

## 2D-QSAR and Molecular Docking Studies of *N*-Substituted Rhodanine Derivatives as Potent Cytotoxic Agents

Medidi Srinivas<sup>1\*</sup>, K Grace Neharika<sup>2</sup>, and K Sri Gowri Priya<sup>2</sup>

<sup>1</sup>Drug Discovery Lab, Geethanjali College of Pharmacy (Affiliated to JNTUH), Cheeryal (V), Hyderabad-501301, India.

<sup>2</sup>Sir CR Reddy College of Pharmaceutical Sciences (Affiliated to AU), Eluru, AP-534007, India.

\*Address for correspondence:

Drug Discovery Lab, Geethanjali College of Pharmacy (Affiliated to JNTUH), Cheeryal (V), Hyderabad-501301, India.

Tel: +91 7013456837; E-mail: drmsr9@gmail.com

### ABSTRACT

*HER2 positive breast cancer is most common malignancy in women globally. Aberrant behavior of HER2 kinase protein is a hallmark of tumorigenesis, and as a result it has been considered as an emerging potential drug target for breast cancer therapy. In the present investigation a series of N-substituted rhodanine derivatives (33 compounds) were subjected to 2D-QSAR studies with the aid of genetic algorithm (GA) method to identify the essential structural features that responsible for cytotoxic activity. Based on the results, the cytotoxic activity of N-substituted rhodanine derivatives can be successfully explained in terms of two-dimensional (ATSC7v), and three-dimensional (geomRadius and RDF45i) descriptors. The obtained model was vigorously validated and passed all validation metrics ( $R^2_{train} = 0.913$ ,  $R^2_{adj} = 0.899$ ,  $Q^2_{LOO} = 0.870$ ,  $R^2_{test} = 0.848$ ). Importantly, the model quality was good based on mean absolute error (MAE) criteria and the results were consistent with proposed limits by Golbraikh and Tropsha. Molecular docking study of the active compounds (ligands 4, 16, and 27) revealed the formation of hydrogen and hydrophobic interactions within the active site of HER2 protein.*

**Keywords:** HER2 proteins, breast cancer, cytotoxic agents, rhodanine, molecular docking, 2D-QSAR.

## 1. INTRODUCTION

Cancer is the second top cause of death globally after cardiovascular diseases. It was estimated that 9.6 million deaths in 2018 were due to cancer [1, 2]. Cancer is considered the leading cause of death in higher-income countries while the second in low and middle-income countries. The most common types of cancer in men are lung, prostate, colorectal, stomach, and liver cancer. In women, cancer prevalence is highest in the breast, colorectal, lung, cervical, and thyroid, respectively [3]. In a recent report of the American Cancer Society, breast cancer was classified as the most frequently diagnosed cancer and the leading cause of cancer death among females [4]. In terms of new cases, an estimated 2.1 million women were newly diagnosed, and 626,679 women died with breast cancer [5]. The global incidence of breast cancer has been rising with annual increases of 3.1% [6, 7]. The human epidermal growth factor receptor (EGFR) protein-tyrosine kinases plays a key role in cell proliferation, survival, differentiation, and metabolism. Disturbances in these vital molecular processes are a hall mark of tumorigenesis [8]. The EGFR family composed of four main members such as ErbB1 (HER1), ErbB2 (HER2), ErbB3 (HER3), and ErbB4 (HER4) [9, 10]. Several cancers are connected with the amplification or increased expression of EGFR family proteins including head and neck, lung, stomach, colorectal, bladder, uterine and pancreatic cancers [11]. In recent years, the Food and Drug Administration (FDA) has been approved thirty-six small-molecule kinase inhibitors for the treatment of many cancer diseases with wide structural diversity [12]. Therefore, protein kinases have been recognized as an attractive therapeutic target for the discovery of anti-cancer drugs, and thus the development of novel kinase targeting agents is very promising in the field of oncology.

Rhodanine (2-thioxothiazolidin-4-one) is considered as privileged scaffold in drug discovery and receiving considerable attention in medicinal chemist community due to their broad therapeutic activities such as antimicrobial [13-15], antiviral [16], antidiabetic [17] and anti-cancer activity [18]. Additionally, rhodanine-based molecules have been popular as small molecule inhibitors of numerous targets such as HCV NS5B protease [19], HCV NS3 protease [20], aldose reductase [21],  $\beta$ -lactamase [22], UDP-*N*-acetylmuramase/*L*-alanine ligase [23], fungal protein mannosyl transferase-1 (PMT1) [24], cathepsin-D [25], anthrax lethal factor protease [26], histidine decarboxylase [27], JNK-stimulating phosphatase-1 (JSP-1) [28] and phosphodiesterase (PDE-4) [29]. Among the thiazolidine derivatives, numerous compounds containing thiazolidine-2,4-dione and rhodanine have been recognized as new potential anti-cancer agents. For example, GSK1059615 (see Fig. 1) is a potent, reversible, ATP-competitive, thiazolidinedione inhibitor of PI3K  $\alpha$ . It inhibits phosphatidylinositol 3-kinase (PI3K) signaling, induces G1 arrest and apoptosis, especially in breast tumor cells [30, 31]. The diverse biological activities of rhodanine and 2,4-thiazolidinone derivatives has attracted attention of medicinal chemists to explore this scaffold in the discovery of new potential anti-cancer agents [32-36].

Quantitative structure activity relationship (QSAR) modeling is a part of computer-aided drug design and plays a pivotal role in lead optimization strategy of drug discovery program [37]. QSAR model is a mathematical linear equation, which explains the relationship between

descriptors (computed properties of chemical structures under study) as an independent variable and a response (experimental activity/toxicity/property) as a dependent variable for set of structurally similar molecules under study [38, 39]. Once a QSAR model is established, it is very useful for exploring the structural features responsible for a biological activity and also feeds essential molecular information to design new molecules with enhanced biological activity of our interest. Additionally, the built QSAR model can be applied to predict the biological activity of newly designed molecules. Therefore, QSAR modeling speeds up the process of development of new drug molecules by saving time, money and more importantly animal sacrifice. The purpose of the present study was to investigate the relationship between the physicochemical parameters responsible for the cytotoxic activity of *N*-substituted rhodanine derivatives. The 2D-QSAR analysis was carried out by employing genetic algorithm method for descriptor optimization and multiple linear regressions (MLRs) analysis for QSAR models development. To the best of our knowledge, till date no 2D-QSAR study has been reported on *N*-substituted rhodanine derivatives as cytotoxic agents as discussed in this article. Hence, we have attempted to develop some robust statistically significant 2D-QSAR model for rhodanine derivatives to correlate cytotoxic activity to its physicochemical properties. The information generated from the 2D-QSAR could be effective to understanding the structure activity relationships of the compounds under consideration and subsequently design a new anti-cancer lead compound. In structure based drug design, molecular docking approach can be help full in predicting the intermolecular interactions between a small molecule and a protein or enzyme at atomic level. This study focused on combining both 2D-QSAR modeling and molecular docking approaches to the assessment of thirty-three *N*-substituted rhodanine derivatives as cytotoxic agents.

## 2. MATERIALS AND METHODS

### 2.1. Dataset

Thirty three (33) already synthesized *N*-substituted rhodanine derivatives with well define activity were obtained from the literature for QSAR study [40]. The biological activity data in the form of % cytotoxicity values were converted to their natural logarithms (%Cytotoxicity = lnCyt) in order to reduce skewness in the activity values used for the 2D-QSAR study. The chemical structures and activity values (lnCyt) of these 33 compounds are represented in Table 1.

### 2.2. Structure Optimization and Molecular Descriptor Calculation

The structure optimization was performed using CS Chem Office version 8.0 software [41]. The molecular structures of the *N*-substituted rhodanine derivatives were properly drawn using ChemDraw Ultra module of the software. The sketched structures without any errors were transferred to Chem3D Ultra module of the same software to create the three-dimensional (3D) structure. These 3D structures were then subjected to energy minimization using molecular mechanics (MM2) server until the root-mean square (RMS) gradient value became smaller than 0.1 kcal/molÅ. All energy minimized molecules were again refined using Austin model-1 (AM1) method using the restricted closed-shell wave function of the molecular orbital package

(MOPAC) module until the RMS gradient value became 0.0001 kcal/molÅ. Geometrical optimization was carried out to obtain the lowest energy structure using EF (Eigenvector Following) routine. Most stable structure for each compound was generated and saved as SDF format, and then exported to PaDEL descriptors software which is a product of Pharmaceutical Data Exploration Laboratory, developed by Yap Chun Wai [42]. The total pool of 1875 descriptors were calculated which includes electrostatic, topological, spatial, autocorrelation, geometrical, constitutional, and thermodynamic descriptors.

### 2.3. Data Pretreatment

After calculating the descriptors, data pretreatment was done using a software tool (Data Pretreatment GUI 1.2 software) [43]. The pool of 1875 descriptors were subjected to data pretreatment to remove constant (variance cut-off < 0.0001) and inter-correlated descriptors (correlation coefficient cut-off  $\geq 0.90$ ) in order to minimize redundant information. If multiple descriptors were inter-correlated, the descriptor with a higher correlation with biological activity was considered for study. After data pretreatment of the descriptors matrix, 745 descriptors remained for 2D-QSAR model development.

### 2.4. Dataset Division

In this study, user friendly QSAR modeling software “QSARINS (QSAR-Insurbia)” developed at the University of Insurbia was used for dataset division, model development and validation. This software allows to generate multiple linear regression (MLR) based QSAR models, by employing genetic algorithm for descriptor selection. This tool is very useful in computing QSAR modeling, which employs an exhaustive double cross-validation approach to select an optimal model for small dataset QSAR modeling [44-46]. After data pretreatment, data set division into training and test sets is considered as most important step in QSAR study because the training and test set should cover the both active and inactive compounds for uniform data sampling. The data set was divided into a test set of 10 compounds (1, 3, 8, 11, 12, 17, 18, 19, 23, and 30), and training set of 23 compounds (2, 4, 5, 6, 7, 9, 10, 13, 14, 15, 16, 20, 21, 22, 24, 25, 26, 27, 28, 29, 31, 32, and 33) by random splitting method in proportion of 70 to 30 % through response (activity) sampling automatically by the software.

### 2.5. Variable Selection

Variable (descriptor) selection process is a key step in development of best QSAR model from the pool of molecular variables, because only few of them correlates well with biological activity with high statistical significance. Therefore, we adopted GFA and MLR approaches for best variable selection as well as for the development of QSAR models.

#### 2.5.1. GFA Method

GFA (genetic function algorithm) is a search optimization process based on the theory of biological evolution. GFA evolved from algorithms known as (i) Holland’s genetic algorithm (ii)

Friedman's multivariate adaptive regression splines (MARS) algorithm [47, 48]. GFA in QSARINS software executes in the following way: Initially,  $N$  regression equations were generated randomly with two distinct variables from the available dataset (set at 500 by default in QSARINS software) then pairs of "parent" equations were chosen randomly from this set of 500 equations and random crossovers were performed to generate progeny equations. The process repeated several times until a reasonable convergence was obtained. The equations usually evaluated by fitness score using inbuilt fitness functions such as LOF,  $Q^2_{LOO}$ ,  $R^2_{adj}$ , and  $RMSE_{cv}$ . In the present study, Friedman's lack of fit (LOF) score was used to assess the fitness of each progeny equation. LOF score penalizes over fitted models and estimates suitable number of variables in the equation. The lack of fit is calculated by using the following formula (1):

$$\text{Lack of Fit} = \text{LSE} / [1 - (c + dp) / m]^2 \quad (1)$$

Where,  $m$  = number of samples in the training set,  $c$  = number of basis functions,  $d$  = smoothing parameter,  $p$  = number of independent variables, and LSE = least square error.

## 2.6. Construction of QSAR Model

The 2D-QSAR model was constructed by employing multiple linear regression (MLR) method using selected variables. The MLR is the traditional and standard approach for multivariate data analysis, it was used to study the relation between one dependent variable (biological activity) and several independent variables (molecular descriptor) [49, 50]. In this study, best model was selected based on statistical quality and with minimum number of reliable descriptors which able to explain variance in the activity.

