QM/MM based 3D QSAR models for potent B-Raf inhibitors

Jae Yoon Chung · Hwan Won Chung ·
Seung Joo Cho · Jung-Mi Hah · Art E. Cho

Abstract Three dimensional (3D) quantitative structure-activity relationship studies of 37 B-Raf inhibitors, pyrazole-based derivatives, were performed. Based on the co-crystallized compound (PDB ID: 3D4Q), several alignment methods were utilized to derive reliable comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) models. Receptor-guided alignment with quantum mechanics/molecular mechanics (QM/MM) minimization led to the best CoMFA model ($q^2 = 0.624$, $r^2 = 0.959$). With the same alignment, a statistically reliable CoMSIA model with steric, H-bond acceptor, and hydrophobic fields was also derived ($q^2 = 0.590$, $r^2 = 0.922$). Both models were validated with an external test set, which gave satisfactory predictive r^2 values of 0.926 and 0.878, respectively. Contour maps from CoMFA and CoMSIA models revealed important structural features responsible for increasing biological activity within the active site and explained the correlation between biological activity and receptor-ligand interactions. New fragments were identified as building blocks which can replace R1-3 groups through combinatorial screening methods. By combining these fragments a compound with a high bioactivity level prediction was found. These results can offer useful information for the design of new B-Raf inhibitors.

Keywords B-Raf · CoMFA · CoMSIA · QM/MM · QSAR · Combinatorial screening

J. Y. Chung · A. E. Cho (✉)
Department of Biotechnology and Bioinformatics, Korea
University, Jochiwon-Eup, Yeongi-Gun, Chungnam 339-700,
Korea
e-mail: artcho@korea.ac.kr

J. Y. Chung · J.-M. Hah
Life Sciences Research Division, Korea Institute of Science
and Technology, PO Box 131, Cheongryang, Seoul 130-650,
South Korea
e-mail: jhah@kist.re.kr

H. W. Chung
Computational Science Center, Future Fusion Technology
Division, Korea Institute of Science and Technology,
PO Box 131, Cheongryang, Seoul 130-650, South Korea

S. J. Cho
Research Center for Resistant Cells, Chosun University,
Gwangju 501-759, Korea

S. J. Cho
Department of Cellular and Molecular Medicine, College of
Medicine, Chosun University, 375 Seosuk-Dong, Dong-Gu,
Gwangju 501-759, Korea

Introduction

B-Raf, a serine/threonine kinase, forms part of a conserved signaling pathway that regulates cell growth, differentiation, and proliferation [1, 2]. Under normal circumstances, B-Raf is activated downstream of receptors in the cell membrane in a RAS small G-protein dependent manner. It then activates the protein kinase MEK, which in turn activates a third protein kinase called ERK. Activated ERK kinases phosphorylate a number of nuclear and cytoplasmic targets to initiate various cellular decisions.

Mutations of the Raf protein kinase have been found in approximately 7% of human cancers, with particularly high frequency in melanoma (50–70%), ovarian (35%), thyroid (30%), and colorectal (10%) cancers [3, 4]. There have been 40 reported B-Raf mutations. The most frequent mutation of Raf protein (>85%) is a valine substitution by glutamic acid at position 600 (V600E), which shows a

500-fold increase in catalytic activity. The discovery of the V600E B-Raf mutation has raised expectations for targeted therapy against human melanoma [5].

Highly potent and selective inhibitors, pyrazole-based derivatives, toward B-Raf kinase were reported by Hansen et al. in 2008 [6]. The scaffold of pyrazole-based derivatives and the biological activities, pIC_{50} (log1/IC_{50}), of 37 compounds are shown in Table 1. The researchers solved one crystal structure of B-Raf inhibitor complex (PDB code 3D4Q), which can be used to understand the important structural features for B-Raf inhibition and help in the design of more potent inhibitors. Figure 1 shows the inhibitor, 19 (Table 1), within the active site of the complex structure (3D4Q). Analyzed structural features are as follows: (1) the H-bond interaction between amino pyridine of 19 and the backbone of CYS532 residue in the hinge region, common for most potent protein kinase inhibitors [7]; (2) the H-bond interaction between the oxime group and LYS483; (3) the H-bond interaction between the piperidine group and ILE463.

The three-dimensional quantitative structure activity relationship (3D-QSAR) methods used here are comparative molecular field analysis (CoMFA), and comparative molecular similarity indices analysis (CoMSIA) [8–11]. The following five different alignments were applied to derive more predictable 3D-QSAR models: *ligand-based alignment*, *pharmacophore-based alignment*, *ligand- & pharmacophore-based alignment* and 2 *receptor-guided alignments with energy minimization (EM) or quantum mechanics (QM) minimization*. The information provided by the derived 3D-QSAR offers insight into the influence of various field interactions and thus assists in the design of new potent B-Raf inhibitors.

We used a QSAR model generated with this procedure to predict the activity of the new compound. Though CoMFA and CoMSIA both have the advantage of intuitive understanding of the models, they are weak in the prediction of compounds which have different structures from the one used for model generation. By performing combinatorial screening based on the fragment library to assemble compounds that resemble those in the data set we were able to obtain new R1-3 groups. We further aligned these compounds by docking and predicted activity through the CoMFA model. Those with high activity values were used in this work.

Materials and computational method

Inhibitor data set

A set of pyrazole-based derivatives was obtained from published data after compounds without clear IC_{50} values

were omitted to derive QSAR models. The pIC_{50} values were used as the dependent variable in QSAR models. A total of 37 compounds were divided into the training set and the test set. The training set consisted of 31 compounds selected randomly to generate QSAR models. Structures and activity values are listed in Table 1.

Preparation of ligands and receptor

The complex structures of 19 and B-Raf protein were obtained from the protein data bank (PDB code 3D4Q) and prepared using the Schrödinger Maestro program. All water molecules were removed from the structure. Because 19 can give biologically active conformations of the scaffold which are reasonable, it was selected as a template. The structures of other compounds were drawn based on the template, and their 3D conformations were generated using the Schrödinger LigPrep program with the OPLS 2005 force field [12].

CoMFA and CoMSIA

Before calculating interaction between CoMFA descriptor and compounds, all compounds were aligned in a certain 3D cubic lattice with grid spacing of 1.0, 1.5, and 2.0 Å. The steric (Lennard-Jones 6–12 potential) and electrostatic (Coulombic potential) field energies were calculated by CoMFA descriptors using a sp^3 carbon probe with van der Waals radius of 1.52 Å and +1 charge at each lattice point [8]. Energy terms below 30 kcal/mol, the default cutoff level, were truncated. A CoMSIA model was also derived based on the same aligned compounds in CoMFA. Similarity indices for a compound's atoms in each grid point were calculated and physicochemical properties (steric, electrostatic, hydrophobic, hydrogen bond donor, and acceptor) were evaluated [9, 10]. CoMSIA, using Gaussian-type distance dependence, provides smooth and interpretable contour maps. A default value of 0.3 was used as the attenuation factor.

