

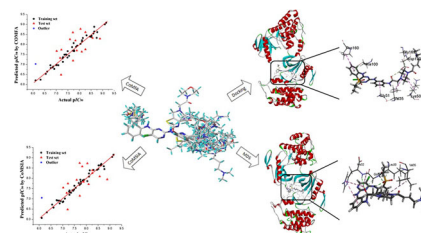
Identification of 2,4-diarylaminopyrimidine analogues as ALK inhibitors by using 3D-QSAR, molecular docking, and molecular dynamics simulations

Dan-Dan Li¹ · Fu-Long Wu¹ · Zhong-Hua Wang^{1,2} · Lei-Lei Huang¹ · Yan Yin¹ · Fan-Hong Wu¹

Received: 9 August 2016 / Accepted: 14 May 2017 / Published online: 26 July 2017
© Springer-Verlag Wien 2017

Abstract Anaplastic lymphoma kinase (ALK) is a particularly promising target for the development of small molecule anti-cancer drugs. In the present study, comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were performed on 60 ALK inhibitors to build three-dimensional quantitative structure–activity relationship models. Both the ligand-based resultants of CoMFA (r^2 0.970, q^2 0.660) and CoMSIA (r^2 0.979, q^2 0.623) models exhibited good predictability. The resulting contour maps illustrated the regions where interactive fields may affect the activity. Molecular docking was then performed to explore the interactions between these inhibitors and the ALK-4DCE protein. A few key residues (His32, Gly31, Gly169, Asp170, Val35, Ala100, Pro160, Lys50, and Leu30) at the binding site of 4DCE were identified. Molecular dynamics simulation further verified the reliability. The information acquired in this work not only provides a better appreciation of interactions between these molecules and the ALK receptor but could also be applied to design more effective ALK inhibitors.

Graphical abstract



Keywords ALK inhibitor · 2,4-Diarylaminopyrimidines · 3D-QSAR · Molecular docking · Molecular dynamics

Introduction

As a member of receptor tyrosine kinase, ALK (anaplastic lymphoma kinase) had attracted high clinical interest in the personalized treatment targeting the anti-cancer field [1], in particular, anaplastic large cell lymphoma (ALCL) [2], inflammatory myofibroblastic tumor [3], diffuse large B cell lymphoma (DLBCL) [4], renal cell carcinoma (RCC) [5], and non-small-cell lung cancer (NSCLC) [6–9].

The first ALK rearrangement was discovered as EML4-ALK fusion oncogene in NSCLC in 2007 [10]. Furthermore, ALK also fused with other proteins like nucleophosphamin (NPM) [11], ALK lymphoma oligomerization on chromosome 17 (ALO17) [12], TRK-fused gene (TFG) [13], moesin (MSN) [14] to form corresponding ALK-fusion proteins which are responsible for tumor growth [15]. Various ALK targeted drugs had been or are being tested in clinical trials and the first-in-class ALK inhibitor crizotinib was approved by FDA in 2011 for

✉ Zhong-Hua Wang
wzhsit@163.com

✉ Fan-Hong Wu
wfh@sit.edu.cn

¹ School of Chemical and Environmental Engineering, Shanghai Institute of Technology, Shanghai, People's Republic of China

² Shanghai Engineering Research Center of Pharmaceutical Progress, Shanghai 201203, China

the ALK-positive NSCLC [16]. The second-generation ALK inhibitor crizotinib was approved by FDA in 2014 not only for its potent ALK inhibitory but also for its anti-drug-resistance to crizotinib [17, 18].

