Heliyon 6 (2020) e03640

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

# QSAR modelling and molecular docking studies for anti-cancer compounds against melanoma cell line SK-MEL-2

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

*Department of Chemistry, Faculty of Physical Sciences, Ahmad Bello University, Zaria, P.M.B.1045 Kaduna State, Nigeria*

ARTICLE INFO

*Keywords:*
Pharmaceutical chemistry
Theoretical chemistry
SK-MEL-2 cell line
Melanoma
QSAR
Binding energy
Pi-Pi interaction
V600E-BRAF

ABSTRACT

A dataset of seventy-two (72) cytotoxic compounds of the National Cancer Institute (NCI) was studied by QSAR and docking approaches to gain deeper insights into ligands selectivity on SK-MEL-2 cell line. The QSAR model was built using fifty (50) molecules and the best-generated model based on multiple linear regression showed, respectively good quality of fits ($R^2$ (0.864), $R^2_{adjusted}$ (0.845), $Q^2_{cv}$ (0.799) and $R^2_{pred}$ (0.706)). The model's predictive ability was determined by a test set of twenty-two (22) compounds. Compounds 30 and 41 were selected as templates for in silico design because they had high $pGI_{50}$ activity and are in the model's applicability domain. The obtained information from the model was explored to design novel molecules by introducing various modifications. Moreover, the designed compounds with better-predicted activity ($pGI_{50}$) values were selected and docked on the active site of the protein (PDB-CODE: 3OG7) which is responsible for melanoma cancer to elucidate their binding mode. AN2 ($-12.1 kcalmol^{-1}$) and AC4 ($-12.4 kcalmol^{-1}$) showed a better binding score for the target when compared with (vemurafenib, $-11.3 kcalmol^{-1}$) the known inhibitor of the target (V600E-BRAF). These findings may be very helpful in early anti-cancer drug development.

## 1. Introduction

Melanoma is identified as one of the most dangerous forms of the skin tumor, having quick metastasizing, progression and a high burden of death, especially when detected late [1]. Even though a significant figure of therapies have been established recently for the late-stage melanoma cancer, this disease has not been defeated yet, resistance develops through cancer heterogeneity, an alternative pathways (signaling) and some serious adverse conditions limiting the potency of the novel treatments [2]. Thus, though therapeutic alternatives are now available and better for the patients with an advanced stage melanoma than before, there is however need to develop new potent drugs that target melanoma and several techniques are being used from exploring a better delivery system for old compounds to assessing new targets [3, 4].

To obtain a complete drug product requires a long period of about twelve years and the estimated expenses for the sold drug are also very high [5, 6]. This lengthy and expensive process may cause delays and failure of drugs development. Therefore, it is very vital to predict the failures before the clinical stage to reduce the costs of drug development [7]. To filter out the possible failures in the drug development stage, several approaches are being used like the in silico, in vitro, etc. An example of an in silico approach is the modeling of a quantitative structure-activity relationships (QSARs) which can be used to screen chemical libraries, understand drugs action and design novel compounds [8, 9]. Combinatorial approaches are an influential technique adopted in the selection to reduce the time of drug development with various mechanisms of action and it has been used to treat cancer [10, 11]. QSAR modeling has now become very crucial in the understanding the biological and physicochemical properties of the molecules [12, 13]. This method is the most essential means adopted in the ligand-based design of drug and have been used severally for the determination of assorted parameters like stability, carcinogenicity, toxicity, retention time, ADME and some other physicochemical parameters apart from the prediction of the biological activity [14, 15, 16, 17, 18].

In the current study, we generate a QSAR model that can be applied to predict the cytotoxic effect of various anticancer compounds against the SK-MEL-2 human melanoma cell-line utilizing the dataset obtained from the database of National Cancer Institute (NCI). A QSAR analysis was conducted on this data to bring out the key structural features responsible for the anti-melanoma activity and also to design a novel template that resembles the most potent bioactive conformation. This QSAR approach served as a fundamental predictive tool and used primarily in the design

* Corresponding author.
*E-mail address:* abdallahbum@yahoo.com (A.B. Umar).

https://doi.org/10.1016/j.heliyon.2020.e03640
Received 21 January 2020; Received in revised form 6 March 2020; Accepted 18 March 2020

of pharmaceuticals [19]. Further, docking simulation studies were conducted on the V600E-BRAF protein receptor which is responsible for melanoma cancer.

## 2. Computational methods

### 2.1. Data collection and structure preparation

Seventy-two (72) set of compounds and their $pGI_{50}$ activities on SK-MEL-2 melanoma cell-line was retrieved from National Cancer Institute (NCI) database. The anticancer activity, chemical name and NSC number of these compounds were presented in Table 1. 2D structures of the studied molecules were converted to the 3D structures utilizing the Spartan 14 Version 1.1.4 from Wavefunction Inc., on the Dell Intel (R) Core ((TM)i7-5500U CPU), 16.00GB RAM @ 2.400GHz 2.400GHz processor, 64-bit Operating system, a ×64-based processor on Windows 8.1 Pro). Molecular optimisation of the molecules was set at the ground state employing the Density Functional Theory (DFT/B3LYP) approach and 6-31G* basis set. The optimized 3D structure was formatted to the SD file and then taken to the PaDEL descriptor tool kit to generate required descriptors for further studies [20].

### 2.2. QSAR model development and validation

The dataset was split into two (2) subsets, the training (modelling) data set and testing (pridicting) data set using Kennard Stone Algorithm [21, 22]. The training data set is utilized in building the QSAR model which contains 70% of the data and the remaining 30% is for the testing data set that was utilized to describe the predictive capability of the model [23]. All the studied molecules were screened using the generated QSAR model for $pGI_{50}$ activity prediction.

Material Studio version 8.0 Software from BIOVIA-Accelrys, was adopted in performing GA and model building. The genetic algorithm (GA) was utilized in chosen proper descriptors as this improves the model accuracy [24]. Multiple Linear Regression (MLR) was applied to the training set to determine the correlation between dependent variable $pGI_{50}$ (Y) and independent variable, descriptors (X). In this regression study, the contingent mean of the $pGI_{50}$ (dependent variable) relies on X (descriptors). The best model QSAR was chosen based on the statistical validation parameters like 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 Eqs. (1), (2), (3), and (4):

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

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

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

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

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 modeling set [23].

### 2.3. Ligand-protein preparation and docking studies

The selected ligands (compounds) were optimized and formatted to PDB files for docking utilizing Spartan 14. The X-ray structure of the V600E-BRAF (receptor) in complex with ligand (vemurafenib) (PDB CODE: 3OG7) [25, 26, 27] was retrieved from (www.rcsb.org). V600E-BRAF was imported into the Discovery Studio Visualizer version 16.1.0.15350 and the PDB file was prepared by updating the hydrogen atoms and removing the excess water molecules present in the X-ray structure. This complex structure comprises two homo-dimeric chains A and B. Our goal was to target the mutated chain A of the V600E-BRAF. Thus, chain B was removed from the structure of 3OG7 and the bound ligand also removed from chain A. All the selected compounds (ligands) were docked into the active kinase domain of V600E-BRAF using Auto Dock vina of PyRx docking and virtual screening programme.

