African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 3, 2020 (pp. 67-89)
www.abjournals.org

DOCKING AND QSAR STUDIES OF NEW IMIDAZO [1,2-A] QUINOXALINE DERIVATIVES USING GENETIC FUNCTION APPROXIMATION (GFA) AGAINST HUMAN MELANOMA

Emmanuel Israel Edache[1*] and Shafiu Saidu[2]

[1]Department of Pure and Applied Chemistry, University of Maiduguri, Brono State, Nigeria
[2]Department of Chemistry, Kaduna State University, Kaduna-Nigeria
[*]Email: edacheson2004@gmail.com, Tel: +2348066776802

ABSTRACT: *In this paper, to comprehend the chemical-biological interactions governing their activities toward antitumor activity, QSAR models of 31 derivatives of New imidazo [1, 2-a] quinoxaline derivatives with inhibitory tumor were developed. The quantitative structure-activity relationship (QSAR) model was built by using the genetic function algorithm (GFA) technique, and the best GFA model has SEE = 0.51748, R^2 = 0.73038 cross-validated, $R^2_{adjusted}$ = 0.63234, F = 7.44967 (DF: 4, 11), and Q^2 = 0.51664 non-cross-validated. The predictive ability of the GFA model was further validated by a test set of 8 compounds, giving R^2_{pred} = 0.73038. Docking studies were used to discover the real conformations of chemicals in the active site, as well as the binding mode shape to the binding site in enzyme. Ligand with PubChem_CID number 44561182 has the least binding affinity with the enzyme. The information provided by the 2D-QSAR model and docking may lead to a better understanding of the structural requirements of 31 New imidazo [1, 2-a] quinoxaline derivatives and help to design potential anti-tumor molecules.*

KEYWORDS: Tumor, Quinoxalinamine Derivatives, GFA-MLR, QSAR, Applicability Domain, Molecular Docking

INTRODUCTION

Regular cells in our bodies follow an organized path of development, separation, and death [1]. The tumor is a class of diseases categorized by out-of-control cell development or irregular development of the cells with the ability to spread in the other parts of the body which damages the body by forming bumps or multitudes of tissue called tumor [1, 2]. Despite increased attention to tumor eradication, the disease causes more than a million deaths each year and is the second cause of death in the world [3, 4]. In spite of several efforts in the handling of a tumor, and the limitations that medications have, this disease became a big problem for the health of societies [2]. Extensive laboratory suggestions from chemical, cell culture, and animal studies indicate that antioxidants may slow or probably prevent the growth of tumors [5]. Quinazoline ring is a versatile lead molecule that has been studied widely which possesses analgesic, anti-inflammatory, antitumor, antimicrobial, anticonvulsant, enzyme inhibition activity, and many other activities [6, 7]. Several studies have shown that the Quinazoline nucleus has potent activity against human cancer particularly by killing the cells in a tumor-specific manner [8]. The nucleus has also been reported to have potent antioxidant activity [5]. Amino acids will reduce the side effects of the metabolite of the parent compound upon

metabolism in the body and enhance the solubility of the synthesized candidates when it is incorporated into pharmacologically active quinazolinone moiety [9]. These observations gave us a great motivation to the search for potential biological active drugs carrying new imidazo[1,2-a] quinoxaline analogs which have been synthesized in good yields via a bimolecular condensation of 2-imidazole carboxylic acid, followed by a coupling and subsequent substitution on the imidazole ring by Suzuki Cross-coupling reaction using microwave assistance [10]. Quantitative Structure-Activity Relationship (QSAR) modeling and docking has become an extensively used tool [11] in computer-aided drug design (CADD). CADD has been used to predictive environmental risk assessment and fate modeling, toxicity and property prediction of chemicals and pharmaceuticals [12, 13] as well as in different modeling problems in material sciences, analytical chemistry and pharmacokinetics/pharmacodynamics profiling of new drug molecules [14]. Computer-aided drug design (CADD), is utilized to achieve the desired results. CADD provides appreciated perceptions into experimental findings and mechanism of action, new suggestions for molecular structures to synthesize, and can help make cost-effective decisions before expensive synthesis is started. Quantitative Structure-Activity Relationship (QSAR) modeling is a ligand-based drug design method for both exploring and exploiting the relationship between chemical structure and its biological action [15, 16, and 17]. To predict the activities of anticancer compounds, quantum chemical descriptors like molecular orbital, dipole moment, charge, etc. and molecular property descriptors like hydrophobic, steric coefficient, etc. have been applied to develop 2D QSAR models [18, 19], while docking study is also performed to explore the binding pocket in the enzyme and to understand the binding mode pattern for each compounds [20].

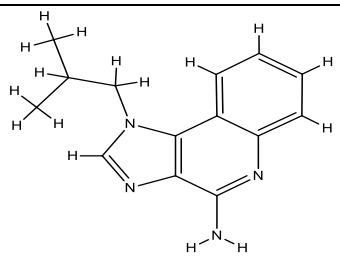
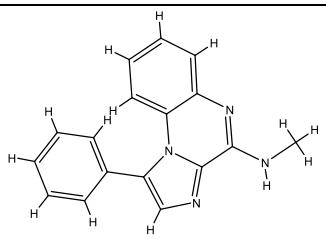
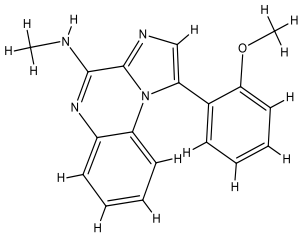
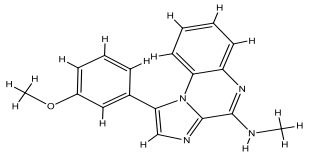
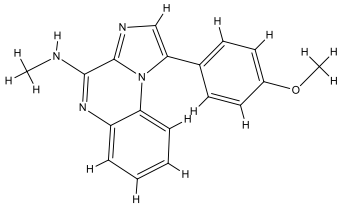
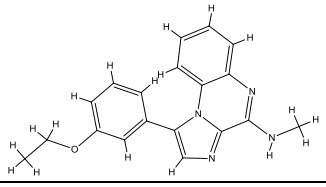
MATERIALS AND METHODS

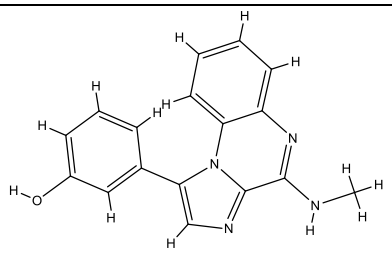
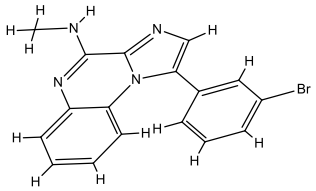
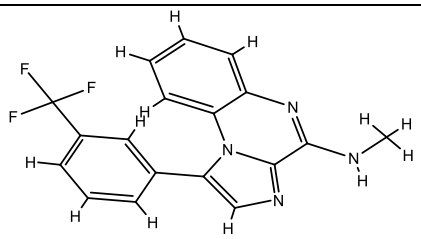
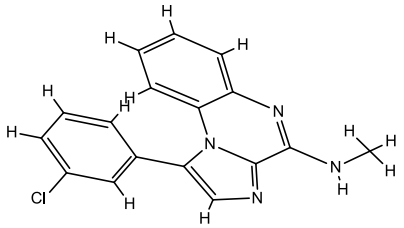
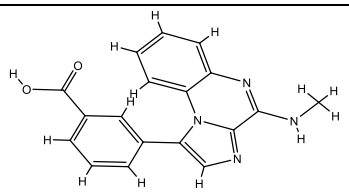
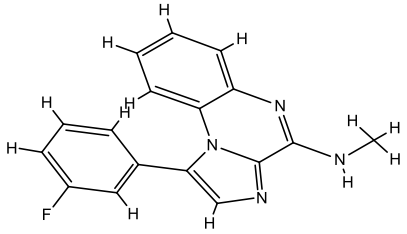
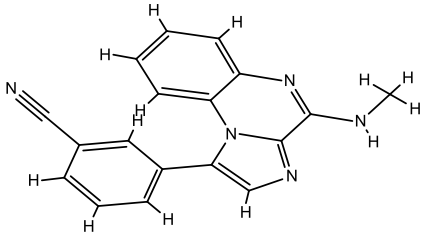
In the present paper, structure-cytotoxic activity relationships of a series of 31 compounds of New imidazo[1,2-a] quinoxaline derivatives against human A375 cells (MTT) reported by PubChem [21, 22] have been used for the 2D-QSAR model development. The Marvin Sketch and Marvin View software package were used to building the chemical structures. Density functional theory was adopted at the level of B3LYP/6-31+G (d, p) [23] to optimize the isolated compounds using Spartan'14 version 1.1.4 software. Descriptor generation was performed using the PaDel package version 2.20 [24]. DTC Lab (new_Dataset Division GUI v1.2 Division) software were used to divide the compounds into training and test set. The Kennard-Stone algorithm was used to split data sets into two distinct subsets with an equal distribution such that no sample from one subset should be too far from any sample of the other subset, and the coverage should start on the boundary of the factor space [25, 26, 27, and 28], followed by the application of statistical methods (GFA-MLR) to determine the main descriptors responsible for the cytotoxic activity of the compounds under investigation. Nowadays, the genetic function approximation (GFA) method [29] is considered superior to other variable selection methods. It is a powerful optimization method that was inspired by evolutionary principles, including survival of the fittest, reproduction, crossover, and mutation. In this study, GA-MLR was used to build the QSAR model. The fitness function utilized herein was the leave-one-out (LOO) cross-validated correlation coefficient (Q^2). Measurement of cytotoxic activity is expressed as half-maximal (50%) inhibitory concentration of a substance (IC_{50}) and values are expressed in Nano molar (nM – 10^{-9}) levels. The values were converted to the pIC_{50} scale (-log IC_{50}) to predict the narrow value wherein higher values indicate exponentially

