ORIGINAL ARTICLE

A C G publications

Org. Commun. XX:X (202X) X-XX

organic communications

Prediction of the antiproliferative effects of some benzimidazole-chalcone derivatives against MCF-7 breast cancer cell lines: QSAR and molecular docking studies

Oluwatoba E. Oyeneyin [1,*], Chiamaka G. Iwegbulam [1]

Nureni Ipinloju [1], Bambo F. Olajide [2] and Abel K. Oyebamiji [3]

[1]*Theoretical and Computational Chemistry Unit, Department of Chemical Sciences, Adekunle Ajasin University, Akungba-Akoko, Ondo State, Nigeria*

[2]*Department of Chemistry, Osun State College of Education, Ilesha, Osun State, Nigeria*

[3]*Department of Basic Sciences, Adeleke University, P.M.B. 250, Ede, Osun State, Nigeria*

(Received March 05, 2022; Revised July 02, 2022; Accepted July 09, 2022)

Abstract: Cancer remains a threat to human existence owing the high number of deaths associated with it. Combating cancer has continued to garner interest from pharmaceutical companies as the resistance and toxicity issues have been observed with the treatment of known drugs. A series of benzimidazole-chalcone derivatives were obtained from literature and investigated for their ability to inhibit MCF-7 Breast Cancer Cell Lines via QSAR and molecular docking techniques. A QSAR model was generated to establish the relationship between molecular descriptors and bioactivity of benzimidazole-chalcone derivatives. The R^2, adjusted R^2, Q^2 and R^2pred values obtained suffice to deem the model fit and reliable. The binding affinities of the lead compounds were obtained via an extra precision docking procedure at the active site of the human serine/threonine-protein kinase receptor, 3FC2. Molecule **21** showed the closest binding affinity (-9.350 kcal/mol) as compared to doxorubicin (-9.305 kcal/mol), and had the best IC_{50} as compared to other compounds as reported earlier in literature. *In-silico* ADME and drug-likeness prediction of all the molecules showed good pharmacokinetic properties, bioavailability and non-toxicity. The model generated here could be used in predicting the bioactivity of novel antiproliferative agents.

Keywords: Breast cancer; MCF-7 cell line; benzimidazoles; quantitative structure activity relationship; molecular docking; pharmacokinetics.

1. Introduction

Recent development in the field of cancer research has led to the renewed interest in the search for new anticancer drugs. Cancer is an uncontrolled growth of the body cells that spreads to other parts of the body. According to WHO, cancer is said to be the leading cause of death worldwide. In 2020, nearly 10 million deaths were recorded[1]. The most common types of cancer include breast, prostate, lung, stomach, colon, rectum and skin cancer. However, breast cancer is the most virulent type of cancer which has been implicated in most of cancer deaths[2]. In 2020, 685000 deaths were recorded as the number of women that died from breast cancer[1]. To date, many researchers have worked on finding a cure for cancer but none have been able to develop a drug molecule that can cure cancer completely, including the non-invasive MCF-7 breast cancer[3,4]. MCF-7 has been proved to be a good cell line for

* Corresponding author:E-Mail: emmanueltoba90@gmail.com and oluwatoba.oyeneyin@aaua.edu.ng

The article was published by ACG Publications
http://www.acgpubs.org/journal/organic-communications © Month-Month 202X EISSN:1307-6175
DOI: http://doi.org/10.25135/acg.oc.132.2203.2374
Available online: July 18, 2022

investigation of breast cancer globally. The recent contribution of the MCF-7 cell line to breast cancer research is its usefulness for the study of the estrogen receptor both *in vivo* and *in vitro* and basic research relies on these cell lines[5,6]. Chemotherapy remains one of the best and fast clinical approaches for the treatment of cancer but due to the toxic effect such as nausea, fatigue, loss of appetite, weight loss, there is a need to develop and design molecules with high potency and low toxicity that can selectively inhibit the growth of abnormal cells[7].

Heterocyclic compounds are cyclic compounds that contain one or more heteroatoms like N, O and S in their ring structure and have been reported for their pharmacological activities[8-10]. Benzimidazole (fused benzene and imidazole ring), is one of the heterocyclic ring systems that possesses several biological applications in drug design and development[4,11,12]. Some benzimidazole derivatives have been reported for their antiviral[13], anticancer[12], antimicrobial[14] and antifungal activities[15]. Recent study showed that N-substituted benzimidazole derivatives enhanced the cytotoxic effects on human breast adenocarcinoma (MCF-7) and human ovarian carcinoma (OVCAR-3) cell lines[16]. Computer-aided drug design (CADD) remains an important technique in drug research as it saves time, cost involved in finding new drug target[17,18]. CADD methods such as molecular docking and quantitative structure-activity relationship (QSAR) methods has been used to calculate the binding affinities of molecules with targeted receptors[19,20] and to predict the bioactivity of molecules via mathematical models[21,22], respectively. In a bid to understand how substituents affect the bioactivity of the molecules and explore the mechanisms of action of some benzimidazole derivatives against MCF-7 breast cancer cell line, the work done by Hsieh and co[16] was extended using a QSAR model to predict the bioactivity of the molecules while molecular docking technique was used to determine the binding affinity and interactions of the lead molecules with a polo-like kinase inhibitor (PLK1).

2. Methodology

2.1 .Computation of Structural Descriptors and QSAR Modelling

A series of twenty-four (24) new derivatives of Benzimidazole-Chalcone as potential anticancer agents with their inhibitory concentration were collected from the literature[16]. The structure of the molecules and their bioactivities are shown in Table 1. The inhibitory concentration (IC_{50}) were normalized by converting to pIC_{50} using (Equation 1).

$$pIC_{50} = -\log10(IC_{50} \times 10^{-6}) \quad \text{Eq. 1}$$

Two-dimensional chemical structures of all the compounds were drawn and using ChemDraw Ultra 10.0[23], saved as cdx file, imported into Spartan 14 software for optimization[24] having initially search for the global minima of the molecules. Optimized structures were saved as sdf format and then imported into PaDEL Descriptor software version 2.20[25] for the calculation of 2D descriptors of all the compounds. Before the generation of the model, the data were pre-treated to remove constant and highly inter-correlated descriptors using the Data Pre-treatment software[26]. The pre-treated data was divided into training and test set by using Kennard and Stone's algorithm[27]. Sixteen (16) compounds constituting 70 % of the data set was used as the training set and validated internally while the remaining 30% (8 compounds) was used as test set to validate the model externally. This was followed with model development via genetic function algorithm (GFA)[28] in material studio while fixing the experimental activities (pIC_{50}) as the dependent variable and the molecular descriptors as the independent variable. The population and Generation were set to 1000 and 2000, the mutation probability was set to 0.1, equation length was set to 4, and the smoothing parameter was set to 0.5.