PLS

Partial least squares (PLS) methodology for cyclic cross-validation with a leave-one-out (LOO) method was used to derive QSAR models [13, 14]. PLS is useful to solve a series of equations with many more independent variables than equations. The descriptors of CoMFA and CoMSIA were used as independent variables, and pIC_{50} activity values were used as dependent variables. The number of independent variables by QSAR descriptors is usually much larger than that of dependent variables which are compounds with pIC_{50} in the data set. PLS can solve those equations and generate QSAR models. A cross-validation

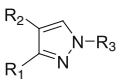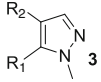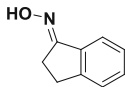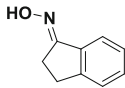**Table 1** The strcutres and observed b-Raf inhibitory activities
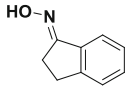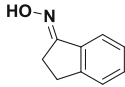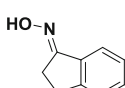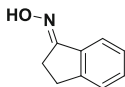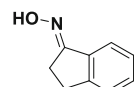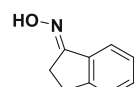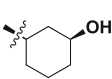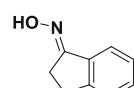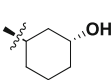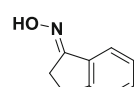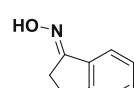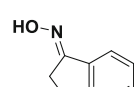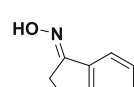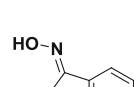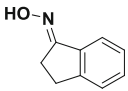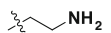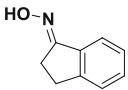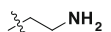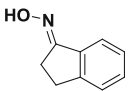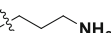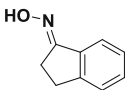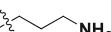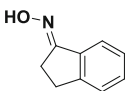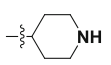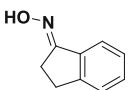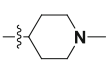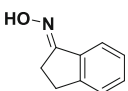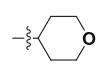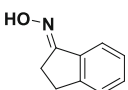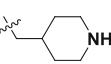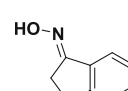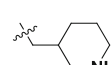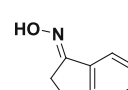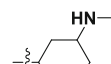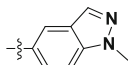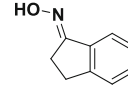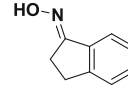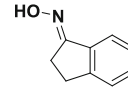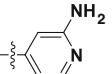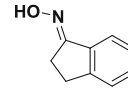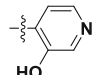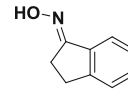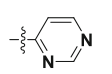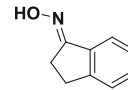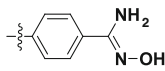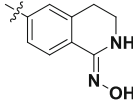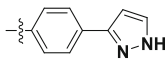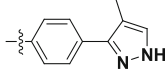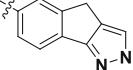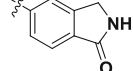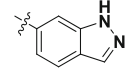
No		R1	R2	R3	IC$_{50}$(M)	pIC50
1	*	4-pyridyl		H	0.15×10^{-9}	6.824
2		4-pyridyl		CH$_3$	0.02×10^{-9}	7.699
3		4-pyridyl		-	0.48×10^{-9}	6.319
4		4-pyridyl		CH$_2$CH$_3$	0.57×10^{-9}	6.244
5		4-pyridyl		Cyclopentyl	3.46×10^{-9}	5.461
6	*	4-pyridyl			3.91×10^{-9}	5.408
7		4-pyridyl			0.33×10^{-9}	6.481
8		4-pyridyl			0.04×10^{-9}	7.398
9		4-pyridyl			0.09×10^{-9}	7.046
10		4-pyridyl		CH$_2$CH$_2$OCH$_3$	1.16×10^{-9}	5.936
11		4-pyridyl		CH$_2$CH$_2$OH	0.13×10^{-9}	6.886
12		4-pyridyl		CH$_2$CH$_2$CH$_2$OH	0.05×10^{-9}	7.301
13	*	4-pyridyl		CH$_2$CH(OH)CH$_2$OH	0.87×10^{-9}	6.060
14		4-pyridyl		CH(CH$_2$OH)$_2$	0.56×10^{-9}	6.252

Table 1 continued

15	*	4-pyridyl	(HO-N indanone oxime)	(ethyl-NH₂)	2.40×10^{-9}	5.620
16		4-pyridyl	(HO-N indanone oxime)	(ethyl-NH₂)	4.02×10^{-9}	5.396
17		4-pyridyl	(HO-N indanone oxime)	(propyl-NH₂)	0.39×10^{-9}	6.409
18		4-pyridyl	(HO-N indanone oxime)	(butyl-NH₂)	0.47×10^{-9}	6.328
19		4-pyridyl	(HO-N indanone oxime)	(piperidine NH)	0.03×10^{-9}	7.523
20		4-pyridyl	(HO-N indanone oxime)	(N-methyl piperidine)	0.34×10^{-9}	6.469
21		4-pyridyl	(HO-N indanone oxime)	(tetrahydropyran O)	0.22×10^{-9}	6.658
22	*	4-pyridyl	(HO-N indanone oxime)	(piperidine NH)	1.20×10^{-9}	5.921
23		4-pyridyl	(HO-N indanone oxime)	(piperidine NH)	0.85×10^{-9}	6.071
24		4-pyridyl	(HO-N indanone oxime)	(cyclohexyl HN—)	0.23×10^{-9}	6.638
25		(1-methyl indazol-5-yl)	(HO-N indanone oxime)	CH₃	969.60×10^{-9}	3.013
26		3-pyridyl	(HO-N indanone oxime)	CH₃	433.80×10^{-9}	3.363
27	*	(oxazol-5-yl)	(HO-N indanone oxime)	CH₃	4.10×10^{-9}	5.387
28		(2-amino-pyridin-4-yl)	(HO-N indanone oxime)	CH₃	2.10×10^{-9}	5.678
29		(3-hydroxy-pyridin-4-yl)	(HO-N indanone oxime)	CH₃	2.90×10^{-9}	5.538
30		(pyrimidin-4-yl)	(HO-N indanone oxime)	CH₃	3.10×10^{-9}	5.509

