JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

Development of Novel 3D-QSAR Combination Approach for Screening and Optimizing B-Raf Inhibitors in silico

Kuei-Chung Shih,[†] Chun-Yuan Lin,[*,‡,§] Jiayi Zhou,[†] Hsiao-Chieh Chi,[‡] Ting-Shou Chen,[∥] Chun-Chung Wang,[∥] Hsiang-Wen Tseng,[∥] and Chuan-Yi Tang[†,⊥]

[†]Department of Computer Science, National Tsing Hua University, Hsinchu 30013, Taiwan

[‡]Department of Computer Science and Information Engineering, Chang Gung University, Taoyuan 33302, Taiwan

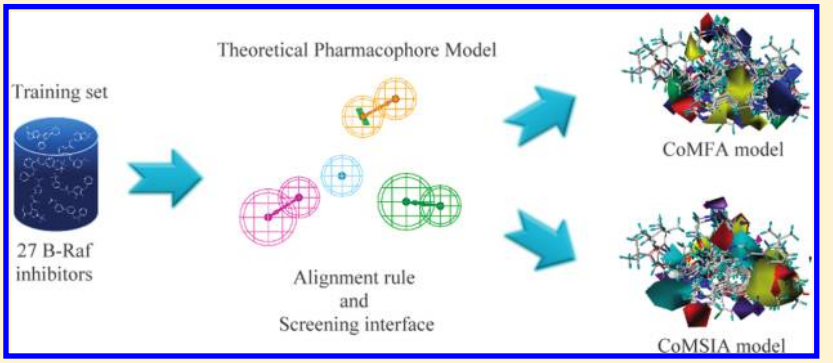
[§]Research Center for Emerging Viral Infections, Chang Gung University, Taoyuan 33302, Taiwan

[∥]Biomedical Engineering Research Laboratories, Industrial Technology Research Institute, Chutung, Hsinchu, 31040, Taiwan

[⊥]Department of Computer Science and Information Engineering, Providence University, Taichung 43301, Taiwan

Supporting Information

ABSTRACT: B-Raf is a member of the RAF family of serine/threonine kinases: it mediates cell division, differentiation, and apoptosis signals through the RAS-RAF-MAPK pathway. Thus, B-Raf is of keen interest in cancer therapy, such as melanoma. In this study, we propose the first combination approach to integrate the pharmacophore (PhModel), CoMFA, and CoMSIA models for B-Raf, and this approach could be used for screening and optimizing potential B-Raf inhibitors in silico. Ten PhModels were generated based on the HypoGen BEST algorithm with the flexible fit method and diverse inhibitor structures. Each PhModel was designated to the alignment rule and screening interface for CoMFA and CoMSIA models. Therefore, CoMFA and CoMSIA models could align and recognize diverse inhibitor structures. We used two quality validation methods to test the predication accuracy of these combination models. In the previously proposed combination approaches, they have a common factor in that the number of training set inhibitors is greater than that of testing set inhibitors. In our study, the 189 known diverse series B-Raf inhibitors, which are 7-fold the number of training set inhibitors, were used as a testing set in the partial least-squares validation. The best validation results were made by the CoMFA09 and CoMSIA09 models based on the Hypo09 alignment model. The predictive r^2_{pred} values of 0.56 and 0.56 were derived from the CoMFA09 and CoMSIA09 models, respectively. The CoMFA09 and CoMSIA09 models also had a satisfied predication accuracy of 77.78% and 80%, and the goodness of hit test score of 0.675 and 0.699, respectively. These results indicate that our combination approach could effectively identify diverse B-Raf inhibitors and predict the activity.



INTRODUCTION

There are three Raf isoforms in humans: A-Raf, B-Raf, and C-Raf (also called c-Raf-1).[1] Evidence showed that B-Raf exhibits higher basal kinase activity than A-Raf and C-Raf and must be phosphorylated to a greater extent within the N-terminal region of the kinase domain to achieve full cascade activation.[2] B-Raf plays a primary MEK activator in Ras-Raf-MEK-ERK pathway.[1] Depletion of B-Raf by RNA interference studies reduced phospho-ERK levels in melanoma cells.[3] The somatic mutation and activation of B-Raf was identified in melanoma (70%), papillary thyroid cancer (50%), and colon cancer (10%).[2] In melanoma, the most common B-Raf mutation is that glutamate substituted for valine at position 600 (V600E) within the activation domain.[4,5]

Because B-Raf mutations have been found in cancers, many small-molecule drugs, such as sorafenib, have been designed to inhibit both mutated and constitutive activity of B-Raf. Sorafenib is a multitarget kinase inhibitor, and the first RAF kinase inhibitor received clinical approval by the United States Food and Drug Administration in renal cell carcinoma and hepatocellular carcinoma.[4] Others, like RAF265,[6] PLX4032,[5,7] and XL281,[8] are small molecules that are orally bioavailable B-RAF kinase inhibitors in clinical trials.[4] Taken together, these findings indicate that design of the inhibitor of the kinase activity of B-Raf would greatly benefit cancer therapy.

Received: September 6, 2010
Published: December 23, 2010

The advantage of Pharmacophore[9−15] (PhModel) is its ability to easily and quickly identify candidate inhibitors for a target protein. When screening compound databases, each compound transformed the structure to 3D conformations for PhModel recognized candidate inhibitors. However, the PhModel could only find the candidate inhibitors fitting the binding site features of protein structure, which could not describe the 3D space limitation of the binding site. By contrast, the other two models, comparative molecular field analysis[16−19] (CoMFA) and comparative molecular similarity index analysis[17−22] (CoMSIA), were not suitable to search compound databases under the common structure limitation, but could easily be used to modify the molecule structure optimization and to delimit the molecular weight range of inhibitor. These two models were also useful for elucidating the chemical characteristics of locations in the 3D space. Therefore, according to the above individual characteristics, we prefer to combine these two different technologies.

We propose the first combination approach to integrate the PhModel, CoMFA, and CoMSIA models based on the same training set of B-Raf inhibitors. This approach could be used for screening and optimizing potential B-Raf inhibitors in silico. In our survey, several combination approaches were proposed on the basis of different pharmacophore methods, phase,[23,24] structure-based,[25,26] Hip-Hop,[27] and HypoGen FAST.[28] They have a common factor in that the number of training set inhibitors is greater than that of testing set inhibitors, and the maximum number of testing set inhibitors is <28.

In our study, 10 PhModels were generated on the basis of the HypoGen BEST algorithm with the flexible fit method and diverse inhibitor structures as the training set, and then set as alignment rules and screening interfaces for CoMFA and CoMSIA models. Therefore, the generated CoMFA and CoMSIA models could align and recognize diverse inhibitor structures. We used two quality validation methods to test the predication accuracy of these combination models. In the partial least-squares[17−19,21,22] (PLS) validation, we were using a large number of diverse series B-Raf inhibitors (189 testing set inhibitors) to test the predication accuracy of the proposed combination approach. In addition, the goodness of hit[29−32] (GH) test is a benchmark for assessing the efficacy of a screening database. We also used the GH test to enhance the confidence of statistical significance when screening compound databases.

