Turkish Computational and Theoretical Chemistry

Turkish Comp Theo Chem (TC&TC)

Volume(Issue): 4(1) – Year: 2020 – Pages: 12-23

e-ISSN: 2602-3237

https://doi.org/10.33435/tcandtc.614263

Received: 02.09.2019 **Accepted**: 10.05.2020 *Research Article*

QSAR and Molecular Docking Studies of novel thiophene, pyrimidine, coumarin, pyrazole and pyridine derivatives as Potential Anti-Breast Cancer Agent

Momohjimoh Ovaku IDRIS[1], Stephen Eyije ABECHI, Gideon Adamu SHALLANGWA, Adamu UZAIRU

Department of Chemistry, Ahmadu Bello University Zaria-Nigeria

Abstract: Quantitative Structure Activity Relationship (QSAR) and molecular docking studies were carried out on some novel compounds to generate a good QSAR model that relate the anti-breast cancer activity values with their molecular structure. Genetic Function Algorithm (GFA) and Multiple Linear Regression Analysis (MLRA) were employed to select the descriptors that were used to build the models. The best model built was found to have statistical validation values of squared correlation coefficient R^2 = 0.9845, adjusted squared correlation coefficient R_{adj}^2 = 0.9814, cross validation coefficient Q_{cv}^2 = 0.9763 and an external squared correlation coefficient R_{ext}^2 = 0.8240 which was used to confirm the validation of the model. The docking results showed that ligands 12 with binding energy (-9.3kcalmol^{-1}) have the highest binding affinity when compared to the reference drug doxorubicin with binding energy (-6.8kcalmol^{-1}). The stability and robustness of the built model showed that new anti-breast cancer agents can be design from these derivatives.

Keywords: *Breast Cancer, QSAR model, Model Validation, Binding Affinity.*

1. Introduction

Cancer is the abnormal growth of the cell and is the second leading cause of death after circulatory diseases. The World Health Organisation (WHO) predicted 15 million death cases by the year 2020 unless a new measure is taking [1].

Breast cancer is the most common cancer in women [2], it usually develop from breast tissue. In Nigeria, cervical cancer was the commonest cause of cancer- related deaths among women for several decades but breast cancer is now the leading cause of cancer related deaths among Nigerian women [3].

Doxorubicin is one of the numerous hypothetical anti-cancer drug used in treatment of all kinds of cancer, the recent emergence of resistance to this available anti-cancer drug calls for immediate need to develop new anti-cancer agents. Development of new drug is by trial and error approach and this is time consuming.

QSAR is computational method that decode the relationship between the structure of a molecule and the activity of such molecule in a numerical form [4]. Application of this technique have been employed in drug discovery to design new drugs and also improved the existing ones because its time saving and lesser cost.

Therapeutic treatment of cancer usually focuses on targeting critical cellular processes involved in DNA replication and cell division. This method consist of different set of agents each targeting different pathways and enzymes. One class of agents, predominantly effective at disrupting cancer cell growth, are drugs targeting DNA topoisomerases [5], this is why this work uses top2A as the receptor.

DNA topoisomerases are a family of enzymes originate in the nucleus and the mitochondria that are responsible for maintaining DNA topology [6]. DNA topology refers to the relationship between

[1] Corresponding Authors

e-mail: eedrismj@gmail.com

the two strands of the double helix and includes the concept of supercoiling [7].

Type II topoisomerases (Topo2α) form a transient double-strand DNA break in one segment which can pass one DNA segment to another through the break prior to ligating the cleaved DNA ends. Type II topoisomerases found in living organisms is divided into IIA and IIB [8]. They vary in terms of structure, mechanism and cofactor. Type II enzymes works either to enhance different chromosomes (e.g., for chromosome segregation and unknotting) or sections of the same chromosome (e.g., during transcription and replication) [6].

Molecular docking is a computational technique used to predict accurately the binding score of a complex (ligand-receptor interaction) [9], information derived can then be used to evaluate the energy profiling, such as binding energy, bond length, bond strength and binding constant. The QSAR models were developed using Drug Theoretic and Cheminformatic (DTC) Laboratory software tool while the docking studies was achieved using the discovery studio and Auto-duck Vina of the PyRx.

The aim of this work is to generate a robust QSAR model and perform a flexible docking studies on those aforementioned compounds that would serve as raw data to the pharmacologist and pharmacist for rational drug designing (structure-based-drug development) of new anti-breast cancer agents with better efficacy [10].