2.1.1. Quality Assurance of the Built Model

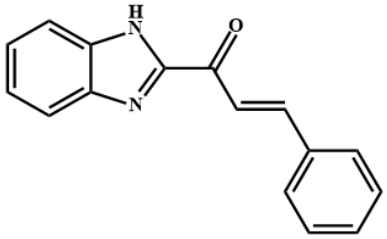
2.1.1.1 Internal Validation

The fitness function of the training model is validated internally using some parameters generated from Material studio. Some of the parameters calculated are Friedman's lack of fit (LOF)[29],

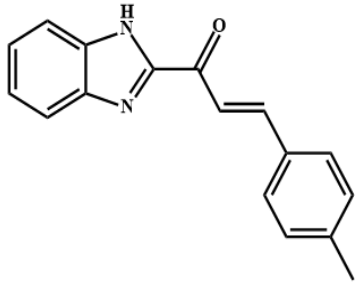
the square of the correlation coefficient (R^2), Adjusted R^2 (R^2_{adj}), Cross validation coefficient (Q^2). The LOF is parameter used to obtain the fitness score of the generated model (Equation 2).

$$LOF = \frac{SEE}{\left(1 - \frac{a+(b\ X\ p)^2}{M}\right)} \qquad \text{Eq. 2}$$

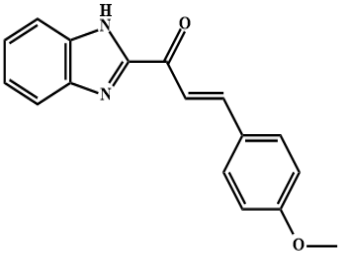
Table 1. Chemical structure of the molecules and experimental activity



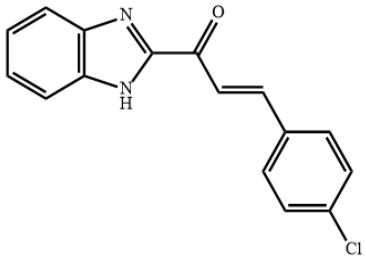
Molecule **1**: pIC_{50} = 4.8699



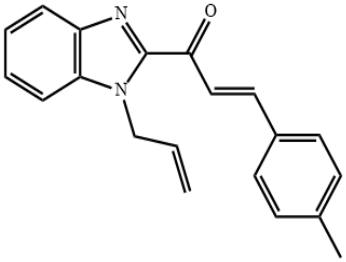
Molecule **2**: pIC_{50} = 4.7425



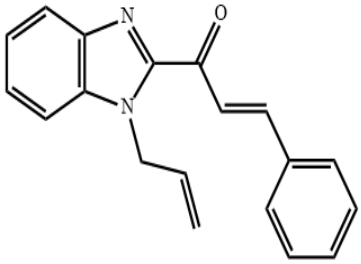
Molecule **3**: pIC_{50} =4.7947



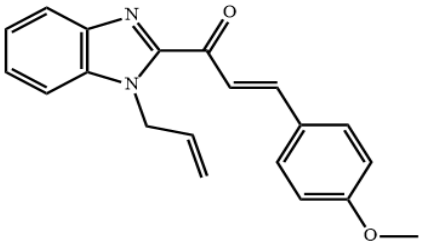
Molecule **4**: pIC_{50} = 4.4874



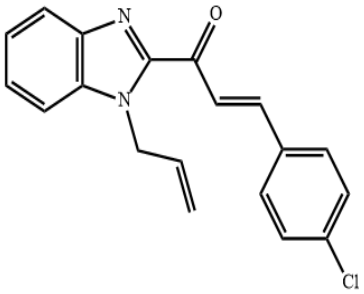
Molecule **5**: pIC_{50} =4.9164



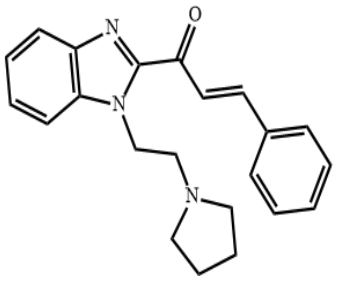
Molecule **6**: pIC_{50} =4.2713



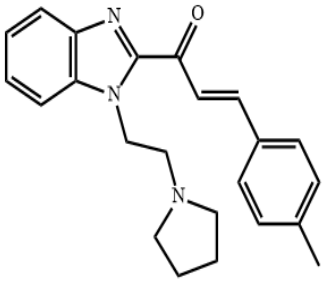
Molecule **7**: pIC_{50} =3.5936



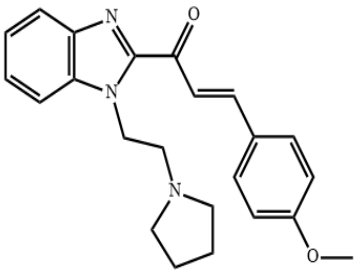
Molecule **8**: pIC_{50} = 4.8722



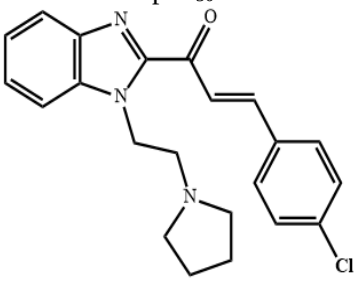
Molecule **9**: pIC_{50} =5.0154



Molecule **10**: pIC_{50} = 4.2741

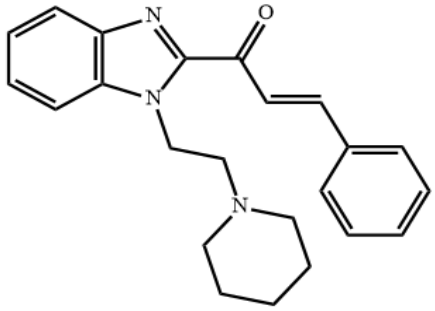


Molecule **11**: pIC_{50} =4.2511

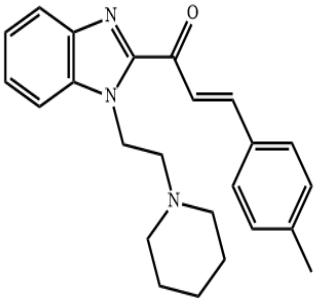


Molecule **12**: pIC_{50} =4.8407

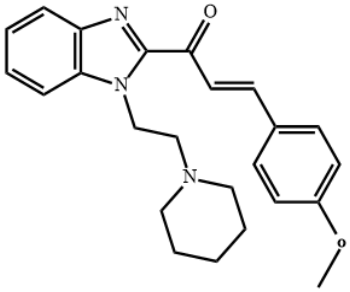
Table 1 continued..



