Available online at www.jcpronline.in

Journal of Current Pharma Research 5 (3), 2015, 1550-1564.

Journal of Current Pharma Research

http://www.jcpronline.in

Original Article

Design of novel pyrazolyl-thiazolinone derivatives as a potential EGFR and HER-2 kinase inhibitors by 2D and 3D QSAR using kNN MFA and molecular docking method.

Anwar R. Shaikh*, Amruta R. Nikam, Sanjay J. Kshirsagar, Mrs. Kankate R.S, Pardeshi Karansingh

Department of pharmaceutical chemistry, Bhujbal Knowledge City, MET's Institute of Pharmacy, Adgaon, Nashik-422003, Maharashtra, India.

Received 29 April 2015; received in revised form 24 May 2015; accepted 24 May 2015

Available online 28 June 2015

Abstract

The epidermal growth factor receptor (EGFR) and human epidermal receptor (HER-2) protein tyrosine kinase are important protein targets for anti-tumor drug discovery to identify potential EGFR and HER-2 inhibitors, we conducted 2D and 3D Quantitative structure-activity relationship (QSAR) studies on the inhibitory activity of recently synthesized pyrazolyl-thiazolinone derivatives (E1-E36) and from best model predicted activity of novel pyrazolyl-thiazolinone derivatives. 2D and 3D QSAR study was performed by using multiple linear regression method and KNN MFA method respectively by using random selection and manual data selection of 36 compound (E1-E36). The statistically best 2D QSAR model (r^2= 0.9125, q^2 = 0.8426, F = 39.63, Pred_r^2 = 0.6736, Pred_r^2 se = 0.2944) was developed using software Vlife MDS version 4.2. from this model novel 20 compounds (MCR 001-020) were designed and activity was predicted. Furthermore, on this novel compound (out of 20/10 compounds) docking studies were conducted to identify receptor ligand interaction (dock score of MCR 005 = -54.25) using Erlotinib as a positive control (dock score = -62.82).

Keywords: Pyrazolyl-thiazolinone, anti-tumor, 2D QSAR, 3D QSAR, EGFR, HER-2, docking.

Introduction

The epidermal growth factor (EGFR) and Human epidermal receptor HER-2, belonging to the family of Receptor tyrosine kinase (RTKs) also called as phosphotyrosine kinase. Receptor tyrosine kinase are single pass transmembrane protein important in intercellular signaling, by translating extracellular signals (ligands or growth factors) into activation of specific cell signaling cascades [2,3] that is they plays an important role in signal transduction pathways that regulate cell division and differentiation. The receptor tyrosine kinase family comprises of four members: EGFR (HER-1/ErbB-1), HER-2 (ErbB-2), HER-3 (ErbB-3) and HER-4 (ErbB-4). [4]

EGFR, a transmembrane protein tyrosine kinase that is activated by ligand induced dimerization, plays a critical role in regulating cell proliferation, differentiation and migration [5]. The Human epidermal growth factor 2 (HER-2) is localized is to chromosome 17q and encodes transmembrane tyrosine kinase receptor protein. [1] A large body of experimental and clinical work to support the view that the EGFR is a relevant target for cancer therapy. Two therapeutic approaches have been shown most promising and are currently being used to inhibit the EGFR: (*a*) MAbs; and (*b*) small molecule inhibitors of the EGFR tyrosine kinase enzymatic activity. Maps are generally directed at the external domain of the EGFR to block ligand binding and receptor activation. TKIs prevent the autophosphorylation of the intracellular tyrosine kinase domain of the EGFR.

*Corresponding author.
E-mail address: pharmacy_2003@rediffmail.com (Anwar R. Shaikh)

EGFR tyrosine kinase-mediate cell growth signaling pathways plays an important role in the formation and development of many types of solid tumors, including head and neck, lung, breast, bladder, prostate and kidney cancers. [6-9] Therefore, EGFR tyrosine kinase represents an attractive target for the development of novel anticancer agents. EGFR and HER-2 are the hottest targets in current cancer research and their over expression or abnormal activation often cause cell malignant transformation. Gefitinib and Erlotinib are the representative drugs for this kind of inhibitors and have been approved by US FDA for the treatment of patient with non small-cell-lung cancer (NSCLC). [10,11]

Thiazolinone and their derivative have attracted continuing interest over the years because of their varied biological activities, such as anti-inflammatory, antimicrobial, antiproliferative, antiviral, anticonvulsant, antifungal and antibacterial. [12,13] Recent years thiazolinone derivative with their antitumor activity have become popular. Havrylyuk et al. reported that thiazolinones containing benzothiazole moiety has anticancer activity on leukemia, lung, colon, CNS, ovarian, renal, prostate and breast cancers cell lines. [14] The pyrazole motif makes up the core structure of numerous biologically active like antiviral/antitumor, antibacterial, antiinflammatory analgesic, fungistatic and anti hypoglycemic activity. [15] Pyrazole derivative containing thiourea skeleton were recently reported as potent anticancer agents targeting EGFR. The pyrazole ring along with the thiazolinone ring the two combined substructure might exhibit synergistic anticancer effect. By integrating these two heterocyclic rings we can synthesize some novel pyrazolyl-thiazolinone derivative as a potential EGFR and HER-2 inhibitory agent.

Quantitative structure–activity relationship (QSAR) is the most popular theoretical method for modeling a compound's biological activity of its chemical structure. [16-18] Using this approach, scientists could predict the activities of a series of newly designed drugs before making the final decision on whether or not to synthesize and assay them. The prediction is based on the structural descriptors of the molecular features that most accounts for the variations in biological activity. Furthermore, this method also can identify and describe the most important structural features of the compounds which are relevant to the variations in molecular properties, thus, it also gains an insight into the structural factors which affect the molecular properties. QSAR models of EGFR inhibitors have been recently investigated with encouraging results. [19-22]

Four QSAR models were constructed from a set of known pyrazolyl-thiazolinone derivatives EGFR and HER-2 inhibitors with multiple linear regression method and kNN MFA method by using random and manual data selection (training and test set). The study sheds the light on the structure-activity relationship of this class of EGFR and HER-2 inhibitors and has the potential predicted ability to identify new EGFR and HER-2 inhibitors. From these four best 2D and 3D QSAR models pIC_{50} value was predicted of newly drawn 20 pyrazoly-thiazolinone derivatives structure and conforming that structure by Chem spider. The pIC_{50} value of these newly designed inhibitors was found to be higher or equal to that of already synthesized inhibitors [E1-E36].

All Docking studies and conformational analysis were performed using the Molecular Design Suite (VLife MDS software package, version 4.2.). The application of computational methods to study the formation of intermolecular complexes is a subject of intensive research. The drug exerts its biological activity by binding to the pocket of receptor molecule (usually protein). In their binding conformations, the molecules exhibit geometric and chemical complementarily, both of which are essential for successful drug activity. The computational process of searching for a ligand that is able to fit both geometrically and energetically into the binding site of a protein is called molecular docking. Molecular docking helps in studying drug/ ligand or receptor/ protein interactions by identifying the suitable active sites in protein, obtaining the best geometry of ligand - receptor complex and calculating the energy of interaction for different ligands to design more effective ligands.