The advantages of our combination approach follow: (i) CoMFA and CoMSIA models could provide 3D space delimitation of inhibitors. In general, the docking method is used to verify the 3D space limitation between inhibitors and the binding site after the PhModel screening. However, our combination approach does not need the docking method to confirm the 3D space limitation of the binding site. CoMFA and CoMSIA models could provide the contours to delimit the 3D molecular weight range to fit the space of the binding site. Besides, the docking method needs the protein crystal structure, and the computation time of CoMFA and CoMSIA models is less than that of the docking method. (ii) The activity prediction ability of CoMFA and CoMSIA models is better than that of PhModel. The PhModel could generate five features at maximum. The contours of CoMFA and CoMSIA models do not have this limitation. Therefore, the contours could reflect the features of active inhibitors as much as possible to provide satisfied prediction results. The inhibitor conformation fits more favored contours in the 3D space, which might increase the activity of the inhibitor. (iii) Inhibitors with diverse structures can be easily aligned into the CoMFA and CoMSIA models without the common structure limitation. Consequently, the CoMFA and CoMSIA models could screen compound databases via the PhModel based on our combination approach. Therefore, when searching compound databases, each compound was transformed to 3D conformations and then aligned by the PhModel. Finally, each compound from the aligned result could be optimized by the chemical characteristics and predicted activity based on CoMFA and CoMSIA models.

In this study, the purpose is to establish the first combination approach as the screening tool for the design or discovery of novel B-Raf inhibitor. We believe our approach can be utilized as follows: (1) for different applications within the drug design process, such as the virtual screening; (2) to perform a 3D query by using the PhModel to identify novel and diverse candidate inhibitors, and then predict the activity and verify the space limitation of protein structure binding site by CoMFA and CoMSIA models; (3) to serve as a structure optimization tool that directly modifies and optimizes the 3D structure of each candidate inhibitor before the synthesis of it.

MATERIALS AND 3D-QSAR MODEL METHODS

Designing a Combination Approach. Models for the approach, namely PhModel, CoMFA, and CoMSIA, were utilized in this study. The PhModel was implemented by the program Accelrys Discovery Studio 2.1, and CoMFA and CoMSIA models were implemented by the program Tripos Sybyl-X 1.0. We first established 10 PhModels and used them to align diverse inhibitor structures for generating the CoMFA and CoMSIA models. We then used the partial least-squares (PLS) method and known B-Raf inhibitors to validate the prediction ability of CoMFA and CoMSIA models. Finally, the goodness of hit (GH) test score was used as a benchmark for appraising the prediction ability of CoMFA and CoMSIA models to screen a compound database.

Biological Data Collection. The B-Raf kinase inhibitors were collected from the Binding Database.[33] After removing the duplicate and equivocal activity inhibitors, 216 inhibitors were ultimately selected for the generation and validation of the PhModel, CoMFA, and CoMSIA models. All inhibitor structures are shown in the Supporting Information (Table S1 and S2). These inhibitors were divided into training (27 inhibitors, Table S1) and testing (189 inhibitors, Table S2) sets. All 216 inhibitors were used to validate the screening database ability of models via the GH test. The activity of all inhibitors is represented as IC_{50} (nM), as divided into three groups: high activity (IC_{50} < 50 nM), moderate activity (100 nM ≤ IC_{50} ≤1000 nM), and inactivity (IC_{50} > 1000 nM).

Generating Ligand-Based Pharmacophore Models. PhModels were generated using the "3D-QSAR Pharmacophore Generation" protocol in the HypoGen algorithm.[34] Selecting the training data plays a crucial part in generating a good-quality PhModel and should meet the following requirements as suggested by the Accelrys Discovery Studio: (i) a minimum of 16 diverse compounds should be selected to ensure that the statistical significance can be assessed, (ii) the biological activity data associated with the compounds should span at least 4 orders of magnitude, (iii) clear and concise information on the selected compounds should be provided to avoid any redundancy or bias in terms of structural features and activity range, and (iv) the most active and inactive compounds should both be included in the training set. On the basis of these criteria above, the 27 B-Raf inhibitors were selected as the training

Table 1. Experimental and Estimated Activity (nM) Values of the Training Set Inhibitors Based on the PhModel, CoMFA, and CoMSIA Models

	PhModel			CoMFA and CoMSIA models		
no.	actual[a] IC_{50}	predicted[a] IC_{50}	error[b]	pIC_{50}	predicted CoMFA	predicted CoMSIA
1	0.09	0.0054	−17	10.046	9.713	9.433
2	0.22	0.79	+3.6	9.658	9.747	9.565
3	2.9	11	+3.9	8.538	8.62	8.823
4	3.5	11	+3.1	8.461	8.592	9.003
5	4.1	18	+4.4	8.387	8.393	8.692
6	4.4	27	+6	8.357	8.246	7.523
7	18	14	−1.3	7.735	7.766	7.958
8	26	540	+21	7.585	7.693	7.377
9	34	17	−2	7.469	7.587	7.639
10	40	570	+14	7.398	7.23	7.382
11	68	28	−2.5	7.165	7.221	6.917
12	650	2800	+4.2	6.187	6.279	6.276
13	790	1700	+2.2	6.102	6.182	6.12
14	870	2600	+3	6.061	6.212	6.293
15	1100	5800	+5.2	5.959	5.822	6.089
16	3800	13 000	+3.5	5.426	5.191	5.58
17	3800	5500	+1.4	5.420	5.312	4.958
18	4200	1900	−2.2	5.377	5.301	4.761
19	8000	8000	−1	5.097	5.288	5.194
20	10 000	16 000	+1.6	5	5.089	5.211
21	16 000	15 000	−1.1	4.796	4.632	4.515
22	58 000	13 000	−4.5	4.237	4.167	3.988
23	58 000	10 000	−5.6	4.237	4.303	4.128
24	65 000	4400	−15	4.187	4.29	4.132
25	99 000	10 000	−9.5	4.004	3.991	4.454
26	100 000	6600	−15	4	3.857	4.262
27	100 000	3500	−28	4	4.165	4.613

[a] Scale: highly active (IC50 < 50 nM), moderately active (100 nM ≤ IC_{50} ≤ 1000 nM), and inactive (IC50 > 1000 nM). [b] + indicates that the predicted IC_{50} is greater than the actual IC_{50}. − indicates that the predicted IC_{50} is lower than the actual IC_{50}.

set, and their structures were shown in Table S1. The biological activities of the training set inhibitors were shown in Table 1. To establish initial program settings, all training set conformations were generated by the "BEST" generation option based on CHARMm-like force field[35] parameters. Moreover, a maximum of 255 conformations was permitted for each inhibitor. In addition, a conformational space lies within a constraint of 20 kcal/mol energy, a threshold for the global minimum energy. Four features, hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic (HY), and ring aromatic (RA), were considered for the PhModel generation during the initial phase. All other parameters were set as default values. In total, 10 PhModels were generated in this study.