## 2.6. QSAR Model Validation

### 2.6.1. Internal validation

The developed QSAR models were evaluated using the following statistical fitting parameters: correlation coefficient ( $R$ ); squared correlation coefficient ( $R^2$ );  $R^2_{adj}$ ;  $SEE$  (standard error of estimate);  $PRESS$  (predicted residual error sum of squares);  $S_{DEP}$  (standard deviation error of prediction);  $F$ -test for statistical significance,  $K_{xx}$  (global correlation among descriptors) and  $LOF$  (lack of fit). The regression coefficient  $R^2$  is a relative measure of fit by the regression equation, and it represents the variation in the observed data that is explained by the regression. Furthermore, the  $R^2$  values are proportional to the number of descriptors in the model which is not reliable in determining the predictive response of the model. The  $R^2_{adj}$  was evaluated based on following formula (2):

$$R^2_{adj} = \frac{R^2 - p(n-1)}{n-p-1} \quad (2)$$

Where  $p$  is the number of the descriptor in the equation and  $n$  is number of compounds in the training set. The number of descriptors in the QSAR model is acceptable when the difference between  $R^2$  and  $R^2_{adj}$  value is less than 0.3 [51]. The  $F$ -test 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 low standard error of estimate (*SEE*) shows absolute quality of fitness of the model. *PRESS* is the sum of overall compounds of the squared differences between the actual and predicted values for the independent variables. The small value of *PRESS* statistics indicates better prediction ability of QSAR equation. The results of fitting parameters of developed QSAR model is represented in Table 5.

The QSAR model generated was internally validated using cross-validation technique [52]. In essence, this technique provides adequate information about the predictive reliability of the QSAR equation. The cross-validated  $Q^2_{\text{LOO}}$  was evaluated based on following formula (3):

$$Q^2_{\text{LOO}} = 1 - \left[ \frac{\sum(Y_{\text{obs}} - Y_{\text{pred}})^2}{\sum(Y_{\text{obs}} - Y_{\text{train}})^2} \right] \quad (3)$$

Where,  $Y_{\text{train}}$  = average observed the concentration of training set,  $Y_{\text{obs}}$  = observed concentration, and  $Y_{\text{pred}}$  = predicted concentration in the training set. The squared correlation coefficient ( $R^2_{\text{pred}}$ ) was determined for the comparison between the predicted concentration by the QSAR equation and observed concentrations from the experiment. However, a QSAR model is considered to be predictive, if the following conditions are satisfied:  $R^2 > 0.6$ ,  $Q^2 > 0.6$  and  $R^2_{\text{pred}} > 0.5$  [52]. The cross validated  $Q^2_{\text{LOO}}$  is ordinarily smaller than the  $R^2$  value of the QSAR model because of its diagnostic means of evaluating the predictive power of the model [53]. In addition, it is important to note that good  $R^2$  and  $Q^2_{\text{LOO}}$  values are not enough measures for validating the model. Therefore, more parameters must be established to point out the predictive capability of the models.

The developed QSAR models were further evaluated using the following statistical internal validation parameters:  $RMSE_{\text{train}}$  (root-mean-square error of the training set);  $RMSE_{\text{cv}}$  (root-mean-square error of the training set estimated through cross validated leave one out method);  $RMSE_{\text{ext}}$  (root-mean-square error of the external validation set);  $CCC$  (concordance correlation coefficient) of the training set ( $CCC_{\text{train}}$ ), LOO cross validated concordance correlation coefficient of test set ( $CCC_{\text{cv}}$ ), and external validation set ( $CCC_{\text{ext}}$ ) [54], and  $PRESS_{\text{cv}}$  (predictive residual sum of squares by employing cross-validated LOO method in the training set, and in the external prediction set ( $PRESS_{\text{ext}}$ )). The robustness of the model was validated by computing scaled  $r^2_{\text{m}}$  metrics [55] such as  $r^2_{\text{m (train)}}$  and  $\Delta r^2_{\text{m (train)}}$  for internal validation;  $r^2_{\text{m (test)}}$ , and  $\Delta r^2_{\text{m (test)}}$  for test set validation. The calculation of  $r^2_{\text{m}}$  metrics is based on the idea that the classical metrics of internal ( $Q^2$ ) and external ( $R^2_{\text{pred}}$ ) validation largely depend on the mean value of the response parameter for the training set compounds [56]. Thus, the values of these metrics become dependent on the range selected for the response parameter and fail to reflect truly the deviation of the predicted response values from the corresponding observed data especially when the y-range of the dataset is quite wide [57]. Acceptable values for these metrics may be obtained as long as the mean response of the training set compounds maintains sufficient distance from the observed data for both the training and the test set molecules, even when poor predictions for some of the compounds in the two sets are present. Thus, large values of these metrics are not always indicative of the most significant model. To

obviate such errors, the modified  $r^2$  ( $r_m^2$ ) metrics were reported by Roy and Roy [55]. The  $r_m^2$  metric may be calculated for both the training and the test sets in addition to the total dataset based on the leave-one out (LOO) predicted (training set) and/or predicted (test set) response and the corresponding observed data. Here, two different variants of the  $r_m^2$  metrics are calculated to assess the predictive ability of the model:  $r_m^2$  and  $\Delta r_m^2$  [56]. It has been suggested that for a model to be considered as predictive one, the value of  $r_m^2$  should be more than 0.5 and the  $\Delta r_m^2$  value should be lower than 0.2 [58]. In the present study we also employed some other statistical validation parameters (external prediction) such as  $Q^2F1$ ,  $Q^2F2$ , and  $Q^2F3$ , to assure the significance of the developed model [59]. The internal and external validation results are given in Table 6 and 7.

The  $Y$ -randomization test is a widely used approach to test the model robustness. In this approach the steps followed during the randomization test are (i) repeatedly scrambling the activity data in the training set molecules, (ii) using the randomized data to generate QSAR models, and (iii) comparing the resulting scores with the score of the original QSAR model generated with non-randomized data. If the activity prediction of the random model is comparable to that of the original model, the set of observations is not sufficient to support the model. The new QSAR models (after several iterations) would be expected to have low  $R^2$  and  $Q^2_{LOO}$  values from the values of actual model. If the opposite happens, then an acceptable QSAR model cannot be obtained for the specific modeling method and data [60]. The results of  $Y$ -randomization test is given in Table 8.

### 2.6.2. External Validation

In the external validation phase, the QSAR model assessed for acceptability criteria proposed by Golbraikh and Tropsha for robustness with good predictive potential [53, 61].

- a.  $Q^2 > 0.5$
- b.  $R^2_{pred} > 0.6$
- c.  $r^2 - r_0^2 / r^2 < 0.1$
- d.  $0.85 < k < 1.15$  or  $0.85 < k' < 1.15$

Where,  $r^2$  = squared correlation coefficient between the observed and predicted activities,  $r_0^2$  = squared correlation coefficient between the predicted and observed activities,  $k$  and  $k'$  are the regression slopes passing through the origin. The  $R^2_{pred}$  was evaluated based on following formula (4):

$$R^2_{pred} = 1 - \left[ \frac{\sum(Y_{pred\ test} - Y_{obs\ test})^2}{\sum(Y_{obs\ test} - Y_{train})^2} \right] \quad (4)$$

Where,  $Y_{pred\ test}$  and  $Y_{obs\ test}$  are the predicted and observed activity of test set compounds respectively.  $Y_{train}$  is the average values of observed activity of the training set compounds. Golbraikh and Tropsha acceptable model validation parameters and their threshold values of the model-1 are shown in the Table 9.

Further, the model prediction quality (training and test sets) was studied using the mean absolute error (MAE)-based criteria [62, 63]. The statistical metrics like MAE<sub>train</sub> (mean absolute

error of the training set), and MAE<sub>ext</sub> (mean absolute error of the external validation set) were computed using a software tool (XternalValidation-Plus 1.1) [43]. The validation tool categorizes the model prediction quality into ‘Good’, ‘Moderate’ and ‘Bad’ considering the values of MAE and standard deviation of the absolute error (AE) values ( $\sigma$ AE) as defined as below:

(i) Good predictions:

$$\text{MAE} \leq [0.1 \times \text{training set range AND MAE}] + [3 \times \sigma\text{AE} \leq 0.2 \times \text{training set range}]$$

(ii) Bad predictions:

$$\text{MAE} > [0.15 \times \text{training set range OR MAE}] + [3 \times \sigma\text{AE} > 0.25 \times \text{training set range}]$$

(iii) Moderate predictions:

The predictions which do not fall under either of the above two conditions are considered as of ‘moderate’ quality. The results of MAE based criteria are shown in Table 10.

### 2.6.3. Development of Applicability Domain (AD)

Applicability domain (AD) analysis helps us to understand the reliability of built QSAR model predictions in the response and chemical structure space. This analysis is used to detect structural and response outliers from the test set and training set respectively [64]. The leverage approach of determining the applicability domain also known as Williams plot was obtained by plotting a scatter plot of standardized residual in the  $Y$ -axis and leverages in the  $X$ -axis of both training, and test sets. By definition, leverage value [65, 66] is determined based on following formula (5):

$$\text{Leverage } (i) = X(i) (x^T x)^{-1} X(i) \quad (5)$$

where  $X(i)$  represents the vector of descriptors of compound ( $i$ ),  $X$  represents the descriptors matrix, and  $x^T$  represents matrix transpose of  $X$ . The threshold leverage ( $F^*$ ) is given by the formula (6):

$$F^* = \frac{3(m+1)}{p} \quad (6)$$

where  $p$  is the number of molecules in the training set and  $m$  is the number of molecular descriptors used in the model. In addition, compounds with higher leverage scores which are greater than threshold leverage ( $i > F^*$ ) tend to have unreliable predictions. However, compounds whose leverage scores are less than the threshold score ( $i < F^*$ ) and the standardized residuals are not greater than  $\pm 3\alpha$  (3 standard deviation units) are said to fall within the applicability domain.

The statistical metrics and their mathematical definitions used in this QSAR study can be found in the literature [55, 67].



## 2.7. Molecular Docking Study

Many human cancers are result of aberrant signaling and over expression of human epidermal growth factor receptor 2 (HER2) of EGFR family having tyrosine kinase activity. On the cell surface, HER proteins exists as monomers and they dimerizes itself, and with other family members upon binding of ligand to their extracellular domains. The dimer undergoes transphosphorylation in cytoplasm, and aberrantly stimulates downstream secondary messenger pathways like phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K), mitogen-activated protein kinase (MAPK), protein kinase C (PKC) and crosstalk's with other membrane signaling pathways which eventually leads to cell proliferation and oncogenesis [68-70]. Inhibition of kinase activity of these enzymes leads to the induction of cell death in many solid tumors including breast and lung cancers. Molecular docking is an important tool in computer-aided drug design and structural molecular biology to predict the predominant binding mode of a ligand with a known 3D-protein structure. In the present investigation, the energy minimized structures of most active compounds (4, 16, and 27), and crystal structure of the Kinase domain of Human HER2 (erbB2 with PDB ID: 3PP0) were used for molecular docking studies [71]. The docking simulation was carried out using Molegro Virtual Docker (MVD 2013.6.0 for Windows 32 program, Molegro Bioinformatics Solutions, Demark, 2013) [72]. The inbuilt cavity detection method was used to determine the possible active site(s)/cavities for the target protein (3PP0). Among them one active site was selected based on the literature and utilized in further docking studies [73]. Prior to the docking, the MVD software helps in assigning the missing charges, hybridization states, bonds and bond orders of the ligands in the study. The MolDock score (GRID) function was used with a grid resolution of 0.30Å and a binding site radius of 15Å. The MolDockSE searching algorithm was employed, 10 runs using a maximum of 1500 iterations with a maximum population size of 50 was applied. The energy threshold used for the minimized final orientation is 100. The maximum 300 steps with a neighbor distance factor 1 were selected as simplex evolution parameters. The poses are sorted according to MolDock score and rerank score and the top-scoring (most negative, thus favourable to binding) poses are kept after execution of docking. Subsequently, the docking results were visualized using the same MVD software to study the protein and ligand interactions. The docking results are represented in Table 11.