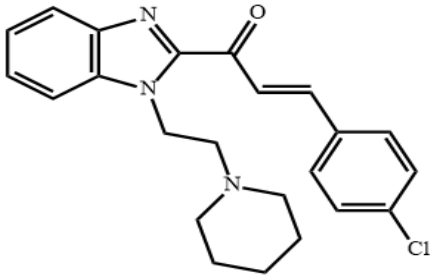
Molecule **13**: pIC_{50} =5.0118



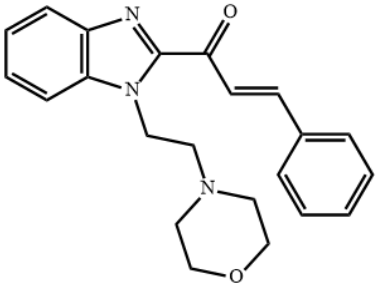
Molecule **14**: pIC_{50} =4.9531



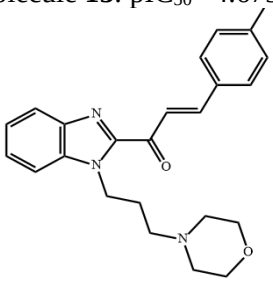
Molecule **15**: pIC_{50} =4.6753



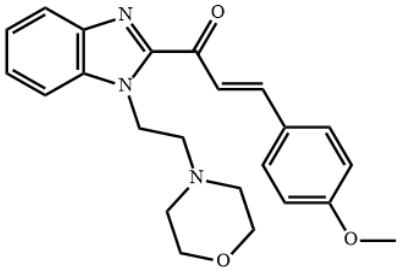
Molecule **16**: pIC_{50} =4.26432



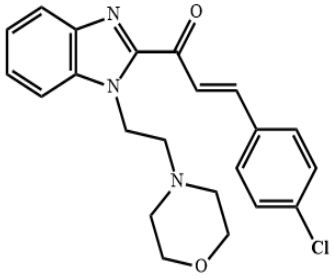
Molecule **17**: pIC_{50} =4.9838



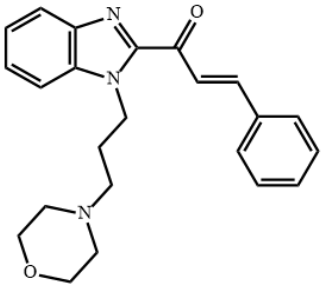
Molecule **18**: pIC_{50} =4.9935



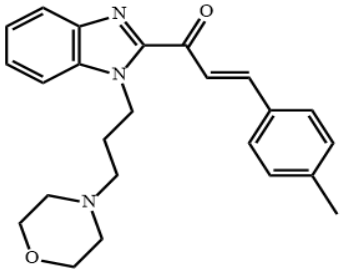
Molecule **19**: pIC_{50} =4.9538



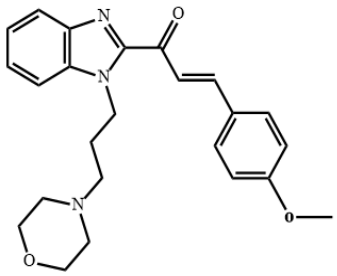
Molecule **20**: pIC_{50} =4.6395



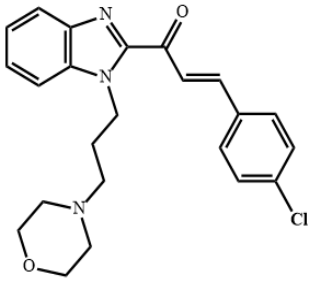
Molecule **21**: pIC_{50} =5.0501



Molecule **22**: pIC_{50} =4.9453



Molecule **23**: pIC_{50} =4.9233



Molecule **24**: pIC_{50} =4.4474

Where SEE is the standard error of estimate and its value must be low for a model to be good; *a* is the number of model terms, *d* is the smoothing parameter and *p* is the number of descriptors. The square of the correlation coefficient (R^2) is most widely used for internal validation of QSAR models. The closer the R^2 to 1, the better the model (Equation 3).

$$R^2 = 1 - \frac{\sum(Y_{exp} - Y_{pred})^2}{\sum(Y_{exp} - Y_{training})^2} \quad \text{Eq. 3}$$

where Yexp is the mean experimental activities, Ypred is the mean predicted activities of the data set. Adjusted (R^2) (Equation 4) measures the reliability and stability of a model. The value generally increases with the number of descriptors.

$$R^2_{adj} = \frac{(R^2-p)\ X\ (n-1)}{n-P-1} \quad \text{Eq. 4}$$

Where where 'n' is the number compounds that made up the training set. Cross validation coefficient (Q^2_{cv}). The strength of the QSAR model was cross-validated (Equation 5).

$$Q^2_{cv} = 1 - \frac{\sum(Y_{pred}-Y_{exp})^2}{\sum(Y_{exp}-Y_{mintraining})^2} \quad \text{Eq. 5}$$

Where Ytraining, Yexp, and Ypred were defined earlier

2.1.1.2. External Validation of the Model

2.1.1.2.1. Predictive R^2 (R^2_{pred})

This is the most widely used parameter to validate a QSAR built model. The closer the R^2_{pred} value to 1, the better the model stability and the reliability of the model. The R^2_{pred} is defined by (Equation 6);

$$\mathbf{R^2_{pred}} = \frac{\sum(Y_{pred_{test}} - Y_{exp_{test}})^2}{\sum(Y_{pred_{test}}-Y_{mintraining})^2} \quad \text{Eq. 6}$$

2.1.1.2.2. Root mean square error (RMSE)

RMSE is another parameter used for evaluating the performance of a model. The closer the value of RMSE to 0, the better the model built. The RMSE is evaluated using (Equation 7);

$$RMSE_{EXT} = \sqrt{\frac{\sum(Y_i-\tilde{Y}_i)^2}{n}} \quad \text{Eq. 7}$$

Where Y_i is the external response observed for the *i*th object and $\tilde{Y}_i$ is the external response predicted by the model.

2.1.1.3 Applicability Domain (AD)

This is a crucial step in determining the predictive potential and dependability of a QSAR model[30], it is described by the descriptors of the training compound. In this work, standardized approached was used. In this approach, a QSAR models learn from the training set descriptors and then use for the prediction of the external set compounds similar in structure to the training set compounds. A small change in the training set descriptors of one or more compounds compare to others indicate that the compound is not properly fixed in the training process. These compounds are outlier and they should be removed in order to build a good and reliable model. If the descriptors of the test compounds are similar to the outlier compounds, then their predictions are not good and they are outside the applicability domain[30,31]. A training set compound is said to be non-outlier if the standardized residue of the descriptors based on mean and standard deviation is less than 3 and if in test set, it is considered to be outside applicability domain.

2.1.2. Quality Assurance Model Generated

The required values for assessing the model is shown in Table 2. The parameters were used in checking the effectiveness of the built model.