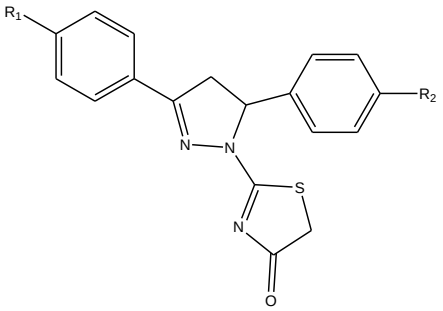

Fig.1. Basic structure of pyrazolyl-thiazolinone derivative.

Materials and Methods

QSAR studies

Selection of molecules

A data set of 36 pyrazolyl-thiazolinone derivatives (Table 1) collected from published literature [] were taken for the present study. The affinity data of inhibitory activities were converted into pIC_{50} values to get the linear relationship in equation using the following formula: pIC_{50}= -log IC_{50}, where IC_{50} values represent inhibitory activity in µM (Table 2). Molecules were rationally divided into the training set and test set by random and manual selection method.

Molecular modeling:

All computational experiments were performed using on HCL computer having genuine Intel Pentium Dual Core Processor and Windows XP operating system using the software Molecular Design Suite (vlifeMDS 4.2.). Structures were drawn using the 2D draw application and converted to 3D structures and subjected to an energy minimization and geometry optimization using Merck Molecular Force Field, force field minimization and charges followed by Austin Model-1 with 10000 as the maximum number of cycles, 0.01 as convergence criteria (root mean square gradient) and 1.0 as constant (medium's dielectric constant which is 1) in dielectric properties. The default values of 30.0 and 10.0 Kcal/mol were used for electrostatic and steric energy cutoff.

2D-QSAR analysis

Calculation of descriptors

The number of descriptors were calculated after optimization or minimization of the energy of the data set molecules. Various types of physicochemical descriptors were calculated: Individual (Molecular weight, H-Acceptor count, H Donor count, XlogP, slogP, SMR, polarisablity, etc.), retention index (Chi), atomic valence connectivity index (ChiV), Path count, Chi chain, ChiV chain, Chain Path Count, Cluster, Path cluster, Kappa, Element count (H, N, C, S count etc.), Distance based topological (Dist Topo, Connectivity Index, Wiener Index, Balaban Index), Estatenumbers (SsCH3count, SdCH2count, SssCH2count, StCHcount, etc.), Estate contribution (SsCH3-index., SdCH2-index, SssCH2-index , StCH index), Information theory based (Ipc, Id etc.) and Polar surface area. More than 200 alignment independent descriptors were also calculated using the following attributes. A few examples are T_2_O_7, T_N_N_5, T_2_2_6, T_C_O_1, T_O_Cl_5 etc. The invariable descriptors (the descriptors that are constant for all the molecules) were removed, as they do not contribute to QSAR.

Generation of training and test sets:

In order to evaluate the QSAR model, the data set was divided into training and test set using random selection and manual selection method. The training set is used to develop the QSAR model for which biological activity data are known. The test set is used to challenge the QSAR model developed based on the training set to assess the predictive power of the model which is not included in model generation.

Random Selection Method:

In order to construct and validate the QSAR models, both internally and externally, the data sets were divided into training [90%-60% (90%, 85%, 80%, 75%, 70%, 65% and 60%) of the total data set] and test sets [10%-40% (10%, 15%, 20%, 30%, 35% and 40%) of the total data set] in a random manner. 10 trials were run in each case.

Manual data selection method:

The data set is divided manually into training and test sets on the basis of the result obtained by the random selection method.

Generation of 2D-QSAR models:

Multiple linear regression method is used for 2D analysis. Multiple regression is the standard method for multivariate data analysis. It is also called as ordinary least squares regression (OLS). This method of regression estimates the values of the regression

coefficients by applying least squares curve fitting method. For getting reliable results, dataset having typically 5 times as many data points (molecules) as independent variables (descriptors) is required.

The regression equation takes the form

$$Y = b1*x1 + b2*x2 + b3*x3 + c,$$

Where, Y is the dependent variable, the 'b' is regression coefficients for corresponding 'x' is independent variable, 'c' is a regression constant or intercept.

Stepwise multiple regression (SMR)

It is an approach to select a subset of variables, when the numbers of independent variables (descriptors) are much more than the number of data points (molecules). It is also a procedure to examine the impact of each variable to the model step by step. Each variable is added to the equation and a new regression is performed. The variable that cannot contribute much to the variance explained would not be added. As a result, SMR generates a single, multiple regression equation.

3D-QSAR analysis:

kNN-MFA

kNN-MFA is novel methodology, unlike conventional QSAR regression methods; this methodology can handle nonlinear relationships of molecular field descriptors with biological activity, thus making it a more accurate predictor of biological activity. The standard kNN method is implemented simply as follows: (i) calculate distances between an unknown object (u) and all the objects in the training set; (ii) select k objects from the training set most similar to object u, according to the calculated distances, (iii) classify object u with the group to which a majority of the k objects belongs. An optimal k value is selected by the optimization through the classification of a test set of samples or by the leave-one out cross-validation. The variables and optimal k values are chosen using different variable selection methods as described below.

kNN-MFA with Simulated Annealing

Simulated Annealing (SA) is another stochastic method for function optimization employed in QSAR. Simulated annealing (SA) is the simulation of a physical process, 'annealing', which involves heating the system to a high temperature and then gradually cooling it down to a preset temperature (e.g., room temperature). During this process, the system samples possible configurations distributed according to the Boltzmann distribution, so that at equilibrium, low energy states are the most populated. kNN-MFA with Stepwise (SW) Variable Selection. This method employs a stepwise variable selection procedure combined with kNN to optimize the number of nearest neighbors (k) and the selection of variables from the original pool as described in simulated annealing

Alignment rules:

Molecular alignment was used to visualize the structural diversity in the given set of molecules. This was followed by generation of common rectangular grid around the molecules. The template structure, i.e. unsubstituted pyrazolyl-thiazolinone was used for alignment by considering the common elements of the series as shown in Figure 2. The reference molecule 28 is chosen high inhibitory effect which made it a valid lead molecule and therefore was chosen as a reference molecule. After optimizing, the template structure and the reference molecule were used to superimpose all molecules from the series using the template alignment method. kNN-MFA method requires suitable alignment of given set of molecules after optimization; alignment was carried out by template based alignment method.

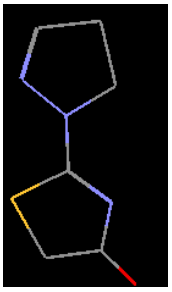

Fig.2. Para substituted pyrazolyl thiazolinone derivatives (template structure)

Creation of interaction energies

Methyl probe with charge 1 and energy cut-off for electrostatic 10 Kcal/mol and for steric 30 Kcal/mol, dielectric constant 1 and charge type Gasteiger-marsili were used to calculate steric and electrostatic fields. The fields were computed at each lattice intersection of a

regularly spaced grid of 2.0 A° within a defined three-dimensional region.

Generation of training and test sets

In order to evaluate the QSAR model, the data set was divided into training and test set using random selection and Manual selection method. The training set is used to develop the QSAR model for which biological activity data are known. The test set is used to challenge the QSAR model developed based on the training set to assess the predictive power of the model which is not included in model generation.

