

In silico exploration of c-KIT inhibitors by pharmaco-informatics methodology: pharmacophore modeling, 3D QSAR, docking studies, and virtual screening

Prashant Chaudhari¹ · Sanjay Bari²

Received: 12 March 2015 / Accepted: 14 September 2015 / Published online: 28 September 2015
© Springer International Publishing Switzerland 2015

Abstract c-KIT is a component of the platelet-derived growth factor receptor family, classified as type-III receptor tyrosine kinase. c-KIT has been reported to be involved in, small cell lung cancer, other malignant human cancers, and inflammatory and autoimmune diseases associated with mast cells. Available c-KIT inhibitors suffer from tribulations of growing resistance or cardiac toxicity. A combined in silico pharmacophore and structure-based virtual screening was performed to identify novel potential c-KIT inhibitors. In the present study, five molecules from the ZINC database were retrieved as new potential c-KIT inhibitors, using Schrödinger's *Maestro* 9.0 molecular modeling suite. An atom-featuring 3D QSAR model was built using previously reported c-KIT inhibitors containing the indolin-2-one scaffold. The developed 3D QSAR model ADHRR.24 was found to be significant ($R^2 = 0.9378$, $Q^2 =$

0.7832) and instituted to be sufficiently robust with good predictive accuracy, as confirmed through external validation approaches, *Y*-randomization and *GH* approach [GH score 0.84 and Enrichment factor (*E*) 4.964]. The present QSAR model was further validated for the OECD principle 3, in that the applicability domain was calculated using a "standardization approach." Molecular docking of the QSAR dataset molecules and final ZINC hits were performed on the c-KIT receptor (PDB ID: 3G0E). Docking interactions were in agreement with the developed 3D QSAR model. Model ADHRR.24 was explored for ligand-based virtual screening followed by in silico ADME prediction studies. Five molecules from the ZINC database were obtained as potential c-KIT inhibitors with high in -silico predicted activity and strong key binding interactions with the c-KIT receptor.

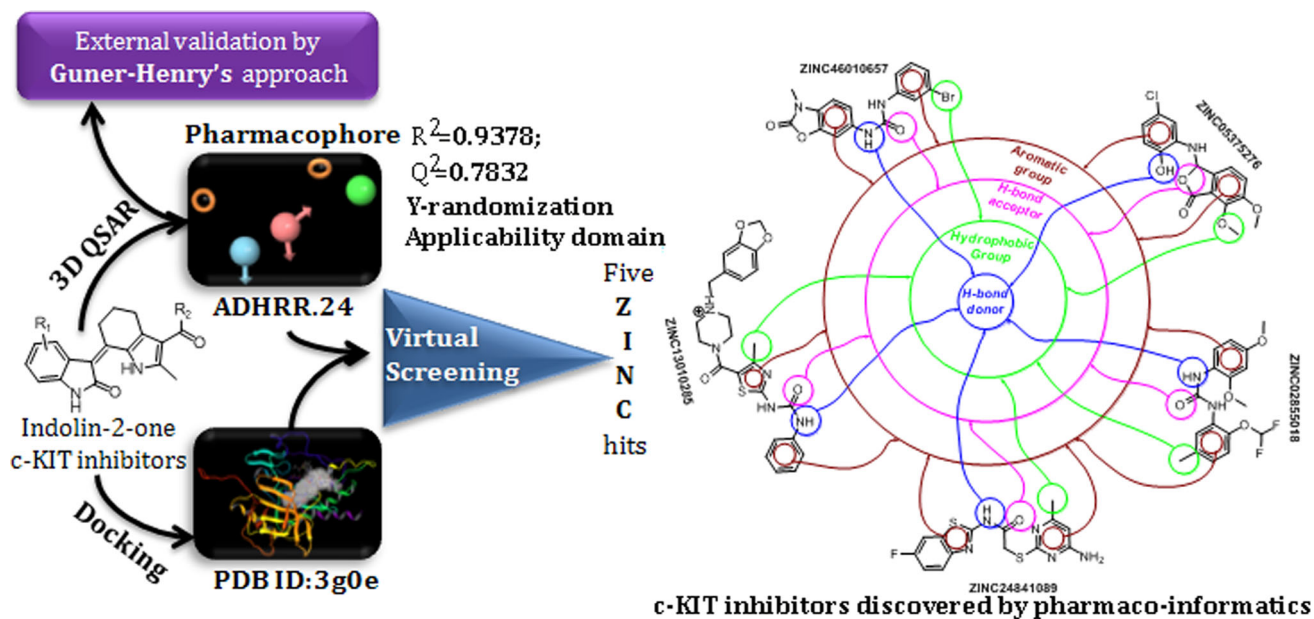
Electronic supplementary material The online version of this article (doi:10.1007/s11030-015-9635-x) contains supplementary material, which is available to authorized users.

✉ Prashant Chaudhari
prashantniperk@gmail.com

¹ Department of Pharmaceutical Chemistry, R. C. Patel Institute of Pharmaceutical Education and Research, Dist-Dhule, Shirpur, Maharashtra 425 405, India

² Department of Pharmaceutical Chemistry, H. R. Patel Institute of Pharmaceutical Education and Research, Dist-Dhule, Shirpur, Maharashtra 425 405, India

Graphical abstract



Keywords c-KIT inhibitors · Indolin-2-one · Pharmacophore · 3D QSAR · ZINC · Docking

Introduction

Protein c-KIT is normally present on the cell surface, wherein it combines with the stem cell factor (SCF) which is involved in the growth of several types of blood cells. Mutation in the c-KIT receptor is observed in certain cancer cells. Quantification of c-KIT from malignant tissue can help in the prognosis and treatment of various types of cancers [1]. Protein tyrosine kinases (PTKs) have critical implications on regulatory signaling mechanisms that are responsible for growth, activation, differentiation, and transformation [2]. c-KIT is also a PTK that is important for mast cell discrimination, propagation, and secretion of cytokines [3]. c-KIT is a component of the platelet-derived growth factor receptor (PDGFR) family which is further categorized as type-III receptor tyrosine kinase (RTK) [4]. A mutation in c-KIT kinase reduces its binding to the stem cell factor. This binding has implications on several human tumors, including gastrointestinal stromal tumors, myeloid leukemia, germ cell tumors, and mastocytosis [5]. Over expression of the c-KIT proto-oncogene has also been observed in case of small cell lung cancer [6]. The percentages of the functional mutation associated with a c-KIT in various cancers are gastrointestinal stromal tumors (90 %), mastocytoma (70 %), sinonasal T-cell lymphoma (16 %), and seminoma/dysgerminoma (9 %) [7].

Imatinib mesylate (Glivec[®]) is the first small-molecule RTK inhibitor that has shown significant results in the management of several cancers. It is effective at the multi-receptor targets including c-KIT, and it has shown remarkable results in the treatment of various tumors, such as dermatofibrosarcoma protuberans and glioblastoma [8]. c-KIT is reported to be involved in the inflammatory and autoimmune diseases related to mast cells [9, 10]. Some preclinical studies revealed that Glivec[®] was efficient against mast cells in rodent arthritis models [11]. Sunitinib malate (Sutent[®]) is another anti-angiogenic multi-targeted PTK inhibitor, effective against c-KIT, PDGFR, Fms-like tyrosine kinase (sFlt1), and VEGFR [8]. Imatinib mesylate and Sunitinib malate are observed to lose efficacy due to drug resistance built by the secondary mutations D816H/V, N822K, Y822K, A829P that are located in the activation loop (A-loop) of c-KIT's catalytic site [12]. Serious cardiotoxicity has been reported for marketed c-KIT inhibitors (Fig. S5 in supplementary material) *viz.* Imatinib (Glivec[®]), Dasatinib (Sprycel[®]), Sunitinib (Sutent[®]), Sorafenib (Nexavar[®]), and Nilotibib (Tasigna[®]) [13].

Sunitinib is one of the first chemical entities from indolin-2-one scaffold as RTKI; however, several reports have reported other c-KIT inhibitors containing the indolin-2-one scaffold [14, 15]. To the best of our knowledge, there are only a few reports that provide an insight into the essential structural features for c-KIT inhibition [16]. In one report, a 3D pharmacophore mapping for the c-KIT was performed using diaryl ureas; however, molecular docking of the QSAR data set molecules on c-KIT was not explored [5].