Other novel ALK inhibitors, including CH5424802 [19], AP26113 [20], NVP-TAE684 [21], LDK378, X-396 [22], and ASP3026 are also in Phase 1 and Phase 2 clinical trials displaying enhanced specificity [23] (Fig. 1). All these

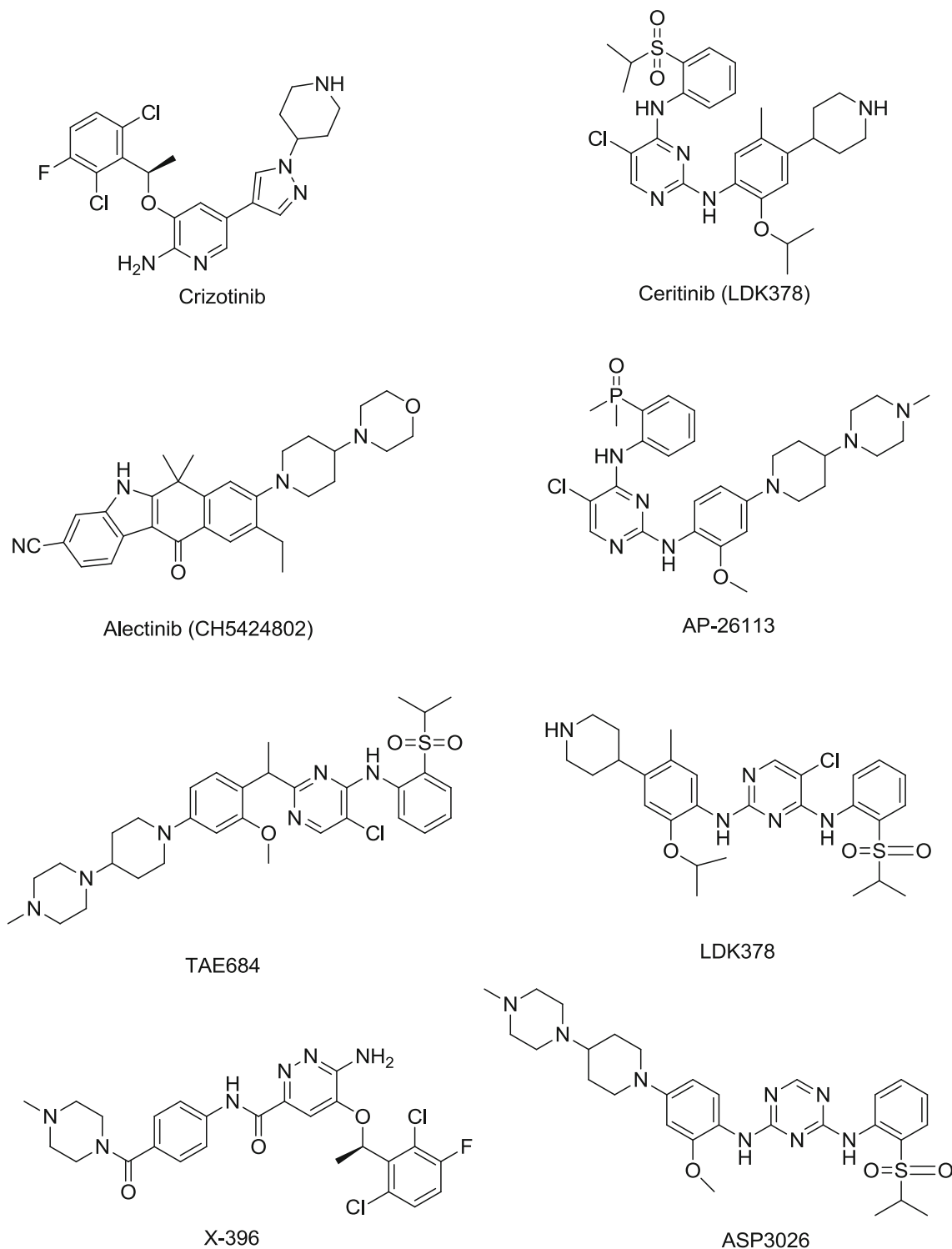


Fig. 1 Structures of anaplastic lymphoma kinase inhibitors

ALK inhibitors showed higher activity as well as enhanced significance in crizotinib-resistant ALK-positive NSCLC. In particular, the series of 2,4-diarylaminopyrimidine (DAAP) analogues showed high inhibitory activity against both c-Met and ALK kinases [24].

