



## Original article

## Molecular modeling studies on imidazo[4,5-b]pyridine derivatives as Aurora A kinase inhibitors using 3D-QSAR and docking approaches

Ping Lan, Wan-Na Chen, Wei-Min Chen\*

Guangdong Province Key Laboratory of Pharmacodynamic Constituents of TCM and New Drugs Research, College of Pharmacy, Jinan University, Guangzhou 510632, PR China

## ARTICLE INFO

## Article history:

Received 16 August 2010

Received in revised form

15 October 2010

Accepted 18 October 2010

Available online 26 October 2010

## Keywords:

Imidazo[4,5-b]pyridine

Aurora A kinase

3D-QSAR

Docking

## ABSTRACT

3D-QSAR and docking studies were performed on sixty imidazo[4,5-b]pyridine derivatives as Aurora A kinase inhibitors. The CoMFA and CoMSIA models using forty-eight molecules in the training set, gave  $r^2_{cv}$  values of 0.774 and 0.800,  $r^2$  values of 0.975 and 0.977, respectively. The external validation indicated that both CoMFA and CoMSIA models possessed high predictive powers with  $r^2_{pred}$  values of 0.933 and 0.959,  $r^2_m$  values of 0.883 and 0.915, respectively. 3D contour maps generated from the two models along with docking binding structures have identified several key structural requirements responsible for the activity. A set of thirty new analogues were proposed by utilizing the results revealed in the present study, and were predicted with significantly improved potencies in the developed models.

© 2010 Elsevier Masson SAS. All rights reserved.

## 1. Introduction

The mammalian Aurora kinases comprise a family of serine/threonine kinases that are essential for coordinated mitotic progression [1]. Three members of the Aurora kinase family have been identified up to now, namely, Aurora A, B and C [2]. The biological roles of Aurora A and B are identified, and the role of Aurora C is still unclear. The expression and activity of Aurora kinases are tightly associated with cell cycle [3]. The two major Aurora kinases, Aurora A and B, have distinct roles in mitosis. Along with its cellular binding partner TPX2 (Target Protein of *Xenopus* kinesin-like protein 2), Aurora A plays an essential role in mitotic spindle formation, centrosome maturation, and segregation; while Aurora B is involved in chromosomal condensation, alignment and separation as well as kinetochore-microtubule attachment and cytokinesis with its binding partners INCENP (Inner Centromere Protein) [3–7]. The Aurora C is required for spermatogenesis, it may complement the functions of Aurora B, but its function in cell cycle progression is still unknown [8,9].

Previous researches indicated that Aurora A and B as attractive targets for pharmacological intervention in oncology because, for instance, both of them, particularly Aurora A, were found to be over-expressed in a wide range of cancer, including colorectal, prostate, ovarian, breast, glioma and other cancers [10–12]. The

initial motivation to direct drug discovery efforts to Aurora A arose from early findings that Aurora A over-expression leads to increased degradation of the natural tumor suppressor p53, inhibition of Aurora A causes defects in centrosome separation, with the formation of characteristic monopolar spindles [13]. In addition, ectopic over-expression of Aurora A can transform the cells. Aurora A over-expression has also been shown to be able to induce resistance to tubulin-disrupting agents, such as paclitaxel, by overriding the mitotic spindle checkpoint [8]. Since Aurora A is up-regulated and amplified largely in many cancer cell lines compared to normal cells, it is gaining immense importance as a potential therapeutic target for cancer [13–16].

Several structurally diverse inhibitors of Aurora kinases with excellent anti-tumor activity have been identified, some of which have reached clinical evaluation, such as VX-680, PHA-739358, AT9283, MNL8054 (Fig. 1) [7,17–19]. Recently, a series of imidazo[4,5-b]pyridine derivatives as orally bioavailable Aurora A inhibitors with excellent potencies were reported [15,16]. In order to investigate the key structural features required to design potential Aurora A inhibitor drug candidates from imidazo[4,5-b]pyridine derivatives, we have performed 3D-QSAR and docking analyses on a set of sixty derivatives.

Structure- and ligand-based drug design approaches have become fundamental components of modern drug discovery [20,21]. During the past decades, QSAR methods especially 3D-QSAR approaches, have been successfully employed to assist the design of new drug candidates, ranging from enzyme inhibitors to

\* Corresponding author. Tel.: +86 2085224497; fax: +86 2085224766.

E-mail address: [twmchen@jnu.edu.cn](mailto:twmchen@jnu.edu.cn) (W.-M. Chen).

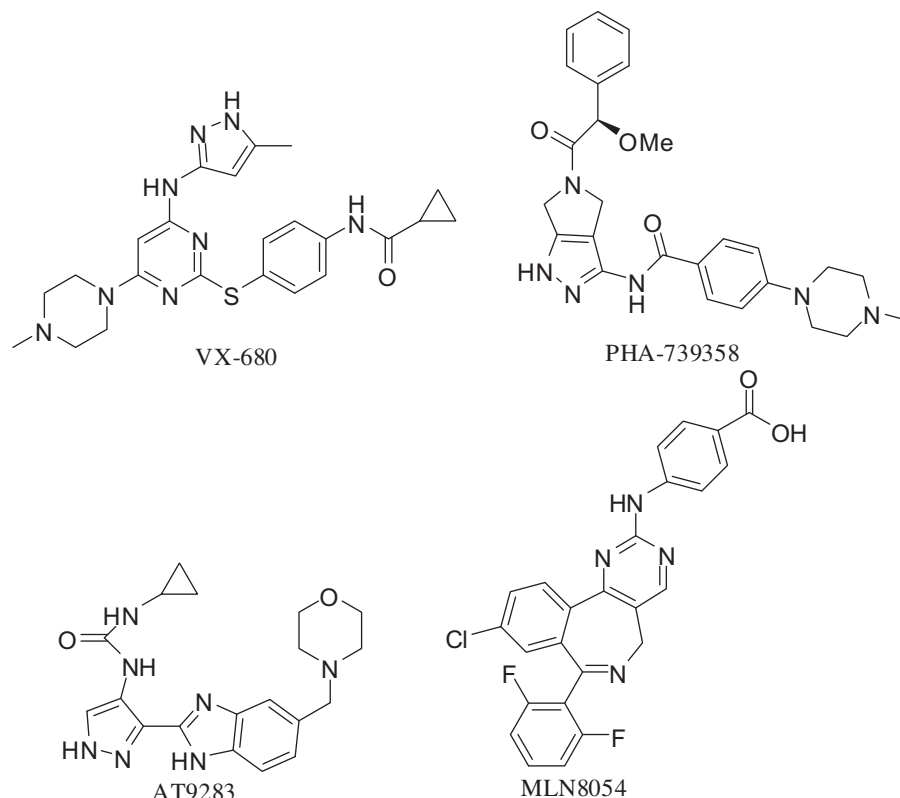


Fig. 1. Aurora kinases inhibitors currently under clinical evaluation.

receptor ligands [22–24]. Furthermore, they have been extensively applied in connection to medicinal chemistry research as well as proteomics, metabolomics, and bioinformatics [25–27]. 3D-QSAR methods including comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were performed to predict the activities of these inhibitors and provided the regions in space where interactive fields may influence the activity. In CoMFA, the biological activity of molecules is correlated with their steric and electrostatic interaction energies. In CoMSIA, similarity indices are calculated at regularly placed grid points for these molecules. CoMSIA includes five molecular descriptors named steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields. To evaluate the true predictive power of the 3D-QSAR models, a systemic external validation was employed. Docking was applied to investigate the Aurora A inhibitor interactions. Based on the good performance of the 3D-QSAR and docking studies, the developed models can not only help in understanding the structure–activity relationship of these compounds but also be served as a useful guide for the design of new inhibitors with better activities. We have designed a number of novel imidazo[4,5-b]pyridine derivatives by utilizing the structure analysis results obtained from present studies, which exhibited excellent predicted activities in the CoMFA and CoMSIA models established. Meanwhile, based on the excellent performance of the external validation, the predicted activities of these newly designed derivatives would be reliable.

## 2. Results and discussions

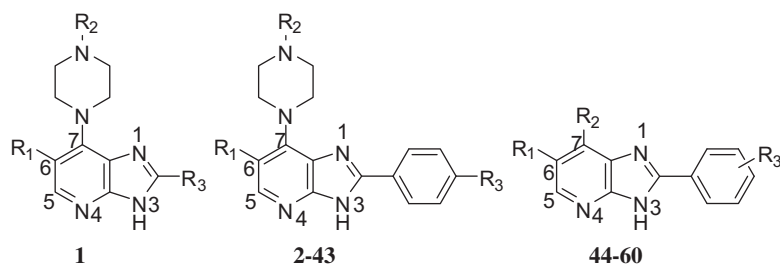
### 2.1. PLS analysis for CoMFA and CoMSIA models

A total set of 60 imidazo[4,5-b]pyridine derivatives were available for 3D-QSAR analysis, all the compounds were randomly

segregated into training and test sets comprising 48 and 12 molecules, respectively. Structures and associated inhibitory activities were listed in Table 1 and Table 2, and the aligned derivatives used in the training set were shown in Fig. 2.

The statistical parameters associated in CoMFA model were listed in Table 3. The CoMFA model gave a good cross-validated correlation coefficient ( $r_{cv}^2$ ) of 0.774 ( $>0.5$ ) with an optimized component of 6, which suggesting that the model should be a useful tool for predicting the  $IC_{50}$  values. A high non-cross-validated correlation coefficient ( $r^2$ ) of 0.975 with a low standard error estimate (SEE) of 0.195, and excellent  $F$  value of 268.426 were obtained. Contributions of steric and electrostatic fields were 0.481 and 0.519, respectively. The actual and predicted  $pIC_{50}$  values of the training set and test set by the CoMFA model were given in Table 2, and the graph of actual activity versus predicted  $pIC_{50}$  of the training set and test set was illustrated in Fig. 3.

The CoMSIA model incorporates steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D) and hydrogen bond acceptor (A) fields. 3D-QSAR models can be generated using these fields in different combinations. The PLS results of CoMSIA analysis using different combinations were depicted in Table 4 and Fig. 5. In most of the combination models, steric, electrostatic and hydrogen bond acceptor fields played important roles for the present series of molecules. The model generated using steric, electrostatic and hydrophobic descriptors (S + E + H) had highest  $r_{cv}^2$ , but showed relative lower  $r^2$  and  $F$  values compared to combinations S + E + H + D + A and S + E + H + D. Meanwhile, the CoMSIA model using steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields (S + E + H + D + A) exhibited better SEE and  $F$  values than that of combination (S + E + H + D). Finally, the combination of steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields was selected as the best model.

