Contents lists available at GrowingScience

Current Chemistry Letters

homepage: www.GrowingScience.com

Development of 2-dimensional and 3-dimensional QSAR models of Indazole derivatives as TTK inhibitors having Anticancer potential

Mithlesh Yadav[a], Balasubramanian Narasimhan[b] and Archana Kapoor[a*]

[a]*Department of Pharmaceutical Sciences, Guru Jambheshwar University of Science and Technology, Hisar, 125001, India*
[b]*Department of Pharmaceutical Sciences, Maharshi Dayanand University, Rohtak, Haryana-124001, India*

C H R O N I C L E

Article history:
Received December 25, 2022
Received in revised form
June 3, 2023
Accepted June 27, 2023
Available online
June 27, 2023

Keywords:
Indazole
Anticancer
2D QSAR
3D QSAR
TRK Inhibitor

A B S T R A C T

The study aimed to explore the anticancer efficacy of indazole pharmacophore by analyzing a series of 109 derivatives of indazole as Tyrosine Threonine Kinase (TTK) inhibitors through quantitative activity relationship analysis using 2D and 3D QSAR techniques. The best 2D-QSAR model was generated by the MLR method, showing a high correlation coefficient (r^2) of 0.9512, and good internal (q^2), and external (pred_r^2) cross-validation regression coefficients of 0.8998, and 0.8661, respectively. The residual values were modest, indicating good agreement between the observed and predicted pIC_{50} values, which suggested that the chosen model was predictably accurate. The 3D QSAR model, built using the SWF kNN approach, displayed a high internal cross-validation regression coefficient (q^2) of 0.9132. Essential structural features/considerations in developing indazole as prospective anticancer medicines have been suggested. The study provides a reliable and predictive model for the prediction of anticancer activity of indazole derivatives. The identified essential structural features/considerations may be useful for the development of prospective anticancer medicines.

© 2024 by the authors; licensee Growing Science, Canada.

1. Introduction

In modern organic chemistry, the importance of computational methods cannot be overstated. These methods allow researchers to predict and understand the properties of organic molecules, such as energy levels, reactivity, and spectroscopic characteristics[1]. Quantum-chemical calculations play a vital role in studying reaction mechanisms by identifying transition states, determining reaction energetics, and exploring different reaction pathways. These calculations enable chemists to uncover the driving forces behind chemical transformations, optimize reaction conditions, and design efficient synthetic strategies[2]. Furthermore, computational methods have expanded their reach into the biological realm, facilitating the investigation of the interactions between chemical compounds and biomolecules. By integrating organic chemistry principles with biology, researchers can explore the potential applications of compounds in drug discovery, bioorganic chemistry, and chemical biology, thereby bridging the gap between chemistry and the life sciences[3].

QSAR/QSPR[4] is a computational method widely utilized for establishing models that capture the relationships between molecular descriptors and physicochemical or physiological characteristics of chemical compounds. These models, derived from numerical representations of the molecular structure, play a vital role in drug identification and optimization[5]. By enabling early *in silico* evaluation of crucial parameters related to compound performance, specificity, and toxicity, QSAR/QSPR models significantly enhance drug discovery programs. The utilization of these models not only improves the effectiveness of drug development but also saves valuable time and resources, making the process more efficient and cost-effective[6].

* Corresponding author.
E-mail address kapoorarchana06@gmail.com (A. Kapoor)

© 2024 by the authors; licensee Growing Science, Canada
doi: 10.5267/j.ccl.2023.6.006

In the realm of cancer research and treatment, computational methods have emerged as essential tools for understanding the complex nature of the disease and guiding the development of effective therapies. Through the integration of computational methods and statistical analyses, we offer insights into the structure-activity relationship and identify key molecular traits and structural features necessary for potent TTK inhibition. These findings can guide the design and synthesis of novel indazole-based TTK inhibitors with improved efficacy, potentially leading to the development of more effective treatments for cancer. Carcinoma was currently the biggest disease-related reason of death in several areas of the globe, and it is predicted to remain so in the upcoming decades[7]. In cancer treatment, various chemicals are used to shrink tumours. Chemotherapy has the advantage of being able to treat malignancies that have spread throughout the body, surgical and radiation treatments, on the other hand, are confined to treating localized tumours[8]. Chemotherapy has piqued the interest of many academicians, and most of the recent studies have been dealing with the development and production of numerous anticancer drugs[9].

Numerous cell growth and checkpoint kinases in the human body have been identified as potential targets for cancer therapy. One such pathway is the spindle assembly checkpoint (SAC), which is a conserved pathway in eukaryotes[10]. Its role is to halt the cell cycle during mitosis until all chromosomes establish stable bipolar attachments to the mitotic spindle. When the SAC malfunctions, it allows mitotic exit even when kinetochores are unattached, leading to chromosome missegregation. Previous studies have shown that inactivating the SAC results in lethal genomic instability in cancer cells. Threonine tyrosine kinase (TTK), also known as Mps1, is a critical component of the SAC and acts as a surveillance mechanism to ensure accurate mitosis and genome stability[11]. TTK is often overexpressed in highly proliferative cancers such as lung cancer, including the most aggressive and deadly forms. Inhibition of TTK leads to premature exit from mitosis, causing chromosome missegregation, aneuploidy, and ultimately cell death of cancerous cell. Decreased TTK activity significantly reduces cell viability. Comprehensive analyses combined with gene expression profiles of human cancers have identified TTK as a promising target for anticancer therapy[12,13]. Reversine its analogue MPI-0479605, NMS-P715, MPS1-IN-1, AZ 314628, SP600125[14], the Genentech and Cancer Research UK inhibitors (CCT251455) are the small molecules of TTK inhibitors (**Fig. 1**). The ubiquitous appearance of a crucial N-phenyl pyrimidin-2-amine motif, indicated in **Fig. 1** is the most remarkable aspect of all of these inhibitors, a variation on this subject can be seen in the Shionogi-1 inhibitor. The strategy was based on hits from a concentrated examination conducted, that led to the discovery of acetamido and carboxamide substituted 3-(1*H*-indazol-3-yl)-benzenesulfonamides (Shionogi-2, **Fig. 2**) [15, 16]. The indazole moiety (1,2-benzodiazole) is a versatile molecule with a variety of biological characteristics, particularly anticancer action. The 5-substituted acetamido and carboxamide derivatives of 3-(1*H*-indazol-3-yl) benzenesulfonamides (**Fig. 2**), respectively, were strong TTK inhibitors with a limited capacity to slow the development of cancer cells. In the context of cancer research, inhibition of TTK has gained significant attention due to its crucial role in mitosis and its association with aneuploidy and chromosomal abnormalities. The development of TTK inhibitors holds promise for the design of novel therapies.

There is a plethora of literature accessible on the various biological actions of naturally and semi-synthetic indazole analogues, however, articles describing the fundamental regulatory aspects influencing anticancer activity are scarce. To address this gap, a quantitative structure-activity relationship (QSAR) investigation was conducted on 109 indazole derivatives as Tyrosine Threonine Kinase (TTK) inhibitors. By analyzing their 2D and 3D QSAR structural requirements, this study aimed to identify essential features for developing novel chemotypes and enhancing TTK inhibitor efficacy. The generated predictive models encompassed steric and electrostatic impacts, providing valuable insights for prospective anticancer drug design. This research has tried to contribute to the understanding of the anticancer efficacy of indazole pharmacophores and offers crucial considerations for developing potent TTK inhibitors with improved therapeutic potential.

Novelty of work and their contribution

By investigating a specific class of compounds, this study provides valuable insights into the structure-activity relationship of indazole-based TTK inhibitors. The study employed comprehensive 2D and 3D QSAR analyses to elucidate the molecular traits and structural features crucial for the compounds' anticancer activity against TTK. Through statistical modeling and analysis, physicochemical and alignment-independent descriptors play a significant role in regulating the alterations in activity. This quantitative understanding of the structural requirements provides a solid foundation for designing and optimizing indazole derivatives with enhanced therapeutic potential. Furthermore, the research contributes to the field by developing reliable predictive models.

Among the three 2D-QSAR methods tested, multiple linear regression (MLR) analysis demonstrated superior predictive power and reliability. The MLR equation derived from the study can serve as a valuable tool for medicinal chemists in the rational design of novel indazole derivatives with improved biological activity. This advancement in predicting structure-activity relationships facilitates the targeted synthesis and optimization of potential TTK inhibitors. The 2D QSAR analyses was further supported by 3D QSAR modeling using the kNN-MFA method. By considering the spatial arrangement of substituents, we identified the shape, particularly the steric descriptors, as the governing factor for the compounds' anticancer activity. This insight expands our understanding of the structural requirements for effective TTK inhibition and offers valuable guidance for future synthetic efforts. Overall, our work contributes to the development of the topic by providing a comprehensive analysis of the anticancer activities of indazole derivatives targeting TTK.

