

DESIGN OF POTENT ANTICANCER MOLECULES COMPRISING PYRAZOLYL-THIAZOLINONE ANALOGS USING MOLECULAR MODELLING STUDIES FOR PHARMACOPHORE OPTIMIZATIONKUNAL RAUT^{1*}, SACHIN KOTHAWADE², VISHAL PANDE², SANDESH BOLE², SAMPADA NETANE², KALYANI AUTADE³, ASHVINI JOSHI¹¹Department of Pharmaceutical Chemistry, RSM's N. N. Sattha College of Pharmacy, Ahmednagar, Maharashtra, India.²Department of Pharmaceutics, RSM's N. N. Sattha College of Pharmacy, Ahmednagar, Maharashtra, India. ³Department of Pharmacology, RSM's N. N. Sattha College of Pharmacy, Ahmednagar, Maharashtra, India.

*Corresponding author: Kunal Raut; Email: rautkunal91@gmail.com

Received: 23 February 2023, Revised and Accepted: 07 April 2023

ABSTRACT

Objectives: Numerous tiny receptor tyrosine kinase inhibitors have been reported as anticancer medications over the past 10 years. However, a lot of them lack effectiveness *in vivo*, selectivity, or do not last long before developing resistance.

Methods: We used molecular modeling research to improve the pharmacophore to get beyond these limitations. For the purpose of linking the chemical makeup of pyrazolyl thiazolinone analogs with their anticancer activity, quantitative structure activity relationship (QSAR) investigations in two dimensions (2D) and three dimensions (3D) were carried out. Pyrazolyl thiazolinone pharmacophore's steric, electronic, and hydrophobic requirements were calculated using 3D QSAR.

Results: By leveraging the findings of QSAR investigations, the pharmacophore was refined and new chemical entities (NCEs) were generated. The r^2 and q^2 values obtained for the best model No. 4 of 2D QSAR were 0.9244 and 0.8701, respectively. A drug-like pharmacokinetic profile was ensured by studying the binding affinities of proposed NCEs on epidermal growth factor receptor-TK using docking studies and estimating their absorption, distribution, metabolism, and excretion features.

Conclusion: When statistical significance is closely examined, predictability of the model and its residuals (actual activity minus predicted activity) is found to be close to zero, leading us to draw the conclusion that the logic behind the design of NCEs was determined to be sound.

Keywords: Molecular modeling, QSAR study, NCE, Docking, EGFR inhibitors, ADME.

© 2023 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) DOI: <http://dx.doi.org/10.22159/ajpcr.2023v16i8.47665>. Journal homepage: <https://innovareacademics.in/journals/index.php/ajpcr>

INTRODUCTION

The phosphotyrosine kinase (PTK) or epidermal growth factor receptor (EGFR) plays a key role in cellular differentiation and proliferation. It is frequently expressed in numerous cancer types. It has shown to be a successful target for the creation of anticancer medications [1].

Small molecule RTK inhibitors have received positive press in the past 10 years as potential cancer treatments. Despite the fact that many of them have strong tyrosine kinase inhibitory actions, they frequently lack selectivity, have poor cellular and *in vivo* potency, or develop drug resistance [2]. Due to the similarities in the ATP catalytic binding sites across the kinase family, developing selective kinase inhibitors is a significant difficulty in the drug discovery and development process [3].

The creation of novel, highly active compounds is made possible by computational approaches. We were able to connect the physicochemical characteristics of chemical compounds through their biological processes thanks to quantitative structure activity relationship (QSAR) investigations, which may have provided insight into the crucial structural elements required for biological activity. The ability of QSAR approach to foretell the biological activity of new chemical entities (NCEs) created *in silico* is one of its key functions [4]. Various techniques for model creation and validation were used to boost the QSAR methodology's predictive potential.

Numerous pyrazole and thiazolinone nuclei have been identified for a variety of activities, including antiviral, antifungal, anti-inflammatory, antibacterial, analgesic, and hypoglycemic [5].

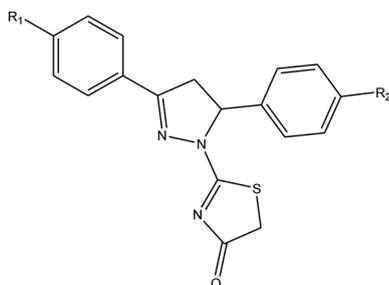
A combination pharmacophore of pyrazole and thiazolinone was described as a strong and selective EGFR-TK receptor inhibitor in various papers [6]. The development of NCEs was therefore expected to benefit from molecular modeling investigations, such as QSAR, absorption, distribution, metabolism, excretion (ADME) prediction, and docking studies on a series of pyrazolyl thiazolinone pharmacophore (NCEs).

In this study, we concentrated on using molecular modeling techniques to create effective anti-cancer chemicals and to create a model that will be able to predict the biological actions of NCEs [7-9]. It incorporates the techniques in building the main QSAR model components, notably the techniques for choosing informative descriptors, and evaluating the model for predicting anticancer activity. On a variety of compounds having the pharmacophore pyrazolyl-thiazolinone, a QSAR model was created to determine the crucial structural fragments needed around both heterocyclic rings for effective anti-proliferative activity. In the current investigations, the predictive QSAR modeling method was used to conduct 2D and 3D QSAR [10]. The NCEs were created using the best model that produced the most accurate and statistically acceptable findings. To further understand the interactions between the proposed NCEs and the EGFR target, molecular docking study was completed. To produce a well-rounded drug design, logical drug design ought to also take into account metabolic and pharmacokinetic data as well as molecular, biochemical, and pharmacological information. Therefore, to guarantee that the proposed NCEs had a drug-like pharmacokinetic profile, the final screen to identify compounds that meet Lipinski's criteria was prediction of ADME characteristics.

METHODS

2D QSAR study

The v-Life Molecular Design Suite Software®, version 3.5, was used for all QSAR investigations [11]. To create a QSAR model reported for anticancer activity, a series of 36 pyrazolyl-thiazolinone compounds (E1-E36) were selected from the literature. For the molecular modeling study, biological activity was scaled to its logarithmic value ($\text{pIC}_{50} = -\log \text{IC}_{50}$) (Table 1) [12]. The 2D draw tool of v-life MDS was used to create all derivatives. All of them were converted to 3D structures using the v-Life tool 2D TO 3D Converter, and then optimized using the Merck Molecular Force Field (MMFF) energy minimization method [13] by maintaining the electrostatic and Vanderwall cutoff at 20 and 10 Kcal/mole, the dielectric constant at 1, the maximum number of cycles allowed to be 10,000, and the convergence criteria of 0.01 (Root mean square gradient).



Experimental design for 2 Dimensional (2D) QSAR

In an effort to boost the QSAR model's prediction power and assure model resilience, the 36-molecule dataset was split into various training and test sets. The plan of the experiment Fig. 1 displays the results of our study. As a result of the randomization test, two models were chosen and given the names Training set-A and Training set-B. By dividing the remaining test set of compounds into two groups, Test set a1 and a2 for Training Set A, and Test set b1, b2 for Training Set B, respectively, these models were exposed to external validation twice. Only this model is used to estimate the activity of the molecules in the second test set (a2) if the model created by the training set satisfies all of the parameters for prediction of the first test set (a1). The entire process, starting with the selection of the training set, is repeated if the model does not fulfill the prediction parameters for the first test set (a1). For the design of NCEs, only models that pass both test sets were chosen [14].