Table 2. Generally accepted values for the validation parameters for a built model

Parameter	Name	Accepted value
R^2	Coefficient of determination	Greater than or equal to 0.6
$P_{(95\%)}$	Confidence interval at 95% confidence level	Less than 0.5
Q^2_{CV}	Cross validation coefficient	Greater than or equal to 0.5
R^2- Q^2_{CV}	Difference between R^2 and Q^2_{CV}	Less than 0.3
N(ext& test set)	Number of test set	Greater than or equal to 5

2.2. Molecular Docking

The PLK1 receptor (PDB: 3FC2) in Figure 1 was downloaded from the Protein Data Bank (www.rcsb.org) and was prepared using the protein preparation tool in Maestro/Schrodinger suite[32]. The structure of the lead compounds, that is compounds with the highest pIC_{50} values (**9, 13, 17, 18** and **21**) and some standard drugs (doxorubicin, fluoracil and chlorambucil) were imported for processing using LigPrep[33]. Subsequently, the compounds were sucessfully docked into each active sites of the receptor to generated docking score using extra precision (XP) docking.

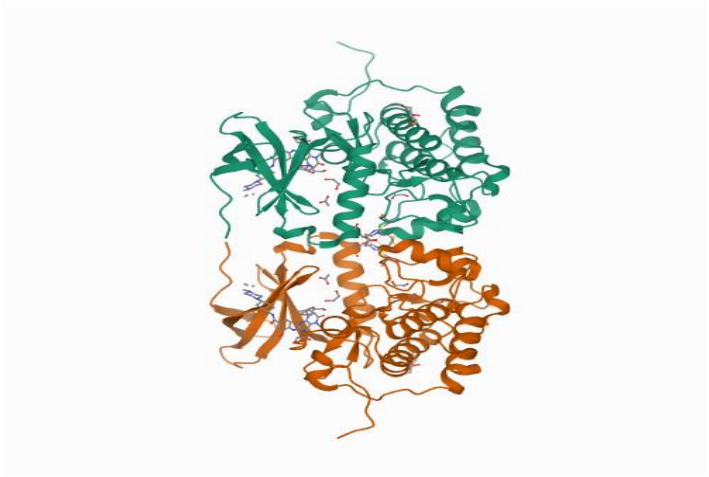


Figure 1. 3FC2

2.3 .Pharmacokinetics

Pharmacokinetics properties of the molecules were determined using the admetSAR prediction tool (http://lmmd.ecust.edu.cn:8000/). The chemical structure of each compound was input in SMILES (simplified molecular-input line-entry system) format to generate different pharmacokinetics parameters.

3. Results and Discussion

3.1. QSAR

QSAR was used to verify the relationship of benzimidazole-chalcone derivatives with its anti-cancer activities. Employing GFA method, equation was built to predict the anti-cancer activities. The mathematical generated for the prediction of bioactivity is:

$$Y = -1.356759595 \times AATS6s) - 0.008948385 \times VR2\_Dzs - 0.174085842 \times nBondsS2 + 9.081215767 \times SpMin7\_Bhe + 6.224213274$$

Table 3 shows the information of the descriptors in the model generated. Table **4** shows the comparison of experimental activity (pIC_{50}), predicted activity (pIC_{50}) and residual of developed model.

Table 5 shows the validation parameters for the built model. Tables **6** and **7** show how the external validation of model calculated using the validation set (test set) and model descriptors. Table **8** shows applicability domain information calculated using standardized approach. Figure 2 shows the plot of experimental activity against predicted activity of the studied compound.

Table 3. List of descriptors, their constructors, description and dimension used in building the QSAR model

S/N	Name	Description	Dimension
1	AATS6s	Average centered bromo -Moreau autocorrelation lag/weighted by I-state	2D
2	VR2_Dzs	Normalized Randic-like eigenvector-based index from Barysz matrix/ weighted by I-state	2D
3	nBondsS2	Total number of single bonds (including bonds to hydrogens, excluding aromatic bonds)	2D
4	SpMin7_Bhe	Smallest absolute eigenvalue of Burden modified matrix - n 7 / weighted by relative Sanderson electronegativities	2D

Table 4. Comparison of Experimental activity (pIC_{50}), predicted activity (pIC_{50}) and residual of developed model

S/N	Experimental activity	Predicted activitity	Residual
1	4.86998800	4.89980200	-0.02981400
2	4.74256100	4.66762700	0.07493500
3	4.79479600	4.91809300	-0.12329700
4	4.48744900	4.51055700	-0.02310800
5	4.91649700	4.86518600	0.05131100
6	4.27132200	4.33452700	-0.06320600
7	3.59363000	3.62750100	-0.03387100
8	4.87224700	4.85350900	0.01873900
9	5.01547300	4.91739400	0.09807900
10*	4.27417	4.379719	-0.10555
11*	4.251115	4.156444	0.094671
12	4.84073400	4.80020100	0.04053300
13*	5.011887	4.970209	0.041679
14*	4.953115	4.914264	0.038851
15*	4.675306	4.379719	0.295587
16	4.26432100	4.21484600	0.04947600
17*	4.983803	4.894414	0.089388
18	4.99353400	5.01728300	-0.02374900
19	4.95389500	4.91301000	0.04088500
20	4.63959600	4.75258900	-0.11299300
21	5.05012200	5.15563100	-0.10550900
22*	4.945387	5.247516	-0.30213
23	4.92336000	4.78176900	0.14159100
24*	4.447453	4.466454	-0.01900

*= test sets

Table 5. Validation parameter for the built model

Parameters	**Equation 1**
Friedman LOF	0.03881100
R-squared	0.95906700
Adjusted R-squared	0.94418300
Cross validated R-squared	0.91380000
Significant Regression	Yes
Significance-of-regression F-value	64.43323300
Critical SOR F-value (95%)	3.40350500
Replicate points	0
Computed experimental error	0.00000000
Lack-of-fit points	11
Min expt. error for non-significant LOF (95%)	0.06760600
RMSE	0.024541
R^2_{pred}	0.76452

Table 6. External validation of developed model

EXP IC_{50}	**AATS6s**	**VR2_Dzs**	**nBondsS2**	**SpMin7_Bhe**	**predicted**	**Error**
4.27417	2.834011	12.85266	41	1.018928	4.379719	-0.10555
4.251115	3.015805	61.82256	39	1.018928	4.156444	0.094671
5.011887	3.097834	23.58998	35	1.018928	4.970209	0.041679
4.953115	2.837269	10.98565	38	1.018928	4.914264	0.038851
4.675306	2.834011	12.85266	41	1.018928	4.379719	0.295587
4.983803	2.838503	13.01666	38	1.018928	4.894414	0.089388
4.945387	3.472222	8.183428	25	0.898519	5.247516	-0.30213
4.447453	3.050701	23.24301	34	0.981316	4.466454	-0.01900

Table 7. External validation of developed model (continue)