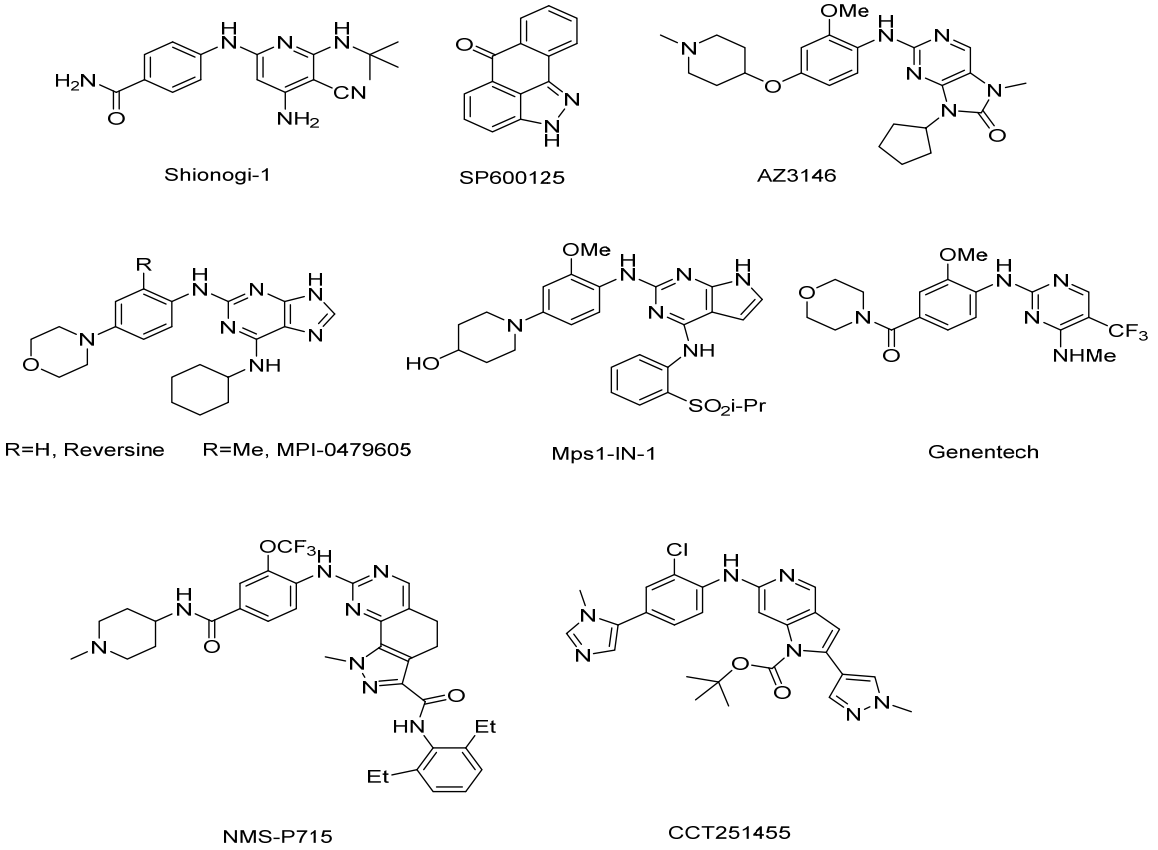

Fig. 1. The reported TTK Inhibitors

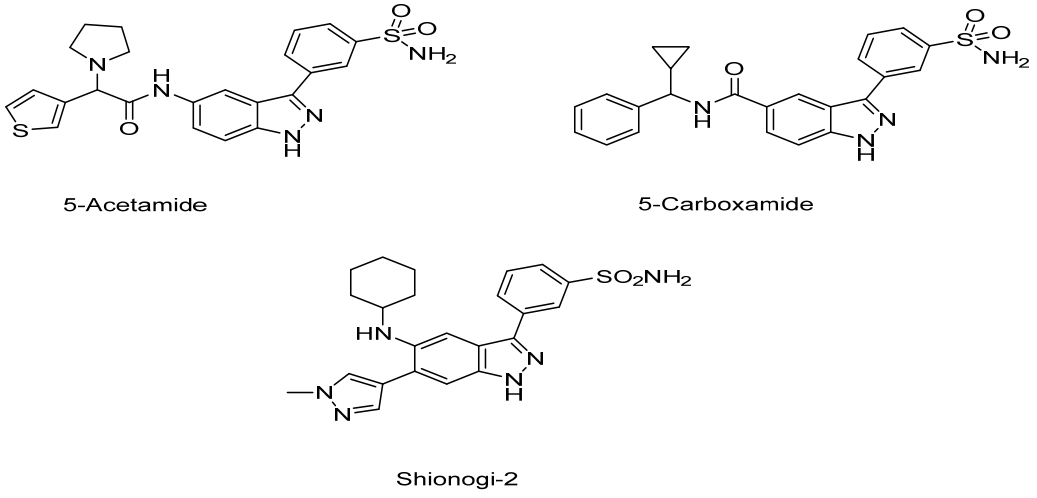

Fig. 2. Reported indazole derivatives from the literature

2. Results and discussion

By applying the novel two-dimensional and three-dimensional QSAR methods to existing sets of compounds, as stated above, the relevance and usefulness of the novel 2D and 3D QSAR methods have been established. The data was divided by a specific splitting procedure i.e., random division method into the training, test, and validation sets. QSAR models were constructed using training sets, and their validation was carried out using test sets. The "actual" predictive power of QSAR models was determined using external validation sets [17].

The set of training and set of tests were chosen by following uni-column statistics, i.e., the test set's maximum is smaller than the training set's maximum, and the test set's minimum is bigger than the training set's minimum, which is required for further QSAR analysis. This indicated that the test is interpolative, meaning that it is generated from the training set's min-

max range. The training and test sets' mean and standard deviation revealed the change in mean and point density distribution between the two sets. The uni-column statistics was reported in **Table. 1**. The electrostatic and steric descriptors, as well as their range, were included in the QSAR models which were created using kNN-MFA to reflect their importance for interaction in the molecular field. 3D QSAR investigations have been represented in models A, B, and C. The series of indazole derivatives yielded numerous QSAR equations. For this discussion, certain statistically meaningful two-dimensional and three-dimensional QSAR models were generated.

Table. 1. For QSAR models, uni-column statistics table used for the set of training and test sets

Data set	Average	Maximum	Minimum	Standard deviation	Total
2D Training	7.404	8.959	4.292	1.239	562.667
Test	7.417	8.918	4.775	1.296	155.749
3D Training	7.356	8.966	4.191	1.173	558.805
Test	7.494	8.921	4.745	1.414	157.371

2.1. 2D QSAR

Using statistical techniques, the 2D-QSAR methodology could help in tying biological activity to their molecular characteristics. Indazole derivatives have shown promise as anti-cancer drugs in the literature. Using various variable selection strategies, various statistically meaningful 2D-QSAR models (PLS, MLR, PCR) were created, resulting in three best models. The prediction performance of the strategy was proposed by the distinct models SWF-MLR, SWF-PLS and SWF-PCR. The training data was predicted by the Model-1 by SWF-MLR, Model-2 by SWF-PLS and Model-3 by SWF-PCR methods with r^2 of 0.9512, 0.9296, and 0.8764 and q^2 of 0.8998, 0.8862, and 0.8156 respectively as shown in in **Table. 2**.

Table 2. Statistical metrics for assessing the prediction power of various 2D-QSAR models

Parameters	**SW-MLR**	**SW-PLS**	**SW-PCR**
N	76	76	76
Degree_of_freedom (DF)	68	70	74
r^2	0.9512	0.9296	0.8764
q^2	0.8998	0.8862	0.8156
F_test	141.3789	68.1808	47.8417
r^2 se	0.4346	0.4694	0.4812
q^2 se	0.4162	0.4467	0.46375
Pred r^2	0.8661	0.8310	0.7486
Pred r^2se	0.2834	0.3082	0.3506
ExternalValidation r^2	0.8486	0.8224	0.7965
ExternalValidation_r^2se	0.2422	0.3201	0.4066
ZScore R^2	16.0227	17.7559	48.7278
ZScore Q^2	10.6728	17.9341	57.7114
best ran R^2	0.5616	0.4669	0.4443
best ran Q^2	0.1649	0.1421	0.1188

MODEL – 1 (MLR Analysis)

The final QSAR equation created during a 2D QSAR investigation utilizing Multiple Linear Regression (MLR) using the forward-backward stepwise variable selection procedure was as follows,

pIC_{50} = 0.3743 (chi4) -0.6916 (T_N_N_7) + 0.2061 (T_2_N_3) - 0.9956 (T_N_N_4) – 0.9505 (T_N_N_2) + 0.1343 (T_2_O_5) – 0.0328 (T_2_C_3) + 2.6257

Degrees of Freedom = 68, N training= 76, N test= 21, r^2= 0.9512, q^2= 0.8998, pred r^2= 0.8661, F test =141.3789

Table 3 summarises the descriptors chosen to describe the indazole derivatives for anticancer activity of **Model-1** and **Table 4** presents the correlation matrix between the descriptors and the biological activity. The anticancer activity of QSAR **Model-1** against TTK inhibitor, with determination coefficient (r^2) of 0.9512 representing 95.12% variability in observed activity data. This model's internal prediction power was 89 percent when cross-validated (q^2 = 0.8998), indicating that it has a strong internal predictive ability. Another statistical parameter used for the predictivity test set compounds was pred_r^2 = 0.8661, which indicated that the model has good external predictive potential. The r^2_se has a low value, suggesting that it is well-fit. The F-test value of 141.3789 indicated that the total scientific significance level of the model is 99.99 percent, implying that the model's failure probability is 1 in 10,000. The contribution chart of Model-1 by SW-MLR is represented in **Fig. 3(a)**. The fitness plot of the expected vs. measured data of Model-1 is represented in **Fig. 3(b), Fig. 3(c)** and **Fig. 3(d)** showed the radar plot visuals that were created to examine the model's quality by comparing experimental and predicted activities in training and test sets and it can be stated that the model is of the highest quality. It can be observed that the model generated by the multiple regression method was governed by the physicochemical as well as alignment-independent (AI) parameters. The physicochemical parameter i.e., retention index (fourth order) chi4, calculated directly from gradient

retention times, has been positively contributing towards the activity. Further, the alignment independent descriptor, T_N_N_7 counts the number of nitrogen atoms isolated from any other nitrogen atom by seven bonds (N-C-C-C-C-C-C-N) contributed negatively to the activity i.e., the presence of amide group at C5 position was not found to be favourable for the activity. The descriptor T_2_N_3 (C-C-C-N) indicated that the activity has been increased if the number of double-bonded atoms (i.e., any double-bonded atom, T 2) were separated by three bond distances from nitrogen atoms in a molecule. So, it can be stated that the existence of the NH group, directly attached to the C5 position of indazole was favourable for the activity. T_N_N_4 descriptor denotes the number of nitrogen atoms separated by four bonds from any other nitrogen atom (N-C-C-C-N), harms biological activity which is evident from the value obtained in the equation. The energy of the molecules correlated by the T_N _N _2 i.e., (N-C-N) descriptor was not found to be favorable for the activity. The presence of descriptor T_2_O_5, means a number of the double-bonded atoms (i.e., any double-bonded atom, T 2) separated from the oxygen atom by 5 bonds (C-C-C-C-C-O), has raised the activity. Thus, the existence of the -OCH_3 group at the C5 position of indazole derivatives was advantageous for activity, while the T_2_C_3 descriptor (count of several double-bonded atoms separated from carbon atom by three bonds) were found to have a negative effect associated with the activity.

