

RESEARCH

Open Access



Chemo-informatics activity prediction, ligand based drug design, Molecular docking and pharmacokinetics studies of some series of 4, 6-diaryl-2-pyrimidinamine derivatives as anti-cancer agents

Sagiru Hamza Abdullahi*, Adamu Uzairu, Muhammad Tukur Ibrahim and Abdullahi Bello Umar

Abstract

Background: The most well-known cause of cancer deaths identified in female is breast cancer. Several drugs approved by the food and drug administration (FDA) for the treatment of breast cancer may have adverse health effects. This research is aimed at developing a QSAR model and utilize it to predict the inhibitive activities of newly designed novel compounds, examine their ADMET and drug-likeness properties and carry out molecular docking studies between the designed compounds and the VEGFR-2 receptors in order to identify the essential amino acid residues involved in protein–ligand interactions and possible mechanism of action of the designed compounds.

Results: The first model was selected as the best because of its fitness statistically with the following assessment parameters: $R^2_{\text{train}} = 0.832$, $R^2_{\text{adj}} = 0.79$, $R^2_{\text{ext}} = 0.62$, $Q^2 = 0.68$, and $\text{LOF} = 0.14509$. Compound 11 was selected as a template to design new powerful compounds based on its low residual and high pIC_{50} values. Majority of the designed compounds has predicted pIC_{50} greater than that of the lead compound and the standard drug (Sunitinib) used as reference. Molecular docking studies results of the designed compounds revealed that they have higher docking scores than the template and the reference drug (Sunitinib) and are found to bind to the VEGFR-2 receptor in a similar manner to the reference drug. Pharmacokinetics and ADMET properties revealed that the designed compounds passed drug-likeness criteria because they did not violate more than 1 Lipinski's rule of Five, They are uniformly distributed to the brain and are assumed to penetrate the central nervous system and finally they are all found to non-toxic and orally bioavailable.

Conclusion: The developed model was therefore found to be efficient in predicting the pIC_{50} of Anti breast cancer compounds that are yet to be synthesized and it also help in reducing the cost and synthetic duration the compounds. The result of this research confirmed that the designed compounds may be developed as novel VEGFR-2 inhibitors.

Keywords: HUVEC cell line, DFT, GFA-MLR, QSAR, Williams plot, VIF, ME, ADME

Background

Cancer is one of the main reasons of death globally nowadays. The mortality rate as a results of numerous kinds of cancer continues to skyrocket globally with an estimated 12 million deaths in 2030 (Solomon et al. 2009). Cancer

*Correspondence: sagirwasai@gmail.com
Department of Chemistry, Faculty of Physical Sciences, Ahmadu Bello University, P.M.B. 1045, Zaria, Kaduna State, Nigeria

cells are unique from exceptional regular counterparts in a variety of biochemical processes, particularly during the cell division and growth control. One attribute of most cancer cells, that distinguishes them from other ordinary cells, is their high proliferative index. As a result, focusing on proliferative pathways which consequences in cell death through apoptosis is regarded as an effective way for fighting this disease (Chandrappa et al. 2009). The most well-known kind of cancers identified in female is breast cancer. Prone populace of breast cancer have frequent characteristics, which include advanced age, low parity, delayed age at first delivery, short period of breastfeeding, overeating, constrained exercising and so on. Breast cancer is a heterogeneous disease and many sub kinds have been defined (Liu et al. 2018). Estrogen receptor α (ER α) which is a member of most of the categories of nuclear receptors has been recognized as the prime internal reasons for the disease, which involves at least 70% breast cancer patients, and this kind of patients are identified as ER positive (ER⁺) (Liu et al. 2018). Tamoxifen, and raloxifene which are selective estrogen receptor modulators (SERMs) that rival with estradiol for binding with ER α in breast tissue are mostly used for the remedy of ER⁺ breast cancer (Wang et al. 2009). However, due to their adverse health effects in other tissues, many SERMs may have acute side effects, such as endometrial cancer (Leeuwen et al. 1994). Moreover, approximately 50% of ER-positive tumors patients either in the beginning does not react or developed resistance to these medications in a period of first five (5) years of remedy (Clarke et al. 2015). Human umbilical vein endothelial cell (HUVEC) are cells derived from the endothelium of veins from the umbilical cord. They play an important function as a standard scenario for the research on the regulation of endothelial cell function and are favorable for the evaluation of anti-angiogenesis impact by anti-proliferative investigation. It is regarded that vascular endothelial growth factor receptor-2 (VEGFR-2) found in HUVEC cells had a significant purpose in the angiogenesis route that take parts in the conversion, acceleration, and infringement of breast cancer cells (Hicklin and Ellis 2005). The receptor was found to play a key role in tamoxifen resistance through the Ras/MAPK route as suggested by many studies (Huang et al. 2008). Low dose of brivanib alaninate, which is a VEGFR-2 inhibitor in combination with tamoxifen was reported to magnify therapeutic efficacy and also to decelerate selective estrogen receptor modulator (SERM) resistant cancer elevation (Patel et al. 2010). The role of the 2-pyrimidinamine frame in medicinal chemistry is popular. Many of these compounds serves as anticancer agent. 4,6-diaryl-2-pyrimidinamine derivatives suggest good activity in the field of medicine. 2-Pyrimidinamine scaffold is

the fundamental frame of pazopanib and JNJ-17029259 which are regarded as VEGFR-2-inhibitors (Liu et al. 2018).

Optimal anti-cancer tablets are supposed to annihilate cancer cells without causing a detrimental damage on normal tissues (Al-Suwaidan et al. 2016). But these drugs can cause damage or completely destroy some ordinary proliferating cells hence, global search for identifying new higher quality drugs that are safe for the prevention and remedy of cancer became necessary.

Quantitative Structure–Activity Relationships (QSAR) are computational relations that correlates the biological activities (response variable) of chemical compounds with their molecular structures (Independent variables) in a quantitative way (Hansch et al. 1995). The molecular descriptors consist of parameters that account for conformational, constitutional, thermodynamic, steric effects and electronic properties of a molecule (Umar et al. 2019).

The approach is mostly employed to determine the properties of new chemical species prior to their synthesis (Abdulfatai et al. 2017).

Furthermore, QSAR strategy reduces the extensive variety of the synthesizable compounds with the aid of assisting in figuring out the most promising candidates and there by means of reducing the prolonged time and cost of drug production (Ibrahim et al. 2020).

The main purpose of this research is to develop a QSAR model with regards to the compounds received from literature and use the model to predict the inhibitive activities of compounds prior to their synthesis, design new novel compounds and predict their activities using the model, perform molecular docking studies of the designed novel compounds with VEGFR-2 receptor kinase and to examine the Absorption, Distribution, metabolism and excretion (ADME) and drug-resemblance properties of the designed novel compounds.

Methods

Data collection

A library of thirty (30) 4, 6-diaryl-2-pyrimidinamine analogues as cancer agents with their inhibitive capacities (IC₅₀) measured in μM are obtained from the research work of Liu et al. (2018). The inhibitive capacities of all the compounds were normalized by taking the negative logarithm to base 10 using Eq. (1) (Ibrahim et al. 2020).

$$\text{PIC}_{50} = -\text{Log}\left(\text{IC}_{50} \times 10^{-6}\right). \quad (1)$$

Geometry optimization

Chemdraw 12.0 software was employed to sketch the 2-Dimensional structure of all the 30 molecules in the

data set. Prior to energy minimization, 2-Dimensional constructions of the molecules were automatically transformed to 3-Dimensional using Spartan 14 software program. Energy minimization was done to reduce constraint in the structures prior to finding the most stable conformation of the considered molecules on potential energy surface (Ibrahim et al. 2020).

Density functional theory (DFT) quantum mechanical calculation utilizing Bee -3- Lee Yang Par (B3LYP) method and 6-31G* basis set present in the Spartan 14.0 software application was used for Geometry optimization of all the thirty (30) compounds. Geometry optimization was conducted in order to locate the most stable structures of all the studied molecules on global minima on the potential energy surface (Ibrahim et al. 2020). New folder was created and the completely optimized structures were saved in Spatial Document File (sdf) format. In order to compute thermodynamic, autocorrelation, topological, electronic, constitutional, and geometric descriptors, the least energy 3D structure in Spatial Document File (sdf) format was then imported into PaDEL descriptor software program and the calculation was conducted (Umar et al. 2019).

Data pretreatment and division

The results of calculated descriptors of all the 30 compounds in the data set have been pretreated using Data pre-treatment 1.2 application software and then manually to get rid of constant and redundant molecular descriptors. Data partitioning software program was then employed to divide the pretreated data set into modelling (training) and validation (test) sets. The model was developed using modeling (training) set while the validation (test) set was used to testify the selected model. In this research work 22 compounds samples were used as the model building (training) set and the remaining 8 molecules as the validation (test) set. This partitioning certify that a related principle can be employed to predict the activity of the validation set. Kennard–Stone Algorithm was utilized for partitioning the data samples into a modelling and validation set (Kennard and Stone 1969; Rajer-Kanduč et al. 2003).

Model development

Genetic function algorithm (GFA) approach present in material studio 8.0 was employed in building the models with actual pIC₅₀ values as the response (dependent) variables and the molecular descriptors as the independent variables. The length of the regression equation was 4, and Population and Generation were set to 1000 and 1000, respectively. The number of top equations returned was 4. Mutation probability was 0.1, and the smoothing parameter

was 0.5. The models were scored primarily based on Friedman's Lack of Fit (LOF) (Khaled 2011).

It is a phenomenal feature of GFA that instead of creating a single model it could create a vast number of models. GFA algorithm, handpicks the most relevant descriptors genetically, develop a far better models than those developed through the utilization of stepwise regression methods. The models were estimated using the LOF, which was measured using a slight variation of the original Friedman formula, so that the best fitness score can be received (Abdufatai et al. 2017). Lack of fit is estimated using the following formula:

$$\text{LOF} = \text{SSE} / \left(1 - \frac{c + dp}{M} \right)^2 \quad (2)$$

where SSE denotes the sum of squares of errors, c is the number of terms in the selected model, apart from the constant term, d is a user defined smoothing parameter, p is the total number of molecular descriptors contained in all model terms (without the constant term) and M is the number of samples in the modeling set. However, for a model to be robust the value of the sum of square of errors must be small. Equation (3) below is used to compute The SSE:

$$\text{SSE} = \frac{(Y_{\text{exp}} - Y_{\text{prep}})^2}{\sqrt{M - P - 1}} \quad (3)$$

where Y_{exp} and Y_{pred} are the actual and computed pIC₅₀ values of the modelling set samples, M is the number of samples in the model building data set and P is the number of independent variables present in the generated model (Troyer 2001).

Model validation

Internal and external validation parameters were utilized in order to testify the reliability and predictive capability of the developed QSAR models (Table 1).

