$) of 0.9816, cross validation coefficient ($Q_{cv}{}^2$) of 0.9727 and an external correlation coefficient square ($R_{test}{}^2$) of 0.6649 was used to validate the model. The built model was a good and robust one for it passed the minimum requirement for generating a QSAR model.*

1. INTRODUCTION

Cancer is a term used to describe the abnormal or irregular growths that occur in the cell. It follows the circulatory diseases when it comes to health issues that take lives. The World Health Organization (WHO) forecast that if a new preventive measure is not adapted, the world may experience about 15 million death case by the year 2020 (Frankish, 2003).

Breast cancer is an irregular growth of the breast cell. It is found mostly in women but in some cases, men can also get it. It is the commonest cancer amongst women; about 1.4 million new breast cancer cases were reported in 2008 to have been diagnosed. Countries with low income were reported to have about 60% death cases resulting from breast cancer (Ferlay et al., 2008). The survival rate of breast cancer cases differs from country to country. An estimated 5 years survival case show that only 40% can survive in low income countries and 60% in high income countries (Coleman et al., 2008).

Quinoline is an aromatic heterocyclic compound having a double ring structure with a fused benzene ring at the two adjacent carbon atoms. It is also referred to as benzo pyridine, benzo[b]pyridine, 1-benzazine. It is a hygroscopic yellowish oily liquid that is slightly soluble in water, alcohol, ether and many other organic solvents (Ferlin et al., 2005). Quinoline derivatives are widely used in the field of medicine and medicinal chemistry because of their anti-malarial, anti-microbial, antitumor, antifungal, antihypertensive, anti-HIV, analgesics and anti-inflammatory activities. Quinoline derivatives represent a large number of anti-proliferative agents exhibiting cytotoxicity through DNA intercalation, causing interference in the replication process (Gasparotto et al., 2006). For the treatment of breast cancer, commercially sold drugs like fulvestrant, lapatinib, eribulin mesylate, pertuzumab, everolimus, doxorubicin and other numerous agents have been approved by the Food Drugs and Administration (FDA) for sub-types treatment. Efforts have been put in place to develop a new and more effective cancer drugs through synthesis and structure modification.

QSAR is a computational technique that shows mathematically the relationship between the inhibitory activity of molecules and their chemical structures. It is the commonly used computational technique for predicting the physicochemical properties of molecules (Wong et al., 2014). QSAR method save cost and resources when developing or designing new drugs and other related substances like fungicides and herbicides (Larif et al., 2013). The aim of this research was to build a QSAR model with improved activity against breast cancer from quinoline derivatives that would give the pharmacologist and pharmacist an insight when to in the design of new breast cancer drugs.

2. MATERIALS AND METHOD

2.1. Data collection:

The quinoline derivatives used in this study were collected from the literature.

2.2. Biological Activities (pIC50)

The biological activities of the compounds were measured and reported in the literature as IC_{50}, it was then converted to logarithm unit (pIC_{50}) using the equation 1 below for simplicity and avoidance of negative IC_{50} value. The IUPAC name of the compounds and their biological activities is presented in Table 1.

$$pIC_{50} = -log\ (IC_{50} \times 10^{-6}) \quad (1)$$

Table 1: Compounds names and their activities.