**Table 1**Chemical structures and IC<sub>50</sub> values of the training and test set molecules.

Compd. No.	Substituent			IC <sub>50</sub> (μM)
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	
<b>1</b>	Br		H	0.517
<b>2</b>	Br		NMe <sub>2</sub>	0.055
<b>3</b>	CN		NMe <sub>2</sub>	0.050
<b>4</b>	Cyclopropyl		OMe	0.053
<b>5</b>	Cl		NMe <sub>2</sub>	0.042
<b>6</b>	Br			0.016
<b>7</b>	Br		H	0.081
<b>8</b>	Cl		OMe	0.052
<b>9</b>	Cl			0.003

(continued on next page)

Table 1 (continued)

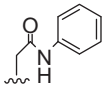
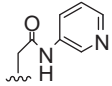
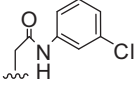
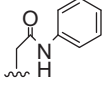
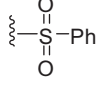

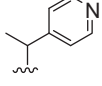
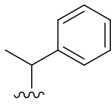
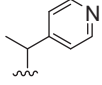
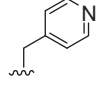
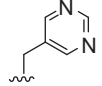
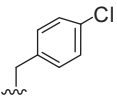

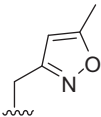
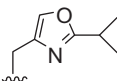



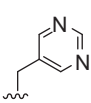
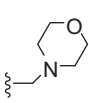
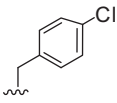
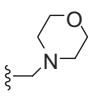
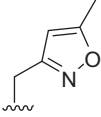
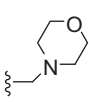




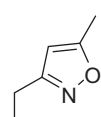
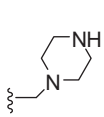
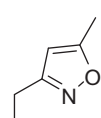
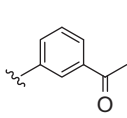
Compd. No.	Substituent			IC <sub>50</sub> (μM)
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	
10	Cl		NMe <sub>2</sub>	0.115
11	Br		NMe <sub>2</sub>	0.118
12	Cl		NMe <sub>2</sub>	0.075
13	Br		NMe <sub>2</sub>	0.277
14	Br		NMe <sub>2</sub>	0.178
15	Br	Ph	NMe <sub>2</sub>	0.258
16	Br		NMe <sub>2</sub>	0.210
17	Br		NMe <sub>2</sub>	0.085
18	Cl		NMe <sub>2</sub>	0.079
19	Br		NMe <sub>2</sub>	0.055
20	Br		OMe	0.009
21	Br		OMe	0.008
22	Br		OMe	0.040

Table 1 (continued)

Compd. No.	Substituent			IC <sub>50</sub> (μM)
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	
23	Br		OMe	0.65
24	Br		OMe	0.017
25	Br		OMe	0.081
26	Br		OMe	0.032
27	Br			0.005
28	Br			0.006
29	Br			0.012
30	Br			0.002
31	Br			0.030
32	Br			0.006
33	Br			0.010
34	Br			0.014

(continued on next page)

Table 1 (continued)

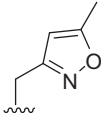
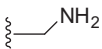
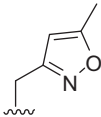
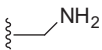
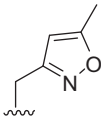
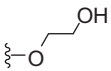
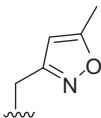
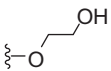
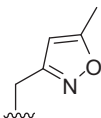

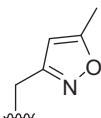
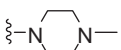
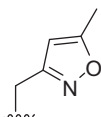
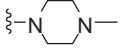
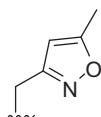

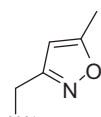

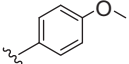
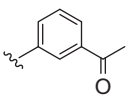
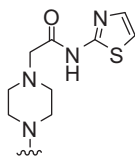
Compd. No.	Substituent			IC <sub>50</sub> (μM)
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	
35	Br			0.010
36	Cl			0.055
37	Br			0.030
38	Cl			0.029
39	Br			0.010
40	Br			0.015
41	Cl			0.010
42	Br			0.010
43	Cl			0.003
44	H	H	<i>p</i> -NMe <sub>2</sub>	4.3
45	H	H	<i>m</i> -NMe <sub>2</sub>	10.0
46	H	H	<i>p</i> -OMe	6.6
47	H	H	<i>m</i> -OMe	16.0
48	H	H	<i>o</i> -OMe	29
49	H	H	<i>p</i> -Pyrrolidin-1-yl	4.6
50	H	H	<i>p</i> -Pyrid-2-yl	4.4
51	Cl	H	<i>p</i> -NMe <sub>2</sub>	0.57
52	H	H	<i>p</i> -NMe <sub>2</sub>	4.3
53	Br	H	<i>p</i> -NMe <sub>2</sub>	0.49
54	Me	H	<i>p</i> -NMe <sub>2</sub>	6.9

Table 1 (continued)

Compd. No.	Substituent			IC <sub>50</sub> (μM)
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	
55		H	<i>p</i> -NMe <sub>2</sub>	20
56		H	<i>p</i> -NMe <sub>2</sub>	9.0
57	Cl	Cl	<i>p</i> -NMe <sub>2</sub>	0.25
58	H	Me	<i>p</i> -NMe <sub>2</sub>	7.0
59	H	Cl	<i>p</i> -NMe <sub>2</sub>	2.4
60	H		<i>p</i> -NMe <sub>2</sub>	0.87

The CoMSIA model gave a cross-validated correlation coefficient ( $r_{cv}^2$ ) of 0.800 ( $>0.5$ ) with an optimized component of 6. A high non-cross-validated correlation coefficient ( $r^2$ ) of 0.977 with a standard error estimate (SEE) of 0.188, and  $F$  value of 290.720 was obtained. Contributions of steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields were 0.191, 0.297, 0.210, 0.089 and 0.213, respectively. The actual and predicted pIC<sub>50</sub> values and residual values for training set and test set compounds by the CoMSIA model were given in Table 2. The relationship between actual and predicted pIC<sub>50</sub> of the training set and test set compounds was illustrated in Fig. 4.

## 2.2. External validation analysis for CoMFA and CoMSIA models

Previous researches indicated that a high cross-validated correlation coefficient ( $r_{cv}^2$ ) may be the necessary condition for a 3D-QSAR model to have a high predictive power. However, it is not a sufficient condition. In fact, the low values of  $r_{cv}^2$  and  $r^2$  can serve as an indicator of a low predictive ability of a model, nevertheless, the opposite is not necessarily true [28]. In many cases, a model with high  $r_{cv}^2$  and  $r^2$  values can be proved to be inaccurate. Even though a model may exhibit a good predictive ability based on the statistics for the test set, it is not always sure that the model will perform well on a new set of data [29]. The only way to estimate the true predictive power of a model is to test it on an external validation. To evaluate the true predictive abilities of the established models, both the CoMFA and CoMSIA models were subjected to systemic external validation process, several statistics such as  $r_{pred}^2$ ,  $r_m^2$ ,  $r_0^2$ ,  $R$ ,  $a$ ,  $b$  and  $k$  were employed. For the ideal model, the slope  $a$  is equal to 1, intercept  $b$  is equal to 0, and correlation coefficient  $R$  is equal to 1. 3D-QSAR models were considered acceptable if they satisfy all of the following conditions:  $r_{cv}^2 > 0.5$ ,  $r^2 > 0.6$ ,  $[(r^2 - r_0^2)/r^2] < 0.1$ ,  $0.85 \leq k \leq 1.15$  and  $r_m^2 > 0.5$ .

The results of the external validation for the CoMFA model were shown in Table 5. The established CoMFA model using 12 molecules in the test set, gave a predictive correlation coefficient ( $r_{pred}^2$ ) of 0.933, slope  $a$  value of 1.034 (close to 1), intercept  $b$  value of -0.281 (close to 0), an excellent  $r_m^2$  value of 0.883 ( $>0.5$ ) as well as high slope of regression lines through the origin ( $k$ ) value of 0.984

( $0.85 \leq k \leq 1.15$ ), and the correlation coefficient ( $R$ ) values of 0.971 (close to 1), the calculated  $[(r^2 - r_0^2)/r^2]$  values of -0.009 ( $<0.1$ ) were obtained. These excellent external validation statistics indicated that the CoMFA model possessed a high accommodating capacity, and it would be reliable for being used to predict the activities of new derivatives.

The external validation results of CoMSIA model using different combinations were depicted in Table 6 and Fig. 5. Both the (S + E + H + D + A) and (S + E + H + D) combinations exhibited excellent results in the external validation, however, the (S + E + H + D + A) was slightly better than (S + E + H + D), therefore, the combination of steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields was selected as the best model.

The best CoMSIA model also using 12 molecules in the test set, gave a predictive correlation coefficient ( $r_{pred}^2$ ) of 0.959, slope  $a$  value of 0.936 (close to 1), intercept  $b$  value of 0.352 (close to 0), an excellent  $r_m^2$  value of 0.915 ( $>0.5$ ) as well as high slope of regression lines through the origin ( $k$ ) value of 0.981 ( $0.85 \leq k \leq 1.15$ ), and the correlation coefficient ( $R$ ) values of 0.984 (close to 1), the calculated  $[(r^2 - r_0^2)/r^2]$  values of -0.004 ( $<0.1$ ) were obtained. It was indicated in this external validation process that the CoMSIA model exhibited better predictive power than CoMFA model, and both the two models would be reliable for being used to predict the potencies of novel derivatives.

## 2.3. Graphical interpretation of CoMFA model

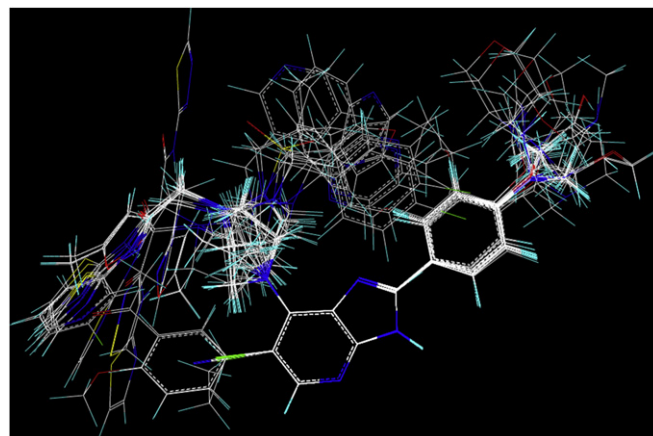
To visualize the information content of the derived 3D-QSAR model, CoMFA contour maps were generated by interpolating the products between the 3D-QSAR coefficients and their associated standard deviations. They could rationalize the regions in 3D space around the molecules where changes in the steric and electrostatic fields were predicted to increase or decrease the activity. The CoMFA steric and electrostatic contour maps were shown in Fig. 6 using compound **30** as a reference structure. The steric field is represented by green and yellow contours, in which green contours (80% contribution) indicate regions where bulky group would be favorable, while the yellow contours (20% contribution) represent

**Table 2**

The actual pIC<sub>50</sub>s, predicted pIC<sub>50</sub>s (Pred.) and their residuals (Res.) of the training and test set molecules.

