2D-QSAR Modeling and Molecular Docking Studies on 1H-Pyrazole-1-carbothioamide Derivatives as EGFR Kinase Inhibitors

Tawassl T. H. Hajalsiddig,* Abu Baker M. Osman, and Ahmed E. M. Saeed

ABSTRACT: Epidermal growth factor receptor (EGFR) kinase has been commonly associated with cancers such as lung, ovarian, hormone-refractory prostate, metastatic colorectal, glioblastoma, pancreatic, and breast cancers. A series of 1H-pyrazole-1-carbothioamide derivatives and their EGFR inhibitory activities were subjected to two-dimensional (2D) quantitative structure–activity relationship (2D-QSAR) studies. The 2D-QSAR models were constructed based on a forward selection of partial least-squares (PLS) and stepwise multiple linear regression (SW-MLR) methods validated by leave-one-out (LOO) and external test set prediction approaches. The stepwise multiple linear regression (SW-MLR) method presented an encouraging result as compared to other methods. The results of the study indicated that the activity of 1H-pyrazole-1-carbothioamide derivatives as an EGFR kinase inhibitor was more influenced by adjacency distance matrix descriptors. The models were improved after outlier removal through the applicability domain. Based on the resultant models, 11 new compounds with high potency were designed as EGFR kinase inhibitors. Molecular docking studies were performed for designing compounds, and they were compared with erlotinib as a reference to predict their interactions in the active site and identify structural features necessary for producing biological activities.

INTRODUCTION

The ERbB or epidermal growth factor (EGF) family of receptor tyrosine kinases (RTKs) is composed of four structurally related members: EGFR/ErbB1/HER, ErbB2/Neu/HER-2, ErbB3/HER3, and ErbB4/HER4.1 EGFR receptor family members are essential in the etiology of numerous tumors, including those of the breast, ovary, lung, colon, nervous system, head and neck, prostate, and pancreas.2 The EGFR is more commonly studied because it plays an important role in many vital processes such as cell proliferation, DNA damage and repair, and DNA replication and transcriptional regulation.

There are currently two types of drugs approved by the Food and Drug Administration (FDA) targeting this family: monoclonal antibodies (such as cetuximab and pertuzumab) and small-molecule inhibitors based on a central 4-aminoarylquinazoline core (gefitinib, erlotinib, lapatinib, and afatinib).3–6

Pyrazoles are heterocyclic compounds and most widely used in the design and synthesis of novel biologically active agents. Pyrazole derivatives display a broad spectrum of biological activities such as antitumor,7 antibacterial,9,10 antifungal,11 antiviral,12 antibacterial,13,14 antimalarial,15 antioxidant,16 anti-inflammatory,17 and analgesic properties18 as well as antihyperglycemic activity.19

Pyrazoles are one of many significant compounds that have been studied and reported for their antitumor activity in vitro and in vivo against a broad range of cancers. New derivatives of benzimidazole-linked pyrazole synthesized and evaluated for their activities in vitro for five human cancer cell lines involving MCF-7, HaCaT, MDA-MB23, HepG2, and A549 have been reported as EGFR receptor inhibitors by Akhtar et al.20 Furthermore, the activity of 1H-pyrazole-1-carboxamide analogues as an EGFR/HER-2 tyrosine kinase inhibitor was evaluated and reported by Tao et al.21 Tumma et al. reported that 1H-pyrazole-1-carbothioamide derivatives showed antiproliferative activity against MCF-7 cell lines as anticancer agents.22 On the other hand, many computational analyses for pyrazole derivatives have been reported, including quantitative structure–relationship activity (QSAR) and molecular docking studies. Sunayana et al. studied the two-dimensional (2D) and three-dimensional (3D) group-based quantitative structure–activity relationship (G-QSAR) for evaluating the activity of a set of thiazolyl-pyrazole derivatives as EGFR inhibitors using the multiple regression method and k-nearest neighbor molecular...
field analysis (KNN-MFA) and then performed a molecular docking study for these derivatives with EGFR kinase as a receptor.23

Lv et al. reported the synthesis of two series of pyrazole derivatives 1H-pyrazole-1-carbothioamide and thiazolyl-pyrazoline and evaluated their activity as EGFR kinase inhibitors.24,25 The importance of 1H-pyrazole-1-carbothioamide derivatives lies in their ability to inhibit the kinase activity of EGFR and potential anticancer activities after binding.24

In this present study, a set of 30 compounds derived from pyrazole containing thiourea skeleton, evaluated as (EGFR) kinase inhibitors, was used to describe 2D-QSAR models developed using a forward selection of partial least-squares (PLS) regression and stepwise multiple linear regression (SW-MLR) methods. New compounds were designed, and their biological activity as EGFR kinase inhibitors was investigated by predicting the 2D-QSAR model. After that, molecular docking of the designed compounds was carried out to investigate their binding pattern with the EGFR kinase protein target.

### RESULTS AND DISCUSSION

#### QSAR Results.

2D-QSAR models were generated in an attempt to determine the effect of structural features of pyrazole derivatives containing thiourea as EGFR kinase inhibitors. The chemical structures of the studied compounds with their EGFR kinase inhibitory activity are shown in Table 1. The correlation matrix for nine descriptors and their values used to generate models for eqs 1–4 are given in Tables 2 and 3, respectively. The methods of partial least-squares (PLS) and multiple linear regression (MLR) were used to perform the 2D-QSAR models by establishing the statistical linear correlation for nine descriptors taken as independent variables and pIC50 as the dependent variable of the training data (24 compounds).

In the partial least-squares (PLS) method, the forward selection was used with descriptors as variables and added one by one in order to obtain the single best improvement of the model. The best model was selected on the basis of statistical parameters via the observed squared correlation coefficient ($r^2$) which is a relative measure of the quality of fit. The cross-validated squared correlation coefficient ($q^2$) should be high as a good indicator for predicting the power of the QSAR model, and

| compounds | R1 | R2 | EGFR (IC50 μM) | compounds | R1 | R2 | EGFR (IC50 μM) |
|-----------|----|----|---------------|-----------|----|----|---------------|
| C1        | 3,4-CH₃ | 4-F | 0.83          | C16       | 3,4-Cl | 4-OH | 5.27          |
| C2        | 3,4-CH₃ | 4-Cl | 1.36          | C17       | 3,4-Cl | 4-NO₂ | 7.78          |
| C3        | 3,4-CH₃ | 4-Br | 2.16          | C18       | 3,4-Cl | 2-F | 6.56          |
| C4        | 3,4-CH₃ | 4-CH₃ | 0.34          | C19       | 3,4-Cl | 2-Cl | 7.28          |
| C5        | 3,4-CH₃ | 4-OCH₃ | 0.07          | C20       | 3,4-Cl | 2-Br | 6.43          |
| C6        | 3,4-CH₃ | 4-OH | 0.13          | C21       | 3,4-Br | 4-F | 8.92          |
| C7        | 3,4-CH₃ | 4-NO₂ | 3.06          | C22       | 3,4-Br | 4-Cl | 8.15          |
| C8        | 3,4-CH₃ | 2-F | 5.19          | C23       | 3,4-Br | 4-Br | 10.06         |
| C9        | 3,4-CH₃ | 2-Cl | 3.87          | C24       | 3,4-Br | 4-CH₃ | 9.83          |
| C10       | 3,4-Cl | 2-F | 4.21          | C25       | 3,4-Br | 4-CH₃ | 8.09          |
| C11       | 3,4-Cl | 4-F | 6.27          | C26       | 3,4-Br | 4-OH | 11.27         |
| C12       | 3,4-Cl | 4-Cl | 7.32          | C27       | 3,4-Br | 4-CH₃ | 10.64         |
| C13       | 3,4-Cl | 4-Br | 6.85          | C28       | 3,4-Br | 2-F | 13.37         |
| C14       | 3,4-Cl | 4-CH₃ | 7.38          | C29       | 3,4-Br | 2-Cl | 12.26         |
| C15       | 3,4-Cl | 4-OCH₃ | 5.74          | C30       | 3,4-Br | 2-Br | 10.97         |