S/N	IUPAC name of compounds	Activities /µmolL^{1}	
		IC_{50}	pIC_{50}
1	2-cyano-3-phenyl-N-(quinolin-3-yl) acrylamide	79.20	4.1013
2	2-cyano-N-(quinolin-3-yl)-3-p-tolylacrylamide	74.40	4.1284
3	2-cyano-3-(4-fluorophenyl-N-(quinolin-3-yl) acrylamide	40.00	4.3979
4	2-cyano-5-phenyl- N-(quinolin-3-yl) penta-2,4-dienamide	63.60	4.1965
5	3-(2-chlorophenyl)-2-cyano-N-(quinolin-3-yl) acrylamide	53.50	4.2716
6	3-(benzo[d] [1,3] dioxol-5-yl)-2cyano-N-(quinolin-3-yl) acrylamide	57.10	4.2434
7	2-cyano-3-(3-nitrophenyl)-N-(quinolin-3-yl) acrylamide	65.20	4.1857
8	2-cyano-3-(4-nitrophenyl)-N-(quinolin-3-yl) acrylamide	63.00	4.2007
9	2-cyano-3-(4-hydroxy-3-methoxyphenyl)-N-(quinolin-3-yl) acrylamide	29.80	4.5258
10	2-cyano-3-(3,4-dimethoxyphenyl)-N-(quinolin-3yl) acrylamide	64.60	4.1898
11	2-cyano-N-(quinolin-3-yl)-3-(2,3,4-trimethoxyphenyl) acrylamide	49.80	4.3028
12	2-cyano-3-(2,4-dichorophenyl)-N-(quinolin-3-yl) acrylamide	57.60	4.2396
13	2-cyano-5-(4-(dimethyl amino) phenyl)-N-(quinolin-3-yl) penta-2,4-dienamide	40.40	4.3936
14	2-cyano-3-(2methoxynaphthalen-1-yl)-N-(quinolin-3-yl) acrylamide	57.50	4.2403
18	7-(trifluoromethyl)-N-(3,4,5-trimethoxyphenyl) quinolin-4-amine	9.380	5.0278
19	N-(3-methyl bicyclo[3.3.1]nonan-3-yl)-7-(trifluoromethyl)quinolin-4-amine	24.10	4.6180
20	7-chloro-N-(4-morpholinophenyl) quinolin-4-amine	31.50	4.5017
21	N-(4-morpholinophenyl)-7-(trifluoromethyl) quinolin-4amine	23.30	4.6326
22	5-(7-(trifluoromethyl) quinolin-4-ylamino) pyrimidin-2,4-(1H,3H)-dione	21.40	4.6696
23	1,3-dimethyl-6-(7-(trifluoromethyl) quinolin-4-ylamino) pyrimidin-2,4-(1H,3H)-dione	23.30	4.6326
24	N-(benzo[d] [1,3] dioxol-5-ylmethyl)-7-chloroquinolin-4-amine	21.10	4.6757
25	N-(benzo[d] [1,3] dioxol-5-ylmethyl)-7-(trifluoromethyl)-quinolin-4-amine	26.20	4.5817
26	N-(5,6-dimethyl-1,2,4-triazin-3-yl)-7-(trifluoromethyl)-quinolin-4-amine	21.80	4.6615
27	N-(7-(trifluoromethyl)-quinolin-4-yl)-quinolin-3-amine	14.20	4.8477
28	2-methyl-N-(7-trifluoromethyl) quinolin-4-yl)-quinolin-3-amine	16.30	4.7878
29	N-(4-(4-aminophenylsulfonyl) phenyl)-7-chloroquinolin-4-amine	18.80	4.7258
30	N-(4-(4-aminophenylsulfonyl) phenyl-7-(trifluoromethyl)-quinolin-4-amine	23.50	4.6289
31	N,N'-(4,4'-sulfonylbis(4,1-phenylene)bis(7-chloroquinolin-4-amine)	23.20	4.6345
32	N,N'-(4,4'sulfonylbis(4,1-phenylene)bis(7-(trifluoromethyl)-quinolin-4-amine)	24.00	4.6198
33	7-Chloro-4-isothiocyanatoquinoline	22.40	4.6498
34	N-(4-(4-aminophenylsulfonyl)phenyl)-N-(7-chloroquinolin-4-yl)-carbamimiodothioic acid	22.70	4.6440

2.3 Data optimization

2D structures of the compounds were drawn with ChemDraw software. The structures were imported into Spartan 14 V1.1.4 Wave Function programming software to obtain the spatial

conformation structures. The software minimizes the energy of the molecules by optimization at Density Functional Theory (DFT) level, utilizing the (B3LYP/6-31G*) basis set. The optimized molecules in Spartan format were then converted to an SD format and were saved, this is because PaDEL-Descriptor software only recognizes SD file format. The saved SD file format was also imported into the PaDEL descriptor software V2.20 to calculate the molecular descriptors.

2.4 Molecular Descriptor Calculation and data pre-treatment

The chemical characteristic of a compound is best described by its descriptors in the form of numerical values. The PaDEL descriptor software V2.20 was therefore used to calculate the descriptors of the compounds and a total number of 931 descriptors were calculated.