Internal and external validation

For the quantitative assessment of the developed QSAR model, the internal and external validation parameters were compared with the minimum recommended values (Veerasamy et al. 2011) as depicted in Table 2. The most commonly used internal validation tool is the squared correlation coefficient (R^2), for an excellent regression equation, the value of R^2 should be close to unity. It is calculated using Eq. (4) below:

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

Table 1 Actual and computed pIC₅₀ values of 4,6-diaryl-2-pyrimidinamine series against HUVEC cancer cell line

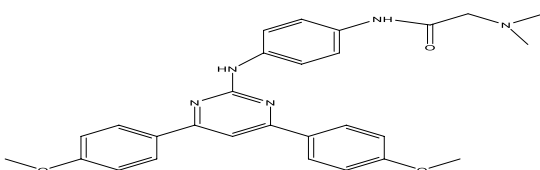
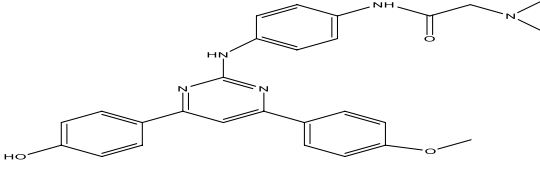
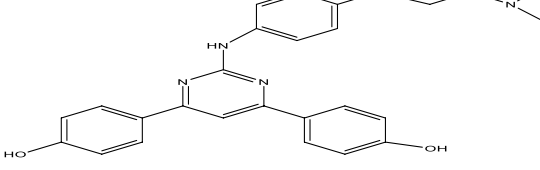
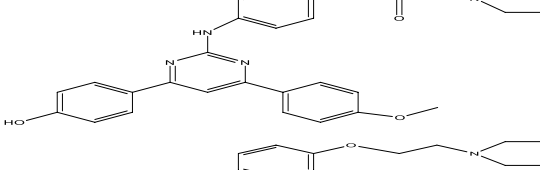
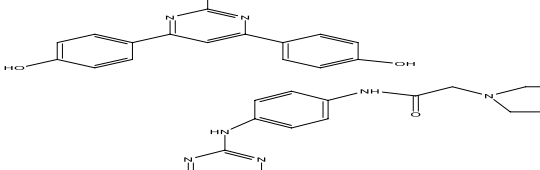
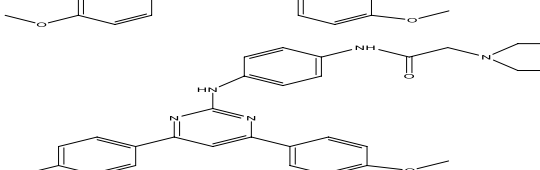
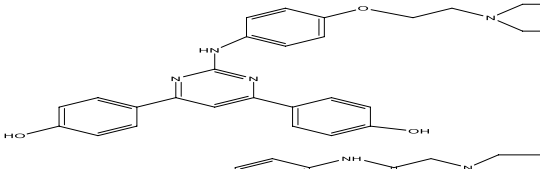
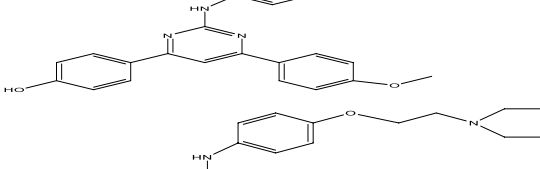
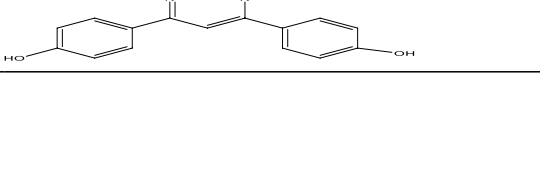

S/no	Formula	pIC ₅₀	Pred pIC ₅₀	Residual
1 ^b		5.55	5.63	-0.08
2 ^a		5.42	5.61	-0.19
3 ^a		4.22	4.41	-0.19
4 ^a		4.66	4.62	0.04
5 ^a		4.72	4.75	-0.03
6 ^a		5.60	5.56	0.04
7 ^a		5.64	5.49	0.15
8 ^b		5.15	4.91	0.24
9 ^a		4.97	5.34	-0.37
10 ^a		5.13	4.98	0.15

Table 1 (continued)

S/no	Formula	pIC50	Pred pIC50	Residual
11 ^a		4.99	5.08	-0.09
12 ^a		4.33	4.45	-0.12
13 ^a		5.3	5.38	-0.08
14 ^a		5.33	5.21	0.12
15 ^a		5.17	5.15	0.02
16 ^a		5.34	5.41	-0.07
17 ^a		5.33	5.19	0.14
18 ^a		5.33	5.17	0.16
20 ^b		5.68	5.60	0.08
21 ^a		5.51	5.43	0.08

Table 1 (continued)

S/no	Formula	pIC50	Pred pIC50	Residual
22 ^a		5.33	5.18	0.15
23 ^b		5.27	5.59	-0.32
24 ^a		5.35	5.48	-0.13
25 ^a		5.34	5.11	0.23
26 ^a		5.11	5.02	0.09
27 ^b		4.13	4.71	-0.58
28 ^a		5.52	5.72	-0.2
29 ^b		5.62	5.43	0.19
30 ^b		5.60	5.20	0.4

^a Training set^b Test set

Table 2 Standard minimum proposed value used to assess a quantifiable QSAR model

Symbol	Name	Value
R^2	Correlation coefficient	≥ 0.6
$P_{(95\%)}$	Confidence interval at 95% confidence level	< 0.05
Q^2	Cross validation coefficient	≥ 0.5
$R^2 - Q^2$	Difference between R^2 and Q^2	≤ 0.3
$N_{\text{ext test set}}$	Minimal number of external validation set	≥ 5
R^2_{ext}	Correlation coefficient for external validation set	≥ 0.6

where Y_{exp} , Y_{pred} and Y_{train} depicts the actual, computed and the mean actual biological capabilities of the modelling set samples (Abdulfatai et al. 2017).

Adjusted R^2 (R^2_{adj}): R^2 value vary literally with the rise in the population of independent variables; thus, R^2 alone is not sufficient recommended criterion for the quality of model fit. Hence, R^2 is altered for the number of elucidative variables in the model (Umar et al. 2019). The altered R^2 is defined as in Eq. (5):

$$R^2_{\text{adj}} = 1 - (1 - R^2) \frac{N - 1}{N - P - 1} = \frac{(N - 1)R^2 - P}{N - P + 1} \quad (5)$$

where P =number of independent variables in the model and N =number of model building data set (Abdulfatai et al. 2017). The standard approved value for this parameter is presented in Table 2. Another important internal validation parameter is the Cross-validation coefficient parameter (Q^2_{cv}) which is computed using Eq. 6.

$$Q^2_{\text{cv}} = 1 - \frac{\sum(Y_{\text{pred}} - Y_{\text{exp}})^2}{\sum(Y_{\text{exp}} - Y_{\text{mtrain}})^2} \quad (6)$$

where Y_{pred} , Y_{exp} and Y_{mtrain} are the predicted, actual and average values of experimental activities of modeling set samples, respectively. It has been proclaimed that excessive estimation of statistical peculiarities is not sufficient to rationalize the functionality of a model, the strategy depicted by Golbraikh and Tropsha (2002), and Roy et al. (2015) were utilized in order to verify the predictive capability of the new QSAR model. The squared correlation coefficient of the test set R^2_{test} was calculated by Eq. (7):

$$R^2_{\text{test}} = 1 - \frac{\sum(Y_{\text{exp}} - Y_{\text{pred}})^2}{\sum(Y_{\text{exp}} - Y_{\text{mtrain}})^2} \quad (7)$$

where Y_{pred} , Y_{exp} are the predicted and experimental values of experimental activities of test set compounds, Y_{mtrain} = Mean value of biological activities of training set compounds, respectively.

Molecular docking studies of the designed compounds against VEGFR-2 receptors

Molecular docking studies between the template, the designed compounds, and the reference drug (Sunitinib) against the VEGFR-2 receptors was carried out to evaluate the fundamental amino acid residues accountable for the protein–ligand interactions and probable mechanism of action of the designed molecules. The geometrically optimized 3D structures of all the ligands were saved in Protein Data Bank (PDB) format. The 3D structures of the VEGFR-2 receptor kinase (pdb code: 4agd) co-crystallized with Sunitinib ligand was downloaded from the Protein data bank (<https://www.rcsb.org>) and prepared using Molegro Virtual Docker (MVD) software via eliminating the extra water molecules and co-crystallized ligand enveloped in the X-Ray structure prior to the docking process. The template, all the newly designed molecules and Sunitinib were docked on to the active site of the VEGFR-2 receptor kinase using the Molegro Virtual Docker 6.0. The docking simulation was run a minimum of 50 times for 5 poses, and the best poses were determined based on the set scoring functions such as the MolDock and rerank score, (Jaworska et al. 2005). A Discovery Studio (DS) Visualizer Version 3.5 was employed to visualize the various intermolecular interactions such as H-bond, hydrophobic, and aryl interactions.

ADME properties and drug likeness prediction of some selected designed compounds

pkCSM an online web server (<http://structure.bioc.cam.ac.uk/pkcsml>), and SwissADME (<http://www.swissadme.ch/index.php>) are accessible web tools that are designed to analyze ADMET and drug-resemblance properties of small molecules (Daina and Michielin 2017). They are utilized to figure out the novel drug candidate, to lower the number of experimental researches and to elevate the success rate. ADMET properties and drug-resemblance prediction of some selected designed compounds as anti-breast cancer agents was conducted by utilizing the web tools. One of the most important parameter at pre-clinical stage of drug discovery is the Lipinski's rule of five, it proposed that for a chemical compound to be permeable or readily absorbed in to the body system it shouldn't violates greater 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) $< 140 \text{ \AA}^2$) (Ismail et al. 2018).

Results

For the accessment of the powerfulness and statistical significance of the developed model, Kennard–Stone algorithm was employed to partition the data set into

modeling and validation set samples. Four model generated were generated by utilizing genetic functional algorithm (GFA) among which model 1 was selected due to its fitness statistically.

Model 1

$$\begin{aligned} \text{PIC}_{50} = & -5.898694472 * \text{AATS2s} \\ & - 7.592394612 * \text{AATSC5s} \\ & - 14.694225303 * \text{MATS3e} \\ & - 5.537580107 * \text{SpMax8_Bhi} \\ & + 38.59916. \end{aligned}$$

$$\begin{aligned} N_{\text{train}} = 22, \quad R_{\text{train}}^2 = 0.832, \quad R_{\text{adj}}^2 = 0.79, \quad R_{\text{ext}}^2 = 0.62, \\ Q^2 = 0.68, \quad \text{LOF} = 0.14509, \quad N_{\text{test}} = 8. \end{aligned}$$

Discussion

Four models were developed among which model 1 was chosen to predict the inhibitive capacity of the molecules because of its quality statistically as it has the best correlation coefficient (R^2) of 0.83, adjusted correlation coefficient (R_{adj}^2) value of 0.79, Leave one out (LOO) cross validation coefficient (Q^2) value of 0.68 and the external confirmation (R_{ext}^2) of 0.62. Referred to Table 2 above, the internal and external attestation parameters of model 1 agreed with the minimal criterion for any credible and powerful QSAR model.

The selected model was employed to compute the anti-proliferative capacity of test set data, and the results was placed in Table 1. The closeness of correlation coefficient (R^2) to unity 1.0 indicates that the model explained a reasonably excessive proportion

of the response variable (descriptor) variation, sufficient enough for a strong QSAR model. Its 0.832 value suggested that 83.2% of the deviation resides in the residual which implies that the model is promising. The excessive modified R^2 (R_{adj}^2) of the model and its proximity in value to the R^2 implies that the model has remarkable descriptive ability to the independent variables (descriptors) in it. Additionally, it reveals the actual impact of utilized descriptors on the pIC_{50} . Also, the excessive and closeness of Q_{cv}^2 to internal R^2 revealed that the model was found not to be over-fitted. The high R_{test}^2 of the model clarified that the model is able to deliver a reliable predictions for newly designed molecules.