Inhibitor Alignment Model for the CoMFA and CoMSIA Models. The inhibitor alignment model for CoMFA and CoMSIA models was based on chemical characteristics of the ligand-based PhModel. The training set for establishing CoMFA and CoMSIA models was the same as that of the PhModel. The 255 3D conformations for each inhibitor were generated by the "BEST" generation option based on CHARMm-like force field parameters. All conformations were aligned to a 3D geometric model performed by the "Ligand Pharmacophore Mapping" protocol in Accelrys Discovery Studio. This aligned result was used to generate and analyze the CoMFA and CoMSIA models.

Generating CoMFA and CoMSIA Models. Gasteiger−Hückel charges were assigned to each of the 3D inhibitor structures mentioned above. The CoMFA model was performed on the basis of the Tripos Sybyl standard field and an sp^3-hybridized carbon atom with a +1 charge as a probe. In addition, the steric (Lennard-Jones potential) and electrostatic (Coulombic potential) fields were calculated on each grid point, and the energy cutoff value was set as 30 kcal/mol. The CoMFA and CoMSIA models share a common probe atom, aligned structure, and training set; however, the five fields, including steric, electrostatic, hydrophobic, hydrogen bond acceptor, and hydrogen bond donor, were all calculated on the basis of an attenuation factor value of 0.3 in the CoMSIA model. The distance dependence function between the probe and each inhibitor atom was employed via the Gaussian type of CoMSIA model.

■ QUALITY VALIDATION METHODS OF THE 3D-QSAR MODEL

We used two quality validation methods for the 3D-QSAR models: (i) PLS was used to assess the model quality and prediction ability, and (ii) the GH test was used to enhance the confidence of statistical significance when screening compound databases.

Table 2. The 10 PhModels Generated by the Training Set of B-Raf Inhibitors

hypo no.	total cost	cost-diff[a]	error cost	rms deviation	training set (r)	feature[b]
1	142.88	417.83	112.91	1.60	0.964	A,D,HY,RA
2	159.88	400.83	133.84	2.03	0.941	A,D,HY,RA
3	177.06	383.65	149.72	2.30	0.924	A,D,HY,RA
4	181.01	379.70	160.63	2.47	0.911	A,D,HY,RA
5	181.02	379.69	156.05	2.40	0.916	A,D,HY,RA
6	182.05	378.66	154.83	2.38	0.918	A,D,HY,RA
7	183.83	376.88	163.88	2.52	0.907	A,D,HY,RA
8	185.01	375.70	160.66	2.47	0.911	A,D,HY,RA
9	187.53	373.18	162.70	2.50	0.909	A,D,HY,RA
10	189.06	371.65	168.47	2.58	0.902	A,D,HY,RA

[a] (Null cost − total cost), null cost = 560.71, fixed cost = 94.15, configuration cost = 14.64. All costs are units of bits. [b] A, hydrogen bond acceptor; D, hydrogen bond donor; HY, hydrophobic; RA, ring aromatic.

PLS Analysis for CoMFA and CoMSIA Models. The PLS[17−19,21,22] method was used to analyze the CoMFA and CoMSIA models, which indicated a linear relationship between the activity and computed fields. The filtering column value was set as the default value, 2.0. We also used the leave-one-out method to determine the optimal number of components, and the coefficient r^2_{cv} (cv, cross-validation) was calculated by the cross-validation analysis. The following text describes the leave-one-out method. By removing one inhibitor from the training set and manipulating a model established by the remaining training set inhibitors, we could predict the activity of the removed inhibitor. When the coefficient r^2_{loo} (loo, leave-one-out) of the leave-one-out method was >0.5, however, the optimal number of components that was considered for the calculation of the coefficient r^2_{ncv} (ncv, noncross-validation) in the noncross-validation analysis and r^2_{cv} in the cross-validation analysis could be derived. As for the CoMFA and CoMSIA models applied here, it was to correctly classify those known activity inhibitors from a testing set to an activity or inactivity one. The predictive r^2 (r^2_{pred}) is an assessed value of CoMFA and CoMSIA models prediction ability. Equation 1 was used to calculate the r^2_{pred} based on the testing set

$$\text{predictive } r^2 = 1 - \frac{\text{PRESS}}{\text{SD}} \quad (1)$$

where SD is the sum of the squared deviations between the actual activity of the testing set and the mean activity of the training set, and PRESS is the sum of the squared deviations between the predicted and actual activity for each inhibitor in the testing set.

Goodness of Hit Test Score. The GH test method was applied to screen databases of known inhibitors to estimate the model quality in silico.[29−32] In this study, the CoMFA and CoMSIA models could be used to screen databases of B-Raf inhibitors with diverse structures via integrating PhModel features to align compound structures. Hence, the GH test could also be applied to estimate the quality of the CoMFA and CoMSIA models. The GH test score ranges from 0 to 1, where a score close to 0 indicates a null model and a score close to 1 indicates an ideal model. To calculate the GH test score, a set of parameters are needed, described as follows: D, the number of inhibitors in the database; A, the number of activity inhibitors in the database; H_t, the number of hit inhibitors in the hit list; H_a, the number of activity inhibitors in the hit list; $\%A$, the percentage ratio of activity inhibitors in the hit list; $\%Y$, the percentage ratio of yield rate of activity inhibitors; E, enrichment factor; and GH, value of the GH test. Equation 2 was used to calculate the GH test score to screen databases of known inhibitors.

$$\text{GH test score} = \left(\frac{H_a(3A + H_t)}{4H_tA}\right) \times \left(1 - \frac{H_t - H_a}{D - A}\right) \quad (2)$$

RESULTS

Pharmacophore Generation Results. Ten PhModels were generated based on the 27 training set inhibitors. Each PhModel included four features: hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic (HY), and ring aromatic (RA). The success of each PhModel was assessed on the basis of two important theoretical cost calculations (represented in bit units) when generating the PhModel. One was the fixed cost, which represents the simplest model that fits all data perfectly. The other one was the null cost, which represents the highest cost of a model with no correlation. The activity estimated by the null cost is the average activity of training set inhibitors. A PhModel was found to be significant once the difference between the null and fixed costs was large. Specifically, a good PhModel should conform to the following conditions. If the difference between the fixed and null costs is >70 bits, there is an excellent chance (>90%) that the experimental and predicted activity data will be positively correlated.[34] In our PhModel results, the null cost among those 10 PhModels was 560.71 bits, whereas the fixed cost was 95.93 bits. Thus, the difference between null and fixed costs was 466.56 bits. The configuration cost in our study was 16.44 bits, which was allowable because it was lower than the threshold value of 17 bits. On the basis of the criteria described above, the total cost of a good-quality PhModel should be close to the fixed cost and distant from the null cost; that means to be an excellent model. In our case, the total cost of the worst PhModel, Hypo10, was 189.06 bits. The cost difference of the Hypo10 model was >70 bits. The correlation coefficient r of 10 PhModels ranged from 0.964 to 0.902. Table 2 presents the information on the 10 PhModels.