#### Table 1. Chemical Structures and EGFR Inhibitory Activity of 3,5-Diphenyl-4,5-dihydro-1H-pyrazole-1-carbothioamide Derivatives C1–C30

#### Table 2. Correlation Matrix for Nine Descriptors Used in Eqs 1–4

| pIC50 (%) | BCUT_PEOE_2 (%) | density (%) | log P(o/w) (%) | BCUT_SMR_1 (%) | diameter (%) | a_cc (%) | PEOE_VSA+1 (%) | a_IC (%) | petitjean (%) |
|----------|-----------------|-------------|---------------|----------------|--------------|----------|---------------|----------|---------------|
| 100      | -93             | -71         | -63           | 61             | 37           | 31       | 19            | -3       | -3            |
| BCUT_PEOE_2 | 100          | 70          | 85            | 100            | 100          | 100      | 100           | 100      | 100           |
| density | -71             | 100         | 100           | 100            | 100          | 100      | 100           | 100      | 100           |
| log P(o/w) | -63            | 100         | 100           | 100            | 100          | 100      | 100           | 100      | 100           |
| BCUT_SMR_1 | 61             | -53         | -31           | -17           | -19          | 43       | 31            | 5        | 5             |
| diameter | 37              | -20         | -40           | -19           | -19          | 41       | 72            | 5        | 5             |
| a_cc     | 31              | -15         | -43           | -43           | -43          | 84       | 41            | 43       | 43            |
| PEOE_VSA+1 | 27             | -6          | -31           | -37           | -37          | 78       | 43            | 100      | 100           |
| a_IC     | 19              | -11         | -51           | -19           | -19          | 72       | 41            | 100      | 100           |
| petitjean | -3              | 17          | 6             | 10             | 10           | 3        | 31            | 3        | -8            |
the difference between this and \( r^2 \) should not be more than 0.3. The standard error of estimate (SEE < 0.3) represents an absolute measure of prediction accuracy. Fischer’s value (\( F \)), or the Fisher ratio, reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the \( F \)-test indicate that the model is statistically significant.  

The value for the null hypothesis, which should not exceed more than 0.05 (\( p < 0.05 \)).  

The linear correlation between experimental biological activities (pIC\(_{50}\)) for 1H-pyrazole-1-carboxamide derivatives as a dependent variable and 2D descriptors (lipophilic, BCUT_PEOE_2, \( a_{\text{acc}} \), and \( a_{\text{IC}} \)) as independent variables expressed in the 2D-QSAR model is illustrated below

\[
pIC_{50} = 11.44972 - 10.28409 \times \text{BCUT_PEOE}_2 + 0.21308 \times \log P_{(o/w)}
\]

\[
a_{\text{acc}} = 0.00322 \times a_{\text{IC}} - 0.00922 \times \log P_{(o/w)}
\]

\[
F_{\text{training set}} = 23, \quad F_{\text{test set}} = 6, \quad r^2 = 0.904, \quad \text{RMSE} = 0.152, \quad q^2 = 0.884, \quad F = 43.05 (p < 0.0001), \quad \text{SEE} = 0.345, \quad \text{and} \quad r^2_{\text{pred}} = 0.879
\]

(1)

And the other 2D-QSAR model for BCUT_PEOE_2, \( a_{\text{acc}} \), density, and \( a_{\text{IC}} \) descriptors is described below

\[
pIC_{50} = 11.36357 - 9.75704 \times \text{BCUT_PEOE}_2 + 0.21094 \times \log P_{(o/w)} + 0.00302 \times a_{\text{IC}} - 0.32127
\]

\[
F_{\text{training set}} = 23, \quad F_{\text{test set}} = 6, \quad r^2 = 0.904, \quad \text{RMSE} = 0.152, \quad q^2 = 0.884, \quad F = 43.05 (p < 0.0001), \quad \text{SEE} = 0.345, \quad \text{and} \quad r^2_{\text{pred}} = 0.879
\]

(2)

In the above QSAR equations, \( n_{\text{training set}} \) and \( n_{\text{test set}} \) are the numbers of compounds of the training and test sets used to derive the QSAR model, respectively; \( r \) is the correlation coefficient of regression for the training set; \( r^2 \) is the square of the correlation coefficient of regression for the training set; \( q^2 \) is the square of LOO, the cross-validated correlation coefficient; \( r^2_{\text{pred}} \) is the predicted correlation coefficient for the external test set; RMSE is the root-mean-square error; \( F \) is the Fisher ratio; \( p \) is the statistical confidence level; and \( \text{SEE} \) is the standard error of the estimate. The experimental pIC\(_{50}\)(Exp.) values of the data set and their new predicted pIC\(_{50}(\text{Pred.}) \) values by eqs 1 and 2 with residues are listed in Table 4.

From the standardized residues of training set compounds shown in Figure 5ab, it is observed that compound C6 has higher standardized residual values, more than –2 and +2, as a cutoff value for accepting predictions; therefore, this is an outlier based on the commonly accepted prediction according to the suggestion of Jalali-Heravi and Kyani.  

Thus, compounds C6 was not considered during the course of obtaining data analysis for eqs 1 and 2. The remaining 23 compounds of the training set...
Table 4. Experimental pIC_{50(Exp.)}, Predicted pIC_{50(Prod.)}, and Residual Values for Eqs 1 and 2a