Figure 1 is a plot of the predicted pIC_{50} for the modelling and validation sets against the experimental pIC_{50} values for the Inhibition of HUVEC cancer cell line. Additionally, the residuals values of the model building and validation sets were plotted against the experimental pIC_{50} values and is presented in Fig. 2. The calculated normalized values of the activities (pIC_{50}) strongly agreed with those of the test set as sighted in the table and figures, therefore the model did not illustrate any relative and systematic error, since the agreement of the residuals on either side of zero is irregular (Table 3).

Variance inflation factor

The degree of inter-correlation between descriptors is detected by Variance inflation factor. For a model to be acceptable VIF values must range from 1 to 10. Computed VIF values that ranges from 1 to 5 indicates the model is acceptable, value less than 1 it suggests that



Fig. 1 Predicted pIC_{50} versus experimental figures of the modelling and validation set

there is no inter-correlation among the descriptors, and a value that exceeds 10, suggests that the model is not acceptable. VIF is computed using Eq. (8) below:

$$VIF = \frac{1}{1 - R^2} \tag{8}$$

where R^2 is the correlation coefficient of the chosen model (Ibrahim et al. 2020). VIF values of the four molecular descriptors that appear in the selected model are highlighted in Table 4. It can be inferred from the table that since all the variables has VIF values is less than 10,

then that the generated model was suggestive statistically, and that the descriptors were found to be reasonably orthogonal(Myers 1990).

Mean effect

The assessment of the role and contribution of each descriptor in the generated model is carried out by adopting their mean effect value. It provides key informations on the impact of the response variables (descriptors) on the generated model, the signal and the magnitude of these descriptors blended with their

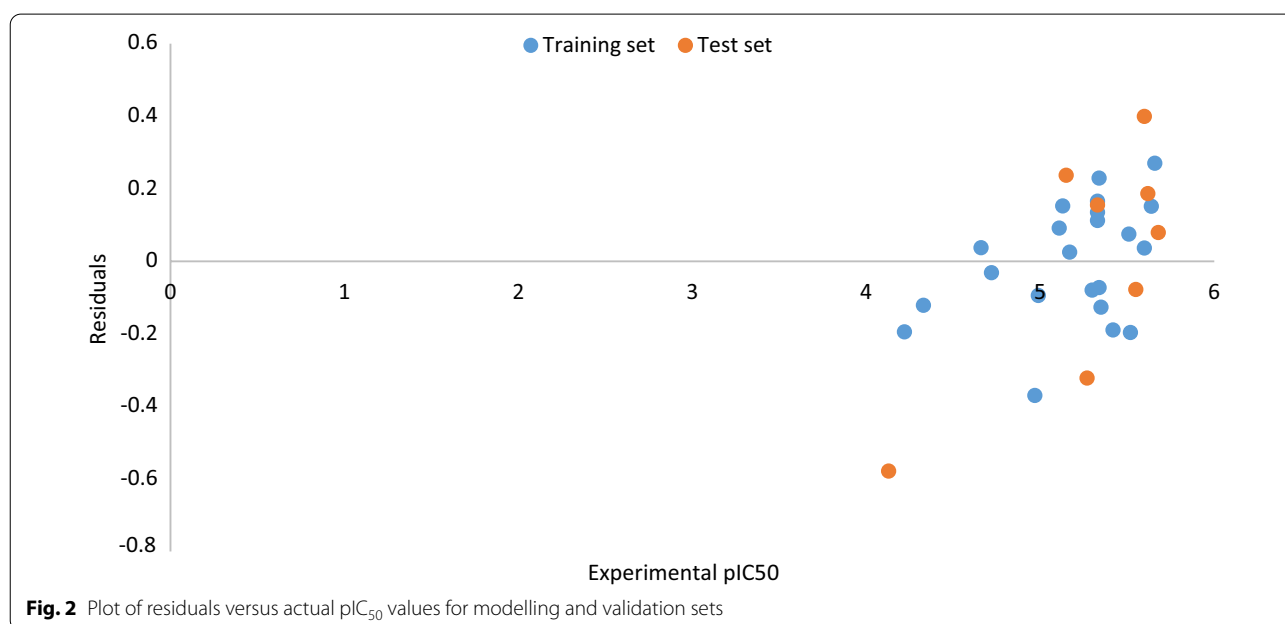


Table 3 List of independent variables (descriptors), their descriptions and class for model 1

S/NO	Symbol	Descriptions	Class
1	AATS2s	Average Broto-Moreau autocorrelation – lag 2/weighted by l-state	2D
2	AATSC5s	Average centered Broto-Moreau autocorrelation – lag 5/weighted by l-state	2D
3	MATS3e	Moranautocorrelation-lag3/weighted by Sanderson electronegativities	2D
4	SpMax8_Bhi	Largest absolute eigenvalue of Burden modified matrix – n 8/weighted by relative first ionization potential	2D

Table 4 Pearson correlation matrix, VIF and ME of the selected model

	AATS2s	AATSC5s	MATS3e	SpMax8_Bhi	VIF	ME
AATS2s	1				7.247738	0.534996
AATSC5s	0.57371	1			3.283985	–0.00206
MATS3e	0.46022	–0.01984	1		2.351116	–0.05061
SpMax8_Bhi	0.61146	0.037807	0.072472	1	3.271821	0.517679

mean effect values signifies their powerfulness in determining the activity of a molecule (Arthur et al. 2018).

Equation (9) below is used to calculate the mean effect of each descriptor:

$$MF_j = \frac{\beta_j \sum_{i=1}^n d_{ij}}{\sum_j^m \beta_j \sum_i^n d_{ij}} \quad (9)$$

where MF_j portrays the mean effect of a descriptor j in a model, β_j represents the coefficient of the descriptor J in the model and d_{ij} is the value of the descriptor in the data matrix per sample in the modelling set, m illustrates the independent variables numbers that turn up in the model and n is the number of samples in the modelling set (Adedirin et al. 2018). A short definitions of each of the response variable are depicted in Table 3 and corresponding mean effect value are in Table 4 respectively.

From Table 4, the most important descriptor is AATS2s because it has the best possible mean effect value, this suggested that it has a remarkable impact on the pIC_{50} values of the molecules. The correlated descriptors were categorized in a sequence in accordance with their offerings towards the standard pIC_{50} of the compounds, in the following decreasing sequence.

AATS2s > SpMax8_Bhi > MATS3e > AATSC5s.

AATS2s descriptor which is defined as Average Broto-Moreau autocorrelation – lag 2/weighted by I-state is an auto correlation descriptor. According to this descriptor atomic masses and electronic dispersion of the atoms that made up the molecule had a tremendous impact on the Anti-cancer capability of the set of molecules. Its largest positive mean effect value indicates that increase in its descriptor's value will elevate the compound's antifroliferative activity against HUVEC cell line. Also the descriptor SpMax8_Bhi has positive signal as indicated in Table 4, the signal suggested that the anti-breast cancer activity of a molecules varies directly with its values. The other descriptors MATS3e, AATSC5s are having negative mean effect values, and it indicates that the activity of the molecules varies inversely with the values of these descriptors.

Applicability domain

One of the approaches used to examine whether there are substantial or irrelevant molecules in a particular set of data is applicability domain. A QSAR model is deemed to be acceptable if it is able to render a dependable formulation of fresh inhibitive capacities of both modelling, validation sets samples when subjected to

the applicability domain (AD) (Golbraikh and Tropsha 2002). Leverage approach is among techniques utilized to evaluate the AD and it is expressed in Eq. (10).

$$h_i = x_i \left(X^T X \right)^{-1} x_i^T \quad (10)$$

The terms x_i , X , and X^T represents the model building set matrix I , $n \times k$ descriptor matrix of the modelling set and the transpose matrix X used in the model development. Lower-limit value of X is the cut-off (h^*) leverage which is presented in the equation below:

$$h^* = \frac{3(p+1)}{N} \quad (11)$$

where p and N are the numbers of independent variables used in developing the model, and the samples used in developing the model.

A plot of the standardized residuals versus the leverage values (h) is called the Williams's plot. This plot is used to analyze the defined applicability domain (AD). A compound whose leverage exceeds the cut-off value has a severe influence on the performance of the model and may be eliminated, but due to the fact that its standardized residual may be minimal it does not meant to be an outlier. Additionally, the cutoff value for accepting predictions of a molecule lies within the range of -3 to $+3$. This is because any points that resides within ± 3 standardized residual from the mean cover ninety-nine percent (99%) of the generally expended data (Jaworska et al. 2005).

Figure 3 shows the Williams plot of the developed. The cutoff leverage for the selected model was 0.682. Four (22, 1, 23 and 27) compounds from the test set were found to have leverage values more than the cut off value (i.e. $h_i > 0.682$), therefore they had been recognized as structurally outliers compounds.

Ligand based drug design

Information derived from the model enabled us to design five new potent compounds. Two molecular descriptors were assumed to play a major role in our design because they have a notable value of mean effect when compared to the other descriptors, AATS2s and SpMax8_Bhi molecular descriptors were the principal descriptors used for our design. Compound 11 was selected as our template for the design because of its low residual and high pIC_{50} values and seemed to stand inside the described domain applicability. Figure 4 represents the template compound (11) and the standard compound's structures that are utilized for our ligand based design. Modification of the compound was achieved by addition and switching of several substituents on the lettered positions (i.e. X , Y and R_1) so

