Moroccan Journal of Chemistry
ISSN: 2351-812X
http://revues.imist.ma/?journal=morjchem&page=login
Umar & al. / Mor. J. Chem. 9 N°1 (2021) 260-273

In silico Studies of some potential anti-cancer agents on M19-MEL cell line

Abdullahi Bello Umar*, Adamu Uzairu, Sani Uba and Gideon Adamu Shallangwa

Department of Chemistry, Faculty of Physical Sciences, Ahmad Bello University, Zaria, Nigeria

** Corresponding author: abdallahbum@yahoo.com*
Received 09 May 2020,
Revised 03 Nov 2020,
Accepted 01 May 2021

Abstract

The resistance of melanoma cancer cells to the known treatments has become a barrier to the success of chemotherapy. In this research, a quantitative evaluation of the structure-activity relationship (QSAR) was carried out on 57 anti-cancer compounds and some selected potent compounds were screened through Lipinski's rule and docked. Genetic function algorithm (GFA) was adopted in variables selection and Multiple linear regression (MLR) was used to generate the model. The built QSAR model showed good statistical parameters ((R^2 (0.904), $R^2_{adjusted}$ (0.885), Q^2_{cv} (0.873) and R^2_{pred} (0.779)). The cR^2_P for Y-randomization is 0.749 and the applicability domain was also determined. The predictive ability of the model was found to be satisfactory and could be used to predict the anti-cancer activity of compounds on M19 MEL cell line. 4 most potent compounds were selected among the data set and screened through Lipinski's rule of five filters for oral bioavailability, ADMET risk filter for a drug like features. Later, V600E-BRAF, a known melanoma cancer target was used for docking. Based on the interaction energy and types of interactions involved, the selected compounds were identified as the best hits against V600E-BRAF. This research would help in the lead identification and design of novel drugs.

Keywords: *Melanoma, M19 cell line, QSAR, Docking, V600E-BRAF*

1. Introduction

Melanoma is one of the most destructive diseases and the prognosis of metastatic melanoma is very poor [1]. The average survival of patients with advanced-stage melanoma is roughly eight-nine months, and the five-year survival rate is less than ten percent (10%) [2]. Chemotherapy is a conventional treatment for patients with melanoma. Dacarbazine (DTIC) is the initial approved chemotherapeutic drug by the US (FDA) for the clinical treatment of melanoma. Nevertheless, the total response rate is less than ten percent (10%) and the median overall survival is less than eight months in patients with melanoma [3]. Therefore, because of melanoma complications to chemotherapy, searching for other drugs to improve the treatment of melanoma becomes necessary. Several in silico methods have shown the potential for filtering chemical databases against the biological targets of interest for the discovery of novel potent leads [4]. One of the most used methods is ligand-based virtual screening which has become popular due to its ability to screen a large number of molecules quickly from the available chemical databases [5]. QSAR modeling is an important method used in the drug discovery arena which correlates chemical structures of molecules with their biological activities [6]. It has become very crucial in the molecular interpretation of biological and physicochemical properties [7, 8]. This method is the most essential tool adopted for the determination of assorted properties like stability, carcinogenicity, toxicity, ADME, retention time and other physicochemical properties apart from biological activity prediction [9-13]. In addition to predicting safety and toxicity, these analyses can foretell interactions between compounds and their receptors, saving money and time during the process of drug filtering. In this research, a QSAR analysis was carried out on a series of compounds collected from the National Cancer Institute (NCI) database against M19-MEL human melanoma cell line, to extract out the key structural features regulating the anti-cancer activity of the compounds. This QSAR method gave a different influence and served as a helpful predictive tool, predominantly in the design of pharmaceuticals [14]. Additionally, the predicted compounds were screened through Lipinski's rule of five for oral bioavailability evaluation and ADMET risk evaluation. Further, the molecular docking simulation studies were conducted on the BRAF receptor as it is associated with nearly sixty-six percent of melanomas [15].

2. Materials and Methods

2.1. QSAR model development

In this research, fifty-seven (57) anti-cancer compounds were collected from the NCI database (https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-60+Growth+Inhibition+Data). The NSC number, chemical name, and experimental activities expressed as pGI_{50} are presented in Table 1. The 2D chemical structures of the compounds were converted to 3D chemical structures and optimize their geometry by using Density Functional Theory (DFT) method at the B3LYP level of theory and 6-31G** basis set with Spartan 14 Version 1.1.4 software package from Wavefunction Inc. 1D, 2D, and 3D descriptors generation was done using paDEL descriptor tool kit [16, 17]. Material Studio version 8.0 Software from BIOVIA-Accelrys and Microsoft excel 2013 was adopted for model building and validation. Descriptors are normally measured in various units and data modeling usually favors data with larger absolute value. To reduce this bias and assign an equal weight, value, and possibility to each descriptor to be part of the final equation, any descriptors variables with constant or near-constant values were deleted and features containing missing values were also deleted because it is likely that for virtual screening purposes models [18, 19] developed with such features will not be suitable for a part of the [20] new compounds. Descriptors highly correlated were also deleted, using a threshold value of the coefficient correlation of 0.80 to minimize the redundancy in the data. Data division has a great influence on the final selected model, and the Kennard-Stone approach was employed for this study. This method is used for diversity sampling of experimental datasets for training and test sets selection. The Kennard-Stone algorithm was applied to split the 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 [21]. The data set of 57 anti-cancer compounds were split into training and test sets. The training set consisted of about 70% of the whole data (39 molecules) and the test set consisted of the remaining 30% of the whole data (18 molecules) [21]. Variable selection for multiple linear regressions (MLR) regression was performed by the genetic function algorithm (GFA) method using Material Studio 8.0 Software. GFA method combined with MLR was employed to generate a QSAR model for the dataset. The best model QSAR was chosen based on the statistical validation parameters such as the correlation coefficient of determination (R^2), Adjusted correlation coefficient R^2 (R^2_{adj}), Cross-validated coefficient of determination (Q^2_{CV}) and correlation coefficient for an external prediction set (R^2_{pred}) all are represented in equations (1-4):

$$R^2 = 1 - \frac{\sum(Y_{exp} - Y_{pred})^2}{\sum(Y_{exp} - Y_{mtraining})^2} \quad (1)$$