## 3. RESULTS AND DISCUSSION

A 2D-QSAR analysis was performed to explore the structure-activity relationship of *N*-substituted rhodanine derivatives acting as cytotoxic agents against MCF-7 breast cancer cell lines. The data set of 33 molecules was divided into a training set of 23 molecules and a test set of 10 molecules based on structural diversity and activity range. In order to select the predominant descriptors that will affect the cytotoxic activities (lnCyt) of these compounds, stepwise multiple linear regression analysis was performed taking the calculated 2D descriptors as independent variables and lnCyt as dependent variable. The training set molecules was then used to generate the 2D-QSAR models, while the test set molecules were selected for model

validation. The GA-MLR analysis led to the derivation of one final model with three descriptors and obtained the following model-1:

$$\ln\text{Cyt} = -1.33468(+/-0.47126)+0.00062(+/-0.00011)\text{ATSC7v} + 0.64348(+/-0.05625) \text{ geomRadius} + 0.03074(+/-0.00901) \text{ RDF45i}$$

(Model-1)

The Table 2 represents the description about descriptors used in model-1. The correlation between the three descriptors was evaluated using correlation matrix (Table 3) and accepted after considering the correlation values. The highest value was obtained for the RDF45i and ATSC7V descriptors (0.204).

The model-1 is a triparametric equation and has good correlation between the cytotoxic activity and descriptors (ATSC7v, geomRadius and RDF45i) as indicated by the high correlation coefficient  $R$  (0.956). In QSAR regression models, the minimum acceptable  $R^2$  is  $\geq 0.60$  (explains 60% variance in activity). Squared correlation coefficient of model-1 ( $R^2 = 0.913$ ) explains 91% variance in cytotoxic activity of the tested compounds. The difference between  $R^2$  and  $R^2_{\text{adj}}$  is less than 0.3 which indicates the number of descriptors involved in the QSAR model is acceptable. RMSE evaluates the goodness of fit (variability not explained by the regression model) than correlation coefficient ( $R$ ). Very low  $\text{RMSE}_{\text{train}}$  (0.107) of model-1 indicates a perfect fit to the data. The model displayed high  $F$ -test value (66.74) with overall significance level better than 99%, which shows statistical significance of the model in terms of relationship between descriptors and the biological activity. The potential for predictive application of the model was confirmed by using leave-one-out (LOO) cross validation method to ensure the robustness of the model and it is usually denoted as  $Q^2_{\text{LOO}}$ . The minimum acceptable value for cross validated squared correlation coefficient ( $Q^2_{\text{LOO}}$ ) is  $\geq 0.5$  (which indicates 50% variation in predictability) and the  $R^2 - Q^2_{\text{LOO}}$  value should be less than 0.3 (higher value indicates overfitting of the regression model). The high  $Q^2$  value (0.870) and  $R^2 - Q^2_{\text{LOO}}$  value (0.045) of model-1 reflects good internal predictive ability of the QSAR model in the present study. The statistical parameters like  $K_{xx}$  (global correlation descriptors) and  $\Delta K$  (difference between  $K_{xx}$  and  $Y$  responsible variable) is used to detect the overfitting of the QSAR model due to inter-correlation between the descriptors. The low  $k_{xx}$  and  $\Delta K \geq 0.05$  implies no chance correlation between descriptors. The generated model was used to predict the test set data, and the predicted results are given Table 4. The predicted values for  $\ln\text{Cyt}$  for the compounds in the training and test set using the model-1 were plotted against the experimental  $\ln\text{Cyt}$  values in Fig. 2. As can be seen from Table 4 and Fig. 2, the calculated values for the  $\ln\text{Cyt}$  are in good agreement with those of the experimental values. Also, the plot of the residual for the predicted values of  $\ln\text{Cyt}$  for both the training and test sets against the experimental  $\ln\text{Cyt}$  values are shown in Fig. 3. As can be seen the model did not show any proportional and systematic error, because the propagations of the residuals on both sides of zero is random.

### 3.1. QSAR Model Validation

A summary of the results of the internal and external validation metrics for model-1 are presented in Table 6 and 7. The developed QSAR model in the present study was accepted after satisfying the following conditions:  $R^2_{\text{ext}} \geq 0.60$ ,  $RMSE_{cv}$  and  $MAE_{cv}$  close to zero,  $RMSE_{\text{train}} < RMSE_{cv}$ , and  $CCC_{\text{ext}} \geq 0.85$ . The model-1 showed  $r^2_{m(\text{Train})}$ ,  $r^2_{m(\text{Test})}$  values greater than 0.50 and the value of  $R^2 Y_{\text{Scr}}$  is greater than  $Q^2 Y_{\text{Scr}}$ , which shows the good predictive ability of the model. The  $Q^2_{F1}/R^2_{\text{pred}}$  (0.846),  $Q^2_{F2}$  (0.807), and  $Q^2_{F3}$  (0.916) parameters reflects the true model quality. Leave-many-out (LMO) cross validation ( $Q^2_{\text{LMO}}$ ) method was also used to test the real robustness of the developed model. The principle involved in this technique is same as like that of  $Q^2_{\text{LOO}}$  and in this process, the model behavior assessed after a large percentage of compounds (30%) are randomly excluded iteratively from the training set. A new model is built after iteration to predict the biological activity of the removed compounds. It is believed that the average of  $Q^2_{\text{LMO}}$  should be close to  $Q^2_{\text{LOO}}$  value of original model. The model-1 expressed a smallest difference between  $Q^2_{\text{LOO}}$  (0.870) and  $Q^2_{\text{LMO}}$  (0.859).

In order to assess the robustness of the model, the Y-randomization test was applied in this study, and the values shown in the Table 8. The developed model demonstrates significant difference in the average  $R^2_r$  and  $Q^2_r$  values of random models compared to the  $R^2$  and  $Q^2$  values of the original model which indicates that original model was not obtained due to a chance correlation or structural dependency in the proposed model. The  ${}^cR_p^2$  value (0.872) is significantly better than the required threshold value (0.50) consequently the model-1 can be considered as a robust model with both high statistical significance and excellent predictive ability.

The quality of predictions based on the MAE-based criteria for the 100% data can mislead the developed model predictivity. In order to obviate this possibility, the MAE based criteria were determined after removing 5% data of compounds with high absolute error values, and the results were found to be “GOOD” for training, and test sets respectively.

The selected model have passed the evaluation criteria proposed by Golbraikh and Tropsha [53, 61] and confirmed the robustness, and stability of the proposed model.

The applicability domain of the model-1 was evaluated by leverage approach to detect the possible outliers using the Williams plot (Fig. 4). In the present study, the number of descriptors is 3 and number of compounds in the training set is 23, the critical leverage ( $h^*$ ) value is  $[3*(3+1)]/23 = 0.5217$ . The Williams plot clearly demonstrates that all the compounds in the training and test set are inside AD, therefore there is no influential compound in the data set.

Furthermore, the QSAR model was assessed based on the multi-collinearity among the descriptors present in the model-1 by computing the variation inflation factor ( $VIF$ ) using the below equation (7):

$$VIF = (1-R^2)^{-1} \quad (7)$$

In the above equation (7),  $R^2$  represents the correlation coefficient of the regression between variables in the model-1. If  $VIF$  equals to 1, then no inter-correlation exists for each variable;

models with a *VIF* value in the range of 1-5 can be accepted; if *VIF* is greater than 5, the related models are unstable and must be discarded [74]. Table 3 shows the correlation matrix and *VIF* scores of the three descriptors in the model-1. The *VIF* scores are within the acceptable range. Thus, there is no colinearity among the descriptors in the model-1. Therefore, the developed model-1 is considered as stable, statistically significant, robust and predictions of the cytotoxic values for new compounds made with this model are reliable.

### 3.2. Interpretation of Descriptors

By interpreting the descriptors contained in the QSAR model, it is possible to gain some insights into factors, which are related to the cytotoxic activity. It is observed that ATSC7v, geomRadius, and RDF45i are the best descriptors in the establishment of the QSAR model-1 for *N*-substituted rhodanine derivatives. For this reason acceptable interpretation of the selected descriptors is provided below. The brief descriptions of descriptors [75] are shown in Table 2.

ATSC7v (Centered Broto-Moreau autocorrelation - lag 7 / weighted by van der Waals volumes) belongs to the 2D autocorrelation descriptors [76]. The ATSC7v descriptor is a graph invariant describing how the property considered is distributed along the topological structure. It indicates that changing the van der Waals volume seven vertices apart affect the biological activities of the molecules, thus changing the type of atoms or increase in the complexity of vertices that are seven bonds apart will affect the cytotoxic activity of the molecules. The positive regression coefficient of ATSC7v descriptor indicates that it contributes positively to the cytotoxic activity, Which means that van der Waals volume seven vertices apart is favorable to the cytotoxicity. For instance, this feature is clearly observed in the most active compounds, like 4, 16, and 27 in the series as shown in Fig. 5.

geomRadius is a Petitjean shape index descriptor classified as 3D geometrical descriptors. This descriptor accounts for the geometrical radius (minimum geometric eccentricity) of the molecular shape. "The molecular shape is a dynamic property that depends upon energy. The higher the energy the larger the family of nuclear configurations and, consequently, the more molecular shapes are available for the molecule" [77]. The geometric, or three-dimensional (3D), shape of a molecule is known to be of crucial importance in determining its biological properties. As suggested by the model-1 equation, this descriptor contributes positively and corroborates well with the descriptor value and shape of R1 substituents, *i.e.*, as the value of R1 substituent increases, the activity values are found to increase (compounds 4, 16, and 27) and the value of R1 substituent decreases, the activity values are found to decrease (compounds 10, 22, and 33) as demonstrated in Fig. 5.

RDF45i descriptor is a radial distribution function at 4.5 interatomic distance weighted by relative first ionization potential [78]. RDF45i descriptor is an independent of the number of atoms, 3D arrangement of the atoms, and the size of a molecule. This descriptor is an invariant against translation and rotation of the entire molecule. In addition, the RDF descriptors can provide information about specific atom types, distribution of interatomic distances, bond distances, ring types, planar and non-planar systems, and steric hindrance to describe

structure/activity properties of molecules. The positive sign of the coefficient of RDF45i suggests that the cytotoxic activity is directly related to this descriptor. It has been found that the compound 11, 13, and 24 show higher cytotoxic activity as their corresponding RDF45i value increases while 5-benzidene compounds (9, 21, and 32) show low cytotoxic activity due to less descriptor value (see Fig. 6).

### 3.3. Molecular Docking Studies

During the initial process of the study, the internal ligand (03Q), and Erlotinib were first docked with the binding site of the receptor, and the same binding site with same dimensions of the cavity were used to compare the docking results of the compounds 4, 16, and 27. The 3D and 2D binding modes of docked compounds is shown in Fig. 7 and 8. The molecular docking study suggests that the internal ligand (03Q) interacted with the active site amino acid residues (Met801, Thr862, and Asp863) through hydrogen bonding interactions and also makes hydrophobic interactions with the side chains of Thr862, Glu770, Met774, Ser783, Leu785, Leu790, Leu796, and Phe864 [79]. The commercially available anti-cancer drug “Erlotinib” displayed hydrogen bonding interactions with Asp863 and Thr862 in the active site. The 3D and 2D binding modes of internal ligand (03Q), and reference drug “Erlotinib” is shown in Fig. 8. Our study revealed that among the docked compounds, the best docking energy for HER2 was exhibited by internal ligand (03Q) with a moldock score -178.77 followed by Erlotinib (-135.91), compound 27 (-131.91), compound 4 (-125.530), and compound 16 (-118.61). The compounds 4, 16, and 27 interacted with active site amino acid residues (Thr862, Asp863) through hydrogen bonding interactions and also shows hydrophobic interactions with the side chains of Leu796, Met774, Leu785 for compound 4, Leu796, Met774 for compound 16, and Leu796, Phe1004 for compound 27 respectively.

## CONCLUSION

In the present study, we have attempted to develop MLR based 2D-QSAR model from 33 diverse compounds having defined cytotoxic activity to investigate the molecular properties essential for the cytotoxic activity. The MLR based QSAR model was developed with simple, meaningful and easily interpretable descriptors such as ATSC7v, geomRadius, and RDF45i. The statistical metrics of the developed model showed that the model is robust, statistically significant with good predictivity based on both internal and external validation parameters. Molecular docking results further revealed that compounds (4, 16, and 27) showed interaction with the active residues (Met801, Thr862, and Asp863) in the binding site of kinase domain of HER2 receptor like the internal ligand (03Q) and Erlotinib. The Moldock scores of the active compounds (4, 16, and 27) were comparable with internal ligand (03Q) as well as with anti-cancer drug (Erlotinib). In addition, the study of molecular interactions involved in binding is very important along with Moldock score to understand the binding mode pattern of any compound in the binding site. Furthermore, the developed 2D-QSAR model can be used as good

model for prediction of activity of more potent analogs of *N*-substituted rhodanines as cytotoxic agents even before their synthesis and evaluation.

