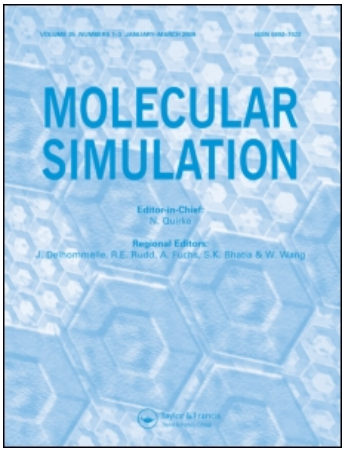


Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Pharmacophore-based 3D-QSAR as a predictive method for the QSAR analysis on a series of potent and selective inhibitors for three kinases of RTK family

Qinglin Jiang[a]; Hongli Liao[a]; Qian Yang[a]; Wang Zan[a]; Zhihe Zang[a]
[a] Department of Pharmacy, Chengdu Medical College, Chengdu, 610000, People's Republic of China

Online publication date: 16 August 2010

To cite this Article Jiang, Qinglin , Liao, Hongli , Yang, Qian , Zan, Wang and Zang, Zhihe(2010) 'Pharmacophore-based 3D-QSAR as a predictive method for the QSAR analysis on a series of potent and selective inhibitors for three kinases of RTK family', Molecular Simulation, 36: 9, 693 — 707
To link to this Article: DOI: 10.1080/08927021003752788
URL: http://dx.doi.org/10.1080/08927021003752788

Molecular Simulation
Vol. 36, No. 9, August 2010, 693–707

Taylor & Francis
Taylor & Francis Group

Pharmacophore-based 3D-QSAR as a predictive method for the QSAR analysis on a series of potent and selective inhibitors for three kinases of RTK family

Qinglin Jiang, Hongli Liao, Qian Yang, Wang Zan and Zhihe Zang*

Department of Pharmacy, Chengdu Medical College, Chengdu, 610000, People's Republic of China

(*Received 14 July 2009; final version received 5 March 2010*)

For targets belonging to the same family of receptors, inhibitors often act at more than one biological target and produce a synergistic effect. Separate pharmacophore-based comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) models were developed from our data-set for the kinase insert domain-containing receptor (KDR), cKit and Flt1 inhibitors. These models showed excellent internal predictability and consistency; validation using test-set compounds yielded a good predictive power for the pIC_{50}-value. The field contour maps (CoMFA and CoMSIA) corresponding to the KDR, cKit and Flt1 kinase subtypes reflected the characteristic similarities and differences between these types. These maps provided valuable information to facilitate structural modifications of the inhibitor to increase selectivity of the KDR over cKit and Flt1.

Keywords: KDR; cKit; Flt1; 3D-QSAR; pharmacophore

1. Introduction

Receptor tyrosine kinases (RTKs) are believed to play critical roles in signal transduction in a number of cellular functions and have been implicated in a variety of pathological conditions including angiogenesis cancer, tumour growth, atherosclerosis, diabetic, retinopathy and inflammatory disease [1–5]. Vascular endothelial growth factor receptor (VEGFR) tyrosine kinase appears to be one of the most important mediators of physiological and pathological angiogenesis and is specifically expressed in vascular endothelial cells including VEGFR1 (Flt1) and kinase insert domain-containing receptor (KDR) (VEGFR2) [6–8]. The activation of the VEGFR family RTKs and the KDR tyrosine kinase by vascular endothelial growth factors (VEGFs) is a potent regulator of vascular endothelial cells and has been directly linked to tumour angiogenesis and blood vessel-dependent metastasis [9–11]. KDR has also been shown to promote the production of immature dendritic cell colonies via its kinase activity [12]. The cKit, a stem cell factor (SCF) receptor belonging to the RTK family, initiates cell growth and proliferation signal transduction cascades in response to SCF binding [13].

VEGFRs play an early role in the differentiation of mesodermal cells into endothelial cells and the proliferation and migration of endothelial cells to form primitive tubular vessels [14–16]. The kinase activity of cKit is tightly regulated throughout its signalling cycle [13]. Often, the biological targets belonging to the same family of receptors raise the problem of selectivity in receptor–ligand binding. Special efforts to design multiple activating drugs can be successfully made using computational methods to support biochemical studies and to design selective ligands for receptor subtypes.

Structure–activity relationship (SAR) analysis is the basis for the perception of the structural features of both the inhibitors and the target receptors concerned in a particular biological process and thus helps in designing more effective inhibitors. SAR analysis can be done using computer-aided drug design techniques which involves either of the two most commonly used techniques: the indirect/ligand-based drug design or the direct/structure-based drug design also known as structure-based design. The latter has assumed increased importance due to rapid advances in the fields of structural and molecular biology. However, it appears relatively difficult to find a reliable predictive model based on the calculated energies obtained by docking [17,18]. To overcome this problem, highly predictive QSAR and comparative molecular field analysis (CoMFA) models are being developed using the technique of structure-based alignments. The 3D-QSAR model development based on cocrystallised conformer, as well as docked conformer-based alignment is not feasible. Since the only possibility left in such a case is to use the global minimised conformer-based alignment (GMCBA) and further, in view of some earlier findings where pharmacophore-based alignment (PBA) has been shown to be suitable for 3D-QSAR model development, it appeared of interest to compare the two alternatives, namely GMCBA and PBA for the 3D-QSAR model development using CoMFA and comparative molecular similarity index analysis (CoMSIA).

*Corresponding author. Email: zangzh2009@163.com

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927021003752788
http://www.informaworld.com

In order to understand the influence of different physicochemical and structural parameters on inhibitor activity at the molecular level, the following strategies were followed. First, common feature pharmacophore models using the Hip-Hop module of CATALYST were developed using different KDR, Flt1 and cKit inhibitors. It was followed by the selection of the best pharmacophore on the basis of the highest rank score, which was then used for the conformational alignment of all molecules or 3D-QSAR model development. To our knowledge, this is the first case of utilising the pharmacophore-based 3D-QSAR method to study the kinase inhibitory selectivity of a series of small molecular compounds. By the differences of the structural requirements for ligand binding to three kinase receptor subtypes, these models would offer utility in guiding the rational design of potent and selective KDR, cKit and cKit inhibitors for therapeutic applications.

2. Materials and methods

2.1 Data-sets

The present study was based on two fundamental assumptions. First, all of the inhibitors bind in the same fashion with three different kinases of the RTK family. The conformations of the compounds bound to the different kinases might be different. In our opinion, however, the assumption here is available when we take the lowest energy conformation as the active conformation. Second, data from different sources are available in our data-set. Discarding inactive compounds and compounds with unspecified inhibitory activity, a data-set composed of 59 KDR inhibitors was taken from two published papers. The activity values, IC_{50}, of these KDR inhibitors were reported previously [19,20] and the pIC_{50} ($-\log IC_{50}$) activity values are calculated and shown in Table 1. The total set is randomly divided into two subsets: a training set of 47 compounds for generating 3D-QSAR models and a test set of 12 compounds for validating the quality of the models. Two additional data-sets are also available: 57 compounds with inhibitory activities of Flt1 and 55 compounds with inhibitory activities of cKit are also reported as IC_{50} values [20]. We randomly divided these data-sets into two subsets. The total set of Flt1 inhibitors (57 compounds) was divided into a training set of 35 inhibitors and a test set of 12 inhibitors, while all the cKit inhibitors (55 compounds) were divided into a training set of 43 compounds and the remaining compounds were assigned to the test set.

2.2 Modelling tools

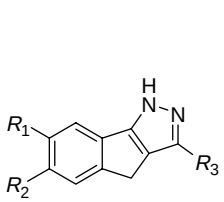
The pharmacophore modelling was accomplished using the Hip-Hop module implemented in CATALYST (release version 4.11; Accelrys, San Diego, CA, USA). The 3D-QSAR studies (CoMFA [21,22] and CoMSIA [23–25]) were done on a Silicon Graphics Octane workstation using SYBYL 6.9 software (Tripos, Inc., St Louis, MO, USA).