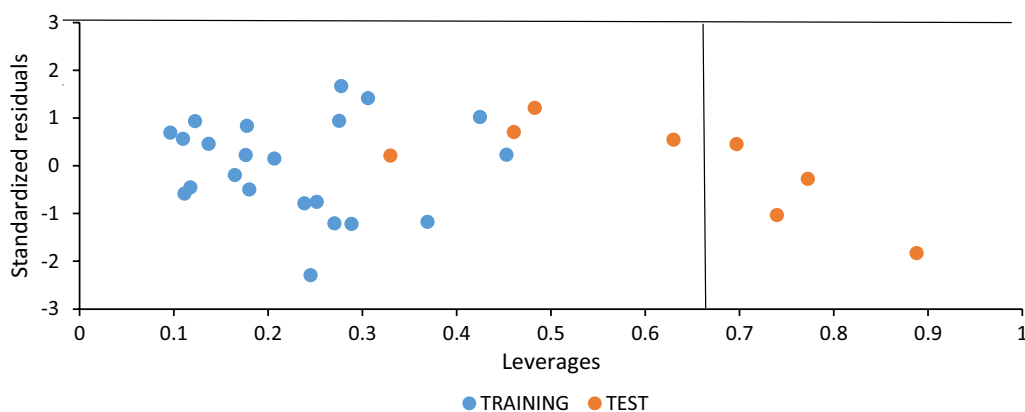


Fig. 3 The Williams Plot of the selected model

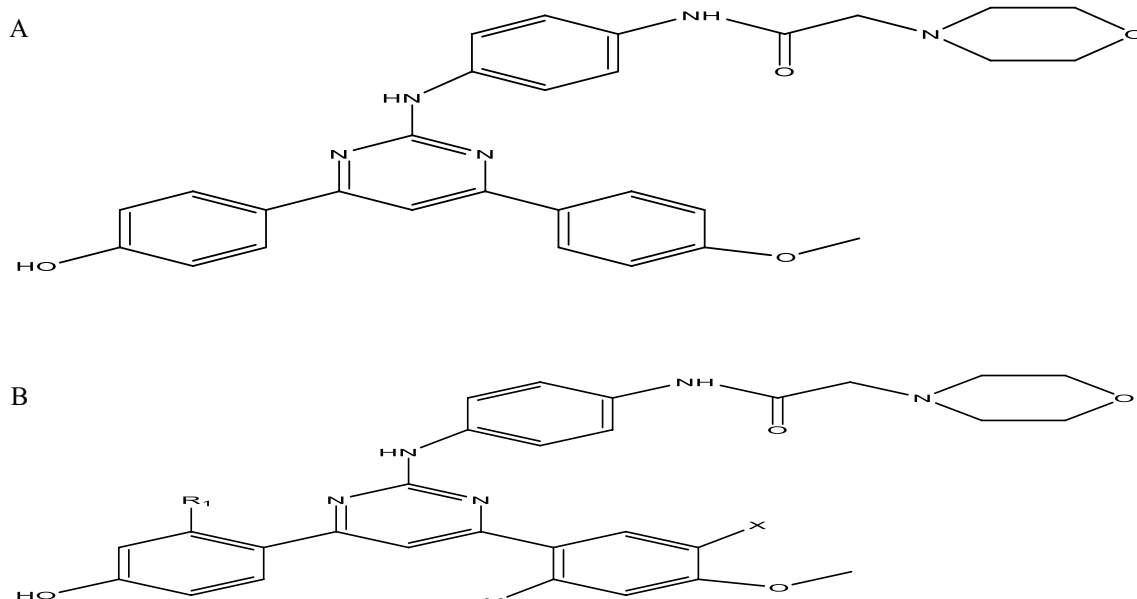


Fig. 4 **A:** Structure of the lead compound (11) used for the design. **B:** structure of the template used for the design

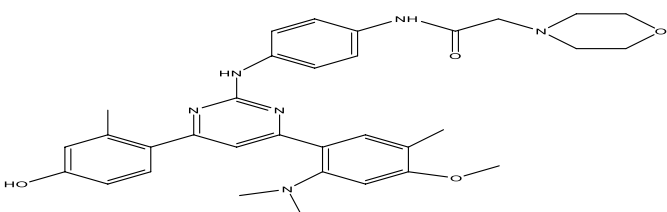
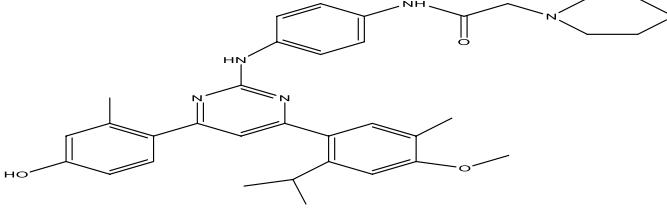
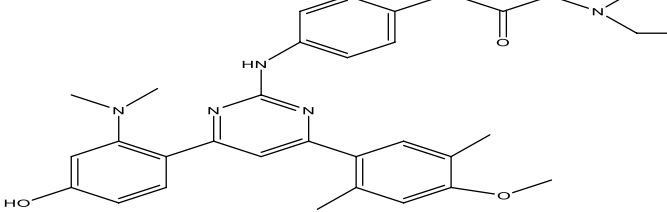
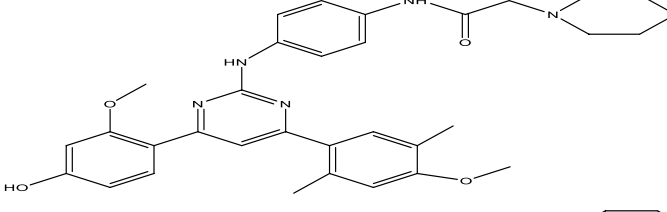
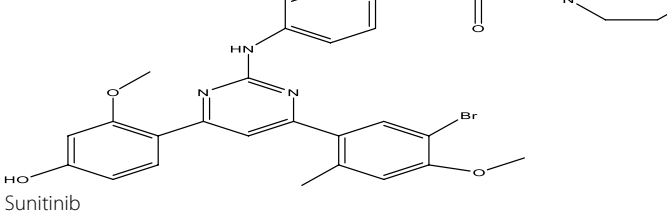
that experimental synthesis of new active molecules will be feasible. Table 5 represents the structure of the newly designed molecules and their predicted pIC₅₀ values, from the table it is observed that majority of the composed molecules had greater predicted activities relative to the principal molecule (11) modified for this design and the reference drug (Sunitinib). Hence, it can be affirmed that a simple QSAR model is able to provide an opportunity of predicting and identifying molecules with satisfactory capability, also to pin point the structurally modified compounds that lies beyond the defined applicability domain. Lastly, Outcomes of this study ascertains how powerful and dependable the

selected QSAR model is and also conveys that together with the application of in silico screening technique, the selected QSAR model is able to perceive new powerful molecules as synthetic targets for drug advancements.

Molecular docking studies of the designed compounds

The potential of the designed compounds to interact with the VEGFR-2 receptor is presented in terms of MolDock score and rerank score respectively. The MolDock and rerank scoring are adopted as the parameters for evaluating the docking results. The outcomes of docking studies of the compounds against VEGFR-2 receptor illustrates that they were docked at the binding site of the receptor

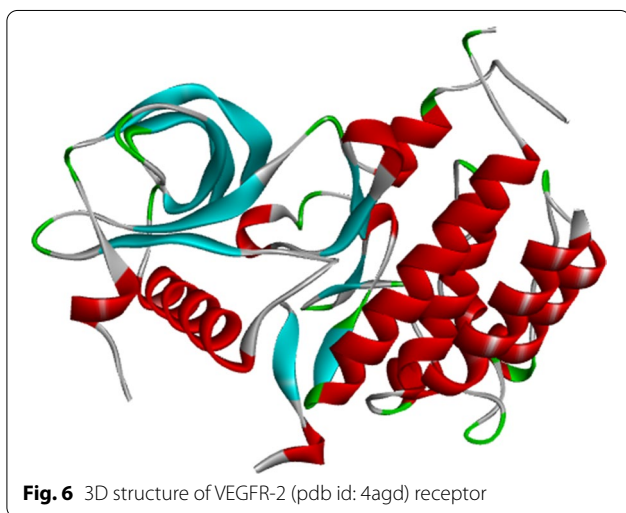
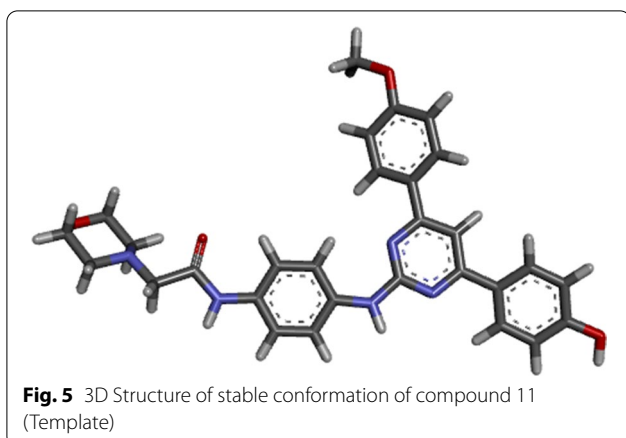
Table 5 2D structures, predicted activities of the designed of 4, 6-diaryl-2-pyrimidinamine derivatives

S/NO	Formula	Pred pIC ₅₀
1		5.858984
2		5.899997
3		5.160579
4		5.305695
5		5.0296465.522879
	Sunitinib	

with a favorable MolDock score and rerank score compared to Sunitinib. 3D structures of the template and the VEGFR-2 receptor (pdb id: 4agd) are shown in Figs. 5 and 6 respectively. Additionally, the docking simulation results and several interactions of the template, the designed compounds and Sunitinib with VEGFR-2 receptor kinase is presented in Table 6 respectively.

The template compound is bound to the receptor via three conventional Hydrogen bonds between Hydrogen atom of the Hydroxyl group with PHE845, and LEU 1049, Hydrogen atom of the Nitrogen attached to the carbonyl group with ARG1027. Four Carbon-Hydrogen bonds

via Carbonyl Oxygen with LEU1067 and PRO1068, Morpholine ring Oxygen with MET1072 and Morpholine ring Hydrogen with ARG1027. One pi-anion electrostatic interactions between the phenyl ring moiety with ASP1028. Two weak alkyl interactions are also observed between the morpholine ring with MET1072 and PRO1068. Other weak pi-Alkyl interactions are also observed (ALA844, LEU1049, ILE1053, ALA1065, ARG1027, and PRO1068). These interactions accounts for the high binding scores between the ligand and the VEGFR-2 receptors. 3D binding mode of the template with VEGFR-2 receptor is shown in Fig. 7 respectively.



Designed compound 1 has a MolDock and rerank score of -161.031 and -70.669 respectively it is bonded to the VEGFR-2 binding pocket via two conventional H-bonds, two alkyl one electrostatic interaction and 6 pi-alkyl bonds. The oxygen atom of the morpholine ring forms one Hydrogen bond with SER1090 and the other is formed between the Hydrogen of the OH- group attached with the phenyl ring and LEU1409. Electrostatic interaction exist between the phenyl ring moiety and ASP1028. Weak interactions such as alkyl with PRO1068 and MET 1072 and six π -alkyl interactions with ALA844, LEU1049, ILE1053, ALA1065, ARG1027 and PRO1068 exist between the Ligand and the receptor, these interactions account for its reasonable binding score. Figure 8 represents the 3D binding interactions of designed compound 1 with VEGFR-2 receptor.

The 3D binding interactions of designed compound 2 is shown in Fig. 9. It has a moldock and rerank score of -144.005 and -97.0301 respectively. It is bound to the receptor through two conventional H-bonds, nine

Carbon-Hydrogen bond, an electrostatic and eight pi-alkyl interactions. The two conventional H-bonds are formed between H-atom of the Hydroxyl group attached with the benzene ring with ASP1052 and Nitrogen atom attached to the carbonyl group with SER877. Two C-H bonds between the Nitrogen atom, and the phenyl ring of the Pyrimidine with LEU802, another between Hydrogen of the morpholine group and that of the Nitrogen attached to the carbonyl group ARG1051, SER877, SER803, ASP1046 and GLY1027 residues with the Amino Hydrogen, Alkoxy Oxygen with ARG1027. The Interactions between the residues and the receptor accounts for the high rerank score of the ligand.

3D interactions of Compound 3 with VEGFR-2 receptor is indicated in Fig. 10. It has a molDock score of -173.02 and rerank score of -86.399 forms two conventional H-bonds, four Carbon-Hydrogen bond, one electrostatic, five alkyl and six pi-alkyl interactions. The Hydrogen atoms of the Nitrogen attached to the carbonyl and that attached to the Hydroxyl groups formed Conventional H-bonds with ARG1027 and LEU1049. Carbonyl oxygen form C-H bond with PRO1068, ASP1028 with the morpholine Hydrogen, ARG1027 with morpholine and the Hydrogen attached to the alpha carbon of the carbonyl group. Electrostatic interaction occur between ALA1065 and delocalized pi electrons of the benzene ring. The other weak interactions occur with PRO1068, MET1072, ILE1053, LEU1067, TYR1054, ALA844, LEU1049, ILE1053, ALA1065, and PRO106.