Descriptor selection

Using the V-Life MDS program, a total of 549 different 2D descriptors were estimated, including topological, polarizability, molecular connectivity, dipole moment, and element count. By eliminating invariables (the ones that do not display variations as per physicochemical variations of the molecules chosen for the experiments), the independent variables (i.e., descriptors) were preprocessed. Only 210 descriptions were left in the spreadsheet after this.

The likelihood of a coincidence between observed and predictive activity has been shown to be high, especially when the number of independent variables (descriptors) employed in any QSAR study is comparable to or exceeds the number of compounds in the dataset [15]. Therefore, reducing the number of descriptors is a crucial step. As a result, we have employed a variety of strategies for variable (descriptor) reduction that may enhance both the performance and predictability of the QSAR model.

Correlation matrix

The relationship between the activity and the description, as well as the relationship between the descriptors themselves, was taken into consideration in the present study [16]. We only took into account the descriptors that were generated for the series of chemicals we chose and that either directly or indirectly correlate with activity. For model generation, only descriptors with activity correlations $\geq 0.3-0.7$ were considered.

Table 1: Selected series of compounds containing the pharmacophore pyrazolyl thiazolinone

Comp code	R ₁	R ₂	IC ₅₀	pIC ₅₀
E1	-H	-H	3.38	-0.52892
E2	-H	-F	4.86	-0.68664
E3	-H	-Cl	3.49	-0.54283
E4	-H	-Br	1.35	-0.13033
E5	-H	-Me	3.03	-0.48144
E6	-H	-OMe	4.27	-0.63043
E7	-F	-H	8.14	-0.91062
E8	-F	-F	16.92	-1.2284
E9	-F	-Cl	10.92	-1.03822
E10	-F	-Br	4.79	-0.68034
E11	-F	-Me	8.36	-0.92221
E12	-F	-OMe	10.69	-1.02898
E13	-Cl	-H	5.34	-0.72754
E14	-Cl	-F	14.21	-1.15259
E15	-Cl	-Cl	8.16	-0.91169
E16	-Cl	-Br	2.28	-0.35793
E17	-Cl	-Me	6.67	-0.82413
E18	-Cl	-OMe	8.58	-0.93349
E19	-Br	-H	3.2	-0.50515
E20	-Br	-F	6.48	-0.81158
E21	-Br	-Cl	4.12	-0.6149
E22	-Br	-Br	2.03	-0.3075
E23	-Br	-Me	5.58	-0.74663
E24	-Br	-OMe	7.96	-0.90091
E25	-Me	-H	1.08	-0.03342
E26	-Me	-F	2.01	-0.3032
E27	-Me	-Cl	1.66	-0.22011
E28	-Me	-Br	0.24	0.619789
E29	-Me	-Me	1.16	-0.06446
E30	-Me	-OMe	4.24	-0.62737
E31	-OMe	-H	2.37	-0.37475
E32	-OMe	-F	5.95	-0.77452
E33	-OMe	-Cl	5.35	-0.72835
E34	-OMe	-Br	1.26	-0.10037
E35	-OMe	-Me	5.46	-0.73719
E36	-OMe	-OMe	8.89	-0.9489

Fitness plot

The following are a few significant elements that we considered when selecting the appropriate descriptors for QSAR model creation:

- The number of data points distributed on either side of the regression line should be equal.
- The slope of the plot between the descriptor and the activity has been shown. Highly correlated descriptors were occasionally deleted since they had a relatively low slope.
- In the case of indicator variables (primarily for topological descriptors), the fitness plot additionally showed the frequency of a specific data point. Each data point in the plot displayed at least three members. This provided data on how frequently each specific substituent appeared in a series. The outcomes are shown in (Fig. 2).

Variance

Using descriptor variance information, variables were minimized [17]. We have thought about the relationship between descriptor and activity here. We discovered that some descriptors constantly showed considerable variance even when the physicochemical qualities did not change all that much, whereas other descriptors, such as the indicator variable, consistently displayed low variance even when there were significant physicochemical changes. This was caused by the fact that each descriptor's calculating process was unique. Descriptors with the highest variance occasionally did not exhibit a strong connection with activity. After carefully examining the results of our study, we have come to the conclusion that, because final outcomes are more dependent on correlation than on variance, one should place more emphasis on the correlation between descriptors and activity than on descriptors with the biggest variance.