Our research group is interested in the discovery of potential novel leads (small molecules) as c-KIT inhibitors using a combined ligand- and structure-based virtual screening approach. Pharmacophore mapping and quantitative structure-activity relationship (QSAR) analysis are the ligand-based molecular modeling techniques used by our group. These techniques apply the general studies associated with the interactions of various molecules with the same target. A correct pharmacophore model provides information regarding hydrogen binding properties (acceptor or donor), hydrophobic properties, and aromatic functionality present in compounds in the dataset. This information could be exploited for the characterization of structurally diverse compounds acting on the same bio-molecules [17]. Three-dimensional (3D) QSAR study covers an entire force field around a molecule instead of only spotlighting pharmacophoric information [18]. 3D QSAR takes into account electrostatic, steric, hydrophobic interactions, and hydrogen-bond donor/acceptor factors for the force field calculations. These calculations give the best results when target-recognizing ligands share a unique structural scaffold [19]. Hence, pharmacophore modeling and 3D QSAR can be useful for lead optimization and/or lead modification by rational drug design approach. Moreover, virtual screening using a pharmacophore has proven to be an efficient strategy for lead compound identification [20]. On this ground, to develop a pharmacophore for c-KIT activity, we chose the 3-pyrrolo[β]cyclohexylene-2-dihydroindolinone series recently reported by Ding et al. for c-KIT inhibition [21]. The series offers a diverse range of biological activity against c-KIT where around 65 % of the compounds were more potent than the standard used (Sunitinib). Moreover, the potency of some compounds was 30- to 40-fold less than that of Sunitinib. This diversity motivated us to perform a 3D QSAR study and pharmacophore development.

Methodology

Methods and computational details

Pharmacophore and structure-based virtual screening using the ZINC database was performed to identify novel potential c-KIT inhibitors. 3D-pharmacophore-based screening was carried out using Schrödinger's *PHASE* 3.1 [22,23,26] and molecular docking studies were performed using Schrödinger's *Glide* 5.5 [24,25] as incorporated in *Maestro* 9.0 installed on a PC Pentium IV 3.06 GHz, Core 2 Quad PC with Windows 7 operating systems. Domain of applicability of the developed 3D QSAR model was determined using the open-access online application 'AD using Standardization approach' version 1.0 [27].

Dataset

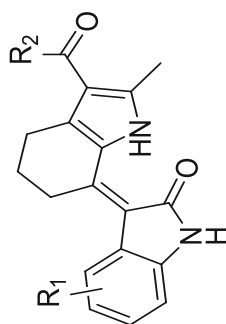
Twenty-five indolin-2-one derivatives previously reported for their c-KIT inhibition profile [21] were taken for the QSAR studies (Table 1). In vitro IC_{50} (inhibition concentrations) values of the molecules against c-KIT were converted to their pIC_{50} values [$-\log(IC_{50})$]. The pIC_{50} values were incorporated as the dependent variable for QSAR model development. The structures of the compounds along with their IC_{50}/pIC_{50} values are specified in Table 1. Out of the 25 compounds used, twenty were randomly selected as the training set entries and the remaining five were selected as the test set entries by using the 'Random training set' option in the *PHASE* 3.1 module. The dataset was divided into the training and test set on the grounds of suggestions given by Golbraikh et al. [28]. A minimum of five compounds was necessary to include in the test set and this selection was done by satisfying three conditions: (1) all representative points of the test set in the multi-dimensional descriptor space must be close to those of in the training set, (2) vice-versa with that of in the training set, (3) the representative points of the training set must be distributed within the whole area occupied by the entire dataset. The training set was used to generate a QSAR model, and the test set was used to validate the generated model [28].

Pharmacophore mapping and 3D QSAR development

All the compounds were sketched using *Maestro* 9.0 and incorporated as a separate entry in the project table. For defining a 'pharma set' in *PHASE*, an activity threshold range was selected in such a way that compounds are active if pIC_{50} value is ≥ -1.1073 and inactive if ≤ -1.500 . This resulted in nineteen 'actives' compounds and six 'inactives' compounds. The steps carried out for pharmacophore and 3D QSAR development are depicted in Fig. 1.

Energy minimization of the structures was performed using Schrödinger's *LigPrep* module version 2.3 selecting the (OPLS)-2005 force field method. The ionization state of the molecules was set to pH 7.0. The majority of the indolin-2-one analogs used in this study were flexible, and all possible conformations were generated to retrieve the active conformer able to bind to c-KIT. A 3-dimensional spatial representation of the structures was necessary for the development of a true pharmacophore model [29].

We chose an atom-based rather than a pharmacophore-based 3D QSAR because an atom-based 3D QSAR model covers the entire structural space around a molecule, whereas the pharmacophore-based model considers only the required pharmacophoric group points [30] (a selection of type of model is separately discussed in the supplementary material). In *PHASE* 3.1, pharmacophoric properties such as a hydrogen-bond donor (D), hydrogen-bond acceptor (A), aro-

**Table 1** Results of 3D QSAR and docking analysis

Compounds	Structure	IC ₅₀ (nM)	Observed pIC ₅₀	Predicted pIC ₅₀	Confirmations	Fitness	Docking score	Glide energy (kcal/mol)
1^b	H	11.7	-1.072 ^d	-0.97	6	2.82	-7.95	-49.48
2^a	NH(CH ₂) ₂ N(CH ₂ CH ₃) ₂	2.4	-0.38 ^d	-0.49	9	2.91	-10.93	-57.25
3^b	NH(CH ₂) ₂ N(CH ₂ CH ₃) ₂	2.7	-0.431 ^d	-0.59	8	2.9	-10.62	-53.90
4^b	NH(CH ₂) ₂ N(CH ₂ CH ₃) ₂	2.3	-0.362 ^d	-0.43	9	2.85	-10.67	-53.28
5^b	NH(CH ₂) ₂ N(CH ₂ CH ₃) ₂	4.1	-0.613 ^d	-0.60	8	2.91	-10.38	-55.73
6^a	NH(CH ₂) ₂ N(CH ₂ CH ₃) ₂	35.5	-1.55 ^c	-1.52	8	2.47	-9.59	-50.99
7^b	NH(CH ₂) ₂ N(CH ₂ CH ₃) ₂	6.9	-0.839 ^d	-0.88	8	2.85	-9.42	-50.19
8^b	NH(CH ₂) ₂ N(CH ₂ CH ₃) ₂	255.7	-2.408 ^c	-2.41	10	2.23	-5.75	-44.42
9^b	NH(CH ₂) ₂ N(CH ₂ CH ₃) ₂	397.6	-2.599 ^c	-2.55	25	2.24	-4.69	-39.46
10^b	NH(CH ₂) ₂ N(CH ₂ CH ₃) ₂	43.1	-1.634 ^c	-1.69	10	2.33	-6.12	-44.09
11^a	NH(CH ₂) ₂ N(CH ₂ CH ₃) ₂	180.8	-2.257 ^c	-2.16	9	2.14	-5.80	-43.81
12^b	NH(CH ₂) ₂ -morpholin-4-yl	2	-0.301 ^d	-0.42	9	2.92	-10.55	-53.77
13^b	NH(CH ₂) ₂ -piperidin-1-yl	5.8	-0.763 ^d	-0.78	8	2.91	-10.59	-54.23
14^b	NH(CH ₂) ₂ -pyrrolidin-1-yl	7.4	-0.869 ^d	-0.78	10	2.94	-9.16	-49.87
15^b	NH(CH ₂) ₂ N(CH ₃) ₂	6.6	-0.82 ^d	-0.92	5	3	-10.54	-51.82
16^b	NH(CH ₂) ₂ OH	1.4	-0.146 ^d	-0.19	23	2.93	-9.77	-46.98
17^b	NH(CH ₂) ₂ -pyridin-2-yl	37.7	-1.576 ^c	-1.67	18	2.46	-6.14	-41.11
18^b	NH(CH ₂) ₃ N(CH ₂ CH ₃) ₂	9.9	-0.996 ^d	-1.02	30	2.84	-8.84	-48.36
19^a	NH(CH ₂) ₃ -morpholin-4-yl	5.7	-0.756 ^d	-0.77	35	2.83	-10.22	-51.83
20^b	NH(CH ₂) ₃ -pyrrolidin-1-yl	10.7	-1.029 ^d	-1.01	27	2.85	-6.58	-45.06
21^b	NHCH ₂ CH(OH)CH ₂ N(CH ₂ CH ₃) ₂	4.1	-0.613 ^d	-0.60	61	2.78	-10.24	-52.87
22^b	Morpholin-4-yl	2.1	-0.322 ^d	-0.37	2	2.68	-10.41	-55.18
23^a		7.8	-0.892 ^d	-0.85	4	2.61	-9.14	-48.66
24^b	5-F	8.5	-0.929 ^d	-0.89	4	2.68	-8.84	-48.36
25^b	Sunitinib_	8.9	-0.949 ^d	-0.95	17	2.69	-10.52	-53.13