## 3. Results and discussions

### 3.1. Developed QSAR model and validation

By using $pGI_{50}$ values (activity) as dependent variables and calculated descriptors as independent variables, Kennard stone algorithm was applied in splitting the data into two subsets, that is, fifty compounds as training data set while twenty-two compounds as the test set as presented in Table 1 and regression were executed for QSAR analysis. The robustness of the generated model was depicted via the activity interactive graph that presents the predicted against experimental ($pGI_{50}$) activity plot as in Figure 1. The best QSAR model is represented by Eq. (5) and the statistical parameters of all the generated models with threshold values were presented in Table 2:

$$\begin{aligned} pGI_{50(SK-MEL-2)} = & -2.360005510(\mathbf{SM1\_Dze}) + 0.010830214(\mathbf{Si}) \\ & -0.131575817(\mathbf{VE3\_Dt}) + 0.385665511\ (\mathbf{MLFER\_BH}) \\ & +42.432325759(\mathbf{JGI4}) + 0.128296620(\mathbf{VE3\_D}) \\ & +2.645302 \end{aligned} \quad (5)$$

$$N_{training} = 50, R^2 = 0.864, R^2_{adjusted.} = 0.845,\ Q^2_{cv.} = 0.799,\ N_{test} = 22,\ R^2_{test} = 0.706$$

Further, the generated model has achieved high activity-descriptor relationship efficiency of 86.4% as shown by the regression-coefficient ($R^2$ = 0.864) and a good activity prediction efficiency of 79.9% as shown by the cross-validated regression-coefficient ($Q^2_{CV}$ = 0.79.9). Knowing the high predictive and descriptive ability, the generated model was considered to be highly robust in predicting the anti-cancer activity of these compounds against the SK-MEL-2 melanoma cancer cell line.

The developed model, which was generated using the training data set compounds, was used to predict the (pGI50) activity of the testing data set of compounds. These predicted activities of the studied compounds for SK-MEL-2 cell line by the built QSAR model are shown in Table 1. Lower values of residual found from both training set and testing set as presented in Table 1 indicate that the model has a high ability to establish the correlation between the activity and structure. The correlation between the experimental activity and predicted activity according to the model was highly significant as determined by statistical analysis. The closeness of regression-coefficient ($R^2$) to 1.0 indicates that the developed model elaborated a great portion of the descriptor-variation large enough for a good QSAR model. The 0.864 value, demonstrates that about 86.4% of the variation is found within the residual indicating that the model highly predictive and very good.

The large adjusted regression-coefficient $R^2$ ($R^2_{adj}$) value presented in the generated QSAR model and its closeness to the value of regression-coefficient ($R^2$) indicates that the developed model has perfect descriptive ability to descriptors in it and it further illustrates the true impact of used descriptors on the $pGI_{50}$. Additionally, the closeness of cross-validated regression-coefficient ($Q^2_{cv}$) to the regression coefficient ($R^2$) and its high value showed that the built model was not overfitted. The high $R^2_{test}$ as shown in the developed model explains

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

| S/N | NSC | Chemical Name | $pGI_{50}$ | $pGI_{50}$ (Pred.) | Residual |
|---|---|---|---|---|---|
| 1t | 267,469 | Deoxydoxorubicin | 6.904 | 6.618 | 0.286 |
| 2 | 269,148 | Menogaril | 5.614 | 6.591 | -0.977 |
| 3 | 268,242 | N,N-Dibenzyldaunorubicin Hydrochloride | 8.000 | 7.143 | 0.857 |
| 4 | 126,771 | Dichloroallyl Lawsone | 4.756 | 5.467 | -0.711 |
| 5 | 136,044 | Rhodomycin A | 7.644 | 7.455 | 0.189 |
| 6 | 140,377 | Arnebin 1 | 5.697 | 5.357 | 0.340 |
| 7 | 196,524 | Epsilon.-Rhodomycinone | 5.506 | 5.206 | 0.300 |
| 8t | 212,509 | 4beta-Hydroxywithanolide | 6.346 | 6.156 | 0.190 |
| 9t | 215,139 | Bikaverin | 6.063 | 5.604 | 0.459 |
| 10 | 236,613 | Plumbagin | 5.732 | 5.564 | 0.168 |
| 11 | 252,844 | Shikalkin | 5.710 | 5.705 | 0.005 |
| 12 | 257,450 | Dermocybin | 4.000 | 4.366 | -0.366 |
| 13 | 143,095 | Pyrozofurin | 4.072 | 4.885 | -0.813 |
| 14 | 629,971 | 9-Aminocamptothecin (R,S) | 6.540 | 6.432 | 0.108 |
| 15t | 606,173 | 11-Hydroxymethyl-20(Rs)-Camptothecin | 4.744 | 5.644 | -0.900 |
| 16 | 364,830 | Camptothecin,(N-Diethyl) Glycinate | 6.418 | 6.247 | 0.171 |
| 17 | 94,600 | Camptothecin | 6.506 | 6.504 | 0.002 |
| 18 | 606,985 | Campothecin Analog | 6.308 | 5.617 | 0.691 |
| 19 | 606,499 | Camptothecin Butylglycinate Ester Hydrochloride | 5.503 | 6.689 | -1.186 |
| 20 | 606,497 | Camptothecinethylglycinate Esterhydrochloride | 5.557 | 5.587 | -0.030 |
| 21t | 176,323 | 9-Methoxycamptothecin | 7.000 | 6.840 | 0.160 |
| 22 | 3,088 | Chlorambucil | 4.498 | 4.636 | -0.138 |
| 23 | 338,947 | Clomesone | 3.463 | 2.401 | 1.062 |
| 24 | 95,678 | Picolinaldehyde | 4.671 | 4.817 | -0.146 |
| 25 | 264,880 | Dihydro-5-Azacytidine | 5.941 | 4.842 | 1.099 |
| 26 | 163,501 | Acivicin | 4.567 | 4.000 | 0.567 |
| 27t | 71,851 | Alpha.-Thiodeoxyguanosine | 3.746 | 3.911 | -0.165 |
| 28 | 132,483 | L-Aspartic Acid | 4.000 | 4.470 | -0.470 |
| 29t | 308,847 | Amonafide | 5.566 | 5.893 | -0.327 |
| 30 | 355,644 | Anthra[1,9-Cd]Pyrazol-6(2h)-One Der | 9.924 | 9.658 | 0.266 |
| 31 | 63,878 | Cytosine, Monohydrochloride | 5.000 | 4.449 | 0.551 |
| 32t | 182,986 | Diaziquone | 5.200 | 6.108 | -0.908 |
| 33t | 139,105 | Triazinate | 4.832 | 5.759 | -0.927 |
| 34 | 409,962 | Carmustine | 4.037 | 3.476 | 0.561 |
| 35* | 337,766 | Bisantrene Hydrochloride | 8.000 | 7.982 | 0.018 |
| 36 | 750 | Busulfan | 3.606 | 3.548 | 0.058 |
| 37t | 95,382 | Camptothecin, Acetate | 4.097 | 4.856 | -0.759 |
| 38 | 107,124 | 10-Hydroxycamptothecin | 6.381 | 6.226 | 0.155 |
| 39 | 79,037 | Lomustine | 4.644 | 4.314 | 0.330 |
| 40 | 132,313 | Dianhydrodulcitol | 4.072 | 4.258 | -0.186 |
| 41 | 376,128 | AC1L2OAS | 9.904 | 10.033 | -0.129 |
| 42 | 73,754 | Fluorodopan | 3.612 | 4.632 | -1.020 |
| 43 | 148,958 | Uracil | 3.101 | 4.345 | -1.244 |
| 44 | 1895 | Guanazole | 2.027 | 2.342 | -0.315 |
| 45 | 329,680 | Hepsulfam | 3.419 | 3.140 | 0.279 |
| 46t | 142,982 | Hycanthone Mesylate | 5.104 | 6.731 | -1.627 |
| 47 | 32,065 | Hydroxyurea | 2.767 | 2.133 | 0.634 |
| 48 | 153,353 | Alanosine Monosodium Salt | 4.013 | 4.090 | -0.077 |
| 49 | 249,992 | Amsacrine | 5.265 | 5.936 | -0.671 |
| 50t* | 740 | Methotrexate | 4.042 | 4.432 | -0.390 |
| 51t | 95,441 | Semustine | 4.691 | 4.654 | 0.037 |
| 52t | 26,980 | Mitomycin C | 5.743 | 6.215 | -0.472 |
| 53 | 353,451 | Mitozolomide | 4.000 | 4.627 | -0.627 |
| 54t* | 268,242 | N,N-Dibenzyldaunorubicin Hydrochloride | 5.695 | 5.913 | -0.218 |
| 55t | 95,466 | Urea, | 3.558 | 3.792 | -0.234 |
| 56 | 25,154 | Pipobroman | 4.023 | 4.528 | -0.505 |
| 57t | 56,410 | Profiromycin | 4.850 | 5.720 | -0.870 |
| 58t | 366,140 | Pyrazoloacridine Mesylate | 6.139 | 6.440 | -0.301 |