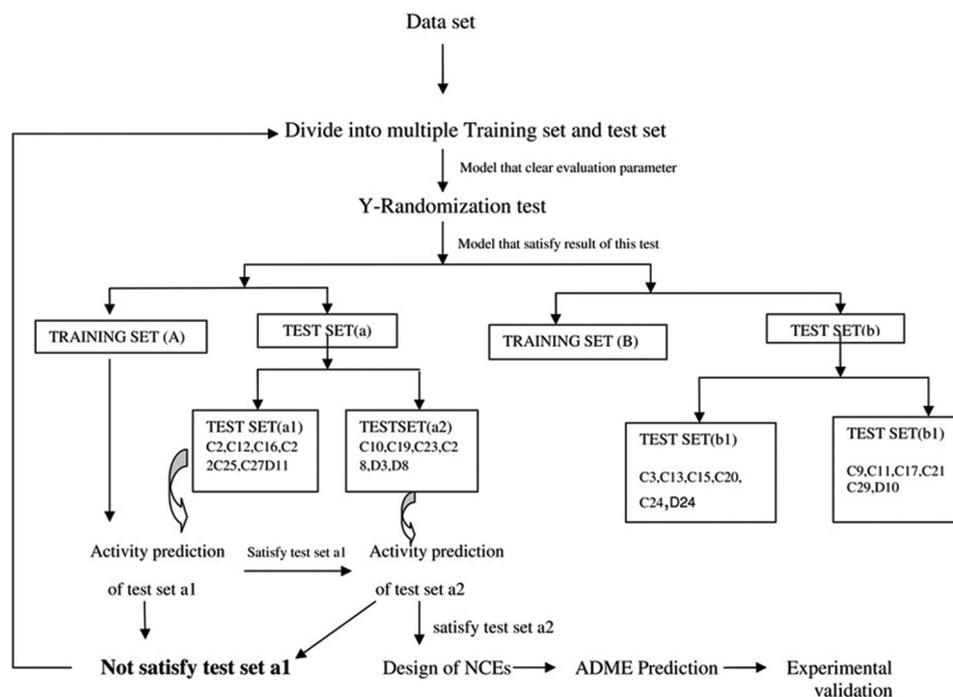


Fig. 1: Experimental design for selection of molecules in training and test sets

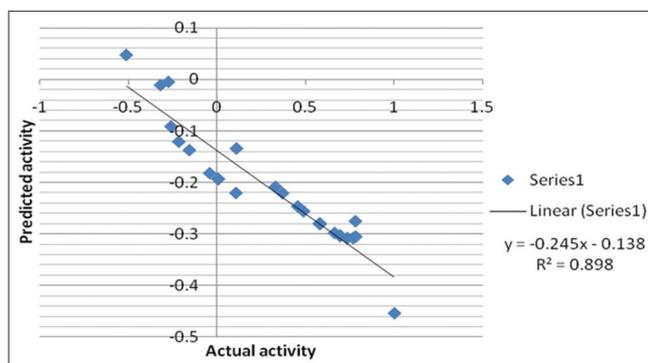


Fig. 2: Fitness plot for descriptor T_2_O_6 with biological activity

The algorithm we followed for variable reduction is as follows

1. Define the proper correlation cutoff value (A_{max}) between the description and the activity. Eliminate any descriptors with values lower than A_{max} .
2. Define the C_{max} value, which is the suitable cross correlation cutoff value between the descriptor and the descriptor. Eliminate all descriptors with values greater than C_{max} .
3. Specify the variance cutoff value for the V_{max} descriptor. Descriptors with variance values below V_{max} were eliminated.

This approach, according to our observations, reduces the number of variables by almost 75%. After that, we used several variable selection techniques to make sure that every last descriptor was significantly influencing the QSAR model. All of the generated descriptors were chosen using the aforementioned methodology and then submitted to various analytical techniques to create models.

QSAR model generation

Using the MLR (multiple linear regressions) approach, QSAR models were created by manually and randomly picking training sets. The methods of step-wise forward-backward and molecular simulation

were employed to choose the variables. The regression model (Equation 1), which describes the activity dependent variable, y , as a linear combination of the X -variables with the coefficients b , is typically fitted using MLR. Remainings, which are represented by the letter e , are the variations between the data (y) and the model (Xb). The obvious relationship between these parameters is shown by the subsequent equation.

$$y = Xb + e \quad (1)$$

3D QSAR by SA-kNN method

k-Nearest Neighbor molecular field analysis (kNN-MFA) studies of 3D-QSAR were carried out utilizing Simulated Annealing (SA) as a variable selection approach [18]. Utilizing k-Nearest Neighbors, there are three steps to calculating grid points: (1) Determine the separation between the k training set known objects and the u unknown object, (2) choose k objects from the training set that are nearby the unidentified object (u), and (3) assign your object to the same group as the majority of your other objects. After optimization, the given set of molecules must be properly aligned according to the kNN-MFA approach; the alignment was done using a template-based alignment method. Molecular alignment was used to view the variety of structural features in the supplied set of molecules. A typical rectangular grid was then generated to surround the molecules. Using a methyl probe with a charge of +1, hydrophobic and electrostatic interaction energies at the grid's lattice points were calculated. The 3D-QSAR model was then constructed using the resulting set of aligned molecules.

Model evaluation

The created QSAR models were assessed using the subsequent statistical metrics: n , the number of molecular observations; k , the number of variables; q^2 , cross-validated r^2 (via leave-one-out [LOO]); $\text{pred } r^2$, r^2 for external test set; r^2 , coefficient of determination; Z score is the result of the randomization test, and the best $\text{rand } q^2$ and best $\text{rand } r^2$ values represent the highest q^2 and highest r^2 values, respectively. SEP stands for standard error of external test set prediction, SEE for standard error of estimate of the model, and SECV for standard error of cross-validation.

Model validation

Internal validation

Internal validation was performed using the (q^2 , LOO) approach [19]. Each molecule in the training set was successively deleted to determine q^2 , after which the model was trained using the same descriptors and the biological activity of the removed molecules was predicted. This effort was conducted to determine the QSAR model's robustness. Every cross-validation study was carried out with the assumption that $q^2 > 0.5$.

External validation

Predicting the activity of a test set of chemicals was done as an external validation of the produced models. As seen below, the pred r^2 value is computed.

$$\text{Pred}_r^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \hat{y}_{\text{mean}})^2} \quad (2)$$

Where y_i and \hat{y}_i are the actual and predicted activity of the i^{th} molecule in the test set, respectively, and y_{mean} is the average activity of all molecules in the training set.

It should be more than 0.5.

Randomization test

This is a practical method to avoid random connection [20]. This approach involves repeatedly permuting the response variable while maintaining the x-variable unaltered. R^2 and q^2 are recorded following each permutation. We may say with some confidence that the original QSAR model is real and was not created by coincidence if the r^2 and q^2 values in each case are significantly lower than the original data.

In our study, we have calculated Z-score to check significance of the model. Following formula was used for the same.

$$\text{Zscore} = \frac{(q^2_{\text{org}} - q^2_{\text{a}})}{q^2_{\text{std}}} \quad (3)$$

Where q^2_{org} is the q^2 value calculated for the actual data set, q^2_{a} is the average q^2 , and q^2_{std} is the standard deviation of q^2 , calculated for various iterations using different randomized data sets.

New chemical entity (NCE) design

The pyrazolyl thiazolinone pharmacophore was optimized using the data from 2D and 3D QSAR experiments to create NCEs with strong anti-cancer efficacy. Using the CombiLib tool of the v-Life MDS 3.5 Software, a new combinatorial library was created. According to a literature review of drug development to date, between 80 and 85% of compounds fail at a later stage of drug discovery due to their inability to display a pharmacokinetic profile consistent with that of a drug. Therefore, we have submitted created NCEs to Lipinski's screening (Rule of Five) to confirm that their pharmacokinetic profiles are similar to those of drugs [21]. The following parameters were used as Lipinski's screen/filters (Values in parenthesis indicate ideal requirements).

1. Number of Hydrogen Bond Acceptor (A) (<10)
2. Number of Hydrogen Bond donor (D) (<5)
3. Number of Rotatable Bond (R) (<10)
4. XlogP (X) (<5)
5. Molecular weight (W) (<500 g/mol)
6. Polar surface area (S) (<140 Å).

Molecular docking studies

The most promising developed NCEs were docked using the molecular docking tool GLIDE (Glide version 5.0) into the receptor tyrosine kinase enzyme binding pocket. Ligprep, which includes the "Ligand preparation wizard" and "Protein preparation wizard" in Maestro wizard 9.0 of the Schrodinger software, was used to prepare all structures for docking.

A controlled impact minimization of the cocrystallized compound was done in the refining component. This aids in the hydroxyl group's side chain reorientation. For this, it makes use of the OPLS-AA force field. The ligand was centered in the crystal structure to determine the length of the grids. The ligands were constructed using the maestro structure builder panel and prepared using the ligprep module, which creates the lowest energy conformers of ligands using the MMFF94 force field. It was extended up to 23A0. Extra precision docking mode was used to pick the ligands' lower energy conformations and dock them into the grid created by the protein structure. Except for the protein's active site, which has a minor degree of flexibility, the ligands and receptors in this docking approach are both flexible. Designed NCEs were also docked using the v-life MDS 3.5 docking tool, and the dock score was recorded.