Docking studies on para-substituted pyrazolyl-thiazolinone derivative-

All Docking studies and conformational analysis were performed using the Molecular Design Suite (VLife MDS software package, version 4.2; from VLife Sciences, Pune, India). 2D structure was drawn by Vlife2Ddraw and converted to 3D. Then this 3D structure is optimized and conformers were generated (Monte Carlo method) and least energy conformers were selected. For preparation of ligand perform the geometry minimization of the ligand.

For preparation of protein first download PDB structures (www.rcsb.org) [1M17] the complex was obtained by Merck molecular force field. For preparation of ligand perform the geometry minimization of the ligand. Merck Molecular Force Fields (MMFF) with default settings were used for the ligand minimization. Further docking study was performed by using Grip method. Docking study was performed in VlifeMDS version 4.3 on Lenovo computer, i3 processors with XP operating system. The GRIP-based ligand docking with genetic algorithm approximated a systematic search of positions, orientations, and conformations of the ligand in the enzyme binding pocket via a series of hierarchical filters. The minimum dock score of example may not be exactly reproducible because this is (GRIP) based run. However, changing the different input parameters in the GRIP parameters dialog box (like No of Generations, Translation, Rotation limits etc.) can result in dock scoring energies within the desired range and improvement in the orientation of the docked ligand as close to that of Co crystallized ligand as possible.

Results and discussion

QSAR:

2D QSAR:

In 2D QSAR analysis, Multiple linear regression analysis (MLR) coupled with stepwise forward, backward variable selection method was applied to generate 2D models. Selection of training and test set was done by Manual data selection, Random selection method. From these models, the one having good q^2 and pred_r^2 values was selected as the best model. Statistically significant 2D QSAR models are shown in Table 3 and 4.

The selection of the best model is based on the values of r^2 (squared correlation coefficient), q^2 (cross-validated correlation coefficient), pred_r^2 (predicted correlation coefficient for the external test set), F (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. r^2se, q^2se and pred_r^2se are the standard error terms for r^2, q^2 and pred_r^2 respectively. The statistically significant 2D-QSAR model is shown as follows.

Model-1: (Test set: 22,11,16,17,18, 21, 23, 24,25,4,5)

pIC50 = 0.506 (prediction) + 0.1070 (T_2_O_7) – 0.0081 (Quadrapole 1) – 0.0001 (Mom Inertia X) – 01664 (T_2_Cl_5) + 0.1812

Statistics: [n = 25; Degree of freedom = 19; r^2 = 0.9125; q^2 = 0.8426; F test = 39.63; r^2se = 0.1030; q^2se = 0.1381; pred_r^2 = 0.2944; pred_r^2se = 0.2944]

Model-2: (Test set: 10, 14,19,25,28,32,34,35,363,9)

pIC50 = 8.5417 (Electronegativity count) + 0.0066 (Quadrapole 2) + 0.1322 (T_C_O_7) + 0.07976 (T_C_C_1) – 0.001 (Mom Inertia X) + 7.7818

Statistics: [n = 25; Degree of freedom = 19; r2 = 0.8675; q2 = 0.7996; F test = 24.88; r2se = 0.1151; q2se = 0.1416; pred_r2 = 0.2875; pred_r2se = 0.1881]

In the above QSAR equations, n is the number of molecules (Training set) used to derive the QSAR model, r^2 is the squared correlation coefficient, q^2 is the cross-validated correlation coefficient, pred_r^2 is the predicted correlation coefficient for the external test set, Fis 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. r^2se, q^2se and pred_r^2se are the standard error terms for r^2, q^2 and pred_r^2 (smaller is better).

Interpretation of the Models

Model-1

From the equation, model 1 explains 91.25% (r^2 = 0.9125) of the total variance in the training set as well as it has internal (q^2) and external (pred_r^2) predictive ability of 84.26 % and 39.63 % respectively. The F test shows the statistical significance of 99.99 % of the model which means that the probability of failure of the model is 1 in 10000. In addition, the randomization test shows confidence of 99.9999. From QSAR model 1, positive coefficient value of T_2_O_7 [count of number of double bonded atoms (i.e. any double bonded atom, T_2) separated from the oxygen atom by seven bonds] on the biological activity indicated that higher values leads to good inhibitory activity while lower value leads to reduced inhibitory activity while a negative coefficient value of T_2_Cl_5 [count of the number of double bonded atoms (i.e. any double bonded atom, T_2) separated from a chlorine atom by five bonds] on the inhibitory activity indicated that lower value leads to better inhibitory activity whereas a higher value leads to decrease inhibitory activity. Mom Inertia X [the descriptor signifies moment of inertia at X axis] lower value or Zero value of Mom Inertia does not affect the activity. The negative coefficient value of Quadrapole 1 [Magnitude of a first tenor of the quadrapole moment] the inhibitory activity indicated that lower values lead to a good inhibitory activity where as higher value leads to decrease inhibitory activity.

Contribution chart for model 1 is represented in Fig. 3, it reveals that the descriptors prediction and T_2_O_7 contributing 27.92% and 22.036% respectively, and other descriptors T_2_Cl_5, Quadrapole 1 and Mom Inertia is contributing inversely 9.284%, 29.67% and 11.076% respectively to biological activity.

Data fitness plot for model 1 is shown in Fig.3. The plot of observed vs. The predicted activity provides an idea about how well the model was trained and how well it predicts the activity of an external test set.

Model 2

From the equation, Model 2 explains 86.75% (r^2 = 0.8675) of the total variance in the training set as well as it has internal (q^2) and external (pred_r^2) predictive ability of 79.96% and 28.75% respectively. The F-test=24.88 .From QSAR model 2, it was observed that Positive coefficient value of Electronegativity count, Quadrapole2 [Magnitude of second tensor of Quadrapole moment],T_C_O_7[count of number of carbon atoms (single, double or triple bonded) separated from any other oxygen atom (single or double or triple bonded) by 7 bond in a molecule] T_C_C_1 [count of number of carbon atoms (single, double or triple bonded) separated from any other carbon atom (single or double or triple bonded) by 1 bonds in a molecule on the biological activity indicated that higher values leads to good inhibitory activity while lower value leads to reduced inhibitory activity. Negative value or Zero value of MomInertia X [Moment of inertia at X axis] does not affect the activity.

Contribution chart for model 2 is represented in Fig.5, it reveals that the descriptor electronegativity count, Quadrapole2, T_C_O_7, T_C_C_1, contributing 36.065%, 23.738%, 17.904% and 12.208 % respectively and Mom Inertia inversely contributing 10.086%.

Data fitness plot for model 2 is shown in Fig.5, the plot of observed vs. The predicted activity provides an idea about how well the model was trained and how well it predicts the activity of an external test set.

The graph of actual vs. The predicted activity of training and test sets for model 2 is shown in Fig. 5, it reveals that the model is able to predict the activity of training set quite well as well as an external test set, providing confidence of the model. Result of the actual vs. The predicted inhibitory activity of the training and test set compounds for the Model 1 and 2 is shown in Table 7.

3D-QSAR:

kNN-MFA samples the steric and electrostatic fields surrounding a set of ligands and constructs 3D-QSAR models by correlating these 3D fields with the corresponding

biological activities. Molecular alignment was used to visualize the structural diversity in the given set of molecules. The template structure, i.e. para-substituted pyrazolyl-thiazolinone derivatives was used for alignment by considering the common elements of the series as shown in Fig.2.

The statistical significant of 3D-QSAR models are shown in Table 5 and 6.