## CONFLICT OF INTEREST

The authors declare that no conflict of interest, financial or otherwise.

## REFERENCES

1. Mattiuzzi C, Lippi G. Current Cancer Epidemiology. *J Epidemiol Glob Health*, (2019), 9(4), 217–222.  
<https://doi.org/10.2991/jegh.k.191008.001>
2. World Health Organization, WHO report on cancer: setting priorities, investing wisely and providing care for all. Geneva, Switzerland, 2020.
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J Clin*, (2018), 68(6), 394–424. <https://doi.org/10.3322/caac.21492>.
4. DeSantis Carol E, Ma Jiemin Gaudet, Mia M, Newman Lisa A, Miller Kimberly D, Goding Sauer A, Jemal A, Siegel RL. Breast cancer statistics, 2019. *CA A Cancer J Clinicians*, (2019), 69(6), 438-451.  
<https://doi.org/10.3322/caac.21583>
5. Azubuike SO, Muirhead C, Hayes L, McNally R. Rising global burden of breast cancer: the case if sub-Saharan Africa (with emphasis on Nigeria) and implications for regional development: a review. *World J Surg Oncol*, (2018), 16(1), 63.  
<https://doi.org/10.1186/s12957-018-1345-2>
6. Bray F, Ferlay J, Laversanne M, Brewster DH, Gombe Mbalawa C, Kohler B, Piñeros M, Steliarova-Foucher E, Swaminathan R, Antoni S, Soerjomataram I, Forman D. Cancer Incidence in Five Continents: Inclusion criteria, highlights from Volume X and the global status of cancer registration. *Int. J. Cancer*, (2015), 137(9), 2060–2071.  
<https://doi.org/10.1002/ijc.29670>
7. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, Ruddy K, Tsang J, Cardoso F. Breast cancer. *Nat. Rev. Dis. Primers*, (2019), 5(1), 66.  
<https://doi.org/10.1038/s41572-019-0111-2>
8. Cohen P. Protein kinases-the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov*, (2002), 1(4), 309–315.  
<https://doi.org/10.1038/nrd773>
9. Roskoski R. ErbB/HER protein-tyrosine kinases: Structures and small molecule inhibitors. *Pharmacol Res*, (2014), 87, 42–59.  
<https://doi.org/10.1016/j.phrs.2014.06.001>
10. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol*, (2001), 2(2), 127–137.  
<https://doi.org/10.1038/35052073>

11. Iqbal N, Iqbal N. Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and Therapeutic Implications. *Mol Biol Int*, (2014), 852748. <https://doi.org/10.1155/2014/852748>
12. Wu P, Nielsen TE, Clausen MH. Small-molecule kinase inhibitors: an analysis of FDA approved drugs. *Drug Discov. Today*, (2016), 21(1), 5–10. <https://doi.org/10.1016/j.drudis.2015.07.008>
13. Pardasani RT, Pardasani P, Sherry D, Chaturvedi V. Synthetic and antibacterial studies of rhodanine derivatives with indol-2,3-diones, *Indian J. Chem*, (2001), 40B, 1275-1278.
14. Zervosen A, Lu WP, Chen Z, White RE, Demuth TP, Frère JM. Interactions between penicillin-binding proteins (PBPs) and two novel classes of PBP inhibitors, arylalkylidene rhodanines and arylalkylidene iminothiazolidin-4-ones. *Antimicrob agents chemother*, (2004), 48(3), 961–969. <https://doi.org/10.1128/aac.48.3.961-969.2004>
15. Habib NS, Rida SM, Badawey EAM, Fahmy HTY, Ghazlan HA. Synthesis and antimicrobial activity of rhodanine derivatives. *Eur. J. Med. Chem*, (1997), 32(9), 759-762. [https://doi.org/10.1016/S0223-5234\(97\)88919-2](https://doi.org/10.1016/S0223-5234(97)88919-2)
16. Dayam R, Sanchez T, Clement O, Shoemaker R, Sei S, Neamati N. Beta-diketo acid pharmacophore hypothesis. 1. Discovery of a novel class of HIV-1 integrase inhibitors. *J Med Chem*, (2005), 48(1), 111–120. <https://doi.org/10.1021/jm0496077>
17. Terashima H, Hama K, Yamamoto R, Tsuboshima M, Kikkawa R, Hatanaka I, Shigeta Y. Effects of a new aldose reductase inhibitor on various tissues in vitro. *J. Pharmacol. Exp. Ther*, (1984), 229(1), 226–230.
18. Ravi S, Chiruvella KK, Rajesh K, Prabhu V, Raghavan SC. 5-Isopropylidene-3-ethyl rhodanine induce growth inhibition followed by apoptosis in leukemia cells. *Eur. J. Med. Chem*, (2010), 45(7), 2748–2752. <https://doi.org/10.1016/j.ejmech.2010.02.054>
19. Powers JP, Piper DE, Li Y, Mayorga V, Anzola J, Chen JM, Jaen JC, Lee G, Liu J, Peterson MG, Tonn GR, Ye Q, Walker NP, Wang Z. SAR and mode of action of novel non-nucleoside inhibitors of hepatitis C NS5b RNA polymerase. *J Med Chem*, (2006), 49(3), 1034–1046. <https://doi.org/10.1021/jm050859x>.
20. Sudo K, Matsumoto Y, Matsushima M, Fujiwara M, Konno K, Shimotohno K, Shigeta S, Yokota T. Novel hepatitis C virus protease inhibitors: thiazolidine derivatives. *Biochem. Biophys. Res. Commun*, (1997), 238(2), 643–647. <https://doi.org/10.1006/bbrc.1997.7358>
21. Fresneau P, Cussac M, Morand JM, Szymonski B, Tranqui D, Leclerc G. Synthesis, activity, and molecular modeling of new 2, 4-dioxo-5-(naphthylmethylene)-3-thiazolidineacetic acids and 2-thioxo analogues as potent aldose reductase inhibitors. *J Med Chem*, (1998), 41(24), 4706–4715. <https://doi.org/10.1021/jm9801399>
22. Grant EB, Guiadeen D, Baum EZ, Foleno BD, Jin H, Montenegro DA, Nelson EA, Bush K, Hlasta DJ. The synthesis and SAR of rhodanines as novel class C beta-lactamase

- inhibitors. *Bioorg. Med. Chem. Lett.*, (2000), 10(19), 2179–2182. [https://doi.org/10.1016/s0960-894x\(00\)00444-3](https://doi.org/10.1016/s0960-894x(00)00444-3)
23. Soltero-Higgin M, Carlson EE, Phillips JH, Kiessling LL. Identification of inhibitors for UDP-galactopyranose mutase. *J. Am. Chem. Soc.*, (2004), 126(34), 10532–10533. <https://doi.org/10.1021/ja048017v>
24. Orchard MG, Neuss JC, Galley CM, Carr A, Porter DW, Smith P, Scopes DI, Haydon D, Vousden K, Stubberfield CR, Young K, Page M. Rhodanine-3-acetic acid derivatives as inhibitors of fungal protein mannosyl transferase (PMT1). *Bioorg. Med. Chem. Lett.*, (2004), 14(15), 3975–3978. <https://doi.org/10.1016/j.bmcl.2004.05.050>
25. Whitesitt CA, Simon RL, Reel JK, Sigmund SK, Phillips ML, Shadle JK, Berry D. Synthesis and structure-activity relationships of benzophenones as inhibitors of cathepsin D. *Bioorg. Med. Chem. Lett.*, (1996), 6(18), 2157-2162. [https://doi.org/10.1016/0960-894X\(96\)00393-9](https://doi.org/10.1016/0960-894X(96)00393-9)
26. Forino M, Johnson S, Wong TY, Rozanov DV, Savinov AY, Li W, Abagyan RA. Efficient synthetic inhibitors of anthrax lethal factor. *Proc. Natl. Acad. Sci. U.S.A.*, (2005), 102(27), 9499-9504. <https://doi.org/10.1073/pnas.0502733102>
27. Free CA, Majchrowicz E, Hess SM. Mechanism of inhibition of histidine decarboxylase by rhodanines. *Biochem. Pharmacol.*, (1971), 20(7), 1421-1428.
28. Cutshall NS, O'Day C, Prezhdo M. Rhodanine derivatives as inhibitors of JSP-1. *Bioorg. Med. Chem. Lett.*, (2005), 15(14), 3374-3379. <https://doi.org/10.1016/j.bmcl.2005.05.034>
29. Irvine MW, Patrick GL, Kewney J, Hastings SF, MacKenzie SJ. Rhodanine derivatives as novel inhibitors of PDE4. *Bioorg. Med. Chem. Lett.*, (2008), 18(6), 2032-2037. <https://doi.org/10.1016/j.bmcl.2008.01.117>
30. Knight SD, Adams ND, Burgess JL, Chaudhari AM., Darcy MG, Donatelli CA, Luengo JI, Newlander KA, Parrish CA, Ridgers LH, Sarpong MA, Schmidt SJ, Van Aller GS, Carson JD, Diamond MA, Elkins PA, Gardiner CM, Garver E, Gilbert SA, Gontarek RR, Dhanak D. Discovery of GSK2126458, a Highly Potent Inhibitor of PI3K and the Mammalian Target of Rapamycin. *ACS Med. Chem. Lett.*, (2010), 1(1), 39–43. <https://doi.org/10.1021/ml900028r>
31. Bei S, Li F, Li H, Li J, Zhang X, Sun Q, Feng L. Inhibiting of gastric cancer cell growth by a PI3K-mTor dual inhibitor GSK1059615. *Biochem Biophys Res Commun.*, (2019), 511(1), 13-20. <https://doi.org/10.1016/j.bbrc.2019.02.032>
32. Liu K, Rao W, Parikh H, Li Q, Guo TL, Grant S, Zhang S. 3, 5-Disubstituted-thiazolidine-2, 4-dione analogs as anticancer agents: design, synthesis and biological characterization. *Eur. J. Med. Chem.*, (2012), 47(1), 125-137. <https://doi.org/10.1016/j.ejmech.2011.10.031>