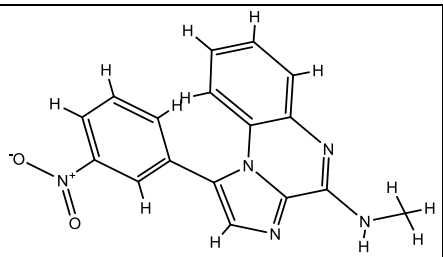
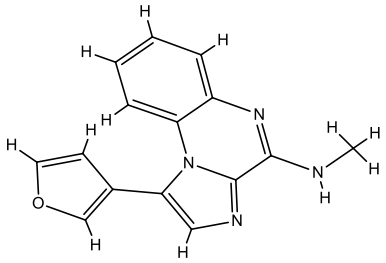
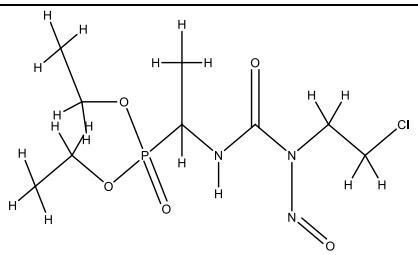
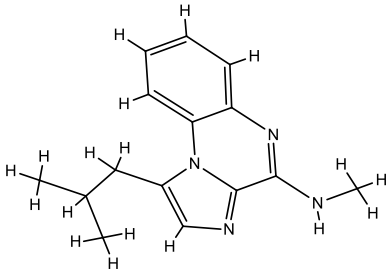
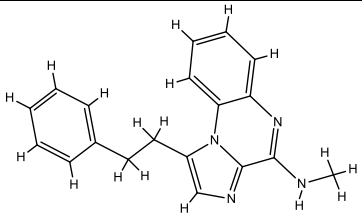
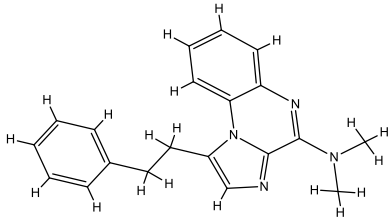
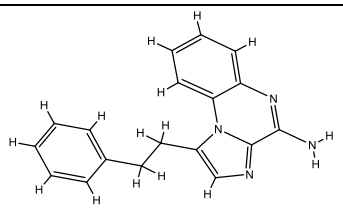
African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 3, 2020 (pp. 67-89)
www.abjournals.org

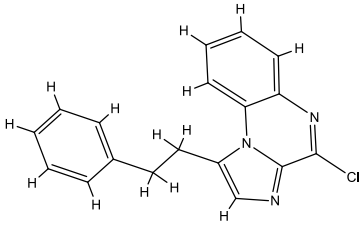
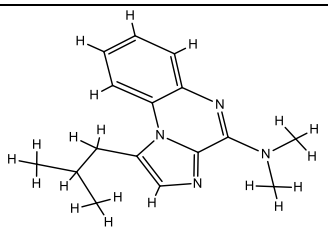
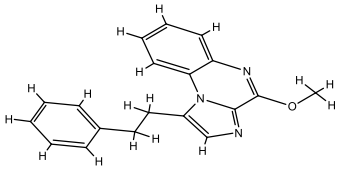
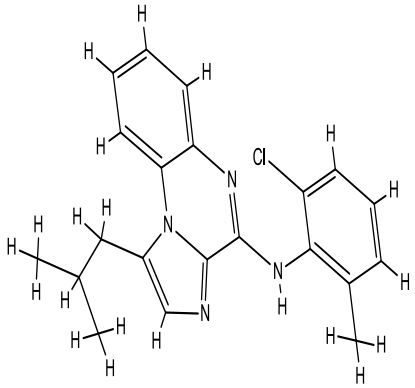
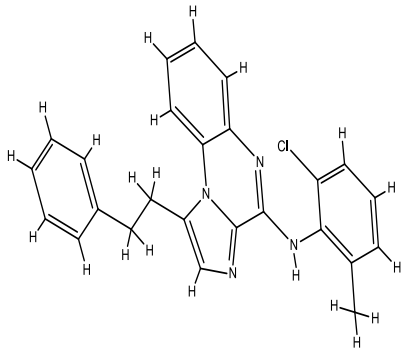
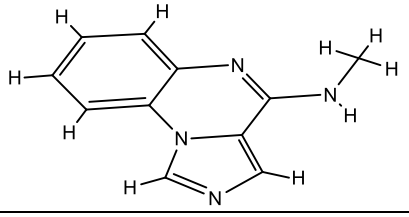
greater potency. The pIC_{50} values were used as the dependent variables to construct the QSAR model. The structures of these compounds and their PubChem_CID numbers are shown in Table 1.

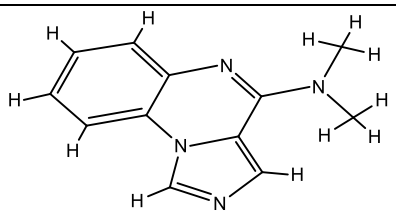
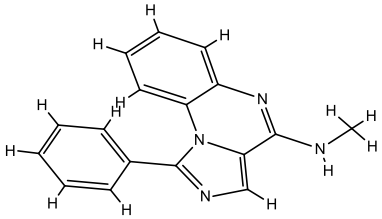
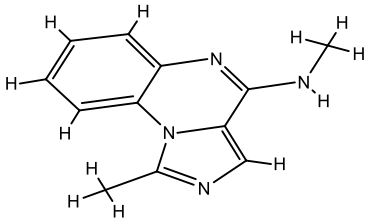
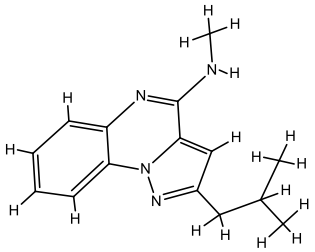
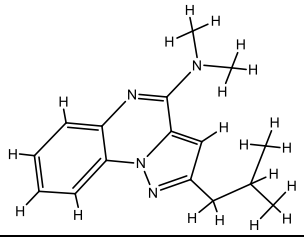
Table 1: Structures of dataset used for GA-MLR QSAR analysis with compound CID, IUPAC CAS name and experiment IC_{50}.

S/N	Structure	PubChem_CID	IUPAC_CAS_NAME	IC_{50}
1[b]		57469	1-(2-methylpropyl)-4-imidazo[4,5-c]quinolinamine	70.3
2[a]		25253254	N-methyl-1-phenyl-4-imidazo[1,2-a]quinoxalinamine	74.1
3[c]		25253255	1-(2-methoxyphenyl)-N-methyl-4-imidazo[1,2-a]quinoxalinamine	0.65
4[b]		25253256	1-(3-methoxyphenyl)-N-methyl-4-imidazo[1,2-a]quinoxalinamine	3.47
5[a]		25253257	1-(4-methoxyphenyl)-N-methyl-4-imidazo[1,2-a]quinoxalinamine	0.3
6[b]		44224743	1-(3-ethoxyphenyl)-N-methyl-4-imidazo[1,2-a]quinoxalinamine	0.56

7[a]		44224744	3-[4-(methylamino)-1-imidazo[1,2-a]quinoxalinyl]phenol	1.28
8[a]		44224745	1-(3-bromophenyl)-N-methyl-4-imidazo[1,2-a]quinoxalinamine	173
9[a]		44224966	N-methyl-1-[3-(trifluoromethyl)phenyl]-4-imidazo[1,2-a]quinoxalinamine	122
10[a]		44224967	1-(3-chlorophenyl)-N-methyl-4-imidazo[1,2-a]quinoxalinamine	0.2
11[b]		44224968	3-[4-(methylamino)-1-imidazo[1,2-a]quinoxalinyl]benzoic acid	0.37
12[b]		44224969	1-(3-fluorophenyl)-N-methyl-4-imidazo[1,2-a]quinoxalinamine	2.19
13[a]		44224970	3-[4-(methylamino)-1-imidazo[1,2-a]quinoxalinyl]benzonitrile	24.9

14[a]		44224971	N-methyl-1-(3-nitrophenyl)-4-imidazo[1,2-a]quinoxalinamine	27
15[b]		44224972	1-(3-furanyl)-N-methyl-4-imidazo[1,2-a]quinoxalinamine	40
16[c]		104799	1-(2-chloroethyl)-3-(1-diethoxyphosphorylethyl)-1-nitrosourea	31.6
17[a]		10244054	N-methyl-1-(2-methylpropyl)-4-imidazo[1,2-a]quinoxalinamine	1.57
18[a]		24779760	N-methyl-1-(2-phenylethyl)-4-imidazo[1,2-a]quinoxalinamine	2.35
19[a]		25254189	N,N-dimethyl-1-(2-phenylethyl)-4-imidazo[1,2-a]quinoxalinamine	24
20[a]		25254190	1-(2-phenylethyl)-4-imidazo[1,2-a]quinoxalinamine	66.3