Compd. No.	pIC <sub>50</sub>	CoMFA		CoMSIA	
	Actual	Pred.	Res.	Pred.	Res.
1	6.287	6.310	−0.023	6.464	−0.178
2 <sup>a</sup>	7.260	7.574	−0.314	7.437	−0.177
3	7.301	7.370	−0.069	7.183	0.118
4	7.276	7.188	0.088	7.080	0.196
5	7.377	7.106	0.271	7.202	0.175
6	7.796	7.870	−0.074	7.834	−0.038
7	7.092	7.259	−0.168	7.210	−0.119
8	7.284	7.451	−0.167	7.447	−0.163
9	8.523	8.586	−0.063	8.480	0.043
10 <sup>a</sup>	6.939	6.904	0.035	7.036	−0.097
11	6.928	6.907	0.021	6.993	−0.065
12	7.125	6.856	0.269	7.088	0.037
13 <sup>a</sup>	6.558	6.511	0.047	6.654	−0.096
14	6.750	6.896	−0.146	6.716	0.034
15	6.588	6.723	−0.135	6.630	−0.042
16	6.678	6.527	0.151	6.537	0.141
17 <sup>a</sup>	7.071	6.962	0.109	7.341	−0.270
18	7.102	7.025	0.077	6.742	0.360
19 <sup>a</sup>	7.260	7.383	−0.123	7.585	−0.325
20	8.046	7.845	0.201	8.157	−0.111
21	8.097	7.975	0.122	7.990	0.107
22	7.398	7.391	0.007	7.379	0.019
23	6.187	6.525	−0.338	6.648	−0.461
24	7.770	7.675	0.095	7.839	−0.069
25	7.092	7.020	0.072	6.952	0.140
26 <sup>a</sup>	7.495	7.252	0.243	7.458	0.037
27	8.301	8.351	−0.050	8.344	−0.043
28	8.222	8.237	−0.015	8.302	−0.080
29	7.921	7.675	0.246	7.913	0.008
30	8.699	8.786	−0.087	8.835	−0.136
31	7.523	7.351	0.172	7.364	0.159
32	7.222	7.279	−0.057	7.154	0.068
33	8.000	8.156	−0.156	7.858	0.142
34 <sup>a</sup>	7.854	8.063	−0.209	7.910	−0.056
35	8.000	7.820	0.180	7.717	0.283
36	7.260	7.446	−0.186	7.426	−0.166
37	7.523	7.710	−0.187	7.698	−0.175
38	7.538	7.328	0.210	7.410	0.128
39	8.000	7.915	0.085	7.911	0.089
40	7.824	7.925	−0.101	8.048	−0.224
41 <sup>a</sup>	8.000	7.550	0.450	7.760	0.240
42 <sup>a</sup>	8.000	8.210	−0.210	8.408	−0.408
43	8.523	8.692	−0.169	8.663	−0.140
44	5.367	5.461	−0.095	5.223	0.144
45	5.000	4.993	0.007	4.777	0.223
46	5.181	5.382	−0.201	5.350	−0.169
47	4.796	4.693	0.103	4.760	0.036
48	4.538	4.393	0.145	4.542	−0.004
49	5.337	5.552	−0.215	5.411	−0.074
50 <sup>a</sup>	5.357	5.696	−0.340	5.498	−0.142
51	6.244	5.991	0.253	5.990	0.254
52 <sup>a</sup>	5.367	5.457	−0.090	5.303	0.064
53	6.310	6.000	0.310	6.186	0.124
54	5.161	5.694	−0.533	5.433	−0.272
55	4.699	4.796	−0.097	5.072	−0.373
56	5.046	4.878	0.168	5.091	−0.045
57	6.602	6.295	0.307	6.245	0.357
58	5.155	5.406	−0.251	5.314	−0.159
59 <sup>a</sup>	5.620	5.772	−0.152	5.559	0.061
60	6.061	6.032	0.029	6.125	−0.064

<sup>a</sup> Test set molecules.

**Fig. 2.** Alignment of the compounds used in the training set.

substituted derivatives (e.g. **1–43**) exhibited significantly improved activities than that with R<sub>1</sub> unsubstituted compounds (e.g. **44–50**, **52** and **58–59**). Compounds **9**, **20**, **21**, **27**, **28**, **30**, **32** and **43** bearing a relative bulky bromo- or chloro-substituent at this position, were the most potential derivatives. A huge green contour around the R<sub>2</sub> position suggested that a bulky group at this position would benefit the potency. This may explain why derivatives **1–43** which possessed a bulky *N*-substituted piperazin-1-yl group showed better potencies than C-6 unsubstituted derivatives **44–56** as well as compounds **58** and **59** which had a relative minor group (e.g. methyl, chloro-) at this site. Another green contour near the R<sub>3</sub> position revealed that a bulky substituent at this site would increase the activity. This consisted with the fact that compounds **2–43** and **60** bearing a bulky substituted or unsubstituted phenyl group at C-2 position showed improved potencies than derivative **1** with a hydrogen atom at C-2. For instance, compounds **1**, **7**, **2**, **6** had an order for the activity of **1** < **7** < **2** < **6**, with the corresponding C-2 substituent -H, phenyl, *p*-dimethylaminophenyl, *p*-dimethylaminomethylphenyl, respectively. Compared derivative **9** with **10**, **30** with **24**, **35** and **34**, their activity discrepancies can be also explained by this green contour.

In Fig. 6(b), the red contour near the R<sub>1</sub> position indicated that an electron-withdrawing substituent at this site would be favored. Compounds **2**, **3**, **5–43**, **51**, **55** and **57** with an electron-withdrawing group (e.g. chloro-, bromo-) at R<sub>1</sub> exhibited significantly improved potencies than derivatives **44–50**, **52**, **54** and **58–60** bearing a hydrogen atom or electron-donating methyl substituent at R<sub>1</sub>. For instance, compounds **51**, **52**, **54** had an order for the activity of

**Table 3**

PLS results of CoMFA and CoMSIA models.

PLS Statistics	CoMFA	CoMSIA
$r^2$ <sup>a</sup>	0.774	0.800
$r^2$ <sup>b</sup>	0.975	0.977
ONC <sup>c</sup>	6	6
SEE <sup>d</sup>	0.195	0.188
F value	268.426	290.720
Field contribution		
Steric	0.481	0.191
Electrostatic	0.519	0.297
Hydrophobic	—	0.210
H-bond donor	—	0.089
H-bond acceptor	—	0.213

<sup>a</sup> Cross-validated correlation coefficient.

<sup>b</sup> Non-cross-validated coefficient.

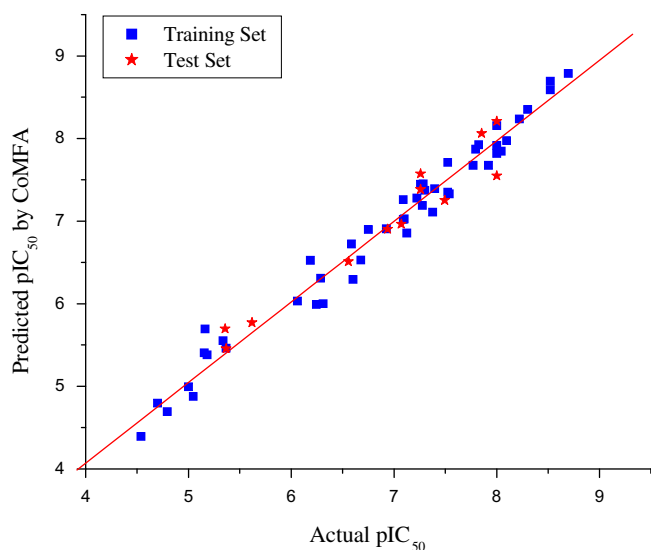
<sup>c</sup> Optimal number of components.

<sup>d</sup> Standard error of estimate.

regions where bulky group would decrease the activity. The electrostatic field is indicated by blue (80% contribution) and red (20% contribution) contours, which demonstrate the regions where electron-donating group and electron-withdrawing group would be favorable respectively.

In Fig. 6(a), the green contour near the R<sub>1</sub> position indicated that a bulky substituent at this site would be favorable. In general, R<sub>1</sub>





**Fig. 3.** Graph of actual versus predicted  $pIC_{50}$  of the training set and the test set using CoMFA.

**51 > 52 > 54**, with the corresponding  $R_1$  substituent chloro-, hydrogen, methyl, respectively. Compared compound **57** with **59** as well as **2** with **60**, their activity discrepancies can be also explained by this red contour. Another small red contour near the terminal of  $R_2$  position suggested that an electron-withdrawing group at this site would improve the activity. Most of the excellent derivatives (**2–9**, **17–22** and **24–43**) all possessed an electron-withdrawing aromatic group (e.g. isoxazolyl, phenyl, thiazolyl, pyridinyl, pyrimidinyl, imidazolyl) at the terminal of  $R_2$  site, meanwhile, those without an electron-withdrawing substituent at this position (**16**, **23**, **44–56** and **58**) were the most inactive compounds. Compared derivative **2** with **16**, **23** with **24** as well as **58** with **60**, their activity discrepancies can be also explained by this red contour. Three blue contours around the  $R_3$  position demonstrated that an electron-donating group at this position would be favorable. Most of the compounds possessed an electron-donating substituent (e.g. dimethylamino, 4-methylpiperazin-1-yl, morpholinylmethyl, methoxyl, dimethylaminomethyl, aminomethyl) at this site, compounds **49** and **50** with an electron-withdrawing group (e.g. *p*-Pyrrolidin-1-yl, *p*-Pyrid-2-yl) at  $R_3$  showed decreased activities.

