Computational modeling of novel quinazoline derivatives as potent epidermal growth factor receptor inhibitors

Muhammad Tukur Ibrahim *, Adamu Uzairu, Sani Uba, Gideon Adamu Shallangwa

Department of Chemistry, Faculty of Physical Science, Ahmadu Bello University, P.M.B 1045, Zaria, Kaduna State, Nigeria

ABSTRACT

QSAR modelling on Thirty (34) novel quinazoline derivatives (EGFRWT inhibitors) as non-small cell lung cancer (NSCLC) agents was performed to develop a model with good predictive power that can predict the activities of newly designed compounds that have not been synthesised. The EGFRWT inhibitors were optimized at B3LYP/6-31G* level of theory using Density Functional Theory (DFT) method. Multi-Linear Regression using Genetic Function Approximation (GFA) method was adopted in building the models. The best one among the models built was selected and reported because it was found to have passed the minimum requirement for the assessment of QSAR models with the following assessment parameters: $R^2$ of 0.965901, $R^2_{adj}$ of 0.893733, $Q^2_{cv}$ of 0.940744, $R^2_{test}$ of 0.818991 and LOF of 0.076739. The high predicted power, reliability, robustness of the reported model was verified further by subjecting it to other assessments such VIF, Y-scrambling test and applicability domain. Molecular docking was also employed to elucidate the binding mode of some selected EGFRWT inhibitors against EGFR receptor (4ZAU) and found that molecule 17 have the highest binding affinity of -9.5 kcal/mol. It was observed that the ligand interacted with the receptor via hydrogen bond, hydrophobic bond, halogen bond, electrostatic bond and others which might me the reason why it has the highest binding affinity. Also, the ADME properties of these selected molecules were predicted and only one molecule (34) was found not orally bioavailable because it violated more than the permissible limit set by Lipinski’s rule of five filters. This findings proposed a guidance for designing new potents EGFRWT inhibitors against their target enzyme.

1. Introduction

Among the foremost cancer issues that results in loss of lives each year in the globe which was estimated for almost one-third of the entire cancer deaths is lung cancer. Non-small cell lung cancer (NSCLC) is the principal subset of lung cancers that estimates for about 85% of the problems raised above [1]. The most common cause of NSCLCs recognised was EGFR kinase. It was found in about 10–15% and 30–40% of the population of patients in Caucasia and Asia. It mostly affects women and cigarette smokers in general [1].

Development of inhibitors for mutant-selective kinase is among the difficulties faced in medicinal chemistry and is the principal concern for EGFR tyrosine kinase inhibitors [2]. The remedy of epidermal growth factor receptor (EGFR) to managed non-small cell lung cancers with the T790M resistance mutation prevails a vital medical necessity [3].

In patients with stimulating modifications of EGFR, EGFR inhibitors show a very high response rate. EGFR inhibitors are categorised into two classes: First generation EGFR inhibitors which are referred to reversible inhibitors and include gefitinib and erlotinib. The second class which consist the second and third generation EGFR inhibitors. The second and third generation EGFR inhibitors are referred to as irreversible inhibitors (examples are afatinib and osimertinib). All these classes of drugs were designed to mitigate the problem of NSCLC most especially the EGFR$^{L858R}$ mutations (First generation EGFR inhibitors were designed to treat this type of mutation), EGFR$^{T790M}$ mutations (while Second generation EGFR inhibitors were designed for the treatment of this type of mutation) and EGFR$^{T790M/L858R}$ double mutations (third generation EGFR inhibitors) were designed to treat this type of mutation [2, 4, 5, 6].

QSAR modelling is a molecular modeling method which quantitatively correlate response variable (biological activities) and molecular descriptors (physicochemical properties) of a molecule [7]. In addition, the QSAR technique of computer-aided drug design plays a significant role in predicting the biological activities of small molecules that have not been synthesised [8]. Another virtual screening method applied in computer aided drug design is molecular docking which give an overview of how the ligand and the receptor interact with one another using their
2.2. Sketching of structures and optimum structure generations

After data collection, the sketching of the 2D structures of the studied molecules was achieved using Chemdraw software version 12.0.2 [13]. After sketching the 2D-structures of the dataset, Spartan 14 software was used to convert the 2D-structures to 3D-structures before energy minimization. Energy minimizing was performed to reduce constrain in the structures before geometry optimization. Geometry optimization is a process of finding the most optimum structure of a molecule on potential energy surface and this was performed by utilizing Spartan 14 software. DFT at B3LYP/6-311G* level of theory was used in finding the most optimum structures of all the studied molecules on global minima on the potential energy surface (PES) [14].

2.3. Descriptors computation, data pre-treatment and dataset splitting

In order to compute the independent variables (descriptors), the most optimum structures obtained in 2.2 above were saved in SDF file format that is been recognized by the software used in computing the descriptors, PaDEL descriptor tool kit. PaDEL descriptor tool kit was used to compute both Fragment count descriptors, Topological descriptors and Geometrical descriptors [15].