2.5 Data Pre-treatment and division

The data was first normalized to give the descriptors equal chance of occurrence using equation 2, after which the data was pre-treated (Singh 2013).

$$X = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (2)$$

Where X_i is the value of each descriptor, X_{max} and X_{min} is the maximum and minimum value of the descriptors in each column X.

The data pre-treatment was carried out using the data pre-treatment software in the DTC Lab, this was done solely to remove any redundant descriptor and none informative descriptors (Shola et al., 2018). The pre-treated data set was then divided into two sub-sets namely, training and test set by employing Kennard and Stone algorithm method (Kennard et al., 1969). The training set contains 70% of the total compounds and was used to build the model while the remaining compounds (test set) were used to validate the built model.

2.6. Internal Validation of Model.

The internal validation of the model was carried out with the Materials Studio V.8.0 software, employing the Genetic Function Approximation (GFA) method. The models were estimated using the LOF (Friedman 1991) which is expressed in equation 3:

$$LOF = \frac{SSE}{(1 - \frac{C + d*P}{M})^2} \quad (3)$$

Where SSE is the sum of squares of errors, C is the number of terms in the model, d is a user-defined smoothing parameter, P is the total number of descriptors in the model, and M is the amount of data in the training set. SSE is defined by equation 4.

$$SSE = \sqrt{\frac{(Y_{exp} - Y_{pre})^2}{N - P - 1}} \quad (4)$$

2.6.1 Correlation coefficient (R^2) and adjusted correlation coefficient ($R_{adj}{}^2$).

The correlation coefficient square (R^2) is the plot of predicted activity against the experimental activity which shows the potency of the model and the efficiency of the selected descriptors. A close value of R^2 to 1.0 indicates a good model. This can be calculated as follows.

$$R^2 = 1 - [\frac{\Sigma(Y_{exp} - Y_{pred})^2}{\Sigma(Y_{exp} - \bar{Y}_{training})^2}] \quad (5)$$

Where Y_{exp} , Y_{pred} and $\overline{Y}_{training}$, are respectively the experimental activity, the predicted activity, and the mean experimental activity of the compounds in the training set. The R^2 value alone cannot be used to affirm the goodness of the model, so R^2 was adjusted for the number of variables in the model. The adjusted R^2 is given as:

$$R^2_{adj} = \frac{R^2 - k(n-1)}{n - P + 1} \quad (6)$$

Where k is the number of independent variables in the model and n is the number of descriptors.

The QSAR equation used to predict the biological activity of the compounds was determined using the leave-one-out cross validation equation (Q^2_{cv}), given as:

$$Q^2_{cv} = 1 - [\frac{\Sigma(Y_{exp} - Y_{pred})^2}{\Sigma(Y_{exp} - \bar{Y}_{training})^2}] \quad (7)$$

Where Y_{exp} , Y_{pred} and $\overline{Y}_{training}$, are respectively the experimental activity, the predicted activity, and the mean experimental activity of the training set.

2.7. External validation of the model.

The external validation of the model was carried out on the test set to ensure the selected descriptors are appropriate and to also confirm the model's robustness. This can be expressed using equation 8.

$$R^2_{test} = 1 - \frac{\Sigma(Y_{pred_{test}} - Y_{exp_{test}})^2}{\Sigma(Y_{pred_{test}} - \bar{Y}_{training})^2} \quad (8)$$

Where $Y_{pred_{test}}$, $Y_{exp_{test}}$ is predicted, experimental activity of the test respectively, and $\bar{Y}_{training}$ is the mean activity of the training set. A good and robust model will have R^2_{test} value ≥ 0.6.

2.8. Y-randomization test.

The Y-Randomization test is an external validation test performed on the training set to confirm the strength of the built model (Tropsha et al., 2003). For a QSAR model to pass Y-Randomization test the $cR_p{}^2$ value must be more than 0.5. The below equation is used for the calculation.

$$cR_p{}^2 = R[R^2 - (R_2)^2]^2 \quad (9)$$

2.9. Variance Inflation Factor (VIF).

The VIF is a measure of the multi-collinearity among the descriptors used in the model and is expressed as:

$$VIF = \frac{1}{1 - R^2} \quad (10)$$

R^2 is the multiple regression correlation coefficients of the variables within the model. If the VIF value falls in the range of 1-5, the model is good and acceptable, if the value is 1, it shows no collinearity among the descriptors and if is above 10, it shows that the model is not good and cannot not be accepted.