$(Ypred-Yobs)^2$	**$Y_{expmean}$**	**$Y_{expmean}$-Y_{exp}**	**$[Y_{expmean}-Y_{exp}]^2$**
0.011141	4.701845	-0.42768	0.182906
0.043295	4.701845	-0.45073	0.203158
0.001737	4.701845	0.310042	0.096126
0.001509	4.701845	0.251269	0.063136
0.087372	4.701845	-0.02654	0.000704
0.00799	4.701845	0.281957	0.0795
0.091282	4.701845	0.243542	0.059313
0.000361	4.701845	-0.25439	0.064715

$\Sigma(Ypred-Yobs)^2 = 0.210355$, $\Sigma(Y_{expmean}-Y_{exp})^2 = 0.749559$, R^2 test = 1- 0.210355/0.749559 = 0.761521

The descriptors in the model equation were defined and classified (Table 3). The negative coefficient of **AATS6s**, **VR2_Dz**, **nBondsS** suggest that a decrease in those descriptors factor would lead to high bioactivities of derivatives while positive coefficient **SpMin7_Bhe** indicates that increase in the descriptor factor will lead to increase in bioactivity of the derivative**.** Both validations (internal and external) are in within the minimum accepTable values used in assessing the equation (Table 5). The low values of residual obtained indicate that there is high correlation between the experimental activity and predicted activity (Table 4). Also, the value of R^2pred (0.7615) calculated indicates that the model is good to predict external data (Table **5**). The graph of experimental activity and predicted activity of all studied compounds are shown in Figure 1. The high value of correlation coefficient square (R^2) of 0.9590, correlation coefficient adjusted square (R^2_{adj}) of 0.9441, cross-validation coefficient (Q^2) of 0.9138 and external validation (R^2_{pred}) of 0.7645 showed that the model generated is great, effective, reliable and can be used to predict the activity of external data set. Table 8 shows Applicability domain

information calculated using standardized approach, as can be seen form the Table, there is no outlier in all the training set compounds and all the test compounds are all within the applicability domain. This indicate that all the compounds descriptors are good in building the model. All these suffice to establish the predictability and reliability of the developed model. This model could be used in predicting the antiproliferative activity of novel molecules.

Table 8. Applicability domain information calculated using standardized approach

Name	Mean standardized value	Outlier info/AD info
1	1.282341	NO
2	0.814213	NO
3	0.765832	NO
4	0.784278	NO
5	0.624497	NO
6	0.68292	NO
7	1.442817	NO
8	0.612305	NO
9	0.512523	NO
10*	0.875378	WAD
11*	0.987632	WAD
12	1.102401	NO
13*	0.480605	WAD
14*	0.818881	WAD
15*	0.875378	WAD
16	0.858572	NO
17*	0.797068	WAD
18	0.928522	NO
19	1.293754	NO
20	0.769446	NO
21	0.502854	NO
22*	0.413329	WAD
23	0.807644	NO
24*	0.401394	WAD

*= test sets, NO: Not outlier, WAD: Within applicability domain

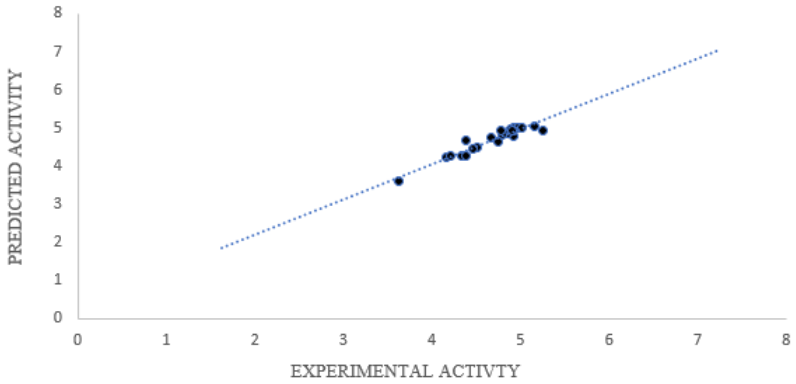


Figure 2. Observed activity vs. predict activity of the studied compounds

3.2. Molecular Docking Results

Molecular docking of the lead molecules with 3FC2 are presented (Table 9 and Figures 3 – 11).

Table 9. The XP docking scores of the lead compounds and some standard drugs

3FC2	**9**	**13**	**17**	**18**	**21**	**doxorubicin**	**chlorambucil**	**5-Fu**
Docking score (kcal/mol**)**	-5.963	-6.681	-8.447	-8.565	-9.350	-9.305	-6.046	-4.740

