

RESEARCH

Open Access



In silico studies of some 2-anilinopyrimidine derivatives as anti-triple-negative breast cancer agents

Hadiza Lawal Abdulrahman*, Adamu Uzairu and Sani Uba

Abstract

Background: Breast cancer is a major form of health problem on the globe and the second cause of death related to cancer amidst women. A prediction of about 1 to 1.3 million cases on cancer of the breast are detected yearly globally. Triple-negative type of breast cancers (TNBCs) are described by the lack of human epidermal growth factor receptor 2 (HER2), estrogen receptor (ER), and progesterone receptor (PR). TNBCs metastasize to the central nervous system and lungs regularly. Such metastatic actions reduce the life expectancy of patients with TNBC than patients with non-TNBC due to non-enhanced inhibitor compounds. The purpose of this research was to explore the anti-proliferative activities of 2-anilinopyrimidine derivatives against triple-negative cancer cell line MDA-MB-468 via in silico studies like QSAR and molecular docking studies to further design and develop new anti-breast cancer drug with high potency and low toxicity.

Results: The quantitative structure–activity relationship QSAR model predicts the bioactivities of the compounds, and molecular docking studies comprehend the interaction between the derivatives (ligand) and thyroid hormone (TR β 1) (receptor). Model 4 was chosen as the best model from the statistical assessment; $R^2 = 0.8760$, $R^2_{adj} = 0.8451$, $Q^2 = 0.6141$, and R^2_{pred} of 0.5390. From the external validation of the QSAR model, the coefficient of the mean effect on the model parameters indicates that decreasing (VR1_Dzv and MOMI-R) and increasing (SpMin1_Bh and C3SP3) would increase the anti-proliferative activities (pIC₅₀) of the compounds. The molecular docking studies revealed that ligands 15 and 18 had the highest docking scores of -7.3 and -7.4 kcal/mol with thyroid hormone receptor (TR β 1). The ligands had docking scores better than the standard anti-breast cancer drug gefitinib (-5.3 kcal/mol).

Conclusions: The results indicate that model 4 can be used in developing new 2-anilinopyrimidine derivatives, with better anti-breast cancer prediction activity and performance. It was proved that some series of 2-anilinopyrimidine derivative compounds bind tightly to the receptor, stabilizing the receptor (TR β 1) which is evident from the receptor–ligand interactions, and these compounds would serve as the most promising inhibitors against TR β 1. This shows a breakthrough for pharmaceutical researchers in designing and developing new anti-triple-negative breast cancer drugs.

Keywords: QSAR, Molecular docking studies, 2-Anilinopyrimidine, Triple-negative breast cancer (TNBC), Thyroid hormone receptor (TR β 1)

* Correspondence: azeezalawal@gmail.com

Department of Chemistry, Ahmadu Bello University, P.M.B. 1044, Zaria, Nigeria

1 Background

After cardiovascular diseases, cancer is the second most deadly disease to the human health [1]. Worldwide, one of seven main death causes is cancer that affects around 14 million people every year. The adoption of lifestyle activities especially in developing countries where almost 82% of the entire population of the world exist has increased higher risk of cancer, due to lack of exercise, smoking, and heredity variation [2]. Breast cancer is the utmost form of cancer on the globe and the second cause of death related to cancer amidst women. A prediction of about 1 to 1.3 million cases on cancer of the breast is detected yearly globally [3].

Triple-negative type of breast cancers (TNBCs) are termed as antagonistic mammary growths, and they are described by the lack of human epidermal growth factor receptor 2 (HER2), estrogen receptor (ER), and progesterone receptor (PR) [4]. TNBCs metastasize to the central nervous system and lung regularly than non-TNBCs, which usually metastasize to the bone. Such metastatic actions reduce the life expectancy of patients with TNBC than patients with non-TNBC patients due to non-enhanced inhibitors compounds [3].

Recently, a novel series of 2-anilinopyrimidine was reported by Jo et al. [4] as inhibitors against MDA-MB-468 cell line. There is also evidence that reduced thyroid hormone receptor manifestation and/or variations in thyroid hormone genes occur frequently in cancer [5], suggesting that the native receptors could act as tumor suppressors and that loss of re-occurrence of this receptor could show a selective lead for cell alteration and advancement of tumor transformation [6].

Unconventional medicine takes prolonged time and effort to be manufactured, thereby not meeting up with the urgency needed for a comprehensive treatment. Computer-aided drug design has been a great success in designing novel drugs with great effectiveness and better potency against diseases. The aim of this study is to explore the anti-proliferative activities of 2-anilinopyrimidine against triple-negative cancer cell line, MDA-MB-468 via in silico studies like QSAR and docking studies that can be used to further develop anti-breast cancer drug candidate.

2 Methods

2.1 Computational information

2.1.1 Hardware

The computer details used in this research is the 7th Generation HP Pavilion Intel R, core i7-7500u RAM 12.00 GB running on a Windows 10 operating system.

2.1.2 Software

The software used to carry out this research includes Spartan'14 (version 1.1.2), Material studio (V8), AutoDock visualiser version 4.2, Pyrex software version, ChemDraw

software version 12.0.2, PADEL-Descriptor Software V2.20 and DTC data lab software, and Microsoft Word Office Excel 2013 version.

2.2 QSAR studies

2.2.1 Data collection

Thirty novel derivative compounds of 2-anilinopyrimidine with their inhibitory concentration (IC_{50}), against triple-negative breast cancer (MDA-MB-468) cell line, were acquired from Jo et al. [4] reports.

2.2.2 Bioactivities

Anti-proliferative activities of 2-anilinopyrimidine derivative compounds were measured in inhibitory concentration (IC_{50}); an IC_{50} (50% inhibitory concentration) value of a chemical compound is defined as the concentration of the compound required to decrease the viability of a given cell line by 50%. The IC_{50} values were normalized using the scale of logarithm to pIC_{50} values to reduce the skew in the activities. The tabulated anti-proliferative activities (IC_{50}) and pIC_{50} of the derivatives are shown in Table 1, and it is measured in the concentration of micromolar (μM). The logarithm scale is given as follows:

$$pIC_{50} = -\log_{10}(IC_{50} \times 10^{-6}).$$

2.2.3 Geometry optimization

The geometry optimization is aimed to earn a more desirable geometric structure that is closer to the actual geometric condition of the molecular structure [2]. The derivative compounds were sketched in 2D on ChemDraw V (12.0.2) and converted on Spartan 14 V (1.1.4) software. Density functional theory (DFT) using the B3LYP, 6-311G basis set, was used for the geometric optimization of the compounds [7–9]. The parent compound is shown in Fig. 1.