Quantitative structure–activity relationship (QSAR) is one of the most widely used rational methods for drug design. In this method, the interactions between molecules and receptor depended on the difference of the molecular field around the compound. On the basis of quantitative molecular field parameters as variables, regression analysis of drug activity can reflect the interaction model between drugs and biological macromolecules, and then new drugs could be designed accordingly [25]. For instance, Vivek as well as the research group of Wang developed three-dimensional quantitative structure–activity relationship (3D-QSAR) models for different sets of compounds including 2-acyliminobenzimidazoles derivatives and piperidine carboxamides derivatives to understand chemical–biological interactions [15, 26]. We have investigated 2-acyliminobenzimidazoles derivatives as potent ALK inhibitors [27]. In the present work, 60 diarylaminopyrimidine (DAAP) derivatives reported [16, 24, 28] as potent and selective ALK inhibitors were collected as a dataset, which were studied using a combination of CoMFA, CoMSIA, molecular docking, and molecular dynamic simulation. The purpose of this study is to establish a reliable 3D-QSAR model by CoMFA and CoMSIA methods to elucidate the structural characteristics of some DAAP derivatives such as ALK inhibitors, which may provide valuable guidance in the rational synthesis of more effective inhibitors.

Results and discussion

CoMFA and CoMSIA statistical result

It is very necessary to make an initial inspection of the inhibitor molecules before establishing the 3D-QSAR models. Compound **50** was considered as an outlier in the CoMFA and CoMSIA models because the r^2 prediction of model was 0.400 on inclusion of this compound, while excluding this compound the r^2 prediction value increased to 0.983. Statistically, an r^2 value >0.3 of the predicted set is usually considered significant, while an r^2 value >0.5 is statistically more significant in CoMFA and CoMSIA studies [29]. The reason for this outlier may be the difference in structure or the different binding conformations, and the larger deviation between the actual and predicted pIC_{50} values. Compound **50** and compounds **53** and **54** were very similar in structure, and the only difference was that the substituent at the nitrogen of azepane was a methyl, which might account for its outlier status since this molecule was the only compound with small volume group in this position.

Based on the internal research of the training set (44 molecules) and the external confirmation of the test set (16 molecules) the CoMFA and CoMSIA models were built. As shown in Table 1, the optimal CoMFA model resulted in a cross-validated q^2 of 0.660, a non-cross-validated correlation coefficient r^2 of 0.970, a standard error (SEE) value of 0.144, and F statistic value (F) of 167.010. For the CoMFA analysis, the q^2 value of 0.623, r^2 value of 0.979, SEE value of 0.120, and F statistic value of 241.162 were calculated, respectively.

Table 1 The best results of the CoMFA and CoMSIA PLS statistical results

	q^2	ONC	r_{ncv}^2	SEE	F	Field contribution/%				
						S	E	D	A	H
CoMFA										
S + E	0.66	7	0.970	0.144	167.010	43.8	56.2			
CoMSIA										
H + E	0.588	5	0.944	0.191	129.053		60.6			39.4
S + E	0.523	5	0.918	0.231	85.347	27.3	72.7			
S + E + D	0.519	5	0.938	0.202	114.891	18.9	52.8	28.2		
S + E + H	0.554	5	0.934	0.208	108.029	17.4	50.1			32.5
H + E + D	0.555	6	0.963	0.158	160.956		48.1	22.7		29.2
S + E + A	0.63	5	0.944	0.191	128.224	18.2	53.4		47.1	
S + E + D + H	0.548	5	0.943	0.194	124.941	14.1	41.4	20.3		24.2
S + E + D + A	0.616	7	0.976	0.128	211.675	13.8	37.1	21.6	27.5	
S + E + D + A + H	0.623	7	0.979	0.120	241.162	10.6	30.0	16.7	23.5	19.2

q^2 cross-validated correlation coefficient after the leave-one-out procedure, r_{ncv}^2 non-cross-validated correlation coefficient, SEE standard error of estimate; F F statistic values, ONC optimal number of components