^a Test set molecules^b Training set molecules^c Pharm set: inactive^d Pharm set: active

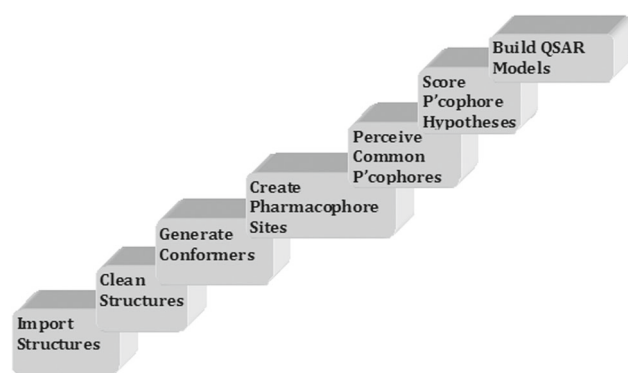


Fig. 1 Steps in 3D QSAR

matic ring (R), hydrophobic moiety (H), negatively ionized group (N), and positively ionized group (P) were taken into consideration. Pharmacophore hypotheses were created via a ‘tree-based partitioning’ method after selecting 1 Å as the terminal box size. The 1 Å sized grid box represented a common pharmacophore covering the least required active-set molecules. Boxes retained by the partitioning procedure were kept, and the rest were eliminated. Common pharmacophore hypotheses (CPHs) indicating at least five sites (common to all 25 molecules) were allowed for the further development.

A total of 630 hypotheses were identified (Table S1 in supplementary material). All common pharmacophores were scrutinized, and a ‘scoring protocol’ was applied to select the pharmacophore from every grid box. Scoring functions offer a grading to the different hypotheses and assist in the selection of proper hypotheses for 3D QSAR model building. Vector and Site alignment scores were first calculated using default parameters in *PHASE*. ‘Survival inactives’ score was also calculated after obtaining a ‘survival score’ for the actives. To refine the hypotheses and minimize a chance of penalty for matching inactives, a re-scoring was performed. As a result of this, we got a post hoc score. Post hoc score is the *survival score* calculated by the sum of the site, volume, vector, and selectivity scores. A total of 154 hypotheses were obtained (Table S2 in supplementary material) and only the top six hypotheses with a post hoc score

value ≥ 4.000 were included in the QSAR model generation step.

A rationale behind a selection of the top six hypotheses is provided in the supplementary material of this article. For the QSAR model building, grid spacing size was kept to 1.00 Å, partial least squares (PLS) factor was kept as four by the rule of $N/5$ [23], where N represents the number of molecules taken in the training set. The single best model was selected after evaluating the statistical results and other validation parameters. The selected model was used for the pharmacophore supported virtual screening.

Validation of 3D QSAR model

Validation parameters were analyzed for hypotheses ADDRR.2, ADDRR.4, ADHRR.23, ADHRR.24, ADDHR.49, and ADDHR.14; of which **ADHRR.24** provided the best results (Table 2). Internal validation at PLS factor four was performed by *leave-n-out* predictions on the training set. This internal validation parameter (R^2) cannot provide a reliable and correct forecasting for the molecules outside the training set. Hence, external validation becomes necessary. External validation was carried out using few extra methods. Q^2 (test set) (q^2 for the predicted activities) and *Pearson R* (the correlation between the predicted and actual activity for test set) were obtained. Other strict requirements for a model acceptance, i.e., R_0^2 , R'^2_0 , k , and k' were also calculated. R_0^2 (Predicted vs. observed activities) and R'^2_0 (observed versus predicted activities) are the coefficients of correlation that obtains by regression lines through the origin with the intercept set to zero. Generally for a model with good predictive ability, R_0^2 or R'^2_0 must be equal to or lower than R^2 . For the slope value k (predicted versus observed activities) or k' (observed versus predicted activities), suitable range is $0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$ [31,32].

The robustness of the developed model was further checked by a *Y*-randomization test. This methodology involves a random scrambling of dependent variable (*Y*) values to produce new dissimilar sets. The newly produced models after random shuffling are expected to have a low

Table 2 Statistics for the selected hypotheses **ADHRR.24**

PLS factor (no.)	SD	R^2	F	p	RMSE	Q^2	Pearson R
1	0.396	0.6514	33.6	3.702e-005	0.2573	0.4577	0.7645
2	0.2408	0.8365	51.3	1.682e-008	0.2496	0.6156	0.7864
3	0.1418	0.9103	79.7	2.025e-011	0.1837	0.7321	0.9004
4	0.109	0.9378	107.4	4.094e-012	0.1729	0.7832	0.9116

SD Standard deviation, R^2 regression coefficient, F variance, p level of significant variance, RMSE residual mean square error, Q^2 value of q^2 for the predicted activities, *Pearson R* correlation coefficient between the predicted and experimental activity for the test set

significant correlation coefficient ($R^2_{Y\text{-randomization}}$) compared to the R^2 of the original model and, if the reverse happens, then the selected QSAR model cannot be considered valid for a particular modeling method and dataset [33]. A scatter graph of the observed versus predicted activity for the test set molecules was drawn and analyzed.

In another method, validation was carried out by the *GH* score approach [34] using an external set of molecules to examine model ADHRR.24 and identify the c-KIT antagonists already included in dataset. In an external dataset, 174 molecules were incorporated: 32 potent c-KIT inhibitors from the freely available BindingDB database (<http://www.bindingdb.org>) [35–37] and 142 molecules dissimilar to the indoline-2-one core randomly taken from the NCI open database [38] (structures of inhibitors and NCI Open molecules are provided in the supplementary material). Once the dataset of all 174 molecule was prepared (pre-processing of the molecules is given in supplementary material), molecular matches for the ADHRR.24 model were searched using *PHASE*. Based on the number of molecular hits obtained, statistical parameters of the *GH* method and *GH* score were calculated.

According to the Organization for Economic Co-operation and Development (OECD) Principle 3, a defined ‘domain of applicability’ must be given to a developed QSAR model [33]. The applicability domain of any QSAR model considers the space of descriptors, scope of biological activity, and knowledge or information on which the training set of a model has been developed. This domain can be utilized for the screening and identification of potential new hits.

An applicability domain assessment can be made using several reliable methods, i.e., ranges in the descriptor space, geometrical methods, distance-based methods, probability density distribution, and range of the response variable [27]. Recently, Melagraki and Afantitis reported the calculation of an applicability domain by similarity node (measures the *Euclidean distances*) and leverages node (measures the *Leverages*) [33,39]. For the present 3D QSAR model (ADHRR.24), we determined the AD using the open-access online application ‘*Applicability domain using standardization approach 1.0*’, developed by Roy et al. [27,40].

Combined virtual screening

In an attempt to identify new hits that could potentially inhibit c-KIT, a screening protocol was followed using the pharmacophore ADHRR.24. The modus operandi for virtual screening is illustrated in Fig. 2 where 236,475 molecules were taken from the freely available ZINC database (<http://zinc.docking.org>) [41] (see supplementary material). These molecules were already available in a pre-filtered form based on molecular weight 150–500 Da, XlogP < 5, rotatable

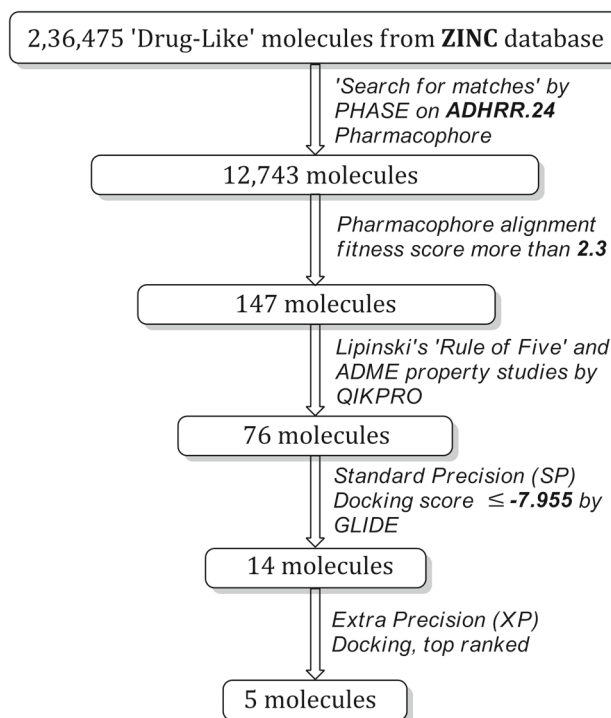


Fig. 2 In silico modus operandi of virtual screening

bonds < 7, topological polar surface area (TPSA) < 150, H-bond donor < 5, H-bond acceptor < 10. The hypothesis ADHRR.24 was put on the query to ‘search for matches’ for 236,475 molecules and 12,743 molecules were retrieved as hits. These hits were sorted further using a *fitness score*. When a fitness score was limited to ≥ 2.3 , a total of 147 molecules remained as the top ranked. The ADME properties for these 147 molecules were calculated using *QikProp* version 3.2 and those molecules falling out of the Lipinski’s ‘Rule of Five’ druggability guidelines [42] were removed from the study. The remaining 76 molecules were subjected to a docking study using standard precision (SP) docking against the c-KIT kinase domain [12]. Molecules with a Glide score value ≤ -7.955 were docked again this time using ‘Extra Precision’ (XP) docking.