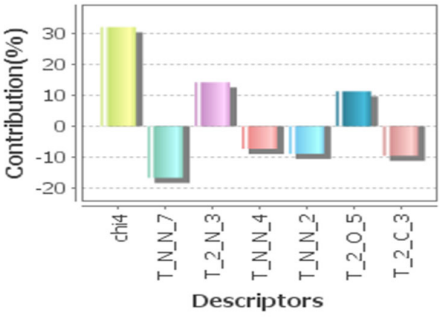

Fig. 3(a). Contribution chart of Model-1 By SW-MLR

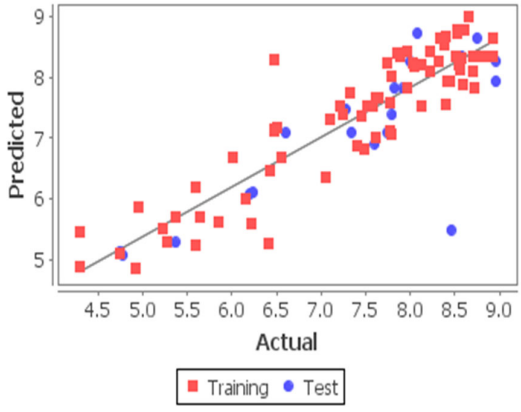

Fig. 3(b). Fitness plot of the expected vs. measured data of Model-1

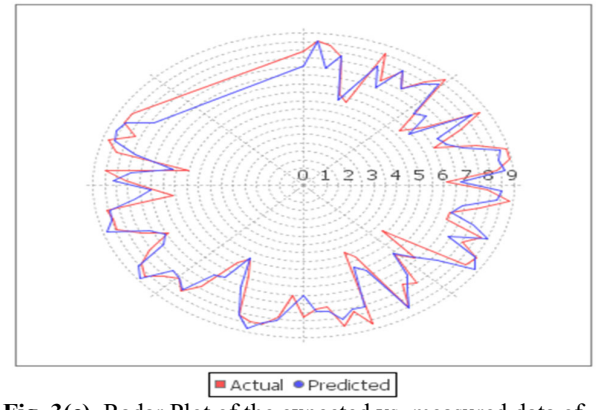

Fig. 3(c). Radar Plot of the expected vs. measured data of training set (Model-1)

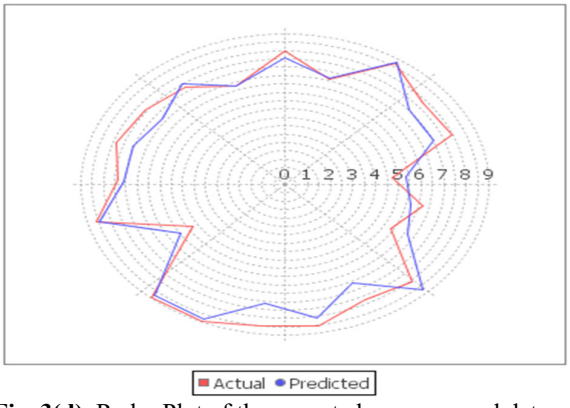

Fig. 3(d). Radar Plot of the expected vs. measured data of test set (Model-1)

MODEL – 2 (PLS Analysis)

In this study, the biological activity was connected with seven chosen descriptors using the PLS approach. The following is the best-retrieved PLS model:

pIC_{50} = 0.3404 (chi4) -0.6908 (T_N_N_7) + 0.2323 (T_2_N_3) – 1.0243 (T_2_N_4) + 0.1606 (T_2_O_5) – 0.9488 (T_N_N_2) – 0.0287 (T_2_T_3) + 2.7849

Degrees of Freedom = 70, N training= 76, N test= 21, r^2= 0.9296, q^2= 0.8862, pred r^2= 0.8310, F test = 68.1808.

With low standard error estimation, **Model-2** explained 92 percent of the variance in activity. It also has 88 percent internal (q^2) and 83 percent external pred_r^2 predictive ability respectively. The chi4, T_2 _N _3, and T _2 _O _5 all contributed positively to the QSAR model. The descriptors T_N_N_7, T_2_N_4, T_N_N_2, and T_2_T_3 contributed negatively to the activity. The contribution chart of **Model-2** by SW-PLS was shown in **Fig. 3(e)**. The Fitness plot of the expected vs. measured data of Model-2 is represented in **Fig. 3(f), Fig. 3(g)** and **Fig. 3(h)** was shown the radar plots of the observed vs. predicted data of pIC_{50}.

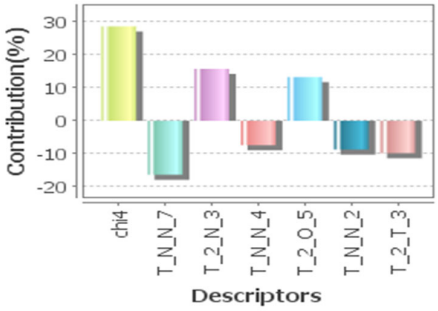

Fig. 3(e). Contribution chart of Model-2 By SW-PLS

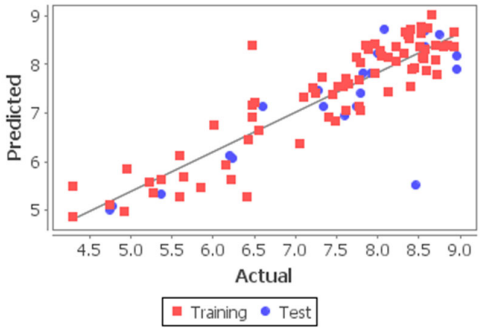

Fig. 3(f). Fitness plot of the expected vs. measured data of Model-2

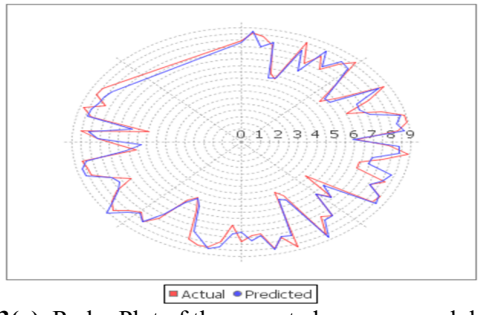

Fig. 3(g). Radar Plot of the expected vs. measured data of training set (Model-2)

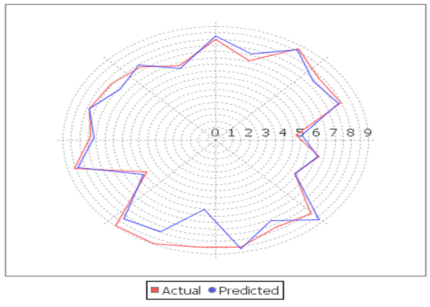

Fig. 3(h). Radar Plot of the expected vs. measured data of the test set (Model-2)

MODEL – 3 (PCR Analysis)
$pIC_{50} = 0.4408$ (chi4) + 2.7102
Degrees of Freedom = 70, N training= 76, N test= 21, r^2= 0.8764, q^2= 0.8156, pred r^2= 0.7486, F test = 47.8417.

The SW-PCR approach produced a coefficient of determination of 0.8764, an internal predictive power of 81 percent, and an outward predictivity of 74 percent for the same data set. The equation was governed by chi4, the retention index (fourth order) according to this model. The contribution chart of **Model-3** by SW-PLS was shown in **Fig. 3(i)**. The fitness plot of the expected vs. measured data of Model-3 is represented in **Fig. 3(j)**. The plots of the observed vs. predicted data of pIC_{50} were shown in **Fig. 3(k)** and **Fig. 3(l)**. From the above-developed models, it can be stated that the contributing descriptors were found to be almost similar while developing equations for Model 1, Model 2, and Model 3. The values of statistical parameters (r^2, q^2, and pred r^2 as well as the intercept to best-fit line) suggested a linear relationship between the dependent and the independent variables, thus stating that the out of three models reported, Model 1 was the best model developed with the predictive ability of 84% (0.8486) respectively. **Table. 2** showed the statistical results, while **Table. 5** showed the actual and predicted activities of the best **Model-1** and also demonstrated the above models which were all verified by estimating the biological activities of the test molecules. The 2D-QSAR investigations were conducted on several indazole compounds that have anticancer action towards TTK and the analysis provided insight into the importance of many physicochemical factors as well as alignment independent descriptors and their connections to the possible anticancer activity: chi4, T_2_N_3, and T_2_O_5 contributed positively towards activity.