Validating the PhModels by Fischer's Test Method. We applied the Fischer's randomization test (Fischer's test) to cross-validate the functions in the "3D-QSAR Pharmacophore Generation" protocol.[36] The Fischer's test was calculated by scrambling and then randomly reassigning the IC_{50} values to

dx.doi.org/10.1021/ci100351s |J. Chem. Inf. Model. 2011, 51, 398–407

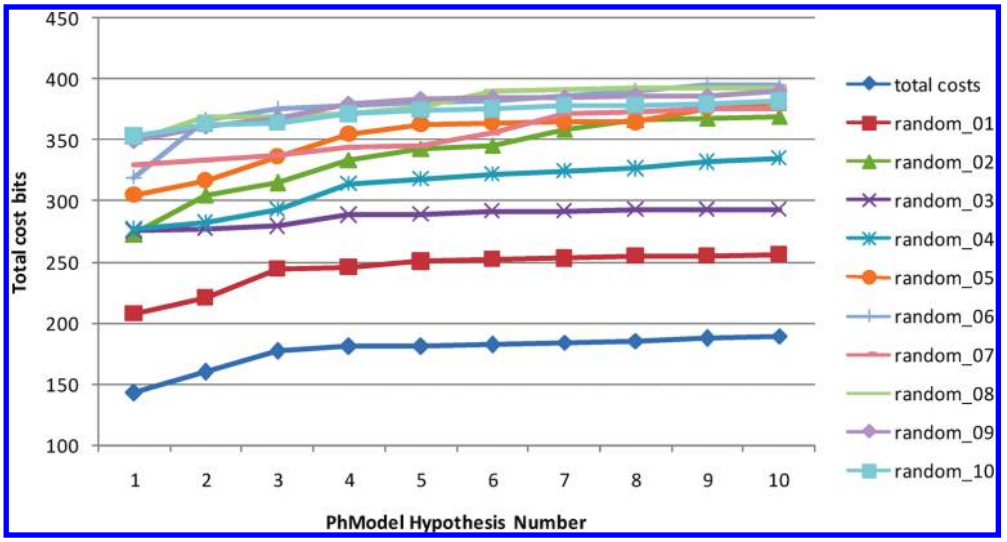


Figure 1. Total costs for the 10 original PhModels and for the 10 results that had the lowest total costs among the 49 randomization procedure results.

compounds in the training set. In the Fischer's test, program parameters of each randomization PhModel were kept with original PhModel generation, and this procedure was repeated 49 times to achieve a confidence level of 98% according to the following equation (eq 3, here x equals 49).

$$\text{confidence, } 98\% = [1-(1+0)/(x+1)] \times 100\% \quad (3)$$

Here, x is the number of times that the procedure was carried out.

The total costs of our 10 PhModels were all lower than that for any of 49 randomization procedures, implying that our PhModels were validated at the 98% confidence level. The 10 lowest total costs resulting from the randomization procedure are shown in Figure 1. It is clearly indicated that the total cost of each of our 10 PhModels was lower than that of 10 models generated by the randomization procedure. As mentioned above, this training set could not be selected using the randomization procedure. These results implied that the 10 PhModels could be used to formulate alignment rules for the generation of the CoMFA and CoMSIA models.

Generating CoMFA and CoMSIA Models. The 10 PhModels were subjected to alignment rules of the 10 CoMFA and 10 CoMSIA models, respectively. The training set of CoMFA and CoMSIA models was the same with the PhModel training set, which IC_{50} values ranged from 0.09 to 100 000 nM. In Table 1, these IC_{50} values were transformed to pIC_{50} values.

Validating CoMFA and CoMSIA Models by Partial Least Squares Analysis. On the basis of the foregoing PhModel results, each PhModel could serve as an alignment rule for the generation of the CoMFA and CoMSIA models, which comprised four features, namely hydrogen bond acceptor, hydrogen bond donor, hydrophobic, and ring aromatic, respectively. Table 3 reports the results of the 10 CoMFA and 10 CoMSIA models. One of the main validation methods was to calculate r^2_{pred}, which reflects the ability of a model to correctly distinguish activity or inactivity inhibitors. In the validation, the testing set utilized 189 known B-Raf inhibitors that were not used in any model generation. Table 3 also reports the validation results in terms of r^2_{pred}. Generally speaking, a good r^2_{pred} calculated from eq 1 should be >0.5.

The steric and electrostatic fields were both calculated in the CoMFA model. In the CoMFA model validation, CoMFA09 was the best one among the 10 CoMFA models, which had the highest r^2_{pred} of 0.56. The PLS analysis of the CoMFA09 model yielded an r^2_{cv} of 0.70, and the optimal number of components was five. In the non-cross-validation analysis, we obtained an r^2_{ncv} value of 0.99 with a standard error of estimate (SEE) of 0.51 and an F-ratio (F) of 757.53. The steric and electrostatic fields comprised 0.44 and 0.56 of the relative contributions, respectively.

The CoMSIA model consists of five fields, namely, steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor. The CoMSIA09 model had the highest r^2_{pred} of 0.56 in the validation procedure of 10 CoMSIA models. In the PLS analysis, the CoMSIA09 model yielded a cross-validated r^2_{cv} of 0.70 and three optimal components. In the non-cross-validated analysis procedure, the r^2_{ncv} value was 0.96 with a SEE of 0.38 and an F of 196.38. The steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor field contributions were 0.13, 0.24, 0.19, 0.18, and 0.26, respectively. In Figure 2a, the Hypo09 model contains one hydrogen bond acceptor feature, one hydrogen bond donor feature, one hydrophobic feature, and one ring aromatic feature. The training set was aligned on the Hypo09 model as shown in Figure 2b. Therefore, the CoMFA09 and CoMSIA09 models both had the best ability to predict the activity that is close to the actual activity based on the statistically validated evidence. The results indicate that the chemical features generated by the training set inhibitors are representative. The Hypo09 model provided the best alignment rule to generate the most efficacious CoMFA and CoMSIA models. The steric and electrostatic contour maps of the CoMFA model are shown in Figures 3 and 4, respectively. The hydrophobic, hydrogen bond donor, and hydrogen bond acceptor contour maps of the CoMSIA model are shown in Figures 5−7, respectively.