2.2.4 Molecular descriptor

Pharmaceutical Data Exploration Laboratory Software V (2.20) was used in calculating molecular descriptors for the 30 optimized compounds of 2-anilinopyrimidine derivatives [10].

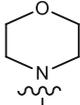
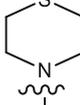
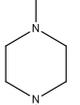
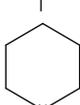
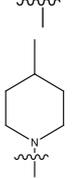
2.2.5 Pretreatment and division of data set

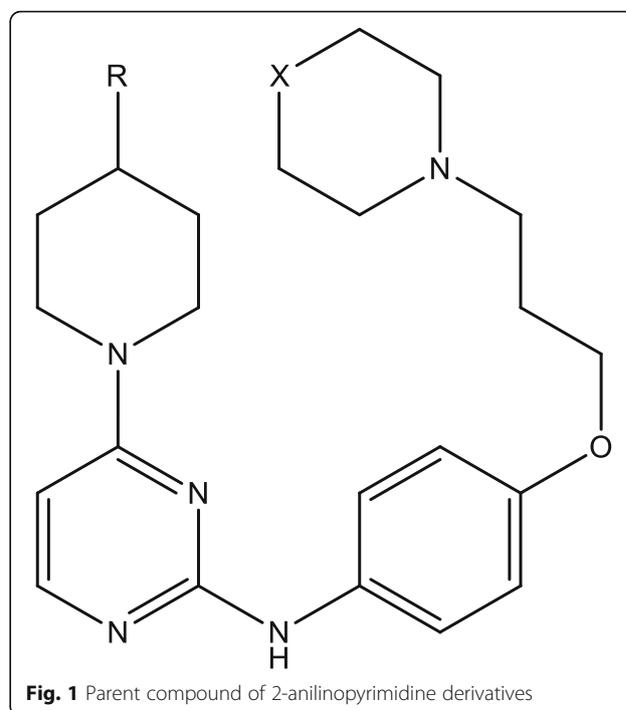
Results obtained from PADEL-software were pretreated using Data Pre-treatment software GUI 1.2 to remove constant values and unwanted descriptors [9, 11]. Kennard-Stone algorithm [12] was used in dividing the derivatives into 21 train and 9 test set to build the model.

2.2.6 Model building and model validation

The internal validation of the train test (twenty-one compounds) was executed in version 8 of Material

Table 1 2-Anilinympyrimidine derivatives with their activities

| Compounds | R | X | Experimental activities (IC ₅₀) | Calculated activities (pIC ₅₀) |
|-----------|---|-----------------|---|--|
| 1 | H | O | 55.2 | 4.26 |
| 2 | - | O | 84.5 | 4.07 |
| 3 | - | S | 57.4 | 4.24 |
| 4 | - | CH ₂ | 86.8 | 4.06 |
| 5 | - | O | 73.4 | 4.13 |
| 6 | - | S | 65.0 | 4.19 |
| 7 | - | CH ₂ | 78.1 | 4.11 |
| 8 | - | O | 80.7 | 4.09 |
| 9 | - | S | 56.5 | 4.25 |
| 10 | - | CH ₂ | 71.7 | 4.14 |
| 11 | H | O | 71.8 | 4.14 |
| 12 | H | CH ₂ | 45.1 | 4.35 |
| 13 | Me | O | 51.3 | 4.29 |
| 14 | Me | S | 61.3 | 4.21 |
| 15 | Cl | CH ₂ | 21.3 | 4.67 |
| 16 | Cl | O | 61.6 | 4.21 |
| 17 | Cl | S | 62.7 | 4.20 |
| 18 | Cl | CH ₂ | 52.1 | 4.28 |
| 19 | - | O | 77.1 | 4.11 |
| 20 | - | S | 78.3 | 4.11 |
| 21 | - | CH ₂ | 53.1 | 4.24 |
| 22 | - | O | 69.7 | 4.16 |
| 23 | - | S | 36.0 | 4.44 |
| 24 | - | CH ₂ | 53.1 | 4.27 |
| 25 |  | - | 88.0 | 4.06 |
| 26 |  | - | 88.0 | 4.06 |
| 27 |  | - | 64.6 | 4.19 |
| 28 |  | - | 69.7 | 4.16 |
| 29 |  | - | 62.0 | 4.21 |
| 30 |  | - | 16.9 | 4.77 |



studio software to construct a model by employing a genetic function approximation technique. Using the Friedman formula, the obtained models were evaluated [13].

$$LOF = \frac{SEE}{M \left[1 - \beta \left(\frac{c+d \times p}{M} \right) \right]^2}$$

where SEE is the standard estimated error. If SEE is low, it implies a better model. SEE is expressed as follows:

$$SEE = \sqrt{\frac{(Y_{\text{exp}} - Y_{\text{pred}})^2}{N - P - 1}}$$

C is the sum of the model terms, p is the total number of model descriptors, M is the sum of train set, and d is a user-defined smoothing parameter [14]. The model is verified using the correlation coefficient (R^2). R^2 value

Table 2 Recommended values used in the assessing of the QSAR model

| Character | Name | Value |
|-----------------------|---|------------|
| R^2 | Coefficient of determination | ≥ 0.6 |
| $P_{(95\%)}$ | Confidence interval at 95% confidence level | < 0.05 |
| Q^2 | Squared cross-validation coefficient | ≥ 0.5 |
| $R^2 - Q^2$ | Difference between R^2 and Q^2 | < 0.3 |
| $N_{\text{test set}}$ | Least number of test set | ≥ 5 |
| R^2_{ext} | Coefficient of determination of external validation | ≥ 0.5 |

Table 3 External validation of model 4

| Name | pIC ₅₀ | VR1_Dzv | SpMin1_Bhs | C3SP3 | MOMIR | Y _{pred} | Y _{pred} - Y _{obs} |
|------|-------------------|----------|------------|-------|---------|-------------------|--------------------------------------|
| 26 | 4.06 | 695.2029 | 1.8450 | 0 | 9.5814 | 4.1940 | - 0.1340 |
| 4 | 4.06 | 505.8137 | 1.8685 | 0 | 10.1391 | 4.1625 | - 0.1025 |
| 2 | 4.07 | 507.3869 | 1.7777 | 0 | 9.7376 | 4.1063 | - 0.0364 |
| 8 | 4.09 | 1589.825 | 1.8089 | 0 | 10.0072 | 4.0429 | 0.0471 |
| 12 | 4.35 | 507.0445 | 1.8648 | 0 | 9.7426 | 4.2090 | 0.1409 |
| 13 | 4.29 | 1135.849 | 1.8954 | 1 | 9.5861 | 4.6054 | - 0.3154 |
| 23 | 4.44 | 723.6867 | 1.8626 | 0 | 9.6193 | 4.2091 | 0.2308 |
| 30 | 4.77 | 461.2732 | 1.9174 | 1 | 9.7076 | 4.6569 | 0.1131 |
| 25 | 4.1 | 650.2583 | 1.7925 | 0 | 9.3087 | 4.1705 | - 0.0704 |

should be close to 1, to obtain an enhanced and effective model. R^2 is given as follows:

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

where Y_{exp} and Y_{pred} are averages of anti-proliferative and predicted activities of the train set [15].