#### 2.4. Graphical interpretation of CoMSIA model

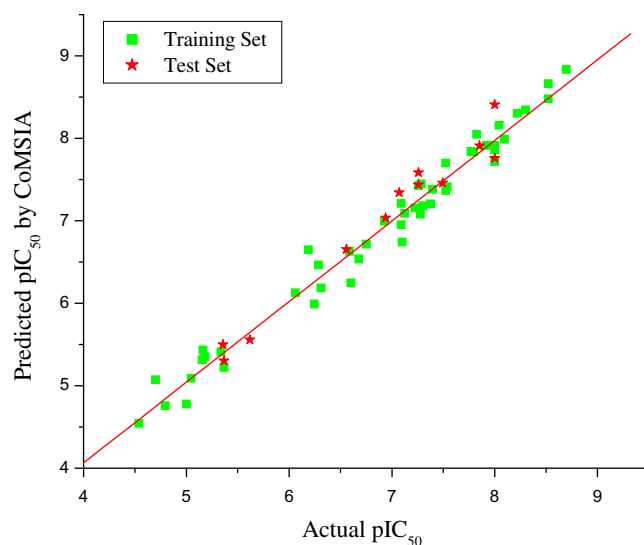
Fig. 7(a–e) provided the steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor contours plots for compound **30** of the CoMSIA model. The CoMSIA steric and electrostatic contour

**Table 4**  
Summary of PLS results of CoMSIA analysis.

	$r_{cv}^2$	$r^2$	ONC	SEE	F value
S + E	0.780	0.969	6	0.220	210.689
S + E + H	0.807	0.971	5	0.210	277.483
S + E + D	0.757	0.961	6	0.245	167.180
S + E + A	0.787	0.964	6	0.236	182.159
H + D + A	0.768	0.936	5	0.310	122.576
S + E + H + D	0.801	0.977	6	0.189	286.075
S + E + H + A	0.806	0.964	5	0.233	223.872
S + E + D + A	0.759	0.960	6	0.246	165.948
<b>S + E + H + D + A<sup>a</sup></b>	<b>0.800</b>	<b>0.977</b>	<b>6</b>	<b>0.188</b>	<b>290.720</b>

S: Steric; E: Electrostatic; H: Hydrophobic; D: H-bond donor; A: H-bond acceptor.

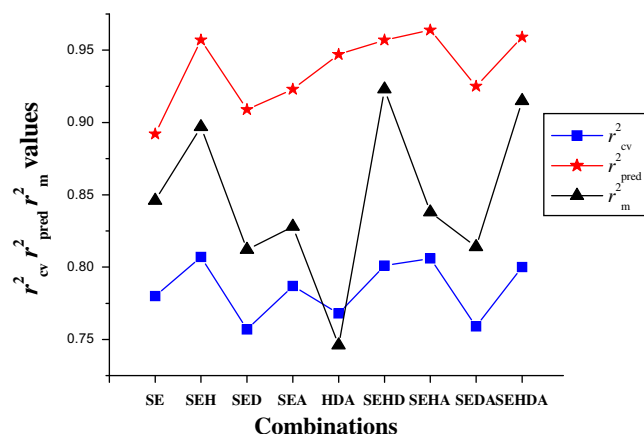
<sup>a</sup> Best model for CoMSIA.



**Fig. 4.** Graph of actual versus predicted  $pIC_{50}$  of the training set and the test set using CoMSIA.

maps were shown in Fig. 7(a) and (b) which were almost the same to the corresponding CoMFA steric and electrostatic contour maps. In hydrophobic field, white (20% contribution) and yellow (80% contribution) contours highlighted areas where hydrophilic and hydrophobic properties were favored. In hydrogen bond donor field, the cyan (80% contribution) and purple (20% contribution) contours indicated favorable and unfavorable hydrogen bond donor groups. In hydrogen bond acceptor field, the magenta (80% contribution) and red (20% contribution) contours identified favorable and unfavorable positions for hydrogen bond acceptors.

In Fig. 7(c), the yellow contour near the  $R_1$  position indicated that a hydrophobic substituent at this site would benefit the activity. All of the derivatives involved in the present study possessed a hydrophobic substituent (e.g. chloro-, bromo-, methyl, cyano, cyclopropyl, 3-methoxyphenyl, 4-methoxyphenyl) at  $R_1$ . Two white contours around the  $R_2$  position suggested that a hydrophilic group at this position may be favored. This consisted with the fact that compounds **9**, **20**, **21**, **27**, **28**, **30**, **32** and **43** bearing a relative hydrophilic groups (e.g. *N*-thiazol-2-yl-acetamino, pyridinylmethyl, pyrimidinylmethyl, thiazolylmethyl, 5-methyl-isoxazol-3-ylmethyl) were the most potential derivatives. Compounds **44–59** without a hydrophilic substituent at this position displayed



**Fig. 5.** Comparative study of different combinations of CoMSIA model.

**Table 5**  
Results of the external validation for CoMFA and CoMSIA models.

Parameters	CoMFA	CoMSIA
$r^2_{\text{pred}}$	0.933	0.959
Slope $a$	1.034	0.936
Intercept $b$	−0.281	0.352
Correlation coefficient $R$	0.971	0.984
Slope $k$	0.984	0.981
$r^2_m$	0.883	0.915
$[(r^2 - r^2_0)/r^2]$	−0.009	−0.004

**Table 6**  
Summary of external validation results of CoMSIA analysis.

	$r^2_{\text{pred}}$	$a$	$b$	$R$	Slope $k$	$[(r^2 - r^2_0)/r^2]$	$r^2_m$
SE	0.892	1.041	−0.411	0.962	0.985	−0.017	0.846
SEH	0.957	0.973	0.114	0.982	0.982	−0.011	0.897
SED	0.909	1.028	−0.350	0.971	0.985	−0.025	0.812
SEA	0.923	1.026	−0.360	0.981	0.984	−0.021	0.828
HDA	0.947	0.815	1.291	0.972	0.977	−0.044	0.746
SEHD	0.957	0.925	0.440	0.982	0.980	−0.003	0.923
SEHA	0.964	0.957	0.223	0.984	0.981	−0.018	0.838
SEDA	0.925	0.997	−0.147	0.980	0.983	−0.024	0.814
<b>SEHDA<sup>a</sup></b>	<b>0.959</b>	<b>0.936</b>	<b>0.352</b>	<b>0.984</b>	<b>0.981</b>	<b>−0.004</b>	<b>0.915</b>

S: Steric; E: Electrostatic; H: Hydrophobic; D: H-bond donor; A: H-bond acceptor.

<sup>a</sup> Best model for CoMSIA.

significantly decreased activities. Moreover, compared compound **23** ( $R_2$  = cyclopropylmethyl) with **21** ( $R_2$  = pyrimidinylmethyl) as well as **16** ( $R_2$  = isobutyl) with **2** ( $R_2$  = *N*-thiazol-2-yl-acetamino), their activity discrepancies can be explained by this red contour. Compounds **58–60** had an order for the activity of **60** > **59** > **58**, with the corresponding C-2 substituent *N*-thiazol-2-yl-acetaminopiperazinyl, chloro-, methyl, respectively. Another white contour around the  $R_3$  position demonstrated that a hydrophilic substituent may be favorable. This may be the reason that why derivatives had a relative hydrophilic (e.g. dimethylamino, 4-methylpiperazin-1-yl, morpholinylmethyl, dimethylaminomethyl, aminomethyl) at this site showed excellent potencies.

In Fig. 7(d), only a purple and a cyan contours around the terminal of the  $R_2$  position, which indicated that the hydrogen bond donor field at this site was not very important for the potency. In fact, the hydrogen bond donor gave the least contribution (0.089) among the five fields (Table 3).

In Fig. 7(e), a purple contour near the  $R_2$  site suggested that a hydrogen bond acceptor substituent at this position would increase the activity. Most of the active derivatives (**2–12**, **17**, **19–21**,

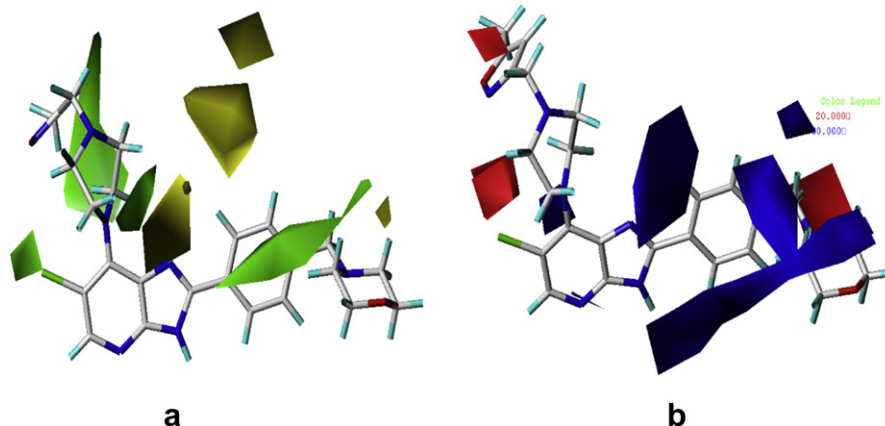
**24–43**) all possessed hydrogen bond acceptor groups such as carbonyl, oxygen or nitrogen atom in the *N*-thiazol-2-yl-acetamino, pyridinylmethyl, pyrimidinylmethyl, thiazolylmethyl, 5-methylisoxazol-3-ylmethyl substituent at  $R_2$ . Compounds **15**, **16**, **23** and **44–59** without a hydrogen bond acceptor substituent at this position, exhibited significantly decreased potencies. Two huge purple contours around the  $R_3$  position also revealed that a hydrogen bond acceptor group would improve the activity. All of the active compounds were found to possess hydrogen bond acceptor nitrogen or oxygen atom in the  $R_3$  site, which validated the observation from Fig. 7(e).