2. Materials and Methods

2.1. Data Collection and Activity Evaluation

The data set used in this work was collected from the literature [11]. The structure of the compounds were drawn with Chem Draw software and optimized with Spartan software to remove the strained energy.

The biological activities of the compounds were provided in the literature as fifty percent growth inhibition concentration (GI_{50}), they were converted to logarithm unit (pGI_{50}) using the equation 1 below for simplicity. The structures of the compounds and their biological activities were presented in Supp. Table S1.

$$pGI_{50} = (GI_{50} \times 10^{-3}) \quad (1)$$

2.2. Molecular Descriptors Calculation and Data Pretreatment

The two dimensional structures (2D) of the compounds presented in the Table 1 were drawn with Chemdraw software version 12.0.2 [12], they were exported to the Spartan 14 V1.1.4 Wave Function programming package software to view the spatial conformers of the compounds, i.e, three dimensional (3D) structures. These 3D structures were geometrically optimized using Density Functional Theory (DFT) method, utilizing the (B3LYP/6-31G*) hybrid function known as Becke's three parameter exchange functional (B3) hybrid with Lee, Yang and Parr correlation functional (LYP) [13, 14]. The optimized molecules in Spartan files format were converted to SD format and saved which was subsequently exported to PaDEL-Descriptor software V2.20 [15] to calculate the molecular descriptors.

Molecular descriptors are numerical description of molecules. The descriptors of all the 34 molecules were calculated using PaDEL-Descriptor software V2.20 and a total of 1875 molecular descriptors were calculated.

The data set was pre-treated with data pre-treatment software from Drug Theoretics and Cheminformatics Laboratory (DTC Lab) so as to remove uninformative data [16].

2.3. Model Generation and Validation

To generate a good QSAR model, the pre-treated data set was divided into two subset (training and test set), using the data division software from DTC Lab [17-19]. The training set comprised of 70% of the total molecules and the rest of the molecules were test set. The training set was used to build the models employing GFA-MLR method from the material studio and the test set were used to validate the model built [20]. The fitness score of the models were evaluated using the leave one out (LOF) giving by the equation 2.

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

Where SEE is the standard error of estimation, C is the number of terms in the model, d is a user defined smoothing parameter, P is the total number of descriptors contained in the model and M is the number of training set data. SEE is defined by equation (3)

$$SEE = \sqrt{\frac{(Y_{exp}-Y_{pre})^2}{N-P-1}} \quad (3)$$

Where Y_{exp} and Y_{pre} are the experimental activity and the predicted activity in the training set

respectively [21]. The squared correlation coefficient is a validation test used to compare the predicted and experimental activities. The closer R^2 value is to 1.0 indicates a good and strong model. R^2 is expressed as:

$$R^2 = 1 - [\frac{\Sigma(Y_{exp}-Y_{pred})^2}{\Sigma(Y_{exp}-\bar{Y}_{training})^2}] \quad (4)$$

Where Y_{exp}, Y_{pred} and $\bar{Y}_{training}$, are respectively the experimental activity, the predicted activity, and the mean experimental activity of the samples in the training set. The R^2 value alone cannot justify the goodness of the model as such it was adjusted to give a stable and reliable value. If the difference between the R^2 and R^2_{adj} value is less than 0.3, it indicates that the number of descriptors used in building the model are appropriate and the model would be accepted but a value greater than 0.3 will be rejected. The adjusted R^2 is givens by:

$$R^2_{adj} = \frac{R^2-k(n-1)}{n-P+1} \quad (5)$$

Where k is the number of descriptors in the model and n is the number of training set compounds [20]. Cross-validation test is used to measure the predictive ability of the model. The cross validation coefficient Q^2_{cv} is defined as:

$$Q^2_{cv} = 1 - [\frac{\Sigma(Y_{exp}-Y_{pred})^2}{\Sigma(Y_{exp}-\bar{Y}_{training})^2}] \quad (6)$$

To be certain that the built model is firm and not infer by chance, the model is further put to an external validation test. This is calculated as thus;

$$R^2_{test} = 1 - \frac{\Sigma(Y_{pred_{test}}-Y_{exp_{test}})^2}{\Sigma(Y_{pred_{test}}-\bar{Y}_{training})^2} \quad (7)$$

Where $Y_{pred_{test}}$ is the predicted activity, $Y_{exp_{test}}$ is the experimental activity of the test set and $\bar{Y}_{training}$ is the mean activity of the training set [21].