| compounds | pIC_{50(Exp.)} | pIC_{50(Prod.)} | residual | pIC_{50(Prod.)} | residual |
|-----------|----------------|----------------|----------|----------------|----------|
| C1        | 6.08           | 5.85           | 0.23     | 5.87           | 0.21     |
| C2        | 5.87           | 5.78           | 0.09     | 5.79           | 0.08     |
| C3        | 5.66           | 5.75           | -0.09    | 5.75           | -0.09    |
| C4T       | 6.47           | 6.32           | 0.14     | 6.33           | 0.14     |
| C5        | 7.15           | 6.99           | 0.16     | 6.96           | 0.19     |
| C6outlier | 6.89           | 5.48           | 1.41     | 5.52           | 1.37     |
| C7        | 5.51           | 5.51           | 0.00     | 5.53           | -0.02    |
| C8        | 5.28           | 5.68           | -0.40    | 5.70           | -0.42    |
| C9T       | 5.41           | 5.67           | -0.26    | 5.69           | -0.28    |
| C10       | 5.37           | 5.67           | -0.30    | 5.67           | -0.30    |
| C11       | 5.20           | 5.03           | 0.17     | 5.05           | 0.15     |
| C12       | 5.14           | 5.01           | 0.13     | 5.03           | 0.11     |
| C13T      | 5.16           | 5.02           | 0.14     | 5.02           | 0.14     |
| C14       | 5.13           | 5.04           | 0.09     | 5.07           | 0.06     |
| C15       | 5.24           | 5.36           | -0.12    | 5.38           | -0.14    |
| C16       | 5.27           | 5.12           | 0.12     | 5.15           | 0.09     |
| C17T      | 5.11           | 4.97           | 0.14     | 4.98           | 0.13     |
| C18       | 5.18           | 5.12           | 0.06     | 5.13           | 0.05     |
| C19       | 5.13           | 5.05           | 0.08     | 5.07           | 0.06     |
| C20       | 5.19           | 5.04           | 0.15     | 5.04           | 0.15     |
| C21       | 5.05           | 5.01           | 0.03     | 4.99           | 0.05     |
| C22       | 5.08           | 5.01           | 0.07     | 4.98           | 0.10     |
| C23       | 4.99           | 4.99           | 0.00     | 4.95           | 0.04     |
| C24       | 5.00           | 5.02           | -0.02    | 5.02           | -0.02    |
| C25T      | 5.09           | 5.34           | -0.25    | 5.32           | -0.23    |
| C26       | 4.95           | 5.11           | -0.16    | 5.09           | -0.14    |
| C27       | 4.97           | 4.95           | 0.02     | 4.93           | 0.04     |
| C28       | 4.87           | 5.11           | -0.24    | 5.08           | -0.21    |
| C29T      | 4.91           | 5.05           | -0.14    | 5.03           | -0.12    |
| C30       | 4.96           | 5.03           | -0.08    | 4.98           | -0.03    |

aT = Test set.

showed more satisfactory QSAR models for predicting the biological activity (pIC_{50}).

Equations 1 and 2 present four descriptors for each equation correlated with the biological activity (pIC_{50}) to indicate that these two predictive equations have good statistical consistency for the training group with high r^2 values (eq 1, r^2 = 0.901; eq 2, r^2 = 0.904). To predict the ability of models as statistically significant by the cross-validated q^2 (eq 1, q^2 = 0.784; eq 2, q^2 = 0.884), the difference between r^2 and q^2 should not be more than 0.3. The value of q^2 is indeed greater than 0.5, which is the essential condition to qualify a QSAR model as valid. To predict the success of both models, each model was applied to the external test set compounds (six compounds); the square of the correlation coefficient r^2_{pred} has been obtained as 0.874 and 0.889 for eqs 1 and 2, respectively. These models showed a good fit with values of RMSE less than 0.3 (eq 1, RMSE = 0.154; eq 2, RMSE = 0.152), low values of SEE, and p-values less than 0.0001. The large calculated F values from the created QSAR models are desirable for meaningful regression.

Adjacency matrix descriptors, originally developed by Burden, are in principle based on producing a molecular identification number out of the lowest eigenvalues of a connectivity matrix. After all hydrogens were deleted and the remaining heavy atoms were numbered, the symmetric matrix was established.29 Pearlman and Smith improved the concept of BCUT descriptors and enlarged it to provide an internally consistent, balanced set of molecular descriptors calculated from the eigenvalues of a modified adjacency matrix.30

The first term in eqs 1 and 2 is BCUTPEOE_2 (a third BCUT descriptor using POE partial charges). POE is the method of partial equalization of orbital electronegativities for calculating atomic partial charges in which charge is transferred between bonded atoms until equilibrium. This descriptor has a very high correlation coefficient (−93%) with pIC_{50} and has dominating influence in both equations with a higher negative descriptor contribution (−10.10430 and −10.679). The BCUTPEOE_2 descriptor is the most outstanding value of the negative contribution with pIC_{50} indicating a strong inverse relationship between them as EGFR kinase inhibitors.

The second term in the above two equations is the a_acc (the number of hydrogen-bond acceptor atoms) descriptor. It is an effective descriptor for the pIC_{50} value of each model with a lower coefficient (31%) and showing a positive contribution (0.21308 and 0.21094). The a_acc descriptor describes polarity for enabling better permeation and absorption, so every increase in the a_acc descriptor value will cause an increase in the pIC_{50} value.

The third descriptor is a_IC (atom information content (total)) is calculated as the entropy of the element distribution in the molecule (ICM) multiplied by n, where n is the sum of the number of occurrences of an atomic number in the molecule) with just a small correlation coefficient (19%) and showing a positive contribution (0.00322 and 0.00302) for each model, meaning that for every change in the a_IC descriptor, the pIC_{50} value will increase.

The fourth term in eq 1 is the log P_{(o/w)} (log of the octanol/water partition coefficient) descriptor with a correlation coefficient of −63% and negative contribution (−0.00922). In general, a decrease in the log P_{(o/w)} value increases the solubility, membrane permeability, and bioavailability of compounds, meaning the increased biological activity of compounds.

However, the fourth term in eq 2 is the density (the ratio of the weight to the van der Waals volume). The weight means the amount of a substance, and the van der Waals volume of a molecule describes the space occupied by the atoms. This descriptor is related to the steric bulk and size of the molecule. The density in eq 2 has a high correlation coefficient (−71%) and negative contribution (−0.32127). Therefore, a decrease in the relation between the steric bulk and size leads to an increase in the activity of the compound.

It can be observed that eq 2 is statistically better than eq 1 because r^2, q^2, and F values increase and the RMSE value becomes much less (<0.3). However, the 2D-QSAR model expressed by eq 2 is more acceptable compared to the one by eq 1.

The plots of the experimental pIC_{50} values versus their predictions of the training set and test set based on the PLS model (eqs 1 and 2) are shown in Figures 1 and 2.

The stepwise multiple linear regression (stepwise-MLR) method was also performed on the same training set chosen for use in the PLS model to select the significant descriptors from 25 descriptors. The good regression model performed by the stepwise-MLR method for biological activity pIC_{50} as a dependent variable with three adjacency and distance matrix descriptors as independent variables is explained below in eq 3
The first term in MLS models that has dominating influence on both eqs 3 and 4 is BCUT_PEOE_2 (a third BCUT descriptor using PEOE partial charges), as in previous PLS models in eqs 1 and 2. This descriptor here has a great effect as well as a high correlation coefficient (−93%) with pIC50 and dominating influence on both eqs 3 and 4 with a higher negative contribution (−12.39 and −10.80). The BCUT_PEOE_2 descriptor has the most prominent negative contribution value that affected the value of pIC50, meaning that there is a strong inverse relationship between them as described in the PLS models above.

BCUT_SMR_1 is the second term in eq 3 with a negative coefficient (−2.56604) similar to the BCUT_PEOE_2 descriptor and has a moderate correlation influence (61%) with pIC50. BCUT_SMR_1 is a BCUT descriptor using the atomic contribution to molar refractivity instead of partial charge.