The selection of the best model is based on the values of q^2 (internal predictive ability of the model) and that of pred_r^2 (the ability of the model to predict the activity of external test set). The statistical significant 3D-QSAR models for pIC_{50} (model-3) and pIC_{50} (model-4) are given below.

Model-3 (Test set-
14,15,18,20,23,30,32,35,3,4,5.**)**

***p*IC$_{50}$**= -S_447 (-0.0322 -0.0321) + E_38 (0.0243 0.0245)

Statistics: [kNN = 2; n = 25; Degree of freedom = 22; q^2 = 0.7219; q^2_se = 0.1622; pred_r^2 = 0.1002; pred_r^2se = 0.4912]

Model-4 (Test set-14,16,18,21,28,33,35,5,6,9**)**

***p*IC$_{50}$**= -S_1252 (-0.2808 -0.0302) + E_499 (0.3253 0.3498)

Statistics: [kNN = 2; n = 25; Degree of freedom = 21; q^2 = 0.0.7185; q^2_se = 0.1238; pred_r^2 = 0.2586; pred_r^2se = 0.2715]

The model 3 explains values of k (2), q^2 (0.7219), pred_r^2 (0.0.1002), q^2_se (0.1622), and pred_r^2 se (0.0.4912) prove that QSAR equation so obtained is statistically significant and shows the predictive power of the model is 72.19% (internal validation). Table 7 represents the predicted inhibitory activity by the model-3 for training and test set.

The data fitness plot for model 3 is shown in Fig.7.The plot of observed vs. The predicted activity provides an idea about how well the model was trained and how well it predicts the activity of the external test set.

Result plot in which 3D-alignment of molecules shown in with the important steric and electrostatic points contributing in the model-3 with ranges of values shown in the parenthesis represented in Fig.7.

It shows the relative position and ranges of the corresponding important steric and electrostatic fields in the model provides guidelines for new molecule design as follows-

(a) Electrostatic field, E_38 (0.0243, 0.0245) has a positive range indicates that positive electrostatic potential is favorable for an increase in the activity and hence the less electronegative substituent group is preferred in that region.

(b) Steric field, S_447 (-0.2808, -0.0321,) has a negative range indicates that negative steric potential is favorable for an increase in the activity and hence less bulky substituent group is preferred in that region.

(c) In model 4, values of kNN (2), q^2 (0.7185), pred_r^2 (0.2586), q^2_se (0.1283), and pred_r^2 se (0.2715) prove that QSAR equation so obtained is statistically significant and shows the predictive power of the model is 71.85% (internal validation). Table 7 represents the predicted inhibitory activity by the model-4 for training and test set.

(d) The data fitness plot for model-4 is shown in Fig. 9. The plot of observed vs. The predicted activity provides an idea about how well the model was trained and how well it predicts the activity of the external test set.

It shows the relative position and ranges of the corresponding important steric and electrostatic fields in the model provides guidelines for new molecule design as follows-

(a) Electrostatic field, E_499 (0.3253, 0.3498) has a positive range indicates that positive electrostatic potential is favorable for an increase in the activity and hence the more electronegative substituent group is preferred in that region.

(b) Steric field, S_1633 (-0.2808, -0.0302,) has a negative range indicates that negative steric potential is favorable for an increase in the activity and hence more bulky substituent group is preferred in that region.

(c) Design and activity prediction of newer derivatives

(d) From the best models obtained, some newer compounds were designed claiming to posses better activity than

the reported one. The structures were not reported earlier anywhere is confirmed by chem. Spider.

Docking studies of newly designed compounds-

To gain a better understanding on the potency of the newly designed compounds (MCR 001-010) we proceeded to examine the interaction of it with EGFR (PDB code: 1M17) by molecular docking, which was performed by simulation of the compound into the ATP binding site in EGFR. The binding model of the compounds and EGFR was depicted in the following figures. The amino acid residues which had interaction with EGFR were labelled. In the binding mode, compound MCR 005 was nicely bound to the ATP binding site of EGFR through hydrophobic interaction and the binding was stabilized by a hydrogen bond and a π-cation interaction.

In the binding model, compound MCR 005 was nicely bound to the EGFR kinase with its carbonyl –group of A721 (LYS 721), forming a more optimal hydrogen bond interaction (distance: N-H… O = 2.58A^0). Based on the favourable EGFR inhibitory activity of pyrazolyl-thiazolinone derivatives, it could be concluded that this H-bond played an important effect in the EGFR inhibitory. The end amino cation of PHE699A was also formed a π-cation interaction with the benzene ring of the compound MCR 005 (distance: 5.02 A^0), which enhanced the binding action between receptor EGFR and the ligand compound MCR 005. The electron donating substituent amino-NH_2 strengthened the π-cation binding and the results indicated that it was primarily due to direct through-space interaction between the subtituent and the cation.

Meanwhile, erlotinib was taken as a positive control in the docking procedure. It was found that there was only one sigma-π bond in the binding pocket. The benzene ring of erlotinib formed one sigma-π bond with PHE699A (distance: 3.97 A^0). The model between compound MCR 005 and the ATP binding site was similar to that with erlotinib. Comparing these models, it was found that the hydrophobic pockets of ATP binding were all nicely occupied by these compounds, and were nicely occupied by these compounds, and the difference was the combination mode. This molecular docking result, along with the biological assay data, suggesting that the compound MCR 005 was a potential inhibitor of EGFR.

Fig.12.3D view –Molecular docking modelling of parent compound erlotinib (Dock score = -62.82) and newly design MCR 005 (dock score = -54.25) with EGFR kinase: for clarity, only interacting residues are displayed. Above: 3Dmodel of the interaction between compound MCR 005 and the ATP binding site. The H-bond (green lines) is displayed as a dotted line, and the π-cation interaction is shown as yellow lines and the hydrophobic interaction is shown as dotted blue lines.

Fig.13. 2Dview- Molecular docking modelling of compound Erlotinib with EGFR kinase: 2D model of the interaction between compound Erlotinib and the ATP binding site. The H-bond (dark blue arrows) is displayed as dotted arrows. And the π-cation interaction is shown as pink lines and hydrophobic bond is shown as faint blue lines.

References

1. J.S. Ross, J.A. Fletcher, K.J. Bloom et al. Targeted therapy in breast cancer: The HER-2/neu gene and protein. Mol Cell Proteomics, 3 (2004) 379 –398.
2. S.R. Hubbard, J.H. Till. Protein Tyrosine Kinase Structure and Function. Annu. Rev. Biochem., 69 (2000) 373-398.
3. J. Schlessinger. Cell Signaling by Receptor Tyrosine Kinases. Cell, 103, (2000) 211-225.
4. K.S. Kolibaba, B.J. Druker. Biochim. Biophys. Acta, 21 (1997) 1333.
5. X.W. Zhang, J. Gureasko, K. Shen, P.A. Cole, J. Kuriyan. Cell, 125 (2006) 1137.
6. T.P. Fleming, A. Saxena, W.C. Clark, J.T. Robertson, E.H. Oldfield, S.A. Aaronson, I.U. Ali. Cancer Res., 52 (1992) 4550.
7. P.B. Jensen, T. Hunter. Nature, 411 (2001) 355.
8. H. Kim, W.J. Muller. Exp. Cell Res., 253 (1999) 78.
9. D.K. Moscatello, M. Holgado-Mudruga, A.K. Godwin, G. Ramirez,

G. Gunn, P.W. Zoltick, J.A. Biegel, R.L. Hayes, A.J. Wong. Cancer Res., 55 (1995) 5536.

10. J. Anido, P. Matar, J. Albanell, M. Guzman, F. Rojo, J. Arribas, S. Averbuch, J. Baselga. Clin. Cancer Res., 9 (2003) 1274.