(*continued on next page*)

**Table 1** (*continued*)

| S/N | NSC | Chemical Name | $pGI_{50}$ | $pGI_{50}$ (Pred.) | Residual |
|---|---|---|---|---|---|
| 59 | 51,143 | Pyrazoloimidazole | 2.100 | 3.426 | -1.326 |
| 60t | 172,112 | Spiromustine | 3.611 | 4.607 | -0.996 |
| 61 | 125,973 | Paclitaxel; | 7.295 | 8.423 | -1.128 |
| 62 | 296,934 | Teroxirone | 4.211 | 3.739 | 0.472 |
| 63t | 363,812 | 5-((4-Chlorobenzyl)Thio)-3-(Trifluoromethyl)-1h-1,2,4-Triazole | 5.242 | 4.558 | 0.684 |
| 64 | 361,792 | 3-Demethylthiocolchicine; | 7.656 | 6.152 | 1.504 |
| 65 | 752 | 6-Thioguanine | 5.770 | 6.093 | -0.323 |
| 66 | 6396 | Thiotepa | 4.526 | 4.582 | -0.056 |
| 67 | 9,706 | Triethylenemelamine | 4.746 | 4.409 | 0.337 |
| 68t | 83,265 | Tritylcysteine | 6.265 | 6.046 | 0.219 |
| 69 | 49,842 | Vinblastine Sulfate | 9.015 | 8.059 | 0.956 |
| 70 | 67,574 | Vincristine Sulfate | 6.870 | 7.711 | -0.841 |
| 71t | 757 | Colchicine | 8.402 | 8.029 | 0.373 |
| 72 | 33,410 | N-Benzoyl-Deacetylcolchicine | 7.921 | 7.314 | 0.607 |

't' indicate testing sets.
'*' indicate compounds outside the defined AD of the model.

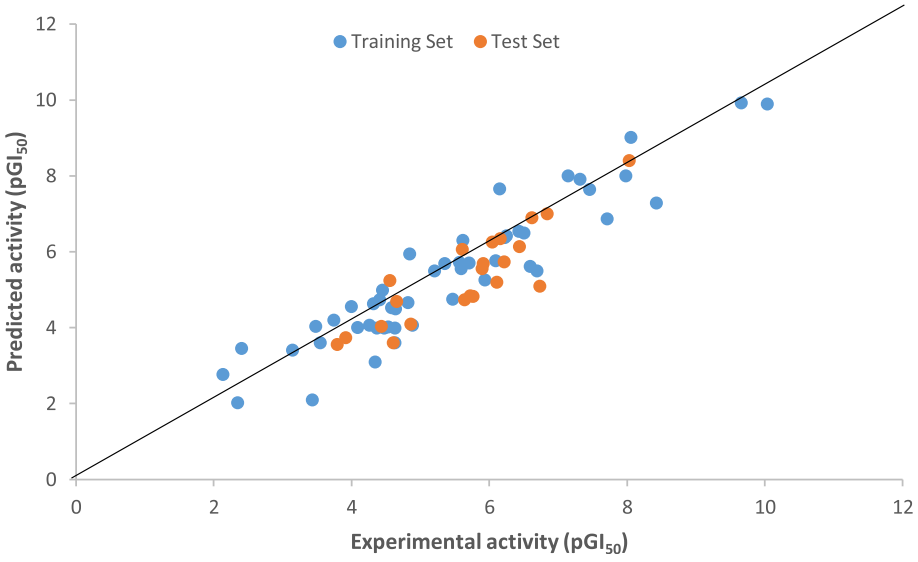


**Figure 1.** Predicted versus experimental $pGI_{50}$ values for both the training and testing sets.

**Table 2.** Statistical parameters of the first Three QSAR models and Threshold values.

| Parameters | Threshold Value | Model 1 | Model 2 | Model 3 |
|---|---|---|---|---|
| Coefficient of determination ($R^2$) | ≥0.6 | 0.864 | 0.863 | 0.863 |
| Cross-validation coefficient ($Q^2_{cv}$) | <0.5 | 0.799 | 0.798 | 0.805 |
| The coefficient of determination for external test set ($R^2_{test}$) | ≥0.6 | 0.706 | 0.688 | 0.650 |
| Difference between $R^2$ and $Q^2$ | ≤0.3 | 0.065 | 0.065 | 0.058 |
| Minimum number of an external test set ($N_{test}$) | ≥5 | 22 | 22 | 22 |

that the generated model can provide a good and valid prediction for the new compounds. A good and acceptable QSAR model must obey the following criteria: regression-coefficient ($R^2$) and adjusted regression-coefficient ($R^2$adj) values close to one. The Cross validated regression-coefficient ($Q^2_{cv}$) > 0.5, $R^2$ - $Q^2_{cv} \leq 0.3$, $R^2_{test} \geq 0.6$, and $N_{test} \geq 5$ [23,28,29]. The generated QSAR model satisfied the criteria and therefore acceptable statistically. We can, therefore, conclude that the developed model will correctly predict the anti-melanoma $pGI_{50}$ activity of a given compound.

### 3.2. Contribution and interpretation of descriptors

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. By defining the descriptors that in the developed model, it is likely to understand the features that are related to anti-melanoma activity. Therefore, concise descriptions of the selected descriptors are presented in Table 3 and an acceptable interpretation is provided. The contribution and significance of every descriptor in the

**Table 3.** Description of the descriptors in the model with their mean effect (ME).

| Descriptors | Description | ME |
|---|---|---|
| SM1_Dze | Spectral moment of order 1 from Barysz matrix/weighted by Sanderson electronegativities | -1.36906 |
| Si | Sum of first first ionization potentials (scaled on carbon atom) | 1.365518 |
| VE3_Dt | Logarithmic coefficient sum of the last eigenvector from detour matrix | 0.768067 |
| MLFER_BH | Overall or summation solute hydrogen bond basicity | 0.311063 |
| JGI4 | Mean topological charge index of order 4 | 0.671447 |
| VE3_D | Logarithmic coefficient sum of the last eigenvector from topological distance matrix | -0.74704 |

built model were evaluated by the computation of the mean effect (ME) value [30] of every descriptor by using Eq. (6). The values for the ME are shown in Table 3:

$$ME_j = \frac{\beta_j \sum_{i=1}^{i=n} d_{ij}}{\sum_j^m \beta_j \sum_i^n d_{ij}} \quad (6)$$

Where: $ME_j$ is the mean effect (ME) of the descriptor $j$, $\beta_j$ 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.

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.

**Si > VE3_Dt > JGI4 > MLFER_BH > VE3_D > SM1_Dze**

Si is the Sum of first ionization potentials (scaled on carbon atom). This descriptor had a positive ME value in Table 3. This indicates its relative significance in influencing the activity of the compounds. Thus, the introduction of the substituent that will improve the ionization of carbon in the molecular structure of a compound might improve $pGI_{50}$ activity for the compound.

VE3_Dt is the Logarithmic coefficient sum of the last eigenvector from detour matrix. The positive value of the mean effect (Table 3) for VE3_Dt suggests a positive contribution to the activity. JGI4 is defined as the Mean topological charge index of order 4. The ME of JGI4, when increased, was found to positively affect the anti-cancer activity of the compound.

MLFER_BH is a 2D molecular linear free energy relation descriptor. The descriptor was found to have positive ME value as seen in Table 3, indicating that the free energy connected with the linearity of the molecule if large would improve the value of $pGI_{50}$ of a compound thereby strengthening its stability but decreasing its entropy and spontaneity in interacting with the protein target. Hence, the less linear the compound is the better its activity.

SM1_Dze is defined as Spectral moment of order 1 from Barysz matrix/weighted by Sanderson electronegativities. The negative ME value of the descriptor (Table 3) shows the decrease in the value of the descriptor would improve the $pGI_{50}$ of the compounds.