2.10. Applicability Domain.

Applicability domain was performed on the compounds to detect an outlier and influential molecules. The leverage approach was employed to describe the applicability domain of the QSAR model (Tropsha et al., 2003). Leverage of a given chemical compound is defined as follows:

$$l_i = X_i(X^TX)^{-1}X_i^T \quad (11)$$

l_i is the leverage of each compounds, X_i is the descriptor row-vector of the query compound i, and X is the (m × n) descriptor matrix of the compounds in the training set that were used to build the model. The warning leverage (l^*) is used to assess the leverages, molecule(s) with value greater than the leverage value is said to be an influential molecule. This was calculated using equation 12.

$$l^* = 3\frac{(k+1)}{n} \quad (12)$$

Where n is the total number of training set compounds and k is the number of descriptors in the model. The Williams plot is a plot of standardized residual against leverage employed to elucidate the relevance area of the model in terms of chemical space. Data is said to be an outlier if the standardized cross-validation residual value generated by the model is greater than ±3.

2.11. Mean Effect of the model (ME).

The mean effect was carried out on the training set to know the relative importance of each descriptors in the model built. This is defined as follows:

$$ME = \frac{B_j\sum_i^n D_j}{\sum_j^m(B_j\sum_i^n D_j)} \quad (13)$$

B_j is the coefficient of the descriptor j in the model, D_j is the value of each descriptors in the data matrix for each molecule in the training set, m and n are respectively the number of descriptors that appears in the model and the number of molecules in the training set (Minovski et al., 2013).

2.12. Strength of the Model.

The strength of the built model was evaluated using both the internal and external validation parameters. Table 2 below show clearly the standard validation parameters for a generally acceptable QSAR model (Veerasamy et al., 2011).

Table 2. Standard Validation Parameters for a good QASR model.

Validation parameters	Meaning	Values
R^2	Coefficient of determination	≥ 0.6
$P_{95\%}$	Confidence interval at 95% confidence level	<0.06
${Q_{cv}}^2$	Cross-validation coefficient	> 0.5
$R^2 - {Q_{cv}}^2$	Difference between R^2 and ${Q_{cv}}^2$	≤ 0.3
$N_{ext.\ test\ set.}$	Minimum number of external test sets	≥ 5
${R_{test}}^2$	Coefficient of determination for external test set	≥ 0.06
${cR_p}^2$	Coefficient of determination for Y-randomization	> 0.5

3. RESULTS AND DISCUSSION

Thirty-one compounds were subjected to an *in silico* studies to develop a QSAR model with a better activity against breast cancer. The compounds were drawn using ChemDraw and optimized using Spartan software 14.1.14 version to obtain the three-dimensional spatial conformers, after which the molecular descriptors were calculated with PaDEL descriptor software V.2.20 and 931 descriptors were calculated. The data were pre-treated to remove those with repeated or same activity and those with empty columns. They were then divided into training and test set. 70% of the total compounds (2, 4, 5, 8 10, 12,13, 14, ,19, 20, 21, 24, 25, 26, 27, ,28, 30, 32, 33, 34) make up the training set while the remaining 30% (1, 3, 6, 7, 9, 11, 18, 23, 29, 31) were the test set. The model was built with the training set utilizing the GFA-MLR from the material studio Software, the model was validated with the test set. Four models were generated and the first model was chosen as the optimum model because of its high potency, affinity, efficacy and selectivity (PAES). Table 3 shows the four equations and their definitions.

Table 3: Models equations and descriptors.