21[b]		44561144	4-chloro-1-(2-phenylethyl)imidazo[1,2-a]quinoxaline	80.1
22[a]		44561145	N,N-dimethyl-1-(2-methylpropyl)-4-imidazo[1,2-a]quinoxalinamine	47.8
23[b]		44561146	4-methoxy-1-(2-phenylethyl)imidazo[1,2-a]quinoxaline	400
24[c]		44561181	N-(2-chloro-6-methylphenyl)-1-(2-methylpropyl)-4-imidazo[1,2-a]quinoxalinamine	329
25[c]		44561182	N-(2-chloro-6-methylphenyl)-1-(2-phenylethyl)-4-imidazo[1,2-a]quinoxalinamine	85.3
26[a]		44561183	N-methyl-4-imidazo[1,5-a]quinoxalinamine	121

27c		44561184	N,N-dimethyl-4-imidazo[1,5-a]quinoxalinamine	400
28[c]		44561185	N-methyl-1-phenyl-4-imidazo[1,5-a]quinoxalinamine	100
29[a]		44561186	N,1-dimethyl-4-imidazo[1,5-a]quinoxalinamine	78.6
30[a]		44561225	N-methyl-2-(2-methylpropyl)-4-pyrazolo[1,5-a]quinoxalinamine	101
31[c]		44561226	N,N-dimethyl-2-(2-methylpropyl)-4-pyrazolo[1,5-a]quinoxalinamine	173

Training set[a]; Test set[b]; outliers[c]

The relationship between dependent variable and independent variables was established by GFA-multiple regression analysis using Material Studio. Significant descriptors were chosen based on the statistical data analysis. Statistical quality of the generated QSAR equation was judged based on the parameters like Friedman LOF and Cross-validation R-squared. Cross-validation is a popular method used to test the reliability of QSAR models. In this study, leave-one-out (LOO) method was applied to generate a number of improved data sets by removing the first row and its value predicted using the rest of the data. Also, each row is left in turn, so that the value of each row is predicted from all others. The model is judged based on these predictions. Predictive ability of the generated model was assessed externally by predicting the activities of test set. This condition may not be sufficient for a QSAR model to be truly predictive [30]. An additional condition for high predictive ability of QSAR model is based on external set cross-validation R^2, (R^2_{pred}) and the regression of observed activities against predicted activities and vice versa for validation set, if the conditions of Golbraikh and Tropsha

are satisfied [30]. Calculations relating to R^2_{pred}, R_0^2 and the slopes, k and k' are based on regression of observed values against predicted values and vice versa.

Molecular Descriptors

In spite of great advances in the field of drug design, the use of descriptors to define the molecular structure of biologically active compounds is the key method used to discover new lead molecules. Descriptors are the biochemical representative of a molecule in numerical form, used for QSAR studies. The information encoded by descriptors normally depends on the kind of molecular representation and the defined algorithm for its calculation. Some of these include: topological indices, and geometrical, constitutional and physicochemical descriptors. Constitutional descriptors are simple, commonly used descriptors reflecting the molecular composition of a compound without any information about its topology. The most common constitution descriptors are number of atoms, bond count, atom type, ring count, and molecular weight (MW). These descriptors are inert to any conformation change and, therefore, do not differentiate among isomers [31].

Domain of Applicability

Applicability domain of a QSAR model must be clear if the model is to be used for screening new compounds. Predictive ability of the model may be considered reliable if the data set falls into this domain [32, 33, and 34]. One simple approach is based on Euclidean based and the calculation of leverages for compounds are used in the study.

Y-randomization

Another simple way of proving that the structure-activity relationships established in this study do not result from coincidence involves checking their robustness by means of the so-called Y-randomization [35]. This test ensures the robustness of a QSAR model [32] and to assess the multiple linear regression models obtained by descriptor selection. In y-randomization test, the dependent variable (pIC_{50}) is randomly shuffled and a new QSAR model is developed keeping X-data (descriptors) intact. The new models are expected to have low R^2 and Q^2 values, which determine the statistical significance of the original model.

Molecular Docking

Docking studies were performed to further explain the results of our 2D-QSAR study and explore the possible binding modes of these compounds. Docking is often utilized to forecast the binding orientation of small molecules (drug candidates) to their macromolecular target (such as protein, carbohydrate and nucleic acid) with the aim to determine their tentative binding parameters [36]. This establishes raw data for the rational drug designing (structure-based-drug development) of new agents with better efficacy and more specificity [37]. The research in the drug discovery process involves virtual screening (VS) which is a computational technique used for the rapid exploration of huge collections of chemical structures in order to recognize those structures that are most likely to bind to a drug target, usually a protein receptor or enzyme [38]. Identification of possible protein targets of small chemical molecules is a main step for unravelling their primary causes of actions at the molecular level [39]. Docking studies were employed to locate the appropriate binding orientations and conformations of these New imidazo[1,2-a] quinoxaline derivatives interacting with ligand using the docking program Molegro Virtual Docker, iGemdock and PyRx. By default, the docking program produces

docked structures for each New imidazo[1,2-a] quinoxaline derivatives. The conformation with the lowest docking energy in the most inhabited cluster is selected as the possible 'active' conformation against the protein active site. The X-ray crystal structure, taken from the Protein Data Bank (pdb: 6EBE) was used to dock. At the beginning of docking, all the water molecules were removed and hydrogen atoms added to the protein were applied. The structures were docked by the active site defined through a grid box (Vina Search Space: center_x = -9.7952, center_y = -1.7878, center_z = 16.0475), Dimensions (Angstrom): size_x = 47.5025488281, size_y = 44.5045238495, size_z = 54.4464433479 and exhaustiveness = 8.

RESULTS AND DISCUSSIONS

pIC_{50} = -7.20917 - 1.4154ALogP - 0.06171AMR +2.92517SC-3 + 3.74747maxaasC ----Model 1

Description about selected variables are as follows:

ALogP (PaDEL; 2D): Ghose-Crippen LogKow

AMR (Dragon; Molecular properties): Ghose-Crippen molar refractivity

SC-3(PaDEL; 2D): Simple cluster, order 3

MaxaasC (PaDEL; 2D): Maximum atom-type E-State: C: -

SEE :0.51748, r^2 :0.73038, r^2 adjusted :0.63234, F :7.44967 (DF :4, 11), Q2 :0.51664, LOF: 1.2666, PRESS :5.2809, SDEP :0.5745, r^2 :0.73038, r0^2 :0.73038, reverse r0^2:0.67893, rm^2(test) :0.73038, reverse rm^2(test) :0.56471, average rm^2(test) :0.64755, delta rm^2(test) :0.16567, rmsep:0.42907, rpred^2 :0.73038, Q2f1 :0.73038, Q2f2 :0.73038

Some External Validation Parameters

	Without scaling	After scaling
rm^2(overall):	0.59546	0.57463
reverse rm^2(overall):	0.47544	0.48104
average rm^2(overall):	0.53545	0.52783
delta rm^2(overall):	0.12001	0.09359

Golbraikh and Tropsha acceptable model criteria's:

1. Q^2 0.51664, Passed (Threshold value Q^2>0.5)

2. r^2 0.73038, Passed (Threshold value r^2>0.6)

3. |r0^2-r'0^2| 0.05145, Passed (Threshold value |r0^2-r'0^2|<0.3)

4. k = 1, [(r^2-r0^2)/r^2] = 0 OR k' = 0.86976, [(r^2-r'0^2)/r^2] = 0.07044 Passed

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 3, 2020 (pp. 67-89)

www.abjournals.org

(Threshold value: [0.85<k<1.15 and ((r^2-r0^2)/r^2) <0.1] OR [0.85<k'<1.15 and ((r^2-r'0^2)/r^2) <0.1])

Table 2: GA-MLR QSAR analysis with corresponding Experimental and predicted class of tumour inhibitors.