11. V. Chandregowda, G.V. Rao, G.C. Reddy. Org. Process Res. Dev., 11 (2007) 813.

12. R. Ottana, R. Maccari, M.L Barreca, G. Bruno, A. Rotondo, A. Rossi, G. Chiricosta, R. Di Paola, L. Sautebin, S. Cuzzocrea, M.G. Vigorita. Bioorg. Med. Chem., 13 (2005) 4243.

13. E. Rydzik, A. Szadowska, A. Kaminska. Acta Pol. Pharm., 41 (1984) 459.

14. D. Havrylyuk, L. Mosula, B. Zimenkovsky, O. Vasylenko, A. Gzella, R. Lesyk. Eur. J. Med. Chem. 45, 5012 (2010) 5021.

15. R. Sridhar, P.T. Perumal, S. Etti, G. Shanmugam, M.N. Ponnuswamy, V.R. Prabavathy, N. Mathivanan. Bioorg. Med. Chem. Lett., 14 (2004) 6035.

16. R.P. Verma, C. Hansch. QSAR modeling of taxane analogues against colon cancer. Eur J Med Chem., 45 (2010) 1470–1477.

17. F.P. Liu, Y.Z. Liang, C.Z. Cao. QSPR modeling of thermal conductivity detection response factors for diverse organic compound. Chemometrics and Intelligent Laboratory Systems 81 (2006) 120–126.

18. A.S. Mandal, K. Roy. Predictive QSAR modeling of HIV reverse transcriptase inhibitor TIBO derivatives. European Journal of Medicinal Chemistry 44 (2009) 1509–1524.

19. F.A. Pasha, M. Muddassar, A.K. Srivastava, S.J. Cho. In silico QSAR studies of anilinoquinolines as EGFR inhibitors. Journal of Molecular Modeling, 16 (2010) 263–277.

20. A. Vema, S.K. Panigrahi, G. Rambabu, B. Gopalakrishnan, J. Sarma et al. Design of EGFR kinase inhibitors: A ligand-based approach and its confirmation with structure-based studies. Bioorganic & Medicinal Chemistry, 11 (2003) 4643–4653.

21. C. Szantai-Kis, I. Kovesdi, D. Eros, P. Banhegyi, A. Ullrich et al. Prediction oriented QSAR modelling of EGFR inhibition. Current Medicinal Chemistry, 13 (2006) 277–287.

22. C. La Motta, S. Sartini, T. Tuccinardi, E. Nerini, F. Da Settimo et al. Computational studies of epidermal growth factor receptor: Docking reliability, three-dimensional quantitative structure-activity relationship analysis, and virtual screening studies. Journal of Medicinal Chemistry, 52 (2009) 964–975.

23. W.M. Shi, Q. Shen, W. Kong, B.X. Ye. QSAR analysis of tyrosine kinase inhibitor using modified ant colony optimization and multiple linear regression. European Journal of Medicinal Chemistry, 42 (2007) 81–86.

24. D.J. Abraham. Burger's Medicinal Chemistry and Drug Discovery, Docking and Scoring Functions/Virtual Screening. 1, 06 (1998) 281-330.

25. K. Qiu, H. Wang, L. Wang, Y. Luo, X. Yang, X. Wang, H. Zhu. Bioorganic & medicinal chemistry, 20, 6 (2012) 2010-2018.

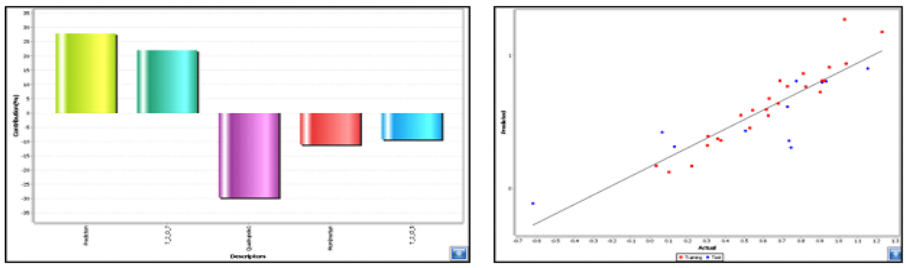

Fig.3. Contribution chart and fitness plot for 2D QSAR model-1.

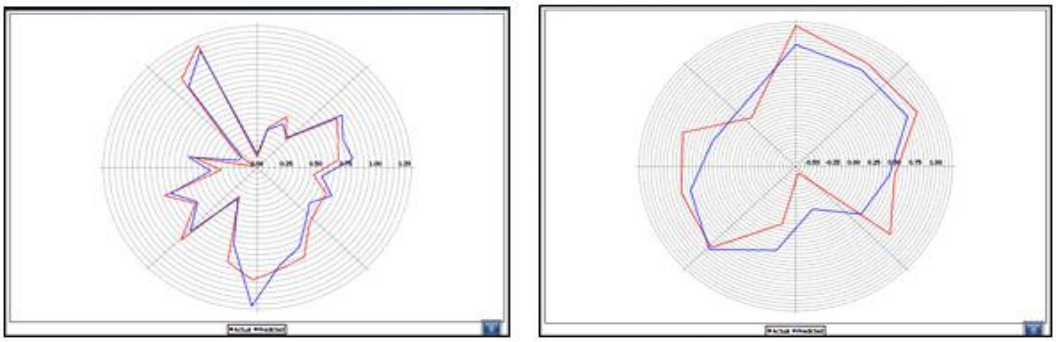

Fig.4. Plot of observed vs. predicted activity of 2D QSAR model-1 (training set and test set).

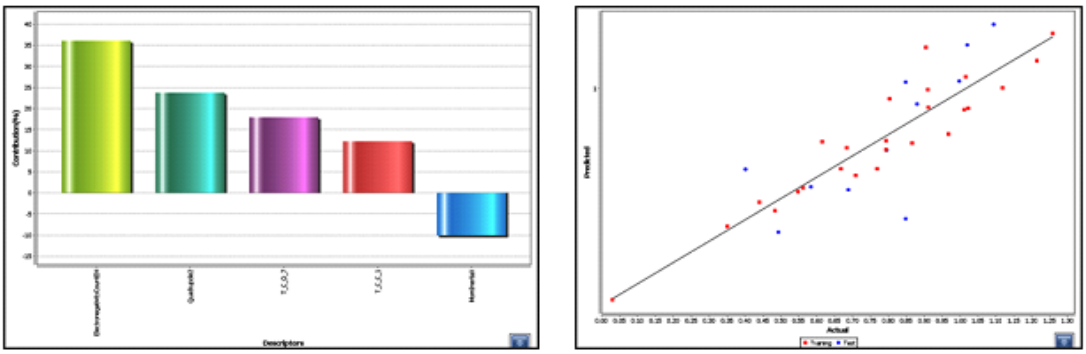

Fig.5. Contribution chart and fitness plot for 2D QSAR model-2.