To eliminate redundant and constant descriptors, data pre-treatment was performed manually in this regard. After pre-treating the data, Data division software was further used in splitting the data into model building and validation set utilizing Kennard-Stone algorithm [16]. The model building which comprise 24 molecules (70%) were used for the generation of the models as the name implies and the validation set which contain 10 molecules (30%) were used for the assessment of the generated models [17].

2.4. Building of the model

The models were built utilizing Genetic Function Approximation (GFA) method with the descriptors as independent parameter and the actual pIC50 as the response parameter. In the case of variable selection, the GFA creates an original population of descriptor sets and determines the most suitable set from it by utilizing evolutionary crossover and mutation speculators which generates a succeeding derivative population of descriptor sets. GFA select most highly correlated descriptors to develop so many models which is one of the distinct characteristic of GFA [18]. The MLR-GFA equation for the model is shown below:

\[ p\text{IC}50 = -\log IC50 \times 10^{-9} \]

where X’s are the descriptors, y’s are the co-efficient of the corresponding descriptors and z is the regression constant.

2.1. Dataset source

Thirty four (34) quinazoline derivatives bearing various 6-benzamide moieties as potent EGFRWT inhibitors with their inhibitory activities (IC50) in nM were selected from the work of Hou et al., for this research [11]. The inhibitory activities (IC50) of all the dataset were then converted to their corresponding negative logarithms (pIC50) using Eq. (1) [12]. Table 1 presents the structures, IC50 and pIC50 for all the data set used in this research.

\[ p\text{IC}50 = -\log IC50 \times 10^{-9} \]

2. In-silico computational method

### Table 1. The Formula, IC50 and pIC50 of the data set.

| S.No. | Formula | IC50 nM | pIC50 (nM) |
|-------|---------|---------|------------|
| D1    | C29H27ClFN5O3 | 27 | 7.568636 |
| D2    | C29H27ClFN5O3 | 29 | 7.537602 |
| D3    | C29H27ClFN5O3 | 13 | 7.886057 |
| D4    | C29H27ClFN5O3 | 30 | 7.522879 |
| D5    | C29H27ClFN5O3 | 24 | 7.619789 |
| D6    | C29H27ClFN5O3 | 14 | 7.853872 |
| D7    | C29H27ClFN5O3 | 5.0 | 8.30103 |
| D8    | C29H27ClFN5O3 | 1.3 | 8.886057 |
| D9    | C29H27ClFN5O3 | 57 | 7.244125 |
| D10   | C29H27ClFN5O3 | 8.3 | 8.080922 |
| D11   | C29H27ClFN5O3 | 43 | 7.366532 |
| D12   | C29H27ClFN5O3 | 365 | 6.437707 |
| D13   | C29H27ClFN5O3 | 3 | 8.522879 |
| D14   | C29H27ClFN5O3 | 56.9 | 7.293282 |
| D15   | C29H27ClFN5O3 | 4.3 | 8.366532 |
| D16   | C29H27ClFN5O3 | 242.4 | 6.615467 |
| D17   | C29H27ClFN5O3 | 44 | 7.356547 |
| D18   | C29H27ClFN5O3 | 68 | 7.167491 |
| D19   | C29H27ClFN5O3 | 2.6 | 8.585027 |
| D20   | C29H27ClFN5O3 | 6.6 | 8.180456 |
| D21   | C29H27ClFN5O3 | 21 | 7.677781 |
| D22   | C29H27ClFN5O3 | 13 | 7.886057 |
| D23   | C29H27ClFN5O3 | 50 | 7.30103 |
| D24   | C29H27ClFN5O3 | 9.2 | 8.036212 |
| D25   | C29H27ClFN5O3 | 6.3 | 8.200659 |
| D26   | C29H27ClFN5O3 | 53 | 7.275724 |
| D27   | C29H27ClFN5O3 | 722 | 6.141463 |
| D28   | C29H27ClFN5O3 | 2426 | 5.615109 |
| D29   | C29H27ClFN5O3 | 172 | 6.764472 |
| D30   | C29H27ClFN5O3 | 503 | 6.298432 |
| D31   | C29H27ClFN5O3 | 374 | 6.427128 |
| D32   | C29H27ClFN5O3 | 390 | 6.408935 |
| D33   | C29H27ClFN5O3 | 169 | 6.727113 |
| D34   | C29H27ClFN5O3 | 39 | 7.408935 |

individual 3D structures [9]. To have an insight on how body response to the administration of drugs there is need to study the ADME and drug likeness of the drugs before it reaches the final (clinical) stage [10].

The aim of this work is to develop a model with good predictive power which could be used to predict the inhibitory activities of newly designed compounds using QSAR technique, study the mode of binding interactions between some selected EGFRWT Inhibitors and EGFR enzyme via docking and also to predict the ADME properties of these selected EGFRWT Inhibitors.