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

$$Q^2_{CV} = 1 - \frac{\sum(Y_{pred} - Y_{exp})^2}{\sum(Y_{exp} - Y_{mtraining})^2} \quad (3)$$

$$R^2_{pred} = 1 - \frac{\sum(Y_{pred} - Y_{exp})^2}{\sum(Y_{exp} - Y_{mntrng})^2} \quad (4)$$

Where P represents the number (total) of descriptors in the QSAR model and N is the sample size. Y_{exp}; Y_{pred}; $Y_{mtraining}$ is the activity (experimental), the activity (predicted) and the mean activity (experimental) of the compounds in the training set[20]. Further, to assess the robustness of the built model, the Y-randomization test was applied to the training data set as suggested by Tropsha et al. [20]. The dependent variable vector (activity data) was randomly shuffled and a new QSAR model was developed using the original independent variable matrix. For the built QSAR model to robust and reliable, the model is expected to have a low R^2 and Q^2 values for several trials. The coefficient of determination cR^2_p for Y-randomization is another parameter calculated which should be greater than 0.5 for passing this test as in equation (5):

$$cR^2_p = R \times [R^2 - R^2_r]^2 \quad (5)$$

cR^2_p is Coefficient of determination for Y-randomization, R is the coefficient of determination for Y-randomization and R_r is average 'R' of random models.

2.2 Lipinski's Rule of Five and ADME properties

Lipinski's rule of five [22] was adopted as a primary screening step for oral bioavailability using SwissADME a free web tool used in evaluating ADME properties and drug-likeness of small molecules. The violating query set predicted active compounds by more than two properties were removed. Later, secondary screening was performed based on the drug-like score (The absorption, solubility, blood-brain barrier (BBB) transport, oral bioavailability, and distribution). The ADMET parameters were also studied so that it can be optimized during the drug designing process.

2.3 Molecular Docking (MD) Simulation

MD simulation was carried out using Molegro Virtual Docker 6.01 (MVD) [23], Some compounds among the dataset that showed higher potency and have low residual values from the QSAR study were docked to the 3D structure of the V600E-BRAF (PDB-code: 3OG7) [24-26]. The selected compounds were docked in the active site of the protein and the best pose of each compound was selected for subsequent ligand-protein interaction energy analysis.

3. Results and Discussion

3.1 QSAR model and Validation

Using the descriptors selected by the GFA method, we have developed a prediction QSAR model by MLR method based on the training set for the prediction of biological activities (pGI_{50}) of the compounds on M19-MEL melanoma cell line. The built QSAR model and the corresponding statistics are shown in equation (6):

$$\boldsymbol{pGI_{50(M19-MEL)}} = 0.756889758\ (\mathbf{SM1\_Dzv}) - 0.000015339\ (\mathbf{VR2\_Dze}) + 5.552189891\ (\mathbf{SpMin8\_Bhe}) - 5.429923889(\mathbf{SpMin7\_Bhi}) - 0.050703095(\mathbf{MIC1}) + 3.034720419(\mathbf{piPC2}) - 2.57198 \quad (6)$$

$$N_{training} = 39\ , R^2 = 0.904, R^2_{adjusted.} = 0.885, Q^2_{cv.} = 0.873, N_{test} = 18, R^2_{pred} = 0.779$$

Table 1. NSC-numbers, Names and pGI_{50} activities of the studied compounds with residuals

S/N	NSC	Chemical name	pGI_{50} (Exp.)	pGI_{50} (Pred.)	Residuals
1t	267,469	Deoxydoxorubicin	7.472	7.131	-0.341
2	269,148	Menogaril	5.949	6.023	-0.074
3	268,242	N,N-Dibenzyldaunorubicin hydrochloride	8.000	7.491	0.509
4	126,771	Dichloroallyl lawsone	5.023	4.646	0.377
5	140,377	Arnebin 1	6.207	6.170	0.037
6	212,509	4beta-Hydroxywithanolide	6.771	6.380	0.391
7	215,139	Bikaverin	6.042	6.411	-0.369
8	143,095	Pyrozofurin	4.880	4.609	0.271
9	629,971	9-aminocamptothecin (r,s)	7.231	6.928	0.303
10	606,173	11-Hydroxymethyl-20(RS)-camptothecin	5.431	6.699	-1.268
11	364,830	Camptothecin,n-diethyl) glycinate	7.045	7.489	-0.444
12	94,600	Camptothecin	7.273	6.977	0.296
13	606,985	Campothecin analog	6.911	6.720	0.191
14	606,499	Camptothecin butylglycinate ester hydrochloride	6.201	6.412	-0.211
15	606,497	Camptothecinethylglycinate esterhydrochloride	6.404	6.492	-0.088
16	3,088	Chlorambucil	4.653	4.730	-0.077
17	338,947	Clomesone	3.525	3.390	0.135
18	95,678	Picolinaldehyde	5.458	5.566	-0.108
19	264,880	Dihydro-5-azacytidine	4.153	3.885	0.268
20	71,851	Alpha.-thiodeoxyguanosine	2.799	2.911	-0.112
21t	132,483	L-Aspartic acid	4.863	5.192	0.329
22t	308,847	Amonafide	5.383	5.888	0.505
23	182,986	Diaziquone	5.262	5.619	-0.357
24	139,105	Triazinate	6.186	6.180	0.006
25	409,962	Carmustine	4.103	4.451	-0.348
26t	750	Busulfan	3.637	3.777	0.140
27t	95,382	Camptothecin, acetate	5.605	5.606	0.001
28t	107,124	10-hydroxycamptothecin	7.348	8.026	0.678
29t	79,037	Lomustine	4.563	4.105	-0.458
30	132,313	Dianhydrodulcitol	4.009	4.000	0.009
31	73,754	Fluorodopan	3.629	3.800	-0.171
32	148,958	Uracil	3.235	3.893	-0.658
33	1895	Guanazole	2.007	1.863	0.144
34	329,680	Hepsulfam	3.417	3.472	-0.055
35	32,065	Hydroxyurea	2.862	2.687	0.175
36	153,353	Alanosine monosodium salt	4.854	4.752	0.102
37t	249,992	Amsacrine	5.812	5.922	0.110