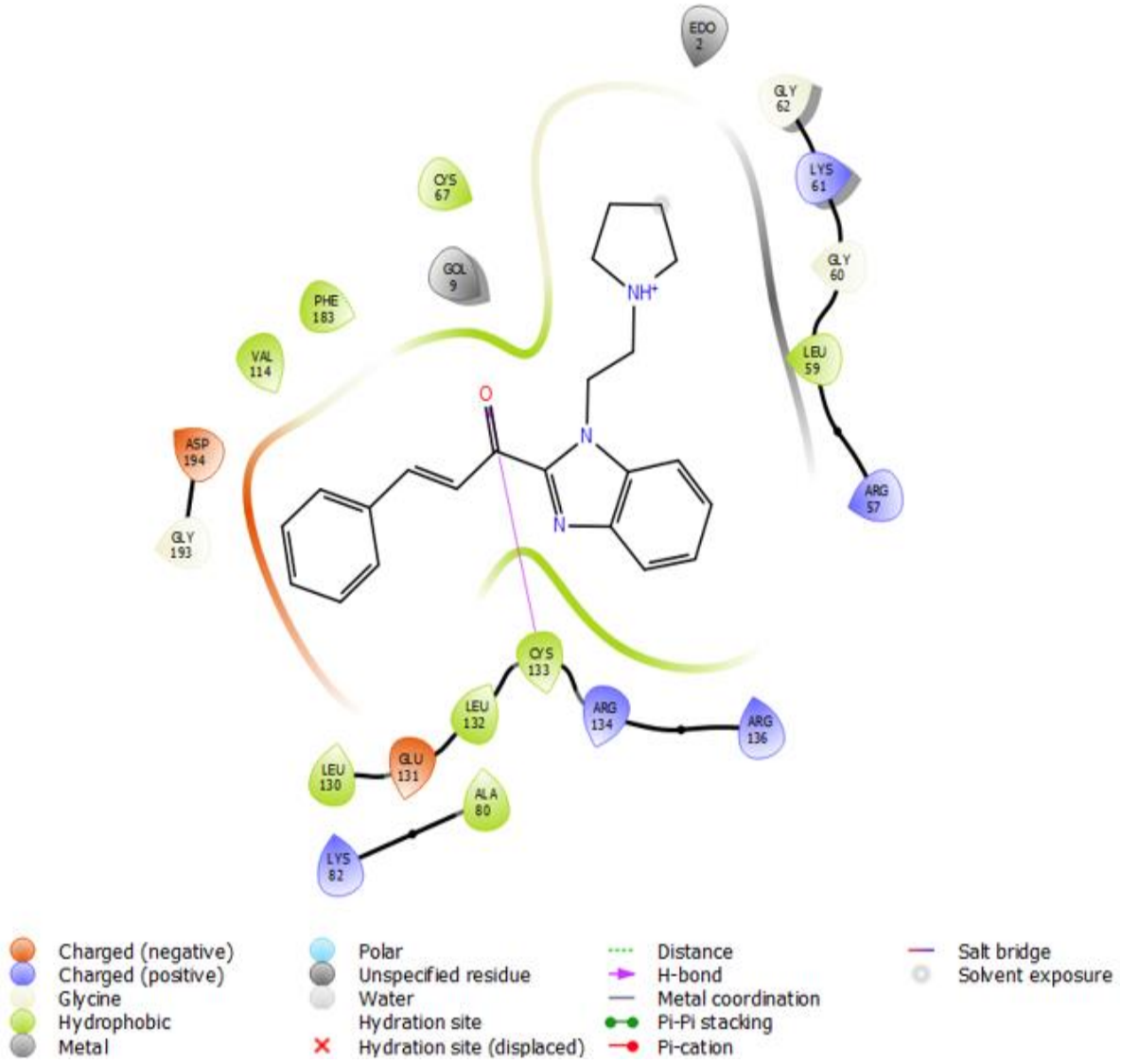


Figure 3. 9 with 3FC2

From Table 9, the results of the study showed that the ligands (**9, 13, 17, 18** and **21**) used in the study compared favourably well with the standard ligands as seen in their binding affinities. All of them have better binding affinities than 5-Fu. All of them except molecule 9 have higher binding affinities than chlorambucil. Molecule **21** is the only molecule having the closest binding affinity (-9.350

kcal/mol) to doxorubicin (-9.305 kcal/mol). This is in line with the experimental result as it showed the best IC_{50} value of all the compounds[16]. Generally, the high binding affinity of the ligands with the receptor indicates that the ligands docked well in the receptor's active site and that a small amount is needed to trigger a physiological response. The high binding affinity of compound **21** is probably due to the interaction with two essential residues within the binding site forming a pi-pi stacking with CYS 133 and hydrogen bond with PHE 183.

From the docking interactions, Compound 9 formed hydrogen bond via its carbonyl group with CYS 133 of the amino acid in the active site of the target protein as shown in Figure 3. Compound **13** formed two interactions, a hydrogen bond from its carbonyl oxygen with CYS 133 and a pi-pi stacking interaction with PHE183 from its imidazole ring as shown in Figure S1. Compound 17 in Figure S2 interacted via its imidazole nitrogen with an unspecified amino acid residue in the active site of the target protein (GOL 9) through hydrogen bonding. Compound 18 interacted via its carbonyl oxygen CYS 133 through hydrogen bonding (Figure S3). Compound 21 formed 2 hydrogen bond interactions with CYS 133 and an unspecified residue (EDO 2) via its carbonyl oxygen and morpholinopropyl oxygen, respectively. It also formed a pi-pi interaction with PHE 183 as shown in Figure S4. Doxorubicin formed 3 hydrogen bonds with three amino acid residues in the active site of the target protein (LEU 59, EDO 2 and LYS 82). İt also formed pi-pi interaction with PHE 183 (Figure S5). Chlorambucil interacted with amino acid ARG 57 through hydrogen bonding and also formed a salt bridge with the amino acid residue as shown in Figure 9. 5-florouracil formed two hydrogen bond interactions with CYS 133 (Figure S6).

3.1.3 Pharmacokinetics Results

The pharmacokinetics results of the lead compounds are presented (Tables 10 and 11).

Table 10. Lipinski properties obtained for the compounds

Compounds	Formulae	MW	HBA	HBD	LogP	PSA	Lipinski violations
9	$C_{22}H_{23}N_3O$	345.45	3	0	5.13	23.45	1
13	$C_{23}H_{25}N_3O$	359.47	3	0	5.55	22.99	1
17	$C_{22}H_{23}N_3O_2$	361.44	4	0	4.42	30.97	0
18	$C_{24}H_{27}N_3O_2$	389.50	4	0	5.19	30.07	1
21	$C_{23}H_{25}N_3O_2$	375.47	4	0	4.70	29.57	0

Table 11. Some selected pharmacokinetics properties of the compounds obtained from Admesar

Compound	BBB permaent	Human intestinal absorptiom	Caco-2 permeability	P-gp substrate	Rat acute toxicity LD50	Acute oral toxicity	Carcinogen
9	+(0.9762)	+(1)	+(05324)	S(0.6880)	2.2853	III(07077)	NC(0.9841)
13	+(0.9711)	+(0.9961)	+(0.5436)	S(0.7077)	2.3588	III(0.7087)	NC(0.9601)
17	+(0.9766)	+(1)	+(0.5163)	S(0.6456)	2.1800	III(0.7041)	NC(0.9529)
18	+(0.9415)	+(1)	+(0.5527)	S(0.6975)	2.2535	III(0.7269)	NC(0.9421)
21	+(0.9568)	+(1)	+(0.5186)	S(0.6199)	2.2129	III(0.7275)	NC(0.9482)
doxoruicin	-(0.9951)	-(0.8092)	-(0.7990)	S(0.7861)	2.6644	III(0.7766)	NC(0.9534)

3. 3. Pharmacokinetics

3.3.1. Drug-likeness Prediction

A molecule can be considered as a drug if it obeys Lipinski's rule of five (RO5)[34,35]. For a molecule to be able to absorb or permeate, it should possess the following properties: molecular mass < 500, logP ≤ 5, HBD ≤ 5 and HBA ≤ 10[34,35]. If two or more of these properties violate the rules, it indicates that the molecule is not orally active. Lipinski's rule obtained for the studied molecules are

depicted in Table 10. It can be seen that all the molecule obey the RO5. This indicates that all the molecules are orally bioactive. Furthermore, polar surface area (PSA) is a crucial property for blood brain barrier and cell membrane permeability. Generally. PSA less than $140A^2$ is required for a molecule to permeate cell membrane while PSA less than $90A^2$ is required for a molecule to permeate blood brain barrier[35]. The PSA of all the molecules are within the accepted value (Table 10), they range from $22.99A^2$ to $30.97A^2$. This shows that all the molecules can permeate cell membrane (they are lipophilic) and blood brain barrier (needed to overcome the incidence of brain metastasis in cancer patients)[37,38]. Also, the molecular weights of all molecules are all less than 500, the HBA and HBD are within accepTable values ($\leq$ 10) and ($\leq$ 5) respectively. Partition coefficient, logP describes the overall lipophilicity of the compounds and helps in hepatetic clearance[35], the values observed for logP of the compounds are effective in terms of lipophilicity. Molecules 9, 13 and 18 having logP higher than 5 could have poor oral absorption and tend to bind to hydrophobic targets[35,39]. All the molecules are in accordance with the Lipinski's rule since none violated more than one of the properties.