2.3 Pharmacophore modelling (Hip-Hop)

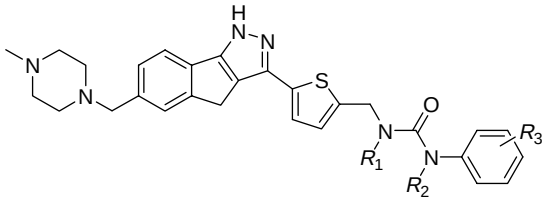
Unlike the CATALYST/HypoGen algorithm, the Hip-Hop algorithm identifies features which are common among a set of compounds under study without considering their biological activities. Considering a collection of generated conformational models of molecules and chemical feature selection as input, CATALYST produces a series of possible molecular alignments and then identifies 3D configurations of features that are common to this set of molecules. The hypotheses are ranked based on (i) the number of molecules fitting the pharmacophore and (ii) the frequency of its occurrence (rank score). During common feature hypothesis generation, all of the possible chemical feature-based hypotheses possessing 10 or fewer features are identified and ranked according to an estimate of their relative selectivity. The most selective hypothesis is given the highest numerical rank. The quality of the mapping between a compound and a hypothesis is indicated by the fit value. There are two control parameters employed in the Hip-Hop algorithm: *Principal* and *Maximum Omit Feat (MaxOmitFeat)*. The values of these control parameter, namely *Principal* and *MaxOmitFeat* as input, set the Hip-Hop algorithm to determine which molecule should be considered while building hypothesis space and which molecule should map to all or some of the features in the final hypothesis, respectively.

All compounds were built using ISIS Draw 2.5, imported to Accelrys Discovery Studio window, and optimised using CHARMm force field. The CATALYST module treats molecular structures as templates consisting of strategically positioned chemical functions that will bind effectively with receptor. The biologically most important binding functions are deduced from a small set of compounds that cover a broad range of activity. CATALYST generates conformational models for each compound using the Poling algorithm. Diverse conformational models for each compound were generated such that the conformers covered accessible conformational space defined within 20 kcal/mol of the estimated global minimum. Among the two types of conformational analyses, fast and best quality provided in the CATALYST, the best option was used specifying 255 as the maximum number of conformers and 20 kcal/mol as an energy cut-off. The estimation of the conformational energy was based on the CHARMm force field. The generated diverse conformations for molecules were submitted to CATALYST for hypothesis generation using the Hip-Hop algorithm.

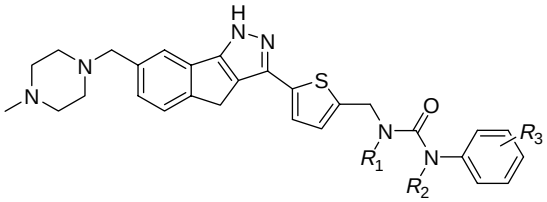
Table 1. Molecular structure and inhibitory activities (pIC_{50}) for KDR, Flt1 and cKit kinases.



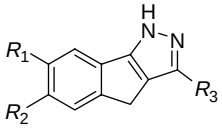
No. 1-3



No. 4-27



No. 28-39



No. 40-59

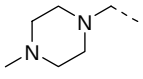
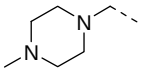
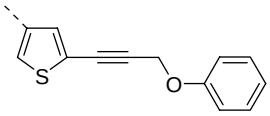
No.	R_1	R_2	R_3	KDR	Flt1	cKit
1		H		6.75	6.01	6.84
2*		H		6.83	6.22	6.99
3[a,b]	H			8.22	8.70	7.59
4	H	H	2-CH_3	6.35	4.60	5.86
5	H	H	3-CH_3	7.21	5.54	7.07
6	H	H	4-CH_3	6.51	4.94	7.62
7	H	H	2-CF_3	5.54	–	4.62
8[a,c]	H	H	3-CF_3	7.47	6.06	7.77
9[c]	H	H	4-CF_3	6.55	5.38	7.89
10	H	H	2-Cl	6.70	–	6.04
11*	H	H	3,5-di-CH_3	7.20	5.78	7.57
12	H	H	3,5-di-Cl	7.10	5.56	7.42
13	CH_3	H	3-CH_3	6.80	5.33	4.83
14	H	CH_3	3-CH_3	5.32	4.84	6.86
15[b,c]	CH_3	CH_3	3-CH_3	5.13	4.55	4.57
16	CH_3	H	2-CH_3	6.91	5.31	5.24
17	CH_3	H	2-Cl	6.74	5.42	–
18[a,c]	CH_3	H	3-F,5-CH_3	7.70	7.48	7.60
19[b]	CH_3	H	3-OCF_3	7.57	7.72	7.25
20	Et	H	3-OCF_3	7.35	6.88	5.66
21[*,c]	n-Pr	H	3-OCF_3	7.46	7.23	7.12
22		H	3-OCF_3	6.91	5.55	4.01
23[a,b]		H	3-CH_3	7.68	7.42	6.87

Table 1 – *continued*

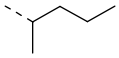
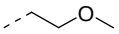
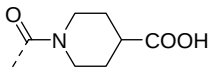
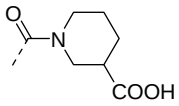
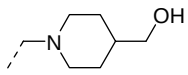
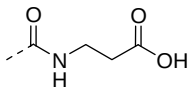
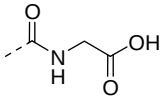
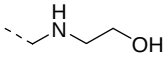
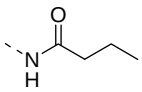
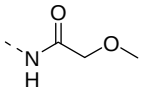
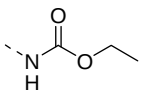
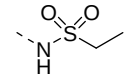
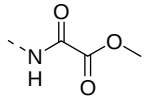
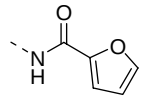
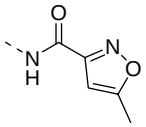
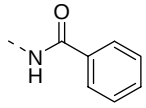
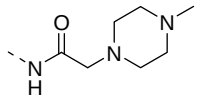
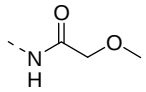
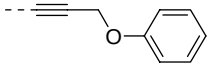
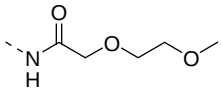
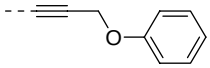
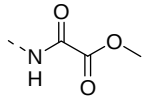
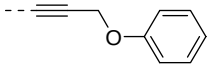
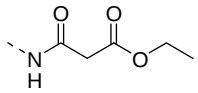
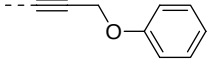
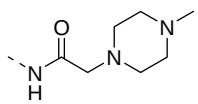
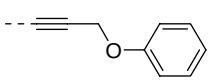
No.	R_1	R_2	R_3	KDR	Flt1	cKit
24		H	3-CH_3	6.79	5.80	4.76
25[a]		H	3-CH_3	6.58	5.44	5.31
26*	—OCH_3	H	3-CH_3	6.59	5.87	6.63
27		H	3-CH_3	6.77	5.65	–
28	H	H	H	6.74	5.42	7.00
29	H	H	2-CH_3	6.80	5.69	6.17
30*	H	H	3-CH_3	7.32	6.23	7.16
31[c]	H	H	4-CH_3	6.98	6.21	7.57
32	CH_3	H	3-CH_3	7.29	7.36	7.02
33[a,b,c]	CH_3	H	3-CF_3	7.96	7.85	7.74
34[a,b]	CH_3	H	3-Cl	7.92	7.70	7.42
35*	CH_3	H	4-F	7.80	7.08	7.23
36[c]	CH_3	H	4-OCF_3	7.77	7.60	7.72
37	CH_3	H	3,4-Cl	7.74	7.44	7.55
38[c]	CH_3	H	4-Br	7.72	7.42	7.77
39[*,a]	CH_3	H	4-Cl	7.68	7.19	7.44
40		H	–	5.15	5.46	–
41		H	–	5.32	5.51	–
42[*,a,b]		H	–	4.55	4.76	5.14
43		H	–	6.24	6.43	5.25
44		H	–	6.39	6.62	5.29
45[b]		H	–	6.36	6.57	5.78
46[c]		H	–	6.25	6.54	5.02
47[b]		H	–	7.15	7.52	5.80
48*		H	–	6.19	6.35	5.59

Table 1 – *continued*

No.	R_1	R_2	R_3	KDR	Flt1	cKit
49		H	–	6.51	6.54	5.36
50[a]		H	–	7.30	7.10	5.95
51		H	–	6.96	7.05	5.96
52		H	–	7.22	7.15	6.05
53[*,c]		H	–	7.00	7.30	6.74
54[a]		H	–	6.04	6.15	5.36
55[b]			–	7.28	7.80	6.76
56			–	6.57	6.87	6.00
57[*,b]			–	7.44	7.70	6.70
58			–	6.58	6.82	6.21
59[*,a,b,c]			–	8.00	8.22	7.31

Note: Compounds marked with * belong to the test sets, marked with a, b, c indicates compounds were used for common feature pharmacophore generation of KDR, Flt1, c-Kit inhibitors, respectively. The unit of IC_{50} is nM.