S/N	Equations	Definitions
1	pIC_{50} = 0.071169725*X505+0.009132493*X751-0.037066466*X758-0.023009609*X845+4.933312035	X505 : SL: minHBint2 X751 : ABX: WPSA3 X758 : ACE: RNCS X845 : AFN: RDF85e
2	pIC_{50} = -0.075172920*X505+0.033710209*X58-0.039221060*X758-0.023180487*X845+4.883406055	X505 : SL: minHBint2 X584:VM:ETAEta_F_L X758 : ACE: RNCS X845 : AFN: RDF85e
3	pIC_{50} = 0.072027698*X505+0.006497927*X580-0.038714612*X758-0.024063116*X845+4.955391798	X505 : SL: minHBint2 X580: VI: ETA_Eta_F X758 : ACE: RNCS X845 : AFN: RDF85e
4	pIC_{50} = -0.070392188*X505+0.000127795*X741-0.039846693*X758-0.025338322*X845+4.992974866	X505 : SL: minHBint2 X741 : ABN : DPSA-2 X758 : ACE : RNCS X845 : AFN : RDF85e

Model one was found to have pIC_{50} = -0.071169725* **minHBint2** + 0.009132493* - **WPSA-3** 0.037066466 * **RNCS** - 0.023009609 * **RDF85e** + 4.933312035. The descriptors used in the model were **minHBint2** which is a 2D structure and is Minimum E-State descriptors of strength for potential Hydrogen Bonds of path length 2, **WPSA-3** is a PPSA-3 * total molecular surface area / 1000**, RNCS** is a 3D Relative negative charge surface area -- most negative surface area * RNCG and **RDF85e** which is also a 3D molecule which means Radial distribution function - 085 / weighted by relative Sanderson electro-negativities. The validation parameters presented in table 4 passed the recommendations for building a good QSAR model when compared to the standard validation parameters; this indicate how potent and robust the model is.

Table 4: Validation parameters (VP)

Validation parameter	Equations			
	1	2	3	4
Friedman LOF	0.0042	0.0045	0.0047	0.0050
R-squared (R^2)	0.9853	0.9842	0.9835	0.9827
Adjusted R-squared (${R_{adj}}^2$)	0.9816	0.9803	0.9793	0.9784
Cross validated R-squared (${Q_{cv}}^2$)	0.9727	0.9719	0.9708	0.9688
Significance-of-regression F-value	268.4242	249.5214	237.7687	227.6479
Min exp. error for none-significant LOF (95%)	0.0238	0.0247	0.0253	0.0258

Table 5: Y- Randomization.

Model	R	R^2	Q^2
Original	0.9926	0.9853	0.9727
Model 1	0.3946	0.1557	-0.4277
Model 2	0.6179	0.3818	-0.2281
Model 3	0.5382	0.2895	-0.4008
Model 4	0.6202	0.3846	-0.1050
Model 5	0.4260	0.1815	0.4988
Model 6	0.6586	0.4338	-0.1035
Model 7	0.4171	0.1740	-0.3489
Model 8	0.4058	0.1647	0.3681
Model 9	0.2344	0.0550	-0.5589
Model10	0.1926	0.0371	-0.7175
Average randomized model			
Average R:	0.4505		
Average R^2:	0.2258		
Average Q^2:	-0.3757		
$cR_p{}^2$:	0.8780		

The Y-randomization result presented in table 5 was a test conducted on the training set to show the robustness of the model. The low values of both R^2 (0.2258) and Q^2 (-0.3757) for several trials affirm that the built model is stable, robust and reliable. While the $cR_p{}^2$ value (0.8780) greater than 0.5 confirm that the built model is powerful and was not deduce by chance. Other statistical analyses carried out on the model's descriptors are Pearson's correlation (PC), mean effect (ME) and the Variance inflation Factor (VIF). The PC shows the inter-correlation between each descriptor, the ME indicates the relative importance of each descriptor on the built model while the VIF shows that the model is strong and statistically acceptable. There was no inter-correlation between the descriptors because all the paired values were less than 1.0. The positive and negative value of the mean effect shows the strength of the model based on their magnitude and signs. Table 6 present this analysis.

Table 6: Pearson's correlation matrix, VIF and mean effect for the QSAR model descriptors.