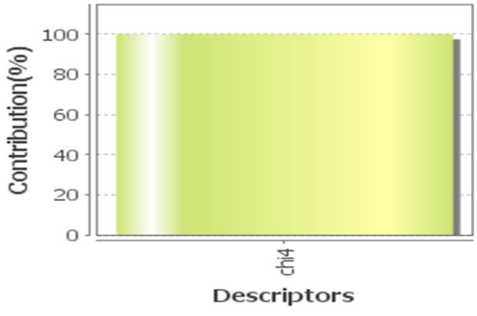

Fig. 3(i). Contribution chart of Model-3 By SW-PCR

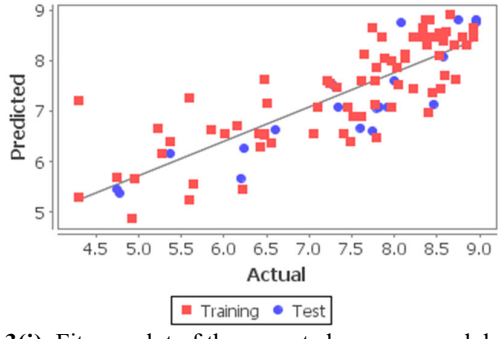

Fig. 3(j). Fitness plot of the expected vs. measured data of Model-3

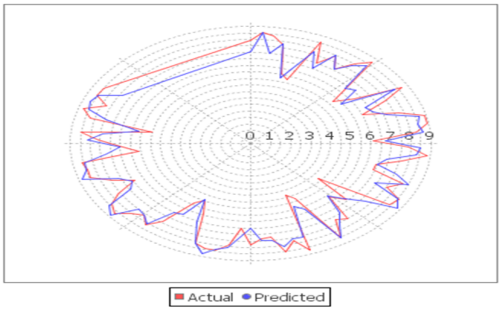

Fig. 3(k). Radar Plot of the expected vs. measured data of training set (Model-3)

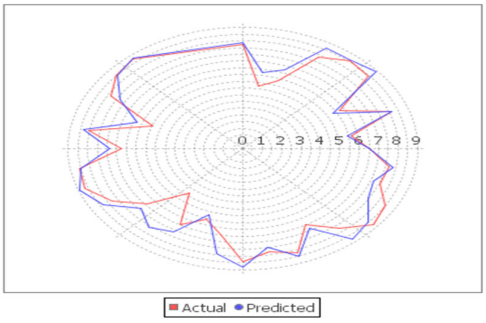

Fig. 3. (l). Radar Plot of the expected vs. measured data of the test set (Model -3)

2.2. 3D QSAR

The 3D QSAR modelling studies have been recognized as an important concept in designing more potent inhibitors, which relates their 3D structural attributes with bioactivity. In comparison to the classical 2D QSAR study, 3D-QSARs offer more promising techniques to deal with steric factors due to the utility of three-dimension descriptors. The main limitations of classical 2D-QSAR method are (a) visualization of the results due to lack of proper three-dimensional graphical outputs, (b) inadequate information regarding drug–receptor interactions. Thus, three-dimensional QSAR (3D-QSAR) has been introduced to overcome such drawbacks[18]. To predict and optimize the structural requirements of the indazole pharmacophore as potential TTK inhibitors, a number of 3D-QSAR models have been created. In order to perform a 3D-QSAR analysis a probe atom was used to calculate the steric and electrostatic fields at each grid point of the created lattice. These models were generated using kNN-SWF, kNN-SA, kNN-GA variable selection methods. Any QSAR model's choice of compounds for the training set and test is a crucial and crucial component. Therefore, attention was made to ensure that all compounds in the test have biological activities that fall within the range of maximum and lowest values for the training set of compounds. The uni-Column Statistics of test and training sets also demonstrated proper test and training set selection in **Table. 1**. The optimal model was chosen using a combination of several statistical parameters q^2 (cross verified r^2) and pred r^2 (the external cross validation regression coefficient) in three fields: steric, electrostatic, and hydrophobic. **Table. 6** showed the statistical findings obtained using SW-kNN MFA techniques. **Table. 7** listed the residuals of the training and test sets compounds as well as the actual/observed activities, anticipated activities by the 3D QSAR model. The propagation of residuals on both sides of the zero-line showed that the kNN-MFA model's evolution was error-free. It has become clear from **Table. 7** that all of the compounds in the test set have predicted activities that are in good correlation with the related experimental activities, yielding the best match. **Table. 8** summarised the descriptors chosen to describe the indazole derivatives for anticancer activity of **Model A** and **Table. 9** presented the correlation matrix between the descriptors and the biological activity of Model A. The kNN-MFA QSAR model's cross-validation was aided by the plots of observed vs predicted activity of the compounds in the training and test sets, which are shown in **Fig. 4(a)**. **Fig. 4(b)** and **Fig. 4(c)** show the radar plots of the observed vs. predicted data of pIC_{50}. The value of q^2, the model's internal predictive capacity, and pred r^2, the model's capacity to forecast the behaviour of the external test set, served as the criteria for selecting a model. Different models were generated like **Model A**, **Model B** and **Model C** by different methods which were shown below:

MODEL – A (kNN-SWF)
pIC_{50} = S_1225 (-0.0061 -0.0030) + S_1185 (-0.3238 -0.2931) + S_846 (30.0000 30.0000) + S_1528(-0.3707 -0.2170)
k Nearest Neighbor = 3, N = 76, Degree of freedom = 71, q^2 = 0.9132, pred_r^2 = 0.9447, External validation r^2 = 0.8912, q^2 se = 0.2062, pred r^2se = 0.1531, External validation r^2 se = 0.3020, Selected Descriptors and range = S_1225 (-0.0061 -0.0030), S_1185 (-0.3238 -0.2931), S_846 (30.0000 30.0000) and S_1528(-0.3707 -0.2170)

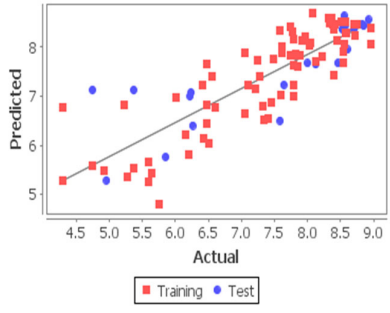

Fig. 4 (a). Fitness plot of the expected vs. measured data of Model A

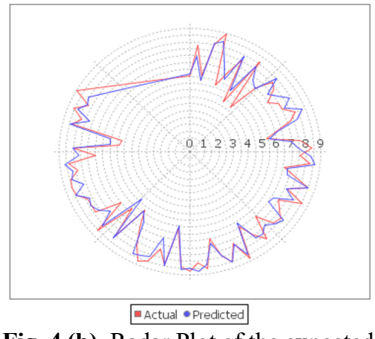

Fig. 4 (b). Radar Plot of the expected vs. measured data of set of training (Model A)

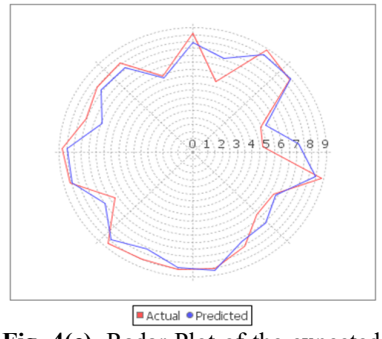

Fig. 4(c). Radar Plot of the expected vs. measured data of set of test (Model A)

Table 6 showed the statistical findings obtained using SW-kNN MFA techniques. **Table 7** listed the residuals of the training and test sets compounds as well as the actual/observed activities, and anticipated activities by the 3D QSAR model. The propagation of residuals on both sides of the zero-line showed that the kNN-MFA model's evolution was error-free. It has become clear from **Table 7** that all of the compounds in the test set have predicted activities that are in good correlation with the related experimental activities, yielding the best match. To predict and optimize the structural requirements of the indazole pharmacophore as a potential TTK inhibitors several 3D-QSAR models have been created. The kNN-MFA QSAR model's cross-validation was aided by the plots of observed vs predicted activity of the compounds in the training and test sets, which are shown in **Fig. 4(a)**. **Fig. 4(b)** and **Fig. 4(c)** show the radar plots of the observed vs. predicted data of pIC_{50}. The value of q^2, the model's internal predictive capacity, and pred r^2, the model's capacity to forecast the behaviour of the external test set, served as the criteria for selecting a model. **Table 8** summarised the descriptors chosen to describe the indazole derivatives for the anticancer activity of **Model A** and **Table 9** presented the correlation matrix between the descriptors and the biological activity of Model A.

Model A was observed to be significantly relevant for external and internal predictive ability, which were found to be 90% (q^2 = 0.9132) and 94% (pred r^2 = 0.9447). In this model (**Fig. 5**), the steric descriptor with a negative coefficient (S_846) was observed to have a negative contribution to the anticancer activity which further highlighted the requirement of less bulky groups at the C5 position of Indazole-based TTK inhibitors. Another steric field parameter had a negative coefficient (S_1225), indicating that bulky substituents on the aryl ring present at C3 of the indazole scaffold are undesirable and their presence reduced the activity. The same observation was made at the C5 position of the indazole with negative steric field parameter (S_1185) which means the bulky substituents cannot be accommodated at this specific site and thus harmed on the activity profile of the indazole TTK inhibitors.