The correlation coefficient R^2 between the experimental and predicted pGI_{50} activity of modeling (training) set was 0.904, while that of predicting (test) set was 0.779, which proved that the built QSAR model was acceptable. The experimental and predicted pGI_{50} values of 57 compounds by the developed QSAR model have been given in Table 1. The plot of predicted pGI_{50} value versus experimental pGI_{50} value for both the training and test sets is presented in Figure 1. Generally, a good and acceptable QSAR model must obey the following criteria: regression-coefficient (R^2) and adjusted regression-coefficient (R^2adj) values close to one. The Cross validated regression-coefficient (Q^2_{cv})> 0.5, R^2 - $Q^2_{cv} \leq 0.3$, $R^2_{pred.} \geq 0.6$, and $N_{test} \geq 5$ [20, 27, 28]. The generated QSAR model satisfied the criteria and therefore acceptable statistically. We can, therefore, conclude that the developed model will correctly predict the anti-cancer pGI_{50} activity of a given compound. Additionally, to further assess the robustness of the build model, the Y-randomization test was employed. The dependent variable vector (inhibitory activity) was randomly shuffled and a new QSAR model was developed using the original independent variable matrix. As was expected the new QSAR models (after several repetitions) have low R^2 and Q^2 values and also, the cR^2_p value was greater than 0.5 as presented in Table 2. This test affirms that the proposed model is powerful and not inferred by chance.

Table 1. (Continued)

S/N	NSC	Chemical name	pGI_{50} (Exp.)	pGI_{50} (Pred.)	Residual
38	740	Methotrexate	6.419	5.402	1.017
39t	95,441	Semustine	4.275	4.769	0.494
40t	26,980	Mitomycin C	6.140	6.098	-0.042
41t	353,451	Mitozolomide	4.000	4.591	0.591
42	268,242	N,N-Dibenzyldaunorubicin hydrochloride	5.925	7.491	-1.566
43t	95,466	Urea,	3.576	3.890	0.314
44t	25,154	Pipobroman	4.040	3.161	-0.879
45	56,410	Profiromycin	5.499	5.527	-0.028
46t	366,140	Pyrazoloacridine mesylate	6.375	6.848	0.473
47	51,143	Pyrazoloimidazole	2.163	2.812	-0.649
48	172,112	Spiromustine	3.699	3.273	0.426
49t	296,934	Teroxirone	4.383	4.837	0.454
50t	363,812	5-((4-chlorobenzyl)thio)-3-(trifluoromethyl)-1H-1,2,4-triazole	6.604	5.709	-0.895
51	361,792	3-demethylthiocolchicine;	8.000	7.225	0.775
52	6396	Thiotepa	4.590	4.685	-0.095
53	9,706	Triethylenemelamine	4.901	4.864	0.037
54t	83,265	Tritylcysteine	5.478	7.142	1.664
55	67,574	Vincristine sulfate	6.819	6.808	0.011
56	757	Colchicine	8.240	7.040	1.200
57t	33,410	N-Benzoyl-deacetylcolchicine	8.000	7.353	-0.647

‘t’ indicate predicting (test) sets ‘*’ indicate compounds outside the defined AD of the model

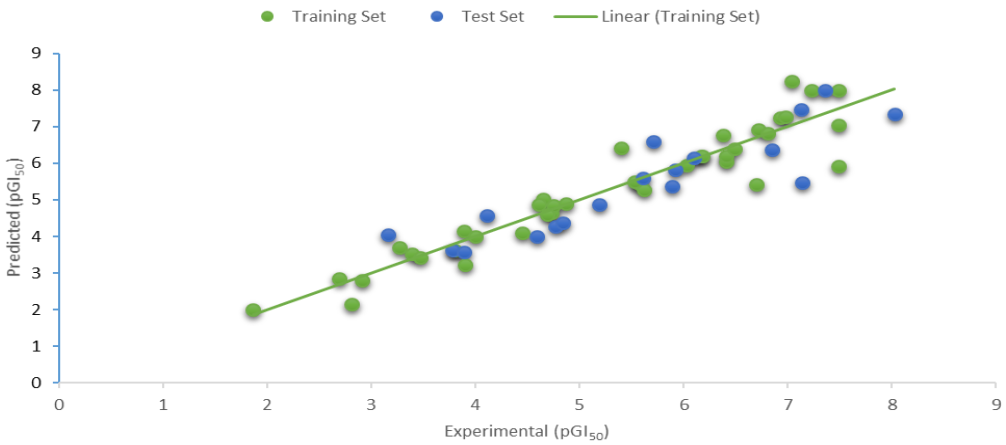

Figure 1. Predicted versus experimental pGI_{50} values for both the training and testing sets

Table 2. R^2 and Q^2 values after several Y-randomization test

Model	R	R^2	Q^2
Original	0.900201	0.810363	0.746851
Random 1	0.335723	0.11271	-0.14688
Random 2	0.484376	0.23462	-0.18531
Random 3	0.337777	0.114093	-0.46568
Random 4	0.467421	0.218482	-0.02434
Random 5	0.494957	0.244983	-0.05292
Random 6	0.21432	0.045933	-0.52451
Random 7	0.302498	0.091505	-0.2948
Random 8	0.252211	0.06361	-0.4286
Random 9	0.177645	0.031558	-0.66118
Random 10	0.353479	0.124947	-0.36429
Random Models Parameters			
Average r :	0.342041		
Average r^2 :	0.128244		
Average Q^2 :	-0.31485		
cR^2_P:	0.749588		