Validating the Ability of the Models to Screen Databases by GH Test. On the basis of the Hypo09 alignment rule, the CoMFA09 and CoMSIA09 models could be used to screen inhibitor databases. The main application of the GH test is to assess the success of a screening database model to correctly identify known highly active inhibitors. For a good screening model, the value of the GH test score must be ≥0.5. CoMFA09 and CoMSIA09 models were used to screen 216 of the known

Table 3. Summary of Analysis Results of the CoMFA and CoMSIA Models

	parameter						contributions				
	r^2_{cv}	NC	r^2_{ncv}	SEE	*F*-value	r^2_{pred}	S	E	H	D	A
CoMFA 01	0.76	3	0.95	0.42	158.31	0.39	0.41	0.59			
CoMFA 02	0.65	3	0.96	0.38	194.53	0.54	0.42	0.58			
CoMFA 03	0.74	4	0.99	0.25	352.55	0.29	0.45	0.55			
CoMFA 04	0.68	3	0.96	0.38	198.11	0.51	0.42	0.58			
CoMFA 05	0.74	3	0.96	0.41	169.92	0.48	0.45	0.55			
CoMFA 06	0.75	5	0.99	0.18	555.42	0.37	0.47	0.53			
CoMFA 07	0.69	2	0.92	0.53	146.35	0.27	0.37	0.63			
CoMFA 08	0.69	4	0.98	0.31	229.89	0.45	0.46	0.54			
CoMFA 09	0.70	5	0.99	0.51	757.53	0.56	0.44	0.56			
CoMFA 10	0.64	3	0.96	0.42	116.84	0.43	0.43	0.57			
CoMSIA 01	0.76	3	0.95	0.44	141.90	0.46	0.13	0.27	0.21	0.15	0.23
CoMSIA 02	0.77	3	0.97	0.36	251.49	0.55	0.12	0.24	0.21	0.19	0.24
CoMSIA 03	0.77	2	0.93	0.52	147.15	0.44	0.11	0.27	0.19	0.19	0.24
CoMSIA 04	0.72	3	0.97	0.33	254.71	0.45	0.09	0.24	0.23	0.19	0.25
CoMSIA 05	0.81	6	0.99	0.19	398.10	0.52	0.11	0.25	0.22	0.20	0.22
CoMSIA 06	0.79	4	0.98	0.28	272.58	0.45	0.11	0.22	0.22	0.26	0.20
CoMSIA 07	0.80	3	0.97	0.36	222.93	0.36	0.12	0.23	0.24	0.18	0.23
CoMSIA 08	0.71	2	0.91	0.59	114.98	0.52	0.12	0.26	0.20	0.17	0.25
CoMSIA 09	0.70	3	0.96	0.38	196.38	0.56	0.13	0.24	0.19	0.18	0.26
CoMSIA 10	0.76	2	0.93	0.52	151.38	0.47	0.11	0.26	0.21	0.16	0.26

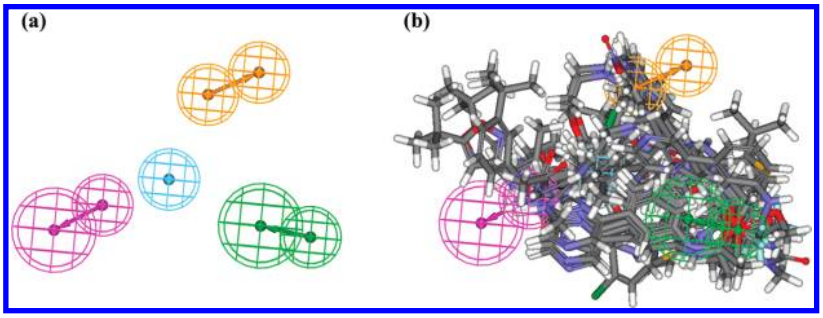


Figure 2. (a) Features of the Hypo09 model. (b) The training set of B-Raf inhibitors was aligned on the basis of the Hypo09 features for the generation of CoMFA and CoMSIA models.

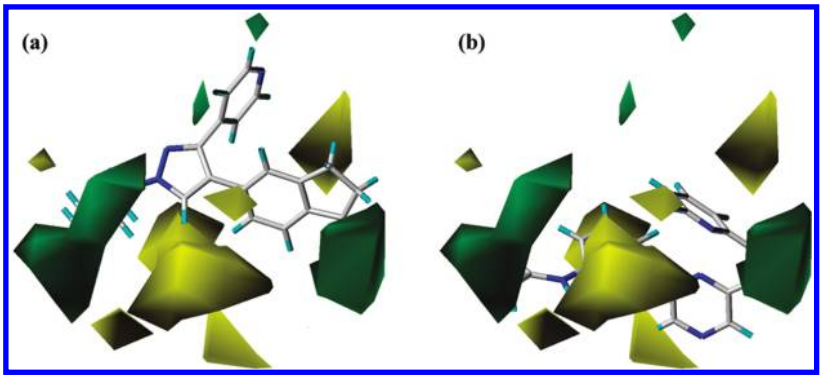


Figure 3. CoMFA steric STDEV*COEFF contour plots of (a) the most activity inhibitor and (b) the inactivity inhibitor aligned based on the Hypo09 model. Sterically favored regions (contribution level, 80%) are represented by green contours. Sterically disfavored regions (contribution level, 20%) are represented by yellow contours.

B-Raf inhibitors in the database. These inhibitors were classified into three groups: 63 highly active inhibitors (IC_{50} < 50 nM), 63 moderately active inhibitors (50 nM ≤ IC_{50} ≤ 1000 nM), and 90 inactive inhibitors (IC_{50} > 1000 nM). The "Ligand Pharmacophore

dx.doi.org/10.1021/ci100351s |*J. Chem. Inf. Model.* 2011, 51, 398–407

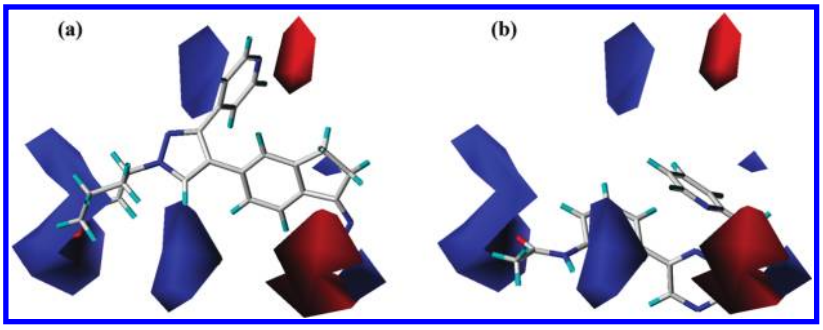


Figure 4. CoMFA electrostatic STDEV*COEFF contour plots of (a) the most activity inhibitor and (b) the inactivity inhibitor aligned based on the Hypo09 model. Positively charged favored regions (contribution level, 80%) are represented by blue contours. Negatively charged favored regions (contribution level, 20%) are represented by red contours.

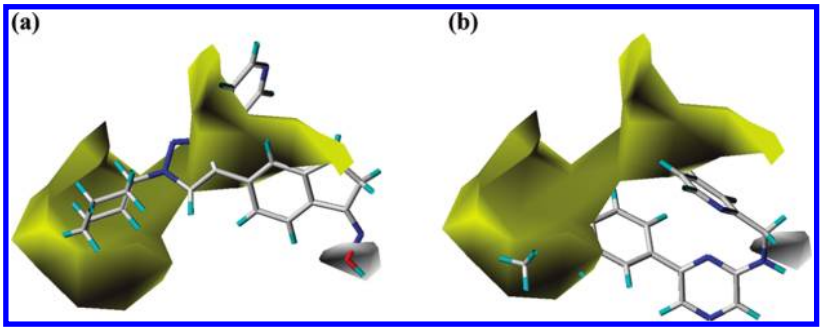


Figure 5. CoMSIA hydrophobic STDEV*COEFF contour plots of (a) the most activity inhibitor and (b) the inactivity inhibitor aligned based on the Hypo09 model. Hydrophobic favored regions (contribution level, 80%) are represented by yellow contours. Hydrophobic disfavored regions (contribution level, 20%) are represented by white contours.