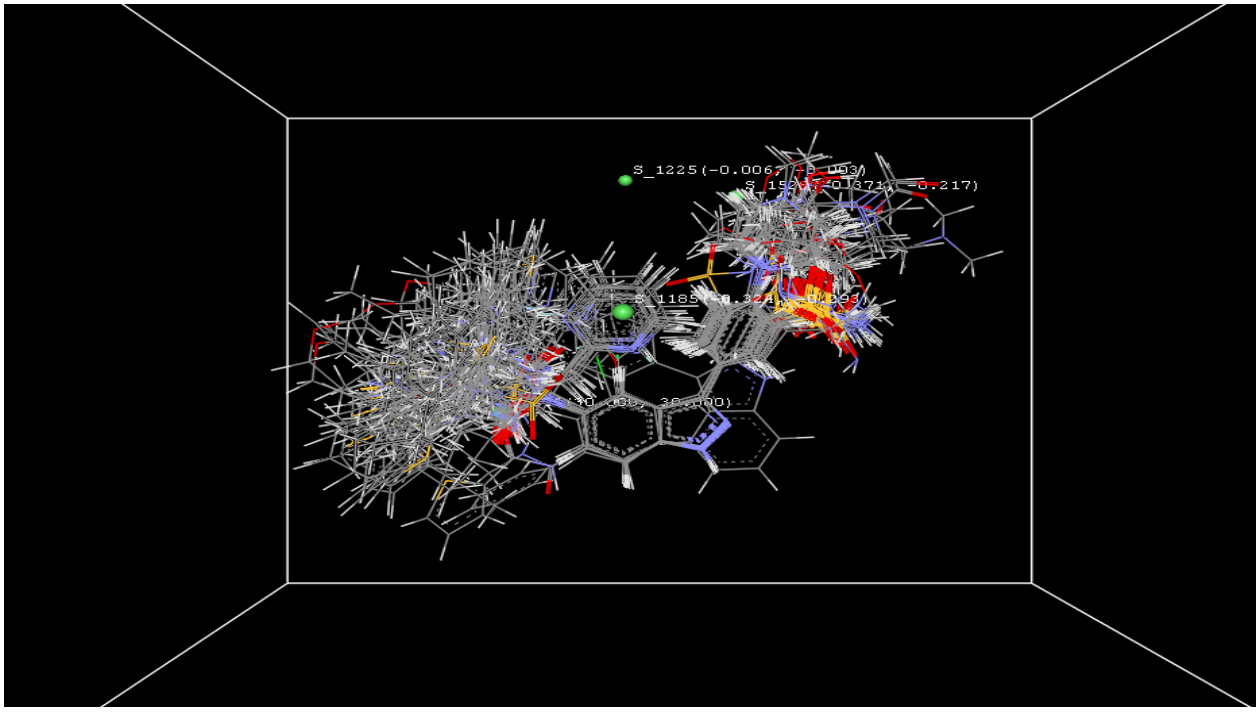

Fig. 5. Stereo view of Model A's developed by SW kNN-MFA method on a molecular rectangular field grid around the superposed molecular units of the indazole series

MODEL – B (kNN-SA)
pIC_{50} = S_690 (-0.1328 -0.0728) + S_ (-0.0574 -0.0111)
k Nearest Neighbor = 5, N = 76, Degree of freedom = 73, q^2 =0.8866, pred_r^2 =0.7339, External validation r^2 =0.7961, q^2 se =0.2731, pred r^2se =0.1805, External validation r^2 se =0.3958, Selected Descriptors and range = S_690 (-0.1328 -0.0728) and S_ 393(-0.0574 -0.0111)

Model B was generated by k Nearest neighbors simulated annealing method and only steric parameters have participated in correlating the biological activity with the independent variables. The negative steric coefficients (S_690) and (S_393) suggested that bulkier groups are not desirable for the anticancer activity of indazole derivatives.

MODEL – C (kNN-GA)
pIC_{50} = S_1199 (-0.0266 -0.0188) + S_1177 (-0.0106 -0.0081) + S_763(-0.0289 -0.0213)

k Nearest Neighbour = 4, N = 76, Degree of freedom = 72, q^2 =0.8627, pred_r^2 =0.7107, External validation r^2 =0.7762, q^2 se =0.6903, pred r^2se =0.6944, External validation r^2 se =0.6157, Selected Descriptors and range = S_1199 (-0.0266 -0.0188), S_1177 (-0.0106 -0.0081) and S_763(-0.0289 -0.0213)

Model C has also been controlled by the negative steric descriptors; a major role plays in depicting the structural requirement for the anticancer activity.

From the above discussion, it can be stated though all three models were ruled by steric parameters only the external and the internal predictive ability of **Model A** was statistically more relevant: having q^2 =0.9132, pred r^2 = 0.9447, pred r^2se = 0.1531, external validation_r^2 = 0.8912 values, hence it was chosen for NCE's design.

Table 6. The comparison of statistical findings of 3D QSAR produced by the kNN-MFA approaches

S.No	Statistical parameter	Stepwise Forward Backward (SWFB) (Model A)	Simulated Annealing (SA) (Model B)	Genetic Algorithm (GA) (Model C)
1	q^2	0.9132	0.8866	0.8627
2	q^2se	0.2062	0.2731	0.3603
3	Pred r^2	0.9447	0.7339	0.7107
4	Pred r^2se	0.1531	0.1805	0.2944
5	ExternalValidation_r^2	0.8912	0.7961	0.7762
6	ExternalValidation_r^2se	0.3020	0.3958	0.4157
7	N	76	76	76
8	K Nearest neighbor	3	5	4
9	Degree of freedom	71	73	72
10	Contributing descriptors	S_1225 -0.0061 -0.0030 S_1185 -0.3238 -0.2931 S_846 30.0000 30.0000 S_1528 -0.3707 -0.2170	S_690 -0.1328 -0.0728 S_393 -0.0574 -0.0111	S_1199 -0.0266 -0.0188 S_1177 -0.0106 -0.0081 S_763 -0.0289 -0.0213

Fig. 5 showed the 3D data sets created around a rectangular area with the range of participation specified in parenthesis utilizing SW KNN-MFA. A grid was created to illustrate the data contained in the 3D-QSAR models and depicted the electrostatic, steric, and hydrophobic characteristics. The SW kNN–MFA 3D-QSAR model creates points. i.e., S_1225 -0.0061 -0.0030, S_1185 -0.3238 -0.2931, S_846 30.0000 30.0000, S_1528 -0.3707 -0.2170 steric interaction at lattice points 1225, 1185, 846, and 1528 respectively. The presence of negative steric descriptors indicated that sterically bulky aryl compounds were not considered for optimum activity.

2.3. Layout Design of Indazole Pharmacophore

The pursuit of good target intensity must not be undertaken without consideration of its implications for accuracy and effectiveness. The method adopted in this research could help in the development of new and promising indazole compounds as anticancer medicines. The results of 2D and 3D QSAR investigations have been summarised. The developed scaffold will be used to synthesize designed analogues with remarkable biological activity in the future. The total substitution in the scheme which is needed around the indazole pharmacophore was presented in **Fig. 6.**

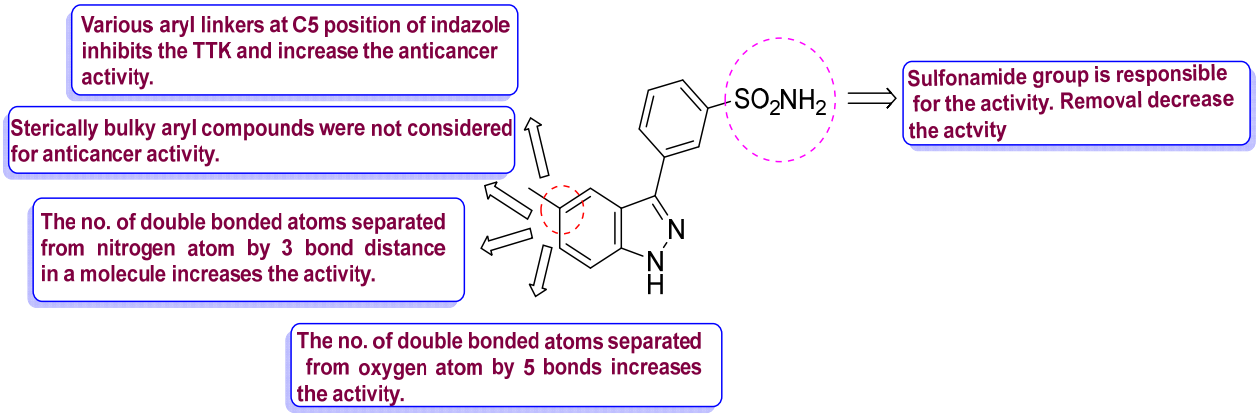

Fig. 6. Structural Insights for the Design of Indazole Analogues

3. Conclusion

This study demonstrated how anticancer activities of several derivatives bearing indazole scaffold can be statistically analysed for the identification of different molecular traits required for high biological activity. 2D and 3D QSAR analyses of various indazole compounds with anticancer activity against TTK have been conducted. The physicochemical and alignment-independent descriptors were discovered to play a significant function in regulating the alteration in activities of 2D QSAR investigations. Among the three 2D-QSAR models (MLR, PCR, and PLS), the results of MLR analysis showed significant predictive power and reliability as compared to the other two methods with an impressive correlation coefficient (r^2) of 0.9512, along with the internal cross-validation regression coefficient (q^2) and the external cross-validation regression coefficient (pred_r^2) displayed good values of 0.8998 and 0.8661, respectively. Thus, demonstrated a high level of predictability and accuracy. The results of 2D QSAR studies were further supported by the 3D QSAR analysis. 3D-QSAR study adopted stepwise forward variable selection algorithms and the models were developed by the kNN–MFA method, which yielded an internal cross-validation regression coefficient (q^2) of 0.9132, further suggesting a strong relationship between the variables in the model. For the dataset of 109 indazole derivatives, steric descriptors of the molecules appeared to be the governing factor for the anticancer activity, thus extending the structural insight. The current research will help synthetic medicinal chemists in creating new, strong TTK inhibitors with increased biological activity in comparison to the reported compounds.

4. Future perspectives

Building upon our findings, future studies could focus on synthesizing and evaluating novel compounds with specific modifications that target the identified structural features associated with enhanced anticancer activity. Additionally, the integration of computational methods, such as molecular dynamics simulations and virtual screening, can aid in the design of novel indazole derivatives with improved potency, selectivity, and pharmacokinetic properties. These approaches can streamline the drug discovery process by providing valuable insights into the interactions between the inhibitors and their target, facilitating the identification of lead compounds for further experimental validation. Furthermore, in vivo studies and preclinical evaluations of the most promising compounds would be conducted to assess their efficacy, safety, and potential for clinical translation. Such studies would provide valuable information regarding the therapeutic potential of indazole-based TTK inhibitors and their applicability in cancer treatment.