2. In-silico computational method

### Table 2. General limit required for the QSAR model assessment.

| Symbol | Name | Recommended Value | Reported Model |
|--------|------|-------------------|----------------|
| R²     | Co-efficient of determination | ≥0.6 | 0.965901 |
| Q²     | Cross-Validation Co-efficient | ≥0.5 | 0.940744 |
| R²-C | Difference between R² and Q² | ≤0.3 | 0.025157 |
| N(test, A & test set) | Minimum number of external and test set | ≥5 | 34 |
| R²ext | Co-efficient of determination of external and test set | ≥0.5 | 0.818991 |
### 2.5. Assessment of the model built

The assessment parameters used in evaluating or validating the quality of a QSAR model are the: Squared correlation coefficient of the training set ($R^2_{\text{training}}$), Adjusted $R^2$ ($R^2_{\text{adj}}$), Cross-validation coefficient ($Q^2_{\text{loocv}}$), and Squared correlation coefficient of the test set ($R^2_{\text{test}}$) [19, 20, 21].

The large value of these parameters seem to be important but not enough [22]. In this regard, the multi-collinearity between descriptors can be identified using their variation inflation factors (VIF), to identify whether these descriptors correlate with each other or not. If the estimated VIF values are equal to 1 it means there is no correlation between them; if it happens to be between 1–5, there is high chance of accepting the model; and if it is greater than 10, the model cannot be accepted is therefore rejected [23]. It can be calculated using the equation below:

$$VIF = \frac{1}{1 - R^2}$$  

(3)

The assessment of importance and participation of each descriptor to the selected model is made using the value of the mean effect (ME) of each descriptor. The equation used in calculating the ME is shown below:

$$ME_j = \frac{B_j \sum_{i=1}^{n} x_{ij}^2}{\sum_{i=1}^{n} B_j x_{ij}^2}$$  

(4)

where ME represents the mean effect of a descriptor $j$ in a model, the coefficient of the descriptor $j$ is represented by $B_j$ in the model and the value of the descriptor in the data matrix for each molecule in the model building set is $x_{ij}$, $n$ is the number of molecules in the model building set and $m$ is the number of descriptor that appear in the model [24].

Y-Scrambling test was performed to assure the robustness of a model and also the model was not achieved by chance correlation. It is done by reshuffling the actual activities and holding the descriptors fixed to generate new QSAR models for many trials, the new built QSAR models were anticipated to give low $Q^2$ and $R^2$ value. The validation parameter for this test is $cR^2_p$ ($cR^2_p > 0.5$) [25].

### 2.6. Applicability domain

The applicability domain (AD) of a model was carried out to determine whether a model can be regarded valid and void if the model can make a good prediction of new activities of the training and test molecules. As such, the model is subjected to AD to find out whether there are influential or outliers molecules in the studied ones [26]. Leverage approach is among the methods used in assessing the AD of QSAR models and thus is given as $h_i$:

$$h_i = \frac{1}{n} \left( X^T X \right)^{-1} x_i (i=A, ..., Z)$$  

(5)

where the model building set matrix $I$ is given by $x_o \times k$ descriptor matrix of the model building set is represented by $X$ and $X^T$ is the transpose matrix $X$ used in generating the model. The threshold for the value of $X$ is the warning threshold ($h^*$) which is presented in the equation below:

$$h^* = 3(x+1)/q$$  

(6)

where the number of chemicals of the model building set is given by $q$, and the number of the descriptors in the model under evaluation is represented by $x$.

### 2.7. Molecular docking analysis

To elucidate the mode of binding interactions between the active site of EGFR enzyme and some selected EGFR$^{\text{WT}}$ inhibitors (ligands), a Dell Latitude E6520 computer system, with the following specification: Intel® Core™ i7 Dual CPU,M330 @2.75 GHz 2.75GHz, 8GB of RAM was

### Table 3. The symbols, descriptions and classes of descriptors for the selected model.

| S/no | Symbol   | Description                                              | Class |
|------|----------|----------------------------------------------------------|-------|
| 1    | ATSC0m   | Centered Broto-Moreau autocorrelation - lag 6/weighted by mass | 2D    |
| 2    | ATSC0e   | Centered Broto-Moreau autocorrelation - lag 8/weighted by Sanderson electronegativities | 2D    |
| 3    | MATS7m   | Moran autocorrelation - lag 7/weighted by mass            | 2D    |
| 4    | SpMax3_Bhp | Largest absolute eigenvalue of Burden modified matrix - n 3/weighted by relative polarizabilities | 2D    |
| 5    | SpMax5_Bhs | Largest absolute eigenvalue of Burden modified matrix - n 5/weighted by relative I-state | 2D    |
| 6    | maxHIb10 | Maximum E-State descriptors of strength for potential Hydrogen Bonds of path length 10 | 2D    |