PubChem_CID (Train)	Yobs	Ypred	(Residual)^2	(Yobs-Ybar)^2
44224745	0.187087	-0.36463	0.304392	1.085594
44224966	-0.10721	-0.39265	0.081477	0.558938
44224967	-0.25042	-0.04276	0.043121	0.365314
44224970	-0.54033	-0.27352	0.071186	0.098912
44224971	-1.3962	-1.58743	0.036567	0.293079
10244054	-1.43136	-0.99634	0.189246	0.332389
25254189	-1.86982	-0.73271	1.293012	1.030197
25254190	-1.49969	-1.38916	0.012215	0.415838
44561145	-0.37107	0.182161	0.306062	0.234028
44561183	-1.38021	-1.45988	0.006347	0.276024
24779760	-1.67943	-1.261	0.175082	0.679959
44561186	-1.93095	-2.97387	1.087687	1.158028
44561225	-0.1959	-0.64536	0.202017	0.434192
25253254	-1.89542	-1.54185	0.125011	1.082829
25253257	0.431798	-0.66553	1.204123	1.655417
44224744	0.251812	-0.1268	0.143351	1.22466
PubChem_CID (Test)	Yobs	Ypred	(Residual)^2	(Yobs-Ybar)^2
57469	-1.84696	-3.39226	2.387952	0.984309
44224968	-1.60206	-1.56728	0.00121	0.55835
44224969	-1.90363	-0.17617	2.984113	1.099983
44224972	-1.82151	-1.005	0.6667	0.934474
44561144	-2	-1.75506	0.059994	1.31141
44561146	-0.34044	-0.50363	0.026629	0.264595
25253256	0.69897	-0.62523	1.753496	2.4143
44224743	0.522879	-1.20436	2.983337	1.898086

A total of 24 compounds (Table 1) were used for the QSAR model generation. It is essential to assess the predictive power of models by using a test set of compounds. This was achieved by setting aside 8 compounds as a test set such that it represented the various functional groups included in the training set and had a regularly distributed biological data. The mean of the biological activity of the training and test set was -0.9049 and -0.9364, respectively. The quality factor (Q) was performed to access the robustness and statistical confidence. The higher value of R^2, RMSEP, Q, and F and lower value of Se, and RMSECV of Model 1 revealed that Model 1 was robust and promising. In the developed Model the value of the coefficient of correlation was significantly high supporting reliability and goodness. The accuracy of the Model 1 was ascertained by correlation coefficient (R^2 = 0.7304), statistical significance more than 95% (against tabulated value F = 3.4035) and low standard error of estimate (0.5175). The model shows that parameter SC-3 and maxaaC showed a positive contribution. The regression model

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 3, 2020 (pp. 67-89)
www.abjournals.org

has small residuals that can be seen in (Table 2). LOO cross-validation analysis revealed that R^2-Q^2_{LOO} < 0.3 (0.7304 – 0.5166 = 0.21). The robustness of the model was justified According to Golbraikh and Tropsha [30], the proposed QSAR model is predictive as it satisfies this conditions like $R^2_{pred} > 0.5$, $R^2 > 0.6$, r^2-$r^2o/r^2 < 0.1$, r^2-$r'^2o/r^2 < 0.1$ and $0.85 \leq K \leq 1.15$ or $0.85 \leq K' \leq 1.15$, but this model satisfy the following criteria $R^2_{pred} = 0.7304 > 0.5$, and $R^2 = 0.7304 > 0.6$. So, this QSAR model is predictive as it's satisfied this condition reported by Golbraikh and Tropsha, [30]. The model also possesses a high value of FIT criterion [40, 41], the internal validation parameter of the model ($Q^2_{cv} = 0.5166$) was also good. It can be observed that the obtained model has sensible internal and external value. Nevertheless, it is always necessary to obtain a model that can relate the physicochemical properties represented by the selected molecular descriptors to the action mechanism of the system under study [42]. The estimation of probable error of the coefficient of correlation (PE) is another requirement for validating the method [43]. This is defined as PE = $2/3(1-R^2/\sqrt{n})$: Where R in multiple correlations is the correlation coefficient and n is the number of compounds under study. It is argued that: (1) if R < PE, then R is not significant; (2) if R > PE, several times; at least 3-times greater correlation is indicated, and (3) if R > 6PE, then the correlation is good. The 6PE = 0.269617 indicates that the proposed correlations are good.

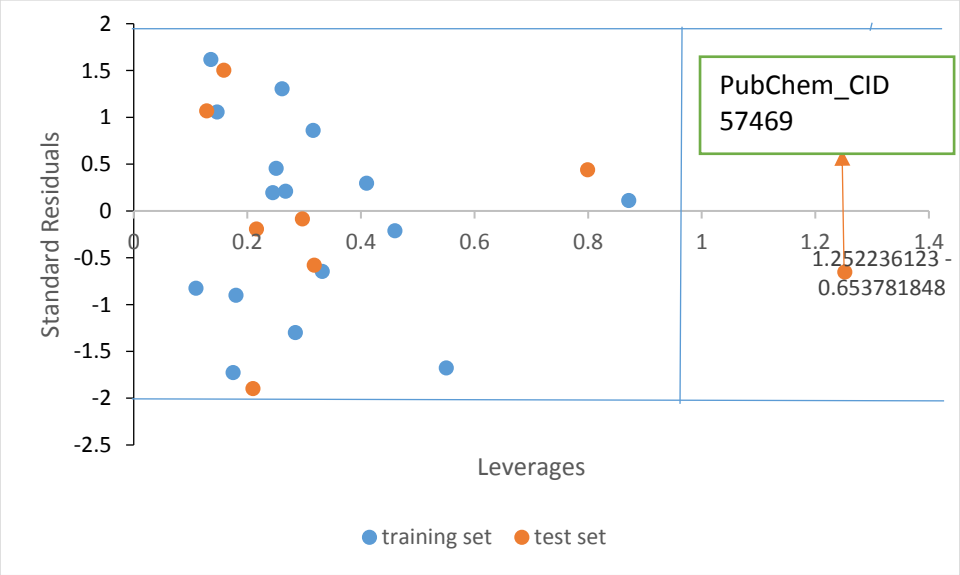


Fig. 1: The Williams plot, the plot of the standardized residuals vs. leverages

Figure 1 shows the standardized residuals (σ) versus the leverage samples plot, and it was used for the identification of outliers. No compound presented residuals higher than 2.5σ. Only one compound presented leverage higher than the leverage cutoff line, but it can be considered acceptable [44]. Therefore, the model can be considered free of outliers, something which guarantees the maximum possible representation in terms of structure and range of inhibitory activity for the dataset under study. Euclidean based applicability domain helps to specify the scope of their proposed models, therefore, defining the model limitations concerning its structural domain and response space. If an external compound is beyond the defined scope of a given model, it is considered outside that model's Applicability Domain (AD) and cannot be associated with a reliable prediction. The resulting model can be reliably applicable for only

those compounds which are inside this domain. Euclidean based applicability domain helps to ensure that the compounds of the test set are representative of the training set compounds used in model development. It is based on distance scores calculated by the Euclidean distance norms. At first, normalized mean distance scores for training set compounds are calculated and these values range from 0 to 1(0 = least diverse, 1 = most diverse training set compound). Then normalized mean distance score for test set are calculated (Table 3), and those test compounds with scores outside 0 to 1 range are said to be outside the applicability domain. This can also be checked by plotting a 'Scatter plot' (normalized mean distance vs. respective activity) including both training and test set as shown in Figure 2. If the test set compounds are inside the domain/area covered by training set compounds that means these compounds are inside the applicability domain otherwise not [32, 34].

Table 3: Euclidean based applicability domain (AD) for Model 1

PubChem_CID (Training Set)	**Distance Score**	**Mean Distance**	**Normalized Mean Distance**
44224745	70.48381	4.405238	0.058052
44224966	68.54478	4.284049	0.043204
44224967	64.68526	4.042829	0.01365
44224970	64.30307	4.018942	0.010723
44224971	102.1794	6.386216	0.30076
10244054	62.90267	3.931417	0
25254189	93.37212	5.835757	0.233318
25254190	193.4945	12.09341	1
44561145	75.42328	4.713955	0.095876
44561183	64.86008	4.053755	0.014989
24779760	67.26472	4.204045	0.033402
44561186	114.4197	7.151229	0.394489
44561225	76.78591	4.79912	0.10631
25253254	193.4887	12.09304	0.999955
25253257	63.43943	3.964964	0.00411
44224744	84.76635	5.297897	0.16742
PubChem_CID (Test Set)	Distance Score	Mean Distance	Normalized Mean Distance
57469	123.8296	7.739348	0.466544
44224968	74.70109	4.668818	0.090346
44224969	144.0752	9.004703	0.621575
44224972	284.4916	17.78073	1.696805
44561144	64.35608	4.022255	0.011129
44561146	114.4033	7.150205	0.394363
25253256	63.47061	3.966913	0.004349
44224743	103.5443	6.471521	0.311211

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 3, 2020 (pp. 67-89)
www.abjournals.org

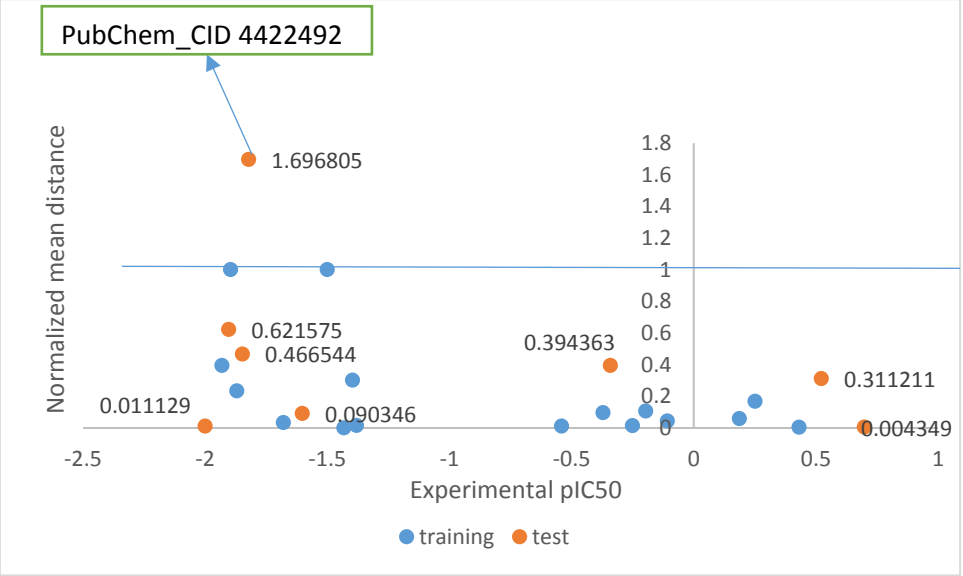


Fig. 2: Euclidean Based applicability domain plot, the plot of the normalized mean distance vs. observed IC_{50}

In addition to Euclidean based applicability domain (AD), the internal predictive ability of the model was further assessed by a Y-randomization performed with 16 derivatives for 10 times. The average of 10 readings was given as average Q^2 as shown in Table 4; the Y-randomization test (Table 4) ensures the robustness of a QSAR model [32] and to assess the GFA models obtained by descriptor selection [45]. In the Y-randomization test, the dependent variable is randomly shuffled and a new QSAR model is developed keeping molecular descriptors intact. The new models are expected to have low R^2 and Q^2_{LOO} values, which determine the statistical significance of the original model. The low R^2 and Q^2_{LOO} values of the random models shown in Table 4 and the value of R^2_p = 0.5894 ($R^2_p \geq 0.5$) indicates that there is no chance of correlation or structural dependency in the proposed model. Consequently, model 1 can be considered as a perfect model with both high statistically significant and excellent predictive ability.