Compound 4 with rerank score -52.7342 is bound to the receptor through 4 conventional H-bonds, 6 Carbon-Hydrogen bond, 1 electrostatic and 8 pi-alkyl interactions. Conventional Hydrogen bonds are formed between Oxygen atom of the OH group with SER1090, Nitrogen attached to the carbonyl group with GLY1048 and ARG1051, Nitrogen attached to the pyrimidine bond with ASP1028. Carbon-Hydrogen bonds are between the Oxygen atom of Hydroxyl group with ALA1065, Hydrogen atom of the morpholine group with ASP1052, Hydrogen of the Alpha carbon attached to the carbonyl with LEU1049, ARG1051 with morpholine and Alpha carbon Hydrogen, SER1086 with the methoxy Hydrogen. Other weak interactions include one pi-sulfur (MET1072), four alkyl interactions (ALA844, PRO1068, ARG1027 and MET1072), and ten π -Alkyl interactions (PHE845, TYR1054, TRP1071, TYR1082, ALA1065, LEU1067, PRO1068, ARG1027, ARG1051 and ILE1053). 3D interactions of designed compound 4 with VEGFR-2 receptor is depicted in Fig. 11.

Compound 5 having rerank score of -84.8435 is bound to the receptor through two conventional H-bonds, five Carbon-Hydrogen bond, one electrostatic, four alkyl and six pi-alkyl interactions. There are two conventional

Table 6 Docking Results and various interactions of the template, the designed compounds and Sunitinib

S/NO	Mol Dock Score	Rerank score	Amino acid types	Bond Type	Interactions	Bond distance
Template (11)	- 161.031	- 70.669	ARG1027	Hydrogen bond	Conventional H-bond	1.50848
			LEU1049	Hydrogen bond	Conventional H-bond	1.83371
			PHE845	Hydrogen bond	Conventional H-bond	1.56865
			LEU1067	Hydrogen bond	Carbon Hydrogen bond	3.09219
			PRO1068	Hydrogen bond	Carbon Hydrogen bond	2.00453
			MET1072	Hydrogen bond	Carbon Hydrogen bond	2.84089
			ARG1027	Hydrogen bond	Carbon Hydrogen bond	2.99178
			ASP1028	Electrostatic	Pi-Anion	4.64831
			PRO1068	Hydrophobic	Alkyl	4.51888
			MET1072	Hydrophobic	Alkyl	3.88605
			ALA844	Hydrophobic	Pi-Alkyl	4.4949
			LEU1049	Hydrophobic	Pi-Alkyl	4.61566
			ILE1053	Hydrophobic	Pi-Alkyl	4.06039
			ALA1065	Hydrophobic	Pi-Alkyl	5.05293
			ARG1027	Hydrophobic	Pi-Alkyl	5.34500
			PRO1068	Hydrophobic	Pi-Alkyl	5.09420
			N1	- 164.847	- 68.0186	SER1090
LEU1049	Hydrogen Bond	Conventional H-bond				1.8799
ASP1028	Electrostatic	Pi-Anion				4.48378
PRO1068	Hydrophobic	Alkyl				5.21075
MET1072	Hydrophobic	Alkyl				4.29418
ALA844	Hydrophobic	Pi-Alkyl				4.58803
LEU1049	Hydrophobic	Pi-Alkyl				4.6593
ILE1053	Hydrophobic	Pi-Alkyl				3.95547
ALA1065	Hydrophobic	Pi-Alkyl				4.96749
ARG1027	Hydrophobic	Pi-Alkyl				5.48065
PRO1068	Hydrophobic	Pi-Alkyl				4.77751
N2	- 144.005	- 97.0301	ASP1052	Hydrogen Bond	Conventional H-bond	1.70417
			SER877	Hydrogen Bond	Conventional H-bond	1.96384
			LEU802	Hydrogen Bond	Carbon Hydrogen Bond	2.92791
			ARG1027	Hydrogen Bond	Carbon Hydrogen Bond	2.86848
			ARG1027	Hydrogen Bond	Carbon Hydrogen Bond	2.53228
			SER877	Hydrogen Bond	Carbon Hydrogen Bond	2.31897
			H22	Hydrogen Bond	Carbon Hydrogen Bond	2.47958
			SER803	Hydrogen Bond	Carbon Hydrogen Bond	2.92212
			ASP1046	Hydrogen Bond	Carbon Hydrogen Bond	2.97715
			GLY1048	Hydrogen Bond	Carbon Hydrogen Bond	2.20309
			ARG1051	Hydrogen Bond	Carbon Hydrogen Bond	2.89556
			LEU802	Electrostatic	Pi-Cation	4.84151
			ARG880	Hydrophobic	Alkyl	4.72145
			PHE845	Hydrophobic	Pi-Alkyl	4.91766
			ALA844	Hydrophobic	Pi-Alkyl	3.58993
			ILE1053	Hydrophobic	Pi-Alkyl	5.10813
			LEU1049	Hydrophobic	Pi-Alkyl	5.22761
ILE1053	Hydrophobic	Pi-Alkyl	5.36263			
ALA881	Hydrophobic	Pi-Alkyl	3.2445			
LEU1049	Hydrophobic	Pi-Alkyl	4.87718			

Table 6 (continued)

S/NO	Mol Dock Score	Rerank score	Amino acid types	Bond Type	Interactions	Bond distance			
N3	- 173.302	- 86.3997	LEU1049	Hydrogen Bond	Conventional H-bond	1.95757			
			ARG1027	Hydrogen Bond	Conventional H-bond	1.45923			
			PRO1068	Hydrogen Bond	Carbon Hydrogen Bond	2.31537			
			ARG1027	Hydrogen Bond	Carbon Hydrogen Bond	2.6565			
			ARG1027	Hydrogen Bond	Carbon Hydrogen Bond	1.44017			
			ASP1028	Hydrogen Bond	Carbon Hydrogen Bond	2.83918			
			AASP1028	Electrostatic	Pi-Anion	4.56026			
			ALA1065	Hydrophobic	Alkyl	4.35837			
			PRO1068	Hydrophobic	Alkyl	4.29098			
			MET1072	Hydrophobic	Alkyl	3.84271			
			ILE1053	Hydrophobic	Alkyl	3.53044			
			LEU1067	Hydrophobic	Alkyl	4.49437			
			TYR1054	Hydrophobic	Pi-Alkyl	4.45661			
			ALA844	Hydrophobic	Pi-Alkyl	4.43002			
			LEU1049	Hydrophobic	Pi-Alkyl	4.61885			
			ILE1053	Hydrophobic	Pi-Alkyl	3.82494			
			ALA1065	Hydrophobic	Pi-Alkyl	5.27784			
			PRO1068	Hydrophobic	Pi-Alkyl	4.90503			
			N4	- 163.074	- 52.7342	SER1090	Hydrogen Bond	Conventional H-bond	2.57474
						ASP1028	Hydrogen Bond	Conventional H-bond	2.13103
GLY1048	Hydrogen Bond	Conventional H-bond				2.45948			
ARG1051	Hydrogen Bond	Conventional H-bond				1.84307			
ALA1065	Hydrogen Bond	Carbon Hydrogen Bond				2.21613			
SER1086	Hydrogen Bond	Carbon Hydrogen Bond				2.62078			
LEU1049	Hydrogen Bond	Carbon Hydrogen Bond				2.48759			
ARG1051	Hydrogen Bond	Carbon Hydrogen Bond				2.43736			
ARG1051	Hydrogen Bond	Carbon Hydrogen Bond				2.32368			
ASP1052	Hydrogen Bond	Carbon Hydrogen Bond				2.50523			
MET1072	Other	Pi-Sulfur				4.23317			
ALA844	Hydrophobic	Alkyl				3.30182			
PRO1068	Hydrophobic	Alkyl				4.04401			
ARG1027	Hydrophobic	Alkyl				4.37398			
MET1072	Hydrophobic	Alkyl				4.60028			
PHE845	Hydrophobic	Pi-Alkyl				5.14405			
TYR1054	Hydrophobic	Pi-Alkyl				5.25941			
TRP1071	Hydrophobic	Pi-Alkyl				5.04143			
TYR1082	Hydrophobic	Pi-Alkyl				3.65961			
ALA1065	Hydrophobic	Pi-Alkyl				4.63313			
LEU1067	Hydrophobic	Pi-Alkyl				4.88646			
PRO1068	Hydrophobic	Pi-Alkyl				4.15487			
ARG1027	Hydrophobic	Pi-Alkyl				5.08995			
ARG1051	Hydrophobic	Pi-Alkyl				4.99503			
ILE1053	Hydrophobic	Pi-Alkyl				4.63802			

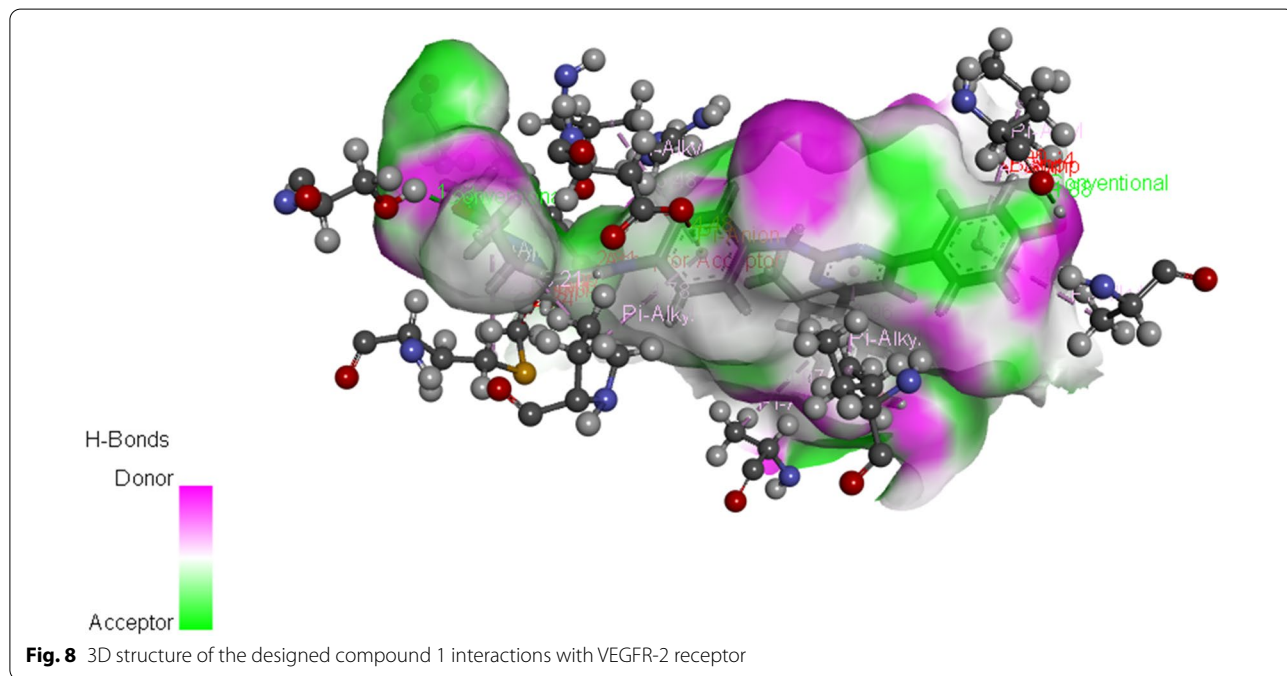
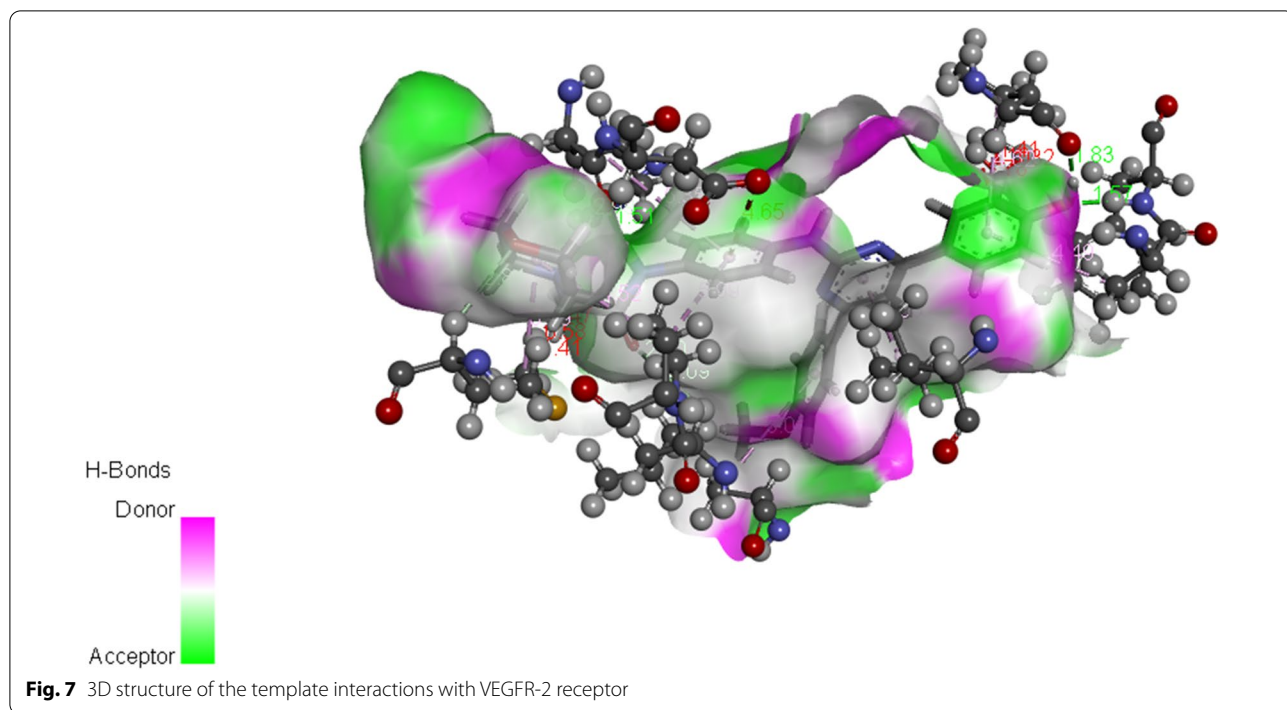
Table 6 (continued)