Prediction of ADME and Central nervous system (CNS) Toxicity properties

Using the Schrodinger software's Qikprop tool, the ADME properties were determined. Both physicochemically significant descriptors and pharmacokinetically significant characteristics are predicted. To guarantee a drug-like pharmacokinetic profile when utilizing rational drug design, it also determines whether analogs are acceptable using Lipinski's [22] rule of five.

RESULTS AND DISCUSSION

2D QSAR

2D QSAR models

After applying MLR method, the five meaningful descriptor were generated, namely, H Acceptor Count, T_2_Br_7, Chi3Cluster, T_2_O_6, and Chlorine count (Fig. 3a).

Out of these five descriptors, T_2_O_6 shows up to 28% contribution with activity as shown in Eq. No. 4.

$$\text{pIC}_{50} = (-0.245) \text{T}_2\text{O}_6 - 0.138 \quad (4)$$

$$r^2 = 0.89, q^2 = 0.65, F\text{-test} = 23.21, \text{Pred}_r^2 = 0.71$$

We learned from the findings that the T 2 O 6 descriptor alone satisfies all assessment criteria. The aforementioned descriptions exhibit the strongest relationship with activity and a balanced distribution of data points. We have created various combinations of the chosen descriptors while keeping T 2 O 6 as a constant descriptor to boost the prediction capability.

Three training sets A, B, and C (using an experimental design) were given multiple models, out of which two were chosen for training set A and training set C based on statistical significance. For Training set-A and C, respectively, outcome shown Fig. 3b and c and in Tables 2 and 3 display the outcomes of all these models.

It noticed that Test set c2 containing derivatives (E1, E2, E7, E11, E15, E19, E21, and E32) gives good result. On the basis of the statistical parameters, namely, $r^2 > 0.7$, cross-validated r^2 , that is, $q^2 > 0.7$ and parameter to assess external validation, that is, $\text{pred}_r^2 > 0.5$; the generated regression equation of model 4 of training set C was used for further studies. Following regression equation was used to design NCEs.

$$\text{pIC}_{50} = -0.247499(\text{H-Acceptor Count}) + 0.132802 (\text{T}_2\text{Br}_7) - 0.554286 (\text{chi3Cluster}) + 0.279116(\text{T}_2\text{O}_6) - 0.11112 (\text{Chlorines Count}) + 0.00182425 \quad (5)$$

Uni-column statistics

Table 4 reports the obtained comparative unicolun statistical parameters for the training and test sets. While the minimum activity value of the test set was more than the minimum of the training set in all models, the maximum activity value of the test set was less than the maximum of the training set. The test sets a1, c1, and a2, c2 as well as training sets A and C's standard deviation were discovered to be remarkably similar. This

Table 2: Test set a₁, a₂ statistical parameters of developed QSAR models

Training set A	Test set a ₁		Test set a ₂	
	Model		Model	
	1	2	3	4
r ²	0.8632	0.7001	0.8750	0.8101
q ²	0.7245	0.5475	0.7887	0.6663
F test	20.1893	10.5030	22.4102	13.6535
r ² se	0.1613	0.2267	0.1724	0.2152
q ² se	0.2290	0.2785	0.2241	0.2853
Pred_r ²	0.6111	0.6374	0.8166	0.5245
Pred_r ² se	0.2212	0.2140	0.1556	0.2367
Best rand r ²	0.41344	0.33649	0.36170	0.19695
Best rand q ²	0.13126	-0.11025	-0.15391	-0.32574
Z Score _{ran_r²}	5.96253	3.59846	4.85464	5.75572
Z Score _{ran_q²}	5.53726	4.05107	4.91963	5.68119
α _{ran_r²}	0.00000	0.00100	0.00002	0.00000
α _{ran_q²}	0.00000	0.00041	0.00001	0.00000
Descriptors	T_2_Br_7	HAcceptorCount	HAcceptorCount	HAcceptorCount
	T_C_N_6	ChlorinesCount	T_2_Br_7	ChlorinesCount
	T_T_T_7	T_2_Br_7	T_2_Cl_5	T_2_Br_7
	T_2_Cl_5	T_2_Br_5	T_2_F_5	T_2_F_5
	T_2_Br_5		T_2_Br_5	T_2_Br_5
Coefficient	0.246232	-0.286907	-0.255198	-0.253403
	0.239912-0	-0.349796	0.15707	-0.393246
	.0626532-0	0.138057	-0.536102	0.141559
	.342318-0.	-0.275964	-0.479	-0.471403
	234524		-0.38092	-0.311254

Table 3: Statistical parameters of developed QSAR models for external test set c₁, c₂

Training set C	Test set c ₁		Test set c ₂	
	Model		Model	
	1	2	3	4
r ²	0.8157	0.9244	0.8814	0.9244
q ²	0.6791	0.8701	0.8050	0.8701
F test	15.0522	39.1315	29.7194	39.1315
r ² se	0.2008	0.0981	0.1099	0.0981
q ² se	0.2650	0.1285	0.1409	0.1285
Pred_r ²	0.5738	0.5721	0.7827	0.7721
Pred_r ² se	0.2339	0.1156	0.1355	0.1156
Best rand r ²	0.58423	0.42601	0.36384	0.42601
Best rand q ²	0.28204	0.02785	0.04835	0.02785
Z Score _{ran_r²}	3.48120	5.68977	5.93013	5.68977
Z Score _{ran_q²}	3.29056	4.97327	4.98157	4.97328
α _{ran_r²}	0.00100	0.00000	0.00000	0.00000
α _{ran_q²}	0.00100	0.00001	0.00001	0.00001
Descriptors	H AcceptorCount	HAcceptorCount	HAcceptorCount	HAcceptorCount
	T_2_Br_7	T_2_Br_7	T_2_Br_7	T_2_Br_7
	slogp	chi3Cluster	chi3Cluster	chi3Cluster
	T_2_Cl_5	T_2_O_6	T_2_O_6	T_2_O_6
	chi3Cluster	ChlorinesCount		ChlorinesCount
Coefficient	-0.441874	-0.247499	-0.230682	-0.247499
	0.228188	0.132802	0.130664	0.132802
	-0.478782	-0.554286	-0.553897	-0.554286
	-0.256814	0.279116	0.259916	0.279116
	0.405305	-0.11112		-0.11112

Table 4: Unicolumn statistics for training set and test set

Column name	Training set A	Test set a ₂	Training set B	Test set b ₁	Training set C	Test set c ₂
Average	-0.5814	-0.7631	-0.5297	-0.6670	-0.7006	-0.7318
Max	0.6198	-0.0645	0.6198	-0.0334	-0.1004	-0.5051
Min	-1.2284	-1.0382	-1.2284	-0.9222	-1.2284	-0.9222
SD	0.4256	0.3172	0.4145	0.2787	0.3113	0.3035
sum	-12.7912	-6.1044	-11.6526	-5.3362	-15.4128	-5.8546

Table 5: Comparison of the various statistical results of 3D QSAR generated by SA kNN-MFA methods