Table 1 continued

31	4-pyridyl	[structure: phenyl with C(NH₂)=N–OH amidoxime]	CH₃	18.40×10^{-9}	4.735
32	4-pyridyl	[structure: tetrahydroisoquinoline with N–OH]	CH₃	16.00×10^{-9}	4.796
33	4-pyridyl	[structure: phenyl-pyrazole]	CH₃	68.40×10^{-9}	4.165
34	4-pyridyl	[structure: phenyl-methylpyrazole]	CH₃	4.40×10^{-9}	5.357
35	4-pyridyl	[structure: indeno-pyrazole]	CH₃	1.50×10^{-9}	5.824
36	4-pyridyl	[structure: isoindolinone]	CH₃	68.40×10^{-9}	4.165
37	4-pyridyl	[structure: indazole]	CH₃	70.00×10^{-9}	4.155

* Test set compounds
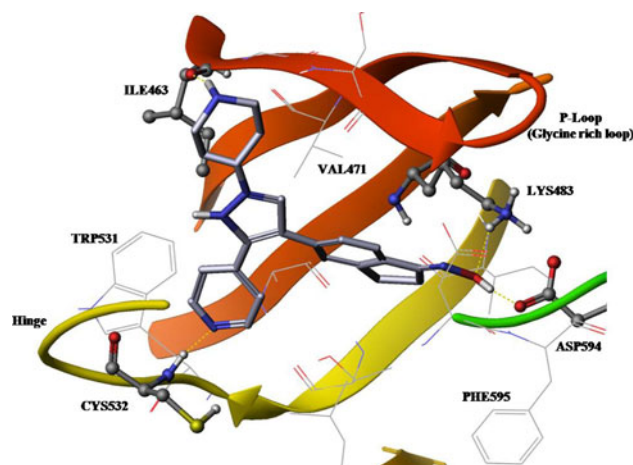
Fig. 1 X-ray crystal structure of b-Raf and its inhibitor, **19**

procedure evaluates the quality of the generated QSAR model. The cross-validation analysis was performed using the LOO method [15]. This procedure repeatedly reproduces new models with data sets where one or more compounds are omitted; then the reproduced models predict the omitted observations. The cross-validated coefficient, q^2, is calculated using Eq. (1):

$$q^2 = 1 - \frac{\sum (Y_{predicted} - Y_{observed})^2}{\sum (Y_{observed} - Y_{mean})^2} \qquad (1)$$

where $Y_{predicted}$, $Y_{observed}$, and Y_{mean} are predicted, actual, and mean values of the target property (pIC₅₀), respectively. $\Sigma(Y_{predicted} - Y_{observed})^2$ is the predictive sum of squares (PRESS). Deciding the best number of PLS components is needed to avoid data overfitting, and the corresponding number with the lowest PRESS value is used to derive the final PLS regression models.

Alignments for CoMFA and CoMSIA

The steric and electrostatic energies from CoMFA and CoMSIA were quite effective for understanding important non-covalent and shape-dependent interactions between small compounds and their target protein because CoMFA and CoMSIA models are derived based upon the assumption that 3D structures of input compounds are aligned according to their biologically active conformations when compounds bind to and interact with their target protein [16, 17]. Although there is no particular alignment method which produces biologically active conformations of compounds, plausible bioactive conformations of compounds can be produced by several different alignment methods: *ligand-based*, *pharmacophore-based*, and *receptor-guided*. The receptor-guided alignment can be accomplished when experimental or homologous structures of target proteins are available. Recently, various receptor-guided alignment

methods have been developed to mimic the biologically active conformations of compounds within the receptor structures using computational modeling methods. In addition to ligand-based and 2 pharmacophore-based alignments, energy minimization (EM) and quantum mechanics/molecular mechanics (QM/MM) energy minimization based alignments were used as receptor-guided alignments in the present study.

Ligand-based alignment

The first method, ligand-based alignment, assumes that relevant bioactive conformations of compounds are similar to optimized geometries at ground state energy level. The structures of all compounds, except **19,** were fully optimized using Schrödinger's MacroModel application [18]. The optimized compounds were then aligned over the core, pyrazole moiety. Figure 2a shows the ligand-based alignment.

Pharmacophore-based alignment

Schrödinger's Phase program was used to develop a common pharmacophore hypothesis and align all compounds [19]. Using ConfGen in Schrödinger package, 619 of various conformers for all compounds, except **19,** were generated with distance dependent dielectric solvation treatment and OPLS 2005 force field [20]. Each compound in a set of conformers was retained with maximum energy difference of 10 kcal/mol compared to global energy minimum conformer. In Phase, pharmacophore features (**A**), hydrogen bond donor (**D**), hydrophobic group (**H**), negatively charged group (**N**), positively charged group (**P**), and aromatic ring (**R**), were defined to figure out chemical structure patterns as SMART queries and were mapped for each compound using a set of pharmacophore features. Compounds with pIC_{50} values more than 7.0 were selected as active compounds, except the test set, for producing the pharmacophore hypotheses. **AAHRR** pharmacophore features were selected as the common pharmacophore hypothesis (CPH). By applying the scoring function for the CPH, many alignment rules were identified. PLS analyses carried out using Phase, produced pharmacophore-based models using grid spacing of 1.0 Å with the training set. The best statistics pharmacophore model was selected for the alignment of all compounds except **19**. Figure 2b shows the pharmacophore-based alignment.

Ligand- and pharmacophore-based alignment

This alignment method is similar to the pharmacophore alignment. The difference between the two methods is in generation of conformers for each compound in the data set. While the previous method finds the best combination among various conformers for each compound based on selected pharmacophore features only, this alignment method uses the quantum mechanically optimized structures in ligand-based alignment as a conformer for each compound to be aligned to the pharmacophore features. Compounds whose pIC_{50} values are more than 7.0 were selected as active compounds, except for the test set which is used for producing the pharmacophore hypotheses. As in pharmacophore-based alignment, **AAHRR** pharmacophore features were selected as the CPH. Figure 2c shows the ligand- & pharmacophore-based alignment results.

Receptor-guided alignment with energy minimization

Since docking programs can generate and select bioactive conformations among a huge number of possible binding poses, a receptor-guided alignment with docking method is useful for a large data set of more than 100 compounds [21, 22]. In a small data set, however, it is possible to manually produce adequate bioactive conformations for each compound using other modeling tools. When the co-crystallized compound has a common scaffold with other compounds to be studied, manually produced poses can show high correlation to biological activity data. In receptor-guided alignment with EM, one can focus on stable conformations of different substituents for each compound within the active site. The input compounds are aligned based on the 3D conformations of the common scaffold within the active site, after which, possible conformations of substituents for each compound within the active site are determined using EM. Those possible conformations of compounds avoid steric hindrance in the active site. MacroModel program in the Schrödinger package was used for the EM method [18]. The top-ranked binding pose by MacroModel was chosen for each compound. According to the above procedure of receptor-guided alignment using EM, several plausible conformations of input compounds were produced using tMacroModel program based on the conformation of common scaffold in **19**. Figure 2d shows the results of receptor-guided alignment with EM.