2.4 Conformational sampling and alignment

Structural alignment is perhaps the most subjective, yet critical, step in the CoMFA study. From experience, it can be shown that the resulting 3D-QSAR models are often sensitive to a particular alignment scheme. However, in contrast to CoMFA, CoMSIA is relatively less sensitive to changes in the orientation of the superimposed molecules in the lattice. The reliability and efficiency of CoMFA or CoMSIA results depend on the correct alignment of the input ligands. The pharmacophore mapping study is considered to be useful in the selection of both a proposed bioactive conformer and a superposition rule. In the CATALYST, the studied molecules are automatically overlaid over the best pharmacophore model, and for each ligand, one aligned conformer based on the lowest rms deviation of feature atom coordinates from those of the corresponding reference features is superimposed on the best hypothesis.

The relative alignment of the molecules is the most critical step in CoMFA and CoMSIA analyses which affects the final outcome; also, the activity of the molecules is a result of the 3D disposition of features present in them or, in other words, their ability to assume a conformation for favourable interactions with the receptor. Therefore, the best hypothesis generated by the Hip-Hop algorithm was used for the alignment of all 52 compounds. The resulting alignment was then exported to SYBYL6.9 for CoMFA and CoMSIA studies. The partial charges for all the compounds were calculated using the Gasteiger–Huckel method.

2.5 CoMFA and CoMSIA analyses

Unlike the traditional 2D-QSAR methods, which rely on the substituent parameters, the CoMFA exploits the active conformer and superposition rule for a set of molecules to provide quantitative correlation of the particular activity with the parameters in terms of steric and electrostatic fields. The CoMFA procedure can be summarised into three steps. (i) First, all molecules under study are aligned using molecular alignment methods, namely common substructure based, multi-atom/point based (multi-fit), field based (field-fit) and pharmacophore based. (ii) Then, a 3D cubic lattice with a grid spacing of 2.0 Å in x, y and z directions is generated to enclose the molecule aggregate. (iii) A probe atom, e.g. sp^3 carbon with $+1$ or -1 charge, is placed at every lattice point to measure the electrostatic (Columbic with $1/r$ dielectric) or steric (6–12 Lennard-Jones potential) field by using molecular mechanics. To avoid too high and unrealistic energy values inside the molecule, a 30 kcal/mol as an energy cut-off was applied. Finally, the results from the field samplings combined with the biological activities of the molecules are put into a spreadsheet, and the partial least-squares (PLS) analysis is applied to get the final CoMFA model. Another molecular modelling technique, CoMSIA, is devised to overcome the problems associated with CoMFA of very rapidly changing steric fields near the atomic nuclei and scaling the two fields (steric and electrostatic) for PLS analyses. CoMSIA avoids the need of defining strict cut-off limits which can result in exclusion of several important data points and more interpretable contour maps. The CoMSIA is immune to small changes in the relative orientation of the aligned molecules with respect to the lattice, which is another well-known problem in CoMFA. Moreover, in addition to the steric and electrostatic fields, the CoMSIA includes hydrogen bonds and hydrophobic interactions which are not sufficiently described by the steric and electrostatic fields, and also includes an entropy component. In CoMSIA analyses, the standard settings (probe with charge +1, radius 1 Å, hydrophobicity +1, hydrogen-bond donating +1, hydrogen-bond acceptor +1, attenuation factor $R = 0.3$ and grid spacing 2.0 Å) were used to calculate five fields, namely steric, electrostatic, hydrophobic, acceptor and donor.

Generally, a leave-one-out (LOO) cross-validated r^2 (q^2) is used as a quantitative measure for CoMFA and CoMSIA models. The graphical representations (contours) of CoMFA and CoMSIA results are indicated by the regions, where the variations in steric and electrostatic properties (in CoMFA) along with hydrophobic and hydrogen bonds (in CoMSIA) of different molecules in a data-set are correlated with the variations in the biological activity.

2.6 Statistical analyses/PLS

PLS was used to correlate the KDR, Flt1 and cKit inhibitory activities with the CoMFA and CoMSIA values containing a magnitude of steric, electrostatic and other potentials. The generated models were assessed by using a LOO cross-validation procedure by the SAMPLS method as implied in SYBYL. A strict criterion for selection of an optimal number of components was applied by selecting the lowest predicted residual sum of squares (PRESS) value, and also wherever the last added component increased q^2 less than 5%, the less complex model was chosen. In addition to LOO cross-validation, the cross-validation in groups using 10 groups, repeating the procedure 10 times, was also carried out. The mean of 10 readings is given as r^2_{cv} (mean). The PLS algorithm was used to generate the final CoMFA and CoMSIA models with an optimal number of components obtained using LOO cross-validation. The minimum standard deviation threshold was set at 2.0 kcal/mol to speed up the analyses and to reduce the noise. The r^2_{cv} (q^2), PRESS, r^2 and standard error of estimate (SEE) values were computed as defined in SYBYL6.9.

3. Results and discussion

3.1 Common feature pharmacophore modelling

For KDR, Flt1 and cKit inhibitory activities, 12 compounds were used to generate 10 optimal hypotheses (Hypo) using the Hip-Hop module implemented in the CATALYST. The rank scores of these 30 optimal hypotheses were in the broad range of 262–127 (Tables 2–4), indicative of a greater probability of finding true correlation. Among these hypotheses shown in the descending order of rank score, the three Hypo1 were found to be the best pharmacophores with the highest rank score of 262.27, 142.59 and 226.47, respectively.

It mapped all the important features of the most active KDR inhibitor and exhibited the highest fit value of 5.68 (Figure S1 of the Supporting Information). The summary of the common feature pharmacophore hypothesis run of KDR inhibitors is described in Table 2. Since the top five hypotheses had the same five features, namely one hydrophobic (H), one hydrogen bond acceptor (HBA), two ring aromatic (RA 1–2) and one hydrogen bond donor (HBD) features, the Hypo1 with the highest rank score was chosen for pharmacophore-based conformational alignment of 59 molecules. Figure 1(a) also depicts the distance and angular constraints between the features in the best pharmacophore (Hypo1). The conformation of the most active compound **3** superimposed on the best hypothesis (Hypo1) is shown in Figure 1(b). All 59 compounds were mapped onto Hypo1 to obtain the pharmacophore-based molecular alignment (Figure S1 of the Supporting Information), which was further used for the CoMFA and CoMSIA studies.

It also mapped all the important features of the most active Flt1 inhibitor and exhibited the highest fit value of 4.82 (Figure 2). The summary of the common feature pharmacophore hypothesis run of Flt1 inhibitors is described in Table 3. Almost all hypotheses had the same five features, namely one hydrophobic (H), one HBA and three ring aromatic (RA 1–3) features; the Hypo1 with the highest rank score was chosen for pharmacophore-based conformational alignment of 57 molecules. Figure 2(a), (b) displays the distance and angular constraints between the features in the best pharmacophore (Hypo1) and conformation of the most active compound 3 superimposed on the best hypothesis (Hypo1), respectively. All 57 Flt1 inhibitors were mapped onto Hypo1 to obtain the pharmacophore-based molecular alignment (Figure S2 of the Supporting Information), which was further used for the

Table 2. Summary of common feature hypothesis of KDR inhibitors run using Hip-Hop.