Statistical parameter	Simulated annealing-k nearest neighbor molecular field analysis	
	Training set A	Training set B
	Model A	Model B
q^2	0.5628	0.5899
q^2_{se}	0.2550	0.2477
$Pred_r^2$	0.7947	0.5579
$Pred_r^2_{se}$	0.1690	0.2383
N	28	28
K nearest neighbor	2	4
Contributing descriptors	E_477 (-1.05499, -0.81627) H_271 (0.115775, 0.129451) E_20 (0.187479, 0.250906) E_115 (0.377103, 0.400051)	E_75 (-1.72254 0.794403) E_358 (3.71312 6.75997) S_178 (-0.108042-0.105661) E_115 (0.346376 0.432693)

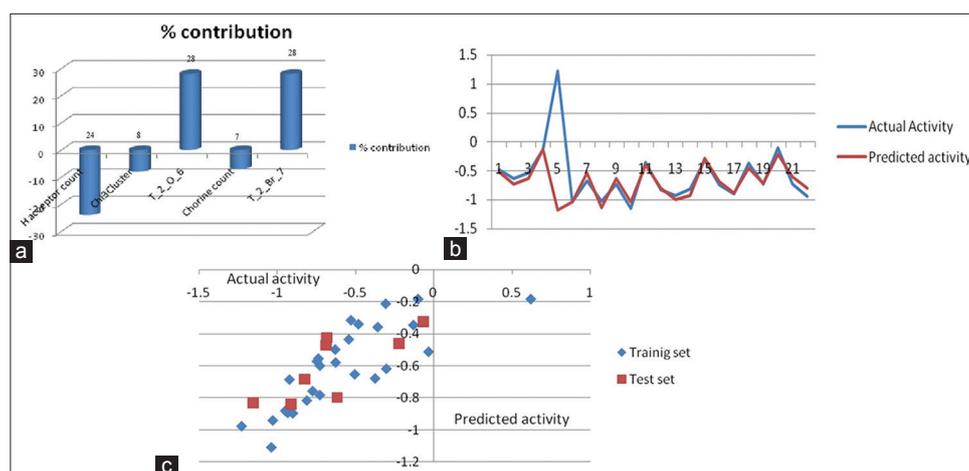


Fig. 3: (a) Contribution plot of significantly contributed descriptors. (b) Plot of actual versus predicted activity for training set A. (c) Plot of actual versus predicted activity for training set C

showed that despite the fact that the molecules selected for the training or test set are different, the distribution pattern for the biological activity of the molecules in the two selection processes is strikingly comparable.

Interpretation of 2D QSAR

The present QSAR model reveals that topological T_2_O_6 descriptor has major contribution in explaining variation in activity. Descriptors T_X_Y_Z can be defined as count of fragments formed with atom types X and Y separated by topological distance of Z bonds.

The definition for the descriptors with their % contribution in QSAR models is given:

H-acceptor count (24%)

This descriptor is negatively contributing. It signifies that number of hydrogen bond acceptor atoms should be less in designed NCEs.

chi3Cluster (8%)

This descriptor is negatively contributing toward activity. This descriptor signifies simple third order cluster chi index in a compound. It belongs to molecular connectivity indices which describes about required group connection like linear or branched or tertiary in substitution. This descriptor indicates that substitution by methyl or ethyl group at R2 position decreases anticancer activity.

Chlorines count (7%)

This descriptor signifies that number of chlorine atoms in a compound should be less as it is negatively contributing towards activity.

It indicates that by avoiding chlorine substitution at R₁ and R₂ position, one can increase anticancer activity.

T_2_O_6 (28%)

This is the count of number of double bounded atoms (i.e., any double bonded atom, T₂) separated from Oxygen atom by six bonds in a molecule. This descriptor contributes positively. It indicate that substitution by hydroxy or oxygen containing group like methoxy and ethoxy at R2 position separated from doubly bounded carbon atom may increase anticancer activity. Essential presence of keto oxygen of thiazolinone ring can be explained by this descriptor.

T_2_Br_7 (28%)

This is the count of number of double bounded atoms (i.e., any double bonded atom, T₂) separated from Bromine atom by seven bonds. This descriptor contributes positively. It describes that the presence of bromine substitution at both R₁ and R₂ position is positively contributing toward activity.

3-Dimensional (3D) QSAR studies

In Model A, error occurred in predictivity was low ($pred_r^2=0.1690$) and predictive power of model B ($pred_r^2=0.5579$) which is inferior to Model A in terms of quality. The q^2 , $pred_r^2$, and k values of model B (SA kNN MFA model A) were found to be 0.5899, 0.5579, and 4, respectively. Although the Model B demonstrated good internal and external predictability, there was a higher error in forecasting the activity when compared to Model A. ($q^2_{se}=0.2550$, $pred\ r^2=0.7947$). Fig. 4 shows the 3D data points produced around a rectangular grid using the SA KNN-MFA Model A, together with the range of

contribution noted in parenthesis, statistical results are shown in Table 5.

Interpretation of 3D QSAR model

The electrostatic, steric, and hydrophobic requirements around the pyrazolyl thiazolinone pharmacophore were optimized using 3D-QSAR. Points generated in SA kNN-MFA 3D-QSAR model(A) are E_477 (-1.05499, -0.81627), H_271 (0.115775, 0.129451), E_20 (0.187479, 0.250906), E_115 (0.377103, 0.400051), that is, electronic and hydrophobic interaction at lattice points 477, 271, 20, and 115, respectively. Negative values in electrostatic data point indicate that electronegative potential (group) is preferred at that position to increase activity, positive range indicates that group that impart positive electrostatic potential is favorable for activity so electropositive (electron withdrawing) group is preferred at that position.

In Model A, small positive hydrophobic value H_271 (0.115775, 0.129451) indicates that the less hydrophobic groups are required to increase activity at R1substitution. Positive and negative values in electrostatic field descriptors E_477 (-1.05499, -0.81627), E_20 (0.187479, 0.250906), and E_115 (0.377103, 0.400051) indicated the requirement of positive and negative electrostatic potential, respectively, at R₁ and R₂ substitution to enhance the anticancer activity of pyrazolyl thiazolinone derivatives. Thus KNN-MFA models helped us to identify various local interacting molecular features responsible for activity variation and hence provided direction for design of new molecules in a convenient way.

Designing NCEs with the pharmacophore pyrazolyl thiazolinone

The information obtained from 2D to 3D QSAR studies has helped a lot in optimizing pyrazolyl thiazolinone pharmacophore and for design of NCEs. Substitution pattern required around pharmacophore is showed in Fig. 5 was used to design NCEs. Fig. 6 shows common template used to design NCEs. Results of combiLib generated are shown in Table 6. All designed NCEs showed reduced hydrogen bond acceptor (A) count and increased hydrogen bond donor count (D) as required in 2D QSAR. All NCEs scored 6 (ADRXWS) for Lipinskis screen. NCEs showed rotatable bond (R), partition coefficient (X), molecular weight (W), and polar surface area(S) in acceptable range. Designed NCEs showed increased predicted activity than most potent of original series (E 28). Top 14 designed NCEs were selected for further screening.