Table 2 The actual and predicted pIC_{50} values of all compounds

Compd.	Actual	COMFA		COMSIA	
		Predicted	Residues	Predicted	Residues
1	6.553	6.529	0.024	6.717	-0.164
2	7.095	6.96	0.135	7.009	0.086
3	7.484	7.345	0.139	7.449	0.035
4	8.569	8.325	0.244	8.088	0.481
5	6.842	6.85	-0.008	6.912	-0.070
6	7.024	6.971	0.053	6.870	0.154
7	8.114	8.211	-0.097	8.230	-0.116
8	8.745	8.329	0.416	8.582	0.163
9	7.827	8.195	-0.368	8.571	-0.744
10	7.294	7.364	-0.070	7.403	-0.109
11	8.260	7.752	0.508	8.252	0.008
12	7.553	7.431	0.122	7.534	0.019
13	8.509	8.377	0.132	8.453	0.056
14	7.719	8.601	-0.882	8.740	-1.021
15	7.355	7.49	-0.135	7.298	0.057
16	7.684	7.866	-0.182	7.766	-0.082
17	9.097	9.003	0.094	9.039	0.058
18	7.75	8.067	-0.317	7.926	-0.176
19	7.593	8.258	-0.665	7.921	-0.328
20	8.602	8.545	0.057	8.639	-0.037
21	8.77	8.614	0.156	8.445	0.325
22	8.31	8.452	-0.142	8.399	-0.029
23	9.155	9.084	0.071	9.139	-0.016
24	8.046	8.774	-0.728	8.722	-0.676
25	8.638	8.644	-0.006	8.694	-0.056
26	7.29	7.131	0.159	7.195	0.095
27	8.523	8.437	0.086	8.625	-0.102
28	8.638	8.883	-0.245	8.638	0.000
29	8.260	8.613	-0.353	8.388	-0.128
30	7.496	7.662	-0.166	7.636	-0.140
31	7.021	7.670	-0.649	7.542	-0.521
32	8.046	8.028	0.018	7.860	0.186
33	8.102	7.639	0.463	7.708	0.394
34	7.42	7.498	-0.078	7.415	0.005
35	7.301	7.349	-0.048	7.306	-0.005
36	7.276	7.305	-0.029	7.352	-0.076
37	8.456	8.397	0.059	8.407	0.049
38	8.149	8.319	-0.170	8.384	-0.235
39	8.886	8.288	0.598	7.944	0.942
40	7.854	7.878	-0.024	7.892	-0.038
41	8.114	8.144	-0.03	8.084	0.030
42	7.721	7.690	0.031	7.678	0.043
43	7.222	7.733	-0.511	7.743	-0.521
44	7.770	7.695	0.075	7.829	-0.059
45	7.886	7.696	0.190	7.829	0.057
46	7.770	7.661	0.109	7.737	0.033
47	7.444	7.479	-0.035	7.391	0.053

Table 2 continued

Compd.	Actual	COMFA		COMSIA	
		Predicted	Residues	Predicted	Residues
48	7.209	6.484	0.725	6.476	0.733
49	7.004	6.963	0.041	6.840	0.164
50	6.112	7.023	-0.911	7.049	-0.937
51	6.767	6.820	-0.053	6.915	-0.148
52	6.750	6.959	-0.209	6.931	-0.181
53	7.726	7.188	0.538	7.144	0.582
54	7.907	7.939	-0.032	7.92	-0.013
55	7.169	7.324	-0.155	7.286	-0.117
56	6.917	6.776	0.141	6.884	0.033
57	6.327	6.237	0.090	6.339	-0.012
58	6.111	6.199	-0.088	6.066	0.045
59	6.873	6.891	-0.018	6.774	0.099
60	6.614	6.641	-0.027	6.567	0.047