VE3_D is described as the Logarithmic coefficient sum of the last eigenvector from the topological distance matrix. It has a negative ME of high magnitude, meaning a decrease in its value favors an increase in the anti-cancer activity value of the molecule. VE3_D is an index of branching with a lower value implying to increase branching. The descriptors selected for developing the model in this research encoded electronic, topological, and other geometrical aspects of the compounds. The presence of the descriptors in the QSAR model indicates the role of steric and

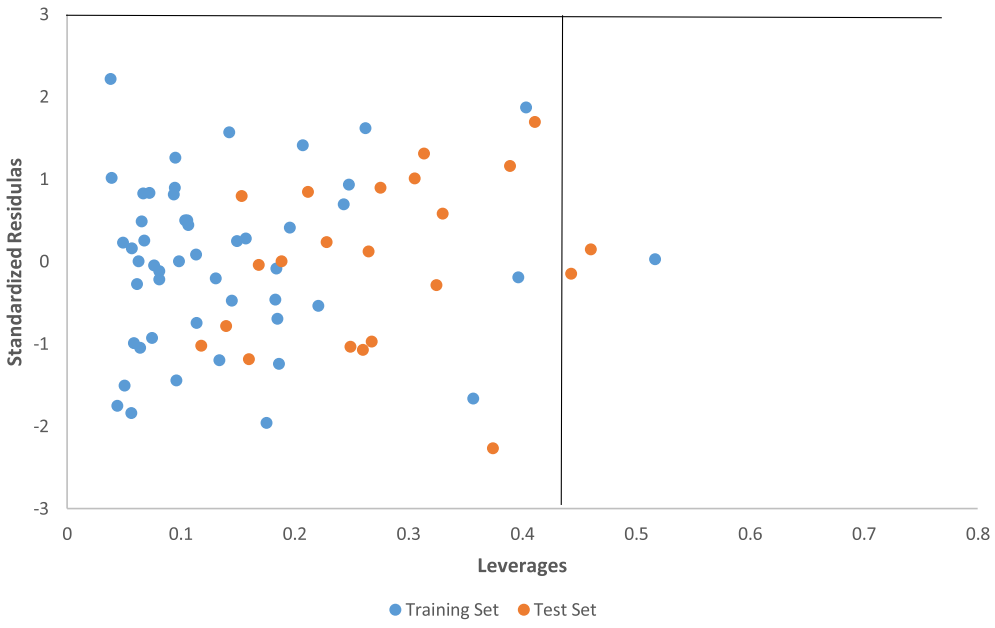


**Figure 2.** The plot of the standardized residuals against the leverages (Williams plot).

electronic interactions in influencing anti-melanoma $pGI_{50}$ activity on SK-MEL-2 cell line.

### 3.3. Applicability domain (AD) and in-silico screening

The applicability domain, AD of the model is the theoretical space in the chemical region comprising of both the descriptors of the QSAR model and the modeled response. This domain permits prediction of uncertainty in the identification of a specific molecule based on the data set of the compounds used in developing the model. The AD is also applied to define the X-outliers in case of the training data set and identify the molecules residing outside the defined AD in case of the testing set utilizing the fundamental theory of standardization method [29]. Several techniques had been employed to define AD of a QSAR model [31]. The generally used one was demonstrated by Gramatica [32], which employed the leverages for each of the compounds of the data set. The leveraged approach enables the evaluation of the status of a novel compound in a model [32]. Therefore, Leverage method is utilized and is shown as $h_i$ in Eq. (7):

$$h_i = x_i(X^T X)^{-1} x_i^T \tag{7}$$

Where x: represents descriptor's vector of the selected molecule, X: refers to the descriptor matrix obtained from the training data set descriptor values and the h* (warning leverage) was calculated as in Eq. (8):

$$h^* = \frac{3(p+1)}{N} \tag{8}$$

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

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 $hi > h^*$ seriously affects the QSAR model performance and can be eliminated from the AD. Further, ±3 value range of standardized residuals is always accepted as a threshold value for affirming predictions of a molecule, because points that lie within ±3 of standardized residuals from the mean cover 99% of the normally distributed data [33]. In this regard, the leverage approach with standardized residuals was combined for the determination and characterization and of the AD. The Williams plot of the constructed model is presented in Figure 2. 0.420 was found to be the $h^*$ (warning leverage) for the built model. According to leverages ($hi > 0.420$), only one training set compound (35) and two test set compounds (50 and 54) was outside the AD (Figure 2) of the constructed model, thus, were recognized as structurally influential compounds based on the large leverage values ($hi > h^*$) they posses [34].

The in silico screening method is a valuable tool for predicting and recognizing novel biologically potent molecules with enhanced properties before their actual synthesis [23, 34]. Consequently, the in silico method decreases the cost and time involved in recognizing potent leads. Virtual screening was achieved by deletion, insertion, and substitution of various substitutes at different positions on the original templates of molecules [29, 35] and the results of the structural adjustments on the biological activity were studied. Later, the domain of applicability of the QSAR model was established to apply the developed model in screening novel compounds. Therefore, the in silico screening was employed to design novel compounds with good $pGI_{50}$ based on the built model and was validated by the developed QSAR model. For this purpose, compound 30 and 41 listed in Table 1, were selected as templates, because they had relatively high anti-melanoma activity ($pGI_{50,}$ 9.924 and 9.904); they are within the model's AD and changes can simply be done around the benzene ring of the main structure. The structure of the templates used for modifications is presented in Figures 3 and 4 respectively. The molecules were adjusted in such a way that their synthesis was experimentally achievable. Next, in silico screen was employed by replacing various groups in the $R_1$ and $R_2$ sites of the ring; which lead to compounds with improved predicted anti-melanoma activity values as shown in Tables 4 and 5 respectively. Before the prediction of the $pGI_{50}$ activity values of newly designed compounds by the