Molecular docking studies

The receptor tyrosine kinase enzyme binding pocket was studied using the molecular docking software GLIDE® (Schrodinger Inc., USA). The RCSB protein data library provided the EGFR-PTK crystal structure. (1M17 in PDB) [23]. Erlotinib mimics ATP and binds to the ATP-binding area of the kinase active site, it was discovered. Two significant hydrogen-binding interactions between the purine base of ATP and the protein backbone between amino acids Gln-767 and Met-769 are involved in the

binding of ATP itself. In the cleavage created between the two lobes, ATP bonds. Four significant structural components make up the cleft:

Activation loop, hinge region, catalytic site, and kinase specificity pocket. For ATP and inhibitor binding to the ATP site, the hinge's interaction with

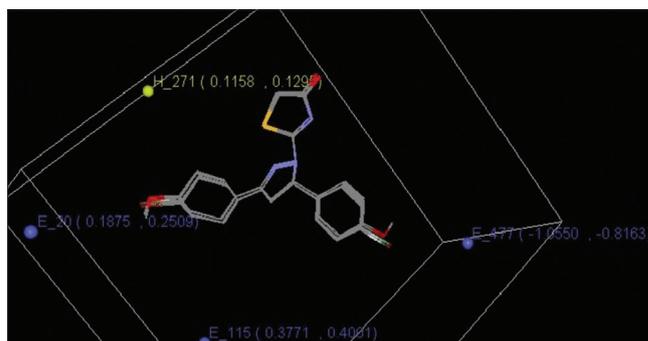


Fig. 4: The 3D rectangular grid of data points created using the kNN-MFA approach (3D-QSAR) demonstrates the contributions of electrostatic, steric, and hydrophobic functional groups for considerable anticancer action

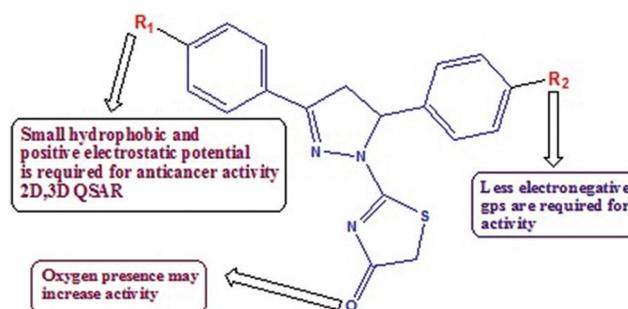


Fig. 5: Pharmacophore requirements around pyrazolyl thiazolinone

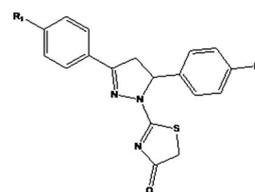


Fig. 6 Common template used to design NCEs

Table 6: Structures of designed NCEs along with predicted activity obtained by MLR equation generated by 2D-QSAR models

Molecule name	R ₁	R ₂	H accept bond	H donnr Bond	xLogP	M.W.	Polar surface area	screen result	Screen score	Predicted activity
D 1	-H	-allyl	1	1	4.542	342.3166	83.75	ADRXWS	6	0.9542
D 2	-vinyl	-NH ₂	2	1	3.258	344.3124	86.99	ADRXWS	6	0.9088
D 3	-vinyl	-F	2	0	3.659	349.304	83.75	ADRXWS	6	0.8385
D 4	-vinyl	-OH	2	1	3.798	346.305	92.98	ADRXWS	6	0.8228
D 5	-H	-NH ₂	2	2	3.952	320.2904	86.99	ADRXWS	6	0.8023
D 6	-H	-COCl	2	1	4.004	334.294	100.82	ADRXWS	6	0.7918
D 7	-CH ₃	-NH ₂	2	2	4.125	332.3014	86.99	ADRXWS	6	0.7848
D 8	-H	-OH	2	1	3.745	322.283	92.98	ADRXWS	6	0.6579
D 9	-CH ₃	-OH	2	1	3.652	334.294	92.98	ADRXWS	6	0.6058
D 10	-NH ₂	-F	3	1	4.023	339.2888	86.99	ADRXWS	6	0.5954
D 11	-NH ₂	-OH	3	2	3.487	336.2898	96.22	ADRXWS	6	0.5885
D 12	-F	-OH	3	1	4.033	341.2814	92.98	ADRXWS	6	0.5754
D 13	-NO ₂	-OH	3	2	3.913	352.1547	112.32	ADRXWS	6	0.5698
D 14	-NO ₂	-F	3	2	4.054	383.365	98.26	ADRXWS	6	0.5618

Table 7: Results of molecular docking studies performed using extra precision mode of glide

Molecule name	R ₁	R ₂	G score	Hydrogen Bond	Contacts			v-Life Dock score
					Good	Bad	Ugly	
D 1	-H	-allyl	-7.16202	1	195	2	0	-6.4354
D 2	-vinyl	-NH ₂	-6.12965	1	203	2	0	-6.3567
D 3	-vinyl	-F	-6.406401	1	200	2	0	-5.2658
D 4	-vinyl	-OH	-7.866976	1	250	2	0	-6.1457
D 5	-H	-NH ₂	-7.764808	2	233	2	0	-6.6927
D 6	-H	-COCl	-5.765891	2	226	4	0	-5.5687
D 7	-CH ₃	-NH ₂	-6.907455	2	274	1	0	-5.7513
D 8	-H	-OH	-7.218316	1	198	6	0	-6.1462
D 9	-CH ₃	-OH	-7.978312	2	242	2	0	-6.9662
D 10	-NH ₂	-F	-7.597398	2	207	5	0	-6.9654
D 11	-NH ₂	-OH	-6.402405	2	213	5	0	-5.2358
D 12	-F	-OH	-7.968543	1	208	1	0	-6.9507
D 13	-NO ₂	-OH	-7.839369	1	215	1	0	-6.2465
D 14	-NO ₂	-F	-6.745977	1	265	1	0	-5.2138
Erlotinib	-	-	-8.575483	2	344	5	0	-8.4845

the Met-769 backbone -NH is crucial [24]. The kinase specificity pocket is a pocket lined by side chains of residues Met742, Cys751, Leu764, Thr766, Thr830, Phe832, and part of Lys721. ATP does not interact with residues within this pocket, but the pocket is important for the binding of inhibitors at the ATP site. The moieties interacting with this pocket modulate inhibitory activity of ATP site antagonists [25]. The final evaluation was done with glide score, that is, G-score (docking score) mentioned in eq.6 and single best pose was generated as the output file for particular ligand. Docking results are shown in Table 7.

$$G \text{ Score} = 0.065 * \text{vdW} + 0.130 * \text{Coul} + \text{Lipo} + \text{Hbond} + \text{Metal} + \text{BuryP} + \text{RotB} + \text{Site} \quad (6)$$

Where, vdW: Van der Waal energy; Coul: Coulomb energy;

Lipo: Lipophilic contact term; HBond: Hydrogen-bonding term;

Metal: Metal-binding term; BuryP: Penalty for buried polar groups;

RotB: Penalty for freezing rotatable bonds;

Site: Polar interactions at the active site; and

The coefficients of vdW and Coul are: a=0.065, b=0.130.