Hypo	Feature	Rank	Direct hit mask	Partial hit mask
1	RRHDA	262.2704	DH: 111111111111	PH: 000000000000
2	RRHDA	257.1355	DH: 111111111111	PH: 000000000000
3	RRHDA	253.3783	DH: 111111111111	PH: 000000000000
4	RRHDA	249.2567	DH: 111111111111	PH: 000000000000
5	RRHDA	246.0196	DH: 111111111111	PH: 000000000000
6	RRRHA	234.0431	DH: 111111111111	PH: 000000000000
7	RRRHA	225.5414	DH: 111111111111	PH: 000000000000
8	RRRHA	217.026	DH: 111111111111	PH: 000000000000
9	RRRHA	208.975	DH: 111111111111	PH: 000000000000
10	RRRHA	192.9	DH: 111111111111	PH: 000000000000

Notes: R, ring aromatic; A, hydrogen bond acceptor; H, hydrophobic; D, hydrogen bond donor; direct hit = all the features are mapped; direct hit = 1 means yes; partial hit = partial mapping of the hypothesis; partial hit = 0 means no.

Table 3. Summary of common feature hypothesis of Flt1 inhibitors run using Hip-Hop.

Hypo	Feature	Rank	Direct hit mask	Partial hit mask
1	RRRHA	142.59	DH: 1111111	PH: 0000000
2	RRRHA	142.5075	DH: 1111111	PH: 0000000
3	RRRHA	140.6074	DH: 1111111	PH: 0000000
4	RRRHA	137.7573	DH: 1111111	PH: 0000000
5	RRRHA	137.007	DH: 1111111	PH: 0000000
6	RRHAA	132.7575	DH: 1110111	PH: 0001000
7	RRRHA	131.7945	DH: 1110111	PH: 0001000
8	RRHAA	131.7615	DH: 1110111	PH: 0001000
9	RRHAA	131.6505	DH: 1110111	PH: 0001000
10	RRHAA	127.2622	DH: 1110111	PH: 0001000

Notes: R, ring aromatic; A, hydrogen bond acceptor; H, hydrophobic; D, hydrogen bond donor; direct hit = all the features are mapped; direct hit = 1 means yes; partial hit = partial mapping of the hypothesis; partial hit = 0 means no.

Table 4. Summary of common feature hypothesis of cKit inhibitors run using Hip-Hop.

Hypo	Feature	Rank	Direct hit mask	Partial hit mask
1	RRHDA	226.47	DH: 111111111111	PH: 000000000000
2	RRRHA	217.90	DH: 111111111111	PH: 000000000000
3	RRHDA	209.68	DH: 111111111111	PH: 000000000000
4	RRHDA	206.45	DH: 111111111111	PH: 000000000000
5	RRRHA	203.07	DH: 111111111111	PH: 000000000000
6	RRRHA	199.84	DH: 111111111111	PH: 000000000000
7	RRHDA	198.18	DH: 101111111111	PH: 010000000000
8	RRRHA	196.44	DH: 111111111111	PH: 000000000000
9	RRHDA	194.72	DH: 111111111111	PH: 000000000000
10	RRRHA	193.06	DH: 101111111111	PH: 010000000000

Notes: R, ring aromatic; A, hydrogen bond acceptor; H, hydrophobic; D, hydrogen bond donor; direct hit = all the features are mapped; direct hit = 1 means yes; partial hit = partial mapping of the hypothesis; partial hit = 0 means no.

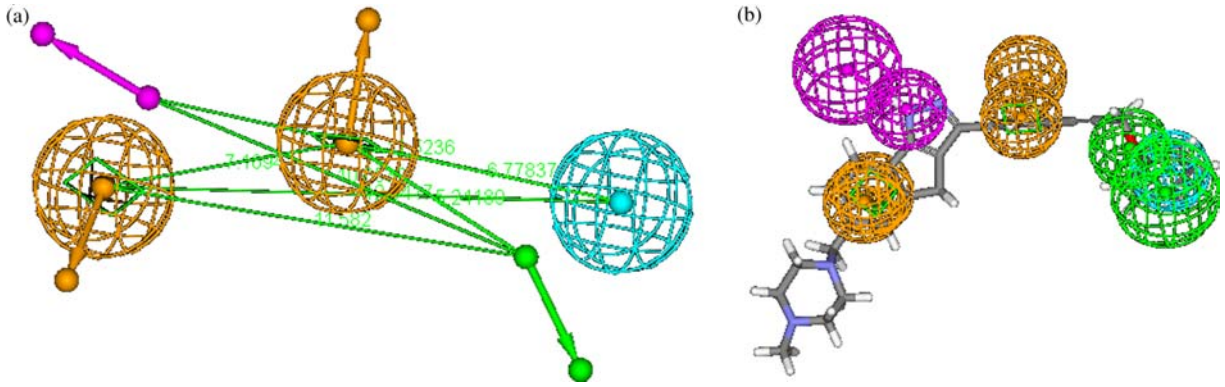


Figure 1. (a) Common feature-based pharmacophore model (Hypo1) of KDR inhibitors with distance and angular constraints and (b) mapping of the most active molecule 1 to Hypo1 with fit value of 5.68. [Hydrophobic (cyan), HBA (green), ring aromatic (orange) and HBD (purple), the distance between chemical features are shown in Å].

CoMFA and CoMSIA studies. Table 4 shows the summary of the common feature pharmacophore hypothesis run of cKit inhibitors with the best fit value of 4.35. The best hypothesis also had the five features, namely one hydrophobic (H), one HBA, two ring aromatic (RA 1–2) and one HBD feature, which was chosen for pharmacophore-based conformational alignment of 55 inhibitors. The distance and angular constraints between the features in the best pharmacophore (Hypo1) and conformation of the most active compound **10** superimposed on the best hypothesis (Hypo1) are depicted in Figure 3. Figure S3 of the Supporting Information shows that the 55 cKit

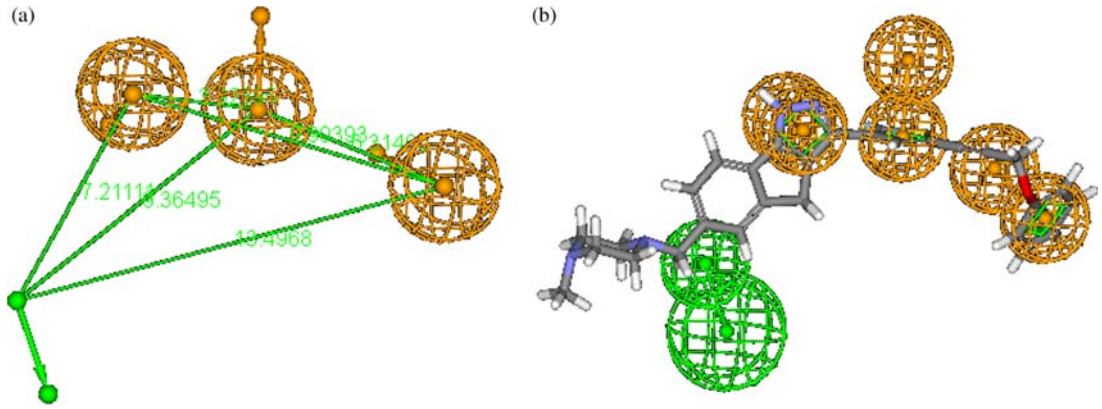


Figure 2. (a) Common feature-based pharmacophore model (Hypo1) of Flt1 inhibitors with distance and angular constraints and (b) mapping of the most active molecule 1 to Hypo1 with fit value of 4.82.