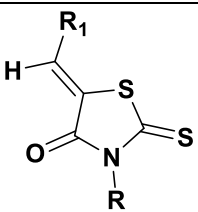
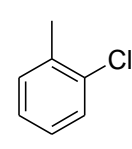
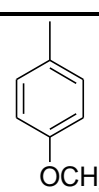
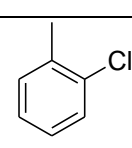
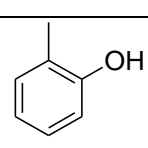
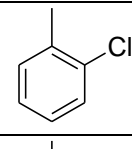
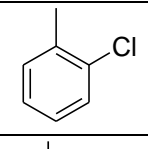
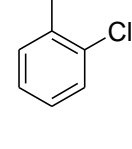
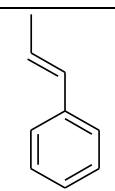
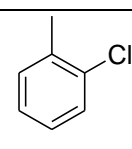
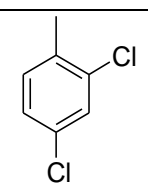
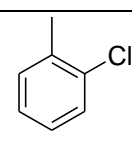
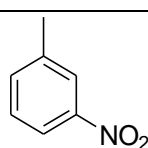
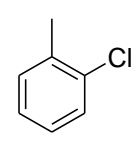
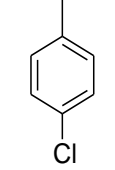
33. Smith AG, Beaumont KA, Smit DJ, Thurber AE, Cook AL, Boyle GM, Muscat GE. PPAR $\gamma$  agonists attenuate proliferation and modulate Wnt/ $\beta$ -catenin signalling in melanoma cells. *Int. J. Biochem. Cell Biol*, (2009), 41(4), 844-852.  
<https://doi.org/10.1016/j.biocel.2008.08.037>
34. Dai Y, Wang WH. Peroxisome proliferator-activated receptor  $\gamma$  and colorectal cancer. *World J Gastrointest Oncol*, (2010), 2(3), 159.  
<https://doi:10.4251/wjgo.v2.i3.159>
35. Tsujie M, Nakamori S, Okami J, Hayashi N, Hiraoka N, Nagano H, Monden M. Thiazolidinediones inhibit growth of gastrointestinal, biliary, and pancreatic adenocarcinoma cells through activation of the peroxisome proliferator-activated receptor  $\gamma$ /retinoid X receptor  $\alpha$  pathway. *Experimental cell research*, (2003), 289(1), 143-151.  
[https://doi.org/10.1016/S0014-4827\(03\)00263-5](https://doi.org/10.1016/S0014-4827(03)00263-5)
36. Asati V, Mahapatra DK, Bharti SK. Thiazolidine-2, 4-diones as multi-targeted scaffold in medicinal chemistry: Potential anticancer agents. *Eur. J. Med. Chem*, (2014), 87, 814-833.  
<https://doi.org/10.1016/j.ejmech.2014.10.025>
37. Jorgensen WL. The many roles of computation in drug discovery. *Science*, (2004), 303(5665), 1813-1818.  
<https://doi.org/10.1126/science.1096361>
38. Xiang M, Cao Y, Fan W, Chen L, Mo Y. Computer-aided drug design: lead discovery and optimization. *Comb Chem High Throughput Screen*, (2012), 15(4), 328-337.  
<https://doi.org/10.2174/138620712799361825>
39. Hansch C, Kurup A, Garg R, Gao H. Chem-bioinformatics and QSAR: a review of QSAR lacking positive hydrophobic terms. *Chem. Rev*, (2001), 101(3), 619-672.  
<https://doi.org/10.1021/cr0000067>
40. Mandal SP, Garg A, Sahetya SS, Nagendra SR, Sripath HS, Manjunath MM, Kumar BP. Novel rhodanines with anticancer activity: design, synthesis and CoMSIA study. *RSC Adv*, (2016), 6(63), 58641-58653.  
<https://doi.org/10.1039/C6RA08785J>
41. Mendelsohn LD. ChemDraw 8 ultra, windows and macintosh versions. *J. Chem. Inf. Model*, (2004), 44(6), 2225-2226.  
<https://doi.org/10.1021/ci040123t>
42. Yap CW. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem*, (2011), 32(7), 1466-1474.  
<https://doi.org/10.1002/jcc.21707>
43. The simple, user-friendly and reliable online standalone tools freely available at [http://teqip.jdvu.ac.in/QSAR\\_Tools/](http://teqip.jdvu.ac.in/QSAR_Tools/) and <http://dtclab.webs.com/software-tools>.
44. Gramatica P, Chirico N, Papa E, Kovarich S, Cassani S. QSARINS, software for QSAR MLR model development and validation, from <http://www.qsar.it>.

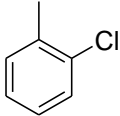
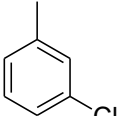
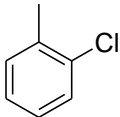
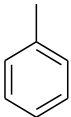
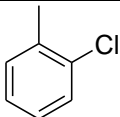
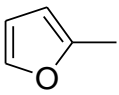
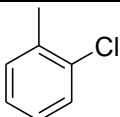
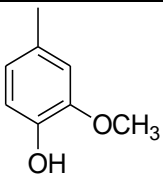
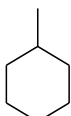
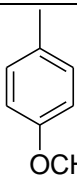
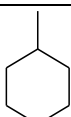
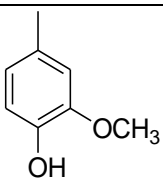
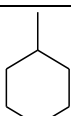
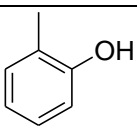
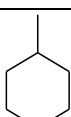
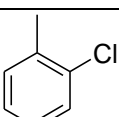
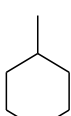
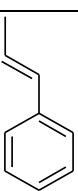
45. Gramatica P, Chirico N, Papa E, Kovarich S, Cassani S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem*, (2013), 34(24), 2121-2132.  
<https://doi.org/10.1002/jcc.23361>
46. Gramatica P, Cassani S, Chirico N. QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. *J. Comput. Chem*, (2014), 35(13), 1036-1044.  
<https://doi.org/10.1002/jcc.23576>
47. Holland JH. *Adaptation in Natural and Artificial Systems*, 2<sup>nd</sup> ed.; The University of Michigan Press: Boston, (1992).
48. Friedman JH. Multivariate adaptive regression splines. *Ann Stat*, (1991), 19, 1-67.
49. Chukhrova N, Johannssen A. Fuzzy regression analysis: systematic review and bibliography. *Applied Soft Computing*, (2019), 84, 105708.  
<https://doi.org/10.1016/j.asoc.2019.105708>
50. Srinivas M, Neharika K. QSAR Analysis of Quinazolinyl-arylurea Derivatives as Potential Anti-Cancer Agents: GA-MLR Chemometric Approach. *Current Chinese Chemistry*, (2021), 1(2), e040821191622 .  
<https://dx.doi.org/10.2174/2666001601666210222090518>
51. Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform*, (2010), 29(6-7), 476-488.  
<https://doi.org/10.1002/minf.201000061>
52. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK. Validation of QSAR models-strategies and importance. *Int. J. Drug Des. Discov*, (2011), 3, 511-519.
53. Golbraikh A, Tropsha A. Beware of q<sup>2</sup>!. *J Mol Graph Model*, (2002), 20(4), 269-276.  
[https://doi:10.1016/s1093-3263\(01\)00123-1](https://doi:10.1016/s1093-3263(01)00123-1)
54. Chirico N, Gramatica P. Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *J Chem Inf Model*, (2011), 51(9), 2320–2335.  
<https://doi.org/10.1021/ci200211n>
55. Roy PP, Roy K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci*, (2008), 27(3), 302-313.  
<https://doi.org/10.1002/qsar.200710043>
56. Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN. Some case studies on application of "r(m)<sup>2</sup>" metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. *J. Comput. Chem*, (2013), 34(12), 1071–1082.  
<https://doi.org/10.1002/jcc.23231>
57. Pratim RP, Paul S, Mitra I, Roy K. On two novel parameters for validation of predictive QSAR models. *Molecules* (Basel, Switzerland), (2009), 14(5), 1660–1701.  
<https://doi.org/10.3390/molecules14051660>
58. Ojha PK, Mitra I, Das RN, Roy K. Further exploring rm<sup>2</sup> metrics for validation of QSPR models, *Chemom. Intell. Lab. Syst*, (2011), 107(1), 194-205.

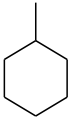
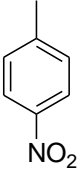
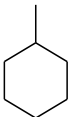
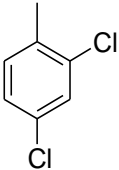
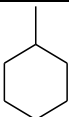
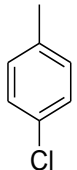
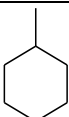
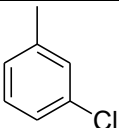
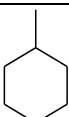
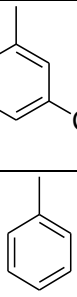
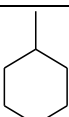
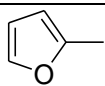
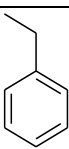
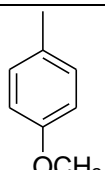
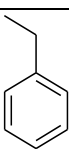
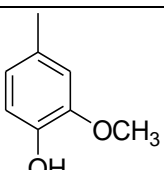
- <https://doi.org/10.1016/j.chemolab.2011.03.011>
59. Schüürmann G, Ebert RU, Chen J, Wang B, Kühne R. External validation and prediction employing the predictive squared correlation coefficient test set activity mean vs training set activity mean. *J. Chem. Inf. Model*, (2008), 48(11), 2140–2145.  
<https://doi.org/10.1021/ci800253u>
60. Mitra I, Saha A, Roy K. Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. *Molecular Simulation*, (2010), 36(13), 1067-1079.  
<https://doi.org/10.1080/08927022.2010.503326>
61. Oloff S, Mailman RB, Tropsha A. Application of validated QSAR models of D1 dopaminergic antagonists for database mining. *J Med Chem*, (2005), 48(23), 7322–7332.  
<https://doi.org/10.1021/jm049116m>
62. Roy K, Das RN, Ambure P, Aher RB. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemom. Intell. Lab. Syst*, (2016), 152, 18-33.  
<https://doi.org/10.1016/j.chemolab.2016.01.008>
63. Ambure P, Roy K. Understanding the structural requirements of cyclic sulfone hydroxyethylamines as hBACE1 inhibitors against A $\beta$  plaques in Alzheimer's disease: a predictive QSAR approach. *RSC Adv*, (2016), 6(34), 28171-28186.  
<https://doi.org/10.1039/C6RA04104C>
64. Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci*, (2003), 22(1), 69-77.  
<https://doi.org/10.1002/qsar.200390007>
65. Gramatica P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci*, (2007), 26(5), 694-701.  
<https://doi.org/10.1002/qsar.200610151>
66. 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.  
<https://doi.org/10.1016/j.chemolab.2015.04.013>
67. Kiralj R, Ferreira MMC. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *J.Braz. Chem.Soc*, (2009), 20, 770-787.
68. Moasser MM. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene*, (2007), 26(45), 6469–6487.  
<https://doi.org/10.1038/sj.onc.1210477>
69. Gutierrez C, Schiff R. HER2: biology, detection, and clinical implications. *Arch Pathol Lab Med*, (2011), 135(1), 55–62.  
<https://doi.org/10.1043/2010-0454-RAR.1>
70. Wang J, Xu B. Targeted therapeutic options and future perspectives for HER2-positive breast cancer. *Sig transduct target ther*, (2019), 4, 34.  
<https://doi.org/10.1038/s41392-019-0069-2>

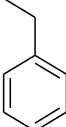
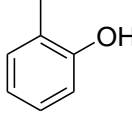
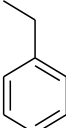
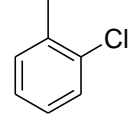
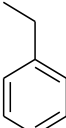
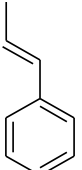
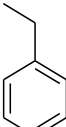
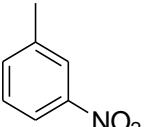
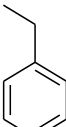
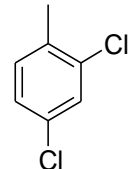
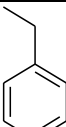
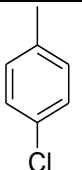
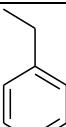
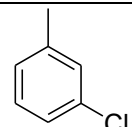
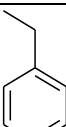
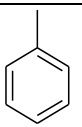
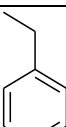
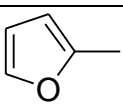
71. Crystal Structure of the Kinase domain of Human HER2 (erbB2); (PDB ID: 3PP0), <https://www.rcsb.org/structure/3pp0> (Accessed on 15/05/2022).
72. Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. *J Med Chem*, (2006), 49(11), 3315–3321. <https://doi.org/10.1021/jm051197e>
73. Misra A, Kishore D, Verma VP, Dubey S, Chander S, Gupta N, Bhagyawant S, Dwivedi J, Alothman ZA, Wabaidur SM, Sharma S. Synthesis, biological evaluation and molecular docking of pyrimidine and quinazoline derivatives of 1,5-benzodiazepine as potential anticancer agents. *Journal of King Saud University-Science*, (2020), 32(2), 1486-95. <https://doi.org/10.1016/j.jksus.2019.12.002>
74. Adhikari N, Halder AK, Saha A, Das SK, Jha T. Structural findings of phenyl indoles as cytotoxic antimitotic agents in human breast cancer cell lines through multiple validated QSAR studies. *Toxicol In Vitro*, (2015), 29(7), 1392–1404. <https://doi.org/10.1016/j.tiv.2015.05.017>
75. Todeschini R, Consonni V. *Handbook of molecular descriptors*, John Wiley Sons, (2008).
76. Alisi IO, Uzairu A, Abechi SE, Idris SO. Evaluation of the antioxidant properties of curcumin derivatives by genetic function algorithm. *J Adv Res*, (2018), 12, 47–54. <https://doi.org/10.1016/j.jare.2018.03.003>
77. Bath PA, Poirrette AR, Willett P, Allen FH. The extent of the relationship between the graph-theoretical and the geometrical shape coefficients of chemical compounds. *J. Chem. Inf. Model*, (1995), 35(4), 714-716. <https://doi.org/10.1021/ci00026a007>
78. González MP, Terán C, Tejeira M, Helguera AM. Radial distribution function descriptors: an alternative for predicting A2 A adenosine receptors agonists. *Eur. J. Med. Chem*, (2006), 41(1), 56-62. <https://doi.org/10.1016/j.ejmech.2005.08.004>
79. Aertgeerts K, Skene R, Yano J, Sang BC, Zou H, Snell G, Jennings A, Iwamoto K, Habuka N, Hirokawa A, Ishikawa T, Tanaka T, Miki H, Ohta Y, Sogabe S. Structural analysis of the mechanism of inhibition and allosteric activation of the kinase domain of HER2 protein. *J. Biol. Chem*, (2011), 286(21), 18756–18765. <https://doi.org/10.1074/jbc.M110.206193>.