For the CoMFA model, the contributions of the steric and electrostatic fields were calculated to be 43.8 and 56.2%, respectively; thus, the electrostatic field has more influence compared to the steric field. For the optimal CoMSIA model, five descriptor fields were considered including the steric, electrostatic, hydrophobic, hydrogen bond-donor, and hydrogen bond-acceptor. Their contributions were 10.6, 30.0, 19.2, 16.7, and 23.5%. Table 2 listed the actual and predicted pIC_{50} values of the training and test set as well as the residues between them.

3D-QSAR contour maps

Through the superposition of the most active molecule **23** with the contour maps generated by CoMFA and CoMSIA, we explored the field effects on the target compounds in 3D space. These contour maps have great significance in explaining the relationship between molecular structure and biological activity because the regions displayed in 3D maps showed the influence of different substituents on the molecular activity.

CoMFA contour maps

The steric and electrostatic contour maps generated by the CoMFA model are shown in Fig. 2. The green polyhedrons represent bulk substituents which are beneficial to the potency, while yellow polyhedrons represent steric bulk groups that would decrease the activity (Fig. 2a).

The medium yellow contour occurring at single side of the aromatic R¹ ring indicated the compounds with bulk substituents at this site would decrease biological activity. The compound **38** ($pIC_{50} = 8.149$) with cyclopropyl group at *ortho*-site of R¹ aromatic ring possesses lower biological

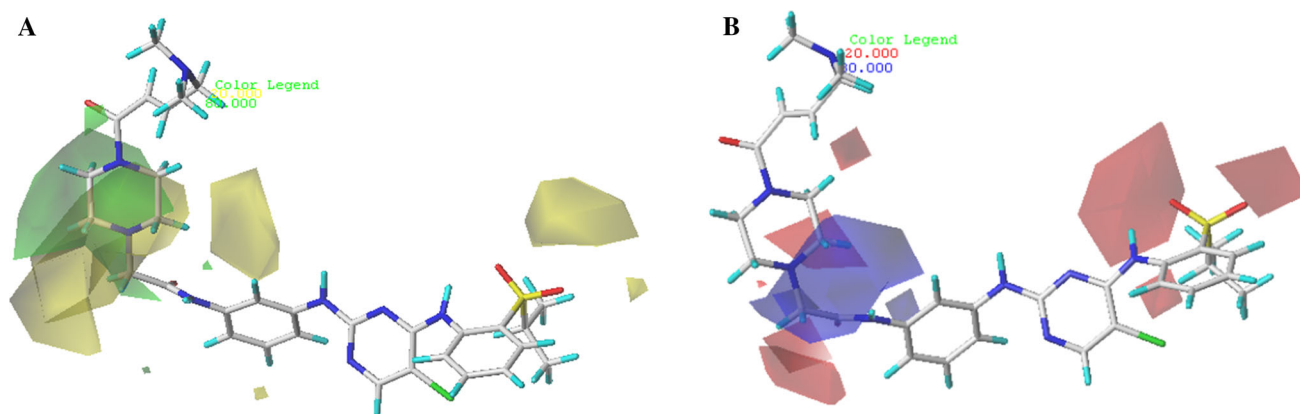


Fig. 2 CoMFA contour maps. Compound **23** is shown inside. **a** Steric field: favored (*green*) and disfavored (*yellow*). **b** Electrostatic field: electropositive (*blue*) and electronegative (*red*) (color figure online)

activity than the corresponding methyl and isopropyl group substituted compounds **37** ($pIC_{50} = 8.456$) and **39** ($pIC_{50} = 8.886$). In contrast, a big negative steric (yellow) and a big positive steric (green) emerged at the *meta*-site of the aromatic ring connected to R^2 . In addition, medium yellow contours appeared above of R^2 position, which suggested that a bulky substituent in this region would decrease the biological activities of the molecule, such as **8** ($pIC_{50} = 8.745$), ($pIC_{50} = 7.827$), and **10** ($pIC_{50} = 7.294$).