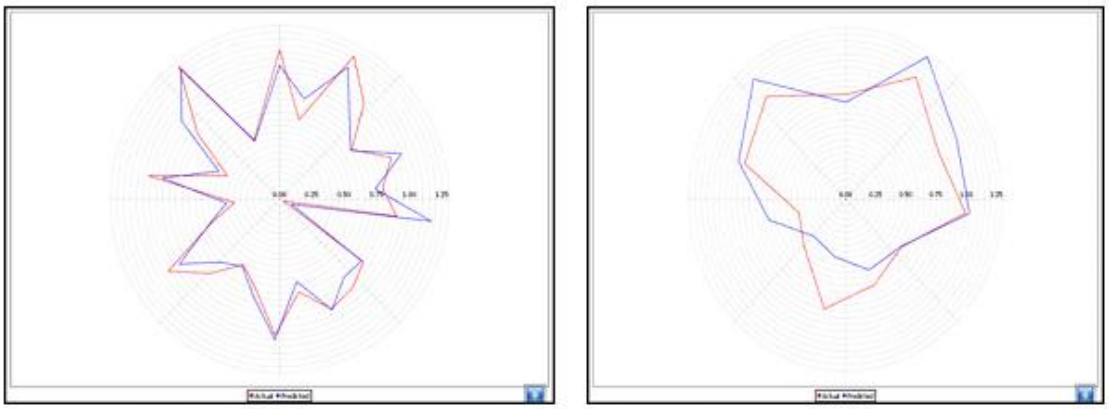

Fig.6. Plot of observed vs. predicted activity of 2D QSAR model-2 (training set and test set).

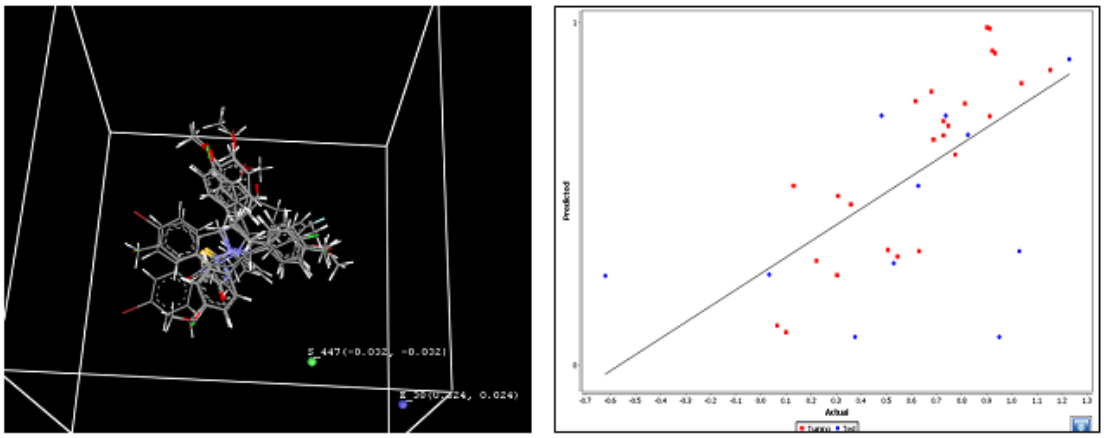

Fig.7. 3D Alignment of molecule and fitness plot for 3D QSAR model-3.

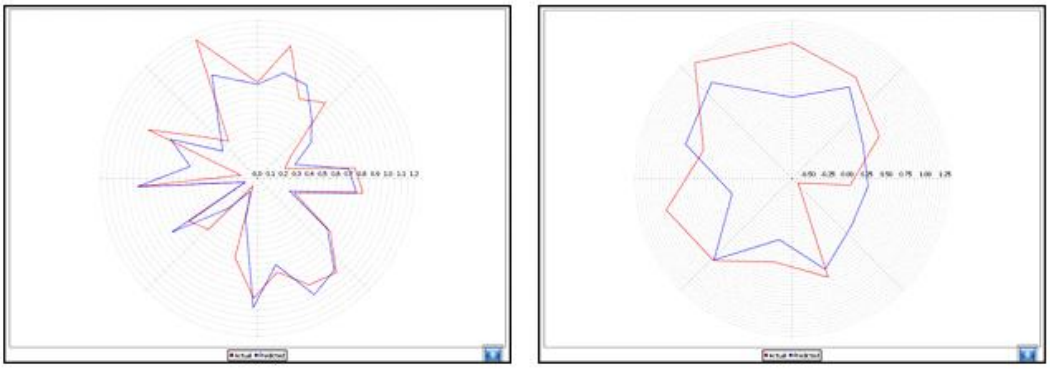

Fig.8. Plot of observed vs. predicted activity of 3D QSAR model-3 (training set and test set).

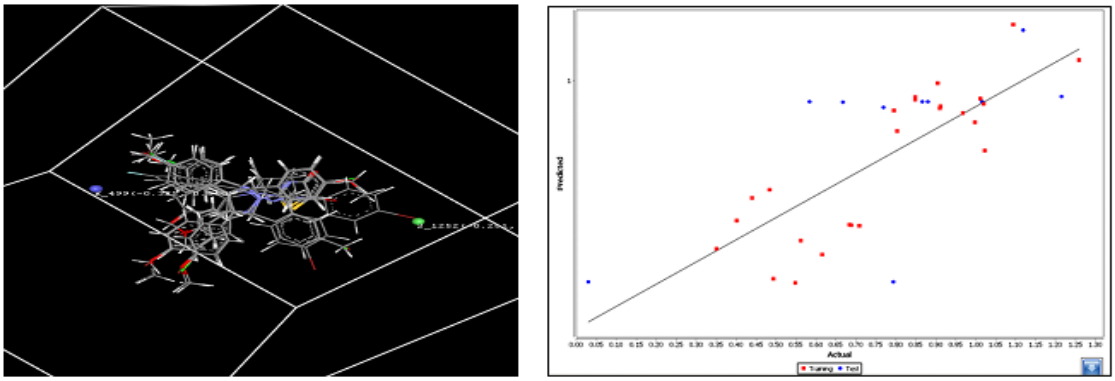

Fig.9. 3D Alignment of molecule and fitness plot for 3D QSAR model-4.

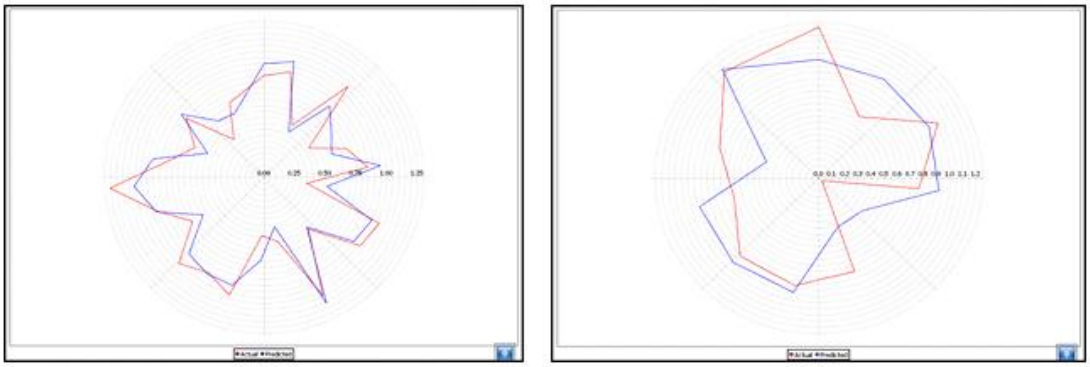

Fig.10. Plot of observed vs. predicted activity of 3D QSAR model-4 (training set and test set).