5. Experimental

5.1. Materials and Methods

Numerous domains such as drug design, predicated toxicology, and threat assessment, benefited from the use of Quantitative structure-activity relationship (QSAR). The physiological, structural, and biological areas are all included in the Organization for Economic Co-operation and Development (OECD) definition of the utility area for QSAR models[19]. The two key goals for QSAR development are as follows: predicting the biological activity of untested molecules and developing a predictive and reliable QSAR model with a specific chemical domain. The QSAR can also be used as an informative tool by extracting significant patterns from descriptors related to the measured biological activity, which helps to understand the mechanisms underlying a given biological activity. Such details may aid in the suggestion of new lead molecule designs with enhanced activity profiles[20]. To carry out the computational investigation, (2D and 3D QSAR) a total of 109 Tyrosine Threonine Kinase (TTK) inhibitors reporting anticancer activity and having known IC_{50} values were selected from the literature [15,16]. All the derivatives selected for the experiment followed the same assay technique and represented similar properties although their pharmacophore and efficacy differed significantly. The inhibition value of the molecules in the data set vary from 0.016 to 50 µm, which were then transformed to pIC_{50} by the following mathematical formula (Eq. 1):

$$pIC_{50} = -log10(IC_{50}) \quad (1)$$

All the molecules in the data set have their structures which were drawn using the sketching capabilities of Marvin sketch saved in MDL Mol file format. The compound's energy is minimized using the MMFF is Merck molecular force field[21] and V life MDS 4.6 software was used, with a dielectric constant of 1.0 and (RMS) a root means square gradient of 0.01 kcal/molÅ. It is necessary to reduce the energy of the molecules for the ligand to bind well to its target receptor[22]. The conformer was selected for each compound and was used for further study having low energy. The random selection method was used to divide the entire data set into training, test sets, and validation sets[23]. By doing QSAR research on a collection of data of indazole analogues, the current work seeks to examine the two-dimensional and three-dimensional structural requirements of QSAR. The models that result from these experiments will give the designer, a new insight into the possible requirements for the design of effective TTK inhibitors.

5.1.1. Dataset for 2D - QSAR Study

Data on anticancer activity, in the form of structure and percentage inhibition of the compounds **(Table. 10)** from the literature, was taken. The random selection (RS) method was used to divide a set of a total of one hundred nine analogs of indazole before the QSPR model development. By following the random selection method, the whole data was split into three sections i.e., Training set, test set, and validation set. In this study, the inclusion of a separate validation set is an attempt to increase the external predictive ability of the QSAR model to be developed[24]. A training set comprising 76 molecules, a test set of 21 molecules, and 10% molecules (12 molecules) for validation of created QSAR model which was used for the QSAR study. The development process of the test and training set after the compound's selection was very important in the generation of QSAR models. A uni-column statistics for training and test set were created to examine the accuracy of selection basis of training and test set compounds [25, 26] given in **Table. 1.** Furthermore, the training set was subjected to a leave-one-out strategy for autonomous internal validation of the models created.

5.1.2. Descriptor Analyses

The Minimum energy form is the most stable form for the drug's ability to attach to the receptor in its efficient sites. The numerical depiction of molecules is called theoretical molecular descriptors. It represents the chemical information included within molecular structures. In the field of theoretical molecular design and advanced research, these descriptors are quite useful. Physiochemical metrics such as polar surface area, estate contributions, element count, hydrophobicity, estate numbers, dipole-moment, $logP_A$; topological, $XlogP_A$, Element Count Chi, Estate number, Path-cluster, and alignment-independent topological descriptors and hydrophobicity have been computed[27,28]. Taking the physiochemical descriptor acting as an independent variable and pIC_{50} (biological activity data) as the dependent variable, a total of 240 descriptors was calculated with the help of the molecular design suite of Vlife Sciences in the QSAR module tool. Descriptors with the same or nearly identical values or that are significantly connected with other descriptors were initially excluded using invariable column selection since they do not play any part in the QSAR evaluation. To further exclude non-significant descriptors, the stepwise forward and backward variable selection method (SWFB) along with other statistical techniques such as (MLR) Multiple Linear Regression, (PLS) Partial Least Squares regression, and (PCR) Principal Component regression[29] was utilized and the number of descriptors was reduced to a total number of 160 descriptors.

5.1.3. Validation and Statistical Techniques

Numerous two-dimensional QSAR models were created by MLR, PLS, and PCR-based regression/algorithms by using variable selection methods such as Stepwise (SW), Genetic Algorithm (GA), Simulated Annealing (SA) having descriptors with the highest correlation with biological activity. The variables were positioned in such a way that the regression equation produced five times fewer autonomous variables than derivatives to produce an effective result. The program evaluates the right model based on the coefficient of determination (r^2), which represents the variability in the experimental action values; (q^2) the internal cross-validation regression coefficient, which is a relative criterion of quality of fit; (pred_r^2) the external cross- validation regression coefficient, which represents the external explanatory power; External Validation_r^2; the F-ratio, which represents the variance of calculated and observed activity and the calculated value of the F-test has a higher degree of statistical significance (99.99 percent) than the tabulated value of the F-test; standard error (r^2_se) representing an absolute measure of the quality of fit; standard error of cross-validated square correlation co-efficient (q^2_se); standard error of predicted squared regression (pred_r^2se) to estimate the predictive potential of the models respectively. The absolute robustness of the model is represented by a low standard error of pred_r^2, q^2, and r^2. For the assessment of predictive ability within the model, the validation of QSPR models was carried out which was further ensured by pred_r^2 (internal validation) and q^2 (cross-validation) parameters[30]. The fitness plot backs up the statistical significance of the two-dimensional QSPR model. It gives information that how much the predicted activity of the external set was performed and how successfully the model was trained. The contribution chart indicated the percentage contribution of the descriptors and the value of the descriptors used for model development [31].

5.1.4. Cross-Validation

Leave-one-out (LOO) was used to perform internal validation. In this method, a molecule from the training set was removed and Eq. (2) was used to predict the activity by considering the average of the activities of the k most similar molecules.

$$\hat{y}_i = \sum w_i \, y_i \tag{2}$$

The elimination step was repeated until each molecule had been removed, predicting its activity.
Further Eq. (3) was used to calculate r^2 and q^2 where y_i and $\hat{y}_i$ refers to the actual and the predicted activities of the *i*th molecule and y_{mean} is the mean of observed activity of all the molecules in the training set [32].

$$q^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - y_{mean})^2} \quad (3)$$

The value 0.9132 of q^2 indicated the internal stability of the developed model which represented the effectiveness of kNN-SWF method and 3 was found to be the optimum value of *k*.

5.1.5. External Validation

For carrying out the external validation, the Eq. (4) was followed:

$$\text{pred_}r^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - y_{mean})^2} \quad (4)$$

wherein, y_i and $\hat{y}_i$ represents the actual and the predicted activities of the *i*th molecule in the test set and y_{mean} represents the mean of observed activity of all the molecules in the training set [33].

5.1.6. Randomization Test

Y-scrambling or the randomization test is a common method of statistical validation. One-tail hypothesis testing is frequently used to assess the statistical significance of the QSAR model for an actual dataset. The robustness of the models for training sets is examined by comparing these models to those derived for random datasets. Random sets are generated by rearranging the activities of the molecules in the training set. To create the statistical model, numerous randomly rearranged actions (random sets) with the chosen descriptors are used, and the related q2 is calculated. A determined Z score is used to determine the importance of the models that are accordingly obtained.

A Z score value is calculated by the following formula:

$$\text{Z score} = h - \mu/\sigma \quad (5)$$

where h is the q^2 value calculated for the actual dataset, μ the average q^2 and s is its standard deviation calculated for various iterations using models built by different random datasets. The probability (a) of the significance of the randomization test is derived by comparing the Z score value with the critical Z score value as reported, if Z score value is less than 4.0; otherwise, it is calculated by the formula as given in the literature. For example, a Z score value greater than 3.10 indicates that there is a probability (a) of less than 0.001 that the QSAR model constructed for the real dataset is random. The randomization test suggests that all the developed models have a probability of less than 1% that the model is generated by chance [34-36].

5.2. 3D QSAR Study

5.2.1. Dataset of 3D QSAR

To estimate and analyse the anticancer activity of indazole derivatives, three-dimensional QSAR models by SW kNN MFA techniques were generated. For developing predictive QSAR models in kNN QSAR investigations the data set is commonly divided into training, test, and external validation sets. A set of training having 76 analogues, a set of tests having 21 derivatives, and 10% molecules (12 molecules) for the validation set have been used to build the models. The optimal model was chosen using a collection of several statistical parameters q^2 (cross- validated r^2) and pred_r^2 in three fields: steric, electrostatic, and hydrophobic. **Table. 6** displays the statistical results of the SW-kNN MFA techniques. A total of 109 molecules were chosen from the literature and divided into subsets such as training: test: validation set: 76:21:12 which were selected by random selection method and accuracy of which was ensured by uni-column statistics as mentioned in the development of 2D QSAR models. The various conformations of each compound were created and used by the Monte Carlo conformation search process. The random selection approach, which used the metropolis condition to delete or accept created conformers, was used to search for conformations of the compound's structure. Four thousand nine hundred and fourteen (4914) descriptors were created, and any descriptors with a similar value or zero value were eliminated before the model generation, resulting in 4863 descriptors for all chemicals being utilized in subsequent investigations. In the 3D-QSPR approach, the descriptors were defined by computing the molecular characteristics at the point of intersection of a 3D frame. The ligand-based 3D-QSPR modelling approach relies on molecular alignment to get accurate results. Generally, there should be geometric alignment between the bioactive conformations and modeled structures of the selected molecules.