The electrostatic contour map of CoMFA is shown in Fig. 2b. The electrostatic field is represented by blue-colored and red-colored contours, in which the blue contours denote that the electropositive groups are favorable to the activity and the red regions indicate that the electronegative groups are positive to the activity. Three red contours at the site of R^1 indicate that electronegative groups in this area are positive to the activity. More electronegative sulfonyl at R^1 of compound **4** ($pIC_{50} = 8.569$) leads to its high activity, while the less electronegative amide group at the same position of compound **3** ($pIC_{50} = 7.484$), respectively, showed inactivity. As shown in the electrostatic contour maps, a big blue color contour encompassed the piperazine ring, which indicates that electropositive groups in this region would be beneficial to the activity. Small red color contours that appear near the methylene connecting the piperazine ring and benzene indicate that negatively charged substituents at this position also have a little influence on the activity.

CoMSIA contour maps

The CoMSIA-generated maps indicate that the presence of a group with a special physicochemical property in the designated area would be beneficial or detrimental to good inhibitory activity. The CoMSIA not only calculated both steric and electrostatic fields the same way as the CoMFA, but also covered hydrophobic, H-bond donor (HBD), and

H-bond acceptor (HBA) fields. Favorable and unfavorable contributions were fixed at 80 and 20%, respectively. Once again, we choose the most active compound **23** to analyze the effects of the five force fields.

Figure 3a displays the steric plot represented by yellow and green color contours. The whole area of R^2 is covered by the yellow color contour, which shows that compounds with bulk substituents in this area would decrease biological activity. The difference between the activities of **1** ($pIC_{50} = 6.553$) and **3** ($pIC_{50} = 7.484$) is due to the presence of small volume of imide in **3**, whereas **1** has a sterically more demanding substitution at this position. The second adverse steric contour was discovered near the R^1 ring indicating the adverse effect of steric bulk, while there are also two green contours at the opposite sites of the yellow polyhedral suggesting that bulky groups were acceptable at this position.

Figure 3b shows the influence of the electrostatic field in the CoMSIA model. The red contour overlapping the amide at the *meta*-site of the R^2 aromatic means that the electronegative groups in this region could improve the inhibition. The large blue contour encompassing the piperazine ring denotes that the electropositive substituents in this area have a positive effect on the molecular activity. For example, the activity of **50** ($pIC_{50} = 6.112$) with methyl was less potent than compound **52** ($pIC_{50} = 6.750$), which possesses an ester at this position.

In Fig. 3c, yellow and gray color contours represent the effect of hydrophobicity on the molecular activity. One yellow color contour can be seen covering the methylene and ketone located in the *meta*-site of the R^2 ring, which suggests that hydrophobic groups in this region contribute to the enhancement of the inhibition. But an equal volume of gray color contour also can be seen near the piperazine ring. In general, both of the hydrophobic favored and hydrophilic contours emerge at the same area indicating that the two groups are in equilibrium in this region. For