Table 4: The average R, R^2 and Q2LOO values after several Y-Randomization

Model 1	R	R^2	Q^2
Original	0.854625	0.730383	0.51664
Random 1	0.344168	0.118452	-1.21506
Random 2	0.299327	0.089597	-1.58454
Random 3	0.411688	0.169487	-1.23794
Random 4	0.608113	0.369801	-0.36964
Random 5	0.459707	0.211331	-0.76204
Random 6	0.730684	0.5339	-0.99338
Random 7	0.371123	0.137732	-1.61845
Random 8	0.62597	0.391838	-0.00437
Random 9	0.550345	0.30288	-0.33447
Random 10	0.646824	0.418382	-0.7894

Random Models Parameters	Results
Average r :	0.504795
Average r^2 :	0.27434
Average Q^2 :	-0.89093
cRp^2 :	0.58936

Interpretation of descriptors

By understanding the descriptors contained in the QSAR model, it is possible to gain certain insights into issues, which are related to the antitumor activity. For this reason, an acceptable interpretation of the selected descriptors is provided below. The brief descriptions of descriptors are shown in model 1 and Table 5. To observe the significance as well as the involvement of each descriptor in the model, the value of the mean effect (MF) was calculated for each descriptor [46, 47]. The MF value shows the relative importance of a descriptor, related with the other descriptors in the model. Its sign designates the variation direction in the values of the activities as a result of the increase (or decrease) of the descriptor values.

Table 5: Variance Inflation Factor (VIF) and Mean effect (MF) of each Descriptors

Descriptors	**VIF**	**MF**
AlogP	2.030022475	-0.006133863
AMR	1.328239136	-0.23656501
SC-3	2.401363209	0.593332196
maxaasC	2.580941216	0.649366678

AlogP –thermodynamic descriptor. AlogP is the partition coefficient calculated using atom-based approach and represents the hydrophobicity of the molecules [48]. A negative mean effect of this descriptor illustrates that the activity increases with decreasing the value of AlogP, which means that the partition coefficient calculated using atom-based approach and represents the hydrophobicity of the molecules will benefit the activity.

Quantum mechanical calculations have become routine even for large molecular systems and therefore the information related to the structure and electronic distribution can be easily and efficiently used in deriving new descriptors and explaining the properties of molecules. AMR Ghose-Crippen molar refractivity, a quantum mechanical descriptor, a measure of the total polarizability of a mole of a substance, can be estimated, is a common molecular descriptor accounting for molecular size and polarizability [49]. The negative sign of this descriptor (Table 2) indicates that the pIC_{50} value is indirectly related to this descriptor. Hence, it was concluded that by decreasing the molecular size and polarizability, the value of this descriptor decreased, causing an increase in its pIC_{50} value.

The selected parameter SC-3 (PaDEL; 2D): Simple cluster, order 3 encode structure information specifically based on a branch point, emphasizing the immediate branch point environment. SC-3 exhibits the second largest positive mean effect (MF) Table 5 (0.5933) to pIC_{50}. Since the value of SC-3 (PaDEL; 2D): Simple cluster, order 3 is positive, increases in values of SC-3 (PaDEL; 2D): Simple cluster, order 3 are advantageous to improving the antitumor activity.

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 3, 2020 (pp. 67-89)
www.abjournals.org

The desirability and suggested advantage of E-states over simple counts of the equivalent atom types is that E-states values for each atom in a given molecule 'reflect' the steric and electronic effects of the surrounding atoms and as such, could be best described as information-rich atomic descriptors [50, 51]. Electro topological state atom type descriptor MaxaasC represents Maximum atom-type E-State: C: -; this descriptor contributes positively MF (Table 5) which indicates that inhibitory activity of New imidazo [1, 2-a] quinoxaline derivatives will increase with Maximum atom-type E-State: C: -. The corresponding VIF values of the four descriptors are presented in Table 5. As can be seen from this table, all the variables have VIF values of less than five, indicating that the obtained model has statistical significance, and the descriptors were found to be reasonably orthogonal [52, 53].

Docking Analysis

One application of molecular docking is to design therapeutic in-silico by optimizing targeted lead candidates against the protein. The lead candidates can be found using a docking algorithm that aims to identify the optimal binding mode of a small molecule (ligand) to the active site of the macromolecular target. Twenty-four New imidazo [1, 2-a] quinoxaline derivatives to obtain more effective compounds as inhibitors of the tumor. Molegro virtual docker (MVD) was used to predict various orientations or conformations of the drugs against the protein targets. The conformations with the least binding energies were selected and saved. The average binding energies were calculated for each ligand after ten simulations with MVD. The Hydrogen bond score, Number of Hydrogen bond, and interacting residues of the protein with the ligands were also analyzed using the software. In the case of the crystal structures PDB code: 6EBE complexes [54], the program generally identified three different binding sites (Figure 3).

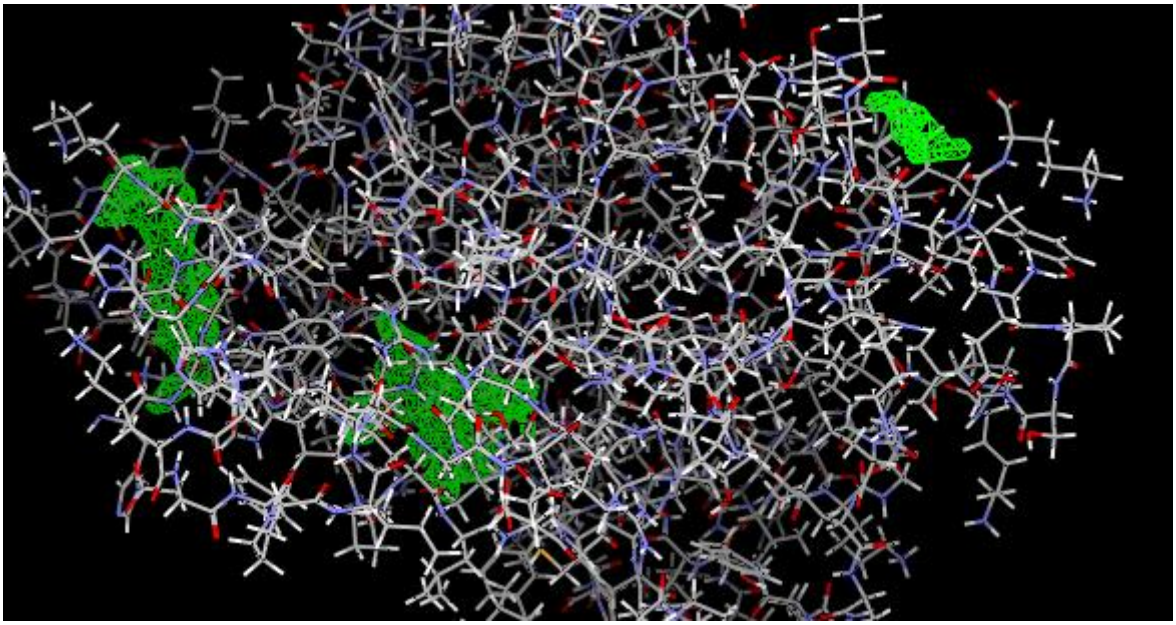


Fig. 3: The three cavity MVD-detected cavities in anti-tumor, (PDB code 6EBE), detected cavity green, carbon atoms grey, oxygen atoms red, nitrogen atoms blue.