## 2.5. Docking analysis

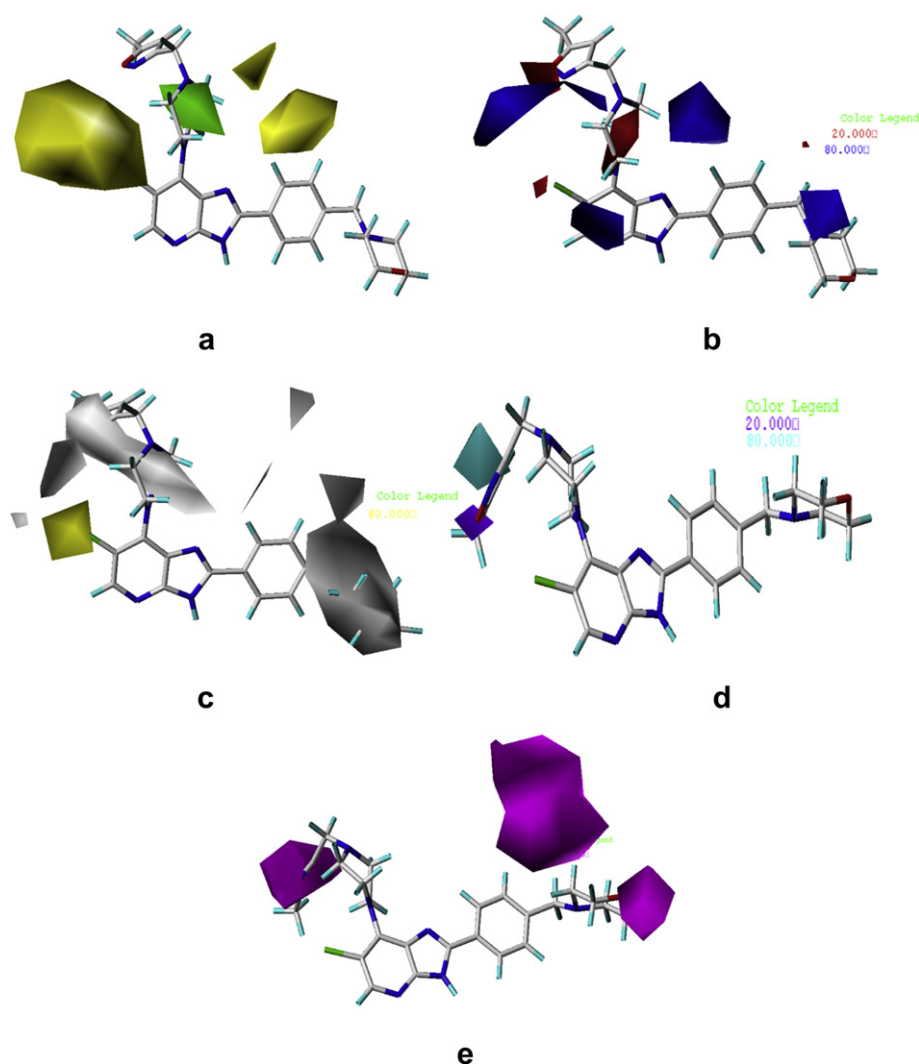
Docking was employed to explore the binding mode between these imidazo[4,5-*b*]pyridine derivatives and the Aurora A, furthermore, to examine the stability of 3D-QSAR models previous established. In order to visualize secondary structure elements, the MOLCAD Robbin surfaces program was applied. Fig. 8 showed the secondary structure of the most active compound **30** in complex with the ATP pocket, alpha helices were shown as helices or cylinders, while beta sheets were shown as arrows and the loop regions as tubes. The key residues and hydrogen bonds were labeled. As shown in Fig. 8, the N-1 atom acted as a hydrogen bond acceptor by forming two H-bonds with the  $-NH_3$  group of the Lys162 residue; the  $-NH$  group of the imidazolyl acted as a hydrogen bond donor and formed two H-bonds with the carbonyl groups of the Asp274 and Glu181, respectively; the N atom of morpholinylmethyl substituent at  $R_3$  position also acted as a hydrogen bond acceptor and formed an H-bond with the Lys143 residue. The observations taken from Fig. 8 were in agreement with the corresponding CoMSIA hydrogen bond contour maps.

The MOLCAD surface of the binding area was also developed and displayed with cavity depth (CD), lipophilic potential (LP) and hydrogen bond site (HB) to further explore the interaction between these inhibitors and the receptor. These potentials on a protein surface can be used to find the sites that act attractively on ligands by matching opposite colors.

Fig. 9(a) and (b) depicted the MOLCAD Robbin and Multi-Channel cavity depth potential surfaces structure of the binding site within the compound **30**. The cavity depth color ramp ranges from blue (low depth values = outside of the pocket) to light red (high depth values = cavities deep inside the pocket). In both Fig. 9(a) and (b), the morpholinylmethyl substituent at  $R_3$  position of compound



**Fig. 6.** Std\* coeff contour maps of CoMFA analysis with 2 Å grid spacing in combination with compound **30**. (a) Steric fields: green contours indicate regions where bulky groups increase activity, while yellow contours indicate regions where bulky groups decrease activity, and (b) Electrostatic fields: blue contours represent regions where electron-donating groups increase activity, while red contours represent regions where electron-withdrawing groups increase activity (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).



**Fig. 7.** Std\* coeff contour maps of CoMSIA analysis with 2 Å grid spacing in combination with compound 30. (a) Steric contour map. Green and yellow contours refer to sterically favored and unfavored regions. (b) Electrostatic contour map. Blue and red contours refer to regions where electron-donating and electron-withdrawing groups are favored. (c) Hydrophobic contour map. White and yellow contours refer to regions where hydrophilic and hydrophobic substituent are favored. (d) Hydrogen bond donor contour map. The cyan and purple contours indicate favorable and unfavorable hydrogen bond donor groups. (e) Hydrogen bond acceptor contour map. The magenta and red contours demonstrated favorable and unfavorable hydrogen bond acceptor groups (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

**30** as found in cyan area which indicated that the terminal of R<sub>3</sub> position was anchored outside the ATP pocket. The rest parts of compound **30** were oriented in a light red region which demonstrated that the majority parts of the molecule were anchored deep inside the pocket.

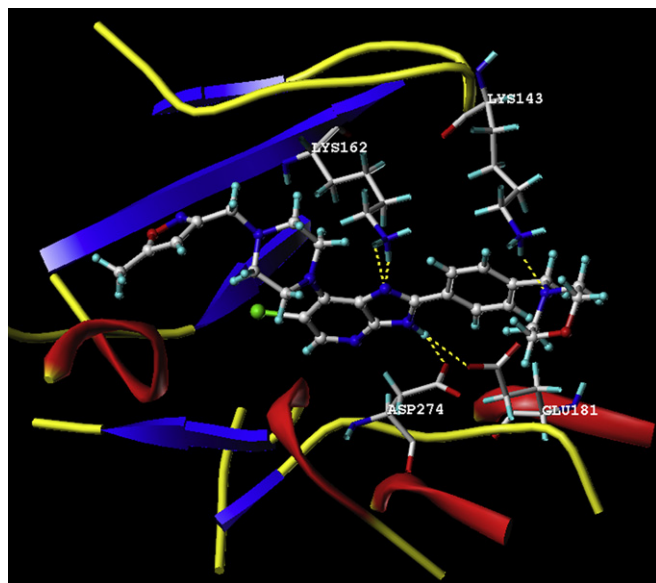
Fig. 10 showed the MOLCAD lipophilic potential surface of the binding area, the color ramp for LP ranges from brown (highest lipophilic area of the surface) to blue (highest hydrophilic area). The R<sub>1</sub> position was oriented to a brown region, suggesting that a hydrophobic substituent may be favored; the R<sub>2</sub> site was oriented to a white area which indicated that a hydrophilic group would be favorable; the R<sub>3</sub> position was surrounded by a blue surface which demonstrated that a hydrophilic group would benefit the potency. The observations taken from Fig. 10 satisfactorily matched those of the CoMSIA hydrophobic contour map.

Fig. 11 displayed the MOLCAD hydrogen bonding sites of the binding surfaces, ligands can be docked to proteins by matching the patterns displayed on the surface, the color ramp for HB ranges from red (hydrogen donors) to blue (hydrogen acceptors). As shown in Fig. 11, N-1 site was oriented to a red surface, which indicated that the surface of this site were hydrogen bond donors,

and a hydrogen bond acceptor substituent would be favorable; the –NH of imidazolyl group was anchored to a blue area, which indicated that the surface of this region were hydrogen bond acceptors, and a hydrogen bond donor property may be favored; and the N atom of morpholinylmethyl substituent at R<sub>3</sub> position was oriented to a red surface, which indicated that the surface of this region were hydrogen bond donors, and a hydrogen bond acceptor property may be favorable. The observations taken from this hydrogen bonding sites satisfactorily matched the corresponding CoMSIA hydrogen bond contour maps.

## 2.6. Summary of structure–activity relationship

The structure–activity relationship revealed by 3D-QSAR and molecular docking studies were illustrated in Fig. 12. In detail, the bulky, electron-withdrawing and hydrophobic groups at R<sub>1</sub> position are favorable; the bulky, electron-withdrawing, hydrophilic and hydrogen bond acceptor substituent at R<sub>2</sub> position would increase the activity; the bulky, electron-donating, hydrophilic and hydrogen bond acceptor groups at R<sub>3</sub> position may benefit the



**Fig. 8.** MOLCAD robbin surfaces structure of selected compound 30 in complex with pocket. Key residues and hydrogen bonds were labeled. The alpha helices were shown as helices or cylinders, while beta sheets were shown as arrows and the loop regions as tubes.

potency. The imidazo[4,5-*b*]pyridine skeleton was essential for binding to the ATP pocket.

### 2.7. Design of novel derivatives

By utilizing the structure–activity relationship revealed by this study as well as the synthetic availability of these derivatives, a set of thirty novel derivatives were designed, these molecules were aligned to the database and their activities were predicted by the CoMFA and CoMSIA models previously established. These molecules were designed by introducing bulky, electron-withdrawing and hydrophobic groups (trifluoromethyl, bromo-, iodo-) at R<sub>1</sub>; bulky, electron-withdrawing, hydrophilic and hydrogen bond acceptor substituent (substituted pyrazolyl, isothiazolyl, oxazolyl) at R<sub>2</sub>; bulky, electron-donating, hydrophilic and hydrogen bond

acceptor groups (*N*-morpholinylsulfonyl, *N*-morpholinylmethyl) at R<sub>3</sub> position. The phenyl group at C-2 site was also replaced with a more hydrophilic pyridinyl substituent (**D29**, **D30**). The chemical structures and predicted pIC<sub>50</sub> values of these compounds were shown in Table 7, and the graph of their predicted pIC<sub>50</sub> values versus the most active compound **30** was shown in Fig. 13. All of the designed molecules exhibited better predicted pIC<sub>50</sub> values than compound **30** in CoMFA or CoMSIA models. Molecules **D1–D12** displayed significantly improved predicted activities than compound **30** in both the CoMFA and CoMSIA models. The most potential designed molecule **D8** was nearly 6–7 fold more active than compound **30**. The results validated the structure–activity relationship obtained by this study.

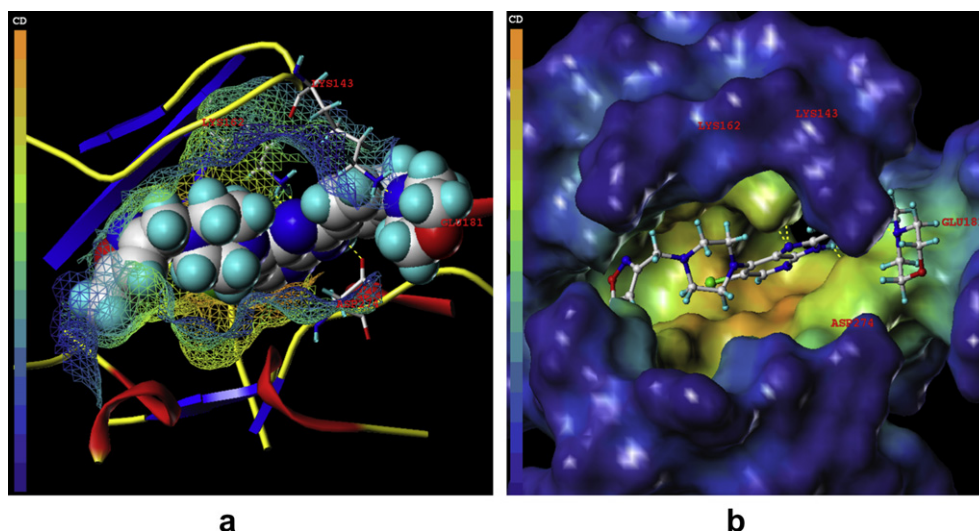
## 3. Conclusion

Aurora A kinase inhibitors have been regarded as potential attractive anti-tumor agents. We employed 3D-QSAR and docking methods to explore the structure–activity relationship for a series of recently reported imidazo[4,5-*b*]pyridine derivatives with excellent Aurora A kinase inhibitory activities. The 3D-QSAR study yielded stable and statistically significant predictive models as indicated by high cross-correlation coefficients. Furthermore, the CoMFA and CoMSIA contour maps along with the docking results offered enough information to understand the structure–activity relationship and identified several important structural features influencing the inhibitory activity. A set of thirty novel derivatives were designed by utilizing the structure–activity relationship taken from present study, based on the excellent performance of the external validation, the predicted activities of these newly designed molecules would be reliable. These molecules can be synthesized to generate a greater number of imidazo[4,5-*b*]pyridine derivatives with required pharmacokinetics for further clinical evaluations.