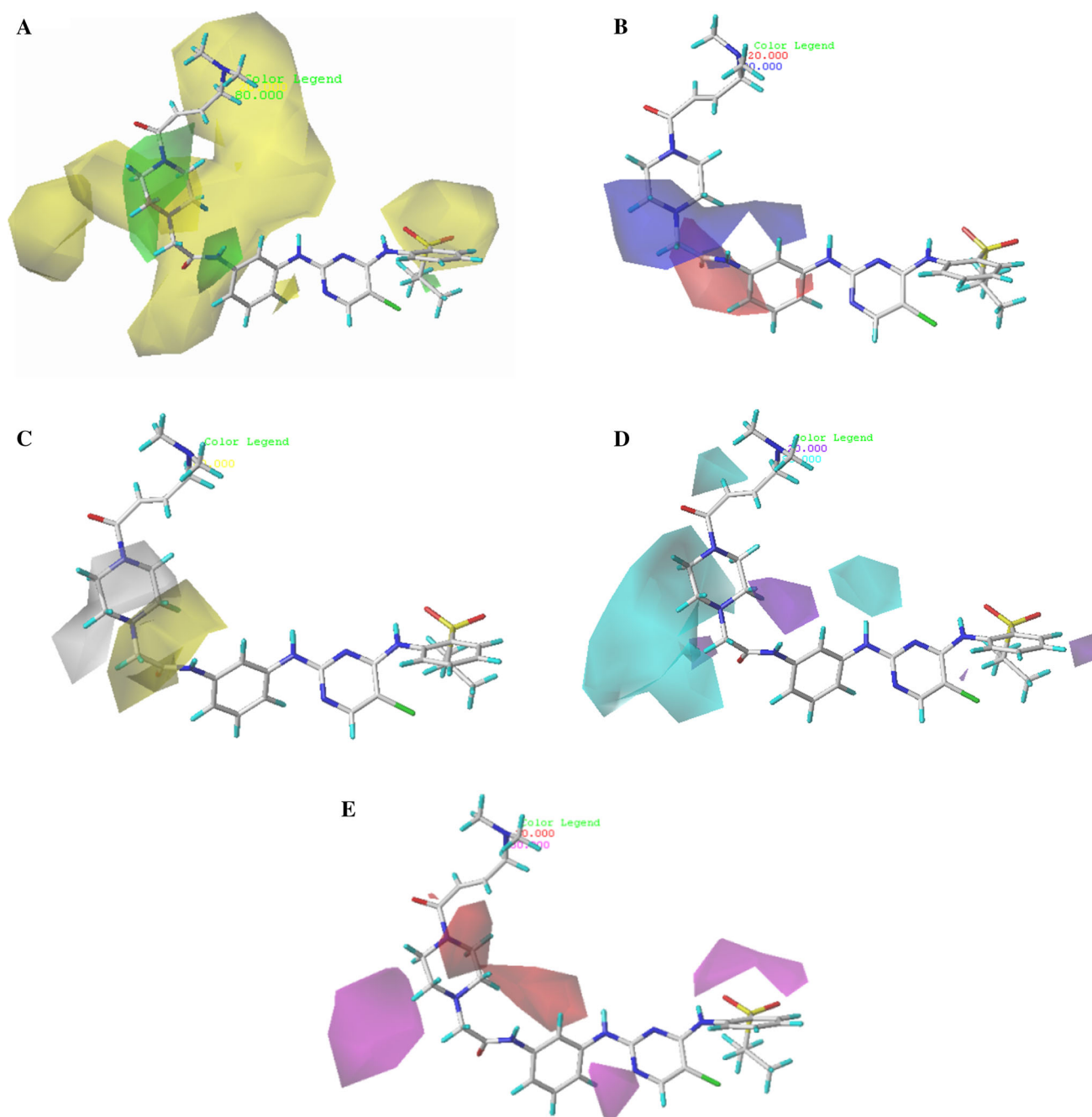


Fig. 3 CoMSIA contour maps. Compound **23** is shown inside. **a** Steric field: favored (green) and disfavored (yellow). **b** Electrostatic field: electropositive (blue) and electronegative (red). **c** Hydrophobic field:

example, both values of **51** ($pIC_{50} = 6.767$) and **52** ($pIC_{50} = 6.750$) are basically the same.

Hydrogen bond donor (HBD) groups represented by cyan (favorable) and purple (unfavorable) contour maps are shown in Fig. 3d. From the contour map of the HBD, a large cyan color contour can be seen in the vicinity of the piperazine ring. For example, in the most active molecule **23**, there is an imine between R^1 and piperazine which could form H-bonds with residues of the protein, indicating

that a hydrogen atom in this position is favorable to the activity of molecule. The alignment of the blue polyhedron has a small purple contour, which indicates that hydrogen bond acceptor groups have little effect on the molecular activity in this position.