Receptor-guided alignment with quantum mechanics/molecular mechanics minimization

The procedure of receptor-guided alignment with quantum mechanics/molecular mechanics (QM/MM) minimization is similar to the receptor-guided alignment with EM. Based on the conformation of **19**'s scaffold, conformations for other compounds were generated using QM/MM minimization. After placing the compounds manually in the active

site while overlapping the scaffolds, we performed QM/MM geometry optimization on each protein–ligand complex with fixed receptor and ligand only as QM regions. At the end of QM/MM minimizations, we calculated the atomic charges on ligands by electrostatic potential (ESP) fitting [23, 24]. QM/MM minimization can find more stable binding pose within the active site than molecular mechanic (MM) based. By calculating atomic charges at the QM/MM minimized pose and feeding it to CoMFA, the effect of QM/MM minimization can be maximized. For QM/MM, we used QSite version 3.5 with 6-31G* basis set, B3LYP functional density functional theory (DFT) for QM, and OPLS 2001 force field for MM [25]. Shown in Fig. 2e are the receptor-guided alignment results with QM/MM minimization within the active site.

(a) Ligand-based   **(b)** Pharmacophore-based

(c) Ligand- & pharmacophore-based   **(d)** EM-based

(e) QM/MM based
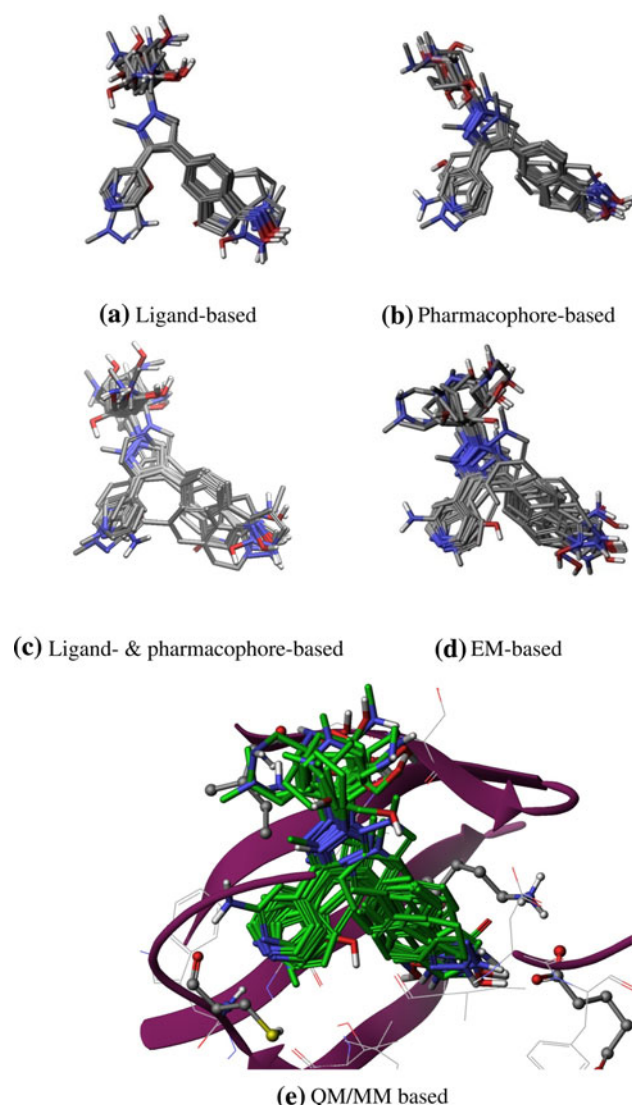
Fig. 2 Different alignment results

Training and test set selection

Most QSAR methods have drawbacks in that even though the model is created with the same data set, results can vary wildly depending on the training set. In case of CoMFA, even if the compound alignment in the data set is not adequate, one can obtain good statistic values with high predictive power depending on the training set. However, our model shows that various test sets as well as the 1 presented earlier can lead to good statistics with the alignment generated here (Fig. 3).

Combinatorial screening

CombiGlide, a part of the Schrödinger package, was used for searching various terminal groups (R1-3). A library was prepared of various fragment-sized molecules with structure files downloaded from the ZINC website. These were broken into fragments based on a simple rule that single bonds or any bonds in ring systems are never to be broken. CombiGlide attaches library molecules to the scaffold and fits them in the active site, after which binding affinities are calculated. CoMFA can only predict the differences in activities for various R-groups attached to the common scaffold in our data set. Therefore, to enhance the accuracy of the prediction by CoMFA, instead of screening the whole drug-like molecule library using simple docking, we used a combinatorial docking technique to screen the fragment based library for R-group substitutes with the scaffold in the data set. By doing this, we were able to find several terminal groups for R1-3 which have more potential to increase biological activities (Table 6).

Results and discussion

CoMFA and CoMSIA analyses

Based on four different alignments, CoMFA models were derived with different grid spacing by PLS analyses. The statistics summaries for different CoMFA models are listed in Table 2. The best CoMFA model, **10,** was obtained from QM/MM based alignment and grid spacing 1.0 Å ($q^2 = 0.600$, $r^2 = 0.950$, $F = 94.587$, SEE $= 0.292$). With the CoMFA model **13**, region focusing (RF) strategy was applied by reducing the number of independent variables in the CoMFA calculation in order to obtain a more predictable CoMFA model. RF-CoMFA model **16** with better statistics was produced ($q^2 = 0.624$, $r^2 = 0.959$, $F = 93.082$, SEE $= 0.270$). Table 3 shows the statistics summary for RF-CoMFA model **16**. To verify the statistical confidence and robustness of CoMFA models, 100 bootstrapping samplings were performed. A bootstrapped r^2_{boot} of 0.985 and a

Fig. 3 PLS statistics result according to selection of different test sets
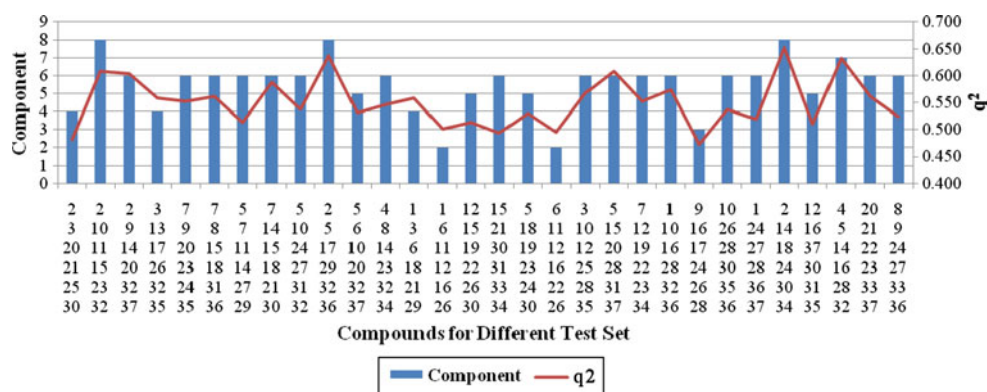
Table 2 Statistics of generated CoMFA models using different alignment methods