Molecular descriptors imply the physicochemical and structural information in the form of numbers, each descriptor represents specific information that can be implored to enhance the overall biological activity of a molecule. Therefore, concise descriptions of the selected descriptors are presented in Table 3. The contribution and significance of every descriptor in the built model were evaluated by the computation of the mean effect (ME) value [29] of every descriptor by using equation (7). The values for the ME are shown in Table 3:

$$ME_j = \frac{\beta_j \sum_{i=1}^{i=n} d_{ij}}{\sum_j^m \beta_j \sum_i^n d_{ij}} \qquad (7)$$

Where: ME_j is the mean effect (ME) of the descriptor *j*, β_j represents the coefficient of the descriptor *j*, d_{ij} represents the value of the selected descriptors of each compound and *m* is the total number of the descriptors in the generated model.

Table 3. Specification of entered descriptors and their mean effect (ME)

Descriptor	Description	ME
SM1_Dzv	Spectral moment of order 1 from Barysz matrix / weighted by van der Waals volumes	-0.2492
VR2_Dze	Normalized Randic-like eigenvector-based index from Barysz matrix / weighted by Sanderson electronegativities	-0.0385
SpMin8_Bhe	Smallest absolute eigenvalue of Burden modified matrix - n 8 / weighted by relative Sanderson electronegativities	0.5701
SpMin7_Bhi	Smallest absolute eigenvalue of Burden modified matrix - n 7 / weighted by relative first ionization potential	-0.5594
MIC1	Modified information content index (neighborhood symmetry of 1-order)	-0.2208
piPC2	Conventional bond order ID number of order 2 (ln(1+x)	1.4977

The ME value indicates the significance of a specific descriptor when compare to the other descriptors. Descriptors found to have high ME values influences anti-melanoma activity (pGI_{50}). The pGI_{50} changes with the ME values of a descriptor, as presented in Table 3. According to ME values, the selected descriptors were arranged in order about their contributions towards the overall pGI_{50} of the studied compounds, in following the increasing order of pGI_{50} of the compounds. Based on ME values, the descriptors were arranged in a sequential order about their contributions towards the overall pGI_{50} of the studied compounds an increasing sequence of pGI_{50} of the compounds.

piPC2 > SpMin8_Bhe > VR2_Dze > MIC1 > SM1_Dzv > SpMin7_Bhi

The applicability domain (AD) is an essential tool for the reliable application of QSAR models, and the characterization of interpolation space is important in defining the applicability domain [30]. In this study, we use the standardization approach introduced by Roy et al. to determine the application domain of our QSAR model [31], which employed the leverages and standardized residuals for each of the compounds of the data set. The leverage (h_i) value of each molecule in the data set is calculated using equation (8):

$$h_i = x_i(X^T X)^{-1} x_i^T \quad (8)$$

Where x: represents descriptor's vector of the selected molecule, X: refers to the descriptor matrix obtained from the training data set descriptor values. The warning leverage (h*) of the model is defined by equation (9)

$$h^* = \frac{3(p+1)}{N} \quad (9)$$

Where N represents the number of the training set compounds and p refers to the number of the descriptors in the built model.

This approach is used to define outliers (in case of the training set) and the compounds residing outside the applicability domain (in case of the test set) of the built QSAR model [32]. The defined AD was then visualized with Williams plot, the plot of the standardized residuals versus the leverages (h) of the molecules. A molecule with *h*i > *h** seriously affects the QSAR model performance and is considered to be an influential molecule and ±3 value range of standardized residuals is always accepted as a threshold value for affirming predictions of a molecule, any molecule with standardized residuals value not within ±3 range is considered to be an outlier. The warning leverage (*h**) for the built model was 0.538 according to equation 9. It can be observed from the Williams plot (Figure 2), no outlier molecule was identified for the data set. However, based on the leverages (*h*i > 0.420), two training set compounds (35 and 52) and two test set compounds (44 and 54) was found to be outside the defined AD (Figure 2) of the model, thus, were recognized as structurally influential compounds based on their high leverages.

Figure 2. The Williams plot, the plot of the standardized residuals versus the leverage value

3.2 Drug likeness properties

To further screen the compound in the data set, four most potent compounds (1, 3, 51 and 56) were selected from the data set due to their good pGI_{50} activity values compared to other compounds within the data set as presented in Table 1. The selected compounds were evaluated for ADME properties and drug-likeness using SwissADME a free web tool used in evaluating ADME properties and drug-likeness of small molecules [33]. The Lipinski's rule of five is useful at the pre-clinical stage of drug discovery which state that if any chemical violates more than 2 of these criteria (Molecular weight < 500, Number of hydrogen bond donors (nHBD) ≤ 5, Number of hydrogen bond acceptors (nHBA) ≤ 10, Calculated Log p ≤ 5 and Polar surface area (PSA) < 140 $Å^2$), the chemical is said to be impermeable or badly absorbed [33]. From the results presented in Table 4, it can be seen that none of the molecules violate more than 3 of the criteria stated by Lipinski's rule of five, it means there is a high tendency all of these molecules might be pharmacologically active. It can also be observed from Table 5 that none of the molecules possess the BBB permeant. Thus, these molecules are said to have good absorption, low toxicity level, orally bioavailable and permeable. The Bioavailability Radar gives an overview of the drug-likeness of a molecule (Figure 3). The region painted pink indicates the range for each property.