R^2 value increases as the descriptor number increases; thus, R^2 value is not guaranteed in terms of the model's strength. The R^2 is altered to obtain a robust and strong model, which is given as follows:

$$R^2 \text{ adj} = \frac{R^2 - P(n-1)}{n-p+1}$$

where p and n are the numbers of generated descriptors and train set. The stability of the model derivatives was assessed using validation coefficient test (Q^2_{cv}) given as:

$$Q^2_{\text{cv}} = 1 - \frac{\sum (Y_{\text{pred}} - Y_{\text{exp}})^2}{\sum (Y_{\text{exp}} - Y_{\text{training}})^2}$$

Table 4 Continuation of descriptor values of the test set used in external validation of model 4

| $(Y_{\text{pred}} - Y_{\text{obs}})^2$ | Y_{mintrain} | $Y_{\text{ob}} - Y_{\text{mintrain}}$ | $(Y_{\text{ob}} - Y_{\text{mintrain}})^2$ |
|--|-----------------------|---------------------------------------|---|
| 0.018212 | 4.213333 | - 0.15333 | 0.023511 |
| 0.010513 | 4.213333 | - 0.15333 | 0.023511 |
| 0.001323 | 4.213333 | - 0.14333 | 0.020544 |
| 0.002219 | 4.213333 | - 0.12333 | 0.015211 |
| 0.019867 | 4.213333 | 0.136667 | 0.018678 |
| 0.099508 | 4.213333 | 0.076667 | 0.005878 |
| 0.053288 | 4.213333 | 0.226667 | 0.051378 |
| 0.012788 | 4.213333 | 0.556667 | 0.309878 |
| 0.004966 | 4.213333 | - 0.11333 | 0.012844 |
| Sum = 0.222685 | | | Sum = 0.481433 |

Table 5 Differentiation of bio-activities (pIC₅₀), prediction inhibition, and residual of model 4

| Name | pIC ₅₀ | Predicted IC ₅₀ | Residual |
|------|-------------------|----------------------------|----------|
| 1 | 4.26 | 4.1979 | 0.0621 |
| 2* | 4.07 | 4.1064 | - 0.0364 |
| 3 | 4.24 | 4.1818 | 0.0582 |
| 4* | 4.06 | 4.1625 | - 0.1025 |
| 5 | 4.13 | 4.1719 | 0.0419 |
| 6 | 4.19 | 4.1616 | 0.0284 |
| 7 | 4.11 | 4.2327 | - 0.1227 |
| 8* | 4.09 | 4.0429 | 0.0471 |
| 9 | 4.25 | 4.2148 | 0.0352 |
| 10 | 4.14 | 4.1386 | 0.0014 |
| 11 | 4.14 | 4.1807 | - 0.0407 |
| 12* | 4.35 | 4.2090 | 0.1400 |
| 13* | 4.29 | 4.6054 | - 0.3155 |
| 14 | 4.21 | 4.2225 | - 0.0125 |
| 15 | 4.67 | 4.6575 | 0.0125 |
| 16 | 4.21 | 4.1851 | 0.0249 |
| 17 | 4.20 | 4.1708 | 0.0292 |
| 18 | 4.28 | 4.2991 | - 0.0191 |
| 19 | 4.11 | 4.1178 | - 0.0078 |
| 20 | 4.11 | 4.1578 | - 0.0478 |
| 21 | 4.24 | 4.2204 | 0.0196 |
| 22 | 4.16 | 4.1488 | 0.0112 |
| 23* | 4.44 | 4.2092 | 0.2308 |
| 24 | 4.27 | 4.2534 | 0.0166 |
| 25* | 4.10 | 4.1705 | - 0.0704 |
| 26* | 4.60 | 4.1940 | - 0.1340 |
| 27 | 4.19 | 4.2215 | - 0.0315 |
| 28 | 4.16 | 4.1405 | 0.0195 |
| 29 | 4.21 | 4.2047 | 0.0053 |
| 30* | 4.77 | 4.6569 | 0.1131 |

The compounds with an asterisk symbol (*) are the test set while the compounds without an asterisk symbol (*) are the train set

Table 6 Definition of descriptors (model parameters) and their classes for model 4

| Descriptors | Description | Class |
|-------------|---|-------|
| VR1_Dzv | Randic-like eigenvector-based index from Barysz matrix/weighted by Van der Waals volumes | 2D |
| spMin1_Bhs | Smallest absolute eigenvalue of burden-modified matrix – n 1/weighted by relative I-state | 2D |
| C3SP3 | Singly bound carbon bound to three other carbons | 2D |
| MOMI-R | Radius of gyration | 2D |

Y_{training} , Y_{exp} , and Y_{pred} are the average biological activities (pIC₅₀), biological activities (pIC₅₀), and prediction inhibition concentration of the train set [16].

2.2.7 QSAR modeling evaluation

The models generated were set to undergo statistical parameters such as the cross-validated test, R^2 Fisher's test, and R^2 predicted.

2.2.8 Mean effect

The mean effect relates to the impact of the descriptors and the compound activities in the model. Notations attached to the descriptors show the variant direction in the activity of the compounds, either an increase or a decrease in the descriptor value. It is defined as follows:

$$\text{Mean effect} = \frac{\beta_j \sum_i^n D_j}{\sum (\beta_j \sum_i^n D_j)}$$

where m is the total descriptors in the model, B_j equals to descriptor coefficient j , n is the total molecules in the train set, and D_j is the matrix value of the descriptor in the train set [17].

2.2.9 Variance inflation factor (VIF)

The VIF measures the extent of correlation between one descriptor and the other descriptor in a model. The higher the values show that it is almost impossible and difficult to show the contribution of a descriptor accurately in a model. It is evaluated as follows:

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

R^2 is the correlation coefficient [18].

The higher the value, the greater the correlation between the descriptors. Values of 1–7 are sometimes regarded as being moderate, and it shows the strength and robustness of the model, while values of 10 show the correlation between the descriptors is very high, and therefore, the model is very unstable.