**Table 1. Structure and cytotoxic activities of *N*-substituted rhodanine derivatives.**

				
Compound No	R	R1	% Cytotoxicity	InCyt
1			62	4.127
2			43	3.761
3			34	3.526
4			81	4.393
5			47	3.85
6			40	3.688
7			45	3.806

8			40	3.688
9			38	3.637
10			23	3.135
11			61	4.11
12			55	4.007
13			58	4.06
14			39	3.663
15			31	3.43
16			77	4.343

17			34	3.526
18			43	3.761
19			41	3.713
20			39	3.663
21			28	3.33
22			20	2.995
23			50	3.912
24			55	4.007

25			35	3.555
26			32	3.465
27			71	4.262
28			29	3.367
29			40	3.688
30			30	3.401
31			27	3.295
32			32	3.465
33			28	3.233



**Table 2. List of descriptors used in QSAR model.**

S.No	Descriptors symbols	Name of the Descriptor(s)	Class
1	ATSC7v	Centered Broto-Moreau autocorrelation - lag 7 / weighted by van der Waals volumes	2D
2	geomRadius	Geometrical radius (minimum geometric eccentricity)	3D
3	RDF45i	Radial distribution function - 045 / weighted by relative first ionization potential	3D

**Table 3. Correlation matrix and Variation inflation factor (VIF) of the descriptors.**

	ATSC7v	geomRadius	RDF45i	VIF
ATSC7v	1			1.06
geomRadius	0.113	1		1.02
RDF45i	0.204	-0.024	1	1.05

**Table 4. Observed and predicted activity of N-substituted rhodanine derivatives.**

Comp No	Observed Activity	Predicted Activity	Residual
1*	4.127	4.256	-0.129
2	3.761	3.753	0.008
3*	3.526	3.696	-0.170
4	4.393	4.378	0.015
5	3.85	3.859	-0.009
6	3.688	3.583	0.105
7	3.806	3.707	0.099
8*	3.688	3.605	0.083
9	3.637	3.536	0.101
10	3.135	3.366	-0.231
11*	4.11	4.168	-0.058
12*	4.007	4.078	-0.071
13	4.06	4.019	0.041
14	3.663	3.581	0.082
15	3.43	3.597	-0.167

16	4.343	4.181	0.162
17*	3.526	3.529	-0.003
18*	3.761	3.780	-0.019
19*	3.713	3.607	0.106
20	3.663	3.531	0.132
21	3.33	3.425	-0.095
22	2.995	2.892	0.103
23*	3.912	3.865	0.047
24	4.007	4.066	-0.059
25	3.555	3.558	-0.003
26	3.465	3.565	-0.100
27	4.262	4.344	-0.082
28	3.367	3.429	-0.062
29	3.688	3.724	-0.036
30*	3.401	3.594	-0.193
31	3.295	3.484	-0.189
32	3.465	3.323	0.142
33	3.233	3.192	0.041

\*Represents the test compounds

**Table 5. Statistical Fitting Parameters**

S.No.	Parameter	Model-1
1	$R$	0.956
2	$R^2$	0.913
3	$R^2_{adj}$	0.899
4	$SEE$	0.119
5	$PRESS$	0.267
6	$S_{DEP}$	0.135
7	$F$	66.74(DF:3,19)
8	$LOF$	0.021
9	$K_{xx}$	0.0472
10	$\Delta K$	0.318
11	$RMSE_{train}$	0.1078
12	$CCC_{train}$	0.954

**Table 6. Internal Validation Parameters**

S.No.	Parameter	Model-1
1	$Q^2_{\text{LOO}}$	0.870
2	$Q^2_{\text{LMO}}$	0.857
3	$R^2_{\text{Y}_{\text{scr}}}$	0.135
4	$Q^2_{\text{Y}_{\text{scr}}}$	-0.288
5	$RMSE_{\text{cv}}$	0.132
6	$PRESS_{\text{cv}}$	0.402
7	$CCC_{\text{cv}}$	0.932
8	$\text{Avg } r^2_{\text{m}}(\text{Train})$	0.819
9	$\Delta r^2_{\text{m}}(\text{Train})$	0.047

**Table 7. External Validation Parameters**

S.No.	Parameter	Model-1
1	$R^2_{\text{ext}}$	0.848
2	$Q^2_{\text{F1}}$	0.846
3	$Q^2_{\text{F2}}$	0.807
4	$Q^2_{\text{F3}}$	0.916
5	$RMSE_{\text{ext}}$	0.106
6	$PRESS_{\text{ext}}$	0.112
7	$CCC_{\text{ext}}$	0.908
8	$r^2_0$	0.838
9	$r'^2_0$	0.845
10	$k$	0.988
11	$k'$	1.01
12	$\text{Avg } r^2_{\text{m}}(\text{Test})$	0.790
13	$\Delta r^2_{\text{m}}(\text{Test})$	0.099

**Table 8. Y-randomization study of the Model-1**

No of Y-randomization	$R_r$	$R^2_r$	$Q^2_r$
1	0.494	0.244	-0.187
2	0.171	0.029	-0.314
3	0.177	0.031	-0.433
4	0.483	0.233	-0.134
5	0.501	0.251	-0.201
6	0.333	0.111	-0.277

7	0.200	0.040	-0.444
8	0.181	0.033	-0.286
9	0.176	0.031	-0.401
10	0.133	0.018	-0.441
Random Models Parameters			
Average $R_r$	0.346		
Average $R_r^2$	0.176		
Average $Q_r^2$	-0.204		
${}^cR_p^2$	0.872		

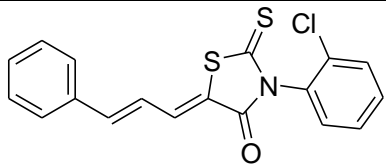
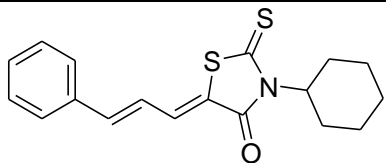
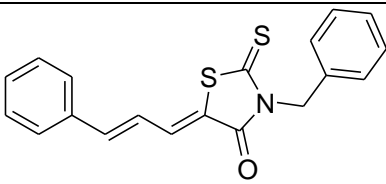
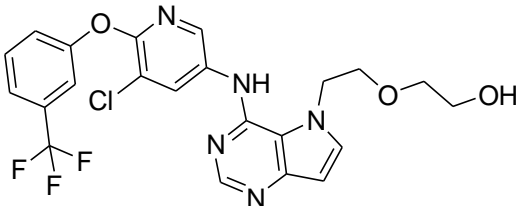
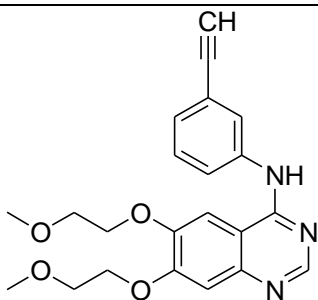
**Table 9. Golbraikh and Tropsha acceptable model validation parameters and their threshold values of the Model-1**

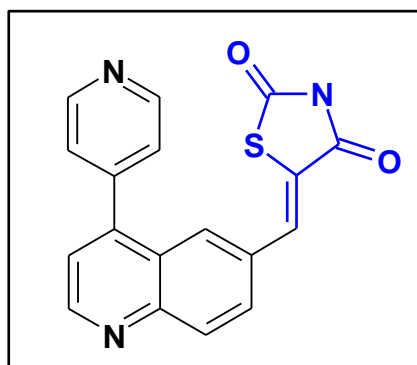
Parameter	Formula	Threshold Score	Model Score	Comment
$Q^2$	$Q^2 = 1 - \left[ \frac{\sum(Y_{obs} - Y_{pred})^2}{\sum(Y_{obs} - Y_{train})^2} \right]$	$Q^2 > 0.5$	0.870	Passed
$R^2_{train}$	Coefficient of determination for the plot of predicted <i>versus</i> observed for the training set by MLR	$R^2_{train} > 0.6$	0.913	Passed
$R^2_{test}$	Coefficient of determination for the plot of predicted <i>versus</i> observed for the test set by MLR	$R^2_{test} > 0.6$	0.848	Passed
$r^2_0$	$r^2$ at zero intercept		0.838	
$r'^2_0$	$r^2$ for the plot of observed <i>versus</i> predicted activity for the test set at zero intercept		0.845	
$ r^2_0 - r'^2_0 $		$ r^2_0 - r'^2_0  < 0.3$	-0.007	Passed
$k$	Slope of the plot of observed activity against calculated activity values at zero intercept	$0.85 < K < 1.15$	0.988	Passed
$k'$	Slope of the plot of calculated activity against observed activity at zero intercept	$0.85 < k' < 1.15$	1.01	Passed
$(r^2 - r^2_0)/r^2$		$(r^2 - r^2_0)/r^2 < 0.1$	0.01211	Passed
$(r^2 - r'^2_0)/r^2$		$(r^2 - r'^2_0)/r^2 < 0.1$	0.00372	Passed

**Table 10. Error-based metrics and prediction quality for the training and test set employed in Model-1**

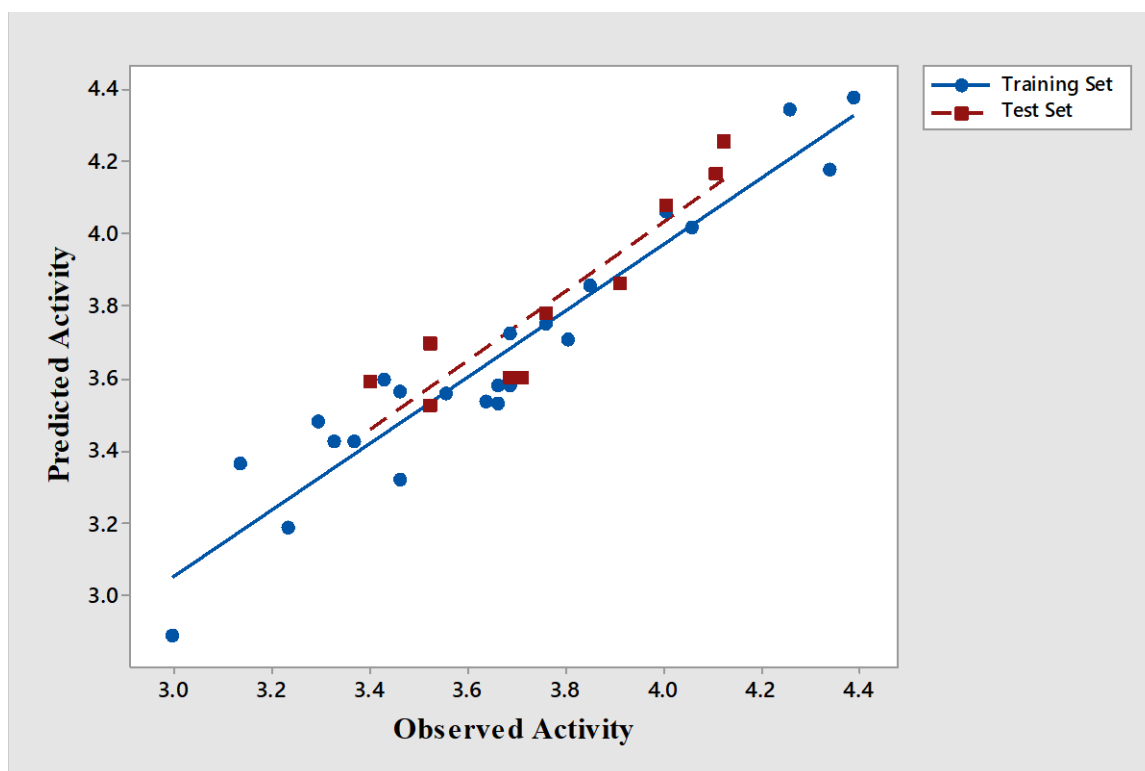
Set	MAE (100%)	$\sigma_{AE}$ (100%)	MAE (95%)	$\sigma_{AE}$ (95%)	MAE + 3* $\sigma_{AE}$ (95%)	Prediction quality
Training	0.0898	0.0610	0.0783	0.0497	0.2274	GOOD
Test	0.0880	0.0622	0.0763	0.0530	0.2353	GOOD