Model	Grid spacing (Å)	Leave-one-out cross-validation			Non-cross-validation			Bootstrap		Predictive R^2	Field contribution	
		q^2	n	SDEP	r^2	SEE	F-value	r^2_{boot}	SD	r^2_{pred}	Steric	Electro.
Alignment 1 (ligand-based alignment)												
1	1.0	0.224	1	0.986	0.371	0.888	20.619	0.421	0.808		0.407	0.593
2	1.5	0.231	1	0.981	0.377	0.883	21.213	0.494	0.831	0.377	0.409	0.591
3	2.0	0.224	1	0.986	0.368	0.890	20.355	0.421	0.810		0.396	0.604
Alignment 2 (pharmacophore-based alignment)												
4	1.0	0.267	1	0.959	0.417	0.855	25.012	0.492	0.741	0.417	0.426	0.574
5	1.5	0.265	1	0.960	0.417	0.855	25.021	0.483	0.795		0.433	0.567
6	2.0	0.265	1	0.960	0.417	0.855	25.019	0.473	0.814		0.433	0.567
Alignment 3 (ligand-& pharmacophore-based alignment												
7	1.0	0.178	2	1.029	0.462	0.833	14.615	0.615	0.101	0.462		
8	1.5	0.090	1	1.068	0.393	0.872	22.672	0.427	0.119		0.678	0.322
9	2.0	0.161	2	1.040	0.490	0.811	16.328	0.659	0.644		0.713	0.287
Alignment 4 (EM-based alignment)												
10	1.0	0.543	3	0.847	0.893	0.384	66.459	0.936	0.297		0.671	0.329
11	1.5	0.511	3	0.757	0.896	0.377	69.005	0.909	0.328	0.896	0.682	0.318
12	2.0	0.543	3	0.847	0.870	0.423	53.412	0.910	0.333		0.614	0.386
Alignment 5 (QMMM based alignment)												
13	1.0	0.600	5	0.824	0.950	0.292	94.587	0.969	0.208	0.905	0.423	0.577
14	1.5	0.572	5	0.852	0.942	0.294	80.601	0.983	0.146		0.419	0.581
15	2.0	0.554	5	0.869	0.964	0.237	105.852	0.987	0.140		0.442	0.558

Table 3 Statistics of the resulting CoMFA models

Model	Resion-focusing	Leave-one-out cross-validation			Non-cross-validation			Bootstrap		Predictive R^2	Field contribution	
		q^2	n	SDEP	r^2	SEE	F-value	r^2_{boot}	SD	r^2_{pred}	Steric	Electro.
Receptor-based CoMFA model												
13	1.0	0.600	5	0.824	0.950	0.292	94.587	0.969	0.208	0.905	0.423	0.577
CoMFA model, region focusing method												
16	PLS	0.624	6	0.815	0.959	0.270	93.082	0.985	0.149	0.926	0.454	0.546

standard deviation of 0.149 from RF-CoMFA model **16** were obtained (Table 3). These statistical values (RF-CoMFA 13) indicate a good internal consistency in the training data set.

The aligned compounds used for this CoMFA model **13** were applied to derive a CoMSIA model. In CoMSIA studies, all possible cases using five fields were calculated. Table 4 shows the statistics summaries of CoMSIA. The important values needed to determine the quality of QSAR models are q^2 and statistical components. The number of statistical components below 7 is preferred. Among CoMSIA models which satisfy these criteria, the CoMSIA model **20** having the highest q^2

value was chosen. The CoMSIA model **20** was produced using three field descriptors (steric, hydrophobic, and H-bond acceptor), the combination of which gave the best statistical result ($q^2 = 0.590$, $r^2 = 0.922$, $F = 77.135$, SEE $= 0.356$).

The observed activities and the predicted activities by the CoMFA, RF-CoMFA, and CoMSIA models are listed in Table 5, which shows that the activities predicted by the produced models are in good agreement with the observed data, suggesting that those three models have satisfactory predictive abilities. Figure 4a–c show plots of the observed data and the data predicted by the CoMFA, RF-CoMFA and CoMSIA models.

Table 4 The regression summary of CoMSIA models

No.	Field					LOO cross-validation		Non-cross-validation			r^2_{pred}
	S	E	H	D	A	q^2	n	r^2	SEE	F	
1	1.000	–	–	–	–	0.448	2	0.670	0.707	28.422	
2	–	1.000	–	–	–	0.554	5	0.931	0.341	67.692	
3	–	–	1.000	–	–	0.455	3	0.837	0.506	46.176	
4	–	–	–	1.000	–	0.272	1	0.430	0.912	21.913	
5	–	–	–	–	1.000	0.582	4	0.838	0.514	33.557	
6	0.180	0.820	–	–	–	0.531	4	0.914	0.373	69.497	
7	0.327	–	0.673	–	–	0.467	3	0.817	0.536	40.168	
8	0.220	–	–	0.780	–	0.374	1	0.512	0.844	30.478	
9	0.327	–	–	–	0.673	0.588	4	0.893	0.418	54.135	
10	–	0.640	0.360	–	–	0.524	2	0.816	0.528	62.133	
11	–	0.667	–	0.333	–	0.456	5	0.913	0.385	52.267	
12	–	0.585	–	–	0.415	0.533	3	0.883	0.428	68.227	
13	–	–	0.519	0.481	–	0.399	10	0.945	0.340	34.636	
14	–	–	0.474	–	0.526	0.589	4	0.913	0.377	68.164	
15	–	–	–	0.427	0.573	0.603	5	0.894	0.424	42.088	
16	0.180	0.541	0.279	–	–	0.523	2	0.803	0.546	56.998	
17	0.150	0.550	–	0.301	–	0.456	7	0.941	0.331	51.967	
18	0.132	0.498	–	–	0.370	0.547	3	0.887	0.421	70.780	
19	0.156	–	0.288	0.556	–	0.390	1	0.535	0.825	33.308	
20	0.182	–	0.368	–	0.450	0.590	4	0.922	0.356	77.135	0.878
21	0.202	–	–	0.369	0.429	0.611	7	0.952	0.297	65.499	
22	–	–	0.288	0.307	0.405	0.573	5	0.924	0.359	60.645	
23	–	0.409	–	0.259	0.332	0.551	4	0.902	0.400	59.795	
24	–	0.442	0.243	–	0.315	0.545	3	0.894	0.408	75.812	
25	–	0.481	0.247	0.272	–	0.453	4	0.907	0.390	63.157	
26	–	0.330	0.180	0.209	0.280	0.532	4	0.911	0.382	66.294	
27	0.130	–	0.247	0.291	0.333	0.586	7	0.959	0.276	76.301	
28	0.099	0.373	–	0.220	0.308	0.537	4	0.910	0.382	65.958	
29	0.102	0.384	0.217	–	0.297	0.551	4	0.930	0.338	86.301	
30	0.103	0.414	0.214	0.268	–	0.446	7	0.950	0.305	61.964	
31	0.088	0.287	0.368	0.217	0.239	0.531	6	0.950	0.296	76.462	