S/NO	Mol Dock Score	Rerank score	Amino acid types	Bond Type	Interactions	Bond distance
N5	- 173.853	- 84.8435	LEU1049	Hydrogen Bond	Conventional H-bond	2.03526
			ARG1027	Hydrogen Bond	Conventional H-bond	1.41506
			PRO1068	Hydrogen Bond	Carbon Hydrogen Bond	2.23911
			ARG1027	Hydrogen Bond	Carbon Hydrogen Bond	2.68609
			ARG1027	Hydrogen Bond	Carbon Hydrogen Bond	3.00698
			ARG1027	Hydrogen Bond	Carbon Hydrogen Bond	1.41551
			ASP1052	Hydrogen Bond	Carbon Hydrogen Bond	2.37466
			ASP1028	Electrostatic	Pi-Anion	4.53676
			PRO1068	Hydrophobic	Alkyl	4.32468
			MET1072	Hydrophobic	Alkyl	3.85875
			ILE1053	Hydrophobic	Alkyl	4.8813
			LEU1067	Hydrophobic	Alkyl	5.42047
			ALA844	Hydrophobic	Pi-Alkyl	4.4595
			LEU1049	Hydrophobic	Pi-Alkyl	4.56107
			ILE1053	Hydrophobic	Pi-Alkyl	3.74416
			ALA1065	Hydrophobic	Pi-Alkyl	5.00924
			ARG1027	Hydrophobic	Pi-Alkyl	5.49749
			PRO1068	Hydrophobic	Pi-Alkyl	4.97686
			Sunitinib	- 134.939	- 5.23604	SER803
GLY1048	Hydrogen Bond	Conventional H-Bond				1.65356
ARG1027	Hydrogen Bond	Carbon Hydrogen Bond				1.49939
ARG1027	Hydrogen Bond	Carbon Hydrogen Bond				2.94602
PRO1068	Hydrogen Bond	Carbon Hydrogen Bond				2.69622
ILE1053	Hydrophobic	Alkyl				4.26599
LEU1067	Hydrophobic	Alkyl				5.42047
PRO1068	Hydrophobic	Alkyl				5.325
PRO1068	Hydrophobic	Alkyl				4.62629
MET1072	Hydrophobic	Alkyl				3.70685
TYR1059	Hydrophobic	Pi-Alkyl				5.45143
TRP1071	Hydrophobic	Pi-Alkyl				5.31488
TYR1082	Hydrophobic	Pi-Alkyl				5.22115
ALA844	Hydrophobic	Pi-Alkyl				5.47748
ILE1053	Hydrophobic	Pi-Alkyl				4.24619
ARG1051	Hydrophobic	Pi-Alkyl				5.06549
ILE1053	Hydrophobic	Pi-Alkyl	4.27781			

Hydrogen bonds between the Nitrogen atom attached to the carbonyl group with ARG1027 and Hydrogen atom of the OH group with LEU1049. Two C-H bonds occur between ARG1027 and Hydrogen atoms of morpholine and alpha carbon, Carbonyl oxygen with PRO1068, Methoxy oxygen with ASP1052 one electrostatic interaction with ASP1023, four weak alkyl interactions with PRO1068, MET1072, ILE1053 and LEU1067. Lastly there

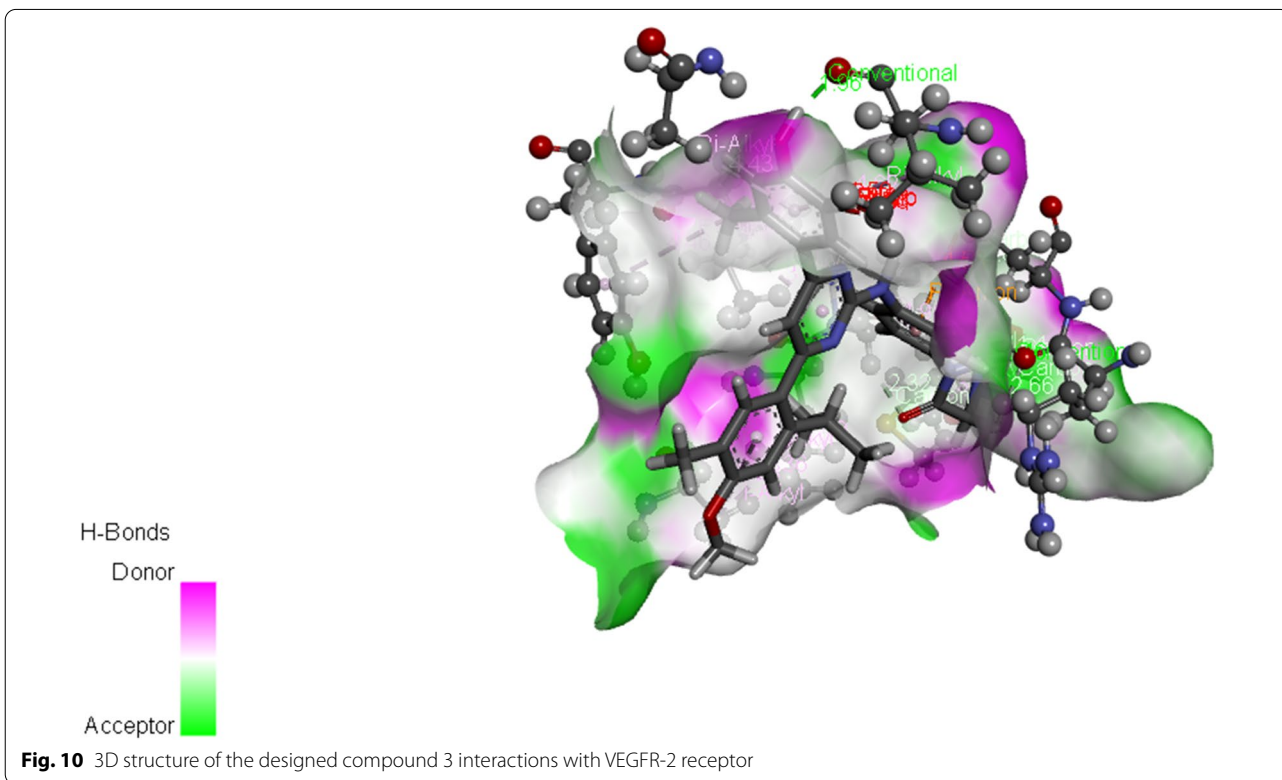
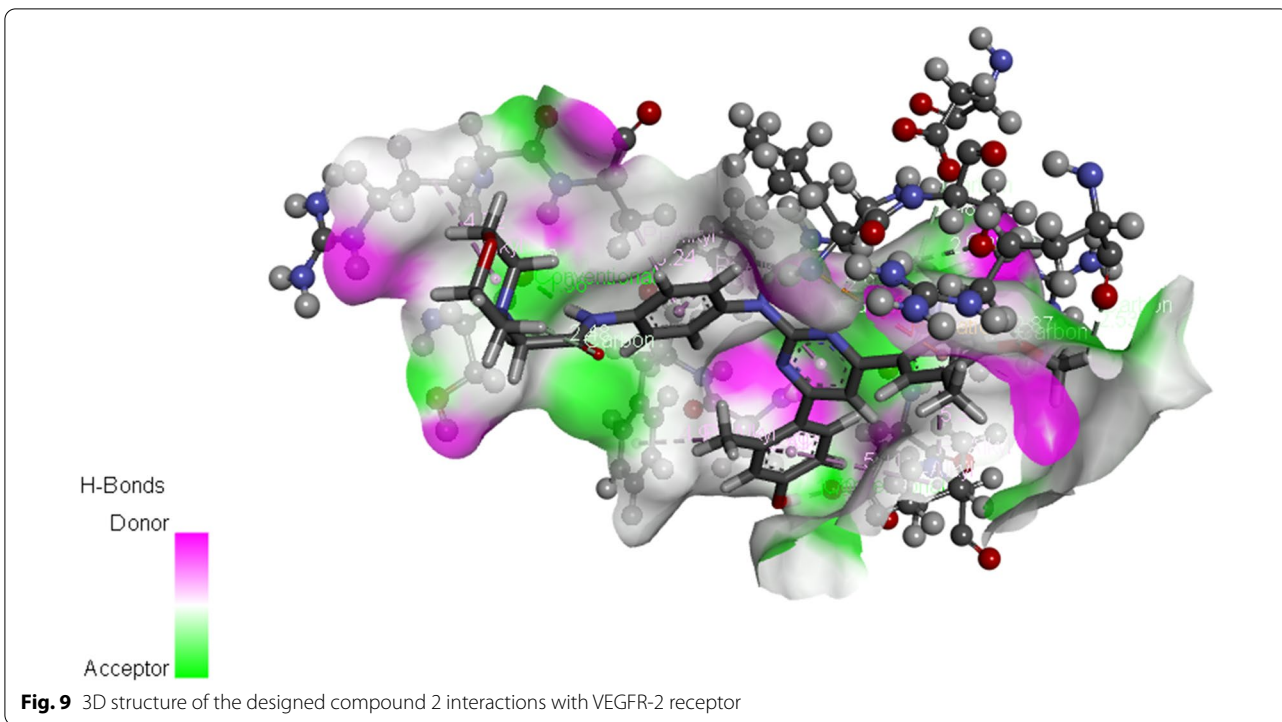
is weak pi-alkyl interactions with ALA844, LEU1049, ILE1053, ALA1065, ARG1027 and PRO 1068. The 3D interactions of designed compound 5 with VEGFR-2 receptor is shown in Fig. 12.