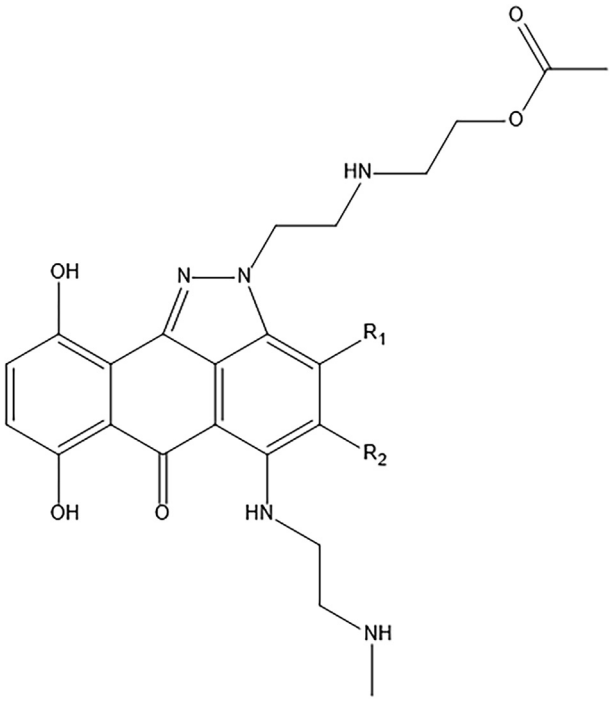


**Figure 3.** The template structure (Compound 30) used for the design.

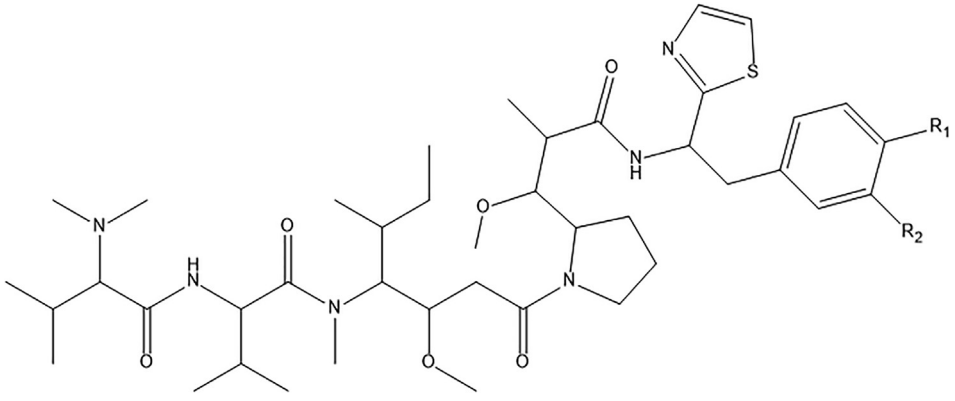


**Figure 4.** The template structure (Compound 41) used for the design.

**Table 4.** Structural alteration of Compound 30 and Predicted $pGI_{50}$ and Leverage limit.

| ID | $R_1$ | $R_2$ | Predicted $pGI_{50}$ | Leverage-limit |
|---|---|---|---|---|
| AN1 | $NH_2$ | $NH_2$ | 10.935 | 0.259 |
| AN2 | $NH_2$ | $CH_3$ | 11.051 | 0.253 |
| AN3 | $CH_3$ | Br | 10.756 | 0.342 |
| AN4 | $OCH_3$ | Cl | 10.802 | 0.148 |
| AN5 | Br | $OCH_3$ | 10.804 | 0.304 |

**Table 5.** Structural alteration of Compound 41 and Predicted $pGI_{50}$ and Leverage limit.

| ID | $R_1$ | $R_2$ | Predicted $pGI_{50}$ | Leverage-limit |
|---|---|---|---|---|
| AC1 | $NH_2$ | $NH_2$ | 11.877 | 0.334 |
| AC2 | Cl | $CH_3$ | 11.967 | 0.178 |
| AC3 | $NH_2$ | $CH_3$ | 11.699 | 0.372 |
| AC4 | $NH_2$ | $NHSO_2CH_3$ | 12.014 | 0.283 |
| AC5 | Cl | Br | 11.274 | 0.289 |

developed model, their geometry has been optimized, descriptors computed and their leverages determined as explained for the training set. The warning leverage (limit) for the model ($h^* = 0.420$) was used as the threshold value to screen the designed compounds. Among various molecules designed, the compound AC4 revealed the best activity ($pGI_{50} = 12.014$). Consequently, it is confirming that adopting a simple QSAR model, it is feasible to concurrently identify compounds with enhanced activity and to discover the structural alterations that don't fall within the defined applicability domain of the model. Lastly, this result proves the reliability of the QSAR models and it confirms that with the development of the QSAR model and application of in silico screening method, it is feasible to recognize new synthetic targets for drug development.

### 3.4. Molecular docking simulation studies

Designed molecules AN1, AN2, AC2, and AC4 were chosen for the molecular docking simulation study to represent molecules shown in Tables 4 and 5 respectively as they have the highest predicted $pGI_{50}$ activity values with low leverage when compared with the model warning leverage limit. The best docking result for the selected compounds was shown in Table 6. To validate the docking protocol and productivity, the co-crystalized ligand (vemurafenib) was also docked to the binding site of V600E-BRAF with the binding affinity of −11.3 kcal $mol^{-1}$ and the RMSD value for both upper and lower bounds were measured (0.0) which proved the docking protocol and productivity.

The obtained docking poses from the Discovery Studio visualizer were shown in Figures 5, 6, 7, 8, and 9 respectively. The designed molecules had negative free energy of binding higher in magnitude when compared with vemurafenib and this indicates a better binding affinity with the receptor. This shows that the designed compounds could be used as an anti-melanoma drug. The docking poses of the designed compounds showed that they interacted with the binding pocket of a protein target in a way similar to vemurafenib with additional number interactions.

AN1 docks with the V600E-BRAF domain with the binding free energy of –11.9 kcal $mol^{-1}$ as presented in Figure 5. Six H-bonds were present between the receptor and molecule (ligand), five of which were the conventional H-bonding with GLN530, THR529, CYS532, PHE595 and GLY596 of the protein with a bond lengths of 2.32456, 2.73479, 2.56856, 2.43786 and 2.106 respectively, from the ligand and one C–H bond with CYS532. Besides, VAL471, ALA481, and CYS532 formed pi alkyl bonds with the ligand, while TRP531 and PHE583 formed pi-pi stacked interaction with the ligand.

The docked structure of AN2 shown in Figure 6 indicates a negative free binding energy of ($-12.1 kcal mol^{-1}$), suggesting that binding is practicable, because most of the interactions (energies) are of H-bond type with these amino acids (ASP594, THR529, PHE595, and GLY596), thus ensuing in the total negative value. The complex stability result can

**Table 6.** Molecular interactions of V600E-BRAF (PDB ID: 3OG7) with some designed Anti-melanoma compounds.

| Molecular System | Binding Free Energy (Kcal/mol) | Hydrogen Bond (HB) | Bond Length (Å) for HB | Alkyl | Pi-Sigma | Pi-Pi | Pi-Alkyl | Pi-sulphur/ cation | C–H |
|---|---|---|---|---|---|---|---|---|---|
| BRAF/AN1 | −11.9 | GLN530<br>THR529<br>CYS532<br>PHE595<br>GLY596 | 2.32456<br>2.73479<br>2.56856<br>2.43786<br>2.106 | | | PHE583<br>TRP531 | VAL471<br>ALA481<br>CYS532 | | CYS532 |
| BRAF/AN2 | −12.1 | ASP594<br>THR529<br>PHE595<br>GLY596 | 2.50828<br>2.8518<br>2.41353<br>2.1919 | | VAL471 | TRP531<br>PHE583 | PHE468<br>VAL471<br>ALA481<br>CYS532 | | ASN580<br>LYS483 |
| BRAF/AC2 | −11.6 | CYS53<br>SER536<br>ASN580 | 2.11094<br>2.75709<br>2.13187 | VAL471<br>ALA481<br>LEU514<br>CYS532<br>LEU514 | PHE583 | TRP531<br>PHE583 | VAL471<br>ALA481<br>ILE463<br>PHE583 | CYS532<br>TRP531 | SER465<br>GLY534<br>ILE463<br>GLY466<br>SER535 |
| BRAF/AC4 | −12.4 | THR529<br>LYS578<br>LYS578<br>ASN580<br>ARG662<br>THR529<br>GLN530<br>THR529<br>GLN530 | 2.35841<br>2.26453<br>2.47978<br>2.51789<br>2.61203<br>2.6315<br>2.52962<br>2.68757<br>2.40948 | ILE617<br>ARG662 | | TRP531 | PHE468 | PHE583<br>VAL471<br>ALA481 | LYS483<br>SER616<br>ASN581<br>ASP594<br>SER465 |
| BRAF/Vemurafenib | −11.3 | CYS532<br>GLN530ASP594<br>PHE595<br>GLY596 | 3.04242<br>2.44521<br>2.29258<br>2.67992<br>2.14527 | CYS532 | | TRP531<br>PHE583 | TRP531<br>PHE583<br>ALA481<br>LEU514<br>CYS532<br>LYS483<br>ILE463 | LYS483 | |