## 4. Computational methods

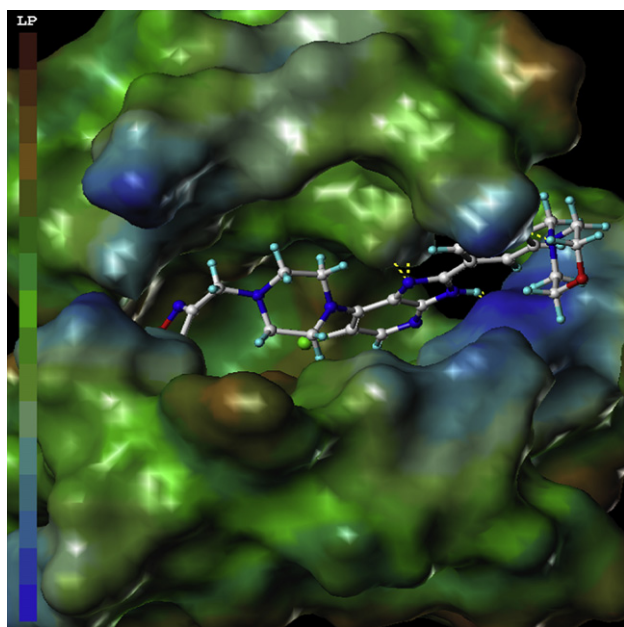
### 4.1. Data set and molecular modeling

The sixty imidazo[4,5-*b*]pyridine derivatives used in this study were originated from literature [15,16]. The inhibitory activity data



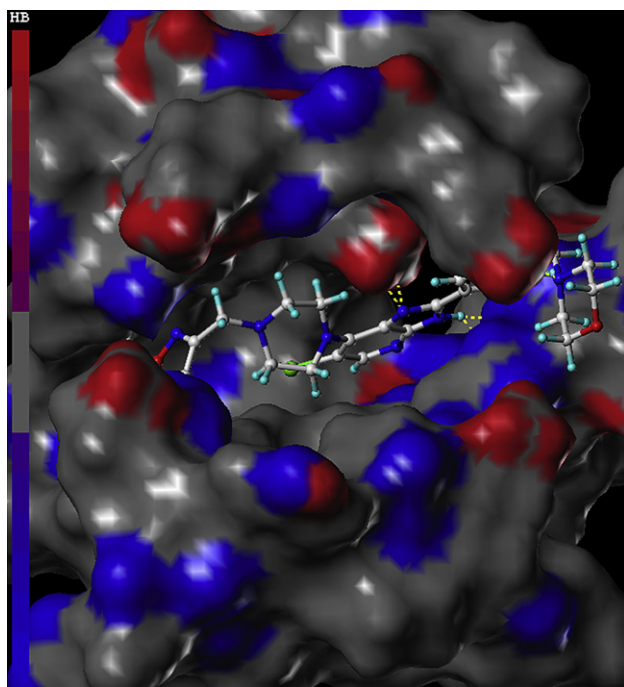
**Fig. 9.** The MOLCAD Robbin and Multi-Channel (a) and Molecule (b) surfaces structure displayed with cavity depth potential of the binding site within the compound 30. Key residues and hydrogen bonds were labeled. The cavity depth color ramp ranges from blue (low depth values = outside of the pocket) to light red (high depth values = cavities deep inside the pocket) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).



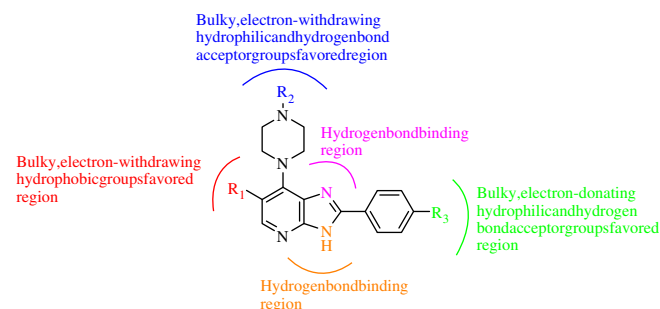


**Fig. 10.** The MOLCAD lipophilic potential surface of the pocket within the compound 30. The color ramp for LP ranges from brown (highest lipophilic area of the surface) to blue (highest hydrophilic area) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

were reported as  $IC_{50}$  against Aurora A kinase.  $IC_{50}$  was converted to  $pIC_{50}$  by taking  $\text{Log}(1/IC_{50})$ , the  $pIC_{50}$  values were used as the dependent variables in all the models developed. Structures of entire sets of imidazo[4,5-b]pyridine derivatives were built using SYBYL 8.1 program package of Tripos, Inc. [30]. 3D structures of all compounds were constructed using the Sketch Molecule module.



**Fig. 11.** The MOLCAD hydrogen binding sites surface of the binding site within the compound 30. The color ramp for HP ranges from red (hydrogen bond donors) to blue (hydrogen bond acceptors) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).



**Fig. 12.** Structure–activity relationship taken from present 3D-QSAR and docking studies.

Structural energy minimization was performed using the standard Tripos molecular mechanics force field and Gasteiger–Hückel charge, the max iterations for the minimization was set to 2000. The minimization was terminated when the energy gradient convergence criterion of 0.05 kcal/mol Å was reached [31,32]. All of the structures were aligned into a lattice box by fitting with imidazo[4,5-b]pyridine skeleton as a common structure using compound 30 as a template, which was the most active compound. The aligned molecules were shown in Fig. 2.

#### 4.2. CoMFA and CoMSIA settings

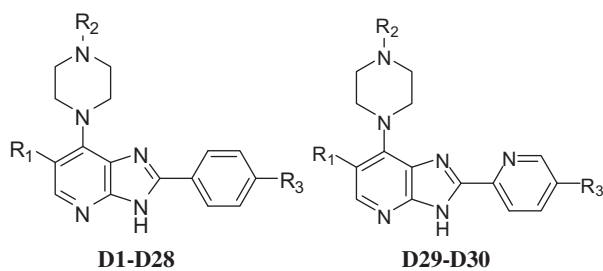
The CoMFA steric and electrostatic field energies were calculated using the Lennard-Jones and the Coulomb potentials, respectively. A  $sp^3$  carbon atom with a radius of 1.53 and a charge of +1.0 was used as a probe to calculate the molecules using the Tripos force field. The truncation for both the steric and the electrostatic energies were set to 30 kcal/mol. The CoMSIA method incorporating steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields was used for the analysis. A  $sp^3$  hybridized carbon atom was used as a probe atom to generate steric field energies and a charge of +1.0 to generate electrostatic field energies. Other similarity indices descriptors were calculated using a  $sp^3$  hybridized carbon atom as a probe atom with a +1 hydrophobicity, +1 H-bond donor and +1 H-bond acceptor properties. The attenuation factor was set to the default value of 0.3 [33–35].

#### 4.3. Partial least squares (PLS) analysis for CoMFA and CoMSIA

The initial PLS analysis was performed using the leave-one-out method for all 3D-QSAR analyses [36]. PLS was conjunct with the cross-validation option to determine the optimum number of components (ONC) which were then used in deriving the final CoMFA and CoMSIA model without cross-validation. The ONC was the number of components resulted in highest cross-validated correlated correlation coefficient ( $r_{cv}^2$ ) [37]. Column filtering was used at the default value of 2.0 kcal/mol in the cross-validation part. The final models were developed with ONC by using non-cross-validated analysis equal yielded the highest correlation coefficient ( $r^2$ ) [38].

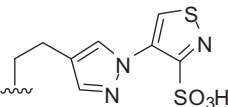
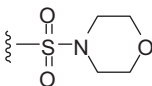
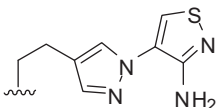
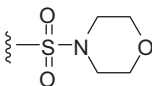
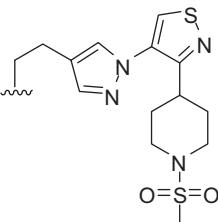
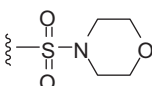
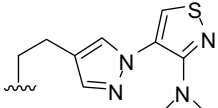
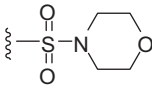
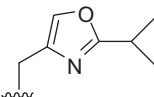

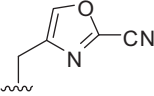

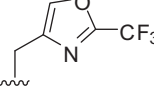

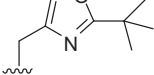

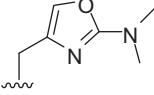

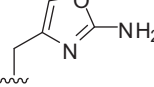

#### 4.4. External validation for CoMFA and CoMSIA models

The preliminary predictive abilities were determined from a test set of 12 compounds that were not included in the training set. These molecules were aligned to the template and their  $pIC_{50}$  values were predicted. The predictive correlation coefficient  $r_{pred}^2$  values of 0.933 and 0.959,  $r_{m}^2$ , based on the molecules of test set, was defined as follows:

**Table 7**Structures and predicted pIC<sub>50</sub> values of newly designed derivatives.

Compd. No.	Substituent			Predicted pIC <sub>50</sub>	
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CoMFA	CoMSIA
<b>D1</b>	–CF <sub>3</sub>			9.238	9.441
<b>D2</b>	–CF <sub>3</sub>			9.179	9.293
<b>D3</b>	–CF <sub>3</sub>			9.118	9.207
<b>D4</b>	–CF <sub>3</sub>			9.068	9.129
<b>D5</b>	–CF <sub>3</sub>			9.068	9.066
<b>D6</b>	–CF <sub>3</sub>			9.218	9.492
<b>D7</b>	–CF <sub>3</sub>			9.285	9.448
<b>D8</b>	–CF <sub>3</sub>			9.442	9.447