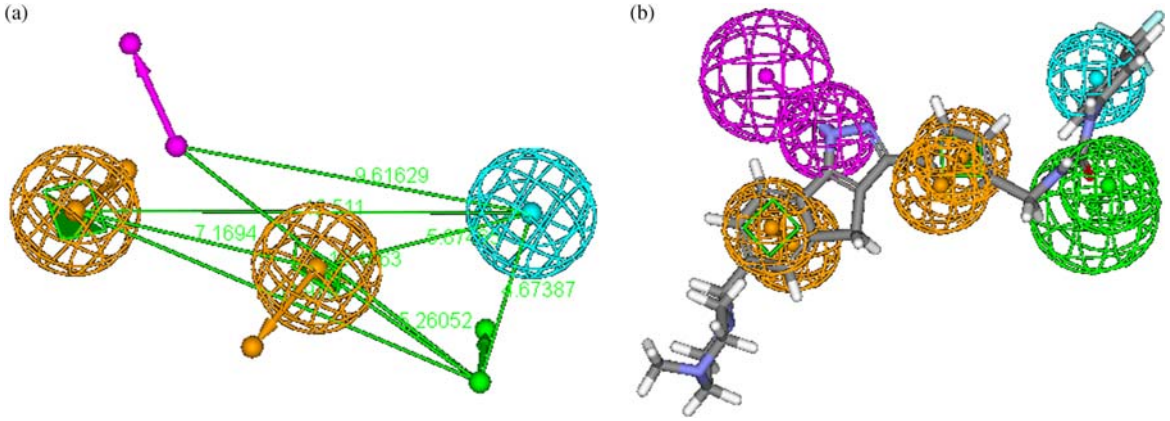


Figure 3. (a) Common feature-based pharmacophore model (Hypo1) of cKit inhibitors with distance and angular constraints and (b) mapping of the most active molecule 1 to Hypo1 with fit value of 4.35.

inhibitors are mapped onto Hypo1 to obtain the pharmacophore-based molecular alignment which was further used for the CoMFA and CoMSIA studies.

3.2 *Maximum common substructure-based alignment*

This alignment approach is implemented in SYBYL6.9 along with two other alignment methods, namely multi- and field-fit methods. The first step of the maximum common substructure (MCS)-based alignment is to explore the substructure which is common to all molecules in the data-set. This is based on the assumption that the common substructure contributes equally to the biological activity of the molecules and can be assumed together as a constant [26]. The next step of the MCS-based alignment is to superimpose the rest of the parts of each molecule in such a way that the overall conformation of each molecule lies near to its local minima, and the substituents of each molecule at a particular site used for major modulations should be as closer as possible with respect to the rest of the molecules. In this paper, the MCS-based alignment was based upon a substructure which then generates molecular alignment as shown in Figure S4 of the Supporting Information.

3.3 *CoMFA and CoMSIA models of KDR inhibitors*

The final results of the CoMFA analyses with a 2.0 Å grid spacing, using the above two alignments, are shown in Table 5; the PLS analyses yielded consistent CoMFA results with the PBA compared to MCS-based alignment. The optimal components that produced the best cross-validation linear regression coefficient were used to get the non-cross-validated model. In 3D-QSAR CoMFA and CoMSIA studies, a q^2 value of 0.4 is considered statistically significant. However, a q^2 value of 0.5 is generally considered better. In view of it, the CoMFA models having $q^2 > 0.6$ were considered highly statistically significant. The LOO cross-validated PLS analysis resulted in a q^2 value of 0.663 using five principal components. The non-cross-validated PLS analysis using the same five principal components yielded an expected higher r^2 value of 0.967 (SEE = 0.141). Although the MCS-based alignment also provided a statistically significant LOO cross-validated q^2 value of 0.584, it was found to be less viable in the prediction with the F value of 136.72, obviously lower than the PBA ones. The observed and estimated KDR kinase inhibitory activity (pIC_{50}) and the residual values for the training set and the test set, using the developed CoMFA model, are described in the Supporting Information, while the graphical plot between the observed vs. estimated (predicted) KDR inhibitory activity for both training and test sets is shown in Figure 4(a).

Various CoMSIA models were generated considering all possible combinations of CoMSIA field descriptors, namely steric (S), electrostatic (E), hydrophobic (H), donor (D) and acceptor (A). In our recent study, steric, electrostatic, hydrophobic and donor field descriptors were found to play an important role in the modulation of KDR inhibitory activity, and the variation of KDR inhibitory activities among the studied inhibitors was well explained. The PLS analyses yielded consistent results of high statistical significance (Table 5) with the alignment generated using the Hip-Hop algorithm, and the results obtained using the other alignment (MCS based) were less significant which further strengthens the robustness of the Hip-Hop-generated conformational alignment. This may be due to the fact that the Hip-Hop algorithm considers a large number of conformations of the active class of compounds, while in MCS, only the conformation near to its local minima of the most active molecule is considered. The best CoMSIA models based on PBA explained well the variation in KDR inhibitory activities for both the training and the test set compounds. These three models

Table 5. Summary of PLS statistic of the best CoMFA and CoMSIA models based on PBA and MCS-based alignment of the three kinases family inhibitors.

	KDR inhibitory activity				Flt1 inhibitory activity				cKit inhibitory activity			
	MCS-based alignment		Pharmacophore-based alignment		MCS-based alignment		Pharmacophore-based alignment		MCS-based alignment		Pharmacophore-based alignment	
	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
Parameters												
q^2	0.584	0.477	0.663	0.723	0.441	0.448	0.736	0.691	0.378	0.258	0.603	0.582
r^2	0.950	0.951	0.967	0.978	0.943	0.935	0.973	0.965	0.951	0.908	0.946	0.954
SEE	0.164	0.182	0.141	0.105	0.204	0.236	0.107	0.126	0.203	0.233	0.130	0.141
F	136.72	86.77	249.31	281.43	118.87	102.71	296.09	270.83	116.69	79.24	196.55	175.39
N	5	6	5	6	5	6	6	5	4	4	3	3
Fractions												
S	0.486	0.194	0.491	0.165	0.419	0.175	0.507	0.162	0.488	0.243	0.465	0.277
E	0.513	0.322	0.509	0.327	0.581	0.284	0.493	0.321	0.512		0.535	
H		0.250		0.264		0.292		0.246		0.420		0.385
D		0.233		0.243		0.249		0.271		0.337		0.338

Notes: q^2, cross-validation correlation; PRESS, predicted residual sum of squares; r^2, regression coefficient; SEE, non-cross-validated standard error of estimate; F, Fisher's F-value; N, optimal number of components; S, steric; E, electrostatic; H, hydrophobic field.

were SEHD developed using S, E, H and D field descriptors, respectively. The PLS statistics of these three best CoMSIA models is summarised in Table 5. The 'SEHD' model gave the LOO cross-validated q^2 value of 0.723 with six principal components and a non-cross-validated r^2 value of 0.978 (SEE = 0.105). The KDR inhibitory activity (pIC_{50}) and the residual values for the training set and the test set compounds using the best CoMSIA (steric, hydrophobic and H-bond acceptor fields in CoMSIA) model are given in the Supporting Information. The graphical plot of the observed vs. calculated KDR inhibitory activity for both the training and test sets is shown in Figure 4(b).

3.4 CoMFA and CoMSIA models of Flt1 and cKit inhibitors

The CoMFA models of Flt1 and cKit inhibitors were also built after model development and validation based on internal predictions of the training set and the external predictions of the test set. PLS analyses of the Flt1 and cKit inhibitor based on the Hip-Hop algorithm showed a high cross-validated q^2 value of 0.736 and 0.603 using six and three principal components, respectively. The non-cross-validated r^2 values were 0.973 and 0.976. In comparison with the CoMFA models based on MCS alignment, the pharmacophore-based models for Flt1 and cKit are inferior in predictable ability with the cross-validation q^2 value (0.441 and 0.378 for the MCS-based model, respectively) and the non-cross-validation r^2 value (0.943 and 0.951 for the MCS-based model, respectively). The statistical parameters for these CoMFA models developed are presented in Table 5. From Table 5, we can see that all compounds in the three test sets yield a good predicted pIC_{50} within 1 log unit of the experimental value. Figure 4 shows the actual vs. predicted values from three CoMFA models for training and test sets.