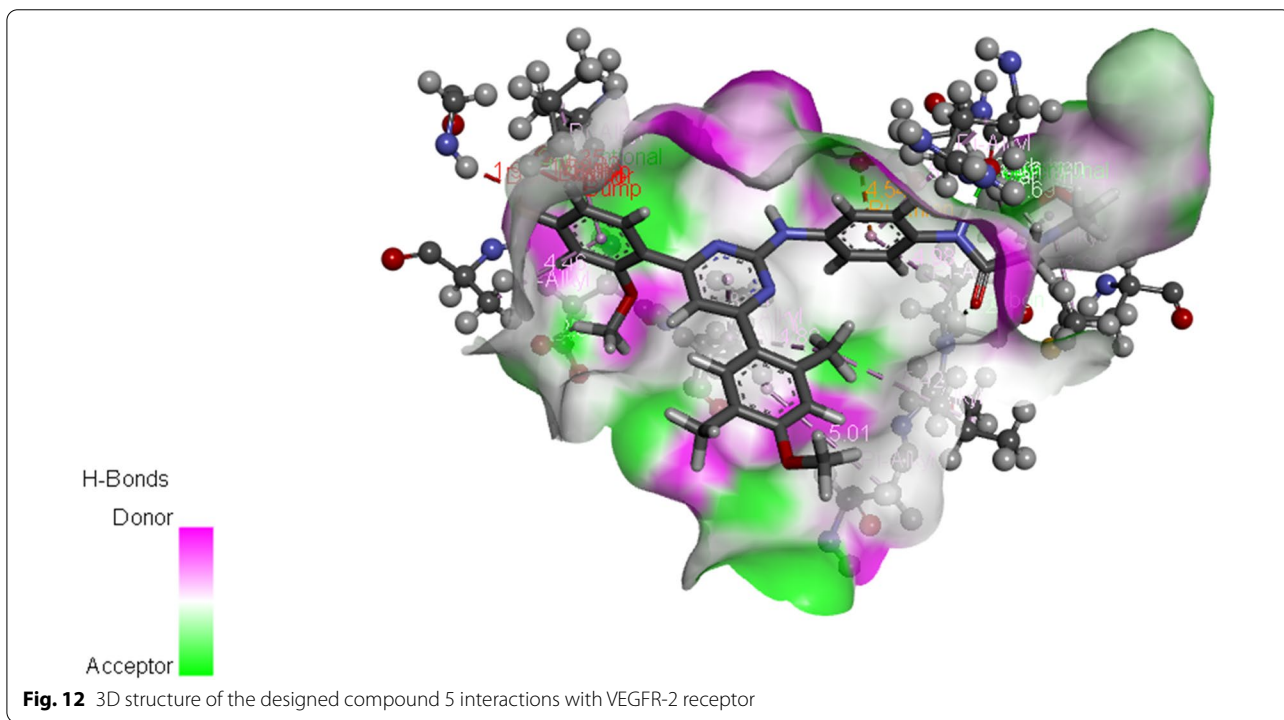
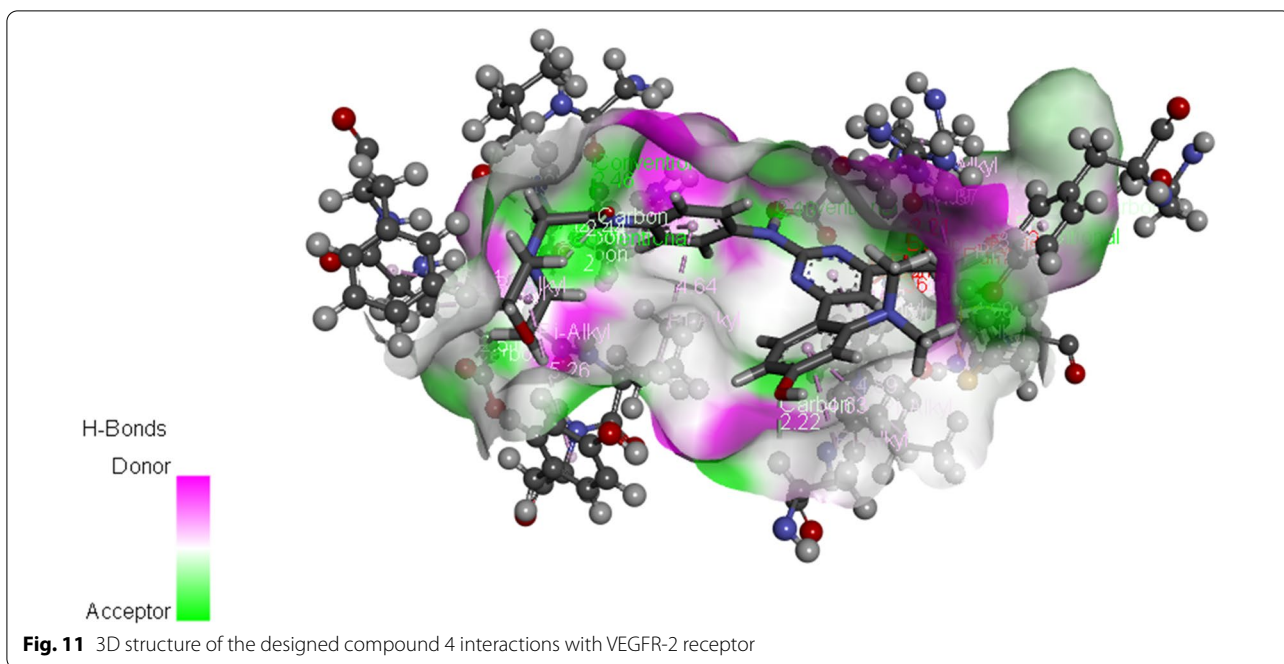
Sunitinib interactions with VEGFR-2 receptor is presented in Fig. 13. The reference drug which is also the co-crystallized ligand in the receptor (Sunitinib) was also re-docked to the active site of the 4AGD target receptor



to verify the preciseness of the docking procedure, and to test whether the designed inhibitors fit well in the active site of the target. It was found to have a Moldock and Rerank score of -134.939 and -5.23604 and interact with the following amino acid residues in the active site of the target receptor SER803, GLY1048, ARG1027,

PRO1068, ILE1053, LEU1067, PRO1068, MET1072, TYR1059, TRP1071, ALA844, ILE1053 and ARG1051 respectively (Fig. 13). Vascular endothelial growth factor-2 (VEGFR-2) receptor is a well-validated target for Breast cancer treatment, many researchers in different literatures had used VEGFR as target in breast cancer





therapy, to mention but few the researches reported by Liu et al. (2018), Luo et al. (2018) and Tahia et al. (2019) respectively. All of the amino acid residues are virtually common in all the designed compounds, the template and the reference drug (Sunitinib).

ADMET and drug-likeness properties of the designed compounds

Tables 7 and 8 represent the ADMET and Drug-likeness properties of the ligand based designed molecules. As recommended by Lipinski's rule of five, there is an increased

expectation that these molecules might be pharmacologically effective, on the account that all of them breach at most one (1) of the criteria. Additionally, the designed molecules exhibit absorbance value between 80.522 and 88.437% which passed the least approved values of 30% and consequently demonstrate promising human intestinal absorption. >0.3 to <-1 Log BB and >-2 to <-3 Log PS, are the minimum recommended values for the

blood–brain barrier (BBB) and central nervous system permeability respectively. The designed compounds have Log BB of >-1 and Log PS >-2 which is a clear indication that the compounds are uniformly dispersed to the brain, and assumed to permeate the central nervous system. Enzymatic metabolism is used to explain the biotransformation of a drug in the body hence, it became necessary to consider the drug's metabolism. Cytochrome

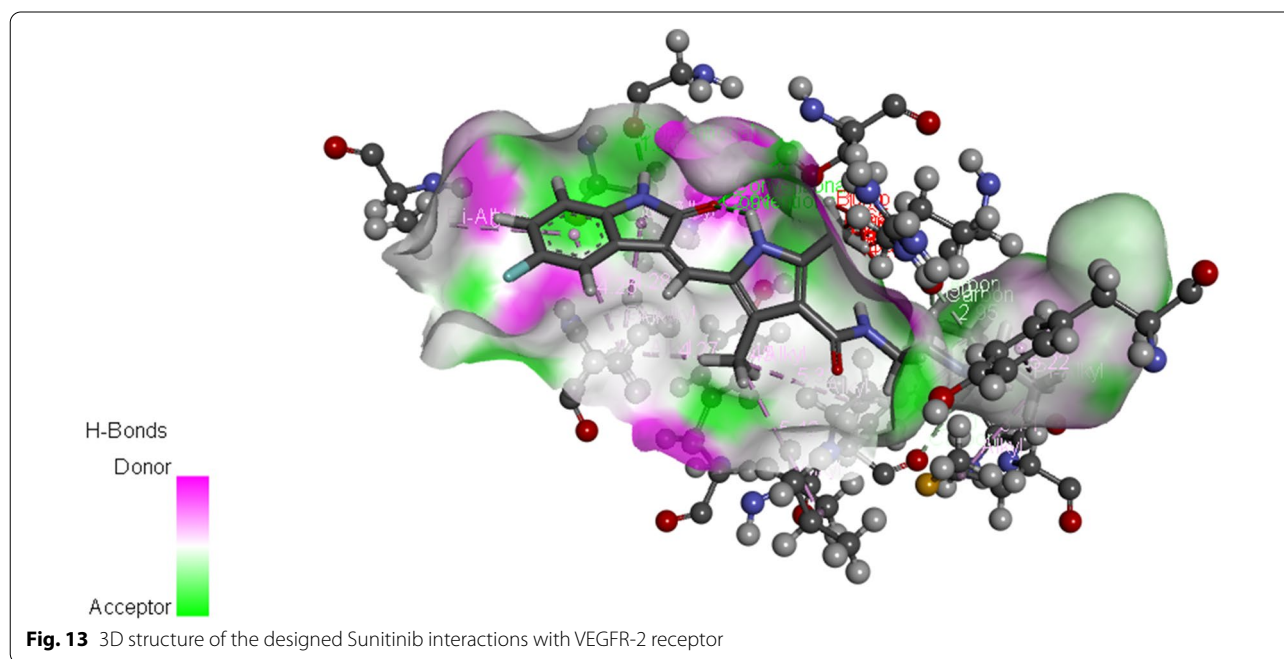


Table 7 Predicted ADMET properties of the designed compounds

S/NO	Absorption Intestinal Absorption	Distribution BBB permeability (Log BB)	CNS permeability (Log PS)	Metabolism substrate inhibitors						Excretion Tolerance clearance	Toxicity AMES toxicity	
				2D6	3A4	1A2	2C19	2C9	2D6			3A4
N1	83.841	-1.571	-3.111	NO	NO	NO	YES	YES	NO	YES	0.631	NO
N2	88.437	-1.532	-2.930	NO	NO	NO	YES	YES	NO	YES	0.730	NO
N3	83.763	-1.593	-3.112	NO	NO	YES	YES	YES	NO	YES	0.635	NO
N4	80.522	-1.665	-3.198	NO	NO	NO	YES	NO	NO	YES	0.695	NO
N5	80.966	-1.839	-3.172	NO	NO	NO	YES	NO	NO	YES	0.631	NO