The diameter and petitjean are descriptors for the distance matrix for graphs that were described by Petitjean; these are calculated as the largest value in the distance matrix and the difference between the diameter and radius divided by the diameter, respectively.32

The diameter (the largest value in the distance matrix) is the third term in eq 3 with a little influence on the correlation coefficient (25%) with pIC50 and a low positive influence on the model (0.14).

Petitjean (the value of the difference between the diameter and radius divided by the diameter using the distance matrix) is the second term in eq 4 with just a small correlation coefficient (17%) with pIC50 and a positive contribution (3.95).

It is noted that both BCUT_PEOE_2 and BCUT_SMR_1 (BCUT descriptors) are inversely related to pIC50, while the diameter and petitjean (distance matrix descriptors) have a direct relationship.

PEOE_VSA+1 is the third term in eq 3 with a low positive contribution influence on the model (0.015) and just a small correlation coefficient (17%). The PEOE_VSA+1 descriptor is defined as a total positive 1 of the van der Waals surface areas using PEOE partial charges in range (0.05, 0.1). This descriptor has shown a direct relationship with the value of pIC50. The experimental pIC50 values of the data set and their new predicted pIC50 values by eqs 3 and 4 with residuals are listed in Table 5.

The plots of experimental pIC50 values versus their predicted pIC50 values by eqs 3 and 4 with residuals are shown in Figures 3 and 4, respectively.

Among these four statistical models, stepwise multiple linear regression (stepwise-MLR) has shown an encouraging result as compared to other partial least-squares (PLS) regression methods.

In general, from this 2D-QSAR study, it was observed that the most active compounds C1, C4, and C5 have very clear smaller values for BCUT, density, and log P(0/0) descriptors, as shown in Table 2. In other words, there is a strong inverse relationship between these descriptors and biological activity values, especially BCUT_PEOE_2 that has a strong inverse relationship.
Table 5. Experimental pIC_{50}(Exp.), Predicted pIC_{50}(Pred.), and Residual Values for eqs 3 and 4

| compounds | pIC_{50}(Exp.) | pIC_{50}(Pred.) | residual | pIC_{50}(Pred.) | residual |
|-----------|----------------|----------------|----------|----------------|----------|
| C1        | 6.08           | 5.97           | 0.11     | 5.77           | 0.31     |
| C2        | 5.87           | 5.70           | 0.17     | 5.87           | 0.00     |
| C3        | 5.66           | 5.62           | 0.04     | 5.77           | -0.11    |
| C4^T      | 6.47           | 6.30           | 0.16     | 7.52           | -1.05    |
| C5        | 7.15           | 7.10           | 0.05     | 7.02           | 0.13     |
| C6^Outlier| 6.89           | 5.30           | 1.59     | 6.05           | 0.84     |
| C7        | 5.51           | 5.55           | -0.04    | 5.59           | -0.08    |
| C8        | 5.28           | 5.62           | -0.34    | 5.77           | -0.49    |
| C9^T      | 5.41           | 5.56           | -0.15    | 5.87           | -0.46    |
| C10       | 5.37           | 5.52           | -0.15    | 5.77           | -0.40    |
| C11       | 5.20           | 5.14           | 0.06     | 5.23           | -0.03    |
| C12       | 5.14           | 5.09           | 0.05     | 5.14           | 0.00     |
| C13^T     | 5.16           | 5.09           | 0.07     | 5.23           | -0.07    |
| C14       | 5.13           | 5.11           | 0.02     | 5.18           | -0.05    |
| C15       | 5.24           | 5.26           | -0.02    | 5.12           | 0.12     |
| C16       | 5.27           | 5.09           | 0.15     | 5.17           | 0.07     |
| C17^T     | 5.11           | 5.23           | -0.12    | 5.19           | -0.08    |
| C18       | 5.18           | 5.11           | 0.07     | 5.23           | -0.05    |
| C19       | 5.13           | 5.01           | 0.12     | 5.14           | 0.01     |
| C20       | 5.19           | 4.95           | 0.24     | 5.23           | -0.04    |
| C21       | 5.05           | 5.09           | -0.05    | 5.02           | 0.02     |
| C22       | 5.08           | 5.03           | 0.05     | 5.13           | -0.05    |
| C23       | 4.99           | 5.03           | -0.04    | 4.83           | 0.16     |
| C24       | 5.00           | 5.05           | -0.05    | 4.97           | 0.03     |
| C25^T     | 5.09           | 5.22           | -0.13    | 4.91           | 0.18     |
| C26       | 4.95           | 5.04           | -0.09    | 4.96           | -0.01    |
| C27       | 4.97           | 5.17           | -0.20    | 4.98           | -0.01    |
| C28       | 4.87           | 5.06           | -0.19    | 5.02           | -0.15    |
| C29^T     | 4.91           | 4.98           | -0.07    | 5.13           | -0.22    |
| C30       | 4.96           | 4.94           | 0.01     | 4.83           | 0.12     |

"T = Test set.

Figure 3. Plot of the predicted training set and test set versus experimental pIC_{50} values for eq 3.

Figure 4. Plot of the predicted training set and test set versus experimental pIC_{50} values for eq 4.

■ DEFINING THE MODEL APPLICABILITY DOMAIN

The applicability domain is an important tool for reliable applications of QSAR models, and the characterization of interpolation space is significant for defining the applicability domain.33

In this work, we used the method developed by Roy et al.34 for the determination of the applicability domain (AD). The leveraged approach allows the determination of the position of a new chemical in the QSAR model, i.e., whether the new chemical will lie within the structural model domain or outside of it. Furthermore, the applicability domain was determined for all QSAR models by the leverage approach, along with the Williams plot.35

The leverage $h_i$ for each compound was used in the QSAR model to construct the Williams plot, and the warning leverage ($h^*_w$) was determined as36

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

where $n$ is the number of training compounds and $p$ is the number of predictor variables.

The defined applicability domain (AD) was then visualized via the Williams plot, the plot of the standardized residuals versus the leverage values ($h_i$). A compound with $h_i > h^*_w$ seriously influences the regression performance and may be excluded from the applicability domain, but it does not appear to be an outlier because its standardized residual may be small.

The Williams plot of the standardized residuals against the leverages is illustrated in Figure 5. The warning leverage ($h^*_w$) was found to be 0.62 for the developed 2D-QSAR models of eqs 1 and 2 that were expressed by the PLS method. Compound C5 has a higher leverage value than $h^*_w$ and is considered to be outside of the defined AD for eqs 1 and 2. The warning leverage ($h^*_w$) of models expressed by the stepwise-MLR method was found to be 0.5. Based on the leverages ($h_i > h^*_w$), compound C5 here was also found to be outside of the defined AD for both eqs 3 and 4. So, this compound C5 with $h_i > h^*_w$ affects the goodness of models (eqs 1–4) but it does not appear to be an outlier because its standardized residual is small.