Docking studies

Initially, all 25 molecules from the QSAR data set, in the “3D conformers” form were docked into c-KIT (PDB ID: 3G0E; resolution, 1.6 Å). This docking was performed to observe the interactions of all active and inactive compounds with the receptor and to examine the agreement between the docking pattern and predicted activity of the validated pharmacophore. Seventy-six molecules retrieved after pharmacophore-based screenings were subjected to receptor-based virtual screening by docking. For docking, c-KIT was analyzed and preprocessed using *Maestro*’s “Pro-

tein Preparation Wizard.” A receptor grid was then generated (methodology for receptor grid generation is provided in the supplementary material). All ligands were processed using the *LigPrep* protocol.

The generated receptor grid and 3D structures of the ligands were selected in the “Ligand Docking” panel. At this stage of the experiment, all molecules from the QSAR dataset were selected. Furthermore, 76 molecules obtained from the pharmacophore-based virtual screening were selected for the SP docking mode. Ligands were docked flexibly, and the nonplanar conformation of amide bonds was penalized. In flexible docking, 5000 poses per ligand were allowed and only the top 1 % was retained for energy minimization. On the basis of an SP Glide score of ≤ -7.955 , only 14 out of 76 molecules were included in a XP docking study.

Results and discussion

Development and validation of 3D QSAR model

To produce a 3D QSAR model for ligand-based virtual screening, we mapped a pharmacophore using previously reported indolin-2-one derivatives as c-KIT inhibitors. After completing rigorous steps toward finding a common pharmacophore hypothesis (CPH); six hypotheses namely ADDRR.2, ADDRR.4, ADHRR.23, ADHRR.24, ADDHR.49, and ADDHR.14 (Table S2 in supplementary material) were selected to build our 3D QSAR model. These six hypotheses were selected by top *survival score* (≥ 3.8) and

post hoc (≥ 4.0). At PLS factor four, 3D QSAR models were generated. A statistical analysis of the results identified hypothesis ADHRR.24 to be the best one. ADHRR.24 was granted on the grounds of the statistical results obtained for the internal validation against the training set and external validation through the test set molecules.

The regression coefficient R^2 (0.9378) was high for the model; however, as suggested by Golbraikh et al. [28], this parameter alone is not sufficient to explain the robustness of a 3D QSAR model. Hence, additional parameters were considered. Q^2 is a cross-validated correlation coefficient that is obtained using the leave-one/*N*-out method and was high (0.7832). The Q^2 parameter is more reliable than the regression coefficient because it is calculated using an external validation method employing the test set molecules. Standard deviation (SD) of the regression was low (0.109). Moreover, the residual mean square error (RMSE) was within the limit (0.1729), and the *Pearson R* correlation coefficient of the test set for the correlation between the observed and predicted activities was high (0.9116). All the statistical results for the best model are listed in Table 2. The predicted activities of the dataset molecules obtained using ADHRR.24 hypothesis are provided in Table 1. Other statistical measurements, i.e., R_0^2 and $R'0^2$ were 0.9174 and 0.9263, respectively, whereas k and k' were 0.96 and 0.94, respectively. All the parameters were within an acceptable range, and a small residual difference between the observed and predicted activities proved that we selected the best 3D QSAR model.

Angles and distances between the sites for the ADHRR.24 model are shown in Fig. 3a, b. QSAR visualization of the

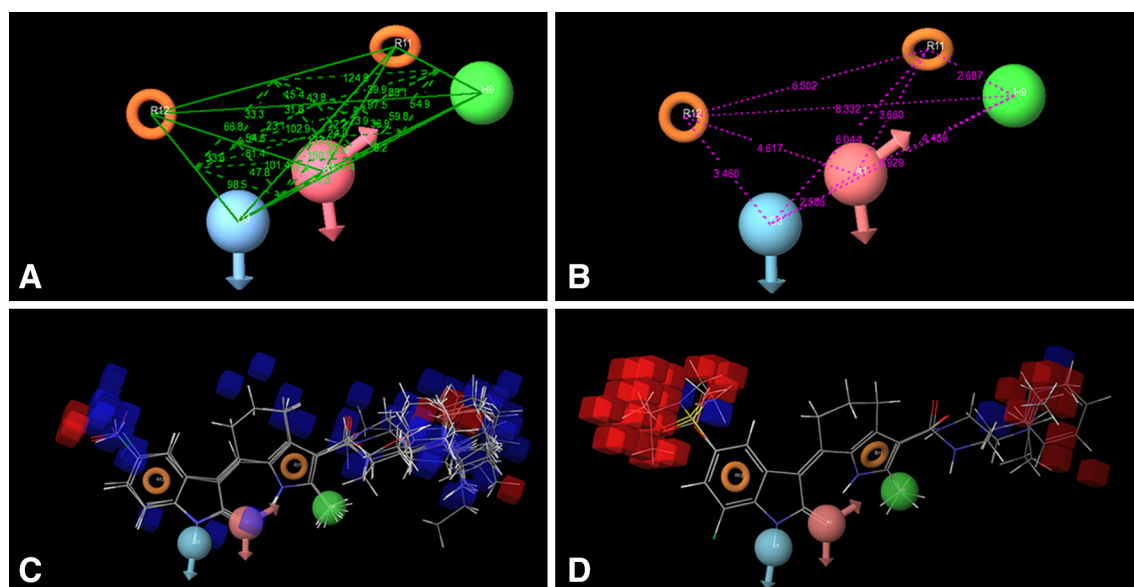


Fig. 3 **a** Angles between sites; **b** distances between sites D3: H-bond donor, A1: H-bond acceptor, R11: aromatic ring, R12: aromatic Ring, H9: Hydrophobic group; **c** atom-based 3D QSAR model ADHRR.24

visualization, *blue cubes* signify favorable sections from active molecules in the data set; *d red cubes* signify unfavorable sections from the inactive molecules in the dataset. (Colour figure online)

active molecules (Fig. 3c) revealed that D3 (H-bond donor feature) arises because of the $-NH$ bond of 2-oxoindoline moiety and A1 (H-bond acceptor feature) arises from $>C=O$ of the 2-oxoindoline moiety. H9 (hydrophobic group) was aligned on the free $-CH_3$ group of the 2-methyl-4,5,6,7-tetrahydro-1H-indole moiety. The aromatic rings R11 and R12 were visualized on the benzene ring of 2-oxoindoline and pyrrole of 2-methyl-4,5,6,7-tetrahydro-1H-indole respectively. The collective effects of (A), (D), (H), (R) and other features were observed by a QSAR visualization panel of PHASE. The blue cubes in Fig. 3c indicate the favorable features causative to ligand interactions and the red cubes in Fig. 3d indicate unfavorable features mainly present in the inactives for the c-KIT receptor. To our observation, carboxylate or *N,N*-diethylsulfamoyl substitutions at the fifth position and methoxy substitution at the 4th or 7th position of the 2-oxoindoline moiety were responsible for the unfavorable interactions in the active site of c-KIT. Moreover, a fitness score value of the inactives was low (Table 1). Here, it could be emphasized that the fitness score is a crucial parameter in studying the demarcation lines between the active and inactive molecules because it is an indicative measure of mapping pharmacophore site points towards the ligands. It also measures how well vector characteristics [(A), (H), (N), (D), (P), and (R)] overlay on the selected hypothesis, and how the conformations superimpose in an overall sense. The regression line equation (observed versus predicted activity) obtained through the scatter graph (Fig. 4) for the combined training and test set molecules was, $y = 0.94x + -0.02$ ($R^2 = 0.94$).