Table 7 (continued)

Compd. No.	Substituent			Predicted pIC <sub>50</sub>	
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CoMFA	CoMSIA
<b>D9</b>	–CF <sub>3</sub>			9.252	9.537
<b>D10</b>	–CF <sub>3</sub>			9.198	9.392
<b>D11</b>	–CF <sub>3</sub>			9.075	9.410
<b>D12</b>	–CF <sub>3</sub>			9.049	9.316
<b>D13</b>	–CF <sub>3</sub>			9.188	8.752
<b>D14</b>	–CF <sub>3</sub>			9.324	8.700
<b>D15</b>	–CF <sub>3</sub>			9.381	8.718
<b>D16</b>	–CF <sub>3</sub>			9.480	8.942
<b>D17</b>	–CF <sub>3</sub>			9.297	8.749
<b>D18</b>	–CF <sub>3</sub>			9.171	8.535

(continued on next page)

Table 7 (continued)

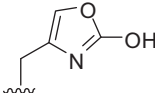

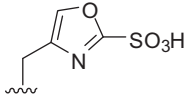

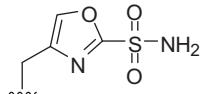

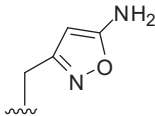
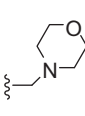
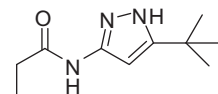

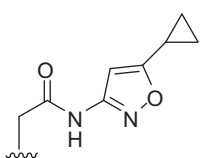
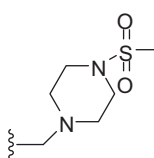
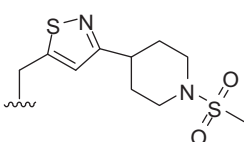
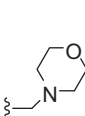
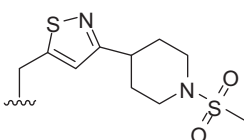
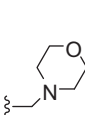
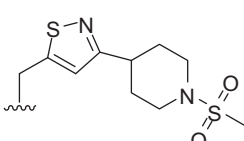
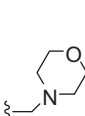
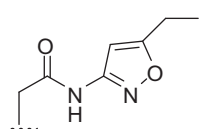
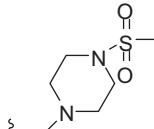
Compd. No.	Substituent			Predicted pIC <sub>50</sub>	
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CoMFA	CoMSIA
D19	–CF <sub>3</sub>			9.117	8.633
D20	–CF <sub>3</sub>			9.143	8.602
D21	–CF <sub>3</sub>			9.148	8.679
D22	–CF <sub>3</sub>			9.186	8.726
D23	–CF <sub>3</sub>			9.172	8.533
D24	–CF <sub>3</sub>			9.048	8.724
D25	–CF <sub>3</sub>			9.490	8.731
D26	–I			9.230	8.947
D27	–Br			9.160	8.531
D28	–CF <sub>3</sub>			9.041	8.717

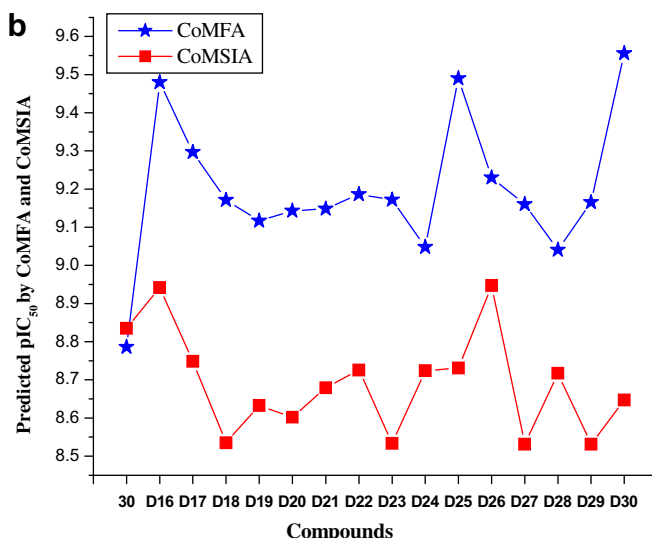
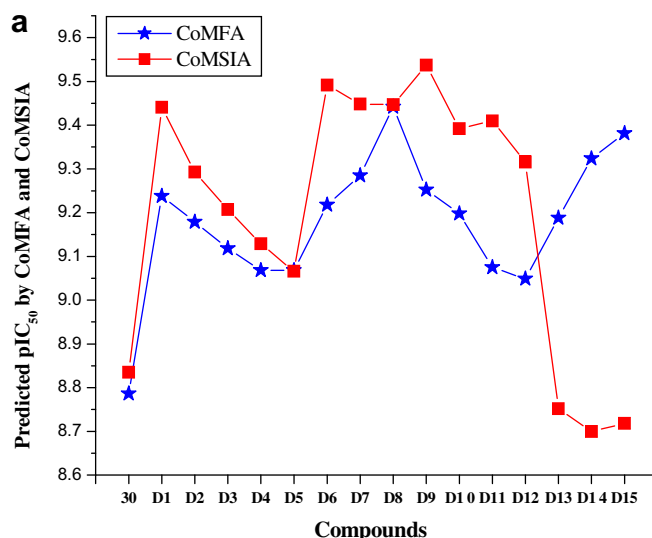


Table 7 (continued)

Compd. No.	Substituent			Predicted pIC <sub>50</sub>	
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CoMFA	CoMSIA
D29	–CF <sub>3</sub>			9.165	8.531
D30	–CF <sub>3</sub>			9.556	8.647

$$r_{\text{pred}}^2 = (\text{SD} - \text{PRESS})/\text{SD}$$

SD is the sum of the squared deviations between the inhibitory activities of the test set and mean activities of the training molecules and PRESS is the sum of squared deviations between predicted and actual activity values for each molecule in the test set [39–42].

Fig. 13. Graph of the predicted pIC<sub>50</sub> of the designed molecules versus compound 30.

Further external validation was obtained as follows. We defined  $\tilde{y}_i$  and  $y_i$  as the predicted and observed pIC<sub>50</sub> values of the test set molecules, respectively. If we plot  $y$  versus  $\tilde{y}$  for the ideal QSAR model, the regression line will bisect the angle formed by positive directions of the orthogonal axes  $\tilde{y}$  and  $y$ . The regression line can be described by expression  $y^r = a\tilde{y} + b$ , where

$$a = \frac{\sum (y_i - \bar{y}_o)(\tilde{y}_i - \bar{y}_p)}{\sum (\tilde{y}_i - \bar{y}_p)^2}$$

and

$$b = \bar{y}_o - a\bar{y}_p$$

In these equations,  $\bar{y}_o$  and  $\bar{y}_p$  are the average values of the observed and predicted pIC<sub>50</sub> values of the test set molecules. For the ideal model, the slope  $a$  is equal to 1, intercept  $b$  is equal to 0, and correlation coefficient:

$$R = \frac{\sum (y_i - \bar{y}_o)(\tilde{y}_i - \bar{y}_p)}{\sqrt{\sum (y_i - \bar{y}_o)^2 \sum (\tilde{y}_i - \bar{y}_p)^2}}$$

for the regression is equal to 1. A 3D-QSAR model may have a high predictive power, if it is close to the ideal one. Meanwhile, the regression of  $y$  against  $\tilde{y}$  through the origin:  $y_i^{r_0} = k\tilde{y}_i$ , should be characterized by  $k$  close to 1. Slope  $k$  is calculated as follow:

$$k = \frac{\sum y_i \tilde{y}_i}{\sum \tilde{y}_i^2}$$

Another essential parameter  $r_m^2$  was defined as follow:

$$r_m^2 = r^2 \left( 1 - \sqrt{|r^2 - r_0^2|} \right)$$

Where, the  $r^2$  was the non-cross-validated correlation coefficient obtained from the PLS process, and the  $r_0^2$  was calculated as follows:

$$r_0^2 = 1 - \frac{\sum (\tilde{y}_i - y_i^{r_0})^2}{\sum (\tilde{y}_i - \tilde{y}_p)^2}$$

Where, the  $y_{r_0}^i$  was obtained by this formula:

$$y_{r_0}^i = k\tilde{y}_i$$

3D-QSAR models were considered acceptable if they satisfy all of the following conditions [28,29]:

$$r_{cv}^2 > 0.5, \quad r^2 > 0.6, \quad \left[ (r^2 - r_0^2) / r^2 \right] < 0.1, \quad 0.85 \leq k \leq 1.15 \quad \text{and} \quad r_m^2 > 0.5.$$

#### 4.5. Docking settings

The Surflex-Dock using an empirical scoring function and a patented search engine to dock ligands into a protein's binding site was applied to study molecular docking [30]. Crystal structure of Aurora A kinase was retrieved from RCSB Protein Data Bank (PDB entry code: 2WTW). The Aurora A kinase structure was utilized in subsequent docking experiments without energy minimization. The inhibitors were docked into corresponding protein's binding site by an empirical scoring function and a patented search engine in Surflex-Dock [30]. All ligands and water molecules have been removed and the polar hydrogen atoms were added. To visualize the binding mode between the protein and ligand, the MOLCAD (Molecular Computer Aided Design) program was employed. MOLCAD calculates and displays the surfaces of channels and cavities, as well as the separating surface between protein subunits [30]. MOLCAD program provides several types to create a molecular surface. In this paper, the fast Connolly method which uses a marching cube algorithm to generate the surface was applied, meanwhile the MOLCAD Robbin and Multi-Channel surfaces program displayed with several potentials were established. Other parameters were established by default in software.

#### Acknowledgements

We gratefully acknowledge support for this research from Natural Science Foundation of Guangdong Province (No. 9151063201000053), China.