We redocked Erlotinib into the enzyme's active site and then replaced it with our compounds to compare the binding modes of both the standard ligand Erlotinib and the test compounds based on our understanding of the structures of similar active sites (Fig. 6). These docking experiments have shown that the NH of Met-769's backbone connects with the oxygen of the hydroxy group in the R2 position through a hydrogen bond, while in certain ligands, the oxygen of the keto group of the thiazolinone ring binds with the NH of Met-769. The created NCEs had a comparable binding mechanism to erlotinib. Hydrogen bond, it was shown that there was a distance between amino acids and NCEs of between 1.8 and 2.03 Å, which is comparable to that of erlotinib. These interactions show the significance of the hydroxy and keto groups for binding and the ensuing inhibitory capability. Similar to Erlotinib, the R1 substituted phenyl ring of the pyrazolyl pharmacophore is located in the same deep hydrophobic binding pocket (Fig. 7). Some NCEs exhibit binding to Leu 764 in the kinase pocket of the receptor with a hydrogen bond distance of 2.12–2.213 Å. The Vander walls radii of good connections should match experimentally measured values. Those contacts that are experimentally inappropriate are bad contacts. In comparison to erlotinib, all NCEs displayed good, poor, and ugly contacts within range Key interactions of functional group of designed NCEs with specific amino acid residue are shown in Table 8. Each NCE displayed a strong v-life dock score. According to a docking research, NCEs block the EGFR TK receptor.

Table 8: Key interactions of functional group of designed NCEs with specific amino acid residue

Comp	R ₁	R ₂	H-bond interaction with distance in Å	Bonding group	
				Functional group of molecule	Amino acid
D1	-H	-Allyl	Met769 2.089	C=O	N-H
D2	-vinyl	-NH ₂	Met769 2.061	N-H	C=O
D7	-CH ₃	-NH ₂	Met769 1.996	C=O	NH
			LEU 764 2.456	N-H	C=O
D11	-NH ₂	-OH	Met769 2.017	C=O	N-H
			LEU 764 2.295	O-H	O=C
D12	-F	-OH	Met769 1.891	O-H	N-H
D13	-NO ₂	-OH	Met769 2.107	O-H	N-H
D14	-NO ₂	-F	Met769 2.028	C=O	N-H

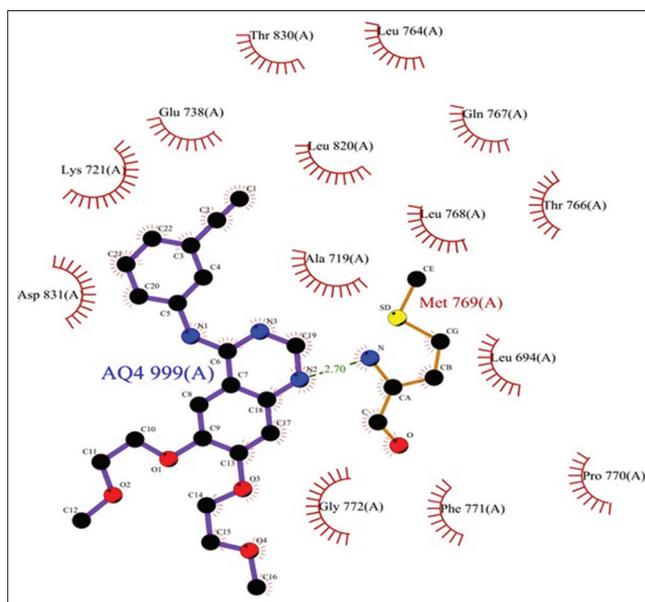


Fig. 7: Binding pose of compound D1, D2, D7, D11, D12, erlotinib, and erlotinib pdb sum in receptor binding pocket

ADME and CNS toxicity predictions

Sometimes substances that have extremely high activity *in vitro* assays fail to exhibit any activity *in vivo* or turn out to be extremely hazardous

Table 9: Results of ADME properties prediction for designed NCEs using Qikprop 2.2 tool of schrodinger

R ₁	R ₂	Mol. Wt. (g/mol)	Donor HB	Accept HB	QPlog Po/w	% Human Oral Absorption	CNS	No.of possible metabolites
-H	-allyl	378.589	1	5.25	3.842	95.325	-1	3
-vinyl	-NH ₂	370.698	1	5.25	3.974	96.237	-1	4
-vinyl	-F	363.433	1	5.25	3.743	100	-1	4
-vinyl	-OH	355.386	1	5.25	3.427	93.888	-1	4
-H	-NH ₂	363.433	1	5.25	3.581	95.336	-1	4
-H	-COCl	350.437	1.5	5.5	3.49	96.661	-2	5
-CH ₃	-NH ₂	382.393	1	6.25	2.428	81.237	-2	5
-H	-OH	382.393	1	6.25	2.547	84.149	-2	5
-CH ₃	-OH	384.384	0	5.5	3.299	87.657	-2	4
-NH ₂	-F	350.437	1.5	5.5	3.254	90.986	0	5
-NH ₂	-OH	355.386	1	5.25	3.602	100	-2	4
-F	-OH	384.384	0	5.5	3.376	88.761	-1	4
-NO ₂	-OH	385.126	1	5.6	3.568	85.687	-1	4
-NO ₂	-F	354.564	1.5	5.4	3.879	82.84	-1	4
Erlotinib	-	393.441	1.5	7.4	4.378	100	0	6

in *in vivo* models. Unwanted pharmacokinetic characteristics may be to blame for the lack of *in vivo* activity, and reactive metabolite production may be the cause of toxicity. Thus, all designed compounds were filtered by predicting their ADME and CNS toxicity properties by use of Qikprop 2.2 Tool of, Schrodinger [28]. Lot many numbers of properties of designed analogues were predicted by Qikprop tool but here, we have reported some descriptors which contribute significantly for predicting pharmacokinetic profile of the molecule. These properties are as follow: Figures in parenthesis indicated ideal values in order the test compounds to have drug like pharmacokinetic properties [26].

Rule of five: It includes (1) Molecular Weight (mol_MW) (<500), Predicted octanol/water partition coefficient. (QPlogPo/w <5), estimated number of hydrogen bond donor (donorHB ≤5), estimated number of hydrogen bond acceptor (acptHB≤10), Polar surface area (<140Å²). Compounds that satisfy these rules were expected to have drug like pharmacokinetic profile. (2) Brain/blood partition coefficient (CNS) (-2-2). (3) Percent human oral absorption (>80% is high, <25% is poor). (4) Number of possible metabolites (should range from 1 to 8).