The *Y*-randomization test was carried out by shuffling the observed biological activity (pIC_{50}) at 100 random trials for the same number of training set molecules and similar group

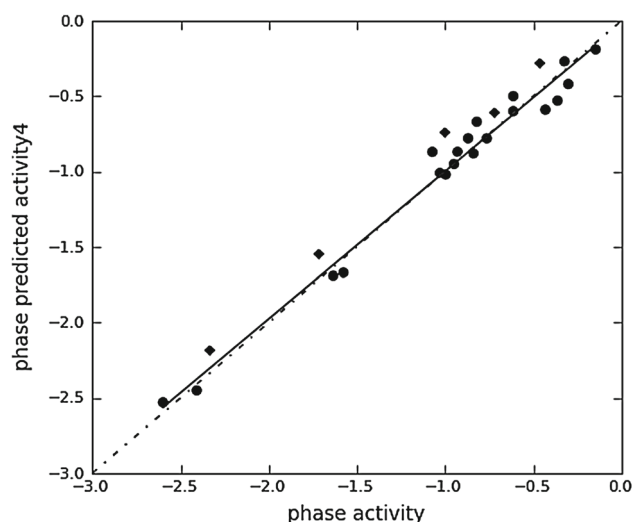


Fig. 4 Scatter graph for training and test set molecules; *filled spheres* training set molecules; *filled diamonds* test set molecules

features. The new correlation coefficients ($R_{Y-randomization}^2$) were in the range of 0.08 to 0.4329, and were found to be lower than the original correlation coefficient ($R^2 = 0.9378$). This proves that our model ADHRR.24 was not obtained by mere chance.

Further, we assessed the forecasting ability of the ADHRR.24 hypothesis using the Guner and Henry's 'goodness of hits' approach. The 174 molecules taken for this method were put on a search query using the ADHRR.24 hypothesis. The number of hits (Ht) obtained was 23 (a list of active hits (Ht) along with their PubChem ID and IUPAC name is given in Table S3 of supplementary material). Out of the 23 hits, 21 molecules (Ha) were from the 32 c-KIT inhibitors. The two molecules (Ht–Ha) were not from the active c-KIT inhibitors hence, considered as 'False positives'. Nine molecules (A–Ha) that were active against c-KIT were identified as inactive by our model. Hence, these nine molecules were classified as 'False negatives' in our study. The GH protocol [43] involves some calculation where %A is a ratio of actives [(Ha/A) \times 100] was found to be 65.63 %, and %Y is a yield of inhibitors [(Ha/Ht) \times 100] was 91.30 %. The enrichment of the concentration of active inhibitors by the model when compared to a random screening devoid of any other hypothesis approach, i.e., enrichment factor (*E*) was found to be (4.964) (calculated as [(Ha/Ht)/(A/D)]). The GH score obtained by formula [Ha (3A+Ht)/4Aht] \times [1–(Ht–Ha/D–A)] was 0.84. The range of the GH score is 0 to 1, where a score closer to 1 indicates a model to be near to ideality; thus based on our *E* and GH score, our model was valid to be utilized for virtual screening.

The applicability domain was determined by a 'standardization approach'. All descriptors of the QSAR dataset molecules, their predicted biological activities (pIC_{50}) and residual were included in determining the AD. The results suggested that no compound in the training set was observed as an X-outlier, and none of the molecules from the test set were obtained as 'Outside AD' (Table S4, supplementary material).

Docking analysis of dataset molecules

We performed an SP mode docking for all indolin-2-ones from the QSAR dataset against the c-KIT kinase domain (PDB ID 3G0E) using GLIDE program. We noted that residues CYS673, GLU671, ALA621, VAL603, CYS809, PHE858, GLY676, CYS674, LEU595, and ASP677 were involved in the ATP recognition at the catalytic site in c-KIT. Most of the active compounds at the energy-minimized states showed the same binding pattern as that of the co-crystallized inhibitor Sunitinib. From these, compound **14** showed similar H-bonding interactions as that of Sunitinib (Fig. 5a).

In the hinge region of the ATP binding site, the 2-oxoindoline moiety displayed two key hydrogen bonding

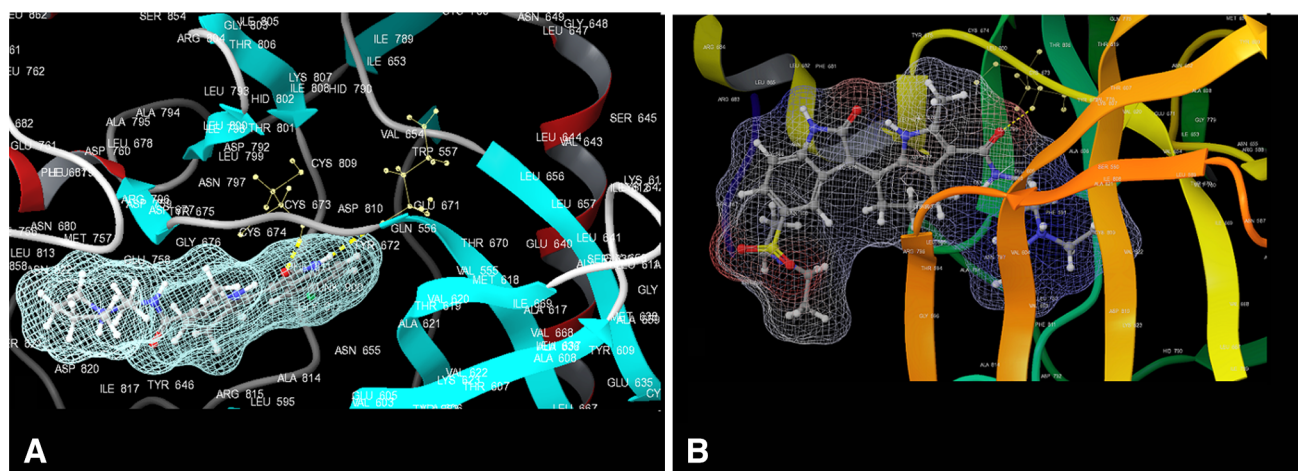


Fig. 5 **a** Compound **14** ($IC_{50} = 7.4$ nM) showing H-bond interaction (yellow dash lines) with CYS673 & GLU671; **b** most inactive Compound **9** interacts only with CYS673 (shown by yellow balls with sticks). (Colour figure online)

interactions: the carbonyl fraction of the 2-oxoindoline ring formed a bond with the amide –NH of CYS673 and the neighboring –NH group donated a hydrogen bond to the carbonyl of GLU671. An intra-molecular hydrogen bonding was also observed between the ‘O’ atom of indolin-2-one and the NH of 4,5,6,7-tetrahydro-1H-indole. This demonstrates that if the 4,5,6,7-tetrahydro-1H-indole moiety remains in an association with 2-oxoindoline, the rotation of bond is restricted, helping the molecules to maintain both intra and inter H-bond interaction. These facts are observed in the majority of the active compounds (Table 1).

The amino group in most of the compounds did not engage in with the receptor. However, compounds **12**, **13**, **14**, **19**, and **22** (with cyclic structures at the tail) showed hydrophobic interactions with LEU813, ILE817, PHE858, TYR646, and VAL514 residues. In compound **9**, a sulfamido substitution at the 5th carbon of 2-oxoindoline demonstrated the steric obstacles with the amino acid residues, that resulted in molecules positioned away from the desired binding interaction site (Fig. 5b). Surprisingly, compound **16**, which possesses a hydroxyl group at the tail portion, exhibited a lower docking score (-9.77) and glide energy (-46.98) than other active compounds (Table 1).

Virtual screening

Initially, we formulated a ligand-based design to identify the new potential c-KIT inhibitors for cancer. We put our ADHRR.24 hypothesis on a search query for ‘drug-like’ molecules from the ZINC database. A PHASE search found 12,743 molecules as matches for the selected pharmacophore (ADHRR.24). The *fitness* score filter (≥ 2.3) returned 147 molecules, and the hits were further subjected to Lipinski’s ‘Rule of Five’ and ADME studies. From these, 76 molecules were retrieved through a pharmacophore-based virtual

screening. An SP docking method was performed on these 76 molecules and only 14 molecules (glide score ≤ -7.955) were further included for XP docking. Finally, 5 hits were obtained (Fig. 6).