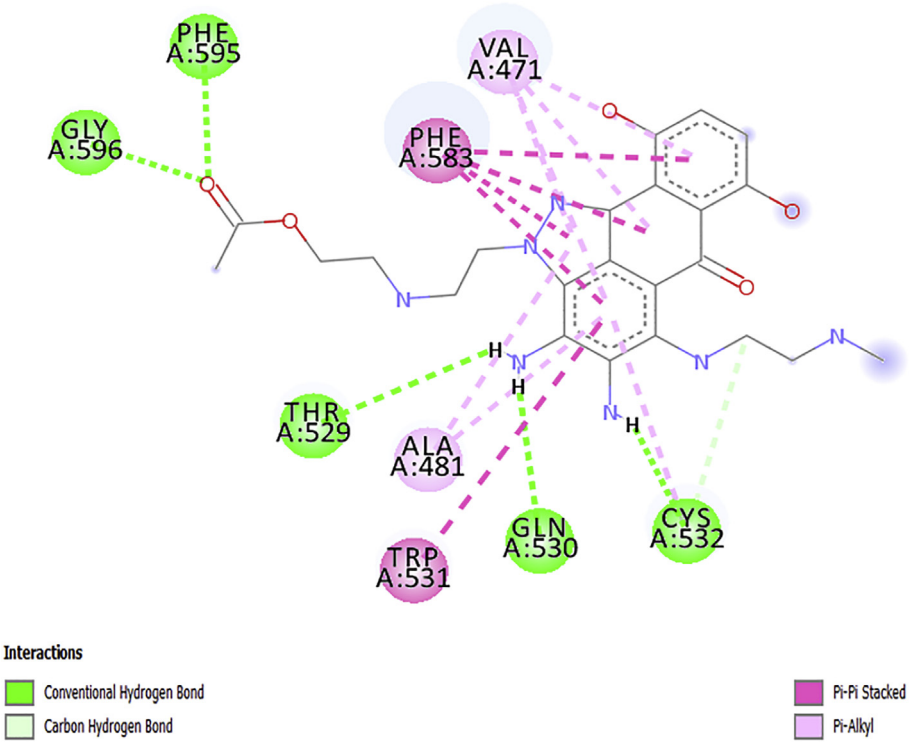


**Figure 5.** 2D interactions diagram of AN1 with V600E-BRAF.

be connected with an extra Pi/sigma interaction associated with VAL471, Pi/alkyl interactions associated with (PHE468, VAL471, ALA481, and CYS532) and Pi-Pi interactions (TRP531 and PHE583) as reported in Table 6.

The docked structure of AC2 depicted in Figure 7 shows negative free binding energy of ($-11.6 kcalmol^{-1}$) which indicates the possibility of stable interactions between the ligand and the protein target. There are three conventional Hbonds identified between this ligand and receptor;

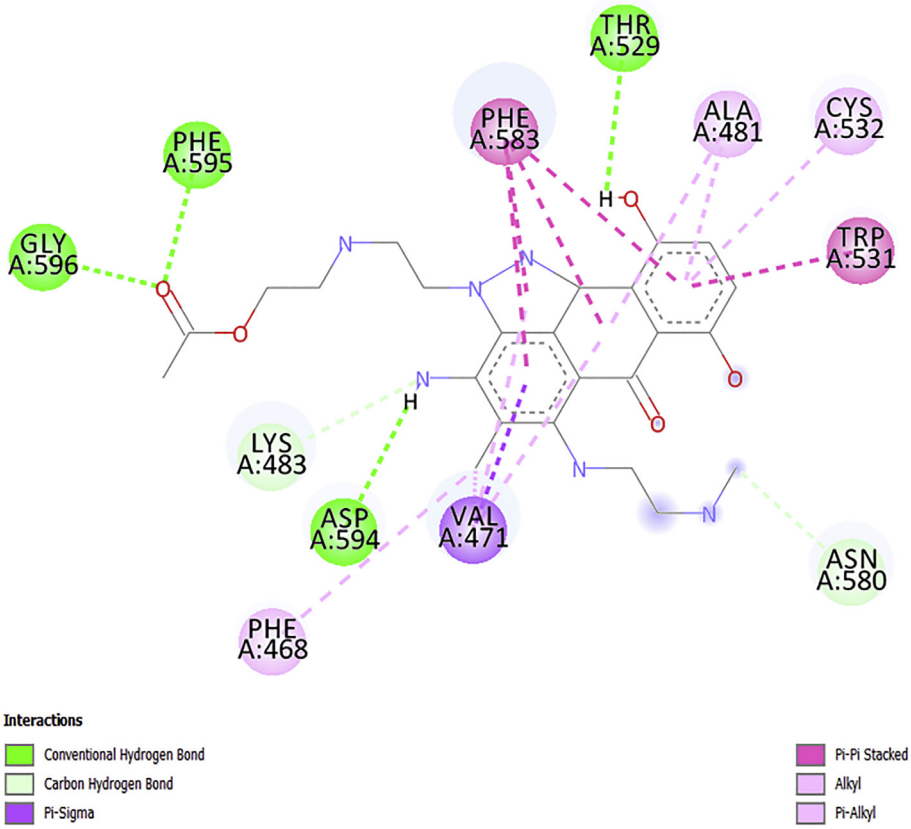


**Figure 6.** 2D interactions diagram of AN2 with V600E-BRAF.

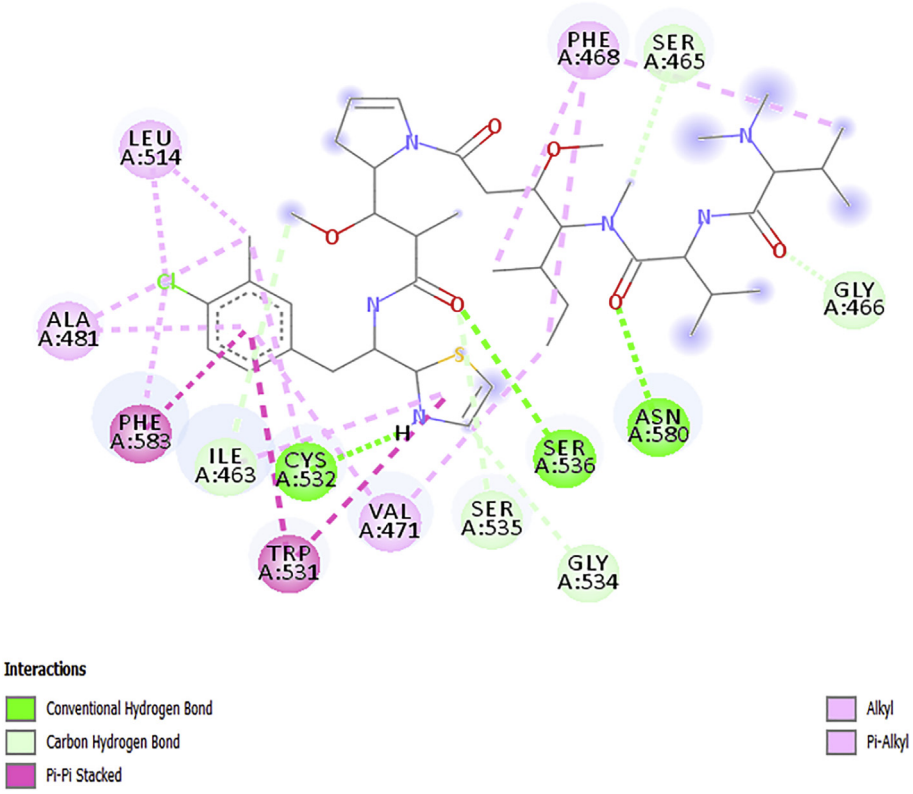


**Figure 7.** 2D interactions diagram of AC2 with V600E-BRAF.

CYS53, SER536 and ASN580 and five carbon-hydrogen with SER465, GLY534, ILE463, GLY466, and SER535. The complex stability may be attributed with an extra alkyl interaction with VAL471, LEU514 and CYS532, Pi-sigma interaction with PHE583, Pi-alkyl interactions (PHE583), Pi-Pi interactions (TRP531 and PHE583) and additional Pi-Sulphur interactions with CYS532, TRP531 as reported in Table 6.