2.4. Y- Randomization Test

Y-randomization test is a test performed on the training data set to ascertain that the descriptors used to build the models were appropriate and to also know how strong the built model is. The test was done by randomly mixing the activity data which was taking as the dependent variable and the descriptors as the independent variable. After several trials, the new QSAR models generated were found to have very low R^2, Q^2 and a randomized square correlation coefficient $(cR_p{}^2)$ with value greater than (0.5) that confirmed the robustness of the models.

$$cR_p{}^2 = R \times [R^2 - (R_r)^2]^2 \quad (10)$$

Where $cR_p{}^2$ is the coefficient of determination for Y-randomization and R_r is the average 'R' of random models [20].

2.5. Mean Effect of the Model

The mean effect of the model is a test used to show the comparative importance of each descriptors present in the model. This was calculated using equation (11)

$$ME = \frac{B_j \Sigma_i^n D_j}{\Sigma_j^m (B_j \Sigma_i^n D_j)} \quad (11)$$

Where B_j is the coefficient of the descriptor *j* in the model, D_j is the value of each descriptor in the data matrix for each of the training set data, m and n are respectively the number of descriptors that appears in the model and the number of molecules in the training set [22].

2.6. The Predictive Power of the Model

The selectivity, efficacy and potency, (SEP) of the developed models were evaluated using both internal and external validation parameters, its applicability domain and the Variance Inflation Factor (VIF). Table 2 below show clearly the standard validation parameters for a generally acceptable QSAR model [23].

The applicability domain is a test performed on the training set to confirm the robustness of the built models. The leverage approach was employed to describe the applicability domain of the QSAR model [24]. Leverage of a given chemical compound is defined as:

$$l_i[X_i(X^TX)^{-1}X_i{}^T] \quad (12)$$

Table 2. Minimum Validation Parameters for generating good QASR model.

Validation parameter	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	≥ 0.6
${cR_p}^2$	Coefficient of determination for *Y*-randomization	> 0.5

Where l_i is the leverage of each molecule is, X_i is the descriptor row-vector of the query compound *i*, and *X* is the (m × n) descriptor matrix of the training set molecules used in building the model. The warning leverage (${l_i}^*$) showed the molecule(s) that exceeded the leverage value. This can be calculated using equation 13.

$$ {l_i}^* = \frac{3(k+1)}{n} \qquad (13) $$

Where is n the number of training set molecules and k is the number of descriptors in the model. The Williams plot was the plot of standardized residual against leverage employed to elucidate the relevance area of the model in terms of chemical space. Any data in the plot with value greater than ±3 would be treated as outlier.

The VIF is a measure of multi-collinearity between the descriptors used to generate the model and is expressed as:

$$ \text{VIF} = \frac{1}{1-R^2} \qquad (14) $$

Where R^2 is the correlation coefficient of the multiple regression between the variables within the model. A good and acceptable model would have its VIF values ranges from 1-5.

2.7. Docking Studies

Molecular docking was carried out to evaluate the binding affinity of the ligands to the receptor. The 3D-structure of the receptor (Top2$\boldsymbol{\alpha}$) was downloaded from RCSB PDB (http://www.rcsb.org/pdb/home/ home.do) with code 4fm9 [25]. Removable residue like cofactors, ligands, water molecule were found absent. The ligands were prepared by converting the optimized 3D structures from SD format to protein data bank format (pdbqt). The prepared receptor and ligands were docked together with the Auto Dock Vina of the PyrX software and the complex was visualized utilizing the discovery studio software visualizer. Figure 1 shows the 3D structure of the receptor (topoisomerase).

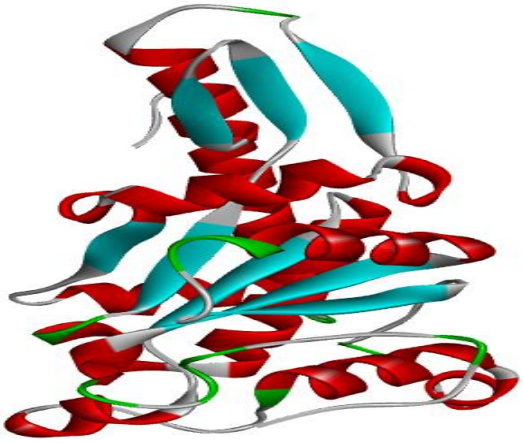


Figure1. 3D structure of Topoisomerase (ii).