Predicted pIC_{50} values of the final hits for c-KIT inhibition are given in Table 3. The hit compound ZINC46010657 (1-(3-bromophenyl)-3-(3-methyl-2-oxo-2,3-dihydrobenzo[d]oxazol-6-yl)urea) was predicted to be the most active ($pIC_{50} = -0.467$) against c-KIT. Moreover, the fitness value of molecule ZINC05375276 (1-((5-chloro-2-hydroxyphenyl) amino)-4,5-dimethoxyisobenzofuran-3(1H)-one) for the ADHRR.24 hypothesis was found to be the highest (2.58) among the final five hits. All ADME parameters for compounds ZINC46010657, ZINC05375276, ZINC02855018, ZINC13010285, and ZINC24841089 were calculated using *Qikprop* and shown to be were within an acceptable range (Table 3).

The binding interactions and alignments of the final five hits from the developed pharmacophore are illustrated in Fig. 7. A recent report on the wild-type and D816V mutant-type c-KIT receptor inhibitors suggested that compounds with a bidentate bonding with CYS673 in the hinge region exhibit more potent inhibitory activity against the D816V-mutated c-KIT receptor than that of one singly bonded with CYS673. The mutant-type D816V c-KIT receptor is resistant to marketed drugs viz. Imatinib and Sunitinib [44]. The hit compound ZINC02855018 (1-(2-(difluoromethoxy)-5-methylphenyl)-3-(2,4-dimethoxyphenyl)urea) demonstrated a bidentate hydrogen binding profile where the amide bond of a diphenyl urea moiety showed two hydrogen bonds with CYS673 (purple) and, to our surprise, the fluorine (F) atom also showed a strong hydrogen bond with an amino group of LYS593 (green) (Fig. 7e). Hence, the possibility of ZINC02855018 being explored for the mutant-type D816V c-KIT receptor cannot be neglected.

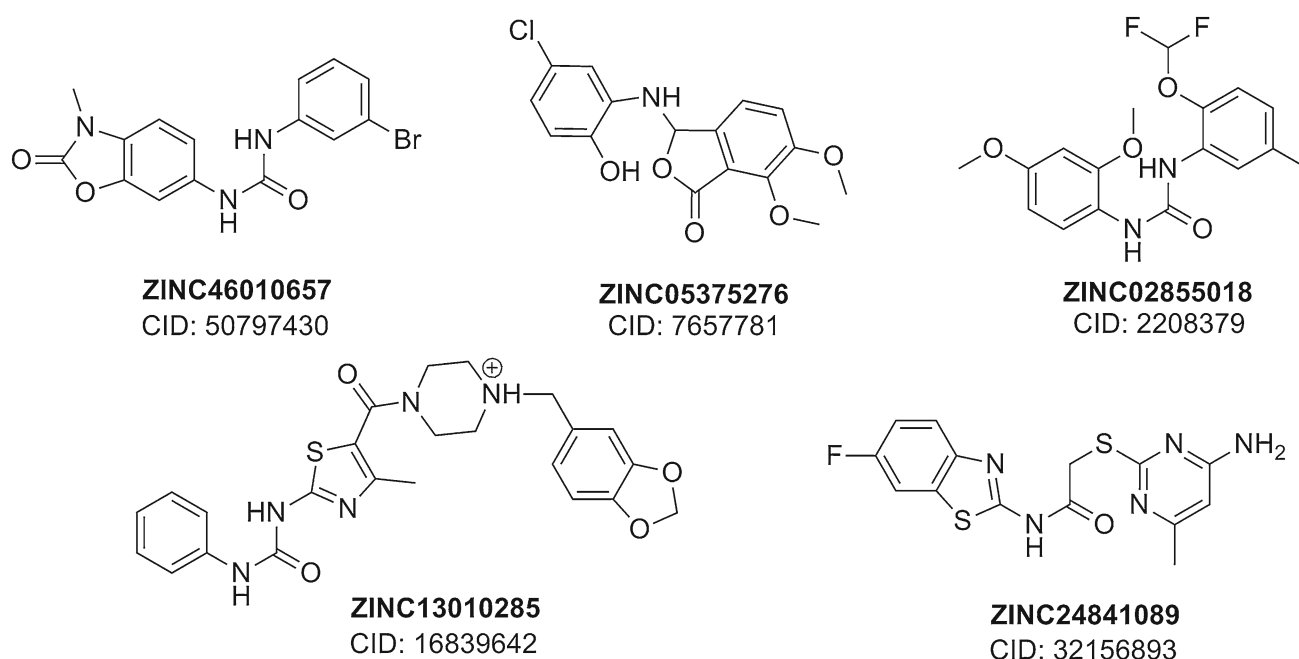


Fig. 6 Structures and PubChem IDs of final 5 ZINC hits obtained through ligand-based and structure-based virtual screening

Table 3 Results of final 5 ZINC hits by virtual screening

ZINC comp IDs	3D QSAR results		XP docking results		ADME results by <i>QikProp</i> (Schrodinger 9.0, LLC, NY, USA)				
	Fitness score	Predicted activity (pIC ₅₀)	Docking score	EvdW (XP docking) ^a	Mol. wt	logP ^b o/w	logS ^c	MDCK ^d	% human oral absorption ^e
46010657	2.41	−0.467	−9.64	−5.25	362.182	2.19	−3.485	693.071	78.071
05375276	2.58	−0.61	−8.89	−6.46	335.74	2.86	−2.961	1520.122	76.243
02855018	2.39	−0.674	−8.74	−5.9	352.337	3.237	−4.173	3588.125	100
13010285	2.33	−0.87	−8.12	−5.55	479.553	2.194	−3.511	143.315	75.474
24841089	2.44	−0.937	−7.53	−4.82	349.4	2.235	−4.385	869.02	83.225

^a Lipophilic term resulting from hydrophobic grid potential and fraction of the total protein ligand vdW energy

^b Predicted octanol/water partition coefficient logP (range: −2.0 to 6.5)

^c Predicted aqueous solubility (S) in mol/L (range: −6.5 to 0.5)

^d Predicted apparent MDCK cell permeability in nm/s.

^e Percentage of human oral absorption (25 % ≥ poor; 80 % ≤ high)

Molecule ZINC24841089 (2-((4-amino-6-methylpyrimidin-2-yl)thio)-*N*-(6-fluorobenzo[d]thiazol-2-yl)acetamide) was well aligned on the developed pharmacophore (Fig. 7j); however, docking studies revealed that a nitrogen atom of benzthiazole moiety and free −NH of amide functionality form a bidentate hydrogen bonding with the backbone scaffold of CYS673 (Fig. 7i). Hence, this hit molecule has a strong potential to exhibit good in vitro activity against the D816V mutant c-KIT receptor.

The five ZINC hits identified as potential c-KIT inhibitors (Fig. 6) were further subjected to a PubChem structure query search [45]. Our search indicated that our virtual hits have not

been reported as inhibitors of wild-type and/or mutant-type c-KIT.

Conclusion

Pharmacophore identification and 3D QSAR model development were carried out using a series of previously reported indolin-2-ones with well defined c-KIT inhibitory activity. The common pharmacophore hypothesis (ADHRR.24) was generated containing five features; two aromatic rings (R) mainly from indolin-2-one and pyrrole, one hydrophobic (H) group (−CH₃), one acceptor (A) group (C=O), and one donor