2.2.10 QSAR applicability domain of the model

The goal of applicability domain methods is for estimating individually the consistency of each model generated [19], and it aimed at predicting the uncertainty of a compound depending on its similarities to the compounds used in building the model and also the distance of both train and test set. The leverage is used in defining the applicability domain of the generated models [20]. It is formulated as follows:

$$H_i = x_i (X^T X)^{-k} x_i^T \quad (i = K \dots, P)$$

where X is the $n \times k$ matrix of train set descriptors, X^T is the matrix transpose of X used in building the model, and X_i is the matrix of train compounds of I . (h^*) is the warning leverage, and it is a prediction tool that checks for outliers. It is written as follows:

$$h^* = \frac{3(p+1)}{n}$$

n is equal to the total train set and p equals to the total descriptors from the model generated. William's plot is generated by plotting the standardized residuals versus the leverage of both the train and test set. Molecules that fall within the warning leverages on the plot are the predicted compounds that fall within the threshold. The reliability of the QSAR model was assessed using the minimum accepted values as shown in Table 3 [21].

2.3 Molecular docking

Molecular docking studies were implemented on the derivative compounds of 2-anilino pyrimidine (ligand) and thyroid hormone receptor (TRβ1). The receptor was gotten from protein data bank with the code (PDB: 1Y0X).

Table 7 Statistical analysis of the descriptors used in the QSAR model 4

| | VR1_Dzv | SpMin1_Bhs | C3SP3 | MOMI-R | VIF | Mean effect (ME) | P values |
|------------|----------|------------|----------|--------|--------|------------------|----------|
| VR1_Dzv | 1 | | | | 2.5789 | – 0.0755 | 0.000204 |
| SpMin1_Bhs | 0.449216 | 1 | | | 1.2807 | 2.3585 | 0.004158 |
| C3SP3 | 0.756905 | 0.419607 | 1 | | 2.5062 | 0.0387 | 3.4E–0.8 |
| MOMI-R | 0.047909 | – 0.05524 | – 0.1127 | 1 | 1.0608 | – 1.3217 | 0.000358 |

The docking scores of the ligand–receptor were obtained with Autodock Vina of PyRx software [11]. The detailed interactions between the ligand and the receptor were visualized using Discovery Studio Visualizer.

3 Results

3.1 QSAR of 2-anilinopyrimidine derivatives

Four QSAR models were generated using the Genetic Function Approximation (GFA) technique to predict the anti-proliferative activities. Model 4 passed the internal validation test which confirmed with the least requirement for QSAR modeling as shown in Table 2.

3.1.1 Model 1

$$pIC_{50} = -0.000041993 \times VR1_Dzv + 0.430022665 \times C3SP3 - 0.029366849 \times RDF125m - 0.013797643 \times RDF105p + 4.414124338$$

3.1.2 Model 2

$$pIC_{50} = 0.019329185 \times ALogp2 + 0.209407843 \times C3SP3 + 0.013676289 \times RDF40i - 0.000911095 \times Vm + 3.736702488$$

3.1.3 Model 3

$$pIC_{50} = -0.015741625 \times VR3_Dzv + 0.385603503 \times C3SP3 - 0.036977855 \times RDF125m - 0.016463267 \times RDF105p + 4.720613746$$

3.1.4 Model 4

$$pIC_{50} = -0.000060824 \times VR1_Dzv + 1.185723768 \times SpMin1_Bhs + 0.378178925 \times C3SP3 - 0.128667903 \times MOMI-R + 3.282331294$$

Tables 3 and 4 show the calculation of the external validation of the QSAR model using the model parameters of model 4. The external validation (R^2_{pred}) was calculated as 0.5390, which also conforms to the minimum required values for QSAR modeling, and makes the model very robust and highly potent. The meaning of

each model parameter used in validating model 4 is given in Table 6.

$$\begin{aligned} \sum (Y_{ob} - Y_{pred})^2 &= 0.2227 \sum (Y_{obs} - Y_{train})^2 \\ &= 0.4814 \therefore R^2_{test} \\ &= 1 - (0.2227 / 0.4814) = 0.5390 \end{aligned}$$

The experimental, predicted, and the residual values of 2-anilinopyrimidine derivatives are shown in Table 5. The residual values were obtained from the calculated activities statistically. All the derivative compounds had low residual values indicating the degree of effectiveness of the QSAR model 4.

Table 6 shows the four model parameters (descriptors) that were used in building the QSAR model 4 and were also used in evaluating the strength of the model externally. The descriptors are defined and classified in Table 6.

Table 7 shows the statistical evaluation (VIF, mean effect, P values) of the model parameters. The VIF shows the degree of co-linearity between the descriptors, and it was calculated using the following equation:

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

R^2 is the correlation coefficient [18].

The mean effect shows the contribution of each descriptor to the built model, and the signs of the values show if the descriptors give a negative or positive contribution in the model. The P values evaluate the statistical significance between the model parameters.

Figure 2 shows a straight line graph of calculated activities (predicted activities) against experimental activities of 2-anilinopyrimidine derivative compounds as tabulated in Table 5. Both the experimental and predicted activities showed a good relationship as proven by the graph.

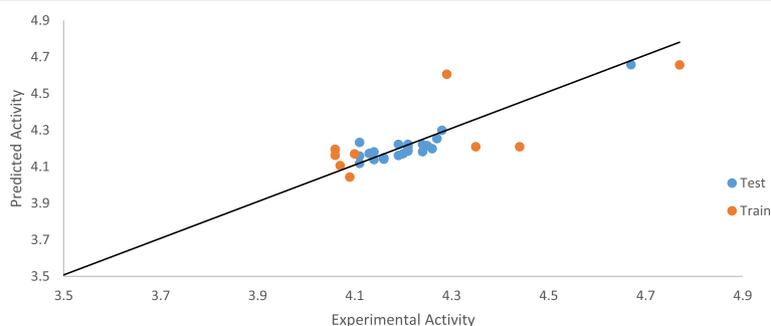


Fig. 2 Graph of predicted activities versus experimental activities (bioactivities)

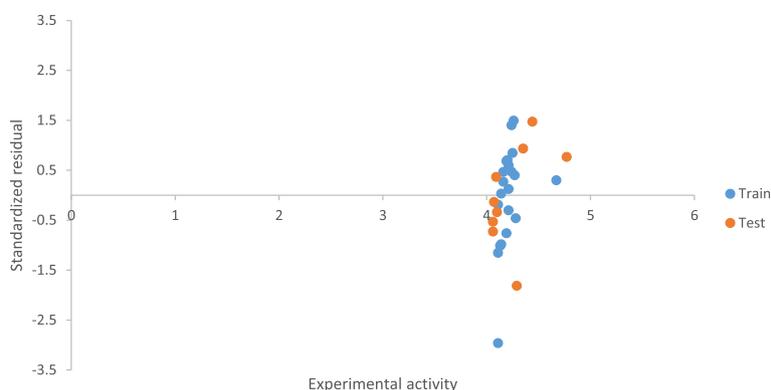


Fig. 3 Graph of standardized residual against inhibition concentration (experimental activity)

Figure 3 shows a graph of standardized residual against inhibition concentration of both the train and test set. All the values were well distributed on both sides of the y -axis, showing the effectiveness of model 4.