The CoMSIA models obtained with PLS analyses comprising steric, electrostatic, hydrophobic and HBD fields exhibited significant q^2 values and were used for the predictive performance on the test set. Using the default CoMSIA parameters and all of the compounds in the Flt1 and cKit inhibitors, we derived models with q^2 value of 0.691 and 0.582 with five and three components, respectively. Compared with the q^2 value for the CoMFA models, the CoMSIA model showed lower q^2 value. However, the performance for the prediction of the activities of the training set was still reliable. The CoMSIA models built from cKit and Flt1 inhibitor training sets showed significance with the F values (175.39 and 270.83, respectively), and non-cross-validation values of r^2 (0.954 and 0.965, respectively), much higher than the MCS-based alignment methods. All of the parameters obtained from the CoMSIA models are presented in Table 5.

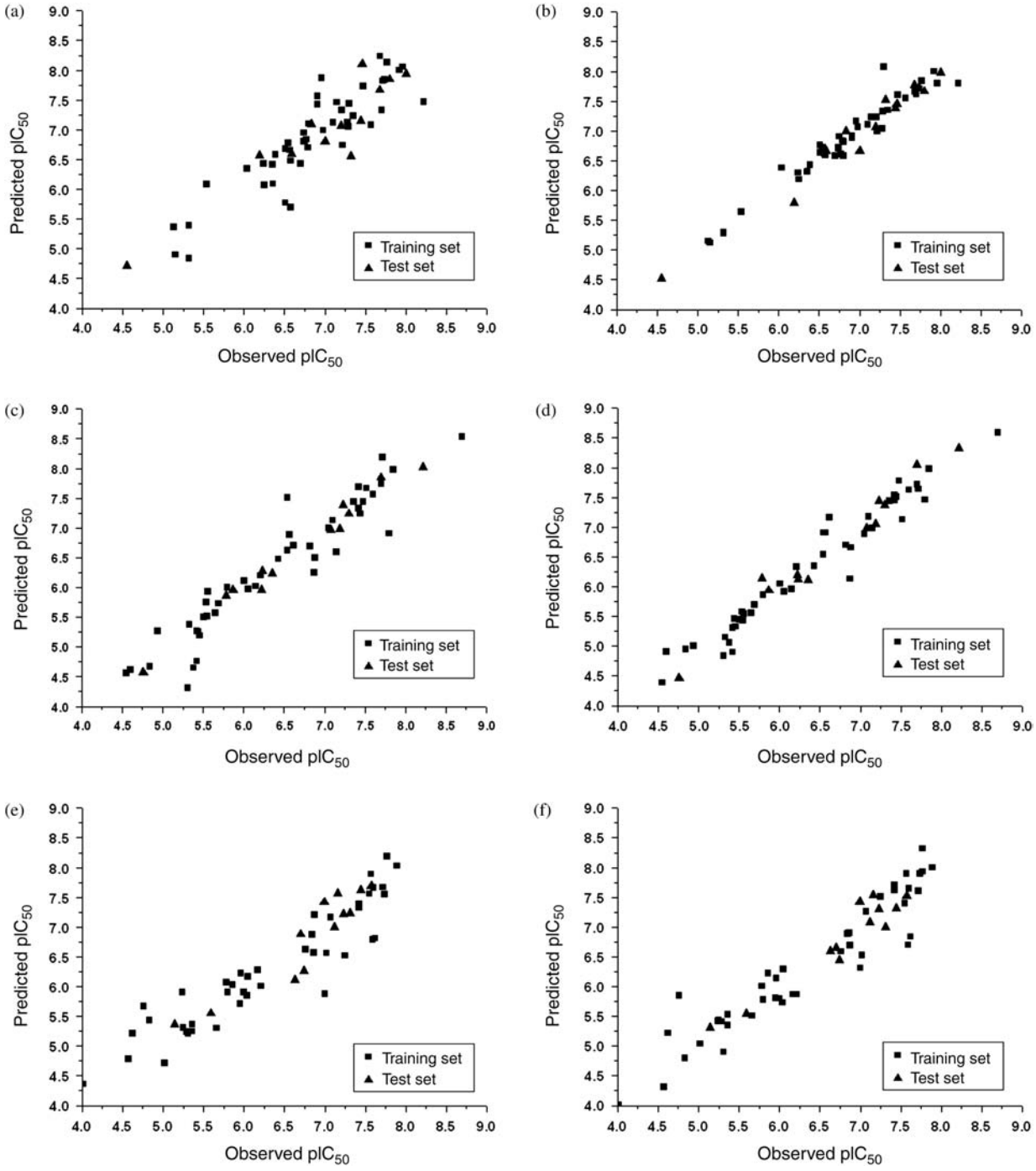


Figure 4. (a) The graph of observed vs. predicted KDR inhibitory activities (pIC_{50}; nM) of both training-set and test-set compounds using the best CoMFA model. (b) The graph of observed vs. predicted KDR inhibitory activities (pIC_{50}; nM) of both training-set and test-set compounds using the best CoMSIA model. (c) The graph of observed vs. predicted Flt1 inhibitory activities (pIC_{50}; nM) of both training-set and test-set compounds using the best CoMFA model. (d) The graph of observed vs. predicted Flt1 inhibitory activities (pIC_{50}; nM) of both training-set and test-set compounds using the best CoMSIA model. (e) The graph of observed vs. predicted cKit inhibitory activities (pIC_{50}; nM) of both training-set and test-set compounds using the best CoMFA model. (f) The graph of observed vs. predicted cKit inhibitory activities (pIC_{50}; nM) of both training-set and test-set compounds using the best CoMSIA model.

Since in the MCS-based alignment, the lowest energy conformation of compound **3** was used as the template of KDR and Flt1 inhibitors (compound **9** was used as the template of cKit inhibitors), and for other compounds the lowest energy conformations were chosen as the active conformations, the same aligned method was used for

building the models of cKit and Flt1 inhibitors. This method assumed that all compounds interacted with the three kinases with the same binding mode. Considering most compounds in our manuscript had obvious different inhibitory activities for the three kinases, this assumption was very suspicious. The results, together with the statistically significant correlations between the actual and the predicted kinase inhibitory activities, demonstrated the power of a common feature pharmacophore algorithm-based conformational alignment scheme and proved superior over the MCS-based conformational alignment scheme implemented in the SYBYL software.

3.5 Graphical interpretation of the results

The contour maps of CoMFA denote the region in the space where the aligned molecules would favourably or unfavourably interact with the receptor, while the CoMSIA contour maps denote these areas within the specified region where the presence of a group with a particular physicochemical activity binds to the receptor. The CoMFA/CoMSIA results were graphically interpreted by field contribution maps using the 'STDEV*COEFF' field type.

3.5.1 CoMFA contour maps

Figure S5 of the Supporting Information (A–F) shows the contour maps derived from the CoMFA PLS model. The more potent analogue, compound **3**, was embedded in the maps A–D, while compound **9** was embedded in maps E and F to demonstrate its affinity for the steric and electrostatic regions of inhibitors. Yellow areas indicate regions of steric hindrance to activity, while green areas indicate a steric contribution to potency. The blue regions indicate positive electrostatic charge potential associated with increased activity, while red regions show negative charge with increased activity. All of the contours represented the default 80 and 20% level contributions for favoured and disfavoured regions, respectively.