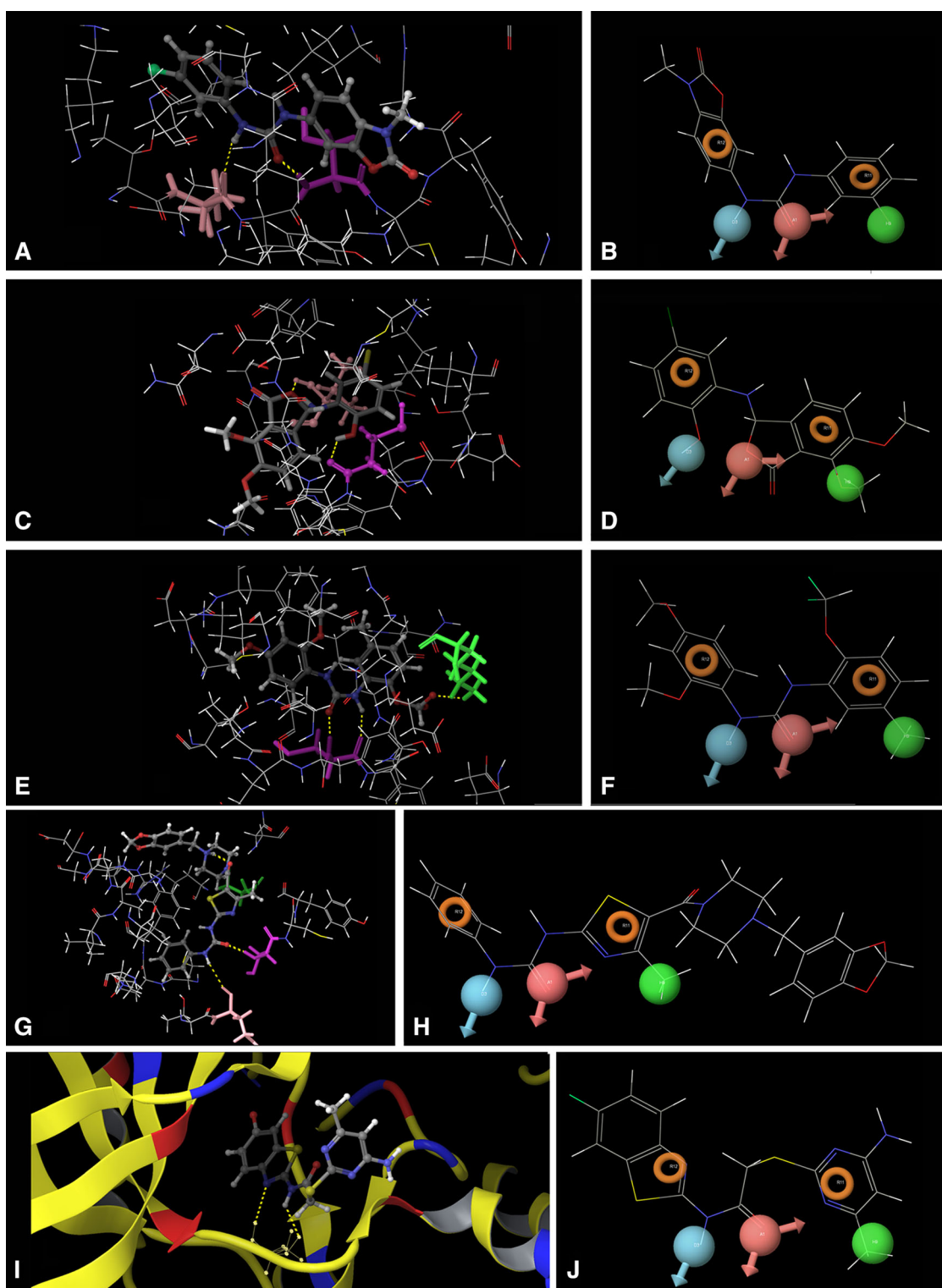


Fig. 7 **a** H-bond (yellow dash line) of ZINC46010657 with GLU671 (pink) & CYS673 (purple); **b** alignment of ZINC46010657 on ADHRR.24; **c** H-bond (yellow dash line) of ZINC05375276 with GLU671 (pink) & CYS673 (purple); **d** alignment of ZINC05375276 on ADHRR.24; **e** bidentate H-bond (yellow dash line) of hit ZINC02855018 with CYS673 (purple) & single with LYS593 (green);

f alignment of ZINC02855018 on ADHRR.24; **g** H-bond (yellow dash line) of ZINC13010285 with GLU671 (pink) & CYS673 (purple); **h** alignment of ZINC13010285 on ADHRR.24; **i** bidentate H-bond (yellow dash lines) of ZINC24841089 with CYS673 (shown by yellow balls with sticks); **j** Alignment of ZINC24841089 on ADHRR.24. (Colour figure online)

(D) from –NH of indoline-2-one. The present 3D QSAR model was able to provide information regarding favorable and unfavorable structural features responsible for c-KIT activity. The model (ADHRR.24) was found to be statistically significant ($R^2 = 0.9378$, $Q^2 = 0.7832$). Other recommended parameters, i.e., R_0^2 , R'^2_0 , k , and k' were found to be within an acceptable range. The model was further validated through external validation methods (*Y*-randomization and ‘goodness-of-hits’ approach). The robustness of the developed model was determined by the applicability domain (AD) calculation using a ‘standardization approach’. A docking study was performed on the indoline-2-ones and a strong agreement was attained between the docking results and the developed 3D QSAR model. Ligand-based and structure-based virtual screenings were performed using compounds from the ZINC database identifying five hits with high potential activity against c-KIT. The ADME properties of the hits were calculated and found to be within an acceptable range. Our molecular docking study indicates that hydrogen-bond interactions with CYS673 and GLU671, and a bidentate hydrogen bond with the CYS673 residue are key for c-KIT affinity.

Acknowledgments Authors are thankful to Prof. Sanjay J. Surana, Principal, R. C. Patel Institute of Pharmaceutical Education and Research, Shirpur, India, for availing the facility to carry out the computational work and his valuable support. Authors are also grateful to the North Maharashtra University, Jalgaon, Maharashtra, for providing financial assistance under the scheme “Vice chancellor Research Motivation Scheme (VCRMS),” Sanction No. *NMU/11A/VCRMS/Budget-2014-15/Pharmacy-15/170/2015*.

Compliance with ethical standards

Conflicts of Interest Corresponding author Prashant J. Chaudhari presented a part of this manuscript at ‘State Level AVISHKAR-2014: The inter-university research convention, Maharashtra’.

References

- Zsebo KM, Williams DA, Geissler EN, Broudy VC, Martin FH, Atkins HL, Hsu RY, Birkett NC, Okino KH, Murdock DC, Jacobsen FW, Langley KE, Smith KA, Takeish T, Cattanach BM, Galli SJ (1990) Stem cell factor is encoded at the Sl locus of the mouse and is the ligand for the c-kit tyrosine kinase receptor. *Cell* 63:213–224. doi:10.1016/0092-8674(90)90302-U
- Isakov N, Biesinger B (2000) Lck protein tyrosine kinase is a key regulator of T-cell activation and a target for signal intervention by Herpesvirus saimiri and other viral gene products. *Eur J Biochem* 267:3413–3421. doi:10.1046/j.1432-1327.2000.01412.x
- Drube S, Schmitz F, Gopfert C, Weber F, Kamradt T (2012) C-Kit controls IL-1 β -induced effector functions in HMC-cells. *Eur J Pharmacol* 675:57–62. doi:10.1016/j.ejphar.2011.11.035
- Robert R Jr (2005) Signaling by Kit protein-tyrosine kinase: the stem cell factor receptor. *Biochem Biophys Res Commun* 337:1–13. doi:10.1016/j.bbrc.2005.08.055
- Kansal N, Silakari O, Ravikumar M (2010) Three dimensional pharmacophore modelling for c-Kit receptor tyrosine kinase inhibitors. *Eur J Med Chem* 45:393–404. doi:10.1016/j.ejmech.2009.09.013
- Wang WL, Healy ME, Sattler M, Verma S, Lin J, Maulik G, Stiles CD, James DG, Johnson BE, Salgia R (2000) Growth inhibition and modulation of kinase pathways of small cell lung cancer cell lines by the novel tyrosine kinase inhibitor STI 571. *Oncogene* 19:3521–3528. doi:10.1038/sj.onc.1203698
- Heinrich MC, Blanke CD, Druker BJ, Corless CL (2002) Inhibition of KIT tyrosine kinase activity: a novel molecular approach to the treatment of KIT-positive malignancies. *J Clin Oncol* 20:1692–1703. doi:10.1200/JCO.20.6.1692
- Imai K, Takaoka A (2006) Comparing antibody and small-molecule therapies for cancer. *Nat Rev Cancer* 6:714–727. doi:10.1038/nrc1913
- Eklund KK (2007) Mast cells in the pathogenesis of rheumatic diseases and as potential targets for anti-rheumatic therapy. *Immunol Rev* 217:38–52. doi:10.1111/j.1600-065X.2007.00504.x
- Jensen BM, Metcalfe DD, Gilfillan AM (2007) Targeting kit activation: a potential therapeutic approach in the treatment of allergic inflammation. *Inflamm Allergy Drug Targets* 6:57–62. doi:10.2174/187152807780077255
- Chen N, Burlin RW, Neira S, Hungate R, Zhang D, Yu V, Nguyen Y, Tudor Y, Plant M, Flynn S, Meagher KL, Lee MR, Zhang X, Itano A, Schrag M, Xu Y, Gordon YN, Hu E (2008) Discovery of a potent and selective c-Kit inhibitor for the treatment of inflammatory diseases. *Bioorg Med Chem Lett* 18:4137–4141. doi:10.1016/j.bmcl.2008.05.089
- Gajiwala KS, Wu JC, Christensen J, Deshmukh GD, Diehl W, DiNitto JP, English JM, Greig YH, Jacques SL, Lunney EA, McTigue M, Molina D, Quenzer T, Wells PA, Yu X, Zhang Y, Zou A, Emmett MR, Marshall AG, Zhang HM, Demetri GD (2009) KIT kinase mutants show unique mechanisms of drug resistance to imatinib and sunitinib in gastrointestinal stromal tumor patients. *Proc Natl Acad Sci USA* 106:1542–1547. doi:10.1073/pnas.0812413106
- Force T, Krause DS, Van Etten RA (2007) Molecular mechanisms of cardiotoxicity of tyrosine kinase inhibition. *Nat Rev Cancer* 7:332–344. doi:10.1038/nrc2106
- Zhang L, Zheng Q, Yang Y, Zhou H, Gong X, Zhao S, Fan C (2014) Synthesis and in vivo SAR study of indolin-2-one-based multi-targeted inhibitors as potential anticancer agents. *Eur J Med Chem* 82:139–151. doi:10.1016/j.ejmech.2014.05.051
- Cho TP, Dong SY, Jun F, Hong FJ, Liang YJ, Lu X, Hua PJ, Li LY, Lei Z, Bing H, Ying Z, Qiong LF, Bei FB, Guang LL, Shen GA, Hong SG, Hong SW, Tai MX (2010) Novel potent orally active multitargeted receptor tyrosine kinase inhibitors: synthesis, structure-activity relationships, and antitumor activities of 2-indolinone derivatives. *J Med Chem* 53:8140–8149. doi:10.1021/jm101036c
- Almerico AM, Tutone M, Lauria A (2012) Receptor-guided 3D-QSAR approach for the discovery of c-kit tyrosine kinase inhibitors. *J Mol Model* 18:2885–2895. doi:10.1007/s00894-011-1304-0
- Pan Y, Wang Y, Bryant SH (2013) Pharmacophore and 3D-QSAR characterization of 6-arylquinazolin-4-amines as Cdc2-like kinase 4 (Cdk4) and dual specificity tyrosine-phosphorylation-regulated-kinase 1A (Dyrk1A) inhibitors. *J Chem Inf Model* 53:938–947. doi:10.1021/ci300625c
- Cruciani G, Watson KA (1994) Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. *J Med Chem* 37:2589–2601. doi:10.1021/jm00042a012
- Ballante F, Ragno R (2012) 3-D QSAR Autogrid/R: an alternative procedure to build 3-D QSAR models. *Methodologies and applications. J Chem Inf Model* 52:1674–1685. doi:10.1021/ci300123x