### Table 4. The pIC50, Predicted pIC50 and the residual values for the studied molecules.

| S/No | pIC50 (nM) | Predicted pIC50 | Residual values |
|------|------------|----------------|----------------|
| 1    | 7.568636   | 7.567486       | 0.00115        |
| 2    | 7.537602   | 7.606153       | -0.06855       |
| 3    | 7.568636   | 7.567486       | 0.00115        |
| 4    | 7.522879   | 7.608232       | -0.08535       |
| 5    | 7.19789    | 7.674599       | 0.05477        |
| 6    | 7.853872   | 7.912141       | -0.05827       |
| 7    | 8.30103    | 8.065812       | 0.235218       |
| 8    | 8.886057   | 8.031683       | -0.85437       |
| 9    | 7.244125   | 7.727673       | 0.483548       |
| 10   | 8.089222   | 8.081978       | -0.00826       |
| 11   | 7.36532    | 7.364126       | 0.002406       |
| 12   | 6.437707   | 6.340232       | 0.097457       |
| 13   | 8.522879   | 8.30143        | 0.221449       |
| 14   | 7.292822   | 7.433613       | -0.14033       |
| 15   | 8.36532    | 8.392319       | -0.02579       |
| 16   | 6.15467    | 6.52079        | 0.094747       |
| 17   | 7.356547   | 7.367566       | -0.01111       |
| 18   | 7.167491   | 7.755543       | 0.588052       |
| 19   | 8.585027   | 8.461748       | 0.123279       |
| 20   | 8.180456   | 8.16481        | 0.015646       |
| 21   | 7.77781    | 7.640802       | 0.036979       |
| 22   | 7.886057   | 7.873194       | 0.012166       |
| 23   | 7.30103    | 7.32366        | 0.03163        |
| 24   | 8.036212   | 8.086822       | -0.04462       |
| 25   | 8.206659   | 8.331855       | -0.1312        |
| 26   | 7.275724   | 7.494222       | -0.2185        |
| 27   | 6.141463   | 6.35151        | 0.210047       |
| 28   | 5.615109   | 5.551742       | -0.06357       |
| 29   | 6.764672   | 6.226603       | -0.33787       |
| 30   | 6.298432   | 6.288255       | 0.010177       |
| 31   | 6.427128   | 6.061555       | -0.36557       |
| 32   | 6.408935   | 6.312643       | 0.096292       |
| 33   | 6.772113   | 6.892031       | -0.11992       |
| 34   | 7.408935   | 7.44403        | -0.0351        |

* z = Test set.
Utilized with the help of Pyrex virtual screening software, Chimera, PyMOL and Discovery studio.

2.8. Ligands and EGFR enzyme preparation for the molecular docking computational analysis

The first thing to do in any molecular docking analysis is ligands preparation. The preparation of the ligands was adopted from the optimized structures in 2.2 above saved in pdb file format using Spartan '14 wave software [27]. The next thing ought to be done is the retrieval of 3D structure of the EGFR enzyme to be used in this study. The EGFR enzyme with pdb code: 4zau was downloaded from the Protein Data Bank (RSCPDB). Discovery Studio Visualizer was utilized in preparing the EGFR enzyme for the docking analysis, in the course of the preparation, hydrogen was added, water molecule, heteroatoms and co-ligands present on the crystal structure were completely eliminated and saved in pdb file.

2.9. Execution of the molecular docking computational analysis

Autodock vina of Pyrex software was used for the docking of the ligands to the active site of EGFR enzyme (pdb ID: 4zau) [28]. Re-coupling of the ligand-receptor (complexes) for further investigation was done with the help of Chimera software [29]. The elucidation of the binding mode interactions of the complexes was achieved using PyMOL and Discovery studio visualizer [30, 31].

2.10. ADME and drug-likeness properties prediction

SwissADME a free online web tool used in evaluating ADME and drug-likeness properties of small molecules was used to predict the ADME and drug-likeness properties of some selected EGFRWT inhibitors among the data set [32]. SMILES is the input file for SwissADME which contains a molecule per line separated by a space with a name (optional). Molecules can be inserted in SMILES format or pasted, or drawn using the molecular sketcher available in the web tool. If the molecule is ready, the calculations can be setup by clicking on the “Run” button [32].

The Lipinski's rule of five filter is very useful at pre-clinical stage of drug discovery which state that if any compound violate more than 2 of these criteria (Molecular weight < 500, Number of hydrogen bond donors ≤5, Number of hydrogen bond acceptors ≤10, Calculated Log p ≤5 and Polar surface area (PSA) <140 Å²), the compound is said to be impermeable or badly absorbed [33].

3. Result and discussion

3.1. QSAR modeling

The reported model was observed to have excelled the limit for the evaluation of a good model with the following evaluation parameters: R²
The negative coefficients of these descriptors (ATSC8e and SpMax5_Bhs) highlighted their negative correlation to the inhibitory activities of the quinazoline derivatives (EGFRWT inhibitors). The lesser the number of these descriptors in the structures of these EGFRWT inhibitors, the higher the potency of these EGFRWT inhibitors toward their target EGFRWT enzyme. On the other hand, the positive coefficient of ATSC6m, MATS7m, SpMax3_Bhp and maxHBint10 descriptors in the reported model gives the positive correlation of these descriptors to the inhibitory activities of EGFRWT inhibitors. That is, the more the presence of these types of descriptors in the structures of these EGFRWT inhibitors the more the inhibitory activities of the EGFRWT inhibitors toward their target enzyme.