The results of Qikprop studies are reported in Table 9. The normal range for the CNS parameter is -2 to +2, where -2 indicates no CNS penetration and +2 indicates active CNS penetration and hence CNS toxicity. The CNS parameter is connected to the absorption of entities over the blood brain barrier. All of the planned entities produced results that were acceptable and within the range. A high oral bioavailability profile and a percentage of oral absorption indicate that oral administration is an appropriate route of medication administration for patient compliance. Here, oral absorption rates for all newly developed chemical entities exceeded 80%. Every designed NCE exhibits minimal or no CNS toxicity. In donor and recipient, it was discovered the range of hydrogen bonding. NCEs demonstrated the creation of 1-6 metabolites, indicating that derivatives are less toxic and excretable from the body. Therefore, it is important to note that all developed molecules have a pharmacokinetic profile similar to that of a medication.

CONCLUSION

The goal of the present study was to optimize the chosen pyrazolyl thiazolinone pharmacophore using molecular modeling studies and it was discovered that this goal had been accomplished because the predicted activity of NCEs was found to be significantly greater (D1=0.95) than the most potent compound of the original series (E28=0.61). The resulting QSAR models were found to have produced statistically significant good results, that is, ($r^2 > 0.8$), cross-validation ($q^2 > 0.7$), and external validation (pred $r^2 > 0.7$), which suggests high predictive capacity of all models. All things considered, it is important to note that the reasoning behind the optimization of the pyrazolyl

thiazolinone pharmacophore employing 2D, 3D QSAR, ADMET, and molecular docking investigations was found to be quite correct.

The most promising compounds will next be put through additional wet laboratory testing, including synthesis, structural characterization, and biological evaluation of anticancer screening using the MCF-7 cell line.

AUTHOR CONTRIBUTIONS

Kunal Raut contributed to design of the research. Material preparation, writing the manuscript, editing, and designed the table and figures done by Sandesh Bole and Sampada Netane. Sample characterization with spectroscopy was done by Ashvini Joshia and Kalyani Autade. Dr. Vishal Pande and Dr. Sachin Kothawade were involved in supervision of all work. All authors read and approved the final manuscript.

CONFLICTS OF INTERESTS

No conflicts of interest were reported by the author in this experimental work.

AUTHORS FUNDING

No funding was received during the study. The study was done through self-finance.

REFERENCES

- Levitzi A. Protein tyrosine kinase inhibitors as novel therapeutic agents. *Pharmacol Ther* 1999;82:231-9. doi: 10.1016/s0163-7258(98)00066-7, PMID 10454200
- Caffrey DR, Lunney EA, Moshinsky DJ. Prediction of specificity-determining residues for small-molecule kinase inhibitors. *BMC Bioinformatics* 2008;9:491. doi: 10.1186/1471-2105-9-491, PMID 19032760
- Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer* 2009;9:28-39. doi: 10.1038/nrc2559, PMID 19104514
- MacKerell AD, Roux B. In: Becker OM, Watanabe M, editors. *Computational Biochemistry and Biophysics*. New York: Marcel Dekker; 2001.
- Menozzi G, Mosti L, Fossa P, Mattioli F, Ghia MW. ω-Dialkylaminoalkyl ethers of phenyl-(5-substituted 1-phenyl-1H-pyrazol-4-yl) methanols with analgesic and anti-inflammatory activity. *J Heterocycl Chem* 1997;34:963-8. doi: 10.1002/jhet.5570340339
- Lv PC, Li HQ, Sun J, Zhou Y, Zhu HL. Synthesis and biological evaluation of pyrazole derivatives containing thiourea skeleton as anticancer agents. *Bioorg Med Chem* 2010;18:4606-14. doi: 10.1016/j.bmc.2010.05.034, PMID 20627597
- Dudek AZ, Arodz T, Gálvez J. Computational methods in developing quantitative structure-activity relationships (QSAR): A review. *Comb Chem High Throughput Screen* 2006;9:213-28. doi: 10.2174/138620706776055539, PMID 16533155

8. Liu Q, Wang HG. Anti-cancer drug discovery and development: Bcl-2 family small molecule inhibitors. *Commun Integr Biol* 2012;5:557-65. doi: 10.4161/cib.21554, PMID 23336025
9. Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform* 2010;29:476-88. doi: 10.1002/minf.201000061, PMID 27463326
10. Consonni V, Todeschini R, Puzyn T, Leszczynski J, Cronin MT. Recent advances in QSAR studies-methods and applications. 2010;8:29-93.
11. VLifeMDS. Molecular Design Suite Version 3.5. Pune, India: V-life Sciences Technologies Pvt. Ltd.; 2004.
12. Qiu KM, Wang HH, Wang LM, Luo Y, Yang XH, Wang XM, et al. Design, synthesis and biological evaluation of pyrazolyl-thiazolinone derivatives as potential EGFR and HER-2 kinase inhibitors. *Bioorg Med Chem* 2012;20:2010-8. doi: 10.1016/j.bmc.2012.01.051, PMID 22361272
13. Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comp Chem* 1996;17:490-519. doi: 10.1002/(SICI)1096-987X(199604)17:5<490::AID-JCC1>3.0.CO;2-P
14. Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A. Rational selection of training and test sets for the development of validated QSAR models. *J Comput Aid Mol Des* 2003;17:241-53. doi: 10.1023/a:1025386326946, PMID 13677490
15. Baumann K. Chance correlation in variable subset regression: Influence of the objective function, the selection mechanism, and ensemble averaging. *QSAR Comb Sci* 2005;24:1033-46. doi: 10.1002/qsar.200530134
16. Leach AR, Gillet VJ. An Introduction to Chemoinformatics. Berlin: Springer; 2007. p. 125-70.
17. Topliss JG, Edwards RP. Chance factors in studies of quantitative structure-activity relationships. *J Med Chem* 1979;22:1238-44. doi: 10.1021/jm00196a017, PMID 513071
18. Ajmani S, Jadhav K, Kulkarni SA. Three-dimensional QSAR using the k-nearest neighbor method and its interpretation. *J Chem Inf Model* 2006;46:24-31. doi: 10.1021/ci0501286, PMID 16426036
19. Roy PP, Paul S, Mitra I, Roy K. On two novel parameters for validation of predictive QSAR models. *Molecules* 2009;14:1660-701. doi: 10.3390/molecules14051660, PMID 19471190
20. Sandberg M, Eriksson L, Jonsson J, Sjöström M, Wold S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J Med Chem* 1998;41:2481-91. doi: 10.1021/jm9700575, PMID 9651153
21. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 2001;46:3-26. doi: 10.1016/s0169-409x(00)00129-0, PMID 11259830
22. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 2000;44:235-49. doi: 10.1016/s1056-8719(00)00107-6, PMID 11274893
23. Stamos J, Sliwkowski MX, Eigenbrot C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J Biol Chem* 2002;277:46265-72. doi: 10.1074/jbc.M207135200, PMID 12196540
24. Ogiso H, Ishitani R, Nureki O, Fukai S, Yamanaka M, Kim JH, et al. Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* 2002;110:775-87. doi: 10.1016/s0092-8674(02)00963-7, PMID 12297050
25. Liu Y, Gray NS. Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol* 2006;2:358-64. doi: 10.1038/nchembio799, PMID 16783341
26. Lipinski CA. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discov Today Technol* 2004;1:337-41. doi: 10.1016/j.ddtec.2004.11.007, PMID 24981612