S steric, E electrostatic, H hydrophobic, D hydrogen bond donor A hydrogen bond acceptor, n number of statistical components, q^2 the LOO cross-validated correlation coefficient, r^2 the correlation coefficient, SEE standard estimated error, F the Fisher value, $r^2_{\text{predictive}}$ the correlation coeffient for test set
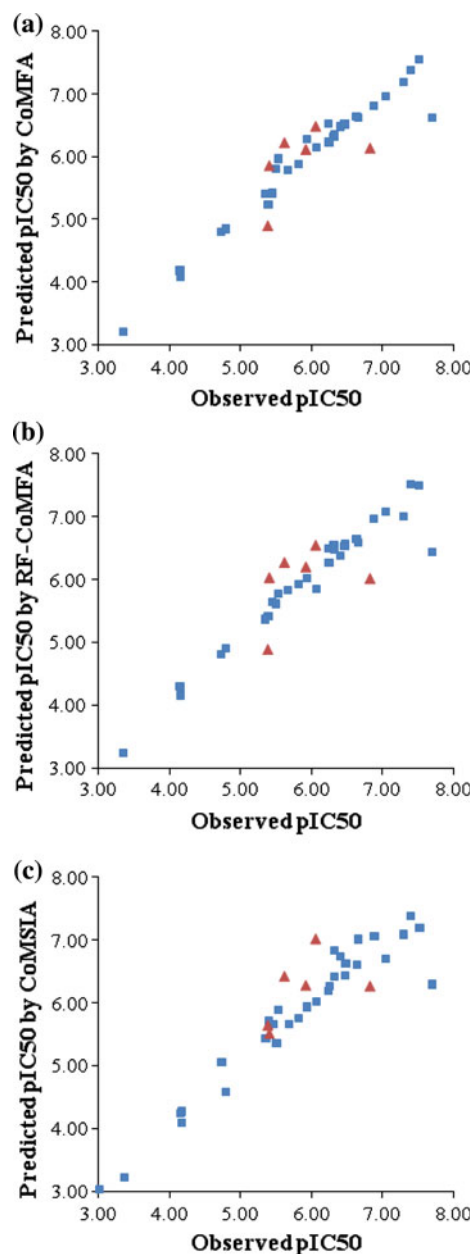
Table 5 The observed vs. predicted activities (pIC50) with deviations by CoMFA, RF-CoMFA, and CoMSIA

No.	pIC50	CoMFA		RF-CoMFA		CoMSIA	
		Prediction	Δ2	Prediction	Δ2	Prediction	Δ2
1*	6.824	6.137	0.472	6.015	0.654	6.264	0.314
2	7.699	6.625	1.153	6.431	1.608	6.298	1.963
3	6.319	6.313	0.000	6.542	0.050	6.840	0.271
4	6.244	6.525	0.079	6.486	0.059	6.196	0.002
5	5.461	5.412	0.002	5.647	0.035	5.659	0.039
6*	5.408	5.857	0.202	6.027	0.383	5.505	0.009
7	6.481	6.529	0.002	6.570	0.008	6.629	0.022
8	7.398	7.376	0.000	7.511	0.013	7.381	0.000
9	7.046	6.968	0.006	7.072	0.001	6.701	0.119
10	5.936	6.279	0.118	6.016	0.006	5.936	0.000
11	6.886	6.804	0.007	6.968	0.007	7.056	0.029
12	7.301	7.191	0.012	7.007	0.086	7.089	0.045
13*	6.060	6.482	0.178	6.545	0.235	7.018	0.918
14	6.252	6.229	0.001	6.261	0.000	6.273	0.000
15*	5.620	6.222	0.362	6.270	0.422	6.421	0.642
16	5.396	5.226	0.029	5.413	0.000	5.707	0.097
17	6.409	6.477	0.005	6.386	0.001	6.741	0.110
18	6.328	6.358	0.001	6.470	0.020	6.417	0.008
19	7.523	7.548	0.001	7.500	0.001	7.195	0.108
20	6.469	6.514	0.002	6.524	0.003	6.444	0.001
21	6.658	6.626	0.001	6.591	0.004	7.016	0.128
22*	5.921	6.116	0.038	6.199	0.077	6.275	0.125
23	6.071	6.138	0.004	5.851	0.048	6.018	0.003
24	6.638	6.633	0.000	6.643	0.000	6.602	0.001
25	3.013	2.979	0.001	2.838	0.031	3.035	0.000
26	3.363	3.203	0.026	3.243	0.014	3.223	0.020
27*	5.387	4.903	0.234	4.882	0.255	5.637	0.063
28	5.678	5.782	0.011	5.835	0.025	5.669	0.000
29	5.538	5.963	0.181	5.767	0.052	5.885	0.120
30	5.509	5.803	0.086	5.614	0.011	5.358	0.023
31	4.735	4.794	0.003	4.814	0.006	5.058	0.104
32	4.796	4.843	0.002	4.900	0.011	4.584	0.045
33	4.165	4.085	0.006	4.176	0.000	4.099	0.004
34	5.357	5.400	0.002	5.369	0.000	5.441	0.007
35	5.824	5.876	0.003	5.919	0.009	5.755	0.005
36	4.165	4.132	0.001	4.152	0.000	4.272	0.011
37	4.155	4.180	0.001	4.299	0.021	4.237	0.007
SD		0.090		0.115		0.149	

* Test set compounds

Fig. 4 Plots of predicted vs. observed values of (**a**) CoMFA, (**b**) RF-CoMFA, and (**c**) CoMSIA

compounds including training and test sets. The predicted r^2_{pred} values for QSAR models are shown in Tables 2, 3, 4. The final r^2_{pred} values are 0.905 for CoMFA, 0.926 for RF-CoMFA, and 0.878 for CoMSIA. These high r^2_{pred} scores imply that our 3D-QSAR models are highly predictable. Figure 4 shows correlation between predicted and observed activity values with these models (Table 6).