■ DESIGN OF NEW (EGFR) KINASE INHIBITORS

Lv et al. observed that the most potent of compounds contained strong electron-donating substituents on II-ring at 4-position, which were explained by the order of potency OH > OCH3 > CH3. Among them, compound C5 presented the most potent EGFR inhibitory activity with IC_{50} of 0.07 μM, which was comparable to erlotinib (IC_{50} = 0.03 μM).24
Based on the above observation, a number of new compounds were designed containing multiple combinations of electron-donating substituents on II-ring at 4-position and multiple combinations of electron-donating and withdrawing substituents on I-ring at 4-position or 3,4-position. The new designed compounds listed in Table 6 were evaluated as EGFR kinase inhibitors by the best acceptable prediction equation (eq 4).

Four compounds from eleven designed compounds have better-predicted activity than the most active compound in the data series (C5, pIC50 = 7.15).

Table 6. Structures of Designed 3,5-Diphenyl-4,5-dihydro-1H-pyrazole-1-carbothioamide Derivatives with Their Predicted pIC50 and BCUT_PEOE_2 Values

| compounds | R1   | R2                      | pIC50 (predicted) | BCUT_PEOE_2 |
|-----------|------|-------------------------|-------------------|-------------|
| 1         | H    | 4-N(CH3)2                | 5.28              | 0.67        |
| 2         | 4-NH2| 4-N(CH3)2                | 5.21              | 0.66        |
| 3         | 4-Br | 4-N(CH3)2                | 5.13              | 0.67        |
| 4         | 4-NO2| 4-N(CH3)2                | 6.49              | 0.58        |
| 5         | 3,4-CH3 | 4-N(CH3)2            | 7.68              | 0.42        |
| 6         | 3,4-Cl | 4-N(CH3)2            | 5.23              | 0.66        |
| 7         | 3,4-Br | 4-N(CH3)2            | 5.08              | 0.66        |
| 8         | 4-NO2| 4-N(CH2CH3)2             | 5.30              | 0.66        |
| 9         | 3,4-CH3 | 4-N(CH2CH3)2          | 8.10              | 0.40        |
| 10        | 3,4-CH3 | 4-CH(OCH2CH3)2          | 8.39              | 0.39        |
| 11        | 3,4-CH3 | 4-O(CH2)3N(CH3)2        | 8.24              | 0.42        |
The high predicted activity of these compounds may be explained by electron-donating substituents present in two rings, leading to an increase in potency of 1H-pyrazole-1-carbothioamide derivatives as EGFR inhibitors. Additionally, it is observed that new compounds were designed to have low values for the dominant BCUT_PEOE_2 descriptor; this confirms the above observation that there is a strong inverse relationship between this descriptor and pIC₅₀. The newly designed compound 10 has the best predicted pIC₅₀ value with the BCUT_PEOE_2 value smaller than the others.

### MOLECULAR DOCKING RESULTS

Computer-aided drug design (CADD) methods have played an essential role in the discovery and design of new drug molecules in the past three decades. Molecular docking used in structure-based drug design (SBDD) is the most important technique in CADD due to its efficiency to predict the conformation of ligands within the suitable target-binding site with a considerable degree of accuracy.

This molecular docking study was executed over the selected protein from the protein data bank (PDB ID: 4hjo). The cocrystal structure of the EGFR kinase protein with the erlotinib ligand in the active site is illustrated in Figure 6.

![Figure 6. Three-dimensional structure of the protein (PDB ID: 4hjo) with the binding site pocket of EGFR kinase.](image)

The active pocket consisted of amino acid residues such as Asp831, Lys704, Thr830, Thr766, Gln767, Cys773, Gly772, Leu820, Leu794, Leu768, Met769, Pro770, Leu694, Phe771, Ala719, and Val702, which play fundamental roles by forming H-bonds and hydrophobic interactions. The docking for erlotinib was performed with its complex cocrystallized protein to validate the binding energy of ligand—protein interactions. The validation results showed a binding energy of −20.498 kcal/mol for erlotinib and a root-mean-square deviation (RMSD) value between 1.8 and 3.0 Å, indicating that this cocrystallized protein can be used for docking studies of other ligands.

Diagram showing the docking of erlotinib with the EGFR kinase.

The docking studies for these four designed compounds having high predicted activities were conducted along with CS, which has high experimental activity. Compounds 5, 9, 10, 11, and CS exhibited binding energies of −20.289, −21.866, −20.976, −18.146, and −18.166 kcal/mol, respectively. Additionally, the lengths of hydrogen bonds formed between ligands and amino acids of the active receptor pocket within the range <3 Å (2.11−2.65 Å) are listed in Table 7.

Three ligands 5, 9, and 11 showed a similar interaction with one of the side pocket Arg817 residues formed by strong hydrogen bonding between H−N of the carbothioamide group and O=C of the carbonyl group in the amino acid, as shown in Figures 7−9 (H-bond interactions are shown as blue dashed lines). Also shown in Figure 10 is ligand 10 interaction with a different mode, by forming three strong hydrogen-bond interactions with two amino acids Lys704 and Lys692 in the side pocket and the third amino acid Cys773 of the essential active site, which may explain the increase of high predicted activity of the designed ligand.

Moreover, CS showed an interaction with Asp831 in the side pocket by one hydrogen bond, as shown in Figure 11. In comparison to erlotinib, as a reference in Figure 12, there is one hydrogen-bond interaction with Met769 in the essential active site.

These results clearly reveal that the interactions of the reference ligand and CS are mostly the same as those observed in the four designed ligand complexes with the protein. These results indicate that 5, 9, 10, and 11 ligands act as selective inhibitors against EGFR kinase activity.

### CONCLUSIONS

A quantitative analysis of the structure–activity relationship (QSAR) was performed on a data set of 30 compounds of pyrazole derivatives containing a thiourea inhibitor of EGFR kinases. The 2D-QSAR model for a series was established using the partial least-squares (PLS) and stepwise multiple linear regression (SW-MLR) methods that yielded a regression model with improved predictive power. The results from the models suggested that descriptors including the atom count and bond count descriptors, lipophilic descriptor, adjacency distance matrix descriptors, and partial charge are the most important descriptors to explain the activity of the studied compounds.

### Table 7. Interactions Data and Binding Free Energies for Investigated Ligands with the Active Site of the Receptor (EGFR) Kinase (4hjo)

| compounds | free binding energy, S | type of bond interacted | interaction group | amino acid interacted | length (Å) |
|-----------|------------------------|-------------------------|------------------|-----------------------|------------|
| erlotinib | −20.498                | H-bond                  | N of the quinazoline ring | Met769               | 2.24       |
| CS        | −18.166                | H-bond                  | H−N of the carbothioamide group | Asp831               | 2.23       |
| 5         | −20.289                | H-bond                  | H−N of the carbothioamide group | Arg817               | 2.30       |
| 9         | −21.866                | H-bond                  | H−N of the carbothioamide group | Arg817               | 2.34       |
| 10        | −20.976                | H-bond                  | O of the OCH₂CH₃ group | Lys704               | 2.13       |
| 11        | −18.146                | H-bond                  | H−N of the carbothioamide group | Cys773               | 2.56       |

https://dx.doi.org/10.1021/acsomega.0c01323
ACS Omega 2020, 5, 18662−18674
There is a reverse relationship between BCUT_PEOE_2, density, $\log P_{(o/w)}$, and PEOE_VSA+1 descriptors for pyrazole-1-carbothioamide derivatives and the biological activity for EGFR kinase inhibitors as anticancer agents, especially the BCUT_PEOE_2 descriptor that has a strong reverse relationship with the biological activity, whereas the relationship is indirect with descriptors a_IC, a_acc, BCUT_SMR_1, diameter, and petitjean.