5.2.2. Generation of three-Dimensional-QSAR Models

The k-Nearest Neighbour Molecular Field Analysis approach is used to create 3D QSAR models for the above-given dataset. The results of the 3D variables such as Electrostatic and Steric parameters are calculated by adjusting the dielectric constant to 1.0, the charge type to Gasteiger-Marsili,[37-39] and sp^3 carbon sensor atom with charge 1.0. For electrostatic and

steric energies, the default cut-off energies are 10.0 kcal/mol and 30 kcal/mol. A cross-correlation limit of 0.5, the number of variables in the equation, term selection criterion of q^2, and F-test are used to create 3D QSAR models. Three selection methods used for developing models are Step Wise Variable Selection Method (SW-kNN MFA), Simulated Annealing Variable Selection Method (SA-kNN MFA), and Genetic Algorithm Variable Selection Method (GA-kNN MFA).

5.2.3. k Nearest Neighbour (kNN) Method

Three-dimensional QSAR research was done by the kNN method by following Forward Stepwise Variable Selection as a variable selection technique. The kNN method simply works on a simple distance learning technique in which an unspecified member is selected among the majority of k nearest neighbours in the training set. The standard kNN-MFA method was used in the following manner:

a. The distances here between the unidentified (u) object and the training sets of other objects were estimated.

b. According to the calculated distances, the number of nearest neighbors (k) was selected.

c. Sort the k objects into the group whereby most of them belong. By using the leave-one-out (LOO) cross-validation method, an optimal k value was selected.

Using the Stepwise Variable Selection method, the variables, and optimal k values were selected. This is to improve

(i) the number of nearest neighbours (k) and

(ii) the selection of parameters from the initial pool, the selection approach is combined with the kNN method.

A trial model was developed in the initial stage with one independent parameter, adding another independent variable one step at a time and observing the fit of the model at each step till no more significant variables left outside the model to generate a best 3D-QSPR model. Descriptors were generated over grid, after generating the test and training sets by applied kNN method. Using a charge +1 methyl probe, all energies (steric, electrostatic, and hydrophobic) were estimated at grid lattice locations [40].

5.2.4. Alignment Procedure

The alignment of molecular structures plays a critical role in 3D-QSAR modelling as it strongly determines the predictive accuracy and statistical quality of any given 3D-QSAR model[41,42]. Template-based alignment method in Vlife MDS 4.6 software was used to align the data set of optimized and energy-minimized molecules[43]. The process of aligning a group of compounds using a template structure, this method requires a moving structure in 3D space, by which the conformational flexibility of the molecules was correlated. A common sub-structure from the series has been used as a template as illustrated in **Fig. 7.** The reference structure was a bioactive and stable conformation of the series on which other molecules were aligned which is shown in **Fig. 8**. The investigation included all the aligned structures and the alignment was utilized to determine the putative pharmacophore for the ligand series[44].

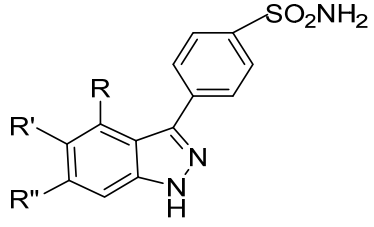

Fig. 7. Structure of Template

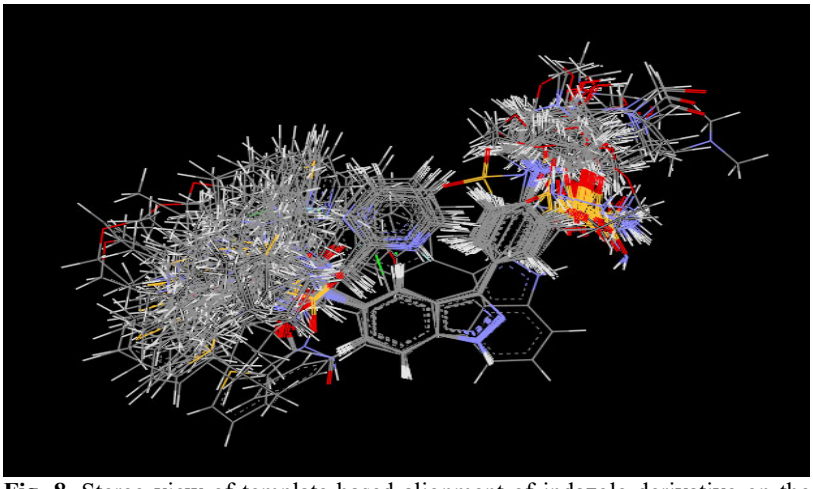

Fig. 8. Stereo view of template-based alignment of indazole derivative on the base template

5.2.5. Model Analysis and Descriptor Estimations

The Vlife MDS 4.6 software was used to count physicochemical parameters of the aligned conformation, after the optimization and energy minimization of the set of molecules. The probe, grid size, and grid interval were chosen to create different descriptors. The distance-dependent dielectric constant is given a value of 1.0, which leads to the calculation of

multiple electrostatic, steric, and hydrophobic descriptors for each compound in separate columns. In 3D-QSPR all invariable columns were removed because they were not playing any role in the 3D-QSAR model creation. The steric and hydrophobic field descriptors had been derived using the Tripos force field and Gasteiger and Marsili charge type electrostatics. Given the distance-dependent dielectric function, the dielectric constant was adjusted to 1.0. A carbon atom with charge 1.0 was the probe set. As a result, descriptors for each electrostatic, steric, and hydrophobic property of each chemical were calculated. The number of data points (molecules) in the training set was used for calculating different parameters such as internal cross-validation regression coefficient (q^2), pred r^2, number of k-nearest neighbours, cross-validated r^2, pred r^2_se, q^2_se, standard error of cross-validation. The model's absolute quality of fitness is demonstrated by the low standard error of pred r^2_se and q^2_se. Finally, the values of q^2 and pred r^2 were utilized to determine which models were the best [44, 45].

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary file]

Acknowledgement

The authors are very thankful for the QSAR study to the V-life sciences technologies Pvt. Ltd.

References

1. Kula K., and Lapczuk-Krygier A. **(2018)** A DFT computational study on the [3+2] cycloaddition between parent thionitrone and nitroethene. *Current Chem. Lett*. 7 (1) 27-34.
2. Harvey J., Himo F., Maseras F., and Perrin L. **(2019)** Scope and Challenge of Computational Methods for Studying Mechanism and Reactivity in Homogeneous Catalysis *ACS Catal.* 9 (8) 6803–6813.
3. Deglmann P., Sch€afer A., and Lennartz C. **(2015)** Application of quantum calculations in the chemical industry—An overview. *Inter. J. Quan. Chem*. 115 107–136.
4. Neves B. J., Braga R. C., Melo-Filho C. C., Moreira-Filho J. T., Muratov E. N., and Andrade C. H. **(2018)** QSAR-based Virtual Screening: Advances and Applications in Drug Discovery. *Front Pharmacol.* 9 1275.
5. Heravi Y. E., Sereshti H., Saboury A. A., Ghasemi J., Amirmostofian M., and Supuran C. T. **(2017)** 3D QSAR studies, pharmacophore modelling and virtual screening of diarylpyrazole–benzenesulfonamide derivatives as a template to obtain new inhibitors, using human carbonic anhydrase II as a model protein. *J. Enzyme Inhib. Med. Chem.* 32 (1) 688–700.
6. Akanksha., Mehta V., Dhingra R., Monika., and Dhingra N. **(2018)** In silico Identification of potential 5α–reductase inhibitors for prostatic diseases: QSAR modelling, molecular docking, and pre ADME predictions. *MOJ. Drug Design Dev. Therapy.* 2 (3) 136–145.
7. Shang C., Hou Y., Meng T., Shi M., and Cui G. **(2021)** The Anticancer Activity of Indazole Compounds: A Mini Review. *Curr. Top. Med. Chem.* 21 (5) 363-376.
8. Chaban T., Rotar D., Panasenko N., Skrobala V., Pokhodylo N., and Matiychuk V. **(2022)** Synthesis, anticancer and antimicrobial properties of some *N*-aryl-2-(5-aryltetrazol-2-yl) acetamides. *Current Chem. Lett*. 11(3) 299-308.
9. Rajora, A. M., Ravishankar D., Zhang H., and Rosenholm J. M. **(2020)** Recent Advances and Impact of Chemotherapeutic and Antiangiogenic Nano formulations for Combination Cancer Therapy. *Pharm.* 12 (6) 592.
10. Akalu Y. T., Rothlin C. V., and Ghosh S. **(2017)** TAM receptor tyrosine kinases as emerging targets of innate immune checkpoint blockade for cancer therapy. *Immunol Rev.* 276 165–177.
11. Zheng L., Chen Z., Kawakami M., Chen Y., Roszik J., Mustachio L. M., Kurie J. M., Villalobos P., Lu W., Behrens C., Mino B., Solis L. M., Silvester J., Thu K. L., Cescon D. W., Rodriguez-Canales J., Wistuba I. I., Mak T. W., Liu X., and Dmitrovsky E. **(2019)** Tyrosine Threonine Kinase Inhibition Eliminates Lung Cancers by Augmenting Apoptosis and Polyploidy. *Mol. Cancer Ther*. 18 1775–1786.
12. Stratford J. K., Yan F., Hill R. A., Major M. B., Graves L. M., Der C. J., and Yeh J. J. **(2017)** Genetic and pharmacological inhibition of TTK impairs pancreatic cancer cell line growth by inducing lethal chromosomal instability. PLoS ONE12: e0174863.
13. Lu N., and Ren L. **(2021)** TTK (threonine tyrosine kinase) regulates the malignant behaviors of cancer cells and is regulated by microRNA-582-5p in ovarian cancer. *Bioengineered.* 12 (1) 5759–5768.
14. Thu K. L., Soria-Bretones I., Mak T. W., and Cescon D. W. **(2018)** Targeting the cell cycle in breast cancer: towards the next phase. *Cell Cycle.* 17 (15) 1871–1885.
15. Liu Y., Lang Y., Patel N. K., Ng G., Laufer R., Li Szi-W., Edwards L., Forrest B., Sampson P. B., Feher M., Ban F., Awrey D. E., Beletskaya I., Mao G., Hodgson R., Plotnikova O., Qiu W., Chirgadze N. Y., Mason J. M., Wei X., Lin D. C. C., Che Y., Kiarash R., Madeira B., Fletcher G. C., Mak T. W., Bray M. R., and Pauls H. W. **(2015)** The Discovery of Orally Bioavailable Tyrosine Threonine Kinase (TTK) Inhibitors: 3-(4-(heterocyclyl) phenyl)- 1*H*- indazole-5-carboxamides as Anticancer Agents. *J. Med. Chem.* 58 (8) 3366-3392.
16. Laufer R., Ng G., Liu Y., Patel N. K. B., Edwards L. G., Lang Y., Li Sze-W., Feher M., Awrey D. E., Leung G., Beletskaya I., Plotnikova O., Mason J. M., Hodgson R., Wei X., Mao G., Luo X., Huang P., Green E., Kiarash R., Lin