Figure 4 is a graph of standardized residuals against the leverage values, and the plot is called William's plot. The plot was used to assess the uncertainty in similarities of the derivative compounds used in building the model. Compounds that fall between the warning leverage tend to be similar structurally. The warning leverage was calculated to be ($h^* = 0.714$) using the formula:

$$h^* = \frac{3(p+1)}{n}$$

3.2 Molecular docking studies

The summary of the docking studies result of some 2-anilinopyrimidine derivative compounds is given in Table 8. The docking score was obtained using PyRx software while the docking interactions between the receptor and the ligand to form complexes which include hydrophobic bond, hydrogen bond, and the bonding distances were visualized using Discovery Studio Software.

The hydrogen and hydrophobic interaction that occurred between 2-anilinopyrimidine derivative compounds (ligand) and the active pocket of (TR β 1) receptor in 3D format for complexes 15 and 18 are shown in 2D format in Figs. 6 and 7, while Fig. 8 shows the same interaction in a 3D format.

4 Discussion

4.1 QSAR of 2-anilinopyrimidine

QSAR modeling was used to validate quantitatively the structure relationship of 2-anilinopyrimidine derivatives with its anti-proliferative activities. The robustness of the QSAR models was assessed by the fitness of the train set and predicted pIC_{50} of the test set. Four QSAR models generated using the Genetic Function Approximation (GFA) technique to predict the anti-proliferative activities. Model 4 passed the internal validation with correlation coefficient squared (R^2) of 0.8760, correlation coefficient adjusted squared (R^2_{adj}) of 0.8451, cross-validation coefficient (Q^2) of 0.6141, and external validation (R^2_{pred}) of 0.5390. All the values obtained were in accordance with the least proposed value used in the evaluation of QSAR model as shown in Table 2. The obtained values (R^2 , R^2_{adj} , Q^2 , and R^2_{pred}) indicate the

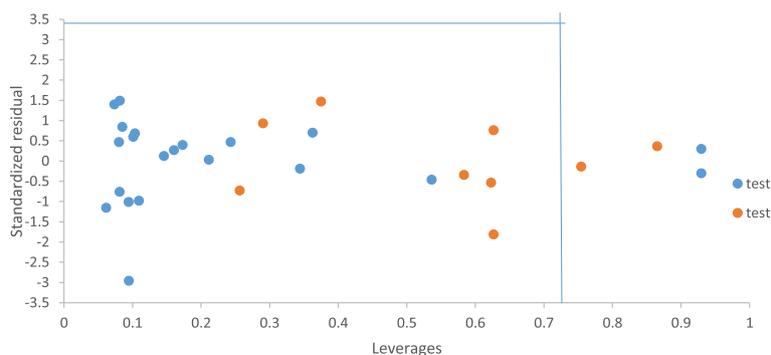


Fig. 4 A plot of standardized residual versus leverages (William's plot)

Table 8 Binding affinity, interaction type, bond type, and distances in between some compounds and the receptor

| Complex | Binding affinity (kcal/mol) | Amino acid | Bond type | Interaction | Distance (Å ⁰) |
|---------|-----------------------------|------------|---------------|----------------------------|----------------------------|
| 30 | - 5.9 | LEU360 | Hydrogen bond | Conventional hydrogen bond | 2.19527 |
| | | LEU360 | Hydrogen bond | Conventional hydrogen bond | 2.08848 |
| | | VAL319 | Hydrogen bond | Carbon hydrogen bond | 3.03818 |
| | | SER361 | Hydrogen bond | Pi-donor hydrogen bond | 3.21568 |
| | | VAL319 | Hydrophobic | Alkyl | 5.40031 |
| | | TRP239 | Hydrophobic | Pi-alkyl | 5.30319 |
| | | TRP239 | Hydrophobic | Pi-alkyl | 4.57531 |
| 12 | - 7.2 | GLY432 | Hydrogen bond | Conventional hydrogen bond | 2.96575 |
| | | GLY432 | Hydrophobic | Amide-Pi stacked | 3.88004 |
| | | ILE303 | Hydrophobic | Alkyl | 5.23513 |
| | | LYS306 | Hydrophobic | Alkyl | 4.84663 |
| | | ARG383 | Hydrophobic | Alkyl | 5.07109 |
| | | PRO384 | Hydrophobic | Alkyl | 5.29712 |
| | | ALA433 | Hydrophobic | Pi-alkyl | 4.14051 |
| | | ALA436 | Hydrophobic | Pi-alkyl | 5.48801 |
| 18 | - 7.3 | GLU311 | Hydrogen bond | Conventional hydrogen bond | 2.10982 |
| | | ARG429 | Hydrogen bond | Conventional hydrogen bond | 2.68544 |
| | | GLY307 | Hydrogen bond | Conventional hydrogen bond | 2.97669 |
| | | GLU311 | Hydrogen bond | Conventional hydrogen bond | 2.85424 |
| | | VAL458 | Hydrogen bond | Carbon hydrogen bond | 3.34145 |
| | | ILE303 | Hydrophobic | Alkyl | 5.28774 |
| | | LYS306 | Hydrophobic | Alkyl | 4.9622 |
| | | ARG383 | Hydrophobic | Alkyl | 5.40494 |
| | | PRO384 | Hydrophobic | Alkyl | 4.84454 |
| | | PRO384 | Hydrophobic | Alkyl | 5.15235 |
| | | ALA436 | Hydrophobic | Pi-alkyl | 4.91503 |
| 15 | - 7.4 | GLU311 | Hydrogen bond | Conventional hydrogen bond | 2.15506 |
| | | ARG429 | Hydrogen bond | Conventional hydrogen bond | 2.57929 |
| | | GLU311 | Hydrogen bond | Conventional hydrogen bond | 2.76094 |
| | | VAL458 | Hydrogen bond | Carbon hydrogen bond | 3.37649 |
| | | ILE303 | Hydrophobic | Alkyl | 5.43788 |
| | | LYS306 | Hydrophobic | Alkyl | 5.04683 |
| | | ARG383 | Hydrophobic | Alkyl | 5.3858 |
| | | PRO384 | Hydrophobic | Alkyl | 5.11068 |
| | | PRO384 | Hydrophobic | Alkyl | 4.78448 |
| PRO384 | Hydrophobic | Alkyl | 4.75312 | | |

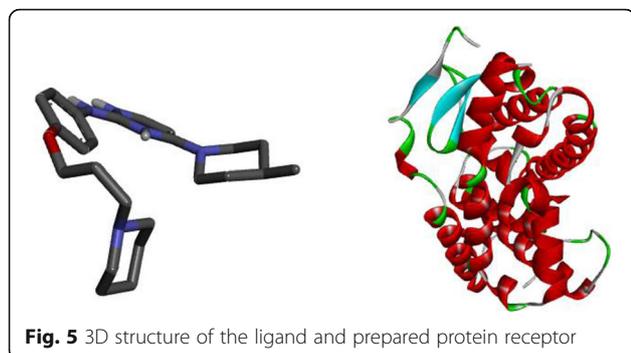
existence of a high correlation between the predicted pIC_{50} along with the biological pIC_{50} of the data set.