Descriptors	Inter-correlation				VIF	Mean Effect
	minHBint2	WPSA-3	RNCS	RDF85e		
minHBint2	1	-0.0748	0.1683	-0.1786	1.0510	0.3973
WPSA-3	-0.0748	1	-0.5931	0.8074	2.9136	-0.4453
RNCS	0.1683	-0.5931	1	0.1683	2.2078	0.3013
RDF85e	-0.1786	0.8074	0.1683	1	4.1810	0.7467

The univariant statistical analysis presented in table 7 below shows that there is no significant difference in the mean, standard deviation and median values of the compounds. This indicates that the inhibitory activity of both training and test set are similar when compared. The insignificant difference in the range values of the two set indicates that the inhibitory activity of the two set are similar. The maximum values (4.8477 -5.0278) and the minimum values (4.1284 - 4.1013) of the training and test set respectively confirmed that the compounds are within the same range and the inhibitory activity of the compounds are interpolative.

Table7: **Statistical analysis**.

Statistical Analysis	Activity	
	Training set	Test set
Number of compounds	21	10
Confidence level (95%)	0.1026	0.2040
Mean	4.4943	4,4772
Median	4.6180	4.4619
Maximum	4.8477	5.0278
Minimum	4.1284	4.1013
Kurtosis	-1.3909	0.1187
Range	0.7193	0.9265
Skewness	-1.3909	0.1187
Standard deviation	0.2254	0.2852
Sample variance	0.0508	0.0814

Table 8 and 9 illustrate the low residual activity values for both training and test set, this confirmed the high predictive power of the built model.

Table 8: **Experimental, predicted, residual and standard residual activity for training set.**

S/N	Experimental activity	Predicted activity	Residual activity	Standardized Residual
2	4.1284	4.1612	-0.0328	-1.1999
10	4.1880	4.1923	-0.0025	-0.0917
4	4.1965	4.2109	-0.0144	-0.5256
12	4.2396	4.2270	0.0126	0.4597
24	4.6757	4.6928	-0.0171	-0.6259
5	4.2716	4.2742	-0.0026	-0.0945
8	4.2007	4.1803	0.0204	0.7458
19	4.6180	4.6773	-0.0593	-2.1731
20	4.5017	4.5122	-0.0106	-0.3866
21	4.6326	4.5946	0.0381	1.3935
22	4.6696	4.6617	0.0079	0,2891
13	4.3936	4.4191	-0.0254	-0.9315
14	4.2403	4.1992	0.0412	1.5071
26	4.6615	4.6923	-0.0307	-1.1248
27	4.8477	4.8075	0.0402	1.4728
30	4.6289	4.5907	0.0382	1.3996
32	4.6198	4.6173	0.0025	0.0907
33	4.6498	4.6498	-0.0001	-0.0025
34	4.6440	4.6391	0.0049	0.1789
25	4.5817	4.6111	-0.0294	-1.0749
28	4.7878	4.7689	0.0190	0.6943

Table 9: Experimental, predicted, residual and standard residual activity for the test set.

S/N	Experimental activity	Predicted activity	Residual activity	Standardized Residual
1	4.1013	4.9957	-0.8944	-1.3558
23	4.6326	4.4106	0.2220	0.7424
6	4.2434	4.8327	-0.5893	-0.8358
29	4.7258	4.2333	0.4925	1.2101
3	4.3979	4.9089	-0.5109	-0.6628
31	4.6289	4.2268	0.4022	1.1071
18	5.0278	4.4661	0.5617	1.3089
7	4.1858	4.9269	-0.7412	-0.9058
9	4.5258	4.7537	-0.2280	-0.1251
11	4.3028	4.5810	-0.2783	-0.4830

Figures 1 and 2 display the plot of experimental activity against the predicted activity for both training and test set. The two plots have R^2 value greater than 0.6 which indicate a strong and reliable model. Figure 3 display the plot of standardize residual against the experimental activities, all the compounds were found to be within the range value of ±2, this confirmed the strength and robustness of the model.

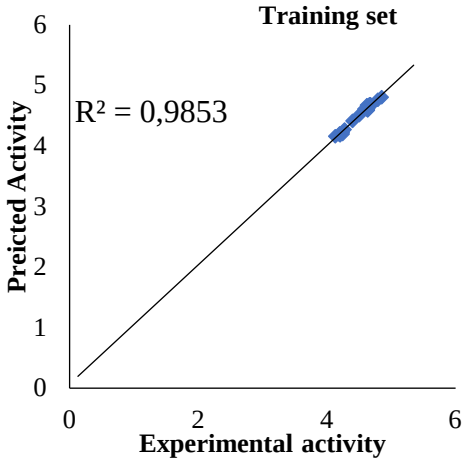

Figure 1 - Plot of experimental activity against predicted activity for training set.