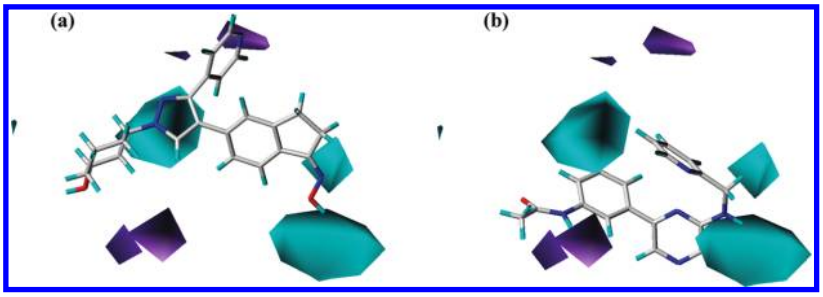


Figure 6. CoMSIA hydrogen bond donor STDEV*COEFF contour plots of (a) the most activity inhibitor and (b) the inactivity inhibitor aligned based on the Hypo09 model. Hydrogen bond donor substituent favored regions (contribution level, 80%) are represented by cyan contours. Hydrogen bond donor substituent disfavored regions (contribution level, 20%) are represented by purple contours.

Mapping" protocol in Accelrys Discovery Studio was applied to the screening and aligning inhibitor programs. Table 4 presents the GH test validation results for the CoMFA09 and CoMSIA09 models. For the CoMFA09 model, 45 B-Raf inhibitors were predicted to be highly active inhibitors, including 35 actual highly active inhibitors and 10 false positive inhibitors. The prediction accuracy was 77.78%, and the GH test score was 0.675. The CoMSIA09 model, with consideration of all five fields, had the best GH test validation result. The validation showed that 45 inhibitors were predicted to be highly active inhibitors, whereas only 36 of the 45 inhibitors actually were. Thus, 80% of the high activity inhibitors were effectively predicted by the CoMSIA09 model, and the GH test score was 0.699, implying that the prediction ability of the CoMSIA09 model was better than that of the CoMFA09 model. Collectively, the GH test results implied the best combination approach is the CoMSIA09 model based on the Hypo09 alignment rule. The approach could be applied to screen inhibitor databases, optimize inhibitor structures, and identify novel highly active B-Raf inhibitors.

Contour Map Analysis of CoMFA and CoMSIA Models. On the basis of the Hypo09 alignment result, each inhibitor was aligned on the features of Hypo09 instead of common structure of inhibitors. A contour map can vividly display the favored and disfavored regions for the inhibitor structure in the 3D space. The CoMFA and CoMSIA models could be used to easily describe the limit range of molecular weights. In addition, these models could be used to improve the compound optimization and identify potential analogues. Inhibitor 1 (most activity inhibitor) and inhibitor 27 (inactivity inhibitor) were used as template structures in Figures 3−7.

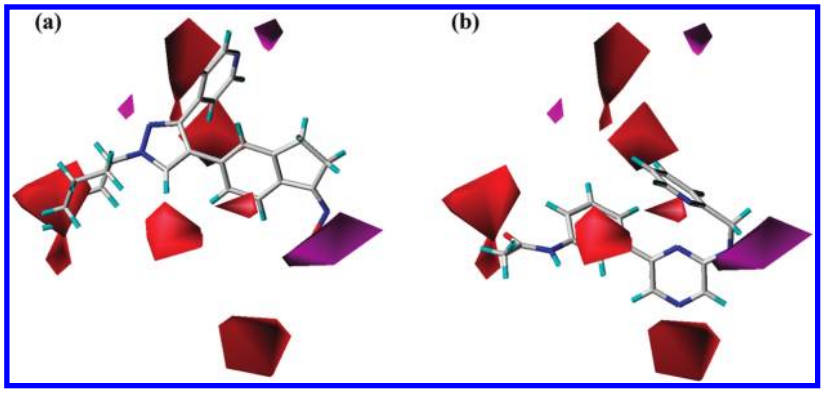


Figure 7. CoMSIA hydrogen bond acceptor STDEV*COEFF contour plots of (a) the most activity inhibitor and (b) the inactivity inhibitor aligned based on the Hypo09 model. Hydrogen bond acceptor substituent favored regions (contribution level, 80%) are represented by magenta contours. Hydrogen bond acceptor substituent disfavored regions (contribution level, 20%) are represented by red contours.

Table 4. Statistical Results of GH Test

serial no.	parameter	CoMFA09	CoMSIA09
1	total inhibitors in database (D)	216	216
2	total number of activity inhibitor in database (A)	63	63
3	total inhibitor hits (H_t)	45	45
4	activity inhibitor hits (H_a)	35	36
5	% yield of activity inhibitor hits $[(H_a/H_t) \times 100]$	77.78	80
6	% ratio of activity inhibitor hits $[(H_a/A) \times 100]$	55.56	57.14
7	enrichment factor (E) $[(H_a \times D)/(H_t \times A)]$	2.67	2.74
8	false negatives $[A - H_a]$	28	27
9	false positives $[H_t - H_a]$	10	9
10	goodness of hit test score[a] (GH)	0.675	0.699

[a] GH test score >0.6 indicates a very good model.

In Figure 3a,b, the green and yellow regions denote regions that are steric allowed and disallowed, respectively. Two main green contour regions are shown in Figure 3a, for which inhibitor 1 was used as the template structure. The first main green contour surrounds the cyclohexanol, indicating that this bulky group would improve the effectiveness of a B-Raf inhibitor. The formaldehyde oxime of 2,3-dihydro-1*H*-indene is oriented to insert into the second main green contour, implying that the steric freedom of this chemical moiety would likely increase the inhibitor activity. Therefore, the presence of these contours could potentially increase the potency of B-Raf inhibition. In Figure 3b, the structure of the inactivity inhibitor 27 has two parts that insert into the two main green contours. However, the aniline group of inhibitor 27 is oriented to insert into the yellow contour, leading to decreased activity of this inhibitor. If the inhibitor structure is too close or directly in contact with the yellow contours, there will be an overall decrease of the inhibitory activity.

Figure 4 presents the electrostatic contour map, which clearly shows the areas of positive and negative electrostatic distributions. The blue and red contours represent positive and negative favored regions, respectively. In Figure 4a, the blue and red contours are clearly distributed around inhibitor 1. Parts of inhibitor 27, however, are not entirely surrounded by blue and red contours, as shown in Figure 4b. These results indicate that the inhibitor conformation fits more favored contours in the 3D space, which might increase the activity of the inhibitor.

In the CoMSIA models, the hydrophobic favored and disfavored regions are clearly distributed. In Figure 5, the big yellow contour indicates the hydrophobic favored region, and the small white contour represents the disfavored region. In Figure 5a, most of the structure of inhibitor 1 is surrounded by the yellow contour, and only the formaldehyde oxime group contacts the white contour. As shown in Figure 5b, only the *N*-phenylacetamide group of inhibitor 27 is close to the yellow contour. According to the foregoing analysis, the hydrophobic favored contour might increase the B-Raf inhibitory activity.