RF-CoMFA contour map

Figures 5 and 6 show 3D contour maps of RF-CoMFA steric and electrostatic interactions within the active site.

Validation of 3D QSAR models

Table 5 shows predicted pIC$_{50}$ values from different QSAR models (CoMFA, RF-CoMFA, and CoMSIA). To validate the predictability and accuracy of the models, predictive correlation coefficient r^2_{pred} was calculated for all
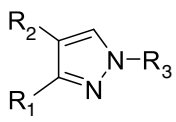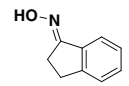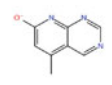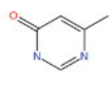
Table 6 The newly designed molecules by CombiGlide and their predicted activities by the RF-COMFA model

No		R1	R2	R3	pIC50
1	*	(bicyclic heteroaromatic)	dihydroindenoxime	4-piperidine	7.019
2		(methylpyrimidinone)	dihydroindenoxime	4-piperidine	7.0180
3		(methyl dioxo)	dihydroindenoxime	4-piperidine	6.803
4		(methylpyrimidinone)	dihydroindenoxime	4-piperidine	6.779
5		4-pyridyl	(purine)	4-piperidine	7.241
6		4-pyridyl	(bicyclic)	4-piperidine	6.477
7		4-pyridyl	dihydroindenoxime	(benzothiadiazine dioxide)	6.749
8		4-pyridyl	dihydroindenoxime	(methyl dioxo)	6.616
9		4-pyridyl	dihydroindenoxime	(quinoxalinedione)	6.247
10		4-pyridyl	dihydroindenoxime	(benzothiadiazine dioxide)	6.273
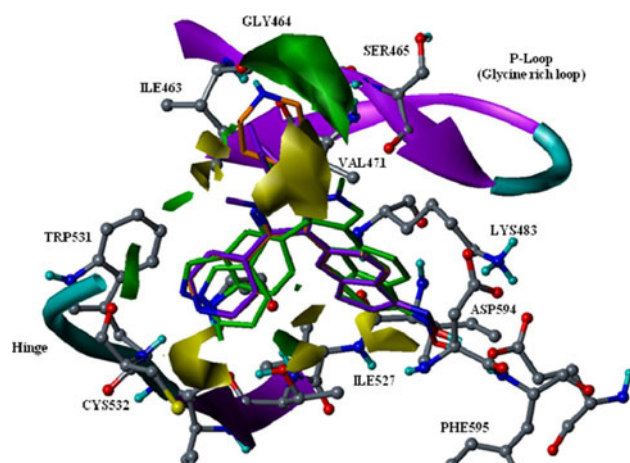
Fig. 5 Interacting residues of the active site and CoMFA steric contour maps with compound **3**, **5** (*purple*), **19** (*orange*), **25**, and **26** (*green*). *Green* contours indicate regions where bulky groups increase activity, whereas *yellow* contours indicate regions where bulky groups decrease activity
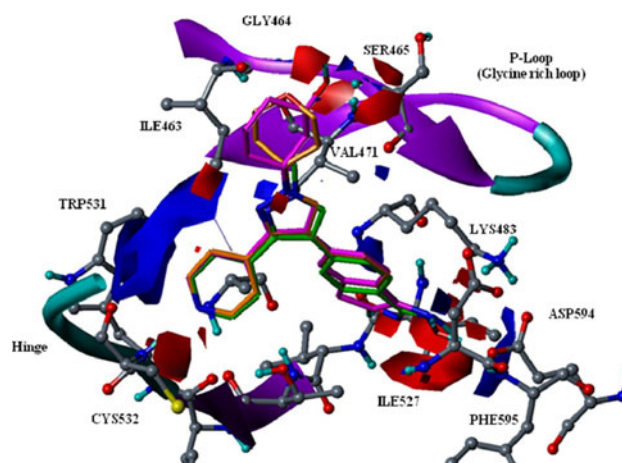
Fig. 6 Interacting residues of the active site and CoMFA steric contour maps with compound **8** (*magenta*), **21** (*orange*), and **31** (*green*). *Blue contours* indicate regions where electron accepting groups increase activity, whereas *red contours* indicate regions where electron donating groups increase activity

These contour maps represent physicochemical properties which effect biological activities. Figures 5 and 6 aid in understanding features of contour maps with respect to receptors. In Fig. 5, a large yellow (sterically unfavored) contour is found near R3 substitution, indicating that a sterically bulky group near the yellow region can decrease biological activity. In fact, biological activity of most compounds among 4–24 are <1 because their R3 substitutions are positioned inside of the yellow contour. However, the biological activities of 9, 12, and 19 whose R3 substitutions are sterically bulky groups are quite high compared with others and their R3 substitutions are oriented toward the green (sterically favored) contour rather than the yellow. This green contour region is in good agreement with the environment of the active site. The green region is surrounded by several hydrophobic residues (ILE463, GLY464, and VAL471). R3 substitutions of 9, 12 and 19 by an H-bond interaction group with the backbone of P-Loop as well as a sterically bulky group are appropriate. Figure 5 also explains why the biological activities of compounds 31–35, with heavier R2 groups, are lower than others. Sterically bulky R2 groups of 31–35 are located in the yellow contour region near ILE527, and this sterically unfavored matching can make H-bond interaction between the R2 group and ASP594 unstable.

In Fig. 6, small blue (positive charge favored) and red (negative charge favored) contour maps near GLY464 and SER465 on P-Loop indicate that H-bond possible groups are preferred as potent R3 substitutions. It is apparent that R3 substitutions of 9, 12, and 19 can satisfy steric and electrostatic conditions. An R3 substitution of 21 is also appropriate for the sterically favored region and H-bond interaction with carbonyl group SER465, as evidenced by the fact that its biological activity is higher than compounds 20–24.

In Fig. 5, there are several small yellow contours around R1 substitutions. RF-CoMFA steric contour map explains that 25's low biological activity results from its R1 substitution. R1 substitution of 25 is close to the yellow contours. In fact, the region of the small yellow contours is blocked by LEU514 and THR529, so R1 substitution of 25 can be sterically hindered by those residues. RF-CoMFA contour map shows another empty region which is favored for sterically bulky groups near TRP531. Sterically bulky groups, which can interact with TRP531 rather than LEU514 or THR529, can increase biological activity. In Fig. 6, the electrostatic contour map shows that blue and red contours and their positions are well matched with H-bond donating (NH) and accepting groups (CO) of the backbone of CYS532 on hinge. These blue and red contours around R1 explain why the 3-pyridyl group of compound 26 is not suitable for having H-bond interaction with CYS532's backbone. It is clear that RF-CoMFA contour maps can explain the different biological activities according to their substitutions. Also, RF-CoMFA suggests appropriate sterically bulky groups which can have H-bond interaction with CYS532 for R1 substitution and with GLY464 and SER465 for R3 substitution.