Additionally, docking of these ligands, New imidazo [1, 2-a]quinoxaline derivatives was performed with the crystal structure and each molecule selects the best position to define the

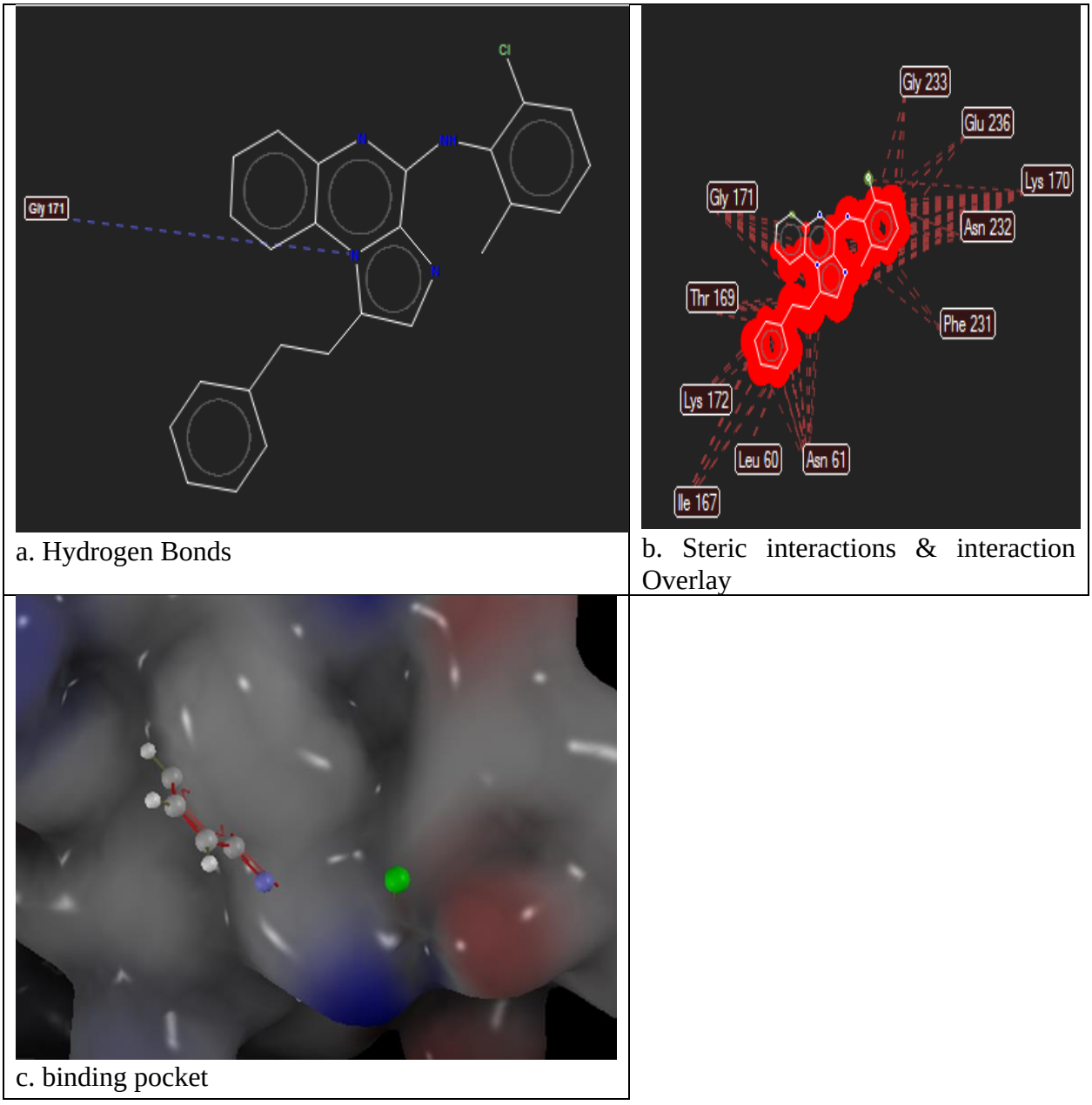
re-rank score. In each docking run, the best poses were selected based on their MVD re-rank scores, and the mean of the six re-rank scores were then computed as the final score for each molecule. The MVD score and the re-rank scores of the best poses for each of the docking studies New imidazo[1,2-a] quinoxaline derivatives with protein PDB code: 6EBE are summarized in Table 6.

Table 6: Plant Score, MolDock Score and Re-rank score (kcal/mol) for New imidazo[1, 2-a]quinoxaline derivatives when docked with 6ebe crystal structure

Pubchem_CID (Ligand)	Plant Score	MolDock Score	Re-rank Score
44561182	-80.1715	-115.862	-88.8733
44561144	-70.2567	-82.5739	-61.3965
25254189	-70.2156	-84.4104	-65.7448
25254190	-69.892	-92.7031	-67.7022
25253257	-63.4977	-100.392	-77.4534
44561145	-61.8085	-78.6937	-65.414

Furthermore, the obtained score is between -78.6937 and -115.862cal/mol. Moldock score of target compound PubChem_CID 44561182 is lower than PubChem_CID 44571145. PubChem_CID 44561182 is a proven active compound as an inhibitor of anti-tumor protein PDB code: 6EBE. By blocking PBD binding to its recognition motif, PubChem_CID 44561182 disrupted the human anti-tumor protein PDB code: 6EBE subcellular localization and eventually arrested the cell cycle [55]. The docking results compound with 6EBE reveals no electrostatic interactions but it has a hydrogen bonding and steric interaction between the ligand to the receptor. The key residues in the binding pocket were Gly233, Glu236, Lys170, Asn232, Phe231, Asn61, Leu60, Ile167, Lys172, Thr169, and Gly171. The ligand formed hydrogen bonding (H-bond) with Gly171 in the binding site. This result indicates that the above-mentioned molecules are predicted to be an antitumor drug candidate. The superposition of ligand as observed in the cavity of the crystallographic structure of 6EBE and the best conformation obtained theoretically are shown in Fig. 4. The result suggests that the software reproduced the appropriate conformation of PubChem_CID 44561182 inside its binding site in the protein (PDB code: 6EBE), the anion binding pocket is blue and specificity pocket is red Fig.4c.

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 3, 2020 (pp. 67-89)
www.abjournals.org



a. Hydrogen Bonds

b. Steric interactions & interaction Overlay

c. binding pocket

Fig.4. the best score docking solution of compound 44561182 with the selected crystal structure of 6ebe. Amino acids in the active site are presented in lines and ligand is presented in lines with fix color blue lines represent the hydrogen bonds between the ligand and the active site of 6ebe.

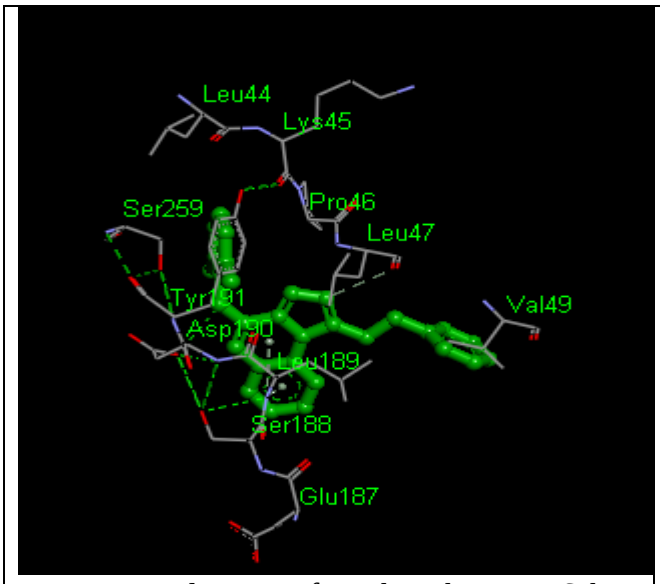
The basic of iGEMDOCK is the GEMDOCK, which is a robust and well-developed tool. By means of iGEMDOCK, the predicted poses produced from the GEMDOCK can be directly visualized by a molecular visualization tool and analyzed by post-analysis tools. iGEMDOCK offers the post-analysis tools by using k-means and hierarchical clustering methods based on the docked poses (i.e. protein-ligand interactions) and compound properties (i.e. atomic compositions). We validated the protein-ligand docking accuracy and screening accuracies of

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 3, 2020 (pp. 67-89)
www.abjournals.org

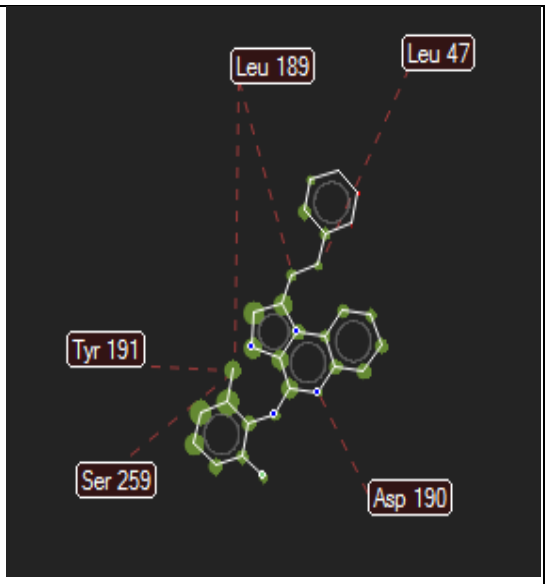
iGEMDOCK by using a training and test set with 24 protein-ligand complexes. Based upon the least binding affinity and other parameters, the best pose was selected and the docked structures were visualized in Discovery Studio and Melogro molecular viewer software (Fig. 5a and 5b) for detailed residue-ligand interactions. The best inhibitor pose energy was obtained -93.6494kcal/mol with a hydrogen bond of -7kcal/mol between the inhibitor and enzyme. Docking of PubChem_CID 44561182 with protein represents that ligands were effectively bound by interacting with Leu189, Leu47, Tyr191, Ser259, and Asp19 which are the residues in the active site of protein (PDB code: 6EBE). The detailed docking results are shown in Table 7.