3.3.2. ADMET Prediction Results

Many of drugs during the development are associated to poor pharmacokinetics. Therefore, monitoring the pharmacokinetics properties at early stage of drug discovery and development minimize the pharmacokinetics related issues**.** Absorption, distribution, metabolism, excretion and toxicity (Admetox) properties of the studied compounds were done using admetSAR. Some selected Admetox investigated are Blood Brain Barrier (BBB) penetration, Human Intestinal Absorption (HIA), Caco-2 cell permeability, p-glycoprotein, rat acute toxicity, acute oral toxicity and Carcinogenicity were calculated. The results are summarized in Table 11. BBB permeation is a vital parameter in drug development and discovery. BBB permeation predicts if a molecule will cross over the BBB and exerts its curative action on the brain. The results of BBB permeate obtained for all the compounds and the standard drug are shown in Table 11. It can be seen from the results that all the compound could pass through the BBB and exerts its therapeutic effects. Doxorubicin has shown negative value, this indicates poor penetration through BBB. The results also showed that all test molecules could be absorbed by the human intestine. Also, all the studied molecule could permeate the Caco-2 cell. Doxorubicin also showed poor penetration through Caco-2 cell. P-glycoprotein (P-gp) is a trans-membrane efflux pump which eliminates bioactive molecules from the cell membrane and cytoplasm and causes molecules to undergo further metabolism and clearance[36]. This study showed that all the molecules are substrates for p-glycoprotein, and non-carcinogenic.

4. Conclusion

In this study, a mathematical model was generated via QSAR to predict the bioactivity of a series of benzimidazole-chalcone derivatives against MCF-7 breast cancer. The QSAR results indicated that all the descriptors in the built model are good as they are within the applicability domain. The R^2, adjusted R^2, Q^2 and R^2pred values were 0.96, 0.94, 0.91 and 0.76, respectively. This is indicative that the built model is good, reliable and can be used to predict the bioactivity of novel benzimidazole-chalcone derivatives. The lead molecules were docked in the active site of a human serine/threonine kinase receptor, 3FC2. The ligands are of higher binding affinities than that of standard ligand (5-FU). All compounds except molecule 9 had higher binding affinities than chlorambucil. The docking score ranged from -5.963 kcal/mol to -9.350 kcal/mol signifying that the molecules can bind very well with the active site of the target. Compound 21 had the closest binding affinity when compared with standard drug, doxorubicin. In addition, the results of the ADME and drug-likeness properties showed that the compounds possess good pharmacokinetic properties which are predicted to be orally bioavailable, less toxic and good absorption. Furthermore, the computational findings were in line with the experimental results reported earlier.

List of abrreviations

- WHO- World Health Organization
- CADD- Computer Aided Drug Design
- QSAR- Quantitative Structural Activity Relationship
- GFA- Genetic Function Algorithm
- LOF- Lack of Fit
- SEE- Standard Error of Estimate
- RMSE- Root Mean Square Error
- AD- Applicability Domain
- PDB- Protein Data Bank
- Smile- Simplified Molecular-Input line-Entry System
- HIA- Human Intestinal Absorption
- BBB- Blood Brain Barrier

Supporting Information

Supporting information accompanies this paper on http://www.acgpubs.org/journal/organic-communications

ORCID
Oluwatoba E. Oyeneyin: 0000-0001-5709-0244
Chiamaka G. Iwegbulam: 0000-0002-0858-7964
Nureni Ipinloju:0000-0002-2683-7146
Bambo F. Olajide:0000-0003-4843-3991
Abel K. Oyebamiji:0000-0002-8932-6327

References

[1] Cancer. (n.d.). https://www.who.int/news-room/fact- sheets/detail/cancer, **2021**.
[2] Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R. L.; Torre, L. A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Canc. J. Clinic*. **2018**, *68*, 394–424.
[3] Milestones in Cancer Research and Discovery. https://www.cancer.gov/research/progress/250-years-milestones. **2015**.
[4] Brishty, S. R.; Hossain, J.; Khandaker, M. U. A Comprehensive Account on Recent Progress in Pharmacological Activities of Benzimidazole Derivatives. *Front. Pharmacol.* **2021**, *12* (762807), 1-49.
[5] Syam, S.; Abdelwahab, S. I.; Al-Mamary, M. A.; Mohan, S. Synthesis of chalcones with anticancer activities. *Molec*. **2012**, *17*, 6179–6195.
[6] Lee, A. V.; Oesterreich, S.; Davidson, N. E. MCF-7 Cells—Changing the Course of Breast Cancer Research and Care for 45 Years. *J. Natl. Cancer. Inst.* **2015**, *107*,1–4.
[7] Iqbal, J.; Ejaz, S. A.; Khan, I.; Ausekle, E.; Miliutina, M.; Langer, P. Exploration of quinolone and quinoline derivatives as potential anticancer agents. DARU *J. Pharmaceut. Sci*. **2019**, *27(2)*, 613-626.
[8] Arulraj, R.; Sivakumar, S.; Suresh, S,; Anitha, K. Synthesis, Vibrational spectra, DFT calculations, Hirshfeld surface analysis and Molecular docking study of 3-chloro-3-methyl-2,6-diphenylpiperidin-4-one. *Spectrochimica Acta Part A: Mol. Biomol. Spectrosc.* **2020**, *232*,118166.
[9] Arulraj, R. Synthesis, Crystal Structure, DFT Calculations and Hirshfeld Surface Analysis of 3-Chloro-3-methyl-r(2),c(6)-bis(p-methoxyphenyl) piperidin-4-one. *J. Chem. Crystallogr*. **2020**, *50*, 41-51.

[10] Oyeneyin, O. E.; Obadawo, B. S.; Metibemu, D. S.; Owolabi, T. O.; Olanrewaju, A. A.; Orimoloye, S. M.; Ipinloju, N.; Olubosede, O. An exploration of the antiproliferative potential of chalcones and dihydropyrazole derivatives in prostate cancer via androgen receptor: combined QSAR, machine learning, and molecular docking techniques. *Phys. Chem. Res.* **2022**, *10(2)*, 211–223.

[11] Serafini, M.; Torre, E.; Aprile, S.; Del Grosso, E.; Gesù, A.; Griglio, A.; Colombo, G.; Travelli, C.; Paiella, S.; Adamo, A. et al. Discovery of highly potent benzimidazole derivatives as indoleamine 2,3-dioxygenase-1 (IDO1) inhibitors: from structure-based virtual screening to in vivo pharmacodynamic activity. *J. Med. Chem*. **2020**, *63(6)*, 3047-3065.

[12] Yadav, S.; Narasimhan, B. Perspectives of Benzimidazole Derivatives as Anticancer Agents in the New Era. *Anti-Cancer Agents Med. Chem*. **2016**, *16*, 1403–1425.