The free binding energy of AC4 with a receptor is –12.4kcalmol$^{-1}$, this interaction was achieved by nine (9) H-bonds with the ligand and Pi/

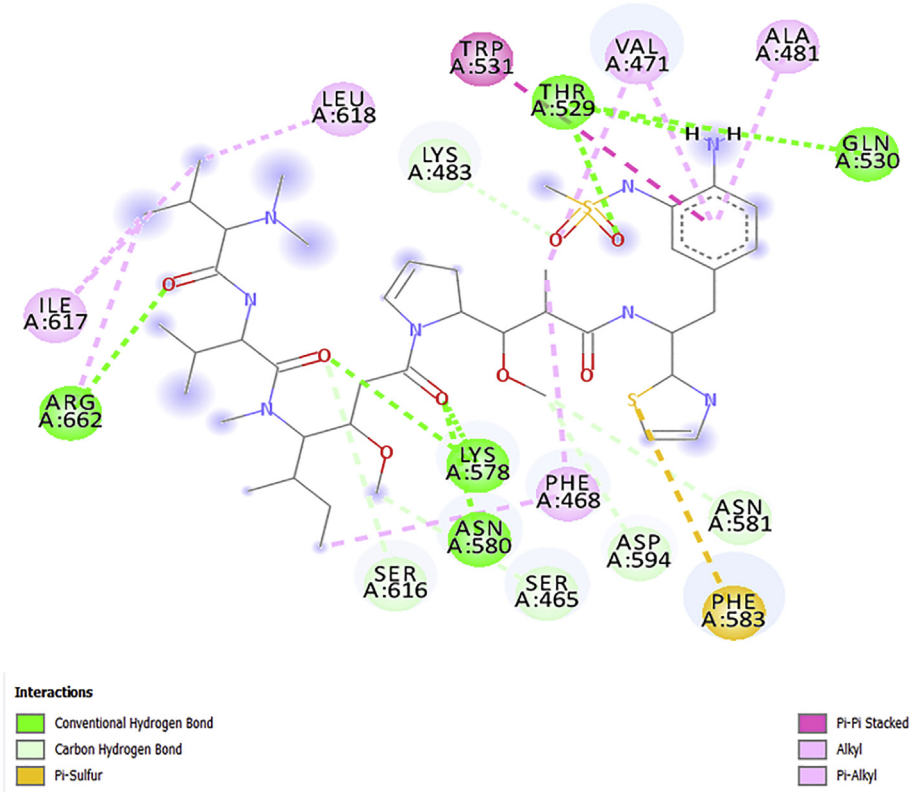


**Figure 8.** 2D interactions diagram of AC4 with V600E-BRAF.

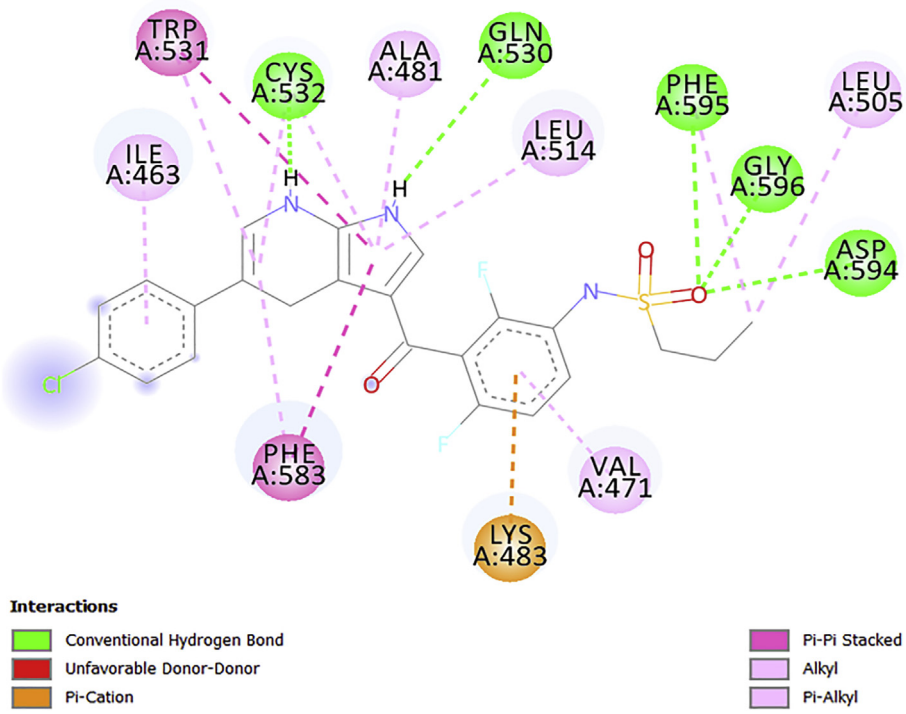


**Figure 9.** 2D interactions diagram of Vemurafenib with V600E-BRAF.

sigma interaction that introduces stabilizing charges responsible for intercalating the drug within the protein (V600E-BRAF) as presented in Figure 8. There were nine (9) conventional H-bonds present in the complex and seven (7) hydrophobic interactions with seven (7) amino acids (Table 6). The Pi/Cation interactions with PHE583, VAL471, and ALA481 which formed a Pi/donor H-bond with the amino acids in the binding segment of the receptor. Other identified interactions are the alkyl interaction with (ILE617 and ARG662), pi/alkyl interaction with PHE468 and Pi/Pi interaction with TRP531 similar to vemurafenib.

It has been reported that H-bonding is the main force regulating the interaction between the docked compounds (ligands) and the receptors and also the binding score of the ligand increases with the number of hydrogen bonding/distance [36, 37]. It can be seen that the number of amino acids involved in the conventional hydrogen bonding with the designed molecules was better than that of vemurafenib as presented in Table 6 and Figures 5, 6, 7, 8, and 9 respectively. This might inform the better free binding affinity of the designed molecules for V600E-BRAF. Additionally, some of the designed molecules interacted more with the target amino acids through strong electrostatic forces especially compound AC4 (Figure 8).

## 4. Conclusion

In this study, a QSAR with molecular docking simulation methods was used to study a series of seventy-two anti-cancer compounds, not only to develop a highly predictive QSAR model but also to identify the key features required to design novel candidates and to explore the interaction mechanism between the compounds and V600E-BRAF protein. The regression coefficients of the model showed an equally good model with sufficient statistical validation keys ($R^2$ (0.864), $R^2_{adjusted}$ (0.845), $Q^2_{cv}$ (0.799) and $R^2_{pred}$ (0.706)) for the internal and external data sets. The information derived from the model suggests that if changes occur in some regions of the molecules, it might be feasible that the inhibitory activity would improve. As for the new templates design, followed the molecular modification approach in the existing data set, molecule 30 and 41 were selected because they had relatively high anti-melanoma activity; they are within the AD of the QSAR model and adjustment can be done easily around their benzene ring. Based on the results, we predicated the $pGI_{50}$ activity of proposed molecules by the built QSAR model. The predicted activities of most of the designed molecules were found to be better than the templates and Compound AC4 was found to be the most active ($pIG_{50}$ 12.014) within all designed compounds. Further, Molecular docking simulation was applied to better understand the binding mechanism and produce the binding poses of some selected proposed compounds (AN1, AN2, AC2, and AC4) into the V600E-BRAF receptor. The newly designed compounds were proved to have a better binding score for V600E-BRAF (PDB CODE: 3OG7) compare to vemurafenib. Thus, these new proposed compounds would be better to investigate the chance of the specificity and higher selectivity toward V600E-BRAF protein. These findings may be helpful in early anti-cancer drug development.

## Declarations

### *Author contribution statement*

Adamu Uzairu: Conceived and designed the experiments.
Abdullahi Bello Umar: Performed the experiments; Wrote the paper.
Gideon Adamu Shallangwa, Sani Uba: Analyzed and interpreted the data.

### *Funding statement*

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### *Competing interest statement*

The authors declare no conflict of interest.

## Additional information

No additional information is available for this paper.

## Acknowledgements

The National Cancer Institute (NCI) was sincerely acknowledged for providing the data used in this research and also Ahmadu Bello University, Zaria for providing the soft wares adopted for the research.