3. Results and Discussion

The 34 compounds comprising of some novel thiophene, pyrimidine, coumarin, pyrazole and pyridine derivatives were subjected to QSAR and molecular docking studies to generate good QSAR model with better activity against breast cancer. The compounds were optimized, their descriptors were calculated and they were divided into training and test set employing the Kennard-Stone method of data division [19]. The training data set were used to develop the model and the test data set were used to validate the built model. The Genetic Algorithm and multi Linear Regression (GA-MLR) from the material studio was employed to build the models and three models were built. Table 3 present the statistical validation parameters of the built models, the first model was carefully chosen and reported as the best model because of its statistical validation values when compared to the minimum required validation parameters presented in Table 2.

Table 4 present the experimental, predicted and residual activity values for both the training and test set. The low residual values (difference between the

experimental activity and predicted activity) indicates the high predictive power of the built model.

Table 3. Validation parameters of the built models.

Validation parameters	Meaning	Values
R^2	Coefficient of determination	0.9845
$P_{95\%}$	Confidence interval at 95% confidence level	0.0195
${Q_{cv}}^2$	Cross-validation coefficient	0.9763
$R^2 - {Q_{cv}}^2$	Difference between R^2 and ${Q_{cv}}^2$	0.0082
$N_{ext.\ test\ set.}$	Minimum number of external test sets	11
${R_{test}}^2$	Coefficient of determination for external test set	0.8240
${cR_p}^2$	Coefficient of determination for *Y*-randomization	0.8200

Best Model

pGI_{50} = 0.709363893 * GATS8 - 4.252846824 *maxHBd - 0.063150018 * TDB10p - 0.153565552 * RNCS + 4.211504042.

Table 4a: Experimental, Predicted and Residual Activity values training set**.**

S/N	Experimental Activity	Predicted activity	Residual activity
2	1.5670	1.5673	-0.0003
3	2.5229	2.4293	0.0936
4	1.3778	1.4520	0.0742
5	1.2899	1.3575	-0.0676
7	1.3188	1.3985	-0.0797
10	4.0458	4.0494	-0.0037
11	1.4214	1.5261	-0.1047
12	1.4056	1.3025	0.1031
13	1.3947	1.3116	0.0831
14	1.4802	1.4158	0.0644
15	1.5376	1.4644	0.0732
16	1.4450	1.4122	0.0328
17	1.3546	1.3360	0.0186
19	1.4353	1.3687	0.0666
20	1.3851	1.4706	-0.0855
21	1.3696	1.2988	0.0708
22	1.4067	1.5193	-0.1126
23	1.3449	1.2428	0.1021
24	1.4157	1.4835	-0.0678
31	1.3862	1.4618	-0.0756
32	1.4012	1.4220	-0.0208
33	2.5086	2.4212	0.0874
34	2.2076	2.3106	-0.1030

Table 4b: Experimental, Predicted and Residual Activities for test set

S/N	Experimental activity	Predicted activity	Residual activity
1	1.6253	1.2498	0.3755
6	1.3546	1.5116	0.1570
8	1.3788	1.3171	0.0617
9	4.0000	3.3794	0.6206
18	1.4572	1.4106	0.0466
25	1.6635	1.8581	-0.1946
26	1.6737	1.5049	0.1688
27	1.7904	1.2847	0.5057
28	1.5988	1.3909	0.2079
29	1.7011	1.0303	0.6708
30	1.6946	1.1253	0.5693

Table 5: Pearson's correlation, VIF and ME.

Descriptors	Inter-correlation				VIF	Mean Effect
	GATS8c	maxHBd	TDB10p	RNCS		
GATS8c	1	0.2208	-0.2053	0.3934	1.2216	-0.2514
MaxHBd	0.2208	1	0.0884	0.1704	1.1071	0.8382
TDB10p	-0.2053	0.0884	1	-0.4878	1.373	0.2425
RNCS	0.3934	0.1704	-0.4878	1	1.5357	0.1706

The result of the Variance Inflation Factor (VIF), Pearson's correlation and the Mean Effect (ME) were presented in table 5. These tests showed the relative importance of each descriptor in the model, the inter-correlation and the collinearity between the descriptors. Their low values infers that the descriptors were well chosen and the built model is said to be statistically satisfactory [26].

The Y-randomization result presented in table 6 is an external validation test conducted on the training data set to confirm the robustness of the model. The coefficient of y-randomization cR_p^2 with value 0.8200 which is greater than the standard value reported in table 2 above clearly shows the built model is highly robust.