Table 8 Predicted Drug-likeness properties of the designed compounds

S/NO	MW	HBD	HBA	WlogP	Lipinski violation	Bio availability score	Drug likeness
N1	582.69	3	7	4.65	1	0.55	YES
N2	581.70	3	7	5.71	1	0.55	YES
N3	582.69	3	7	4.65	1	0.55	YES
N4	569.65	3	8	4.59	1	0.55	YES
N5	634.52	3	8	5.05	1	0.55	YES

P450 is a class of super enzymes that plays a vital role in drug metabolism. 1A2, 2C9, 2C19, 2D6, and 3A4, are the CYP families responsible for drug metabolism. Results presented in Table 7, suggested that most of the designed compounds are the inhibitors of CYP2C19, 2C9, and 3A4 respectively. Total clearance which is an indicator that describes the relationship between the rate of elimination of the drug and its concentration in the body. The designed compounds demonstrate high values of total clearance which are within the acceptable limit of a drug molecule in the body. Furthermore, the five designed compounds are regarded to be nontoxic. The overall ADMET properties of these compounds reveal their good pharmacokinetic profiles.

Conclusions

Four models were generated out of which the first was selected as the best because of its fitness statistically with the following assessment parameters: $R^2_{\text{train}} = 0.832$, $R^2_{\text{adj}} = 0.79$, $R^2_{\text{ext}} = 0.62$, $Q^2 = 0.68$, and $\text{LOF} = 0.14509$ hence, satisfy the criteria of standard QSAR model. Additionally, new potentially active compounds on HUVEC cell line were designed by employing in-silico screening approach and their pIC_{50} was predicted with the aid of the generated QSAR model. The computed activity of most of the designed compounds were found to be greater than that of the lead reference molecule (11) employed in the design. Additionally, molecular docking simulation was conducted so as to gain insight in to the binding mode of the designed compounds against VEGFR-2 receptor. Designed compounds (N2, N3 and N5, with Rerank scores -97.0301 , -86.3997 and -84.8435) were found to have better docking scores than the template (Compound 11 Rerank score -70.669) and the reference drug (Sunitinib Rerank score -5.23604) respectively. High values of the docking scores of the designed compounds is attributed to the substituents that are introduced into the structure of the template during the design.

As recommended by Lipinski's rule of five, there is an increased expectation that these molecules are pharmacologically effective, on the account that all of them breach at most one (1) of the criteria. Consequently, the molecules are assumed to have good absorption, intense toxicity grade, are orally bioavailable and are absorptive. The developed model was therefore found to be efficient in predicting the pIC_{50} of Anti breast cancer agents that are yet to be synthesized and it also help in reducing the cost and synthetic duration the compounds. Finally, synthesis, in vivo, and in vitro evaluation of these ligands is recommended to be carried out so as affirm them as novel VEGFR-2 inhibitors for breast cancer treatment.

Abbreviations

QSAR: Quantitative structure activity relationship; HUVEC: Human umbilical vein endothelial cell; VEGFR-2: Vascular endothelial growth factor receptor-2; ER + : Estrogen receptor; DFT: Density functional theory; PADEL: Pharmaceutical Data Exploration Laboratory; B3LYP: Bee -3- Lee Yang Par; GFA-MLR: Genetic function algorithm-Multi linear regression; VIF: Variance inflation factor; ME: Mean effect; ADME: Absorption, Distribution, metabolism and excretion.

Acknowledgements

The author sincerely acknowledges all the contributors of this exploratory group for their guidance and motivation at some stages of this research work and Ahmadu Bello University for providing the softwares utilized in this analysis.

Authors' contributions

All authors participate accordingly, SHA performed all the computational analysis and wrote the manuscript, AU provided all the softwares for this research and edited the manuscript to ensure that error was minimized before final submission, MT analyzed the results obtained from the softwares, AB assessed the manuscript by using Plagiarism checker as well as arranged the manuscript according to the journal format. All authors have read and approved the manuscript.

Funding

Unrestrained allocation was not received by the authors for this research.

Availability of data and material

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable, because this article does not contain any studies with human or animal subjects.

Consent of publication

Not applicable.

Competing interests

The correspondents did not acknowledge competing of interest.

Received: 23 August 2021 Accepted: 27 September 2021

Published online: 09 October 2021

References

- Abdulfatai U, Uzairu A, Uba S (2017) Quantitative structure-activity relationship and molecular docking studies of a series of quinazolinonyl analogues as inhibitors of gamma amino butyric acid aminotransferase. *J Adv Res* 8:33–43
- Adedirin O, Uzairu A, Shallangwa GA, Abechi SE (2018) QSAR and molecular docking based design of some *n*-benzylacetamide as γ -aminobutyrate-aminotransferase inhibitors. *J Eng Exact Sci* 4(1):0065–0084
- Al-Suwaidan IA, Abdel-Aziz AA-M, Shawer TZ, Ayyad RR, Alanazi AM, El-Morsy AM, Mohamed MA, Abdel-Aziz NI, El-Sayed MA-A, El-Azab AS (2016) Synthesis, antitumor activity and molecular docking study of some novel 3-benzyl-4 (3H) quinazolinone analogues. *J Enzyme Inhib Med Chem* 31(1):78–89
- Arthur DE, Uzairu A, Mamza P, Abechi SE, Shallangwa G (2018) Insilico modeling of quantitative structure-activity relationship of pGI50 anticancer compounds on K-562 cell line. *Cogent Chem* 4:1432520
- Chandrappa S, Kavitha CV, Shahabuddin MS, Vinaya K, Ananda Kumar CS, Ranganatha SR (2009) Synthesis of 2-(5-(4-chlorophenyl) furan-2-yl) methylene)-4-oxo-2 thioxothiazolidin-3-yl) acetic acid derivatives and evaluation of their cytotoxicity and induction of apoptosis in human leukemia cells. *Bioorg Med Chem* 17:2576–2584
- Clarke R, Tyson JJ, Dixon JM (2015) Endocrine resistance in breast cancer—an overview and update. *Mol Cell Endocrinol* 418(3):220–234

- Daina AO, Michielin VZ (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep* 7:42717
- Golbraikh A, Tropsha A (2002) Beware of q²! *J Mol Graph Model* 20(4):269–276
- Hansch C, Leo A, Hoekman DH (1995) Exploring QSAR. Fundamentals and application in chemistry and biology. Am Chem Soc, Washington, DC
- Hicklin DJ, Ellis LM (2005) Role of the vascular endothelial growth factor pathway in tumor growth and angiogenesis. *J Clin Oncol* 23(5):1011–1027
- Huang D, Ding Y, Luo WM, Bender S, Qian CN, Kort E, Kristin Z (2008) Inhibition of MAPK pathways suppressed renal cell carcinoma growth and angiogenesis in vivo. *Can Res* 68(1):81–88
- Ibrahim MT, Uzairu A, Shallangwa GA, Uba S (2020) In-silico activity prediction and docking studies of some 2, 9-disubstituted 8 phenylthio/phenylsulfanyl-9h-purine derivatives as Anti-proliferative agents. *Heliyon* 6:e03158. <https://doi.org/10.1016/j.heliyon.2020.e03158>
- Ismail SY, Uzairu A, Sagagi B, Sabiu M (2018) In-silico molecular docking and pharmacokinetic study of selected phytochemicals with estrogen and progesterone receptors as anticancer agent for breast cancer. *JOTCSA* 5(3):1337–1350
- Jaworska J, Nikolova-Jeliazkova N, Aldenberg T (2005) QSAR applicability domain estimation by projection of the training set descriptor space: a review. *Atla-Nottingham* 33:445
- Kennard RW, Stone LA (1969) Computer aided design of experiments. *Technometrics* 11(1):137–148
- Khaled KF (2011) Modeling corrosion inhibition of iron in acid medium by genetic function approximation method: a QSAR model. *Corros Sci* 53(11):3457–3465
- Liu L, Tang Z, Wu C, Li X, Huang A, Lu X, You Q, Xiang H (2018) Synthesis and biological evaluation of 4, 6-diaryl-2-pyrimidinamine derivatives as anti-breast cancer agents. *Bioorg Med Chem Lett* 28:1138–1142
- Luo G, Tang Z, Lao K, Li X, You Q, Xiang H (2018) Structure-activity relationships of 2, 4-disubstituted pyrimidines as dual ER α /VEGFR-2 ligands with anti-breast cancer activity. *Eur J Med Chem*. <https://doi.org/10.1016/j.ejmech.2018.03.018>
- Myers RH (1990) Classical and modern regression application, 2nd edn. Duxbury Press, CA
- Patel RR, Sengupta S, Kim HR, Klein-Szanto AJ, Pyle JR, Zhu F (2010) Experimental treatment of oestrogen receptor (ER) positive breast cancer with tamoxifen and brivanib alalinate, a VEGFR-2/fgfr-1 kinase inhibitor. *Eur J Cancer* 46(9):1537–1553
- Rajer-Kanduč K, Zupan J, Majcen N (2003) Separation of data on the training and test set for modelling: a case study for modelling of five colour properties of a white pigment. *Chemom Intell Lab Syst* 65(2):221–229
- Roy K, Kar S, Ambure P (2015) On a simple approach for determining applicability domain of QSAR models. *Chemom Intell Lab Syst* 145:22–29. <https://doi.org/10.1016/j.chemolab.2015.04.013>
- Solomon VR, Hua C, Lee H (2009) Hybrid pharmacophore design and synthesis of isatin–benzothiazole analogs for their anti-breast cancer activity. *Bioorg Med Chem* 17:7585–7592
- Tahia KM, Rasha ZB, Samia AE, Al MM, Abeer EM (2019) Synthesis, anticancer effect and molecular modeling of new thiazolylpyrazolyl coumarin derivatives targeting VEGFR-2 kinase and inducing cell cycle arrest and apoptosis. *Bioorg Chem* 85:253–273
- Troyer JR (2001) The multiple discoveries of the first hormone herbicides. *Weed Sci* 49:290–297
- Umar BA, Uzairu A, Shallangwa GA, Sani U (2019) 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 Inform Bioinform* 8:22
- van Leeuwen FE, Benraad J, Coebergh JW, Kiemeneij LA, Gimbrère CH, Otter R, Schouten LJ, Damhuis RA, Bontenbal M, Diepenhorst FW (1994) Risk of endometrial cancer after tamoxifen treatment of breast cancer. *Lancet* 343:448
- Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK (2011) Validation of QSAR models-strategies and importance. *Int J Drug Des Discov* 2:511–519
- Wang T, You Q, Huang FS, Xiang H (2009) Recent advances in selective estrogen receptor modulators for breast cancer. *Mini-Rev Med Chem* 9(10):1191–1201

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)