D. C. C., Harris-Brandts M., Ban F., Nadeem V., Mak T. W., Pan G. J., Qiu W., Chirgadze N. Y., and Pauls H. W. **(2014)** Discovery of Inhibitors of the Mitotic Kinase TTK Based on N- (3- (3-Sulfamoylphenyl)-1*H*-indazol-5-yl)-Acetamides and Carboxamides. *Bioorg. Med. Chem.* 22 (17) 4968-9714.

17. Martin T. M., Harten P., Young D. M., Muratov E. N., Golbraikh A., Zhu H., and Tropsha A. **(2012)** Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling?. *J. Chem. Inf. Model.* 52 (10) 2570-2578.
18. Verma J., Khedkar V. M., and Coutinho E. C. **(2010)** 3D-QSAR in drug design–a review. *Curr. Top. Med. Chem.* 10 (1) 95-115.
19. Zhao M., Wang L., Zheng L., Zhang M., Qiu C., Zhang Y., Du D., and Niu B. **(2017)** 2D-QSAR and 3D-QSAR Analyses for EGFR Inhibitors. *Hindawi BioMed. Res. Inter.* 2017 1-11.
20. Sharma P. K., and Vakil B. V. **(2017)** Predictive QSAR analysis of flavonoid analogues as antipsoriatic agents. *IJPSR.* 8 (12) 5146-5160.
21. Gopinath P., and Kathiravan M. K. **(2022)** molecular field-based qsar studies and docking analysis of mercaptoquinazolinone benzene Sulfonamide derivatives against HCA XII. *RASAYAN J. Chem.* 15 (1) 686-699.
22. Khedkar S. A., Patil J. S., and Sable P. M. **(2017)** 3D quantitative structure activity relationship of tetrahydroimidazo [1,2-a] pyrimidine as antimicrobial agents. *Marmara Pharm. J.* 21 (3) 644-653.
23. Veerasamy R., and Rajak H. **(2021)** QSAR Studies on Neuraminidase Inhibitors as Anti-influenza Agents. *Turk. J. Pharm. Sci.* 18 (2) 151-156.
24. Chitre T. S., Kathiravan M. K., Bothara K. G., Bhandari S. V., and Jalnapurkar R. R. **(2011)** Pharmacophore optimization and design of competitive inhibitors of thymidine monophosphate kinase through molecular modeling studies. *Chem. Biol. Drug. Des.* 362 (78) 826-34.
25. Bhadoriya K. S., Kumawat N. K., Bhavthankar S. V., Avchara M. H., Dhumal D. M., Patil S. D., and Jain S. V. **(2016)** Exploring 2D and 3D QSARs of benzimidazole derivatives as transient receptor potential melastatin 8 (TRPM8) antagonists using 347 MLR and kNN-MFA methodology. *J. Saud. Chem. Soc.* 20 (S) 256–S270.
26. Fadili M. El., Er-Rajy M., Kara M., Assouguem A., Belhassan A., Alotaibi A., Mrabti N. N., Fidan H., Ullah R., Ercisli S., Zarougui S., and Elhallaoui M. **(2022)** QSAR, ADMET In Silico Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the treatment of Schizophrenia. *Pharmaceuticals (Basel).* 15 670.
27. Asati V., Bharti S. K., Rathore A., and Mahapatra D. K. **(2017)** SWFB and GA Strategies for Variable Selection in QSAR Studies for the Validation of Thiazolidine- 2,4-Dione Derivatives as Promising Antitumor Candidates. *Indian J. Pharm. Edu. Res.* 51 (3) 436-451.
28. Hunashal R. D., and Palkar M. B. **(2017)** Rational Design of Antifungal 1,2,4-triazole derivatives by 2D-QSAR Study. *Int. J. New. Tech. Res.* 3 (4) 88-91.
29. Khan M. S., Ul-Haque Z., Taleuzzaman M., Surana S. S., and Maru A. D. **(2022)** Development of 2D and 3D Quantitative Structure Activity Relationship Models of Thiazole Derivatives for Antimicrobial Activity. *Int. J. Pharm. Sci. Drug. Res.* 14 (2) 164-170.
30. Olasupo S. B., Uzairu A., Shallangwa G., and Uba S. **(2020)** QSAR modelling, molecular docking and ADMET/pharmacokinetic studies: a chemometrics approach to search for novel inhibitors of norepinephrine transporter as potent antipsychotic drugs. *J. Iranian Chem. Soc.* 17 1953-1966.
31. Panigrahi D., Mishra A., and Sahu S. K. **(2015)** Rational in silico drug design of HIV-RT inhibitors through G-QSAR and molecular docking study of 4-arylthio and 4-aryloxy-3-iodopyridine-2(1-H)-one derivative. *Beni-Suef. Uni. J. Bas. App. Sci.* 9 (48) 1-18.
32. Panigrahi D., Mishra A., and Sahu S. K. **(2020)** Pharmacophore modelling, QSAR study, molecular docking and in-silico ADME prediction of 1,2,3-triazole and pyrazolopyridones as DprE1 inhibitor antitubercular agents. *SN. App. Sci.* 2 922.
33. Palkar M. B., Noolvi M. N., Patel H. M., Maddi V. S., and Nargund L. V. G. **(2011)** 2D-QSAR study of fluoroquinolone derivatives: an approach to design anti-tubercular agents. *Inter. J. Drug Desgn. Dis.* 3 559-574.
34. Antre R. V., Oswal R. J., Kshirsagar S. S., Kore P. P., and Mutha M. M. **(2012)** 2D-QSAR studies of substituted pyrazolone derivatives as anti-inflammatory agents. *Med. Chem.* 2 (6) 126-130.
35. Abdi H. **(2010)** Partial least squares regression and projection on latent structure regression. Wiley Interdisciplinary Reviews: *Computational Statistics.* 2 (1) 97-106.
36. Bhatia M. S., Pakhare K. D., Choudhari P. B., Jadhav S. D., Dhavale R. P., and Bhatia N.M. **(2017)** Pharmacophore modeling and 3D QSAR studies of aryl amine derivatives as potential lumazine synthase inhibitors. *Arabian J. Chem.* 10 (1) S100-S104.
37. Silva-Junior E. F. D., Aquino T. M. D., and Araujo-Junior J. X. D. **(2017)** 3D-QSAR and Pharmacophore Identification Studies Applied to Pyridazin-3-one Derivatives as Potent PDE4 Inhibitors. *Acta Sci Pharm Sci.* 1 (5) 22-27.
38. Gasteiger J., and Marsili M. **(1980)** Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges. *Tetrahedron.* 36 3219-28.
39. Suhane S., Nerkar G., Modi K., and Sawant S. D. **(2019)** 2d and 3d-qsar analysis of amino (3-((3, 5-difluoro-4-methyl-6-phenoxypyridine-2-yl) oxy) phenyl) methaniminium derivatives as factor Xa inhibitor. *Int. J. Pharm. Pharm. Sci.* 11 (2) 104-114.
40. Al-Attraqchi O. H. A., and Mordi M. N. **(2022)** 2D- and 3D-QSAR, molecular docking, and virtual screening of pyrido [2, 3-d] pyrimidin-7-one-based CDK4 inhibitors. *J. Appl. Pharmaceutical Sci.*12 (01) 165–175.

41. ElMchichi L., Belhassan A., Lakhlifi A., and Bouachrine M. **(2020)** 3D-QSAR study of the chalcone derivatives as anticancer agents. *Hindawi J. Chem.* 2020 1-12.
42. VLife MDS 4.6 **(2018)** Molecular design suite. Vlife Sciences Technologies Pvt. Ltd. Pune, India.
43. Khare S., Subramani P., Choudhari S., Phalle S., Kumbhar S., Kadam A., and Choudhari P. B. **(2016)** k nearest neighbor and 3D QSAR analysis of Thiazolidinone derivatives as antitubercular Agents. *J. Pharm. Res.* 15 (3) 67-72.
44. Wang J. L., Cheng L. P., Wang T. C., Deng W., and Wu F. H. **(2017)** Molecular modelling study of CP- 690550 derivatives as JAK3 kinase inhibitors through combined 3D QSAR, molecular docking, and dynamics simulation techniques. *J. Mol. Graph Model.* 72: 178–186.
45. Bose P., Mishra M., Gajbhiye A., and Kashaw S. K. **(2019)** QSAR Pharmacophore Mapping and Molecular Docking of 2,4-Diaminoquinazoline as Antitubercular Scaffold: A Computational Hybrid Approach. *Indian J. Pharm. Sci.* 81 (6) 1078-1088.

© 2024 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).