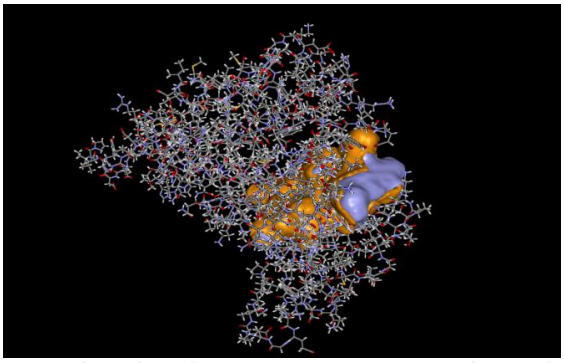

Fig.11.Active site of MCR 005 receptor-cavity-surface.

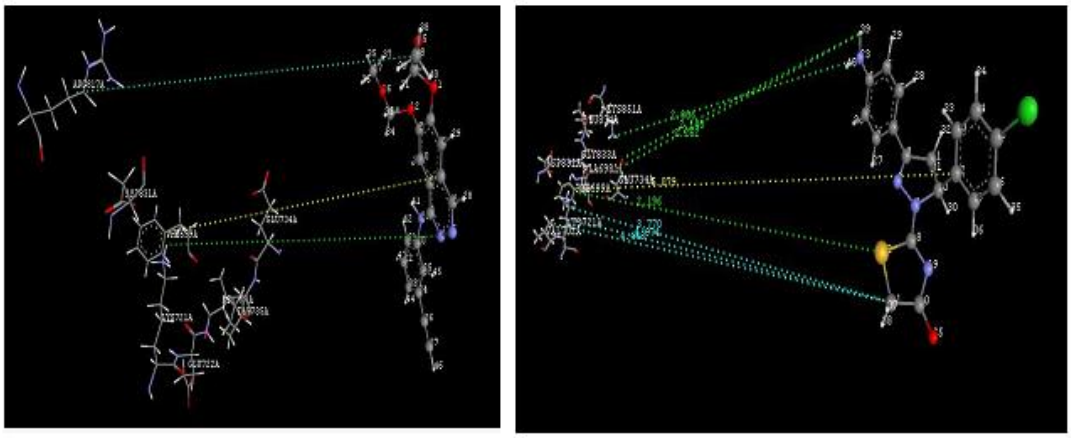

Fig.12.3D view –Molecular docking modelling of parent compound erlotinib.

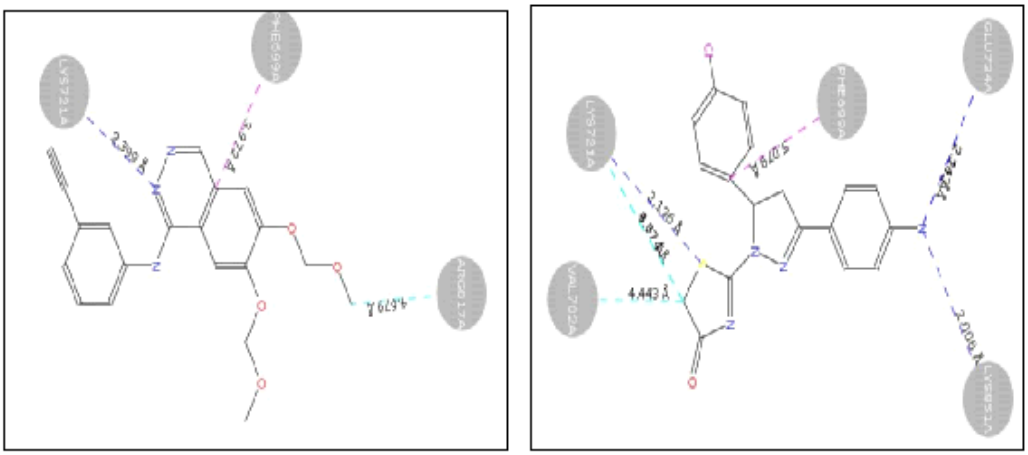

Fig.13. 2Dview- Molecular docking modelling of compound Erlotinib with EGFR kinase.

Table 1: Various substitutions of pyrazolyl-thiazolinone derivatives.

Compound	R_1	R_2	Compound	R_1	R_2
E1	H	H	E19	Br	H
E2	H	F	E20	Br	F
E3	H	Cl	E21	Br	Cl
E4	H	Br	E22	Br	Br
E5	H	Me	E23	Br	Me
E6	H	OMe	E24	Br	OMe
E7	F	H	E25	Me	H
E8	F	F	E26	Me	F
E9	F	Cl	E27	Me	Cl
E10	F	Br	E28	Me	Br
E11	F	Me	E29	Me	Me
E12	F	OMe	E30	Me	OMe
E13	Cl	H	E31	OMe	H
E14	Cl	F	E32	OMe	F
E15	Cl	Cl	E33	OMe	Cl
E16	Cl	Br	E34	OMe	Br
E17	Cl	Me	E35	OMe	Me
E18	Cl	OMe	E36	OMe	OMe

Table 2: IC $_{50}$ value represents the inhibitory activity of 36 compounds (E1-E36) against EGFR and HER-2 respectively and all are converted into *p*IC$_{50}$ value.

Comp	EGFR Activity		HER-2 Activity		Comp.	EGFR Activity		HER-2 Activity	
	IC_{50} (ug/ml)*	pIC_{50}**	IC_{50} (ug/ml)*	pIC_{50}**		IC_{50} (ug/ml)*	pIC_{50}**	IC_{50} (ug/ml)*	pIC_{50}**
E1	3.38	0.529	5.12	0.709	E19	3.20	0.505	4.87	0.688
E2	4.86	0.687	6.35	0.803	E20	6.48	0.812	8.14	0.911
E3	3.49	0.543	4.83	0.684	E21	4.12	0.615	5.87	0.769
E4	1.35	0.130	3.05	0.484	E22	2.03	0.307	3.65	0.562
E5	3.03	0.481	4.64	0.667	E23	5.58	0.747	7.04	0.848
E6	4.27	0.63	6.21	0.793	E24	7.96	0.901	9.27	0.967
E7	8.14	0.911	10.53	1.022	E25	1.08	0.033	2.24	0.35
E8	16.92	1.228	18.12	1.258	E26	2.01	0.303	3.53	0.548
E9	10.92	1.038	13.16	1.119	E27	1.66	0.22	3.11	0.493
E10	4.79	0.68	6.24	0.793	E28	0.24	-0.62	1.07	0.029
E11	8.36	0.922	10.26	1.011	E29	1.16	0.064	2.52	0.401
E12	10.69	1.029	12.43	1.094	E30	4.24	0.627	6.23	0.794
E13	5.34	0.728	7.05	0.848	E31	2.37	0.375	4.12	0.615
E14	14.21	1.153	16.42	1.215	E32	5.95	0.775	8.04	0.905
E15	8.16	0.912	9.96	0.998	E33	5.35	0.728	7.59	0.88
E16	2.28	0.358	3.84	0.584	E34	1.26	1.0	2.75	0.439
E17	6.27	0.824	8.13	0.91	E35	5.49	0.737	7.35	0.866
E18	8.58	0.933	10.34	1.015	E36	8.89	0.949	10.48	1.02

Table 3: Statistical evaluation of 2D-QSAR models of para substituted pyrazolyl-thiazolinone derivative-(For EGFR activity).