Table 4. Lipinski rule of 5 parameters for the studied inhibitors

S/N	MW	nHBD	nHBA	WLOGP	TPSA ($Å^2$)
1	527.52	5	11	0.71	185.84
3	707.76	4	11	4.14	163.06
51	401.48	2	5	2.96	110.16
56	399.44	1	6	2.55	83.09

Table 5. Predicted ADME parameters

S/N	GI absorption	BBB permeant	Pgp substrate	Bioavailability Score
1	Low	No	Yes	0.17
3	Low	No	Yes	0.17
51	High	No	Yes	0.55
56	High	No	Yes	0.55

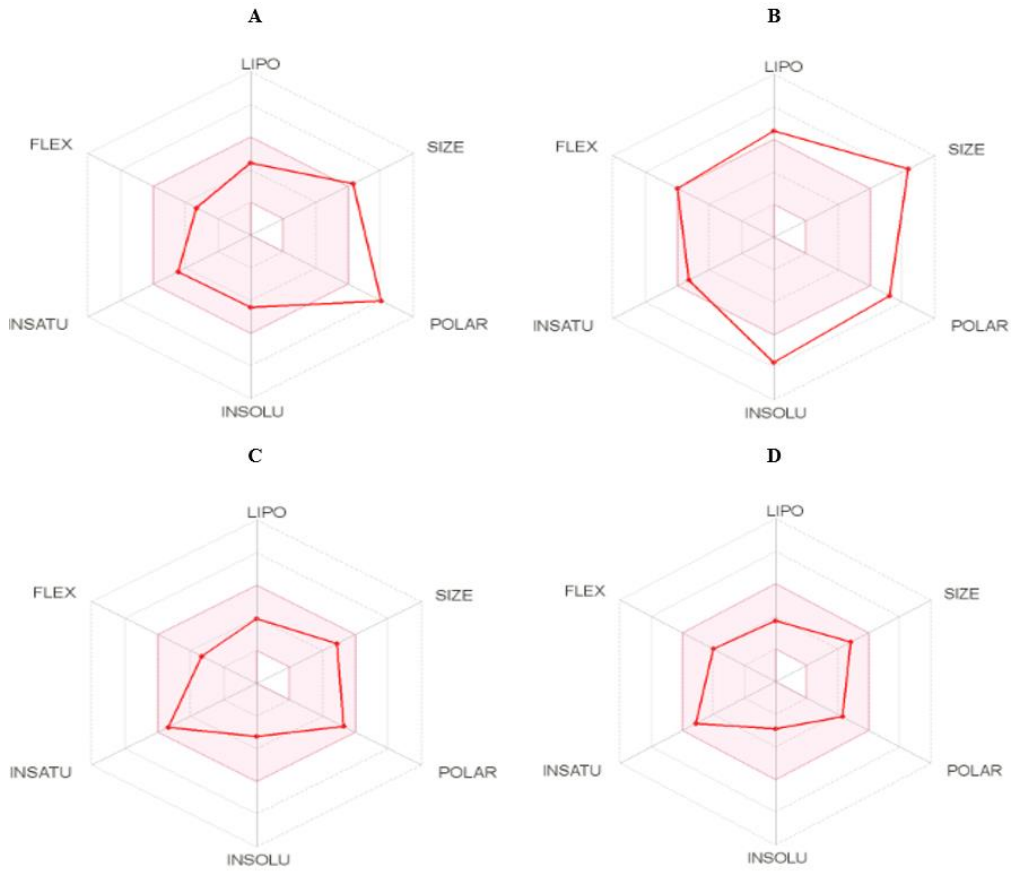

Figure 3. The Bioavailability Radar of compound (A) 1, (B) 3, (C) 51 and (D) 56.

3.3 Molecular docking (MD) simulation studies

After screening through Lipinski's rule of five for oral bioavailability and ADME compliance the selected compounds (1, 3, 51 and 56**)** were further evaluated for target binding affinity through molecular docking simulation studies. MD simulation study was conducted using the MVD to get insights into the possible binding mode of the selected compounds with the V600E-BRAF receptor (V600E-BRAF is a target expressed in melanomas). The detail docking results and types of interactions involved are presented in Table 6. The interaction energy of compound 1, 3, 51 and 56 are -170.804 kcalmol$^{-1,}$ -166.462 kcalmol^{-1}, -121.107 kcalmol^{-1} and -132.685 kcalmol^{-1}, respectively. These findings indicated that the selected compounds had favorable ligand-protein interaction energy at the binding cavity of V600E-BRAF.

The complex structure of the docked ligand-1 with the receptor is shown in Figure 4. This ligand shows the highest interaction energy among the selected compounds as presented in Table 6, which indicates the possibility of stable interactions between the ligand and the protein target. There were five conventional hydrogen bonding identified between this ligand and receptor; ASN580, ASN580, ASN581, ILE527 and CYS532 as presented in Table 6. One alkyl interaction with VAL471 is also present between this ligand and the receptor. The complex stability may be attributed with extra, Pi-alkyl interactions (VAL471, ILE463 and PHE468) and Pi-Pi interactions with TRP531 and PHE583 as shown in Figure 4.

Ligand 3 was docked into the V600E-BRAF-binding pocket. The calculated binding mode (Figure 5) showed that ligand 3 established carbon-hydrogen-bonding interactions with SER465, CYS53, GLY466 and SER53; and demonstrated a favorable pi-sigma interaction with ILE463, GLY464, and VAL471 which is important for selectivity [34]. Besides, it also formed pi-pi interactions with PHE583. The docked structure of ligand 51 is presented in Figure

6. This ligand forms two alkyl interactions with VAL471 and ALA481 residues of the receptor. Other identified interactions are Pi-sigma interaction with VAL471 and pi-alkyl interaction with (VAL471 and PHE583). Lastly, ligand 56 forms carbon-hydrogen bond with ASN581, alkyl and Pi-alkyl interaction with PHE583 and Pi-Pi interaction with PHE583.