## References

[1] N.A. Stueven, N.M. Schlaeger, A.P. Monte, S.-P.L. Hwang, C.-c. Huang, A novel stilbene-like compound that inhibits melanoma growth by regulating melanocyte differentiation and proliferation, Toxicol. Appl. Pharmacol. 337 (2017) 30–38.
[2] A. Marra, C.R. Ferrone, C. Fusciello, G. Scognamiglio, S. Ferrone, S. Pepe, F. Perri, F. Sabbatino, Translational research in cutaneous melanoma: new therapeutic perspectives, Anti Canc. Agents Med. Chem. 18 (2) (2018) 166–181.
[3] M. Mioc, I.Z. Pavel, R. Ghiulai, D.E. Coricovac, C. Farcaş, C.-V. Mihali, C. Oprean, V. Serafim, R.A. Popovici, C.A. Dehelean, The cytotoxic effects of betulin-conjugated gold nanoparticles as stable formulations in normal and melanoma cells, Front. Pharmacol. 9 (2018) 429.
[4] N. Theodosakis, G. Micevic, C.G. Langdon, A. Ventura, R. Means, D.F. Stern, M.W. Bosenberg, p90RSK blockade inhibits dual BRAF and MEK inhibitor-resistant melanoma by targeting protein synthesis, J. Invest. Dermatol. 137 (10) (2017) 2187–2196.
[5] O. Usta, W. McCarty, S. Bale, M. Hegde, R. Jindal, A. Bhushan, I. Golberg, M. Yarmush, Microengineered cell and tissue systems for drug screening and toxicology applications: evolution of in-vitro liver technologies, Technology 3 (1) (2015) 1–26.
[6] S. Kraljevic, P.J. Stambrook, K. Pavelic, Accelerating drug discovery, EMBO Rep. 5 (9) (2004) 837–842.
[7] R.A. Lionberger, FDA critical path initiatives: opportunities for generic drug development, AAPS J. 10 (1) (2008) 103–109.
[8] C. Yap, Y. Xue, Y. Chen, Application of support vector machines to in silico prediction of cytochrome p450 enzyme substrates and inhibitors, Curr. Top. Med. Chem. 6 (15) (2006) 1593–1607.
[9] S. Satbhaiya, O. Chourasia, Scaffold and cell line based approaches for QSAR studies on anticancer agents, RSC Adv. 5 (103) (2015) 84810–84820.
[10] A. Kamal, E.V. Bharathi, M.J. Ramaiah, D. Dastagiri, J.S. Reddy, A. Viswanath, F. Sultana, S. Pushpavalli, M. Pal-Bhadra, H.K. Srivastava, Quinazolinone linked pyrrolo [2, 1-c][1, 4] benzodiazepine (PBD) conjugates: design, synthesis and biological evaluation as potential anticancer agents, Bioorg. Med. Chem. 18 (2) (2010) 526–542.
[11] F. Xie, H. Zhao, L. Zhao, L. Lou, Y. Hu, Synthesis and biological evaluation of novel 2, 4, 5-substituted pyrimidine derivatives for anticancer activity, Bioorg. Med. Chem. Lett 19 (1) (2009) 275–278.
[12] Y.C. Martin, 3D QSAR: current state, scope, and limitations. 3D QSAR in Drug Design, Springer, 1998, pp. 3–23.
[13] P.R. Ashton, M.C. Fyfe, S.K. Hickingbottom, J.F. Stoddart, A.J. White, D.J. Williams, Hammett correlations 'beyond the molecule'1, J. Chem. Soc. Perkin Transact. 2 (10) (1998) 2117–2128.
[14] A.S. Reddy, S.P. Pati, P.P. Kumar, H. Pradeep, G.N. Sastry, Virtual screening in drug discovery-a computational perspective, Curr. Protein Pept. Sci. 8 (4) (2007) 329–351.
[15] P. Srivani, G.N. Sastry, Potential choline kinase inhibitors: a molecular modeling study of bis-quinolinium compounds, J. Mol. Graph. Model. 27 (6) (2009) 676–688.
[16] R. Benigni, A. Giuliani, Putting the predictive toxicology challenge into perspective: reflections on the results, Bioinformatics 19 (10) (2003) 1194–1200.
[17] C. Hansch, A. Leo, S.B. Mekapati, A. Kurup, Qsar and adme, Bioorg. Med. Chem. 12 (12) (2004) 3391–3400.
[18] H.K. Srivastava, M. Chourasia, D. Kumar, G.N. Sastry, Comparison of computational methods to model DNA minor groove binders, J. Chem. Inf. Model. 51 (3) (2011) 558–571.
[19] J. Verma, V.M. Khedkar, E.C. Coutinho, 3D-QSAR in drug design-a review, Curr. Top. Med. Chem. 10 (1) (2010) 95–115.
[20] C.W. Yap, PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints, J. Comput. Chem. 32 (7) (2011) 1466–1474.
[21] K. Rajer-Kanduč, J. Zupan, 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, Chemometr. Intell. Lab. Syst. 65 (2) (2003) 221–229.
[22] R.W. Kennard, L.A. Stone, Computer aided design of experiments, Technometrics 11 (1) (1969) 137–148.
[23] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, Mol. Informat. 22 (1) (2003) 69–77.
[24] R. Leardi, Genetic algorithms in feature selection. Genetic Algorithms in Molecular Modeling, Elsevier, 1996, pp. 67–86.
[25] M.S. Brose, P. Volpe, M. Feldman, M. Kumar, I. Rishi, R. Gerrero, E. Einhorn, M. Herlyn, J. Minna, A. Nicholson, BRAF and RAS mutations in human lung cancer and melanoma, Canc. Res. 62 (23) (2002) 6997–7000.
[26] G. Bollag, P. Hirth, J. Tsai, J. Zhang, P.N. Ibrahim, H. Cho, W. Spevak, C. Zhang, Y. Zhang, G. Habets, Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma, Nature 467 (7315) (2010) 596.
[27] W.-K. Choi, M.I. El-Gamal, H.S. Choi, D. Baek, 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, Eur. J. Med. Chem. 46 (12) (2011) 5754–5762.
[28] W. Wu, C. Zhang, W. Lin, Q. Chen, X. Guo, Y. Qian, L. Zhang, Quantitative structure-property relationship (QSPR) modeling of drug-loaded polymeric micelles via genetic function approximation, PloS One 10 (3) (2015), e0119575.
[29] B.A. Umar, A. Uzairu, G.A. Shallangwa, 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, Netw. Model. Anal. Health Informat. Bioinformat. 8 (1) (2019) 22.
[30] M. Jalali-Heravi, E. Konuze, Use of quantitative structure property relationships in predicting the Kraft point of anionic surfactants, Elec. J. Mol. Des. 1 (2002) 410–417.
[31] L. Eriksson, J. Jaworska, A.P. Worth, M.T. Cronin, R.M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs, Environ. Health Perspect. 111 (10) (2003) 1361–1375.
[32] P. Gramatica, E. Giani, E. Papa, Statistical external validation and consensus modeling: a QSPR case study for Koc prediction, J. Mol. Graph. Model. 25 (6) (2007) 755–766.
[33] J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg, QSAR applicability domain estimation by projection of the training set descriptor space: a review, ATLA-NOTTINGHAM- 33 (5) (2005) 445.
[34] I. Muegge, S. Oloff, Advances in virtual screening, Drug Discov. Today Technol. 3 (4) (2006) 405–411.
[35] G. Melagraki, A. Afantitis, H. Sarimveis, P.A. Koutentis, G. Kollias, O. Igglessi-Markopoulou, Predictive QSAR workflow for the in silico identification and screening of novel HDAC inhibitors, Mol. Divers. 13 (3) (2009) 301–311.
[36] O. Adedirin, A. Uzairu, G.A. Shallangwa, S.E. Abechi, Optimization of the anticonvulsant activity of 2-acetamido-N-benzyl-2-(5-methylfuran-2-yl) acetamide using QSAR modeling and molecular docking techniques, Beni-Suef Univ. J. Basic Appl. Sci. 7 (4) (2018) 430–440.
[37] B.A. Umar, A. Uzairu, G.A. Shallangwa, S. Uba, Rational drug design of potent V600E-BRAF kinase inhibitors through molecular docking simulation, J. Eng. Exact Sci. 5 (5) (2019) 469–481.