In Figure 6, the cyan and purple regions represent the hydrogen bond donor favored and disfavored, respectively. The formaldehyde oxime group of inhibitor 1 is near the cyan contours as shown in Figure 6a. The hydrogen bond donor of inhibitor 27 does not point toward any cyan contours. Furthermore, the *N*-methylacetamide group of inhibitor 27 is oriented toward the purple contour areas (Figure 6b), leading to decreased activity of inhibitor.

In Figure 7, the hydrogen bond acceptor contours of the CoMSIA model are clearly distributed, and the favored and disfavored regions are shown by magenta and red, respectively. There is one main and two minor magenta contours. In Figure 7a, the formaldehyde oxime group of inhibitor 1 is enclosed in the main magenta contour. The nitrogen atoms of pyridine and pyrazole in inhibitor 1 are near the two minor magenta contours. In Figure 7b, the hydrogen bond acceptor group of pyrazine of inhibitor 27 is pointed at two red contours, and the formaldehyde group is pointed toward another red contour. The inhibitor 27, all hydrogen bond acceptor groups point toward disfavored contours. Therefore, the presence of hydrogen bond acceptor

dx.doi.org/10.1021/ci100351s |*J. Chem. Inf. Model.* 2011, 51, 398–407

contours on candidate inhibitors may suggest a relatively higher inhibitory activity toward B-Raf than others.

On the basis of the above contour analyses, if the inhibitor structure is more consistent with the favored contours in each field, then this may enhance B-Raf inhibitor activity. Using our combination approach, each inhibitor structure can be easily aligned into the CoMFA and CoMSIA models without any common structure of inhibitor.

CONCLUSIONS

In this study, we proposed a combination approach to integrate PhModel, CoMFA, and CoMSIA models based on the same training set of B-Raf inhibitors. This approach could be applied to screen inhibitor databases, optimize inhibitor structures, and identify novelty potent or specific B-Raf inhibitors. Theoretical PhModels were set as alignment rules and 3D query interfaces for CoMFA and CoMSIA models when screening compound databases. Therefore, using these alignment rules to generate CoMFA and CoMSIA models could align and recognize diverse inhibitor structures. Moreover, CoMFA and CoMSIA models were used to predict the activity and verify the space limitation of protein structure binding site. In our combination approach, each inhibitor with the diverse structure can be easily aligned into the CoMFA and CoMSIA models without any common structure. We were using a large number of diverse series B-Raf inhibitors to test the predication accuracy of our combination approach. In addition, GH test was used to enhance the confidence of statistical significance when screening compound databases. In the PLS validation, the testing set inhibitors we used are 7-fold the number of training set inhibitors, that can be used to validate the predication accuracy of our combination approach. We used the 189 known diverse series B-Raf inhibitors to validate all of CoMFA and CoMSIA models. The CoMFA09 with Hypo09 alignment model was the best one among the 10 CoMFA models, which had the highest r^2_{pred} value of 0.56. The CoMSIA09 with Hypo09 alignment model had the highest r^2_{pred} value of 0.56 in the validation procedure of 10 CoMSIA models. The PLS validation results indicate this approach could effectively predict the activity of a large number of diverse series B-Raf inhibitors. The results indicate that the chemical features generated by 27 inhibitors as the training set are representative. In the final GH test of the known B-Raf inhibitor database, the CoMFA09 model had a satisfied predication accuracy of 77.78% and a GH test score of 0.675; the CoMSIA09 model with all five fields had 80% predication accuracy and a GH test score of 0.699. On the basis of the statistically validated evidence, the GH test results point out that this approach could effectively identify high activity inhibitors when screening compound databases. The combination approach provides an easily and quickly screening workflow for the design or discovery of B-Raf inhibitors without the docking method. Our study provides a valuable tool for the discovery of new lead compounds with desired biological activities via the combination approach that could help medicinal chemists in their efforts to identify or design new B-Raf inhibitors.

ASSOCIATED CONTENT

Supporting Information. All inhibitor structures in Tables S1 and S2, as well as a training (Table S1) set and testing (Table S2) set of compounds with actual experimental and estimated activity IC50 (nM) values. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: 886-3-2118800 ext 3581; fax: 886-3-2118700; e-mail address: cyulin@mail.cgu.edu.tw.

ACKNOWLEDGMENT

Part of this work was supported by the National Science Council (NSC) with a grant of project contact NSC97-2221-E-182-033-MY3 and by Chang Gung University under Grant UERPD290101. We are grateful to the National Center for High-Performance Computing (NCHC) for computer time and facilities.

REFERENCES

(1) Mercer, K. E.; Pritchard, C. A. Raf proteins and cancer: B-Raf is identified as a mutational target. *Biochim. Biophys. Acta, Rev. Cancer* **2003,** *1653*, 25–40.

(2) Garnett, M. J.; Marais, R. Guilty as charged: B-RAF is a human oncogene. *Cancer Cell* **2004,** *6*, 313–9.

(3) Karasarides, M.; Chiloeches, A.; Hayward, R.; Niculescu-Duvaz, D.; Scanlon, I.; Friedlos, F.; Ogilvie, L.; Hedley, D.; Martin, J.; Marshall, C. J.; Springer, C. J.; Marais, R. B-RAF is a therapeutic target in melanoma. *Oncogene* **2004,** *23*, 6292–8.

(4) Montagut, C.; Settleman, J. Targeting the RAF-MEK-ERK pathway in cancer therapy. *Cancer Lett.* **2009,** *283*, 125–34.

(5) Yang, H.; Higgins, B.; Kolinsky, K.; Packman, K.; Go, Z.; Iyer, R.; Kolis, S.; Zhao, S.; Lee, R.; Grippo, J. F.; Schostack, K.; Simcox, M. E.; Heimbrook, D.; Bollag, G.; Su, F. RG7204 (PLX4032), a selective BRAFV600E inhibitor, displays potent antitumor activity in preclinical melanoma models. *Cancer Res.* **2010,** *70*, 5518–27.

(6) Roberts, P. J.; Der, C. J. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* **2007,** *26*, 3291–310.

(7) Rubinstein, J. C.; Sznol, M.; Pavlick, A. C.; Ariyan, S.; Cheng, E.; Bacchiocchi, A.; Kluger, H. M.; Narayan, D.; Halaban, R. Incidence of the V600K mutation among melanoma patients with BRAF mutations, and potential therapeutic response to the specific BRAF inhibitor PLX4032. *J. Transl. Med.* **2010,** *8*, 67.

(8) Schwartz, G. K.; Yazji, S.; Mendelson, D. S.; Dickson, M. A.; Johnston, S. H.; Wang, E. W.; Shannon, P.; Pace, L.; Gordon, M. S. A phase 1 study of XL281, a potent and selective inhibitor of RAF kinases, administered orally to patients (pts) with advanced solid tumors. *EJC Suppl.* **2008,** *6*, 120–120.