Figure S5 of the Supporting Information shows that the different physicochemical field property contours are mainly distributed within the region surrounding the diaryl urea unit and near the region enclosed by the 6,7-position of the 1,4-dihydroindeno[1,2-c]pyrazoles of the reference inhibitor. This suggests that these functional groups tune the affinity of each ligand. The green contour for the steric properties derived from the CoMFA studies indicates favourable steric interaction in the 7-position of the 1,4-dihydroindeno[1,2-c]pyrazoles. This will benefit the inhibitor for increasing the activity with the KDR receptor (Figure S5(a) of the Supporting Information). As to electrostatic properties, the blue contour presented in the right side of the diaryl urea unit region in the map suggests that positive electrostatic charge groups, e.g. the more positive charged nitrogen may favour enhanced affinity between KDR and its inhibitors (Figure S5(b) of the Supporting Information). The less bulky and positive substitution (e.g. positively charged nitrogen) on the 8-position of 1,4-dihydroindeno[1,2-c]pyrazoles would increase the activity of KDR and cKit inhibitors, while the negative substitution (e.g. functions containing fluorine or oxygen) is favoured for Flt1 of inhibitors. The contributions from the steric and electrostatic fields for the present models are 0.491/0.509, 0.507/0.493 and 0.465/0.535 (Table 2), respectively, for the KDR, Flt1 and cKit CoMFA models. Such contributions of field indicated that the variations in binding affinity among these inhibitors are dominated by steric interactions, but are distributed in different proportions across the binding sites of each of the three kinase receptors. This factor can be applied to design high potent and selective KDR inhibitors.

3.5.2 CoMSIA contour maps

Compared to standard CoMFA, the major advantage of CoMSIA is better ability to visualise and interpret the obtained correlation in terms of the field contributions. Five contribution maps obtained by CoMSIA studies may be used to identify features important for interaction between small molecules and protein. The electrostatic and hydrophobic field contours of CoMSIA are presented in Figure 5. Comparison of CoMSIA electrostatic maps indicates that the presence of a negative charge moiety in the urea region is crucial for kinase selectivity. Differences found in the region of the 6, 7 and 8 positions of the 1,4-dihydroindeno[1,2-c]pyrazole units indicate that electronegative substitutions at this region will increase the selectivity of KDR over Flt1 and cKit receptors.

Table 5 shows that the hydrophobic field contributions for the three CoMSIA models show a relatively high contribution (0.264, 0.246 and 0.385, for the KDR, Flt1 and cKit models, respectively). The yellow and white regions shown in Figure 5 represent those areas of favourable and unfavourable hydrophobic interactions, respectively. The larger white regions at the 6-position rings shown in Figure 5(b), (d), (f) highlight the fact that hydrophobic substitutions will decrease the inhibitor affinity with three kinases. Combined with the electrostatic and other field contour maps, reasonable structural modifications can be carried out to improve the activity and selectivity of KDR inhibitors.

The contributions of the descriptor fields also show the differences of binding mode between inhibitors and three different receptors. The contributions of the steric field are 0.165 and 0.162 for the KDR and Flt1 CoMSIA models, respectively, while a high contribution is 0.277 for the cKit CoMSIA model. The donor field contribution is 0.243 for

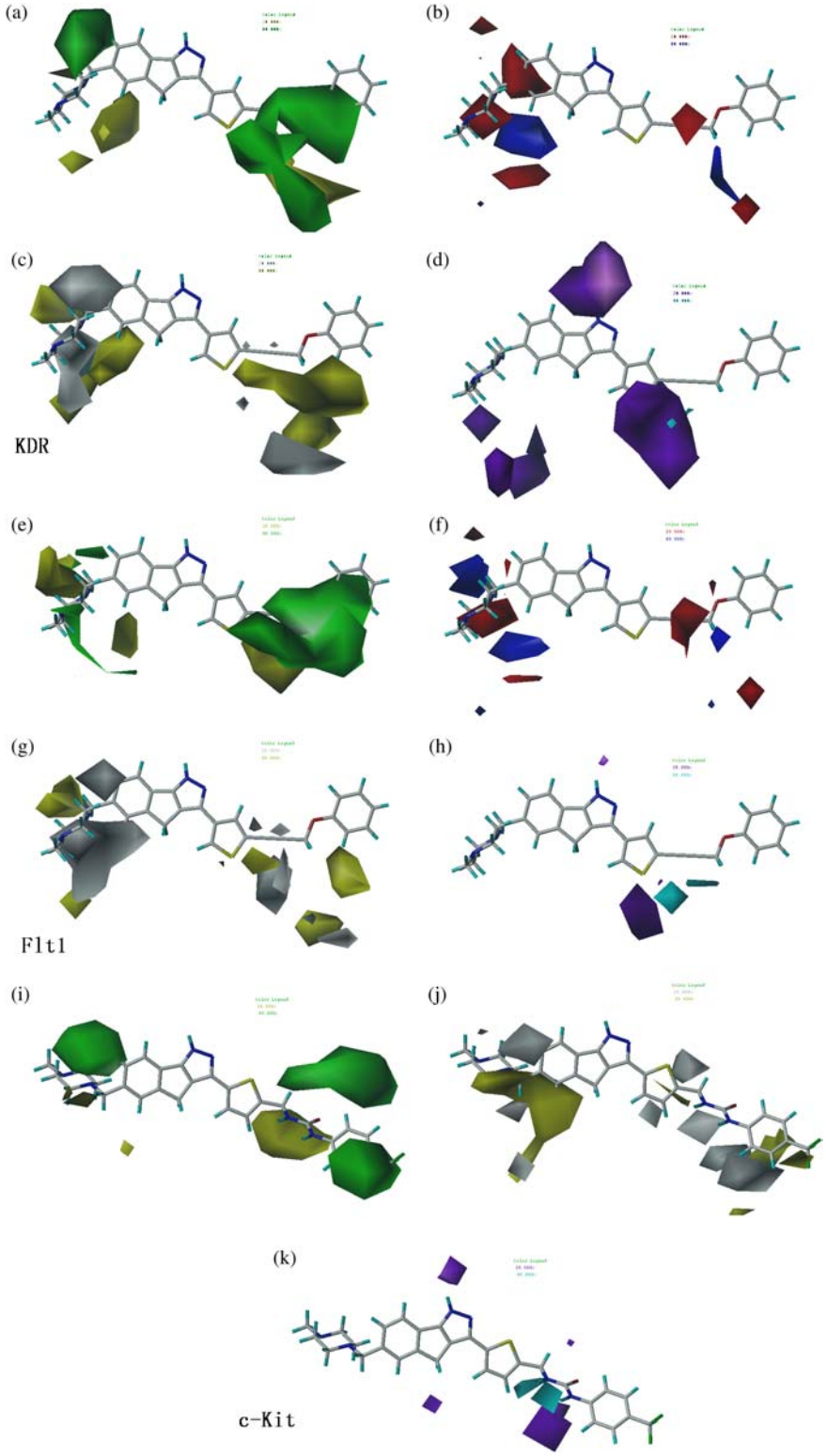


Figure 5. The contour maps of CoMSIA models for three receptors. (a), (e), (i) The CoMFA steric field contour maps, sterically favoured regions are in green; sterically disfavoured regions are in orange. (b), (f) The electrostatic field contour maps, red regions indicate disfavoured areas of positive potential, cyan regions indicate favoured areas of positive potential. (c), (g), (j) The hydrophobic field contour maps. The orange regions indicate favourable hydrophobic interactions and white regions indicate unfavourable hydrophobic interactions. (d), (h), (k) The HBD field contour maps. The cyan regions indicate increased HBD activity and purple regions indicate decreased HBD activity. Compound **3** was embedded in the maps (a) to (h) while compound **9** was embedded in the maps (i), (j) and (k).

KDR, which is lower than the value for Flt1 (0.271) and the value for cKit (0.338). These parameters would alert us to increase the selectivity of inhibitors by changing this field contribution on the right position.