Table 6. Y-scrambling test.

| Model      | R   | R²  | Q²  |
|------------|-----|-----|-----|
| Original   | 0.888285 | 0.789051 | 0.520506 |
| Random 1   | 0.210234 | 0.044198 | -0.64726 |
| Random 2   | 0.517313 | 0.267612 | -0.4446 |
| Random 3   | 0.591649 | 0.350049 | -0.09228 |
| Random 4   | 0.397103 | 0.157691 | -0.47247 |
| Random 5   | 0.485224 | 0.235442 | -0.76149 |
| Random 6   | 0.521054 | 0.271497 | -0.35543 |
| Random 7   | 0.294695 | 0.086845 | -0.59685 |
| Random 8   | 0.490146 | 0.240243 | -0.30231 |
| Random 9   | 0.393085 | 0.154516 | -0.42651 |
| Random 10  | 0.521333 | 0.271788 | -0.17528 |

Random Models Parameters

Average r: 0.442183
Average r²: 0.207988
Average Q²: -0.42745
cRp²: 0.68434

Table 7. The Ligand-Receptor, Binding affinity, Hydrogen bond, Bond distance, Halogen, Hydrophobic and Other Amino Acid Residues of some selected ligands.

| S/No | Ligand-Receptor (4ZAU) Binding affinity (Kcal/mol) Hydrogen Bond | Bond distance (Å) Halogen, Hydrophobic and Other Amino Acid Residues |
|------|---------------------------------------------------------------|---------------------------------------------------------------------|
| D2   | Complex 2 | -9.2 | GLU762 THR790 MET793 PHE723 | MET793, LYS745, LEU718, VAL726, ILE759, ALA743, LEU844 |
| D8   | Complex 8 | -9.0 | THR790 GLU762 MET793 GLU762 GLU762 | MET793, LYS745, LEU844, LEU718, VAL726, ALA743 |
| D13  | Complex 13 | -9.1 | GLU762 THR790 MET793 PHE723 | MET793, LYS745, LEU718, ILE759, LYS745, VAL726, ALA743, LEU844 |
| D17  | Complex 17 | -9.5 | MET793 MET793 THR854 ASP855 | GLN791, LYS745, LEU718, MET776, CYS797, ALA743, ALA743, VAL726 |
| D34  | Complex 34 | -9.4 | ARG841 ASN842 LYS745 UNK1 | LEU788, ASP855, CYS797, VAL726, LYS745, LEU718, ALA743, LEU844 |

Figure 3. Williams Plot of the selected model.
3.1. Description of the descriptors that appear in the reported model

ATSC6m and ATSC5e are Moreau–Broto autocorrelation of a Topological Structure, ATS (The ATS descriptor is a graph invariant describing how the property considered is distributed along the topological structure). These descriptors can be seen as a special case in which other types of descriptors can also be derived from [34]. This is the most known spatial autocorrelation defined on a molecular graph G as

\[ ATS_i = \sum_{j=1}^{n} A_{i,j} \delta(d_{i,j}) \]

where \( A_{i,j} \) is an adjacency matrix of graph G, \( d_{i,j} \) is the distance between the \( i \)-th and \( j \)-th atoms, \( \delta(d_{i,j}) \) is a function of the distance, and \( n \) is the number of atoms.

MATS7m is a Moran autocorrelation which if applied to a molecular graph. Moran coefficient usually takes value in the interval [-1, 1]. Positive autocorrelation corresponds to positive values of the coefficient whereas negative autocorrelation produces negative values [34]. It can be defined as

\[ I_i = \sum_{j=1}^{n} w_{i,j} (x_i - \bar{x})(x_j - \bar{x}) \]

where \( w_{i,j} \) is a weight function of the distance between the \( i \)-th and \( j \)-th atoms.

SpMax3_Bhp and SpMax5_Bhs are the maximum absolute eigenvalue of Burden modified matrix - n 3/ and - n 5/ weighted by relative I-state and relative polarizabilities, called leading eigenvalue or spectral radius, SpMaxA is the maximum absolute value of the spectrum. These kinds of functions were called by Ivanicui matrix spectrum operators. This eigenvalue has been suggested as an index of molecular branching, this descriptor talks about branching in molecules [35]. As seen from the regression equation and ME values (Table 5), SpMax5_Bhs contributes negatively to the inhibitory activities of the studied molecules. It suggests that reducing chain branching in the studied molecules will improve the inhibitory activities of the studied molecules toward their target enzyme.