Table 6. Molecular docking results of the studied inhibitors with V600E-BRAF (PDB ID: 3OG7)

Molecular System	E-inter (Kcalmol^{-1})	Hbond	Carbon-Hbond	Alkyl	Pi-Sigma	Pi-Pi	Pi-Alkyl
BRAF/1	-170.804	ASN580 ASN580 ASN581 ILE527 CYS532		VAL471		PHE583 TRP531 PHE583 TRP531	VAL471 ILE463 ILE463 PHE468
BRAF/3	-166.462		SER465 CYS53 GLY466 SER53		ILE463 GLY464 VAL471	PHE583	
BRAF/51	-121.107			VAL471 ALA481	VAL47		ALA481 PHE583
BRAF/56	-132.685		ASN581	ILE463		PHE583	PHE583

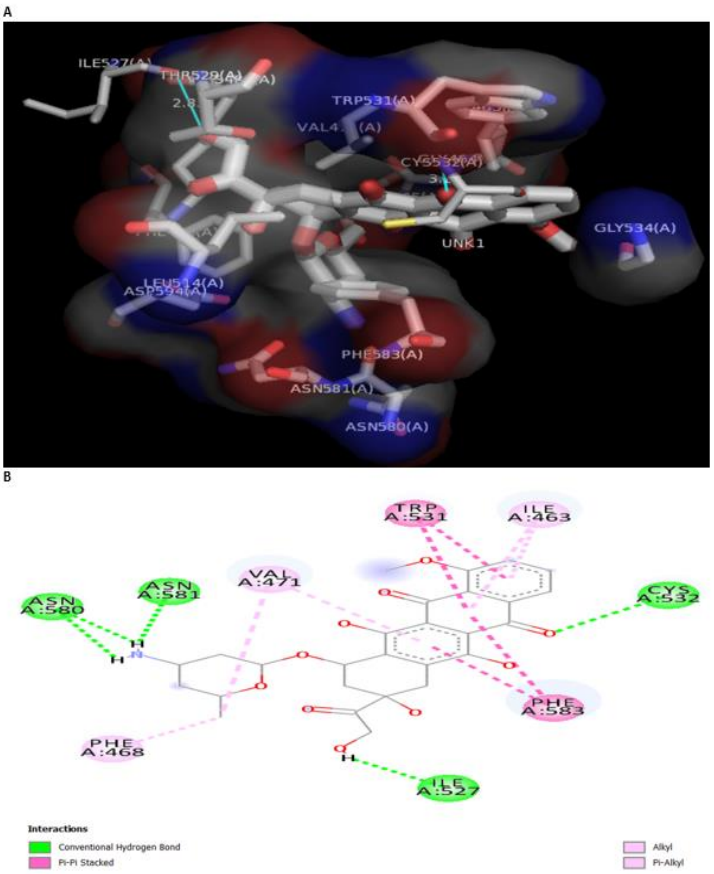

Figure 4. (A) 3D and (B) 2D view of V600E-BRAF /Compound 1 interactions

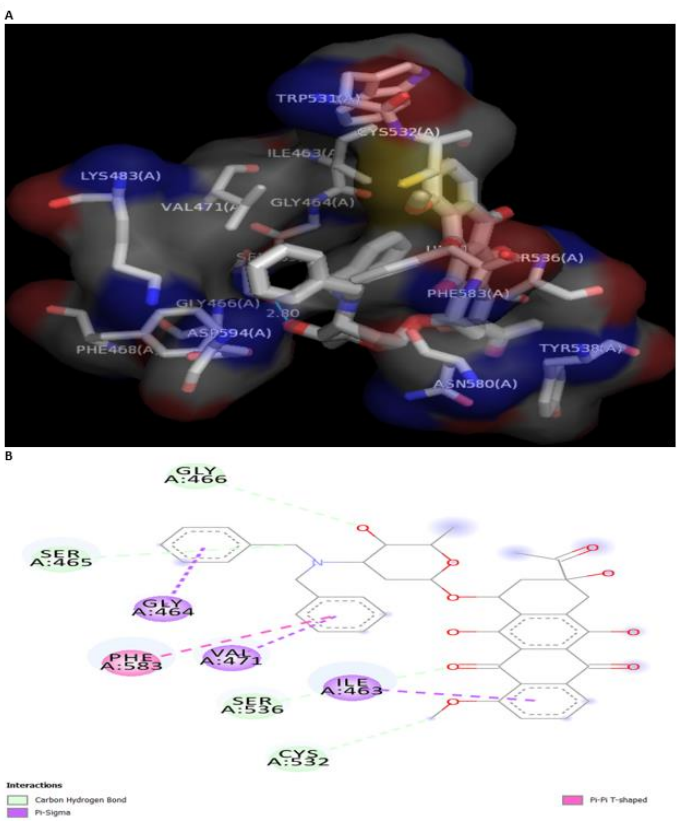

Figure 5. (A) 3D and (B) 2D view of V600E-BRAF /Compound 3 interactions

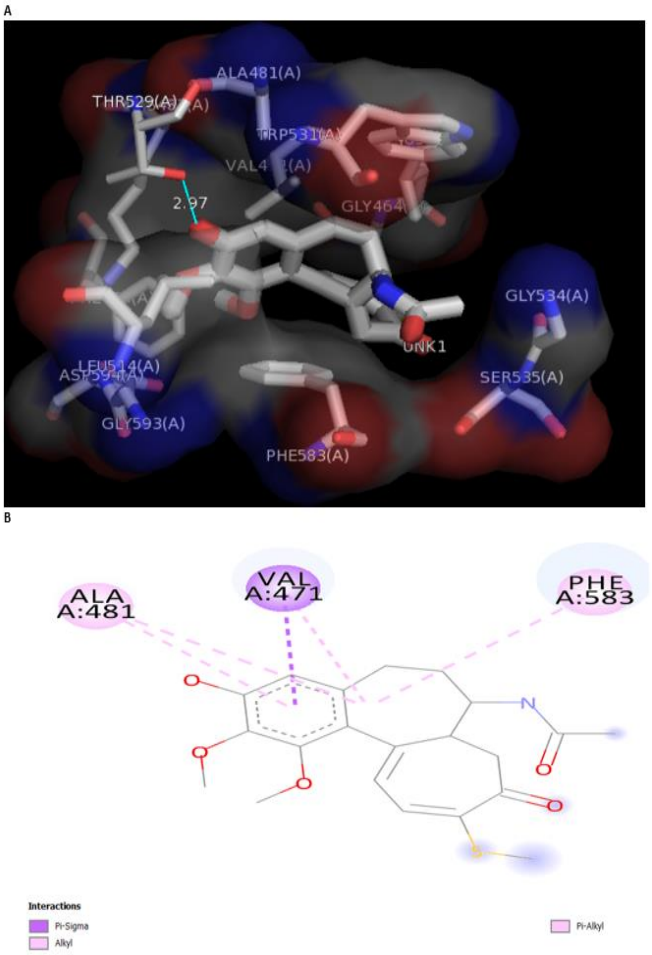

Figure 6. (A) 3D and (B) 2D view of V600E-BRAF /Compound 51 interactions

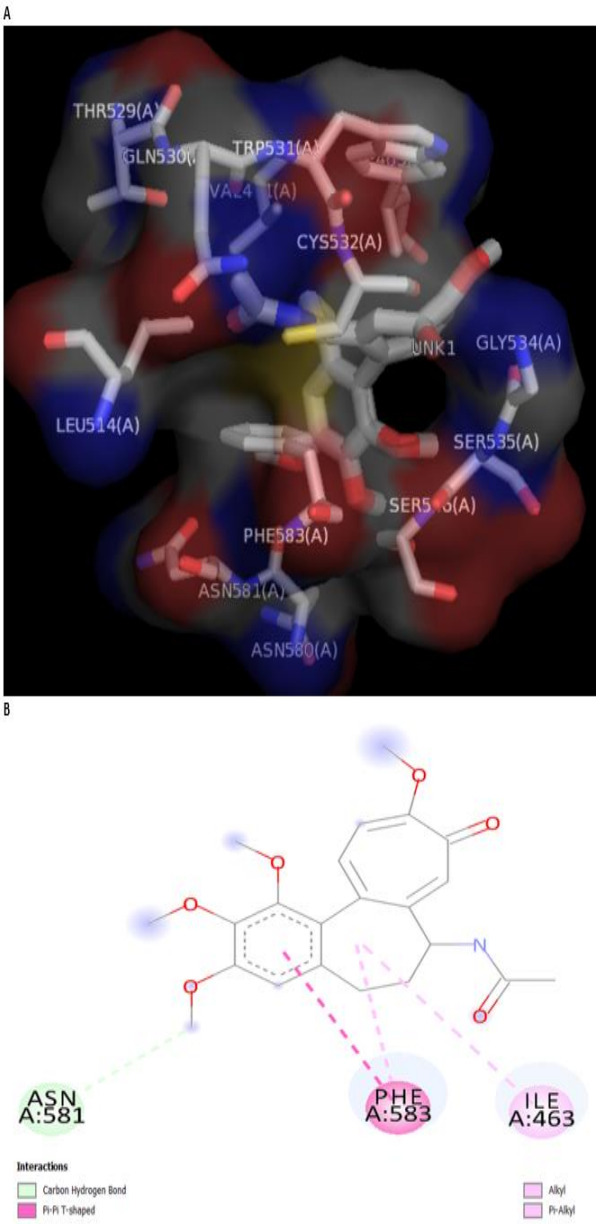

Figure 7. (A) 3D and (B) 2D view of V600E-BRAF /Compound 56 interactions

4. Conclusion

A quantitative analysis of the structure-activity relationship (QSAR) was performed on compounds retrieved from the NCI database. The genetic function algorithm was adopted for variable selection and the QSAR model was developed using the Multiple Linear Regression (MLR). The resulting model can be used to predict the anti-cancer activity of compounds against M19-MEL cell line. The predictive power of the model obtained was confirmed by ((R^2 (0.904), $R^2_{adjusted}$ (0.885), Q^2_{cv} (0.873) and R^2_{pred} (0.779)). A strong correlation was observed between the experimental and predicted values of the biological activities, which indicated the validity and quality of the QSAR model developed in this work. Four (4) most potent compounds out of fifty (57) compounds that showed better pGI_{50} activities were selected and screened through Lipinski's rule of five filters for oral bioavailability, ADMET risk filter for drug-like features, and synthetic accessibility for chemical synthesis. Later, V600E-BRAF, a known anti-melanoma target was used for docking. Based on low energy docking scores, the selected compounds were identified as the best hits against V600E-BRAF.