(9) Al-Masri, I. M.; Mohammad, M. K.; Taha, M. O. Discovery of DPP IV inhibitors by pharmacophore modeling and QSAR analysis followed by in silico screening. *ChemMedChem* **2008,** *3*, 1763–79.

(10) Chopra, M.; Gupta, R.; Gupta, S.; Saluja, D. Molecular modeling study on chemically diverse series of cyclooxygenase-2 selective inhibitors: generation of predictive pharmacophore model using catalyst. *J. Mol. Model.* **2008,** *14*, 1087–99.

(11) Kansal, N.; Silakari, O.; Ravikumar, M. Three dimensional pharmacophore modelling for c-Kit receptor tyrosine kinase inhibitors. *Eur. J. Med. Chem.* **2010,** *45*, 393–404.

(12) Lauria, A.; Ippolito, M.; Fazzari, M.; Tutone, M.; Di Blasi, F.; Mingoia, F.; Almerico, A. M. IKK-beta inhibitors: An analysis of drug-receptor interaction by using Molecular Docking and Pharmacophore 3D-QSAR approaches. *J. Mol. Graphics Modell.* **2010,** *29*, 72–81.

(13) Abu Hammad, A. M.; Taha, M. O. Pharmacophore modeling, quantitative structure-activity relationship analysis, and shape-complemented in silico screening allow access to novel influenza neuraminidase inhibitors. *J. Chem. Inf. Model.* **2009,** *49*, 978–96.

(14) Markt, P.; Feldmann, C.; Rollinger, J. M.; Raduner, S.; Schuster, D.; Kirchmair, J.; Distinto, S.; Spitzer, G. M.; Wolber, G.; Laggner, C.; Altmann, K. H.; Langer, T.; Gertsch, J. Discovery of novel CB2 receptor

ligands by a pharmacophore-based virtual screening workflow. *J. Med. Chem.* **2009,** *52,* 369–78.

(15) Taha, M. O.; Dahabiyeh, L. A.; Bustanji, Y.; Zalloum, H.; Saleh, S. Combining ligand-based pharmacophore modeling, quantitative structure-activity relationship analysis and in silico screening for the discovery of new potent hormone sensitive lipase inhibitors. *J. Med. Chem.* **2008,** *51,* 6478–94.

(16) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988,** *110,* 5959–67.

(17) Jiang, Y. K. Molecular docking and 3D-QSAR studies on beta-phenylalanine derivatives as dipeptidyl peptidase IV inhibitors. *J. Mol. Model.* **2010,** *16,* 1239–49.

(18) Khedkar, V. M.; Ambre, P. K.; Verma, J.; Shaikh, M. S.; Pissurlenkar, R. R.; Coutinho, E. C. Molecular docking and 3D-QSAR studies of HIV-1 protease inhibitors. *J. Mol. Model.* **2010,** *16,* 1251–68.

(19) Lei, B. L.; Du, J.; Li, S. Y.; Liu, H. X.; Ren, Y. Y.; Yao, X. J. Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) of thiazolone derivatives as hepatitis C virus NS5B polymerase allosteric inhibitors. *J. Comput.-Aided Mol. Des.* **2008,** *22,* 711–725.

(20) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994,** *37,* 4130–46.

(21) Lu, P.; Wei, X.; Zhang, R. CoMFA and CoMSIA studies on HIV-1 attachment inhibitors. *Eur. J. Med. Chem.* **2010,** *45,* 1792–8.

(22) Zeng, J.; Liu, G. X.; Tang, Y.; Jiang, H. L. 3D-QSAR studies on fluoropyrrolidine amides as dipeptidyl peptidase IV inhibitors by CoMFA and CoMSIA. *J. Mol. Model.* **2007,** *13,* 993–1000.

(23) Murumkar, P. R.; Zambre, V. P.; Yadav, M. R. Development of predictive pharmacophore model for in silico screening, and 3D QSAR CoMFA and CoMSIA studies for lead optimization, for designing of potent tumor necrosis factor alpha converting enzyme inhibitors. *J. Comput.-Aided Mol. Des.* **2010,** *24,* 143–56.

(24) Pasha, F. A.; Muddassar, M.; Neaz, M. M.; Cho, S. J. Pharmacophore and docking-based combined in-silico study of KDR inhibitors. *J. Mol. Graphics Modell.* **2009,** *28,* 54–61.

(25) Chen, Y.; Li, H.; Tang, W.; Zhu, C.; Jiang, Y.; Zou, J.; Yu, Q.; You, Q. 3D-QSAR studies of HDACs inhibitors using pharmacophore-based alignment. *Eur. J. Med. Chem.* **2009,** *44,* 2868–76.

(26) Lu, X. Y.; Chen, Y. D.; You, Q. D. 3D-QSAR studies of arylcarboxamides with inhibitory activity on InhA using pharmacophore-based alignment. *Chem. Biol. Drug Des.* **2010,** *75,* 195–203.

(27) Chaudhaery, S. S.; Roy, K. K.; Saxena, A. K. Consensus superiority of the pharmacophore-based alignment, over maximum common substructure (MCS): 3D-QSAR studies on carbamates as acetylcholinesterase inhibitors. *J. Chem. Inf. Model.* **2009,** *49,* 1590–601.

(28) Palomer, A.; Pascual, J.; Cabre, F.; Garcia, M. L.; Mauleon, D. Derivation of pharmacophore and CoMFA models for leukotriene D(4) receptor antagonists of the quinolinyl(bridged)aryl series. *J. Med. Chem.* **2000,** *43,* 392–400.

(29) Boppana, K.; Dubey, P. K.; Jagarlapudi, S. A.; Vadivelan, S.; Rambabu, G. Knowledge based identification of MAO-B selective inhibitors using pharmacophore and structure based virtual screening models. *Eur. J. Med. Chem.* **2009,** *44,* 3584–90.

(30) Osman, F. G. D., Henry, R. Metric for analyzing hit lists and pharmacophores. In. *Phamacophore Perception, Development, and Use in Drug Design*; Guner., O. F., Ed.; International University Line: La Jolla, CA, 2000; pp 193−210.

(31) Ravikumar, M.; Pavan, S.; Bairy, S.; Pramod, A. B.; Sumakanth, M.; Kishore, M.; Sumithra, T. Virtual screening of cathepsin k inhibitors using docking and pharmacophore models. *Chem. Biol. Drug Des.* **2008,** *72,* 79–90.

(32) Vadivelan, S.; Sinha, B. N.; Irudayam, S. J.; Jagarlapudi, S. A. Virtual screening studies to design potent CDK2-cyclin A inhibitors. *J. Chem. Inf. Model.* **2007,** *47,* 1526–35.

(33) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **2007,** *35,* D198–201.

(34) Sprague, P. W. Automated chemical hypothesis generation and database searching with Catalyst(R). *Perspect. Drug Discovery Des.* **1995,** *3,* 1–20.

(35) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983,** *4,* 187–217.

(36) *Discovery Studio*, version 2.1; Accelrys: San Diego, CA, 2008.