Table 7. iGEMDOCK Result: Computationally predicted potencies of 11 compounds screened

Cpd CID (Ligand)	Energy	VDW	HBond
25253256	-87.5408	-87.5408	0
25253257	-89.5498	-74.7659	-14.7839
44224743	-85.6956	-85.6956	0
44224970	-81.7628	-81.7628	0
10244054	-80.5062	-80.5062	0
25254189	-86.7377	-83.2393	-3.49839
25254190	-92.7138	-80.7486	-11.9652
44561144	-84.595	-78.595	-6
44561145	-84.6749	-73.3491	-11.3258
44561181	-89.203	-82.203	-7
44561182	-93.6494	-86.6494	-7



a. Visualization of Docking between 6ebe and compound CID 44561182.



b. Visualization of Docking between 6ebe and compound CID 44561182.

Fig. 5. Visualization of Docking between 6ebe and compound CID 44561182 using Discovery Studio and Melogro molecular viewer respectively.

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 3, 2020 (pp. 67-89)
www.abjournals.org

PyRx is an open-source package to accomplish virtual screening. It is a combination of several software such as AutoDock Vina, AutoDock 4.2, Mayavi, Open Babel, etc. PyRx uses Vina and AutoDock 4.2 as docking software.

Table 8. PyRx AutoDock Vina Result of MVD and iGEMDOCK selected compounds (kcal/mol)

PuBChem_CID (Ligand)	Result	PuBChem_CID (Ligand)	Result
25254190	-8.1	-	-
44561182	-8.6	10244054	-6.7
44561144	-7.1	44561181	-7.8
25254189	-8.2	44224970	-9.0
25253257	-7.5	44224743	-7.7
44561145	-6.8	25253256	-7.9

The results of PyRx docking experiments of tumor inhibitors using AutoDock Vina are summarized in Table 8. For each docking experiment, the lowest energy docked conformer was selected from 10 runs. Ligand 44224970 (Fig. 6) showed better inhibition potential than ligand 44561182 (Fig. 7), a potent tumor inhibitor, with binding energy -9.0 kcal/mole (Table 8). Modeling and docking analysis revealed the nature of the active site and some key interactions that enabled the binding of inhibitor ligand 44561182 to the active site. The numbers of the binding modes (compound CID 44561182) were thirteen (n=13) while compound CID 44224970 was twelve (n = 12), respectively. Based on the present molecular docking study, ligand 44561182 appeared as a strong binder to the enzyme (6ebe) active site than the ligand 44224970 and the interacting numbers of amino acids and conventional hydrogen bonds might be critical factors for regulating target protein activity. These data also suggest that computer-aided drug design process using Melogro Virtual Docker, iGemDock and PyRx tools is highly reliable and can be a good example for identifying the action mechanism between the 6ebe (enzyme) and its interacting ligands.

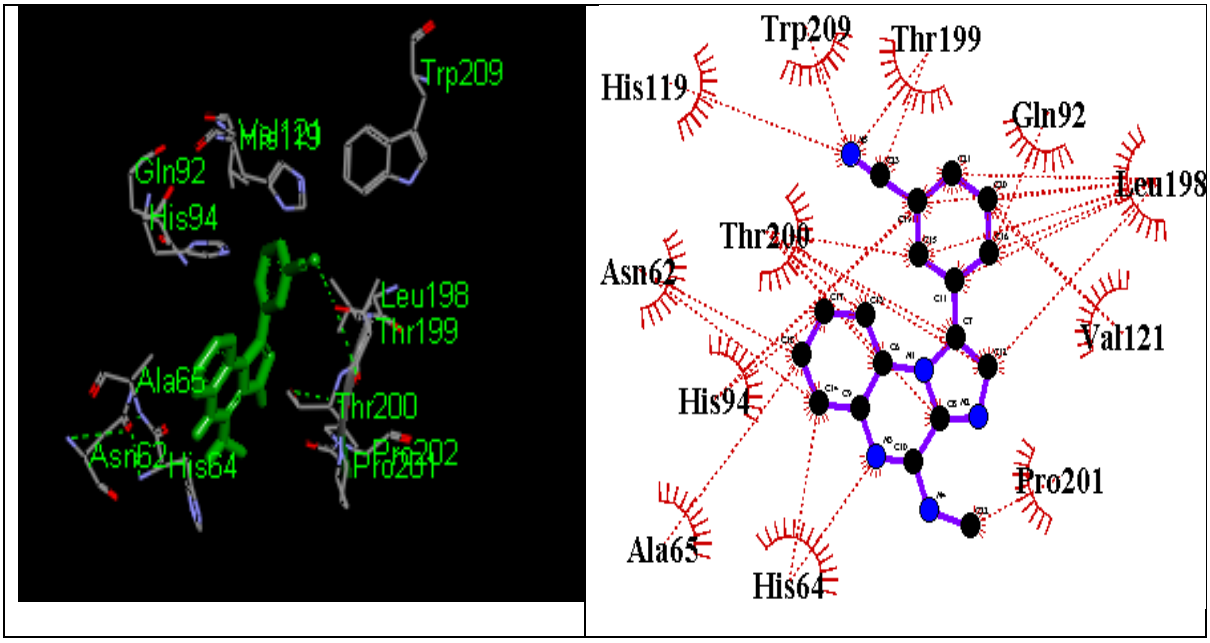


Fig. 6: Binding interaction of compound CID 44561182 with antitumor (PDB ID-6EBE)

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 3, 2020 (pp. 67-89)
www.abjournals.org

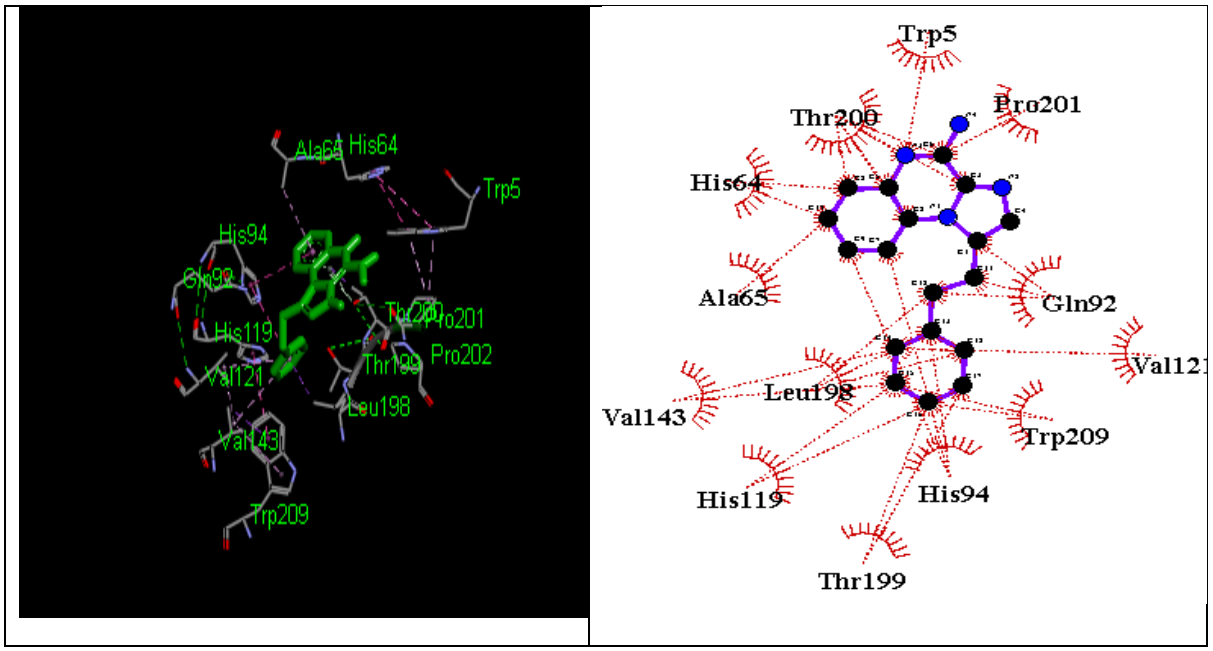


Fig. 7: Binding interaction of top ten least energetic molecules (Ligand 44224970) with anti-tumor (PDB ID-6EBE)

CONCLUSION

Present work aimed to develop a 2D QSAR model for imidazo[1,2-a] quinoxaline to identify the novel molecules that have well in silico predictions (QSAR and docking). GFA method was used for variable selection and model building of imidazo[1,2-a] quinoxaline derivative. The successful model was built with descriptors AlogP, AMR, SC-3, and maxaasC having SEE: 0.51748, r^2:0.73038, r^2 adjusted: 0.63234, F: 7.44967 (DF: 4, 11), Q2:0.51664. The cross-validation method, Y-randomization technique, applicability domain, and external validation indicated that the model is statistically significant and has good internal and external predictability. The reliability of these predictions for screened molecules, which were not part of the QSAR training set, was also assessed by domain of applicability (Leverage and Euclidean based applicability domain). Most of the molecules were followed the same domain as a training set; hence, predictions were reliable. The 2D descriptors were related to topology and 3D arrangement of atoms in molecules that can be used to design new inhibitors with good potency.