Table 6: Y- randomization result

Models	R	R^2	Q^2
Original	0.9436	0.8904	0.7631
Model 1	0.3428	0.1175	-0.2877
Model 2	0.2754	0.0758	-0.5373
Model 3	0.1695	0.0287	-0.4698
Model 4	0.1420	0.0202	-0.2828
Model 5	0.5947	0.3536	-0.1640
Model 6	0.2385	0.0569	-0.1526
Model 7	0.3906	0.1525	-0.1264
Model 8	0.1609	0.0259	-0.2854
Model 9	0.5167	0.2670	-0.8873
Model10	0.8452	0.7144	0.3324
Average randomized model			
Average R:	0.3676		
Average R^2:	0.1813		
Average Q^2:	-0.2861		
cR_p^2 :	0.8200		

Table 7: Details of the descriptors used to build the models

S/N	Descriptors	Description	Number	Class
1	GATS8c	Geary autocorrelation - lag 8/ weighted by charges	346	2D
2	MaxHBd	Maximum E-States for (strong) Hydrogen Bond donors	489	2D
3	TDB10p	3D topological distance based autocorrelation - lag 10 / weighted by polarizabilities	80	3D
4	RNCS	Relative negative charge surface area -- most negative surface area * RNCG	29	3D

Table 8: Binding energy (BE) and the hydrophobicity interaction of the ligands:

S/N	BE (Kcalmol^{-1})	Target	Hydrogen Bond		Hydrophobicity Interactions
			Amino Acids	Bond length (Å)	Amino acids
1	-4.7	Topo2α	TRP608 HIS605 LYS606	2.81 2.48 2.34	VAL610
2	-4.5	Topo2α	GLU623 LYS579 LYS609	2.95 2.24 2.13	
3	-6.3	Topo2α	THR453 HIS567	2.90 2.40, 2.71	PHE569, LEU531, LEU528
4	-5.9	Topo2α	GLU572 GLN542 LYS550 ASP543 SER547	2.87 2.69 2.73 2.20 2.96	
5	-6.3	Topo2α	GLU586 ARG633 ALA588	2.26 1.92 2.17	GLU626, HIS634
6	-5.9	Topo2α	GLU682 LEU680	2.96 2.48	LYS676, ARG672, PRO681, PRO593
7	-5.5	Topo2α	ARG568 ARG635 GLN637	4.87 2.12 1.81, 2.26	ILE636, LYS632
8	-5.8	Topo2α	LEU680 SER600 GLY679	2.42, 2.65 2.58 2.92	ARG672, ARG675, PRO681
9	-6.5	Topo2α	ASP541 GLU461	2.06 2.86	GLY488, ASP543
10	-6.4	Topo2α	ASP541 GLU461	2.10 2.73	LEU616, GLY615, ASP543
11	-9.1	Topo2α	GLU572 LYS550	2.41 2.84	ILE574, ILE636, ILE554, PHE638
12	-9.3	Topo2α	LEU516 TYP518 LYS519	2.52 2.56 2.65	LYS512, ILE511, GLN517
13	-7.1	Topo2α	TYR518 ASN560 VAL493 ALA496	1.96 2.86 2.72, 2.30 2.36, 2.95	HIS559, HIS498
14	-6.9	Topo2α	TYR18	2.38	LYS519, ILE501

			HIS559	2.49	
			ASN560	2.18	
			ALA496	2.42, 2.34	
15	-8.0	Topo2α	ALA496	2.02	ILE501, LYS519
			ASN560	2.68	
			HIS559	2.07	
			VAL493	2.74	
16	-7.9	Topo2α	ASN560	2.34	
			ALA496	2.37	ILE501, LYS519
			HIS559	2.04	
17	-6.7	Topo2α	GLU682	1.94, 3.00	ARG672, ARG675
			PRO593	2.57	
			SER600	1.93	
18	-8.7	Topo2α	GLU682	2.77	ARG675, ARG672,
			GLU596	2.67	PRO681
			LEU592	2.72	
19	-7.1	Topo2α	ASN560	2.25	ILE501, LYS519
			HIS559	2.40	
			TYR518	2.35	
20	-7.5	Topo2α	GLU525	2.42	LYS529
			THR453	2.45	
			LEU528	3.71	
			ARG568	2.40	
			HIS567	1.97	
21	-7.2	Topo2α	ASP543	2.63	ILE577
			GLN542	2.55	
			SER547	2.27	
			LYS550	2.36	
			GLU572	3.03	
22	-8.3	Topo2α	ALA648	2.65	ALA652, LEU565, ILE554, PHE653
23	-7.4	Topo2α			ASP645, ALA648, ILE554, ILE649 LEU565, LYS550, PHE653
24	-6.9	Topo2α	LYS639	2.75	ALA652, PHE653, LYS550, ILE649, ILE554, PHE638
25	-6.2	Topo2α	ALA465	2.61	
			ASP541	2.64	GLU461
			ASP543	2.22, 2.59	
			LYS614	2.40, 2.83	
26	-6.9	Topo2α	GLY617	2.08	GLU461
			LEU616	3.03	
			ASP541	4.98	
			ASP543	2.18	
27	-7.1	Topo2α	ASP630	2.77	ALA629
			TYR590	2.94	
			GLU586	2.71	
			GLU626	5.30	
28	-7.1	Topo2α	ASP630	2.83	ALA629
			TYR590	2.51	
			GLU586	4.39	
			GLU626	2.95	
29	-7.0	Topo2α	ALA588	2.99	HIS634, PHE589,
			GLU586	2.38	MET587
			ARG633	2.39	