[13] Luo, Y.; Yao, J.; Yang, L.; Feng, C.; Tang, W.; Wang, G.; Zuo, J.; Lu, W. Design and synthesis of novel benzimidazole derivatives as inhibitors of hepatitis B virus. *Bioorg. Med. Chem*. **2010**, *18(14)*, 5048–5055.

[14] Tuncbilek, M,; Kiper, T.; Altanlar, N. Synthesis and in vitro antimicrobial activity of some novel substituted benzimidazole derivatives having potent activity against MRSA. *Eur. J. Med. Chem*. **2009**, *44(3)*, 1024–1033.

[15] Zhang, H.; Damu, G. L. V.; Cai, G.; Zhou, C. Design, synthesis and antimicrobial evaluation of novel benzimidazole type of Fluconazole analogues and their synergistic effects with Chloromycin, Norfloxacin and Fluconazole. *Eur. J. Med. Chem*. **2013**, *64*, 329–344.

[16] Hsieh, C.; Ko, P.; Chang, Y.; Kapoor, M.; Liang, Y. Chu, H.; Lin, H.; Horng, J.; Hsu, M. Design and synthesis of benzimidazole-chalcone derivatives as potential anticancer agents. *Molecules* **2019**, *24(18)*, 1–19.

[17] Olanrewaju, A. A.; Ibeji, C. U.; Oyeneyin, O. E. Biological evaluation and molecular docking of some newly synthesized 3d-series metal(II) mixed-ligand complexes of fluoro-naphthyl diketone and dithiocarbamate. *SN Appl. Sci.* **2020**, *2*, 678.

[18] Oyeneyin, O. E., Obadawo, B. S.; Olanrewaju, A. A.; Owolabi, T. O.; Gbadamosi, F. A.; Ipinloju, N.; Modamori, H. O. Predicting the bioactivity of 2- alkoxycarbonylallyl esters as potential antiproliferative agents against pancreatic cancer (MiaPaCa-2) cell lines: GFA-based QSAR and ELM-based models with molecular docking. *J. Gen. Engin. Biotec.* **2021**, *19*, 38.

[19] Oyeneyin, O. E.; Abayomi, T. G.; Agbaffa, E. B.; Akerele, D. D.; Amusa, O. Investigation of Amino chalcone derivatives as anti- proliferative agents against MCF-7 breast cancer cell lines-DFT, molecular docking and pharmacokinetics studies. *Adv. J. Chem.-Sect. A.* **2021**, *4(4)*, 288–299.

[20] Metibemu, D. S.; Oyeneyin, O. E.; Omotoyinbo, D. E.; Adeniran, O. Y.; Metibemu, A. O.; Oyewale, M. B.; Yakubu, F.; Omotuyi, I. O. Molecular docking and quantitative structure activity relationship for the identification of novel phyto-inhibitors of matrix metalloproteinase-2. *Sci. Lett*. **2020**, *8(2)*, 61–68.

[21] Oyeneyin, O. E.; Obadawo, B. S.; Orimoloye, S. M.; Akintemi, E. O.; Ipinloju, N.; Asere, A. M.; Owolabi, T. O. Prediction of inhibition activity of bet bromodomain inhibitors using grid search-based extreme learning machine and molecular docking. *Lett. Drug Des. Discov*. **2021**, *18(11)*, 1039–1049.

[22] Metibemu, D. S. 3D-QSAR and molecular docking approaches for the identification of novel phyto-inhibitors of the cyclin-dependent kinase 4. *Sci. Lett*. **2021**, *9(2)*, 42–48.

[23] Mills, N. ChemDraw Ultra 10.0. CambridgeSoft. *Comp. Soft. Rev*. **2006,** *128(41)*, 13649–13650.

[24] Shao, Y.; Molnar, L. F.; Jung, Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P. et al. *SPARTAN 14'*, build 1.01, (**2014**).

[25] Yap, C. W. PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comp. Chem*. **2011,** *32*, 1466–1474.

[26] Ballabio, D.; Consonni, V.; Mauri, A.; Claeys-Bruno, M.; Sergent, M.; Todeschini, R. A novel variable reduction method adapted from space-filling designs. *Chemomet. Intell. Lab Sys.* **2014**, *136*, 147–154.

[27] Kennard, R. W.; Stone, L. A. Computer aided design of experiments. *Technomet*. **1969**, *11*, 137–148.
[28] Khaled, K. F. Modeling corrosion inhibition of iron in acid medium by genetic function approximation method: A QSAR model. *Corr. Sci*. **2011**, *53(11)*, 3457–3465.
[29] Friedman, J. H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19(1)*: 1–141.
[30] Minovski, N.; Zuperl, S.; Drgan, V.; Novic, M. Assessment of applicability domain for multivariate counter-propagation artificial neural network predictive models by minimum Euclidean distance space analysis: A case study. *Analyt. Chimica Acta.* **2013**, *759*, 28–42.
[31] Roy, K.; Kar, S.; Ambure, P. On a simple approach for determining applicability domain of QSAR models, *Chemomet. Intel. Lab. Sys*. **2015**, *145*, 22-29.
[32] Sastry, M. G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comp-Aid. Molec. Des*. **2013**, *27*, 221–234.
[33] Schrödinger Release 2021-3: LigPrep, Schrödinger, LLC, New York, NY, **2021**, (n.d.).
[34] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev*. **2001**, *46*(*1-3*), 3–26.
[35] Adejoro, I. A.; Waheed, S. O.; Adeboye, O. O.; Akhigbe, F. U. Molecular docking of the inhibitory activities of triterpenoids of *Lonchocarpus cyanescens* against ulcer. *J. Biophys. Chem*. **2017**, *8(1)*, 1–11.
[36] El-Readi, N. Z.; Al-Abd, A. A.; Althubiti, M. A.; Almaimani, R. A.; Al-Amoodi, H. S.; Ashour, M. L.; Wink, M.; Eid, S. Y. Multiple molecular mechanisms to overcome multidrug resistance in cancer by natural secondary metabolites. *Front. Pharmacol*. **2021**, *12*, 658513.
[37] Omoboyowa, D.A.; Omomule, O.M.; Balogun, T.A.; Saibu, O.A.;Metibemu, D.S. Protective potential of ethyl acetate extract of *Abrus precatorius* (Linn) seeds against HCl/EtOH-induced gastric ulcer via pro-inflammatory regulation: *In Vivi* and *in silico* study. *Phytomed. Plus*. **2021**, 1 100145.
[38] Angeli, E.; Nyugen, T.T.; Janin, A.; Bousquet, G. How to make anticancer drugs cross the blood–brain barrier to treat brain metastases. *Int. J. Mol. Sci*. **2020**, *21*, 22.
[39] Singh, Y.P.; Singh, R.A. Theoretical studies of different tautomers of anti cancer drug: dichloroacetate. *Pakist. J. Pharmac. Sci*. **2008**, *21*, 390-395.