**Table 11. Molecular docking results of compounds (4,16, and 27), internal ligand (03Q), and reference drug (Erlotinib)**

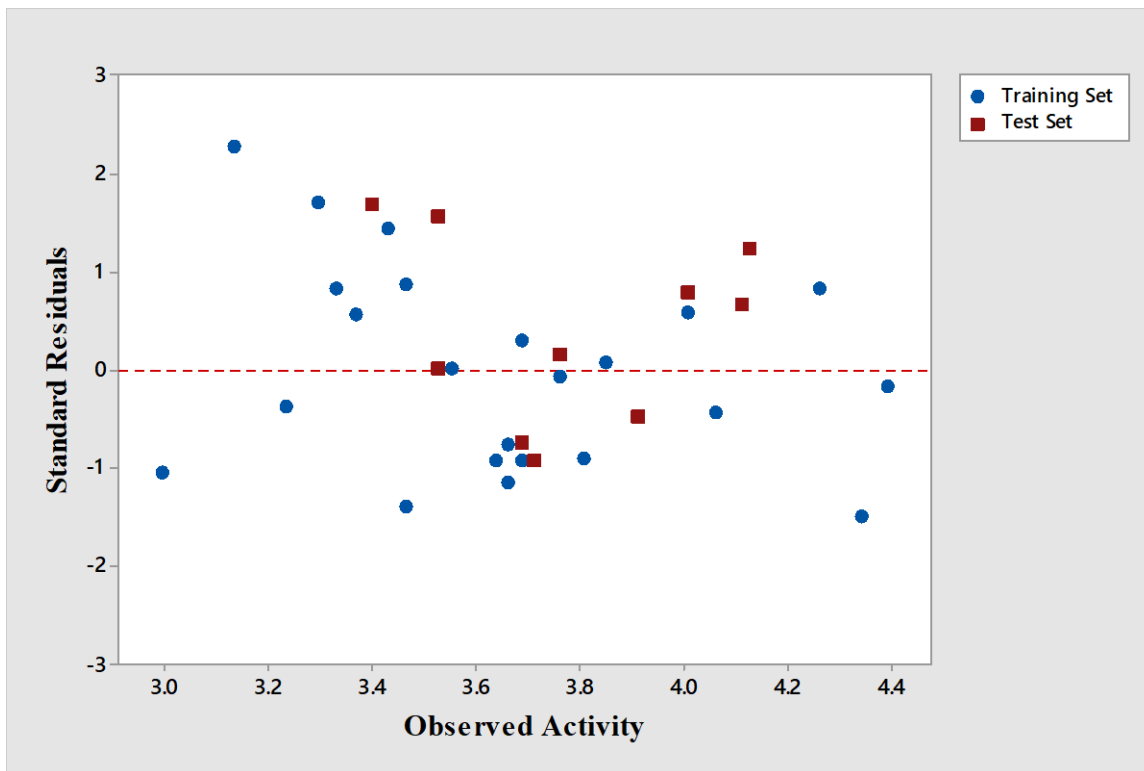
Name	Structure	MolDock Score (kcal/Mol)	Rerank Score (kcal/Mol)
Compound 4		-125.53	-99.58
Compound 16		-118.61	-99.46
Compound 27		-131.91	-111.95
Internal Ligand (03Q)		-178.77	-147.62
Reference Drug (Erlotinib)		-135.31	-101.77



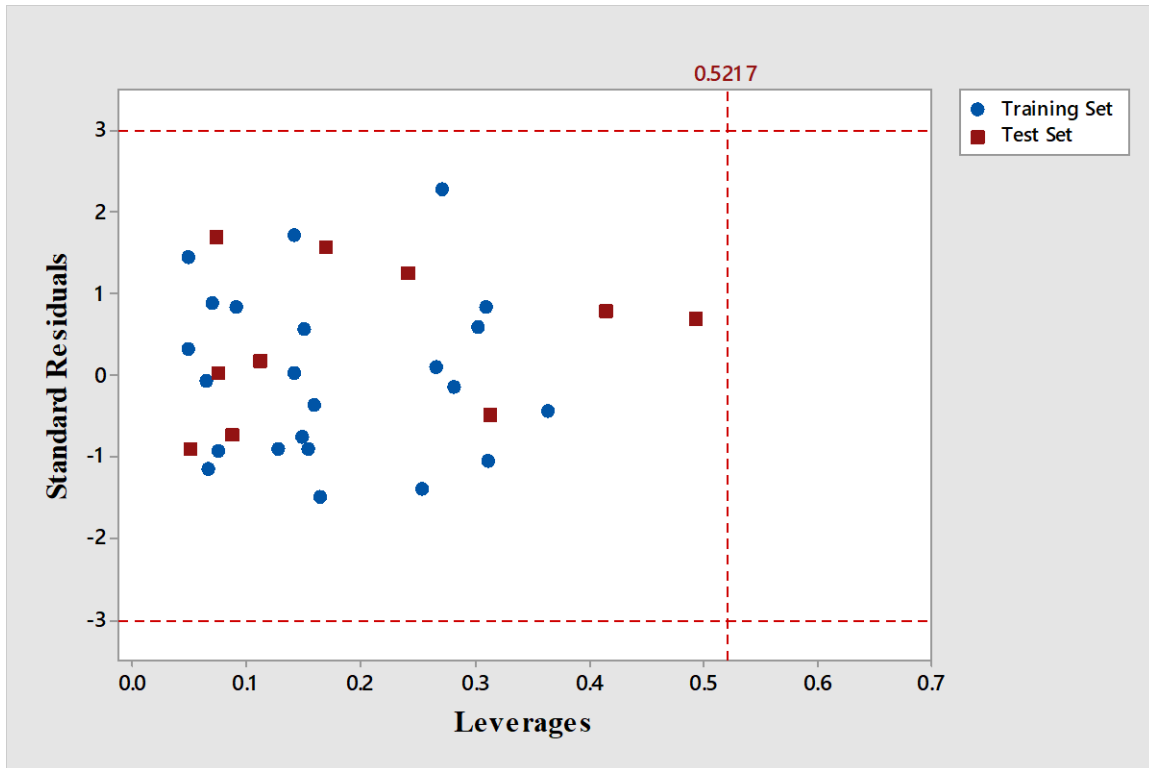
**Fig. (1).** Structure of GSK1059615



**Fig. (2).** Scatter plot of observed vs. predicted activity of training set (blue) and test set (red)

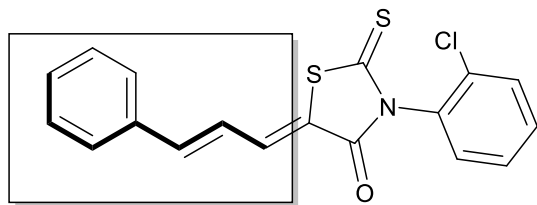


**Fig. (3).** The standard residuals vs. observed activity for the training (blue) and test set (red)

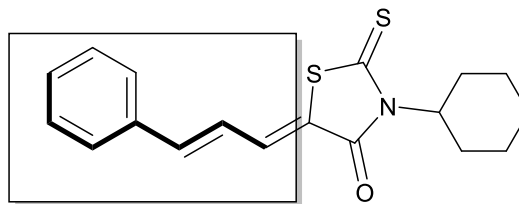


**Fig. (4).** The scatter plot of standard residuals vs leverages (Williams plot)

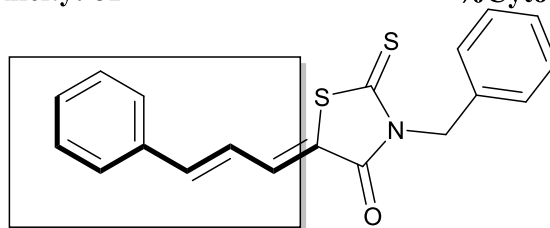




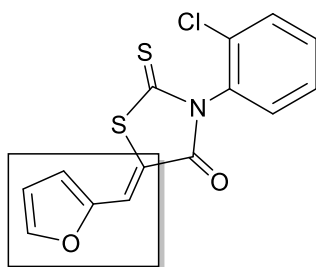
**Compound No: 4**  
**ATSC7v: -416.37**  
**geomRadius: 8.632**  
**%Cytotoxicity: 81**



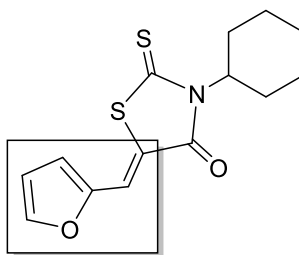
**Compound No: 16**  
**ATSC7v: -532.17**  
**geomRadius: 8.310**  
**%Cytotoxicity: 81**



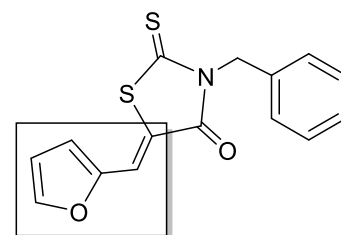
**Compound No: 27**  
**ATSC7v: -447.60**  
**geomRadius: 8.662**  
**%Cytotoxicity: 81**