			TYR590	2.21, 2.70	
			LYS579	2.71	
30	-7.4	Topo2α	GLU589	2.60	ALA629
			GLU626	2.62, 2.56	
			ASP630	2.52	
			TYR590	2.59	
31	-8.9	Topo2α	HIS559	2.01	PRO562
			TYR518	2.43	
			ASN560	2.36	
32	-8.2	Topo2α	HIS559	1.98, 2.45	ILE501
			ASN560	2.43	
			TYR518	2.70	
33	-8.2	Topo2α	HIS567	3.03	LEU528
			THR453	2.34	LEU531
					PHE569
34	-8.1	Topo2α	LYS639	3.04	PHE638, PHE653,
			ALA648	2.59	ILE,
Doxorubicin	-6.8	Topo2α	LEU516	2.39	GLN517, ARG532
			ASN433	2.00	
			THR530	2.95	
			LYS520	2.95	

Figure 2 and 3 represent the plot of predicted activity against experimental activity for both training set and test set. The square correlation coefficient R^2 values for the two plots were greater than 0.5 which passed the minimum requirement for a good QSAR model.

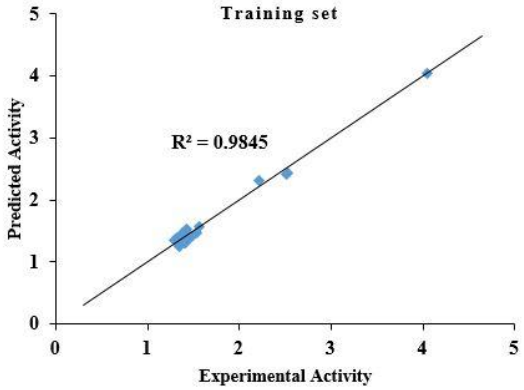


Figure 2: Plot of predicted activity against experimental activity for training set.

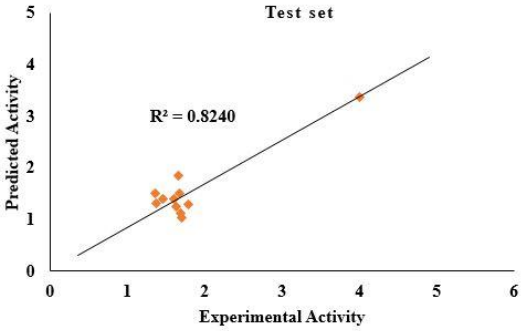


Figure 3. Plot of predicted activity against experimental activity for test set.

Figure 4 show the plot of standardized residual activity against the leverages, this plot is otherwise called Williams plot. This plot basically helps to illustrate the outliers and influential compounds and in this work there were four whose leverage value goes beyond the calculated warning leverage (l^* = **0.65**), and were treated as outliers. Figure 5 display the plot of standardized residual against the experimental activity and for the fact that the scattered plot were all within the base line of the graph, it indicate that there are no significant systematic error.

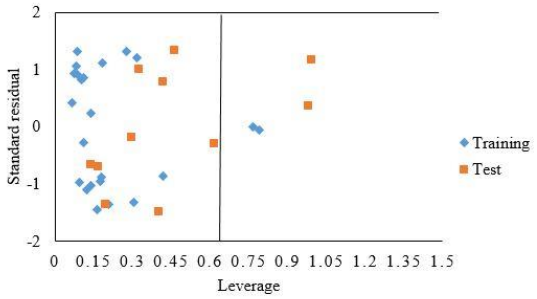


Figure 4: Plot of standardize residual against leverages.