References

1. D. Schadendorf, A.C.J. van Akkooi, C. Berking, K.G. Griewank, R. Gutzmer, A. Hauschild, A. Stang, A. Roesch, and S. Ugurel, *Melanoma.* The Lancet, 392(10151)(2018) 971-984.
2. C. Karimkhani, B.Y. Reddy, R.P. Dellavalle, and S. Sundararajan, *Novel therapies for unresectable and metastatic melanoma.* bmj, 359(2017).
3. T.K. Eigentler, U.M. Caroli, P. Radny, and C. Garbe, *Palliative therapy of disseminated malignant melanoma: a systematic review of 41 randomised clinical trials.* The lancet oncology, 4(12)(2003) 748-759.
4. M. Danishuddin and A.U. Khan, *Structure based virtual screening to discover putative drug candidates: necessary considerations and successful case studies.* Methods, 71(2015) 135-145.
5. M. Danishuddin and A.U. Khan, *Virtual screening strategies: a state of art to combat with multiple drug resistance strains.* MOJ Proteomics Bioinform, 2(2)(2015) 00042.
6. M. Shahlaei, *Descriptor selection methods in quantitative structure–activity relationship studies: a review study.* Chemical reviews, 113(10)(2013) 8093-8103.
7. Y.C. Martin, *3D QSAR: current state, scope, and limitations*, in *3D QSAR in Drug Design*. 1998, Springer. p. 3-23.
8. P.R. Ashton, M.C. Fyfe, S.K. Hickingbottom, J.F. Stoddart, A.J. White, and D.J. Williams, *Hammett correlations 'beyond the molecule'1.* Journal of the Chemical Society, Perkin Transactions 2, (10)(1998) 2117-2128.
9. A.S. Reddy, S.P. Pati, P.P. Kumar, H. Pradeep, and G.N. Sastry, *Virtual screening in drug discovery-a computational perspective.* Current Protein and Peptide Science, 8(4)(2007) 329-351.
10. P. Srivani and G.N. Sastry, *Potential choline kinase inhibitors: a molecular modeling study of bis-quinolinium compounds.* Journal of Molecular Graphics and Modelling, 27(6)(2009) 676-688.
11. R. Benigni and A. Giuliani, *Putting the predictive toxicology challenge into perspective: reflections on the results.* Bioinformatics, 19(10)(2003) 1194-1200.
12. C. Hansch, A. Leo, S.B. Mekapati, and A. Kurup, *Qsar and Adme.* Bioorganic & medicinal chemistry, 12(12)(2004) 3391-3400.
13. H.K. Srivastava, M. Chourasia, D. Kumar, and G.N. Sastry, *Comparison of computational methods to model DNA minor groove binders.* Journal of chemical information and modeling, 51(3)(2011) 558-571.
14. J. Verma, V.M. Khedkar, and E.C. Coutinho, *3D-QSAR in drug design-a review.* Current topics in medicinal chemistry, 10(1)(2010) 95-115.
15. A.S. Dhillon, S. Hagan, O. Rath, and W. Kolch, *MAP kinase signalling pathways in cancer.* Oncogene, 26(22)(2007) 3279.
16. S.A. Amin and S. Gayen, *Modelling the cytotoxic activity of pyrazolo-triazole hybrids using descriptors calculated from the open source tool "PaDEL-descriptor".* Journal of Taibah University for Science, 10(6)(2016) 896-905.
17. C.W. Yap, *PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints.* Journal of computational chemistry, 32(7)(2011) 1466-1474.
18. K. Rajer-Kanduč, J. Zupan, and N. Majcen, *Separation of data on the training and test set for modelling: a case study for modelling of five colour properties of a white pigment.* Chemometrics and intelligent laboratory systems, 65(2)(2003) 221-229.
19. R.W. Kennard and L.A. Stone, *Computer aided design of experiments.* Technometrics, 11(1)(1969) 137-148.
20. A. Tropsha, P. Gramatica, and V.K. Gombar, *The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models.* Molecular Informatics, 22(1)(2003) 69-77.
21. A. Golbraikh and A. Tropsha, *Predictive QSAR modeling based on diversity sampling of experimental datasets for*

the training and test set selection. Molecular diversity, 5(4)(2000) 231-243.
22.C.A. Lipinski, *Lead-and drug-like compounds: the rule-of-five revolution.* Drug Discovery Today: Technologies, 1(4)(2004) 337-341.
23.C.A. Lipinski, F. Lombardo, B.W. Dominy, and P.J. Feeney, *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* Advanced drug delivery reviews, 64(2012) 4-17.
24.M.S. Brose, P. Volpe, M. Feldman, M. Kumar, I. Rishi, R. Gerrero, E. Einhorn, M. Herlyn, J. Minna, and A. Nicholson, *BRAF and RAS mutations in human lung cancer and melanoma.* Cancer research, 62(23)(2002) 6997-7000.
25.G. Bollag, P. Hirth, J. Tsai, J. Zhang, P.N. Ibrahim, H. Cho, W. Spevak, C. Zhang, Y. Zhang, and G. Habets, *Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma.* Nature, 467(7315)(2010) 596.
26.W.-K. Choi, M.I. El-Gamal, H.S. Choi, D. Baek, and C.-H. Oh, *New diarylureas and diarylamides containing 1, 3, 4-triarylpyrazole scaffold: Synthesis, antiproliferative evaluation against melanoma cell lines, ERK kinase inhibition, and molecular docking studies.* European journal of medicinal chemistry, 46(12)(2011) 5754-5762.
27.W. Wu, C. Zhang, W. Lin, Q. Chen, X. Guo, Y. Qian, and L. Zhang, *Quantitative structure-property relationship (QSPR) modeling of drug-loaded polymeric micelles via genetic function approximation.* PloS one, 10(3)(2015) e0119575.
28.B.A. Umar, A. Uzairu, G.A. Shallangwa, and U. Sani, *QSAR modeling for the prediction of pGI50 activity of compounds on LOX IMVI cell line and ligand-based design of potent compounds using in silico virtual screening.* Network Modeling Analysis in Health Informatics and Bioinformatics, 8(1)(2019) 22.
29.M. Jalali-Heravi and E. Konuze, *Use of quantitative structure property relationships in predicting the Kraft point of anionic surfactants.* Electronic Journal of Molecular Design, 1(2002) 410-417.
30.D. Gadaleta, G.F. Mangiatordi, M. Catto, A. Carotti, and O. Nicolotti, *Applicability domain for QSAR models: where theory meets reality.* International Journal of Quantitative Structure-Property Relationships (IJQSPR), 1(1)(2016) 45-63.
31.K. Roy, S. Kar, and R.N. Das, *A primer on QSAR/QSPR modeling: Fundamental concepts*. 2015: Springer.
32.P. Gramatica, E. Giani, and E. Papa, *Statistical external validation and consensus modeling: a QSPR case study for Koc prediction.* Journal of Molecular Graphics and Modelling, 25(6)(2007) 755-766.
33.A. Daina, O. Michielin, and V. Zoete, *SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules.* Scientific reports, 7(2017) 42717.
34.C.R. Guimarães, B.K. Rai, M.J. Munchhof, S. Liu, J. Wang, S.K. Bhattacharya, and L. Buckbinder, *Understanding the impact of the P-loop conformation on kinase selectivity.* Journal of chemical information and modeling, 51(6)(2011) 1199-1204.