20. Schuster D, Langer T (2005) The identification of ligand features essential for PXR activation by pharmacophore modeling. *J Chem Inf Model* 45:431–439. doi:[10.1021/ci049722q](https://doi.org/10.1021/ci049722q)
21. Ding L, Tang F, Huang W, Jin Q, Shen H, Wei P (2013) Design, synthesis, and biological evaluation of novel 3-pyrrolo[b]cyclohexylene-2-dihydroindolinone derivatives as potent receptor tyrosine kinase inhibitors. *Bioorg Med Chem Lett* 23:5630–5633. doi:[10.1016/j.bmcl.2013.08.037](https://doi.org/10.1016/j.bmcl.2013.08.037)
22. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA (2006) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des* 20:647–671. doi:[10.1007/s10822-006-9087-6](https://doi.org/10.1007/s10822-006-9087-6)
23. Dixon SL, Smondyrev AM, Rao SN (2006) PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des* 67:370–372. doi:[10.1111/j.1747-0285.2006.00384.x](https://doi.org/10.1111/j.1747-0285.2006.00384.x)
24. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47:1739–1749. doi:[10.1021/jm0306430](https://doi.org/10.1021/jm0306430)
25. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. Enrichment factors in database screening. *J Med Chem* 47:1750–1759. doi:[10.1021/jm030644s](https://doi.org/10.1021/jm030644s)
26. Schrödinger Suite 2009 Virtual screening workflow; Glide version 5.5; LigPrep 2.3; QikProp 3.2, Schrödinger, LLC, New York. <http://www.schrodinger.com>
27. Roy K, Kar S, Ambure P (2015) On a simple approach for determining applicability domain of QSAR models. *Chemom Intell Lab Syst* 145:22–29. doi:[10.1016/j.chemolab.2015.04.013](https://doi.org/10.1016/j.chemolab.2015.04.013)
28. Golbraikh A, Shen M, Xiao Z, Xiao Y, Lee K, Tropsha A (2003) Rational selection of training and test sets for the development of validated QSAR models. *J Comput Aided Mol Des* 17:241–253. doi:[10.1023/A:3A1025386326946](https://doi.org/10.1023/A:3A1025386326946)
29. Teli MK, Rajanikant GK (2012) Pharmacophore generation and atom based 3D-QSAR of *N*-iso-propyl pyrrole-based derivatives as HMG-CoA reductase inhibitors. *Org Med Chem Lett* 2:1–10. doi:[10.1186/2191-2858-2-25](https://doi.org/10.1186/2191-2858-2-25)
30. Shah UA, Deokar HS, Kadam SS, Kulkarni VM (2010) Pharmacophore generation and atom-based 3D-QSAR of novel 2-(4-methylsulfonylphenyl)pyrimidines as COX-2 inhibitors. *Mol Divers* 14:559–568. doi:[10.1007/s11030-009-9183-3](https://doi.org/10.1007/s11030-009-9183-3)
31. Golbraikh A, Tropsha A (2002) Beware of q^2 !. *J Mol Graph Mod* 20:269–276. doi:[10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1)
32. Zhang S, Golbraikh A, Oloff S, Kohn H, Tropsha A (2006) A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. *J Chem Inf Model* 46:1984–1995. doi:[10.1021/ci060132x](https://doi.org/10.1021/ci060132x)
33. Melagraki G, Afantitis A (2013) Enalos KNIME nodes: exploring corrosion inhibition of steel in acidic medium. *Chemom Intell Lab Syst* 123:9–14. doi:[10.1016/j.chemolab.2013.02.003](https://doi.org/10.1016/j.chemolab.2013.02.003)
34. Guner O, Henry D (1998) Formula for determining the “goodness of hit lists” in 3D database searches. Accelrys/MDL Information Systems, Inc., San Diego/San Leandro. <http://www.netsci.org/Science/Cheminform/feature09.html>. Accessed 12 Aug 2015
35. Chen X, Liu M, Gilson MK (2002) BindingDB: a web-accessible molecular recognition database. *Comb Chem High Throughput Screen* 4:719–725. doi:[10.2174/1386207013330670](https://doi.org/10.2174/1386207013330670)
36. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* 35(Database issue):198–201. doi:[10.1093/nar/gkl999](https://doi.org/10.1093/nar/gkl999)
37. Chen X, Lin Y, Gilson MK (2001) The binding database: overview and user’s guide. *Biopolymers* 61:127–141
38. Compounds: release 4 file series: May 2012. <http://cactus.nci.nih.gov/download/nci>. Accessed 8 Aug 2015
39. Melagraki G, Afantitis A (2014) Enalos InSilicoNano platform: an online decision support tool for the design and virtual screening of nanoparticles. *RSC Adv* 4:50713–50725. doi:[10.1039/C4RA07756C](https://doi.org/10.1039/C4RA07756C)
40. The simple, user-friendly and reliable online application for the AD computation. <http://dtclab.webs.com/software-tools> or http://teqip.jdvu.ac.in/QSAR_Tools. Accessed 16 Aug 2015
41. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG (2012) ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model* 52:1757–1768. doi:[10.1021/ci3001277](https://doi.org/10.1021/ci3001277)
42. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46:3–26. doi:[10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0)
43. Guner OF, Waldman M, Hoffmann RD, Kim JH (2000) Pharmacophore perception, development, and use in drug design, IUL biotechnology series. In: Guner OF (ed) *Strategies for database mining and pharmacophore development*, 1st edn. International University Line, La Jolla, pp 213–236
44. Park H, Lee S, Lee S, Hong S (2014) Structure-based de novo design and identification of D816V mutant-selective c-KIT inhibitors. *Org Biomol Chem* 26:4644–4655. doi:[10.1039/c4ob00053f](https://doi.org/10.1039/c4ob00053f)
45. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Zhou Z, Han L, Karapetyan K, Dracheva S, Shoemaker BA, Bolton E, Gindulyte A, Bryant SH (2012) PubChem’s BioAssay database. *Nucleic Acids Res* 40:400–412. doi:[10.1093/nar/gkr1132](https://doi.org/10.1093/nar/gkr1132)