4.1.1 External validation of QSAR model 4

Model 4 was verified as the best model using the descriptors from the test set of the derivative compounds. Tables 3 and 4 show how the external validation was achieved using the values of the descriptors from the test

set. The experimental, predicted, and the residual values of 2-anilinopyrimidine derivatives are shown in Table 5. The low residual value from biological (anti-proliferative) activities and predicted activities shows the high performance of the model.

Table 6 shows the definition of the descriptors (model parameters). The mean effect result (Table 7) showed the degree of impact of each descriptor on

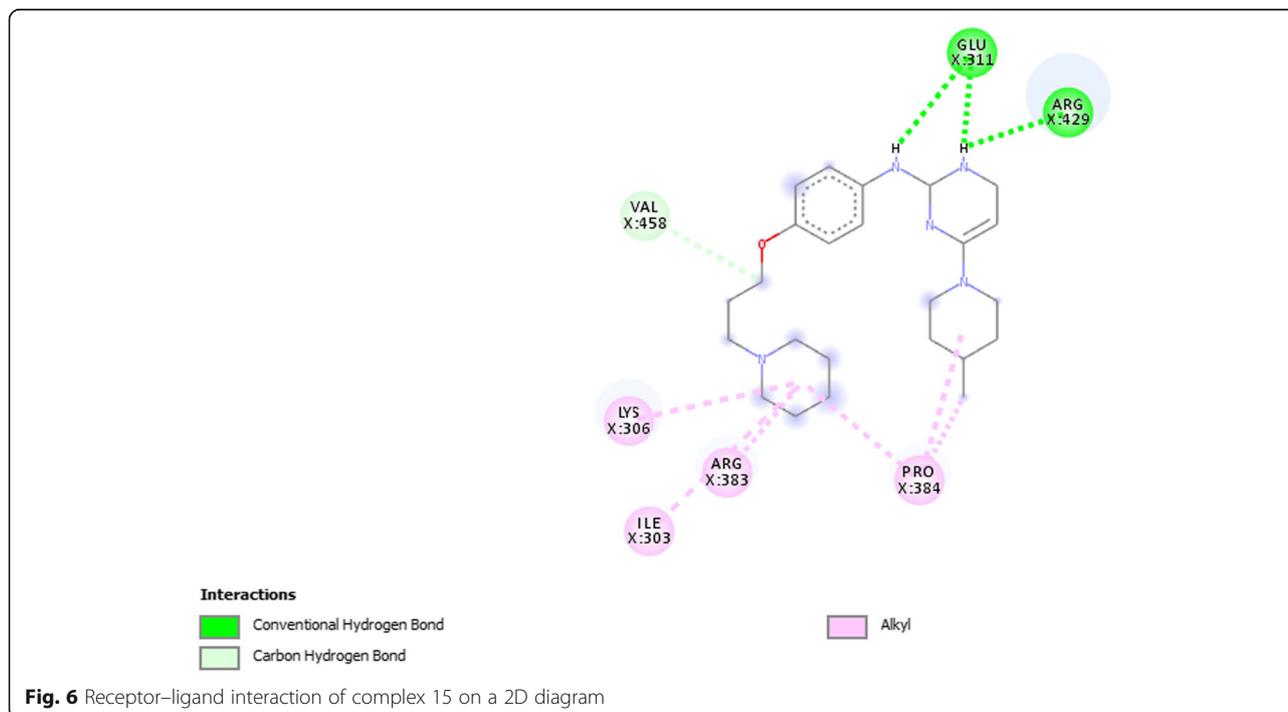


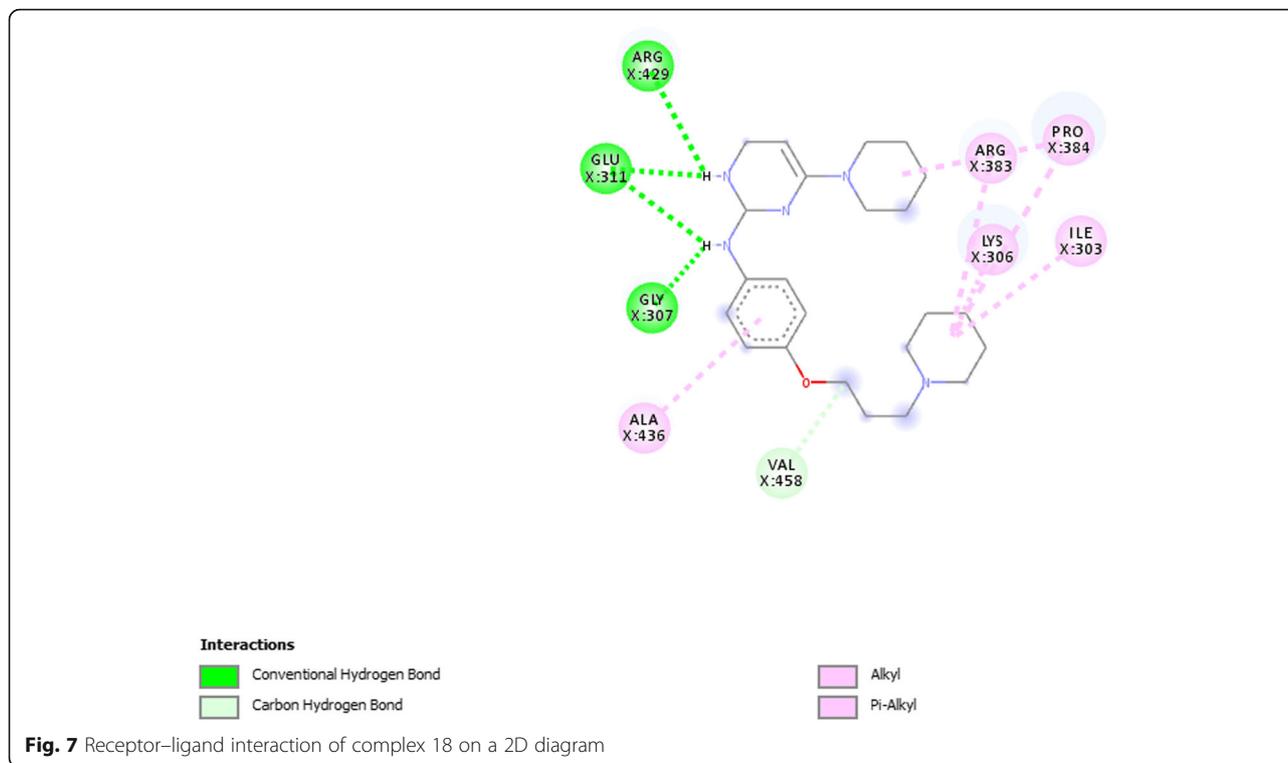
the model, and the values and coefficients of the descriptors show that decreasing MOMI-R and then VR1_Dzv (negative descriptors) would increase the anti-proliferative activities of the derivative compounds while increasing SpMin1_Bh followed by C3SP3 (positive descriptors) which would also increase the anti-proliferative activities of 2-anilinopyrimidine derivative compounds. The variance inflation factor (VIF) showed that there is no much inter-correlation between the descriptors making the model very stable. The null hypothesis shows no significant connection amid the bio-activity and model parameters of the constructed model at $p > 0.05$. At 95% confidence level, the P values of the model parameters were below 0.05. Therefore, the null hypothesis is rejected and the alternative hypothesis is accepted as shown in Table 7.