MaxHiBint10 is a maximum E-State descriptors of strength for potential hydrogen bonds of path length 10. Based on the regression equation and ME values (Table 5), this descriptor gave the highest contribution toward the inhibitory activities of the studied molecules. Increasing the number of hydrogen bond in the molecules might increase their potency against their target protein.

The XY (Scatter) plot of predicted activities of both the test and training sets against the actual pIC50 was shown in Figure 1A & 1B. The significance of the reported model was confirmed by the distribution of the values around the straight line. Also, the R² values from the plots agree with those of the training and test for the internal and external assessment.

On the other hand, the XY (Scatter) plot of actual pIC50 against the residuals of both the model building and validation sets was shown in Figure 2. The unusual occurrence of these residuals on either side of zero on the plot shows the non-existence of methodological error in the reported model.

The pIC50, Predicted pIC50 and the residual values for all the studied molecules were presented in Table 4. The low residual values noted in the table verified the reliability of the reported model.

The correlation statistical analysis of the descriptors in the reported model was performed (Table 5) and the descriptors were found not to correlate with one another. This shows the high performance of the descriptors utilized in generating the reported model. To further confirm whether there is a similarity or not between the descriptors in the reported model, The VIF values of these descriptors in the model building set were estimated and realized to be less than 2 (Table 5) indicating the applicability of the reported model and thus the descriptors were independent of one another. The ME value (Table 5) gives the contribution of a descriptor in opposition to other ones in the reported model. The signs point the various directions of either increase or decrease in the values of these descriptors which will improve the inhibitory activities of the studied molecules. It is observed that from the model and ME values (Table 5), MaxHiBint10 descriptor gives the highest contribution.

The Y-scrambling test was presented in Table 6 for the 10 randomly generated models and the R² and Q² values for the newly generated random models were determined to be very low. This has affirmed the obtainability of the reported model was not by chance and further confirm its robustness.

The plot of leverages against standardized residuals of both the model building and validation sets (Williams plot) presented in Figure 3 identified two (2) influential compounds from which were all in the validation set. It is very paramount to decipher that these molecules (influential compounds) with leverage value greater than the threshold h²(h² = 0.875) are not put into consideration when designing new EGFRWT inhibitors. These molecules might be structurally different from those used to generate the reported model and, thus may have a different mechanism of action.

3.2. Molecular docking analysis

The mode of binding interactions between the active site of EGFR receptor (4zau) and some selected EGFRWT inhibitors (ligands) was elucidated through molecular docking (Table 7). From Table 7, Complex 17 was identified to have the highest binding affinity of -9.5 kcal/mol. With the help of discovery studio visualizer, the ligand was clearly observed to have interacted with the active site of EGFR receptor via Hydrogen bond with the following amino acid residues MET793, MET793, THR854 and ASP855 with bond distances of 2.61394 (Å), 2.18464 (Å), 2.57601 (Å) and 2.68794 (Å). The interaction was not only via hydrogen bond, it also interacted with the active site of the EGFR receptor via halogen bond (GLN791), hydrophobic bond (LEU718, CYS797, LYS745, ALA743, ALA743, and VAL726), electrostatic bond (LYS745) and others (MET766) which might be the reason why it has the highest binding affinity. The next one identified with good binding affinity after the one mentioned above (complex 17) is complex 34. It interacted in the active site of the receptor through hydrogen bond with ARG841, ASN842, LYS745, UNK1 residues with bond distances of 2.7122 (Å), 2.07811 (Å), 2.54982 (Å) and 3.62317 (Å). It also interacted with the active site of the EGFR receptor through halogen bond (LEU788), hydrophobic bond (VAL726, LYS745, LEU718, ALA743 and LEU844), electrostatic bond (ASP855) and others (CYS797). The rest other three complexes interacted in the active site of the receptor through hydrogen bond, halogen bond, electrostatic and hydrophobic bond as shown in Table 7. Figure 4 & Figure 5 showed the 3D and 2D structures of the complexes. Based on the molecular docking results, the most common amino acids to all of the examined compounds were MET793, LEU718, LYS745 and VAL726 (Table 7). The most important amino acids that might be responsible for the higher binding affinity were CYS797, GLN 791 and ASP855 due to their interaction with the molecules with higher binding affinity (Table 7).

Halo substituted molecules (Complex 17) were found to fit better in the active site of the receptor than those with bulkier substituents (Complex 34) as shown in Figure 5.

On comparing the QSAR and docking results, the molecule with the highest activity was among those having higher binding affinity. This means that there is little correlation between the QSAR and the molecular docking studies.