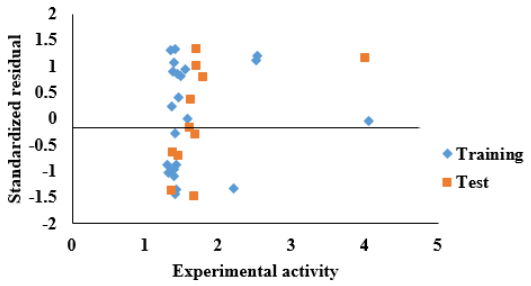
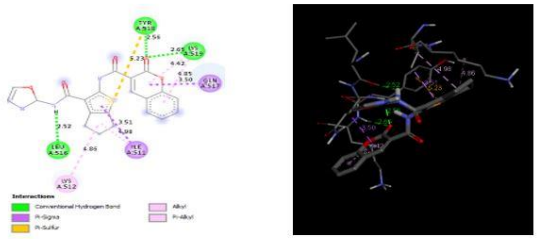


Figure 5: Plot of standardize residual against Experimental activity.



(a) 2D interaction of ligand 12. (b) 3D interaction of ligand 12

Figure 6: 2D and 3D interaction of ligand 12.

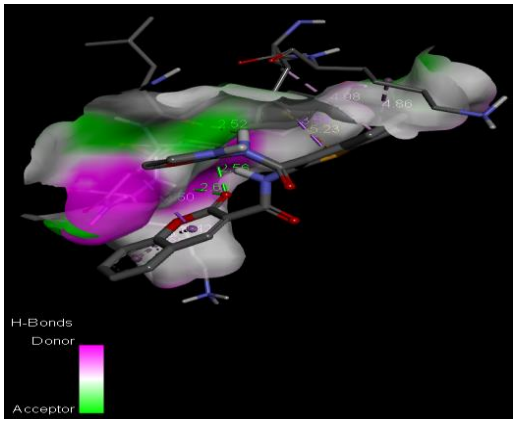


Figure 7: Hydrogen interaction of ligand 12.

Table 7 present the class and nature of the descriptors used to build the model, while table 8 clearly shows the binding energy, hydrogen Bond and Hydrophobicity interactions of the complex. In the table, it clearly shows that the ligands have binding energy that ranges from -4.5 kcal/mol to -9.3 kcal/mol. Ligand 12 was found to have the highest binding energy of -9.3 kcal/mol and bind strongly into the pocket of the receptor than the reference drug Doxorubicin with -6.8 kcal/mol binding energy. The visualized 2D and 3D structure of ligand 12 is presented as figure 6. Figure 7 however present the hydrogen bond interaction between ligand 12 and topo2$\boldsymbol{\alpha}$, this shows three hydrogen bond interaction of bond lengths 2.52 Å, 2.56 Å, and 2.65 Å with LEU516, TYP518 and LYS519 amino acid residues of the target and also three hydrophobicity interaction with LYS512, ILE511 and GLN517 of the target site. The N-H in the 2-methyloxazol-amine of ligand 12, acts as hydrogen acceptor and formed a hydrogen bond with LEU516 of the target. While the C=O in the 2H-chromen-2-one of the ligand acts as hydrogen donor and formed two hydrogen bond with TYR518 and LYS519 of the target.

4. Conclusion

This work has successfully built a good and robust QSAR model that passed all the minimum recommendations for building a good QSAR model. The Williams plot however pointed out four compounds out of the 34 compounds as outliers and may not be considered when designing a new anti-breast cancer agents from the derivatives. Conclusively ligand 12 with the highest binding energy can serve as better drug against breast cancer.

References

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

[2] H.D. Nelson, M.E. Smith, J.C. Griffin, R. Fu, Use of medications to reduce risk for primary breast cancer: a systematic review for the U.S. Preventive Services Task Force. Annals of Internal Medicine 158 (2013) 604-14.

[3] M.N. Okobia, C.H. Bunker, F.E. Okonofua and U. Osime Knowledge, Attitude and Practice of Nigerian Women towards Breast Cancer: A Cross-Sectional Study. World Journal of Surgical Oncology. 4 (2006) 11–15.

[4] Rathod, Antifungal and Antibacterial activities of Imidazolyl pyrimidines

derivatives and their QSAR Studies under Conventional and Microwave-assisted, Int J PharmTech Res. 3 (2011) 1942–1951.