The accuracy and predictability of the proposed models were demonstrated by various criteria, including cross-validation,
Figure 10. 2D and 3D molecular interaction visualizations of ligand 10 with the active site of 4hjo.

Figure 11. 2D and 3D molecular interaction visualizations of C5 with the active site of 4hjo.

Figure 12. 2D and 3D molecular interaction visualizations of erlotinib with the active site of 4hjo.
external evaluation, the root-mean-square error (RMSE), warning leverage, and Williams plot within the applicability domain. All of these results proved the good statistical parameters of models to predict the activity for the development of new EGFR kinase inhibitor compounds as anticancer agents. The result of the developed 2D-QSAR model expressed by eq 4 was used to predict the biological activity (pIC_{50}) of newly designed pyrazole derivatives containing thiourea as the EGFR kinase inhibitors.

The result showed that the four designed ligands 5, 9, 10, and 11 can be potential inhibitors of the EGFR kinase protein. The ligands have lower binding energies and more molecular interactions with pockets. The result indicates that computer-aided drug design (CADD) can be an important discovery method for developing a new drug for various diseases.

Finally, we conclude that the most potent derivatives resulted in this study can be subjected to synthesis and pharmacological evaluations to develop highly potent (EGFR) kinase inhibitors as anticancer agents.

**EXPERIMENTAL SECTION**

**Data Sets.** A group of 30 1H-pyrazole-1-carbothioamide derivatives synthesized and evaluated as EGFR inhibitors by Lv et al. was used to carry out the 2D-QSAR study. The structures of 1H-pyrazole-1-carbothioamide derivatives with their biological activity are shown in Table 1. For the development of the 2D-QSAR model, this data set of compounds was manually divided into a training set (24 compounds) to create the model and a test set (6 compounds) for external validation of the predictive ability of the model.

The biological activities of these studied derivatives expressed in terms of IC_{50} (μM; the half-maximal inhibitory concentration) were converted to the logarithmic scale pIC_{50} (pIC_{50} = log 1/IC_{50} (μM)) and used as dependent variables in the 2D-QSAR model.

**2D-QSAR Model.** The geometry of the compounds was built with ACD/Labs 2018 (freeware v 14.00) and saved in .mol file format. Saved files were opened by MOE (Molecular Operating Environment—MOE 2009); to minimize their energy, a set of 185 2D descriptors for various types were selected (physical, atom count, and bond count, the Kier and Hall kappa molecular shape indices, the adjacency and distance matrices, pharmacophore features, and partial charge descriptors) based on atoms and connection information on structures of 3,5-diphenyl-4,5-dihydro-1H-pyrazole-1-carbothioamide derivatives. These descriptors were checked for the presence of constant or zero values to remove. The remaining descriptors for different types were reduced by eliminating the redundancy existing for descriptors, and the descriptors with a higher correlation with pIC_{50} were kept in the descriptor data matrix. In the partial least-squares (PLS) method, by forwarding selection, the various selected descriptors were correlated with the biological activity to allow the construction equations to be added one at a time.

The first descriptor having the highest correlation with the biological activity was included in the regression model, and then other new descriptors were progressively added to the model. Thus, the optimal model should be with the minimum number of descriptors to avoid overfitting of the model. Good fitting models with high squared correlation coefficient (r² > 0.6) were performed by the best five remaining descriptors: the third BCUT descriptor using PEOE partial charges (BCUT_−PEOE_2) is one of the BCUT descriptors calculated by Pearlman et al. from the eigenvalues of a modified adjacency matrix; the log P octanol/water partition coefficient (log P_{oct/wa}) is expressed by the logarithm of the n-octanol/water partition coefficient; the atom information content (total) (a_IC) is calculated as the atom information content (mean) (a_ICM) multiplied by n, where a_ICM is the entropy of the element distributed in the molecule, including implicit hydrogens but not lone pair pseudatoms and n is the sum of the number of occurrences of an atomic number in the molecule; the number of H-bond acceptor atoms (a-acc); and the mass density (density) is the ratio of the weight to the van der Waals volume. The stepwise multiple linear regression (SW-MLR) method was performed by the IBM SPSS statistics 23 software (version 23.0; SPSS Inc., Chicago, IL) to determine the relationship between these 25 remaining descriptors as independent variables and the biological activity of the compounds of the training set as an independent variable. To make sure that each descriptor is statistically significant in the model, these were determined by the hypothetical calculation of the p-value (sig) of the statistical confidence level (p < 0.05). The suitable descriptors were performed using SW-MLR models with r² > 0.6, including the third BCUT descriptor using PEOE partial charges (BCUT_−PEOE_2), the molar refractivity BCUT (1/3) (BCUT_SM-1) as the second BCUT descriptor using the atomic contribution to molar refractivity instead of partial charge, (PEOE_VSA+1) as the sum of van der Waals surface area with PEOE partial charge in the range (0.05, 0.10), (diameter) as the largest value in the distance matrix, and (petitjean) as the distance matrix and the difference between the diameter and radius divided by the diameter defined by Petitjean.

**Validation of the QSAR Model.** The QSAR model generated the “leave-one-out” (LOO) cross-validation scheme that was used to evaluate the predictive ability. From the LOO cross-validation procedure, the square of the cross-validation coefficient (q²) was obtained, which was used as a criterion to evaluate both the robustness and the predictive ability of the model generated.

The predictive ability of the model was evaluated by dependent variables of an external test set of six compounds that were not used for building the model.

**Modeling of 1H-Pyrazole-1-carbothioamide Derivatives.** About 35 1H-pyrazole-1-carbothioamide derivatives were designed according to structural templates of substituents in 3,4-positions of phenyl rings. All 11 new derivatives were selected to have a log P value less than 5, molecular weight less than 500, hydrogen-bond acceptors less than ten, and hydrogen-bond donors less than five. ACD/Labs 2018 (freeware v 14.00) was used to sketch these new compounds and saved in the .mol format. All .mol formats were opened by MOE 2009.10 software and energy-minimized, and the (fit) format of eq 4 was used to predict the biological activity of new 1H-pyrazole-1-carbothioamide derivatives in terms of pIC_{50}, as shown in Table 5.

**Molecular Docking.** The receptor (EGFR) tyrosine kinase binding with erlotinib used in this study was downloaded from the protein data bank (RCSB-PDB) website (http://www.rcsb.org/pdb/home/home.do) with the PDB ID 4hjo.