3.3. ADME properties prediction

The ADMET properties of these selected EGFRWT inhibitors were predicted and presented in Table 8. From Table 8, it can be observed that only one among these molecules violated more than the maximum permissible limit 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 except molecule 34 which has more than 3 violations. In a null shell the remaining four (4) molecules are said have good absorption, low toxicity level, orally bioavailable and permeable properties. The Bioavailability Radar gives an overview of the drug-likeness of all the selected molecule (Figure 6.). The painted pink area shows the range for each properties (Lipophilicity: XLOGP3 between \(-0.7\) and \(+5.0\), size: MW between 150 and 500 g/mol, polarity: TPSA between 20 and 130 Å², solubility: log \(S\) not higher than 6, saturation: fraction of carbons in the sp3 hybridization not less than 0.25, and flexibility: no more than 9 rotatable bonds). Based on this criteria, all the molecules are said to be orally bioavailable except molecule 34 which is too Flexible, Polar, Lipophilic and Insoluble. The plot of WLOGP against TPSA (Boiled-egg plot) to predict gastrointestinal absorption and brain penetration of the selected molecules was shown in Figure 7. It can be seen from the plot that none of the molecules possess the BBB permeant but they are within the GI absorption region.

4. Conclusion

A very high predictive model was developed using QSAR modelling technique on some EGFRWT inhibitors. The reported model was selected and reported because of its fitness with the following assessment parameters: \(R^2_{\text{trng}} = 0.919035\), \(R^2_{\text{adj}} = 0.893733\), \(Q^2_{\text{cv}} = 0.866475\), \(R^2_{\text{test}} = 0.636217\), and \(\text{LOF} = 0.215884\). The high predicted power, reliability, robustness of the reported model was verified by other assessments such as AD, Y-scrambling test and found to be statistically fit. The molecular docking results of the examined compounds showed that CYS797, GLN791 and ASP855 amino acids might be responsible for the higher binding affinity of molecule 17 (-9.5 kcal/mol) and 34 (-9.4 kcal/mol). Also, the
Figure 5. 2D structures of (A) Complex 2, (B) Complex 8, (C) Complex 13, (D) Complex 17 and (E) Complex 34 with bond distances using Discovery studio visualizer.
Table 8. ADME properties.

| S/N | MW     | HB donor | HB acceptor | WLOGP | TPSA  | Lipinski violations |
|-----|--------|----------|-------------|-------|-------|---------------------|
| D2  | 571.98 | 2        | 7           | 6.49  | 88.61 | 1                   |
| D8  | 579    | 2        | 9           | 5.8   | 112.4 | 1                   |
| D13 | 553.99 | 2        | 8           | 5.93  | 88.61 | 1                   |
| D17 | 473.79 | 2        | 8           | 6.67  | 112.73| 1                   |
| D34 | 705.11 | 2        | 11          | 7.26  | 143.66| 3                   |

Figure 6. The bioavailability radar of (A) molecule D2 (B) molecule D8 (C) molecule D13 (D) molecule D17 and (E) D34.
results of ADME properties predicted indicated that only molecules 34 among others was nor orally bioavailable as it has violated more than the maximum permissible limit for the orally bioavailability of drugs set by Lipinski’s rule of five. This research was able to identified compound 17 as a lead among the studied compounds and proposed when designing new compounds it should be used as template for structural modification.

Declarations

Author contribution statement

Adamu Uzairu: Conceived and designed the experiments.
Muhammad Tukur Ibrahim: Performed the experiments; Wrote the paper.
Sani Uba, Gideon Adamu Shallangwa: 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.

Acknowledgments

Ahmadu Bello University, Zaria was sincerely acknowledge by the authors for its technical support in the course of this research.

References

[1] L.-L. Kong, R. Ma, M.-Y. Yao, X.-E. Yan, S.-J. Zhu, P. Zhao, C.-H. Yan, Structural pharmacological studies on EGFR T790M/C797S, Biochem. Biophys. Res. Commun. 488 (2) (2017) 266–272.

[2] J. Song, S. Jung, J.W. Lee, D. Jung, S. Lee, K.H. Min, Click chemistry for improvement in selectivity of quinazoline-based kinase inhibitors for mutant epidermal growth factor receptors, Bioorg. Med. Chem. Lett. 29 (3) (2019) 477–480.

[3] E.J. Hanan, M. Baumgardner, M.C. Bryan, Y. Chen, C.Eigenbrot, P. Fan, X.-H. Gu, H. La, S. Malek, H.E. Purkey, 4-Aminoindazolyl-dihydrofuro [3, 4-d] pyrimidines as non-covalent inhibitors of mutant epidermal growth factor receptor tyrosine kinase, Bioorg. Med. Chem. Lett. 26 (2) (2016) S34–S39.

[4] D.A. Cross, S.E. Ashton, S. Ghiorghiu, C. Eberlein, C.A. Nebhan, P.J. Spitzler, J.P. Orme, M.R.V. Finlay, R.A. Ward, M.J. Mellor, AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer, Cancer Discov. 4 (9) (2014) 1046–1061.

[5] F. Solca, G. Dahl, A. Zeoghel, G. Bader, M. Sandersen, C. Klein, O. Kraemer, F. Himmelsbach, E. Haaksma, G.R. Adolf, Target binding properties and cellular activity of atafinib (BBBW 2992), an irreversible ErbB family blocker, J. Pharmacol. Exp. Therapeutics. 343 (2) (2012) 342–350.