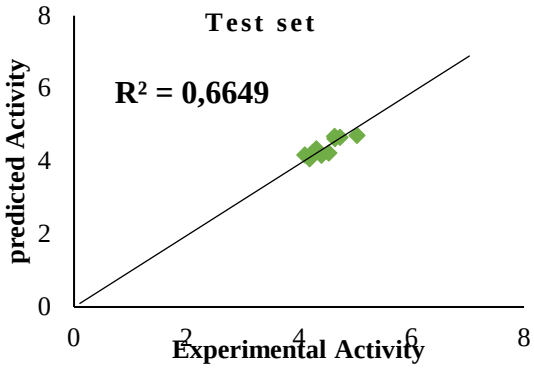

Figure 2 - Plot of experimental activity against predicted activity for test set.

The Williams plot in figure 4 above is a plot of standardized residual against leverage to know the influential compounds and outliers in the model. The result shows that all the compounds were within the limits square of ±3 except compounds 18, 23 and 31 from the test set that exceeded the calculated warning leverage ($l^* = 0.7$). This could be attributed to the difference in the chemical structure of those compounds and as such those compounds are said to be structurally influential compounds.

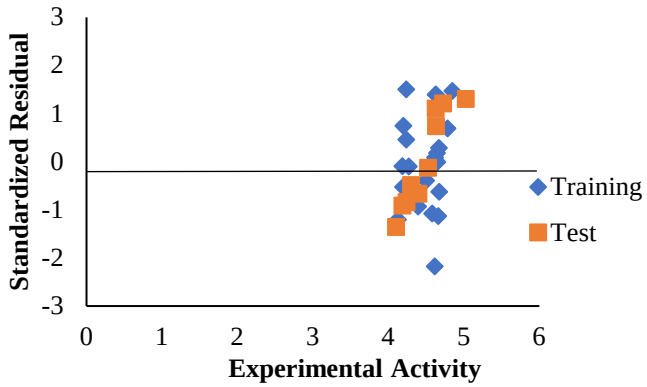

Figure 3 - Plot of standardized residual versus experimental

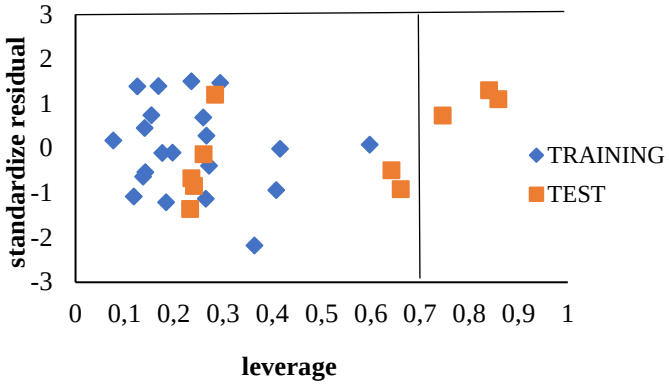

Figure 4 - Plot standardize residual against leverage

4. CONCLUSION

The result of this work in all ramification passed the minimum recommendation for building a good QSAR model, with values of R^2 = 0.9853, adjusted R^2 = 0.9816, $Q_{cv}{}^2$ = 0.9727 and an external validated R^2 = 0.6649. The applicability domain and the low residual values both confirmed that the built model is robust and has a high predictive power which satisfies the research aim. Conclusively, this work would give first-hand information to the medicinal chemist, pharmacist and pharmacologist when developing a new anti-breast cancer agents.