Figure 2 shows the plot of predicted activity (pIC_{50}) versus the experimental activity (IC_{50}) of both the test set and train set of compounds. The plot showed that the predicted activity was in good agreement with its experimental values as shown in Table 2, conforming the effectiveness and strength of the built model. Figure 3 is a plot of standardized residual versus biological activity (inhibition concentration) of both the train set and test set, and it shows the values spread on both sides of the zero point of the plot, showing no systematic errors. Figure 4 is a graph of standardized residuals against the leverage value, and the plot is called William's plot. Almost all the compounds were within the calculated warning leverage (applicability domain) of $h^* = 0.714$, and compounds 2, 8, 15, and 14 were found to be outside the warning leverage which perhaps is because of the slight difference in their structures equated to other molecules of the data set. Both internal and external validation conform model 4 to be very stable, robust, and highly predictive.

4.2 Molecular docking studies

Molecular docking studies on compounds of 2-anilinopyrimidine with the protein target thyroid hormone receptor ($TR\beta 1$) were performed. Amongst all the derivatives, compounds 12, 15, 18, and 30 had high docking scores. The prepared receptor and ligand are shown in Fig. 5. Compounds 15 and 18 had the highest docking score of -7.4 and -7.3 kcal/mol as shown in Table 8. The visual examination of the docked complexes





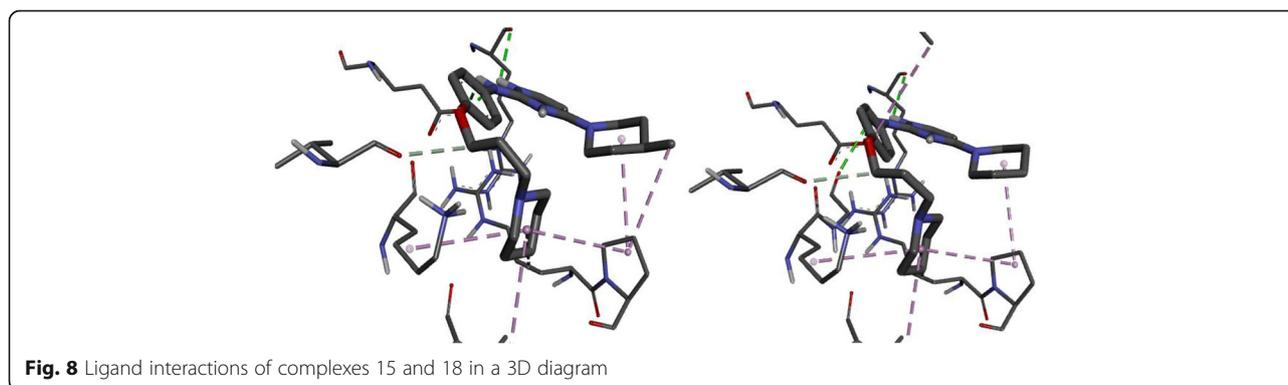
was done by evaluating the hydrogen bond interaction, hydrogen bond length, and hydrophobic interaction.

Compound 15 showed the backbone conventional hydrogen bonding interaction with ARG 429 (2.50 \AA^0) and two amino acids of GLU311 (2.7609 \AA^0 and 2.1551 \AA^0). Again, VAL458 showed carbon–hydrogen interaction with the compound at distance of 3.3765 \AA^0 . Also, the pi-orbital containing delocalized electrons in the benzene ring interact with the alkyl groups of ILE303 (5.4379 \AA^0), LYS306 (5.04683 \AA^0), and ARG383 (5.3858 \AA^0) and three amino acids of PRO384 (5.1107 \AA^0 , 4.7845 \AA^0 , and 4.7531 \AA^0) to form hydrophobic bond.

Compound 18 also showed the same hydrogen bonding with amino acid residues of GLU311 (2.10982 \AA^0), ARG439 (2.68544 \AA^0), GLY307 (2.97669 \AA^0), GLU311

(2.85424 \AA^0), and carbon–hydrogen bonding with VAL458 (3.34145 \AA^0). Furthermore, the benzene ring of the compound interacts with the alkyl groups the amino acid residues formed hydrophobic bond, and they include ILE303 (5.28774 \AA^0), LYS306 (4.9622 \AA^0), ARG383 (5.40494 \AA^0), PRO384 (4.84454 \AA^0 and 5.15235 \AA^0), and ALA436 (4.91503 \AA^0).

Both compounds were adequately docked and their orientation is similar in some instances, validating the good quality of the docking results. Both compounds showed the same hydrogen bond and hydrophobic bond interactions with the amino acid residues of the receptor at different distances. The ligands had docking scores better than the standard drug gefitinib (-5.3 kcal/mol). From the compound interaction with the receptor, it



proves the ability of the compounds to inhibit TR β 1 receptor. Figures 6 and 7 give detailed binding interactions of the receptor with ligands 15 and 18 while Fig. 8 shows how the ligand (compound) binds firmly to the active site of the protein receptor to form complexes in 3D with ligands 15 and 18.