The 3D structure of the receptor protein was prepared using the MOE program according to the following steps. The stepwise multiple linear regression (SW-MLR) method was performed by the IBM SPSS statistics 23 software (version 23.0; SPSS Inc., Chicago, IL) to determine the relationship between these 25 remaining descriptors as independent variables and the biological activity of the compounds of the training set as an independent variable. To make sure that each descriptor is statistically significant in the model, these were determined by the hypothetical calculation of the p-value (sig) of the statistical confidence level (p < 0.05). The suitable descriptors were performed using SW-MLR models with r² > 0.6, including the third BCUT descriptor using PEOE partial charges (BCUT_−PEOE_2), the molar refractivity BCUT (1/3) (BCUT_SM-1) as the second BCUT descriptor using the atomic contribution to molar refractivity instead of partial charge, (PEOE_VSA+1) as the sum of van der Waals surface area with PEOE partial charge in the range (0.05, 0.10), (diameter) as the largest value in the distance matrix, and (petitjean) as the distance matrix and the difference between the diameter and radius divided by the diameter defined by Petitjean.

The predictive ability of the model was evaluated by dependent variables of an external test set of six compounds that were not used for building the model.

**Modeling of 1H-Pyrazole-1-carbothioamide Derivatives.** About 35 1H-pyrazole-1-carbothioamide derivatives were designed according to structural templates of substituents in 3,4-positions of phenyl rings. All 11 new derivatives were selected to have a log P value less than 5, molecular weight less than 500, hydrogen-bond acceptors less than ten, and hydrogen-bond donors less than five. ACD/Labs 2018 (freeware v 14.00) was used to sketch these new compounds and saved in the .mol format. All .mol formats were opened by MOE 2009.10 software and energy-minimized, and the (fit) format of eq 4 was used to predict the biological activity of new 1H-pyrazole-1-carbothioamide derivatives in terms of pIC_{50}, as shown in Table 5.

**Molecular Docking.** The receptor (EGFR) tyrosine kinase binding with erlotinib used in this study was downloaded from the protein data bank (RCSB-PDB) website (http://www.rcsb.org/pdb/home/home.do) with the PDB ID 4hjo.

The 3D structure of the receptor protein was prepared using the MOE program according to the following steps. The first step is protonation with the “Protonate 3D” feature to add the hydrogen atoms into the crystal structure, partial-charged, and energy-minimized, respectively. Other parameters were adjusted by the MMFF94x force field, and finally, the output files were saved in .mol file format to be ready for the docking process.
The 2D structure of ligands was built with ACD/Labs 2018 (freeware v 14.00) and saved in the .mol file format. The database of ligands was made by the MOE database viewer in the MDB file format.

The molecular docking simulation process was performed with the “Dock” feature in the MOE program. The placement method of “Triangle Matcher” was used with a repetition of energy readings per 1,000,000 positions, “London dG” was used as the Rescoring 1 function with the repetition population of 1000, refinement was used as the force field, the first repetition was counted ten times, and the second set shows only one best result of 100 repetitions. After this, the Gibbs free energy (ΔG binding) value and hydrogen-bond interactions of ligands with amino acids were observed and analyzed to determine the best ligand.20 Table 6 presents all information about binding free energies and interactions for ligands.

■ AUTHOR INFORMATION
Corresponding Author
Tawassl T. H. Hajalsiddig — Department of Chemistry, College of Science, Sudan University of Science and Technology, Khartoum, Sudan; orcid.org/0000-0001-8053-6339;
Email: tawasslt@gmail.com

Authors
Abu Baker M. Osman — Department of Chemistry, College of Science & Arts (Suratatabidha), King Khalid University, Abha, Saudi Arabia
Ahmed E. M. Saeed — Department of Chemistry, College of Science, Sudan University of Science and Technology, Khartoum, Sudan

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c01323

Notes
The authors declare no competing financial interest. Author Email: abubker123@gmail.com (A.M.O.), aemsaeed@gmail.com (A.E.M.S.).

■ ACKNOWLEDGMENTS
We thank everyone who contributed and helped with their constructive suggestions during the planning and development of this work.