REFERENCES

COLEMAN, M. P.; QUARESMA, M.; BERRINO, F.; LUTZ, J. M.; DE ANGELIS, R.; CAPOCACCIA, R.; BAILI, P.; RACHET, B.; GATTA, G.; HAKULINEN, T.; MICHELI, A.; SANT, M.; WEIR, H. K.; ELWOOD, J. M.; TSUKUMA, H.; KOIFMAN, S.; SILVA, G. A.; FRANCISCI, S.; SANTAQUILANI, M.; VERDECCHIA, A. STORM, H. H.; YOUNG, J. L. Cancer survival in five

continents. A worldwide population-based study *Lancet Oncol*. 2008, 9:730-756 [PMID: 18639491 DOI:10.1016/S1470-2045(08)70179-7].

FERLAY, J.; SHINH. R.; BRAY, F.; FORMAN, D.; MA THERS, C.; PARKIN, D. M. Estimates of worldwide burden of cancer in **2008**. Globocan 2008, *Int J Cancer*. 2010, 127: 2893- 2917 [PMID: 21351 DOI: 10.1002 /ijc.25516]

FERLIN, M. G.; CHIARELOTTO, G.; GASPAROTTO, V.; DALLA, V. L.; PEZZI, V.; BARZON, L.; PALU`, G.; CASTAGLIUOLO, I. *J. Med. Chem*.**2005**, 48, 3417, DOI: 10.1021/jm049387.

FRANKISH H. 15 Million new cancer cases per year by 2020, says WHO, *Lancet*, 361.**2003**, 1278–1287, DOI: 10.1016/S0140-6736(03)13038-3.

FRIEDMAN, J. H.. Multivariate adaptive regression splines. *Ann. Stat*, **1991**, 1-67.

GASPAROTTO, V.; CASTAGLIUOLO, I.; CHIARELOTTO, G.; PEZZI, V.; MONTANARO, D.; BRUN P.; PALU, G.; VIOLA, G.; FERLIN, M. G. *J. Med.Chem.* **2006**, 49, 1910 DOI:10.1021/jm0510676.

KENNARD, R. W.; STONE, L. A.. Computer aided design of experiments. Technometrics. *J. Sci. Res.***1969,** 11, 137-48, DOI: 10.1080/00401706.1969.10490666.

LARIF, M.; CHTITA, S.; ADAD, A.; HMAMOUCHI, R.; BOUACHRINE, M.; LAKHLIFI, T.. Predicting biological activity of Anticancer Molecules 3-ary hydroxyquinolin-2 (1H)-one by DFT-QSAR models, *interternational Journal of Clinical and Experimental Medicine*, **2013**, vol.3, pp.32– 42, ISSN: 2319-8753.

MINOVSKI, N.; ˇZUPERL, ˇS. DRGAN, V.; NOVIˇC M. Assessment of applicability domain for multivariate Counter propagation artificial neural network predictive models by minimum Euclidean distance space analis. A case study, *Analytica Chimica Acta*. **20** vol. 759, pp. 28–42, DOI:10.1016/j.aca.2012.11.002.

SINGH, P. Quantitative Structure-activity relationship study of substituted-[1,2,4] oxadiazoles as $S1P_1$ agonists, *J. Curr. Chem. Pharm. Sci*.**2013,** 3, 64- 79, ISSN: 2277-2871.

SHOLA, E. A.; UBA, S.; UZAIRU A.. QSAR Modeling and Molecular Docking Analysis of Some Active compou- against Mycobacterium tuberculosis Receptor (MtbC YP121), *J. Pathog*. **2018**, DOI:10.1155/2018/1018694.

TROPSHA, A.; GRAMATICA, P.; GOMBAR, V. K. The importance of being earnest: validation is the absolute essential for successful application and interpret tation of QSPR models. *M Inform.* **2003**, 22, 69-77, DOI:10.1002/qsar.200390007.

VEERASAMY, R.; RAJAK, H.; JAIN, A.; SIVADASAN, S.; VARGHESE, C. P.; AGRAWAL, R. K. Validation of QSAR models-strategiesand importance, *International Journal of Drug Design and Discovery*.**2011**,vol. 3, pp. 511–519.

WONG, K. Y.; MERCADER, A. G.; SAAVEDRA, L. M.; HO NARPARVAR, B.; ROMANELLI, G. P.; DUCHDU- CHOWICZ, P. R. QSAR analysis on related acetyl cholinesterase inhibitors, *Journal of Biomedical science* **2014,** vol. 21, no.1, DOI: 10.1186/s12929- 0140084 -0.