5 Conclusion

2-Anilinopyrimidine derivatives were proven to be a better anti-cancer drug candidate against MDA-MB-468 cell line from both QSAR studies and molecular docking studies that were carried out to predict a better activity from the experimental activity of the derivatives and also comprehend the interaction of the ligand (derivatives) and thyroid hormone receptor (TR β 1). The coefficient and values of the mean effect of QSAR model 4 indicate that increasing Spmin1_Bhs and C3SP3 descriptors will increase the anti-proliferative activities of the derivatives while decreasing VR1-DZv and MOMI-R descriptors would also increase the activities of 2-anilinopyrimidine derivatives as a standard anti-breast cancer agent. The robustness, applicability, and predicted capacity of the model generated were analyzed for both internal and external validation test which conforms to the minimum recommended values. This indicates that model 4 can be used in developing new 2-anilinopyrimidine derivative compounds with better anti-breast cancer activity. The molecular docking result showed that compounds 15 and 18 had the highest docking score of -7.4 and -7.3 kcal/mol, when it is compared to the standard drug gefitinib. From the studies, it is proven that some series of 2-anilinopyrimidine derivative compounds bind tightly to the receptor, stabilizing the receptor (TR β 1) which is evident from the receptor–ligand interactions. The compounds would serve as the most promising inhibitors against TR β 1. This research would be a breakthrough for pharmaceutical researchers in designing and developing new anti-triple-negative breast cancer drugs.

Abbreviations

QSAR: Quantitative structure-activity relationship; TNBCs: Triple-negative breast cancers; TR β 1: Thyroid hormone receptor

Acknowledgements

The authors acknowledge the technical effort of the Chemistry Department, Ahmadu Bello University, Zaria, for their enormous contribution and guidance toward the successful completion of this research article.

Consent of publication

Not applicable

Author's contributions

HAL gathered the data set and analyzed it computationally in accordance with the methodology to get a model of high performance and was a major contributor in drafting the manuscript. AU carried out the statistical analysis on the model to ensure its stability and also participated in the write up, and SU re-edited the work and ensured it conforms with the manuscript's guide. All the authors read and approved the final manuscript.

Funding

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

Availability of data and materials

Not applicable

Ethics approval and consent to participate

Not applicable

Competing interests

The authors declare that they have no competing interests.

Received: 9 October 2019 Accepted: 24 February 2020

Published online: 26 May 2020

References

1. Lan J, Huang L, Lou H, Chen C, Liu T, Hu S et al (2018) Design and synthesis of novel C14-urea-tetrandrone derivatives with potent anti-cancer activity. *European journal of medicinal chemistry* 143:1968–1980
2. Putri, D. E., Pranowo, H. D., & Haryadi, W. I. N. A. R. T. O. (2019). Study on anti-tumor activity of novel 3-substituted 4 anilino-coumarin derivatives using quantitative structure-activity relationship (QSAR). In *Materials Science Forum* (Vol. 948, pp. 101-108). Trans Tech Publications.
3. Ge W, Hao X, Han F, Liu Z, Wang T, Wang M et al (2019) Synthesis and structure-activity relationship studies of parthenolide derivatives as potential anti-triple negative breast cancer agents. *European journal of medicinal chemistry* 166:445–469
4. Jo J, Kim SH, Kim H, Jeong M, Kwak JH, Han YT et al (2019) Discovery and SAR studies of novel 2-anilinopyrimidine-based selective inhibitors against triple-negative breast cancer cell line MDA-MB-468. *Bioorganic & medicinal chemistry letters* 29(1):62–65
5. Aranda A, Martínez-Iglesias O, Ruiz-Llorente L, García-Carpizo V, Zambrano A (2009) Thyroid receptor: roles in cancer. *Trends in Endocrinology & Metabolism* 20(7):318–324
6. Martínez-Iglesias, O., Ruiz-Llorente, L., Jurado, C. C., & Aranda, A. (2014). Thyroid hormone receptors and their role in cell proliferation and cancer. In *Cellular Endocrinology in Health and Disease* (pp. 1-17). Academic Press.
7. Abdullahi M, Uzairu A, Shallangwa GA, Mamza P, Arthur DE, Ibrahim MT (2019) In-silico modelling studies on some C14-urea-tetrandrone derivatives as potent anti-cancer agents against prostate (PC3) cell line. *Journal of King Saud University-Science*
8. Adeniji SE, Uba S, Uzairu A (2018) Quantitative structure-activity relationship and molecular docking of 4-alkoxy-cinnamic analogues as anti-mycobacterium tuberculosis. *J King Saud University-Science*
9. Arthur DE, Uzairu A, Mamza P, Abechi SE, Shallangwa G (2018) Activity and toxicity modelling of some NCI selected compounds against leukemia P388ADR cell line using genetic algorithm-multiple linear regressions. *J.King Saud University-Science*
10. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* 32:1466–1474
11. Abdulfatai U, Uzairu A, Uba S (2018) Molecular docking and quantitative structure-activity relationship study of anticonvulsant activity of aminobenzothiazole derivatives. *Beni-Suef University Journal of Basic and Applied Sciences* 7(2):204–214
12. Kennard RW, Stone LA (1969) Computer aided design of experiments. *Technometrics* 11:137–148
13. Friedman JH (1991) Multivariate adaptive regression splines. *Ann. Stat.* 1–67
14. Khaled KF, Abdel-shafi NS (2011) Quantitative structure and activity relationship modeling study of corrosion inhibitors: genetic function approximation and molecular dynamics simulation methods. *Int. J. Electrochem. Sci.* 6:4077–4094
15. Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *Mol. Inform.* 22:69–77
16. Brandon V, Orr, A., 2015. Comprehensive R archive network (CRAN): <http://CRAN.Rproject.org>.
17. Minovski N, Z'uperl Š, Drgan V, Novič M (2013) Assessment of applicability domain for multivariate counter-propagation artificial neural network predictive models by minimum Euclidean distance space analysis: a case study. *Anal. Chim. Acta* 759:28–42

18. Myers RH (1990) Classical and modern regression application. Duxbury Press, CA
19. Eriksson L, Jaworska J, Worth AP, Cronin MT, McDowell RM, Gramatica P (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. *Environmental health perspectives* 111(10):1361–1375
20. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK (2011) Validation of QSAR models-strategies and importance. *Int.J.DrugDes.Discov* 3:511–519
21. Ibrahim MT, Uzairu A, Shallangwa GA, Ibrahim A (2018) In-silico studies of some oxadiazoles derivatives as anti-diabetic compounds. *J King Saud University-Science*

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