[5] Y. Pommier, E. Leo, H. Zhang, and C. Marchand, DNA Topoisomerases and Their Poisoning by Anticancer and Antibacterial Drugs, Chem. Biol. 17 (2010) 421–433.

[6] Y. Pommier, Y. Sun, S.N. Huang, and J.L. Nitiss, Roles of Eukaryotic Topoisomerases in Transcription, Replication and Genomic Stability, Nat. Rev. Mol. Cell Biol. 17 (2016) 703–721.

[7] J.E. Deweese, and N. Osheroff, The DNA Cleavage Reaction of Topoisomerase II: Wolf in Sheep's Clothing. Nucleic Acids Res. 37 (2009) 738–749.

[8] P. Forterre, and D. Gadelle, Phylogenomics of DNA Topoisomerases: Their Origin and Putative Roles in the Emergence of Modern Organisms, Nucleic Acids Res. 37 (2009) 679–692.

[9] B. Mukesh, and K. Rakesh, Molecular Docking: A Review. IJRAP, 2 (2011)1746-1751.

[10] I.A. Guedes, C.S. de Magalhães and L.E. Dardenne, Receptor–Ligand Molecular Docking: Biophysical Reviews. (2014) 6 75-87

[11] Albratty, M., El-Sharkawy, K. A., and Alam, S., (2017). Synthesis and Antitumor Activity of Some Novel Thiophene, Pyrimidine, Coumarine, Pyrazole and Pyridine Derivatives. Acta Pharm. 67, 15–33, DOI: 10.1515/acph-2017-0004.

[12] Z. Li, H. Wan, Y. Shi, P. Ouyang, Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch, Journal of Chemical Information and Computer Sciences. 44 (2004) 1886–1890.

[13] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, Physical Review B. 37 (1988) 785.

[14] A.D. Becke, Becke's three parameter hybrid method using the LYP correlation functional, J. Chem. Phys. 98 (1993) 5648– 5652.

[15] S.E. Adeniji, S. Uba, A. Uzairu, In Silico Study For Investigating and Predicting the activities of 1, 2, 4-triazole derivaties as potent anti-tubercular agents, The Journal of Engineering and Exact Sciences. 4 (2018) 246–254.

[16] P. Singh, Quantitative Structure-Activity Relationship Study of Substituted-[1, 2, 4] Oxadiazoles as S1P1 Agonists, Journal of Current Chemical and Pharmaceutical Sciences. 3 (2013) 334–345.

[17] G. Melagraki, A. Afantitis, K. Makridima, H. Sarimveis, O. Igglessi-Markopoulou, Prediction of toxicity using a novel RBF neural network training methodology, Journal of Molecular Modeling. 12 (2006) 297–305.

[18] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. IgglessiMarkopoulou, A novel QSAR model for predicting induction of apoptosis by 4-aryl4H-chromenes, Bioorganic & Medicinal Chemistry. 14 (2006) 6686–6694.

[19] A.K. Chakraborti, B. Gopalakrishnan, M.E. Sobhia, A. Malde, 3D-QSAR studies of indole derivatives as phosphodiesterase IV inhibitors, European Journal of Medicinal Chemistry. 38 (2003) 975–982.

[20] K.F. Khaled, Modeling corrosion inhibition of iron in acid medium by genetic function approximation method: A QSAR model, Corrosion Science. 53 (2011) 3457–3465.

[21] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. Mol. Inform. 22, 69-77, 2003, DOI: 10.1002/qsar.200390007.

[22] N. Minovski, Š. Župerl, V. Drgan, M. Novič, Assessment of applicability domain for multivariate counter-propagation artificial neural network predictive models by minimum Euclidean distance space analysis: a case study, Anal. Chim. Acta 759 (2013) 28–42.

[23] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C.P. Varghese, R.K. Agrawal, Validation of QSAR models-strategies and importance, International Journal of Drug Design & Discovery. 3 (2011) 511–519.

[24] A. Tropsha, P. Gramatica, V. K. Gombar, The importance of being earnest: validation is the absolute essential for successful application

and interpretation of QSPR models, Molecular Informatics. 22 (2003) 69–77.

[25] http://www.rcsb.org/pdb/home/home.do

[26] S.E. Adeniji, S. Uba, A. Uzairu, QSAR Modeling and Molecular Docking Analysis of Some Active Compounds against Mycobacterium tuberculosis Receptor (Mtb CYP121), Journal of Pathogens. (2018) 1– 24.