[6] M.-S. Tsao, A. Sakurada, J.-C. Cutz, C.-Q. Zhu, S. Kamel-Reid, J. Squire, L. Lorimer, T. Zhang, N. Liu, M. Daneshmand, Erlotinib in lung cancer—molecular and clinical predictors of outcome, N. Engl. J. Med. 353 (2) (2005) 133–144.

[7] K. Ojha Lokendra, S. Rachana, B.M. Rani, Modern drug design with advancement in QSAR: a review, Int. J. Res. Biosci. 2 (1) (2013) 1–12.

[8] A. Abdulatif, S. Uba, B.A. Umar, M.T. Ibrahim, Molecular design and docking analysis of the inhibitory activities of some α-substituted acetamido-N-benzylacetamide as anticancer agents, SN Appl. Sci. 1 (5) (2019) 499.

[9] D.B. Kitchen, H. Decerzo, J.P. Furr, J. Bejarat, Docking and scoring in virtual screening for drug discovery: methods and applications, Nat. Rev. Drug Discov. 3 (11) (2004) 935.

[10] M.F. Khan, G. Verma, A. Akhtar, M. Shaquisquzzaman, M. Akhter, M.A. Rizvi, M.M. Alam, Pharmaphore modeling, 3D-QSAR, docking study and ADME prediction of acyl 1, 3, 4-thiazole amides and sulfonamides as antitubulin agents, Arab. J. Chem. (2016).

[11] W. Hou, Y. Ren, Z. Zhang, H. Sun, Y. Ma, B. Yan, Novel quinazoline derivatives bearing various 6-benzamide moieties as highly selective and potent EGFR inhibitors, Bioorg. Med. Chem. 26 (8) (2018) 1740–1750.

[12] M. Abdullahia, G.A. Shallangwa, M.T. Ibrahim, A.U. Bello, D.E. Arthura, A. Uzairu, P. Mamzaa, Q SAR Studies on Some C14-Urea Tetrandrine Compounds bearing various 6-benzamide moieties as highly selective and potent EGFR inhibitors, Bioorg. Med. Chem. 37 (14) (2019) 2215–2225.

[13] V. B. T. Zhang, N. Liu, M. Daneshmand, Erlotinib in lung cancer—molecular and clinical predictors of outcome, N. Engl. J. Med. 353 (2) (2005) 133–144.

[14] W. Kohn, A.D. Becke, R.G. Parr, Density functional theory of electronic structure, J. Phys. Chem. 100 (31) (1996) 14071–14084.

[15] T.I. Oprea, C.L. Waller, G.R. Marshall, Three-dimensional quantitative structure–activity relationship of human immunodeficiency virus (I) protease inhibitors. 2. Predictive power using limited exploration of alternate binding modes, J. Med. Chem. 37 (14) (1994) 2206–2215.
[21] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C.P. Varghese, R.K. Agrawal, Validation of QSAR models-strategies and importance, Int. J. Drug Des. Discov. 3 (2011) 511–515.

[22] A. Tropsha, J.R. Bajorath, Computational Methods for Drug Discovery and Design, ACS Publications, 2015.

[23] A. Beheshti, E. Pourbasheer, M. Nekoei, S. Vahdani, QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm–multiple linear regressions, J. Saudi Chem. Soc. 20 (3) (2016) 282–290.

[24] O. Adediran, A. Uzairu, G.A. Shallangwa, S.E. Abechi, QSAR and molecular docking based design of some n-benzylacetamide as γ-aminobutyrate-aminotransferase inhibitors, J. Eng. Exact Sci. 4 (1) (2018) 65–84.

[25] 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, Molec. Inform. 22 (1) (2003) 69–77.

[26] 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, Molec. Inform. 22 (1) (2003) 69–77.

[27] S.E. Adeniji, S. Uba, A. Uzairu, Quantitative structure-activity relationship and molecular docking of 4-Alkoxy-Cinnamic analogues as anti-mycobacterium tuberculosis, J. King Saud Univ. Sci. (2018).

[28] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2) (2010) 455–461.

[29] E.F. Petterson, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (13) (2004) 1605–1612.

[30] J.A. Capra, R.A. Laskowski, J.M. Thornton, M. Singh, T.A. Funkhouser, Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure, PLoS Comput. Biol. 5 (12) (2009) e1000565.

[31] S.M.D. Rizvi, S. Shakil, M. Haneef, A simple click by click protocol to perform docking: AutoDock 4.2 made easy for non-bioinformaticians, EXCLI J. 12 (2013) 831.

[32] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. 7 (2017) 42717.

[33] S.Y. Ismail, A. Uzairu, B. Sagagi, M. Sabiu, In silico molecular docking and pharmacokinetic study of selected phytochemicals with estrogen and progesterone receptors as anticancer agent for breast cancer 5 (3) (2018) 1337–1350.

[34] R. Todeschini, V. Consonni, Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing/volume II: Appendices, References, 41, John Wiley & Sons, 2009.

[35] 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.