CoMSIA contour map

The CoMSIA steric and H-bond acceptor contour maps and the active site of B-Raf were superimposed with the most active compound, 19 (Fig. 7). The steric contour map from CoMSIA analysis (Fig. 7) is generally in agreement with the CoMFA steric contour map (Fig. 5): sterically favored region (green) is close to P-Loop and surrounded by three residues (ILE463, GLY464, and VAL471), and sterically unfavored region (yellow) is outside the P-Loop. As explained in CoMFA steric analysis, the CoMSIA steric contour map also indicates that a sterically bulky group is preferred for R3 substitution. On the H-bond acceptor contour map of CoMSIA (Fig. 7), H-bond acceptor favored regions (magenta) and unfavored region (red) are in accordance with negative charge favored (red) and positive charge favored (blue) regions of the CoMFA electrostatic contour map and properties of some residues in the active site (Fig. 6). The H-bond acceptor favored region (magenta) of CoMSIA was found near the amino group of SER465 in P-Loop where negative charge favored region (red) of CoMFA is positioned. This H-bond acceptor contour map also suggests that for R3 substitution, the H-bond acceptor group, which can interact with the amino group of SER465, can increase biological activity. CoMSIA H-bond acceptor favored (magenta) and unfavored
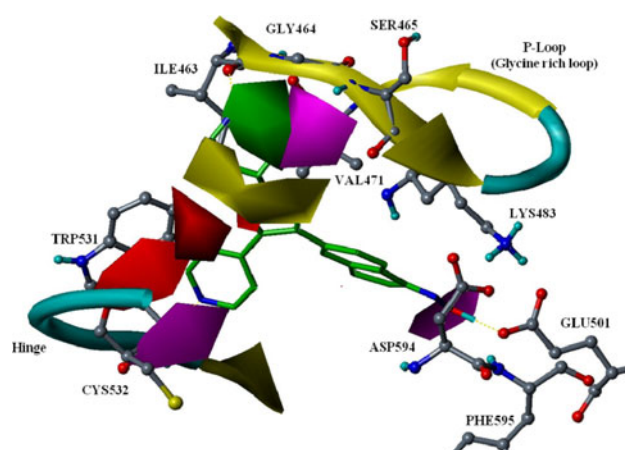
Fig. 7 CoMSIA steric and H-bnod acceptor contour maps with compound 18. *Green contours* indicate regions where bulky groups increase activity; *yellow contours* indicate regions where bulky groups decrease activity. *Magenta* indicates regions where hydrogen bond acceptor substituents enhance activity; *red* indicates hydrogen bond acceptor substituents reduce activity

(red) contours around R1 are well matched with atoms (NH and CO) of CYS532 as observed in the CoMFA electrostatic contour map around R1. The unfavored region (red) near CYS532's carbonyl group indicates that the H-bond donor group can increase the biological activity. Therefore, for R1 substitution, a group which can have H-bond interaction with both amino and carbonyl groups simultaneously is preferred to a 4-pyridyl group which can have H-bond interaction with only the amino group of CYS532.

Conclusion

The 3D-QSAR methods, CoMFA and CoMSIA were applied to a series of pyrazole-based derivatives as inhibitors of B-Raf kinase in an attempt to build a highly predictive model and find significant factors to aid in the design of new B-Raf inhibitors. In order to derive predictive 3D-QSAR models, several alignment methods were utilized including ligand-based, pharmacophore-based, ligand- & pharmacophore-based and two receptor-guided methods. Among these alignment methods, receptor-guided with QM/MM minimization produced the best predictive CoMFA model with sufficient statistic validation. Based on the same alignment method as in the CoMFA model, a CoMSIA model using steric, H-bond acceptor, and hydrophobic fields was also generated with acceptable statistics. Results of those 3D-QSAR models provide an important basis for future drug design of B-Raf inhibitors. Both CoMFA and CoMSIA models indicated that H-bond interactions with the backbone of CYS532 (hinge) around R1 are important to increase biological activity. At the same time, both models suggest that sterically bulky groups, which can have H-bond interaction with a NH backbone group of S465 (P-loop) would be a proper R1 substitution to increase biological activity.

References

1. Wellbrock C, Karasarides M, Marais R (2004) Nat Rev Mol Cell Biol 5:875–885
2. Avruch J, Khokhlatchev A, Kyriakis JM et al (2001) Recent Prog Horm Res 56:127–156
3. Davies H, Bignell GR, Cox C et al (2002) Nature 417:949–954
4. Tuveson DA, Weber BL, Herlyn M (2003) Cancer Cell 4:95–98
5. Houben R, Vetter-Kauczok CS, Ortmann S, Rapp UR, Broecker EB, Becker JC (2008) J Invest Dermatol 128:2003–2012
6. Hansen JD, Grina J, Newhouse B et al (2008) Bioorg Med Chem Lett 18:4692–4695
7. Noble MEM, Endicott JA, Johnson LN (2004) Science 303:1800–1805
8. Cramer R III, Patterson D, Bunce J (1988) J Am Chem Soc 110:5959–5967
9. Klebe G, Abraham U, Mietzner T (1994) J Med Chem 37:4130–4146
10. Klebe G, Abraham U (1999) J Comput Aided Mol Des 13:1–10
11. SYBYL 7.3 TI, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA
12. LigPrep v2.2, Schrödinger, LLC, New York, NY (2005)
13. Wold S, Sjöström M, Eriksson L (2001) Chemom Intell Lab Syst 58:109–130
14. Kettaneh N, Berglund A, Wold S (2005) Comput Stat Data Anal 48:69–85
15. Tetko IV, Tanchuk VY, Villa AEP (2001) J Chem Inf Comput Sci 41:1407–1421
16. Liao SY, Qian L, Miao TF, Lu HL, Zheng KC (2009) Eur J Med Chem 44:2822–2827
17. Chen Y, Li H, Tang W et al (2009) Eur J Med Chem 44:2868–2876
18. MacroModel v9.6, Schrödinger, LLC, New York, NY (2008)
19. Phase v3.0, Schrödinger, LLC, New York, NY (2008)
20. ConfGen v2.0, Schrödinger, LLC, New York, NY (2008)
21. Buolamwini JK, Assefa H (2002) J Med Chem 45:841–852
22. Patel MR, Dimmock JR, Talele TT (2007) J Chem Inf Model 47:2110–2123
23. Philipp DM, Friesner RA (1999) J Comput Chem 20:1468–1494
24. Cho AE, Guallar V, Berne BJ, Friesner R (2005) J Comput Chem 26:915–931
25. QSite v5.0, Schrödinger, LLC, New York, NY (2008)