Trials	r^2	q^2	r^2se	q^2se	Pred_r^2	Pred_r^2se	F test
1(Model-1)	**0.9125**	**0.8426**	**0.1030**	**0.1381**	**0.6736**	**0.2949**	**39.63**
2	0.8350	0.7400	0.1850	0.2323	0.2825	0.2380	25.30
3	0.7816	0.6795	0.1824	0.2210	0.3511	0.3208	39.36
4	0.7392	0.6675	0.1642	0.6675	0.3605	0.3892	31.17

Table 4: Statistical evaluation of 2D-QSAR models of para-substituted pyrazolyl-thiazolinone derivative-(For HER-2 activity).

Trials	r^2	q^2	r^2se	q^2se	Pred_r^2	Pred_r^2se	F test
1(Model-2)	**0.8675**	**0.7996**	**0.1151**	**0.1416**	**0.2875**	**0.1881**	**24.88**
2	0.8518	0.7916	0.1166	0.1383	0.2546	0.1884	40.24
3	0.8107	0.7258	0.1306	0.1572	0.3288	0.1846	29.97
4	0.7961	0.6832	0.1346	0.1677	0.5968	0.1729	19.51

Table 5: Statistical evaluation of 3D-QSAR models of para-substituted pyrazolyl-thiazolinone derivatives (EGFR activity).

Trials	kNN	DOF	q^2	q^2_se	pred_r^2	pred_r^2se
1(Model-3)	**2**	**22**	**0.7219**	**0.1622**	**0.1002**	**0.4912**
2	2	21	0.7052	0.1831	0.4485	0.3547
3	2	21	0.7044	0.1857	0.4104	0.3660
4	2	22	0.6293	0.2422	0.2580	0.2939

Table 6: Statistical evaluation of 3D-QSAR models of para-substituted pyrazolyl-thiazolinone derivatives (HER-2 activity).

Trials	kNN	DOF	q^2	q^2_se	pred_r^2	pred_r^2se
1(Model-4)	**2**	**22**	**0.7185**	**0.1283**	**0.2586**	**0.2715**
2	2	21	0.7531	0.1054	0.4677	0.2685
3	2	22	0.7242	0.1186	0.3236	0.2826

Table 7: Actual and predicted activities for 36 compounds based on the best 2D/3D-QSAR models of para-substituted pyrazolyl-thiazolinone derivatives.

Compounds	Actual EGFR Activity pIC_{50}**	Predicted EGFR Activity (Model 1) pIC_{50}**	Predicted EGFR Activity (Model 3) pIC_{50}	Actual HER-2 Activity pIC_{50}**	Predicted HER-2 Activity (Model 2) pIC_{50}**	Predicted HER-2 Activity (Model 4) pIC_{50}**
E1	**0.529**	0.456	0.297	**0.709**	0.628	0.583
E2	**0.687**	0.811	0.659	**0.803**	0.955	0.855
E3	**0.543**	0.588	0.317	**0.684**	0.746	0.586
E4	**0.130**	0.315	0.524	**0.484**	0.477	0.686
E5	**0.481**	0.552	0.728	**0.667**	0.657	0.939
E6	**0.63**	0.687	0.26	**0.793**	0.777	0.422
E7	**0.911**	0.811	0.727	**1.022**	0.916	0.799
E8	**1.228**	1.171	0.893	**1.258**	1.235	1.059
E9	**1.038**	0.94	0.824	**1.119**	1.003	1.145
E10	**0.68**	0.64	0.799	**0.793**	0.777	0.422
E11	**0.922**	0.81	0.917	**1.011**	0.909	0.948
E12	**1.029**	1.271	0.339	**1.094**	1.273	1.161
E13	**0.728**	0.616	0.672	**0.848**	1.026	0.946
E14	**1.153**	0.903	0.826	**1.215**	1.118	0.954
E15	**0.912**	0.811	0.982	**0.998**	1.032	0.88
E16	**0.358**	0.375	0.469	**0.584**	0.579	0.939
E17	**0.824**	0.765	0.627	**0.91**	0.995	0.92
E18	**0.933**	0.807	0.933	**1.015**	1.049	0.939
E19	**0.505**	0.435	0.363	**0.688**	0.567	0.585
E20	**0.812**	0.865	0.764	**0.911**	0.919	0.926
E21	**0.615**	0.595	0.77	**0.769**	0.773	0.923
E22	**0.307**	0.394	0.494	**0.562**	0.574	0.54
E23	**0.747**	0.308	0.698	**0.848**	1.026	0.952
E24	**0.901**	0.726	0.986	**0.967**	0.804	0.907
E25	**0.033**	0.17	0.264	**0.35**	0.409	0.516
E26	**0.303**	0.323	0.263	**0.548**	0.558	0.42
E27	**0.22**	0.168	0.033	**0.493**	0.384	0.431
E28	**-0.62**	-0.114	0.26	**0.029**	0.096	0.422
E29	**0.064**	0.424	0.115	**0.401**	0.653	0.598
E30	**0.627**	0.548	0.524	**0.794**	0.738	0.422
E31	**0.375**	0.362	0.082	**0.615**	0.773	0.501
E32	**0.775**	0.809	0.713	**0.905**	1.176	0.993
E33	**0.728**	0.775	0.627	**0.88**	0.933	0.939
E34	**0.1**	0.123	0.097	**0.439**	0.513	0.663
E35	**0.737**	0.359	0.728	**0.866**	0.765	0.939
E36	**0.949**	0.913	0.82	**1.02**	1.186	0.934

Table 8: Newly designed molecules of para-substituted pyrazolyl-thiazolinone derivatives.

Compound	**R_1**	**R_2**	**Predicted EGFR activity**	**Predicted HER-2 activity**	**Compound**	R_1	R_2	**Predicted EGFR activity**	**Predicted HER-2 activity**
MCR 001	OH	OH	0.927	1.464	MCR 011	CH_3	NH_2	0.676	0.390
MCR 002	NH_2	H	0.895	0.523	MCR 012	NH_2	CH_3	0.893	0.520
MCR 003	H	OH	0.629	0.768	MCR 013	CH_3	OH	0.891	0.781
MCR 004	NH_2	F	0.263	1.104	MCR 014	OH	CH_3	0.927	0.996
MCR 005	NH_2	Cl	0.671	0.894	MCR 015	NH_2	Br	0.674	0.599
MCR 006	OH	H	0.927	1.011	MCR 016	Br	NH_2	0.893	0.408
MCR 007	F	OH	0.977	1.373	MCR 017	OH	Br	0.927	0.859
MCR 008	OH	OCH_3	0.927	1.380	MCR 018	Br	OH	0.916	0.811
MCR 009	NH_2	OCH_3	0.979	0.874	MCR 019	F	NH_2	0.978	0.901
MCR 010	F	OCH_3	0.927	1.245	MCR 020	Cl	NH_2	0.980	0.672

****