■ REFERENCES
(1) Reise, D. J.; Stern, D. F. Specificity within the EGF Family/ErbB receptor family signaling network. BioEssays 1998, 20, 41–48.
(2) Yardim, Y.; Sliskowski, M. X. Untangling the ErbB signalling network. Nat. Rev. Mol. Cell Biol. 2004, 2, 127–137.
(3) Hsu, J. L.; Hung, M. C. The role of HER2, EGFR, and other receptor tyrosine kinases in breast cancer. Cancer Metastasis Rev. 2016, 35, 575–588.
(4) Hsu, J. L.; Hung, M. C.; Chang, W. C.; Wang, Y. H.; Tsai, C. H.; Hung, M. C. HER2, EGFR, and other receptor tyrosine kinases in breast cancer. Cancer Metastasis Rev. 2016, 35, 575–588.
(5) Hsu, J. L.; Hung, M. C.; Chang, W. C.; Wang, Y. H.; Tsai, C. H.; Hung, M. C. HER2, EGFR, and other receptor tyrosine kinases in breast cancer. Cancer Metastasis Rev. 2016, 35, 575–588.
(6) Hossam, M.; Lasheen, D. S.; Abouzid, K. A. M. Covalent EGFR inhibitors: binding mechanisms, synthetic approaches, and clinical profiles. Arch. Pharm. 2016, 349, 573–593.
(7) Liu, F.; Tang, B.; Liu, H.; Li, L.; Liu, G.; Cheng, Y.; Xu, Y.; Chen, W.; Huang, Y. 4-Aminooquinazoline derivatives with epidermal growth factor receptor inhibitor activity. Anti-Cancer Agents Med. Chem. 2016, 16, 1652–1664.
(8) Park, H. J.; Lee, K.; Park, S. J.; Ahn, B.; Lee, J. C.; Cho, H.; Lee, K. I. Identification of antitumor activity of pyrazole oxime ethers. Bioorg. Med. Chem. Lett. 2005, 15, 3307–3312.
(9) Manna, K.; Agrawal, Y. K. Microwave assisted synthesis of new indophenazine 1,3,5-trisubstituted pyrazoline derivatives of benzofuran and their antimicrobial activity. Bioorg. Med. Chem. Lett. 2009, 19, 2688–2692.
(10) Bekhit, A. A.; Fahmy, H. T.; Rostom, S. V.; Baraka, A. M. Design and synthesis of some substituted 1H-pyrazolyl-thiazolo[4,5-d]-pyrimidines as anti-inflammatory–antimicrobial agents. Eur. J. Med. Chem. 2003, 38, 27–36.
(11) Sridhar, R.; Perumal, P. T.; Etti, S.; Shanmugam, G.; Ponnumwamy, M. N.; Prabavathy, V. R.; Mathivanan, N. Design, synthesis and anti-microbial activity of 1H-pyrazole carboxylates. Bioorg. Med. Chem. Lett. 2004, 14, 6035–6040.
(12) Mohamed, M. S.; Zohny, Y. M.; El-Senouy, W. M.; Abou, A. M. Synthesis and biological screening of novel pyrazoles and their precursors as potential antiviral agents. Pharmaceutophore 2018, 9, 126–139.
(13) Bhattacharjee, D.; Chudasama, C. J.; Patel, K. D. Pyrazole clubbed triazolo[1,5-a]pyrimidine hybrids as anti-tubercular agents: synthesis, in vitro screening and molecular docking study. Bioorg. Med. Chem. 2015, 23, 7711–7716.
(14) Xu, X.; Gao, C.; Ren, Q. C.; Song, X. F.; Feng, L. S.; Lv, Z. S. Recent advances of pyrazole-containing derivatives as anti-tubercular agents. Eur. J. Med. Chem. 2017, 139, 429–440.
(15) Bekhit, A. A.; Saudi, M. N.; Hassan, A. M. M.; Fahmy, S. M.; Ibrahim, T. M.; Ghareeb, D.; Bekhit, A. E. D. A.; et al. Synthesis, in silico experiments and biological evaluation of 1, 3, 4-trisubstituted pyrazole derivatives as antimalarial agents. Eur. J. Med. Chem. 2019, 163, 353–366.
(16) Bandgar, B. P.; Gawande, S. S.; Bodade, R. G.; Gawande, N. M.; Khabragade, C. N. Synthesis and biological evaluation of a novel series of pyrazole chalcones as anti-inflammatory, antioxidant and antimicrobial agents. Bioorg. Med. Chem. 2009, 17, 8168–8173.
(17) Abbaldali, K. R. A.; Falady, W. A.; Kamel, G. M.; Elshaier, Y. A.; El-Magd, M. A. Design, synthesis, modeling studies and biological evaluation of thiazolidine derivatives containing pyrazole core as potential anti-diabetic PPAR-γ agonists and anti-inflammatory COX-2 selective inhibitors. Bioorg. Chem. 2019, 82, 86–99.
(18) Gulnaz, A. R.; Mohammed, Y. H.; Khanum, S. A. Design, synthesis and molecular docking of benzeneophene conjugated with oxadiazole sulphur bridge pyrazole pharmacophores as anti-inflammatory and analgesic agents. Bioorg. Chem. 2019, 92, No. 103220.
(19) Pogaku, V. J.; Gangarapu, K.; Basavaju, S.; Tatapudi, K. K.; Katragadda, S. B. Design, synthesis, molecular modelling, ADMET prediction and anti-hyperglycemic evaluation of new pyrazole-triazolopyrimidine hybrids as potent α-glucosidase inhibitors. Bioorg. Chem. 2019, 93, No. 103307.
(20) Akhtar, M. J.; Khan, A. A.; Ali, Z.; Dewangan, R. P.; Rafi, M.; Hassan, M. Q.; Akhtar, M. D.; Siddiqui, A. A.; Partap, S.; Pasha, S.; Yar, M. S. Synthesis of stable benzimidazole derivatives bearing pyrazole as anticancer and EGFR receptor inhibitors. Bioorg. Chem. 2018, 78, 158–169.
(21) Tao, X. X.; Duan, Y. T.; Chen, L. W.; Tang, D. J.; Yang, M. R.; Wang, P. F.; Xu, C.; Zhu, H. L. Design, synthesis and biological evaluation of pyrazolyl-sitromidazole derivatives as potential EGFR/HER2 kinase inhibitors. Bioorg. Med. Chem. Lett. 2016, 26, 677–683.
(22) Tumma, R.; Vamaraju, H. B. Design, synthesis, molecular docking, and biological evaluation of pyrazole-1-carbethoximide incorporated isoxazole derivatives. Asian J. Pharm. Clin. Res. 2019, 12, 245–250.
(23) Sunayana, G.; Shashikant, B.; Sandeep, W. 2D, 3D, G-QSAR and docking studies of thiazolyl-pyrazoline analogues as potent (epidermal growth factor receptor-tyrosine kinase) EGFR-TK inhibitors. Lett. Drug Des. Discovery 2017, 14, 1228–1238.
(24) Lv, P.-C.; Li, H.-Q.; Sun, J.; Zhou, Y.; Zhu, H.-L. Synthesis and biological evaluation of pyrazole derivatives containing thiourea skeleton as anticancer agents. *Bioorg. Med. Chem.* 2010, 18, 4606−4614.

(25) Lv, P.-C.; Li, D.-D.; Li, Q.-S.; Lu, X.; Xiao, Z.-P.; Zhu, H.-L. Synthesis, molecular docking and evaluation of thiazolyl-pyrazole derivatives as EGFR TK inhibitors and potential anticancer agents. *Bioorg. Med. Chem. Lett.* 2011, 21, 5374−5377.

(26) Sanmati, K. J.; Rahul, J.; Lokesh, S.; Arvind, K. Y. QSAR analysis on 3,5-disubstituted-4, 5-dihydropyrazole-1-carbothioamides as epidermal growth factor receptor (EGFR) kinase inhibitors. *J. Chem. Pharm. Res.* 2012, 4, 3215−3223.

(27) Jalali-Heravi, M.; Kyani, A. Use of computer-assisted methods for the modeling of the retention time of a variety of volatile organic compounds: A PCA-MLR-ANN approach. *J. Chem. Inf. Comput. Sci.* 2004, 44, 1328−1335.

(28) Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inf.* 2010, 29, 476−488.

(29) Burden, F. R. Molecular identification number for substructure searches. *J. Chem. Inf. Model.* 1989, 29, 225−227.

(30) Pearlman, R. S.; Smith, K. M. Novel software tools for chemical diversity. In 3D QSAR in Drug Design; Springer: Dordrecht, 2002; Vol. 2, pp 339−353.

(31) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* 1980, 36, 3219−3228.

(32) Petitjean, M. Applications of the radius-diameter diagram to the classification of topological and geometrical shapes of chemical compounds. *J. Chem. Inf. Model.* 1992, 32, 331−337.

(33) Gadaleta, D.; Mangiardi, G. F.; Catto, M.; Carotti, A.; Nicolotti, O. Applicability domain for QSAR models: where theory meets reality. *Int. J. Quant. Struct.-Prop. Relat.* 2016, 1, 45−63.

(34) Roy, K.; Kar, S.; Ambure, P. On a simple approach for determining applicability domain of QSAR models. *Chemom. Intell. Lab. Syst.* 2015, 145, 22−29.

(35) Asadollahi, T.; Dadfarnia, S.; Shabani, A. M. H.; Ghasemi, J. B.; Sarkhosh, M. QSAR models for CXCR2 receptor antagonists based on the genetic algorithm for data preprocessing prior to application of the PLS linear regression method and design of the new compounds using in silico virtual screening. *Molecules* 2011, 16, 1928−1955.

(36) Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; Mcdowell, R. M.; Gramatica, P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ. Health Perspect.* 2003, 111, 1361−1375.

(37) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational methods in drug discovery. *Pharmacol. Rev.* 2014, 66, 334−395.

(38) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Structure-based virtual screening and lead optimization: methods and applications. *Nat. Rev. Drug Discovery* 2004, 3, 935−949.

(39) Shao, J. Linear model selection by cross-validation. *J. Am. Stat. Assoc.* 1993, 88, 486−494.

(40) Corbeil, C. R.; Williams, C. I.; Labute, P. Variability in docking success rates due to dataset preparation. *J. Comput.-Aided Mol. Des.* 2012, 26, 775−786.