4. Conclusions

The development of RTK inhibitors is an active area of drug discovery within the pharmaceutical industry. These 3D-QSAR studies carried out combining CoMFA and CoMSIA have led to the identification of some of the important regions of the inhibitor that possesses specific steric, hydrophobic, electronic and hydrogen bond interactions with three different kinases. Different 3D-QSAR models have been developed, KDR models, cKit models and Flt1 models, to represent the molecular fields associated with the lead compounds acting at the KDR, cKit and Flt1 receptors, respectively. An analysis of the model parameters and contour maps reveals the factors that influence the activity of those compounds for different receptors. This information is vital to steer the rational design of new drugs. Moreover, the results, together with the statistically significant correlations between the actual and predicted kinase inhibitory activities, demonstrated the power of a common feature pharmacophore (Hip-Hop) algorithm-based conformational alignment scheme and proved superior over the MCS-based conformational alignment scheme implemented in the SYBYL software. The studies suggest that in the development of 3D-QSAR models, the PBA may be useful in getting the robust predictive models which may provide useful information required for proper understanding of the important structural and physicochemical features for designing novel selective kinase inhibitors comprising novel scaffolds leading to the candidate molecules as antitumour agents for drug development.

Acknowledgements

This work was supported by the Science Foundation for Young Scientists of the education office of Sichuan Province, China (No. 2008ZB056). This work was also supported by the National Natural Science Foundation for the Youth of China (Grant No. 30901837). The authors acknowledge the assistance of the State Key Lab of Biotherapy, Sichuan University, for the operation of the Catalyst software on IBM graphic workstation. The authors also acknowledge the assistance of the Kunming Institute of Botany, Chinese Academy of Sciences, for the operation of the SYBYL software on SGI graphic workstation.

References

[1] P. Carmeliet and R.K. Jain, *Angiogenesis in cancer and other diseases*, Nature 407 (2000), pp. 249–257.
[2] J. Folkman, *Clinical applications of research on angiogenesis*, N. Engl. J. Med. 333 (1995), pp. 1757–1763.
[3] J. Folkman, *Angiogenesis in cancer, vascular, rheumatoid and other disease*, Nat. Med. 1 (1995), pp. 27–30.
[4] B.R. Zetter, *Angiogenesis and tumor metastasis*, Annu. Rev. Med. 49 (1998), pp. 407–424.
[5] J.T. Kuethe, A. Wong, C.X. Qu, J. Smitrovich, I.W. Davies, and D.L. Hughes, *Synthesis of 5-substituted-1H-indol-2-yl-1Hquinolin-2-ones: A novel class of KDR kinase inhibitors*, J. Org. Chem. 70 (2005), pp. 2555–2567.
[6] K.A. Thomas, *Vascular endothelial growth factor, a potent and selective angiogenic agent*, J. Biol. Chem. 271 (1996), pp. 603–606.
[7] B.I. Terman, M. Dougher-Vermazen, M.E. Carrion, D. Dimitrov, D.C. Armellino, D. Gospodarowicz, and P. Bohlen, *Identification of the KDR tyrosine kinase as a receptor for vascular endothelial cell growth factor*, Biochem. Biophys. Res. Commun. 187 (1992), pp. 1579–1586.
[8] C. De Vries, J.A. Escobedo, H. Ueno, K. Houck, N. Ferrara, and L.T. Williams, *The fms-like tyrosine kinase, a receptor for vascular endothelial growth factor*, Science 255 (1992), pp. 989–991.
[9] G.D. Yancopoulos, S. Davis, N.W. Gale, J.S. Rudge, S.J. Wiegand, and J. Holash, *Vascular-specific growth factors and blood vessel formation*, Nature 407 (2000), pp. 242–248.
[10] M.T. Bilodeau, M.E. Fraley, and G.D. Hartman, *Kinase insert domain-containing receptor kinase inhibitors as anti-angiogenic agents*, Expert Opin. Investig. Drugs 11 (2002), pp. 737–745.
[11] M. Shibuya and L. Claesson-Welsh, *Signal transduction by VEGF receptors in regulation of angiogenesis and lymphangiogenesis*, Exp. Cell Res. 312 (2006), pp. 549–560.
[12] M.M. Dikov, J.E. Ohm, N. Ray, E.E. Tchekneva, J. Burlison, D. Moghanaki, S. Nadaf, and D.P. Carbone, *Differential roles of vascular endothelial growth factor receptors 1 and 2 in dendritic cell differentiation*, J. Immunol. 174 (2005), pp. 215–222.
[13] C.D. Mol, D.R. Dougan, T.R. Schneider, R.J. Skene, M.L. Kraus, D.N. Scheibe, G.P. Snell, H. Zou, B.C. Sang, and K.P. Wilson, *Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase*, J. Biol. Chem. 279 (2004), pp. 31655–31663.
[14] N. Ferrara and T. Davis-Smyth, *The biology of vascular endothelial growth factor*, Endocr. Rev. 18 (1997), pp. 4–25.
[15] P. Borgstroem, K.J. Hillan, P. Sriramarao, and N. Ferrara, *Complete inhibition of angiogenesis and growth of microtumors by anti-vascular endothelial growth factor neutralizing antibody: Novel concepts of angiostatic therapy from intravital video microscopy*, Cancer Res. 56 (1996), pp. 4032–4039.
[16] A.P. Adamis, D.T. Shima, M.J. Tolentino, E.S. Gragoudas, N. Ferrara, J. Folkman, P.A. D'Amore, and J.W. Miller, *Inhibition of vascular endothelial growth factor prevents retinal ischemia-associated iris neovascularization in a nonhuman primate*, Arch. Ophthalmol. 114 (1996), pp. 66–71.
[17] P. Prathipati, G. Pandey, and A.K. Saxena, *CoMFA and docking studies on glycogen phosphorylase a inhibitors as antidiabetic agents*, J. Chem. Inf. Model. 45 (2005), pp. 136–145.
[18] S.J. Cho, M.L. Garsia, J. Bier, and A. Tropsha, *Structure-based alignment and comparative molecular field analysis of acetylcholinesterase inhibitors*, J. Med. Chem. 39 (1996), pp. 5064–5071.
[19] I. Akritopoulou-Zanze, D.H. Albert, P.F. Bousquet, G.A. Cunha, C.M. Harris, M. Moskey, J. Dinges, K.D. Stewart, and T.J. Sowin, *Synthesis and biological evaluation of 5-substituted 1,4-dihydro-indeno[1,2-c]pyrazoles as multitargeted receptor tyrosine kinase inhibitors*, Bioorg. Med. Chem. Lett. 17 (2007), pp. 3136–3140.
[20] J. Dinges, K.L. Ashworth, I. Akritopolou-Zanze, L.D. Arnold, S.A. Baumeister, P.F. Bousquet, G.A. Cunha, S.K. Davidsen, S.W. Djuric, V.J. Gracias, M.R. Michaelides, P. Rafferty, T.J. Sowin, K.D. Stewart, Z. Xia, and H.Q. Zhang, *1,4-Dihydroindeno[1,2-c]pyrazoles as novel multitargeted receptor tyrosine kinase inhibitors*, Bioorg. Med. Chem. Lett. 16 (2006), pp. 4266–4271.
[21] R.C. Wade, *Molecular interaction fields*, in *3D QSAR in Drug Design*, H. Kubinyi, ed., ESCOM, Leiden, 1993, pp. 486–505.
[22] S. Wold, E. Johansson, and M. Cocchi, *PLS – partial least squares projections to latent structures*, in *3D QSAR in Drug Design, Theory Methods and Applications*, H. Kubinyi, ed., ESCOM, Leiden, 1993, pp. 523–550.
[23] G. Klebe, U. Abraham, and T. Mietzner, *Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to*

correlate and predict their biological activity, J. Med. Chem. 37 (1994), pp. 4130–4146.

[24] G. Klebe and U. Abraham, *Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries*, J. Comput.-Aided Mol. Des. 13 (1999), pp. 1–10.

[25] M. Bohm, J. Sturzebecher, and G. Klebe, *Three-dimensional quantitative structure–activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa*, J. Med. Chem. 42 (1999), pp. 458–477.

[26] S.S. Chaudhaery, K.R. Kuldeep, and A.K. Saxena, *Consensus superiority of the pharmacophore-based alignment, over maximum common substructure (MCS): 3D-QSAR studies on carbamates as acetylcholinesterase inhibitors*, J. Chem. Inf. Model. 49 (2009), pp. 1590–1601.