#### References

- [1] M. Li, A. Jung, U. Ganswindt, P. Marini, A. Friedl, P.T. Daniel, K. Lauber, V. Jendrosseck, C. Belka, *Biochem. Pharmacol.* 79 (2010) 122–129.
- [2] J.R. Medina, S.W. Grant, J.M. Axten, W.H. Miller, C.A. Donatelli, M.A. Hardwicke, C.A. Olejowski, Q. Liao, R. Plant, H. Xiang, *Bioorg. Med. Chem. Lett.* 20 (2010) 2552–2555.
- [3] A.H. Kishore, B.M. Vedamurthy, K. Mantelingu, S. Agrawal, B.A.A. Reddy, S. Roy, K.S. Rangappa, T.K. Kundu, *J. Med. Chem.* 51 (2008) 792–797.
- [4] T.E. Rawson, M. R  th, E. Blackwood, D. Burdick, L. Corson, J. Dotson, J. Drummond, C. Fields, G.J. Georges, B. Goller, J. Halladay, T. Hunsaker, T. Kleinheinz, H.-W. Krell, J. Li, J. Liang, A. Limberg, A. McNutt, J. Moffat, G. Phillips, Y. Ran, B. Safina, M. Ultsch, L. Walker, C. Wiesmann, B. Zhang, A. Zhou, B.-Y. Zhu, P. R  ger, A.G. Cochran, *J. Med. Chem.* 51 (2008) 4465–4475.
- [5] S. He, S. Yang, G. Deng, M. Liu, H. Zhu, W. Zhang, S. Yan, L. Quan, J. Bai, N. Xu, *Cell. Mol. Life Sci.* 67 (2010) 2069–2076.
- [6] F.H. Jung, G. Pasquet, C.L. der Brempt, J.M. Lohmann, N. Warin, F. Renaud, H. Germain, C.D. Savi, N. Roberts, T. Johnson, C. Dousson, G.B. Hill, A.A. Mortlock, N. Heron, R.W. Wilkinson, S.R. Wedge, S.P. Heaton, R. Odedra, N.J. Keen, S. Green, E. Brown, K. Thompson, S. Brightwell, *J. Med. Chem.* 49 (2006) 955–970.
- [7] D. Fancelli, J. Moll, M. Varasi, R. Bravo, R. Artico, D. Berta, S. Bindi, A. Cameron, I. Candiani, P. Cappella, P. Carpinelli, W. Croci, B. Forte, M.L. Giorgini, J. Klapwijk, A. Marsiglio, E. Pesenti, M. Rocchetti, F. Roletto, D. Severino, C. Sconini, P. Storici, R. Tonani, P. Zugnoni, P. Vianello, *J. Med. Chem.* 49 (2006) 7247–7251.
- [8] S. Wang, C.A. Midgley, F. Sca  rou, J.B. Grabarek, G. Griffiths, W. Jackson, G. Kontopidis, S.J. McClue, C. McInnes, C. Meades, M. Mezna, A. Plater, I. Stuart, M.P. Thomas, G. Wood, R.G. Clarke, D.G. Blake, D.I. Zheleva, D.P. Lane, R.C. Jackson, D.M. Glover, P.M. Fischer, *J. Med. Chem.* 53 (2010) 4367–4378.
- [9] I. Aliagas-Martin, D. Burdick, L. Corson, J. Dotson, J. Drummond, C. Fields, O.W. Huang, T. Hunsaker, T. Kleinheinz, E. Krueger, J. Liang, J. Moffat, G. Philips, R. Pulk, T.E. Rawson, M. Ultsch, L. Walker, C. Wiesmann, B. Zhang, B.-Y. Zhu, A.G. Cochran, *J. Med. Chem.* 52 (2009) 3300–3307.
- [10] M.S. Coumar, J.-S. Leou, P. Shukla, J.-S. Wu, A.K. Dixit, W.-H. Lin, C.-Y. Chang, T.-W. Lien, U.-K. Tan, C.-H. Chen, J.T.-A. Hsu, Y.-S. Chao, S.-Y. Wu, H.-P. Hsieh, *J. Med. Chem.* 52 (2009) 1050–1062.
- [11] M. Zhong, M. Bui, W. Shen, S. Baskaran, D.A. Allen, R.A. Elling, W.M. Flanagan, A.D. Fung, E.J. Hana, S.O. Harris, S.A. Heumann, U. Hoch, S.N. Ivy, J.W. Jacobs, S. Lam, H. Lee, R.S. McDowell, J.D. Oslob, H.E. Purkey, M.J. Romanowski, J.A. Silverman, B.T. Tangonan, P. Taverna, W. Yang, J.C. Yoburn, C.H. Yu, K.M. Zimmerman, T. O'Brien, W. Lew, *Bioorg. Med. Chem. Lett.* 19 (2009) 5158–5161.
- [12] K.J. Moriarty, H.K. Koblish, T. Garabrant, J. Maisuria, E. Khalil, F. Ali, I.P. Petrounia, C.S. Cryslar, A.C. Maroney, D.L. Johnson, R.A. Gallemmo, *Bioorg. Med. Chem. Lett.* 16 (2006) 5778–5783.
- [13] N.D. Adams, J.L. Adams, J.L. Burgess, A.M. Chaudhari, R.A. Copeland, C.A. Donatelli, D.H. Drewry, K.E. Fisher, T. Hamajima, M.A. Hardwicke, W.F. Huffman, K.K. Koretke-Brown, Z.V. Lai, O.B. McDonald, H. Nakamura, K.A. Newlander, G.A. Olejowski, C.A. Parrish, D.R. Patrick, R. Plant, M.A. Sarpong, K. Sasaki, S.J. Schmidt, K.J. Silva, D. Sutton, J. Tang, C.S. Thompson, P.J. Tummino, J.C. Wang, H. Xiang, J. Yang, D. Dhanak, *J. Med. Chem.* 53 (2010) 3973–4001.
- [14] J.D. Oslob, S.A. Heumann, C.H. Yu, D.A. Allen, S. Baskaran, M. Bui, E. Delarosa, A.D. Fung, A. Hashash, J. Hau, I. Sheryl, J.W. Jacobs, W. Lew, J. Maung, R.S. McDowell, S. Ritchie, M.J. Romanowski, J.A. Silverman, W. Yang, M. Zhong, T. Fuchs-Knotts, *Bioorg. Med. Chem. Lett.* 19 (2009) 1409–1412.
- [15] V. Bavetsias, J.M. Large, C. Sun, N. Boulloc, M. Kosmopoulou, M. Matteucci, N.E. Wilsher, V. Martins, J. Reynisson, B. Atrash, A. Faisal, F. Urban, M. Valenti, A.H. Brandon, G. Box, F.I. Raynaud, P. Workman, S.A. Eccles, R. Bayliss, J. Blagg, S. Linardopoulos, E. McDonald, *J. Med. Chem.* 53 (2010) 5213–5228.
- [16] V. Bavetsias, C. Sun, N. Boulloc, J. Reynisson, P. Workman, S. Linardopoulos, E. McDonald, *Bioorg. Med. Chem.* 17 (2007) 6567–6571.
- [17] E.A. Harrington, D. Bebbington, J. Moore, R.K. Rasmussen, A.O. Ajose-Adeogun, T. Nakayama, J.A. Graham, C. Demur, T. Hercend, A. D  u-Hercend, M. Su, J.M.C. Golec, K.M. Miller, *Nat. Med.* 10 (2004) 262–267.
- [18] S. Howard, V. Berdini, J.A. Boulstridge, M.G. Carr, D.M. Cross, J. Curry, L.A. Devine, T.R. Early, L. Fazal, A.L. Gill, M. Heathcote, S. Maman, J.E. Matthews, R.L. McMenamin, E.F. Navarro, M.A. O'Brien, M. O'Reilly, D.C. Rees, M. Reule, D. Tisi, G. Williams, M. Vinkovi  , P.G. Wyatt, *J. Med. Chem.* 52 (2009) 379–388.
- [19] M.G. Manfredi, J.A. Exsedy, K.A. Meetze, S.K. Balani, O. Burenkova, W. Chen, K.M. Galvin, K.M. Hoar, J.J. Huck, P.J. LeRoy, E.T. Ray, T.B. Sells, B. Stringer, S.G. Stroud, T.J. Vos, G.S. Weatherhead, D.R. Wysong, M. Zhang, J.B. Bolen, C.F. Claiborne, *Proc. Natl. Acad. Sci. U S A* 104 (2007) 4106–4111.
- [20] S. Vilar, G. Cozza, S. Moro, *Curr. Top. Med. Chem.* 18 (2008) 1555–1572.
- [21] K. Worm, R.E. Dolle, *Curr. Pharm. Des.* 29 (2009) 3345–3366.
- [22] K.C. Chou, *Curr. Drug Metab.* 4 (2010) 369–378.
- [23] G.H.G. Trossini, R.V.C. Guido, G. Oliva, E.I. Ferreira, A.D. Andricopulo, *J. Mol. Graph. Model.* 28 (2009) 3–11.
- [24] H. Gonzalez-Diaz, A. Duado-Sanchez, F.M. Ubeira, F. Prado-Prado, L.G. Perez-Montoto, R. Concu, G. Podda, B. Shen, *Curr. Drug Metab.* 4 (2010) 379–406.
- [25] F.J. Prado-Prado, E. Uriarte, F. Borges, H. Gonzalez-Diaz, *Eur. J. Med. Chem.* 44 (2009) 4516–4521.
- [26] F.J. Prado-Prado, F.M. Ubeira, F. Borges, H. Gonzalez-Diaz, *J. Comput. Chem.* 31 (2009) 164–173.
- [27] L.G. Perez-Montoto, F. Prado-Prado, F.M. Ubeira, H. Gonzalez-Diaz, *Curr. Proteomics* 4 (2009) 246–261.
- [28] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* 20 (2002) 269–276.
- [29] P.P. Roy, K. Roy, *QSAR Comb. Sci.* 27 (2008) 302–313.
- [30] SYBYL 8.1. Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri 63144, USA.
- [31] N. Zhang, R. Zhong, *Eur. J. Med. Chem.* 45 (2010) 292–297.
- [32] S.K. Sivan, V. Manga, *J. Mol. Model.* 16 (2010) 1169–1178.
- [33] K. Roy, S. Paul, *J. Mol. Model.* 16 (2010) 951–964.
- [34] P. Lan, Z.J. Huang, J.R. Sun, W.M. Chen, *Int. J. Mol. Sci.* 11 (2010) 3357–3374.
- [35] P.R. Murmumkar, R. Giridhar, M.R. Yadav, *Chem. Biol. Drug Des.* 71 (2008) 363–373.
- [36] H.-Y.P. Choo, J.-S. Lim, Y. Kam, S.Y. Kim, J. Lee, *Eur. J. Med. Chem.* 36 (2001) 829–836.
- [37] H. Zeng, R. Cao, H. Zhang, *Chem. Biol. Drug Des.* 74 (2009) 596–610.
- [38] R. Hu, F. Barbault, M. Delamar, R. Zhang, *Bioorg. Med. Chem.* 17 (2009) 2400–2409.
- [39] X.-Y. Lu, Y.-D. Chen, N. Sun, Y.-J. Jiang, Q.-D. You, *J. Mol. Model.* 16 (2010) 163–173.
- [40] P. Lu, X. Wei, R. Zhang, *Eur. J. Med. Chem.* 45 (2010) 3413–3419.
- [41] A. Basu, K. Jasu, V. Jayaprakash, N. Mishra, P. Ojha, S. Bhattacharya, *Eur. J. Med. Chem.* 44 (2009) 2400–2407.
- [42] P.P. Roy, S. Paul, I. Mitra, K. Roy, *Molecules* 14 (2009) 1660–1701.