**Compound No: 10**  
**geomRadius: 7.141**  
**%Cytotoxicity: 23**

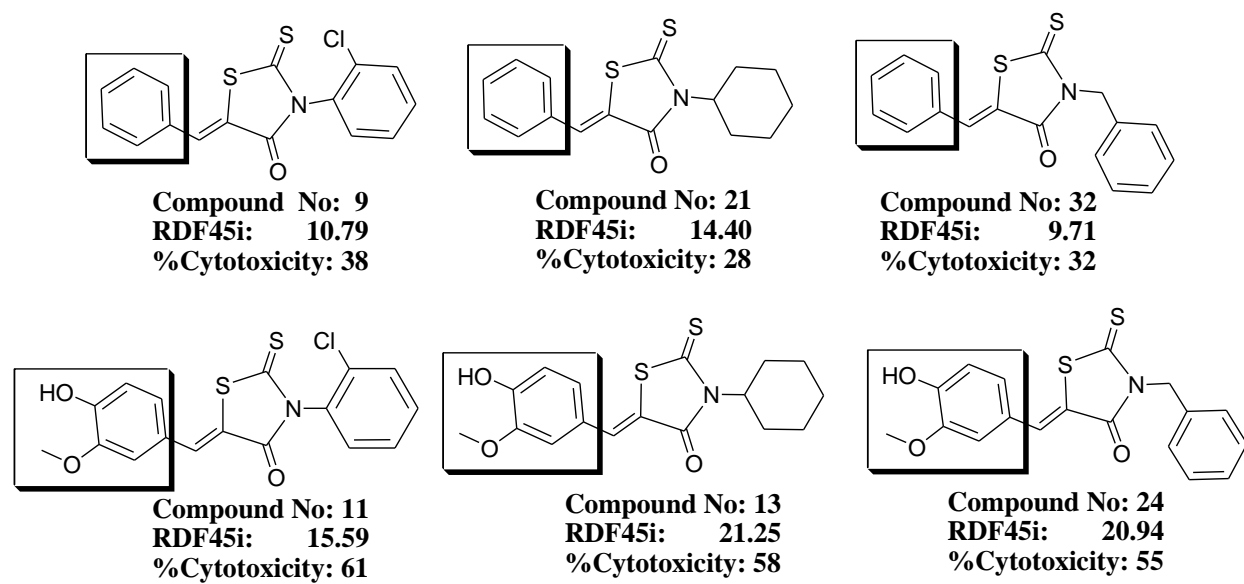


**Compound No: 22**  
**geomRadius: 6.556**  
**%Cytotoxicity :20**

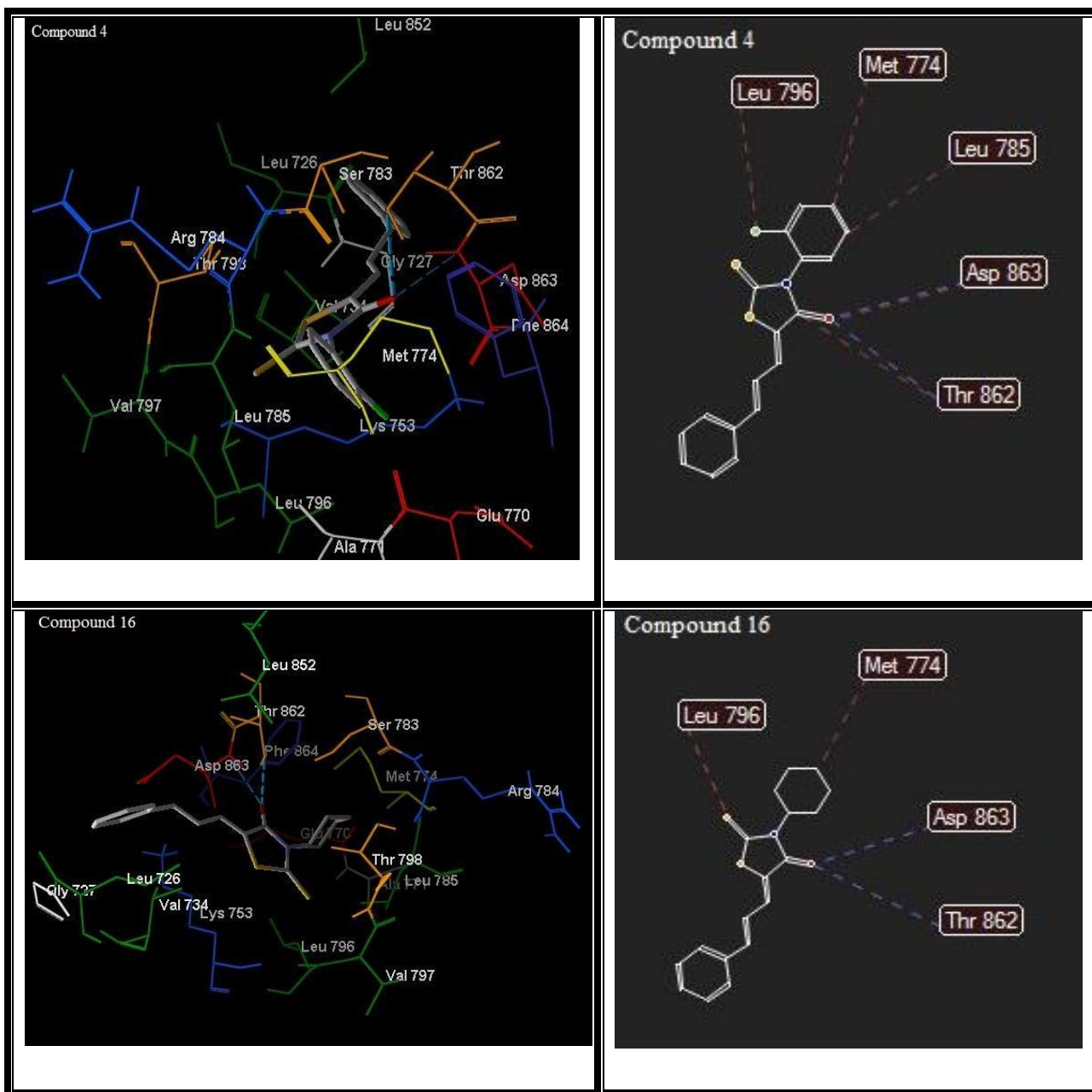


**Compound No: 33**  
**geomRadius: 6.935**  
**%Cytotoxicity: 28**

**Fig. (5).** Contribution of ATSC7v, geomRadius descriptors in compounds 4, 10, 16, 22, 27 and



**Fig. (6).** Contribution of RDF45i descriptor in compounds 9, 11, 13, 21, 24 and 32



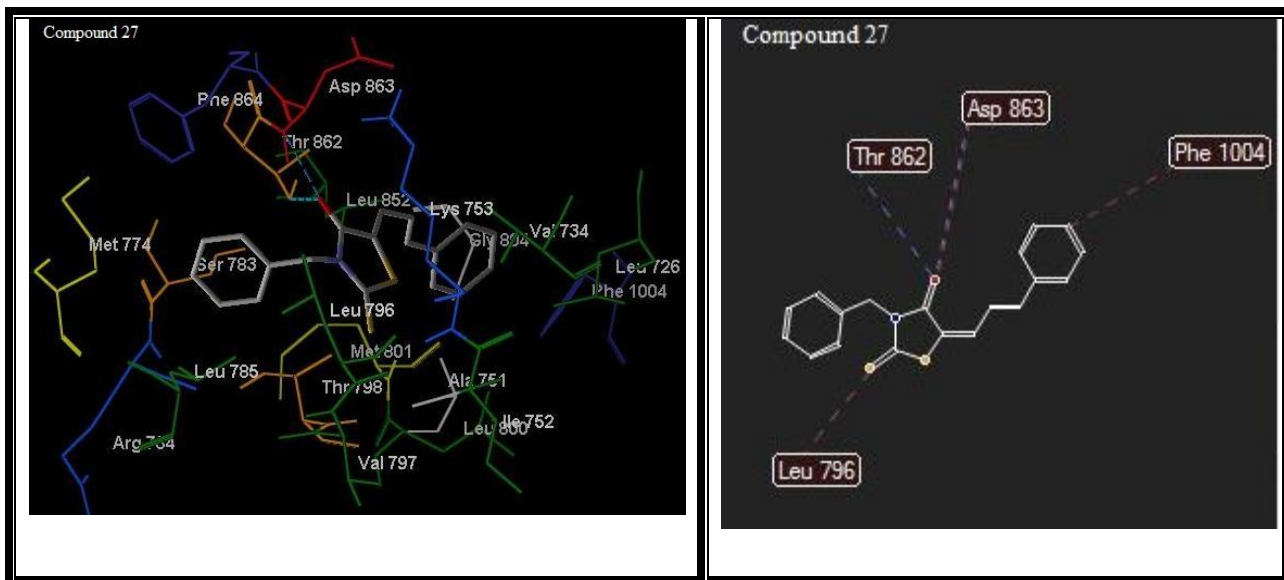
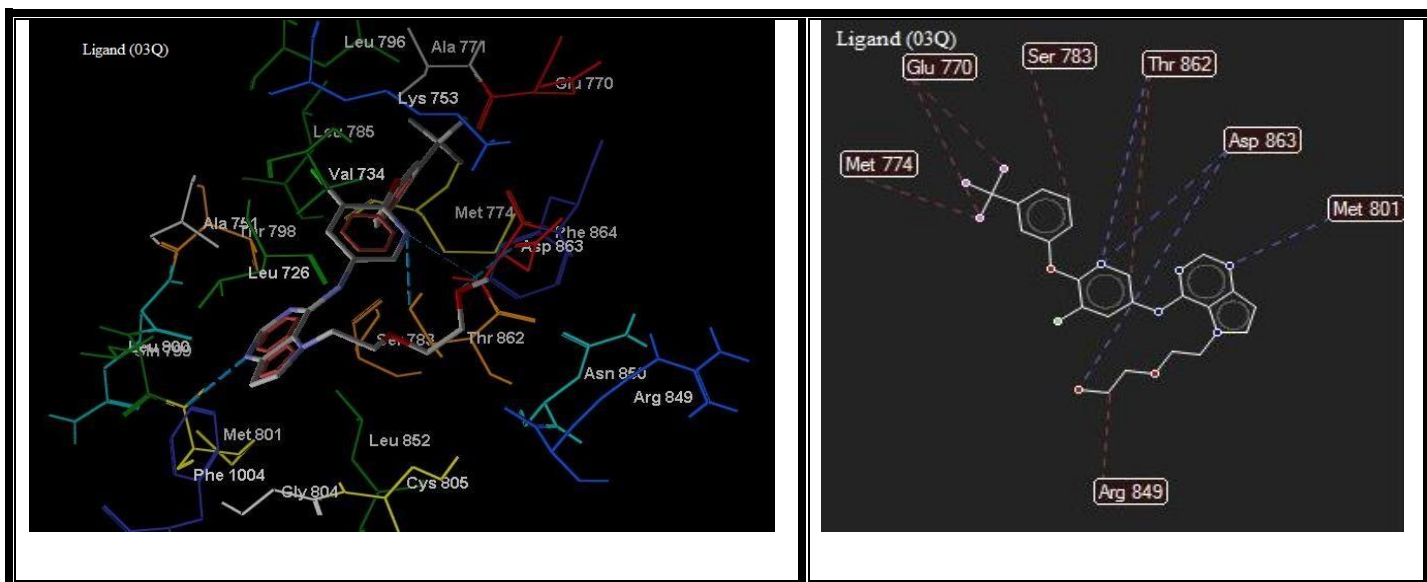
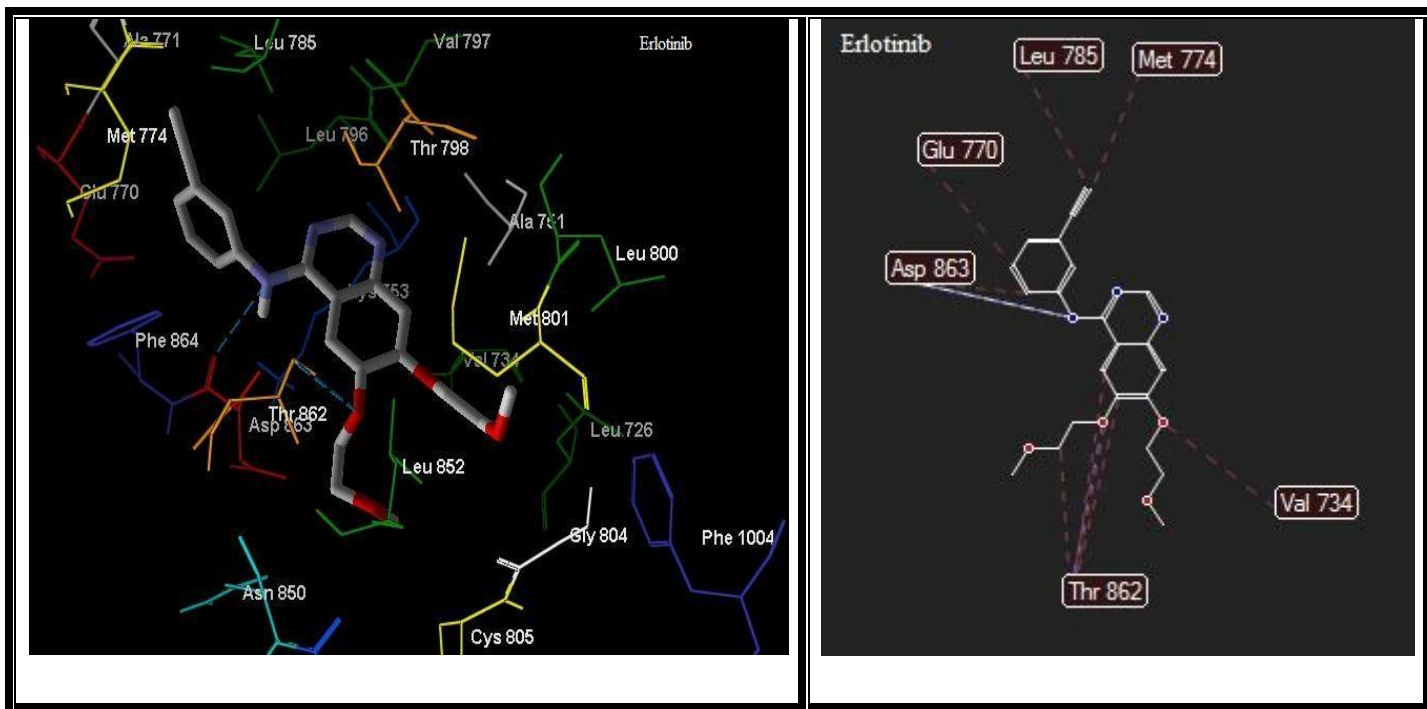


Fig. (7). 3D and 2D binding modes of compounds 4, 16, and 27.





**Fig. (8).** 3D and 2D binding modes of internal ligand (03Q), and reference drug (Erlotinib).