The docking results of screened molecules ligand 44561182 were exhibited consistency in terms of position into the active site and binding modes in all docking runs. As concluded from these studies, imidazo[1,2-a] quinoxaline using known active enabled to identify new hits with good in silico activities and binding poses, but in vitro assay to verify their experimental activity need to be done.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper. Also, they declare that this paper or part of it has not been published elsewhere.

Acknowledgments

The authors gratefully acknowledge the Department of Chemistry, Ahmadu Bello University, Zaria (Samaru, Zaria-Nigeria); for computational studies.

REFERENCES

[1] Abechi, S. E and Edache, E. I. IJPSR. **2016**, 7(04), 204-220.
[2] Agarwal, S.; Mehrotra, R. JSM Chem. **2016**, 4(2), 1024.
[3] Ansari, K. I; Kasiri, S; Grant, J. D; Mandal, S. S. J. of Biomol. Screening, **2010**, 16, 26–35.
[4] Chen, J.; Shen, Y.; Liao, S.; Chen, L.; Zheng, K. Int. J. Quantum Chem. **2007**, 107, 1468–1478.
[5] Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R. J. Med. Chem. **2014**, 57, 4977–5010.
[6] Clark, D. E. Expert Opin Drug Discov. **2006**, 1(2), 103–110.
[7] Cooper, A.; Potter, T.; Luker, T. Drug Metab. Dispos. **2010**, 38, 2218–2225.
[8] Danishuddin and Khan, A.U. Drug Discovery Today, **2016**, 21.
[9] Deleuze-Masquefa, C.; Moarbess, G.; Khier, S.; David, N.; Gayraud-Paniagua, S.; Bressolle, F.; Pinguet, F.; Bonnet, P. A. Eur. J. Med. Chem. **2009**, 9, 3406-11. doi: 10.1016/j.ejmech.2009.02.007.
[10] De-Melo, E. B. Scientia Pharmaceutica. **2012**, 80, 265–281 doi:10.3797/scipharm.1112-21.
[11] Dimitrov, S.; Dimitrova, G.; Pavlov, T.; Dimitrova, N.; Patlewicz, G.; Niemela, J.; Mekenyan, O. A. J. Chem. Infor. Modeling. **2005**, 45, 839–49.
[12] Golbraikh, A.; Tropsha, A. J. Comput. Aided Mol. Des. **2002**, 16, 357–369.
[13] Gramatica, P. QSAR Comb Chem. **2007**, 26, 694–701. http://dx.doi.org/10.1002/qsar.200610151.
[14] Guedes, I. A.; de Magalhães, C. S.; Dardenne, L. E. Biophysical Reviews. **2014**, 6, 75-87.
[15] Hall, L. H.; Mohney, B.; Kier, L. B. J Chem Inf Comput Sci. **1991**, 31, 76-82.
[16] Hassan, G.; Dashtbozorgi, Z.; William E. A. Jr. Eur. J. Pharmaceutical Sci. **2012**, 47(2), 421-429.
[17] Holland, J. H. Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.
[18] Ibezim, E.; Duchowicz, P. R.; Ortiz, E.V.; Castro, E. A. Chemometrics and Intelligent Laboratory Systems. **2012**, 110, 81–88.
[19] Jaiswal, M.; Khadikar, P.V.; Scozzafava, A.; Supuran, C. T. Bioorg. Med. Chem. Lett. **2004**, 14, 3283–3290.
[20] Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. Alternative Laboratory Animals. **2005**, 33, 445–459.

[21] Kar S, and Roy K, Predictive toxicology using QSAR: a perspective, J. Indian Chem. Soc., **2010**, 87, 1455–1515.
[22] Kennard, R.W.; Stone L. A. Technometrics, **1969**, 11, 137– 148.
[23] Khan, N. U. H.; Pandya, N.; Kureshy, R. I. Mol. and Biomol. Spectroscopy. **2009**, 74(1), 113–119.
[24] Kier, L. B.; Hall, L. H. Molecular Structure Descriptors: The Electrotopological State. Academic Press: New York; 1999.
[25] Kubinyi, H. Quant Struct Act Relat. **1994**, 13, 285-294.
[26] Kubinyi, H. Quant Struct Act Relat. **1994**, 13, 393-401.
[27] Kumar, P.; Agarwal, J. C.; Nath, C.; Bhargava, K. P.; Shanker, K. Pharmazie, **1981**, 36, 780.
[28] Kusumaningrum, S.; Budianto, E.; Kosela, S.; Sumaryono, W.; Juniarti, F. JAPS. **2014**, 4 (11), 047-053, DOI: 10.7324/JAPS.2014.4119
[29] Liao, C.; Sitzman, M.; Pugliese, A.; Nicklaus, M. C. Future Med. Chem. **2011**, 3, 1057–1085.
[30] Meyyanathan, M.; Chandrasekhar, K. Indian Drugs, **2006,** 43(6).
[31] Moarbess, G.; Deleuze-Masquefa, C.; Bonnard, V.; Gayraud-Paniagua, S.; Vidal, J. R.; Bressolle, F.; Pinguet, F.; Bonnet, P. A. Bioorg Med Chem. **2008**, 16(13), 6601-10. doi: 10.1016/j.bmc.2008.05.022.
[32] National Center for Biotechnology Information. PubChem Database. Source=ChEMBL, AID=1315713, https://pubchem.ncbi.nlm.nih.gov/bioassay/433907 (accessed on June 30, 2019).
[33] Nocentini, A.; Trallori, E.; Singh, S.; Lomelino, C. L.; Bartolucci, G.; Di Cesare Mannelli, L.; Ghelardini, C.; McKenna, R.; Gratteri, P.; Supuran, C. T. J. Med. Chem. **2018**, 61, 10860-10874.
[34] Noolvi, M. N.; Patel, H. M.; Bhardwaj, B. Digest J. Nanomater. Biostruct. **2010**, 5, 387–401.
[35] Noolvi, M. N.; Patel, H. M.; Bhardwaj, V.; Chauhan, A. Europ. J. Med. Chem. **2011**, 46, 2327e2346
[36] Pereira, F.; Latino D. A. R. S.; Gaudêncio, S. P. Molecules **2015**, 20, 4848-4873; doi:10.3390/molecules20034848.
[37] Pourbasheer, E.; Riahi, S.; Ganjali, M. R.; Norouzi, P. Eur. J. Med. Chem. **2009**, 44, 5023–5028.
[38] Rajasekaran, S; Gopalkrishna, R. Der Pharmacia Lett. **2012**, 4(1), 349-353.
[39] Rekha, S. H.; Basavaraj, R. P.; Dayananda, S. B.; Ramesh, S. V.; Kalagouda B. G. Der Pharma Chemica. **2010**; 2(2), 116-128.
[40] Riahi, S.; Pourbasheer, E.; Ganjali, M. R.; Norouzi, P. J. Hazard. Mater. **2009**, 166, 853-859
[41] Roy, K. Expert Opin. Drug Discov. **2007**, 2, 1567–1577.
[42] Roy, K.; Kar, S.; Ambure, P. Chemometrics and Intelligent Laboratory Systems, **2015**, 145, 22–29.
[43] Shapiro, S.; Guggenheim, B. Quant. Struct.-Act. Relat. **1998**, 17, 327–337.
[44] Singh, J.; Singh, S.; Thakur, S.; Lakhwani, M.; Khadikar, P. V.; Agrawal, V. K.; Supuran, C.T. ARKIVOC **2006**, (xiv), 103-118.
[45] Snarey, M.; Terrett, N. K.; Willet, P.; Wilton, D. J. J. Mol. Graphics Modell. **1997**, 15, 372– 385.
[46] Somanathan, R. R.; Hellberg, L. H. Synth. Commum, **1998**, 28(1), 2077-2086.
[47] Tropsha, A.; Gramatica, P.; Gombar, V. K. QSAR Comb Sci. **2003**, 22, 69–77.

[48] Vaidya, A.; Jain, A.K.; Kumar, B.R.P.; Sastry, G.N.; Kashaw, S.K.; Ram Kishore Agrawal, R.K. Arabian J. Chem. **2017**, 10, S3936–S3946. http://dx.doi.org/10.1016/j.arabjc.2014.05.034.
[49] Vaidya, A.; Jain, S.; Jain, S.; Jain, A. K.; Agrawal, R. K. Journal of Pharmaceutical Sciences and Pharmacology, **2014**, 1, 219–232.
[50] Viswanathan, V. M.; Goshe, A. K.; Revanker, G. R.; Robbins, R. K. J. Chem. Inf. Comput. Sci. **1989**, 29, 163-172.
[51] Walters, W. P.; Stahl, M. T.; Murcko, M. A. Drug.Discov. Today. **1998**, 3(4), 160–178.
[52] Wang, J.; Chu, P.; Chen, C.; Lin, J. Nucleic Acids Research, **2012,** 40.
[53] Wavefunction, (2004) Inc. Spantan'04, Irvin, California, USA.
[54] Wildman, S. A.; Crippen, G. M. J Chem Inf Comput Sci. **1999**, 39, 868-873.
[55] Yap, C.W. J. Computational Chem. **2011**, 37(7), 1466-1474