Figure 3e illustrates the effects of a hydrogen bond acceptor (HBA) in the CoMSIA model. Magenta color (80% contribution) and red color (20% contribution) contours, respectively, representing the HBA are favorable or

unfavorable to the biological activity. One large volume of magenta color contour is discovered near the R¹ ring indicating that the presence of sulfonic groups in this area could act as HBA attacking protons, which further showed that HBA in this position was conducive to improving inhibitory activity. Another magenta color contour emerges below the piperazine ring, while there are two red contours at the opposite position, which suggests a balance between H-bond donor and H-bond acceptor in the same region.

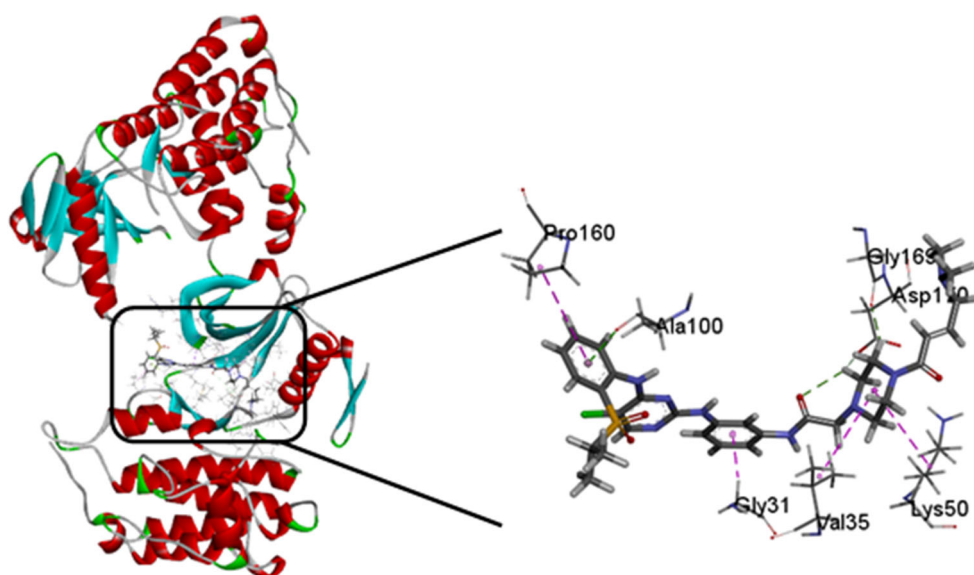
In summary, the structural characteristics for better inhibitory activities from the above-mentioned contour analysis of CoMFA and CoMSIA models are:

1. R¹: medium-sized and electronegative substituents at R¹ site, and hydrogen bond acceptor (favorable).
2. R²: bulk substituents, electropositive and hydrophobic substituents at *meta*-site of R² aromatic ring, and hydrogen bond donor near the piperazine ring (favorable).

Docking analysis

Through molecular docking, we found that the activity of molecules is related to the free energy changes in the process of binding with the protein. The most active inhibitor **23** was selected to dock with the ALK-4DCE protein and the results explain the interaction mechanism between the ligand and the receptor, which is shown in Fig. 6. The benzene ring among the common skeleton formed a π -alkyl interaction with Pro160 (2.45 Å). On the other hand, the benzene ring at R² forms a π - σ interaction with Gly31 (2.42 Å). Alkyl hydrophobic interactions form between the piperazine ring and Val35 (5.23 Å) and Lys50 (4.73 Å), which coincides with the hydrophobic contour map

Fig. 4 Docking result of the representative ligand **23** into the binding site of the ALK protein. Ligands and the important residues for binding interaction are depicted by *stick* and *line* models

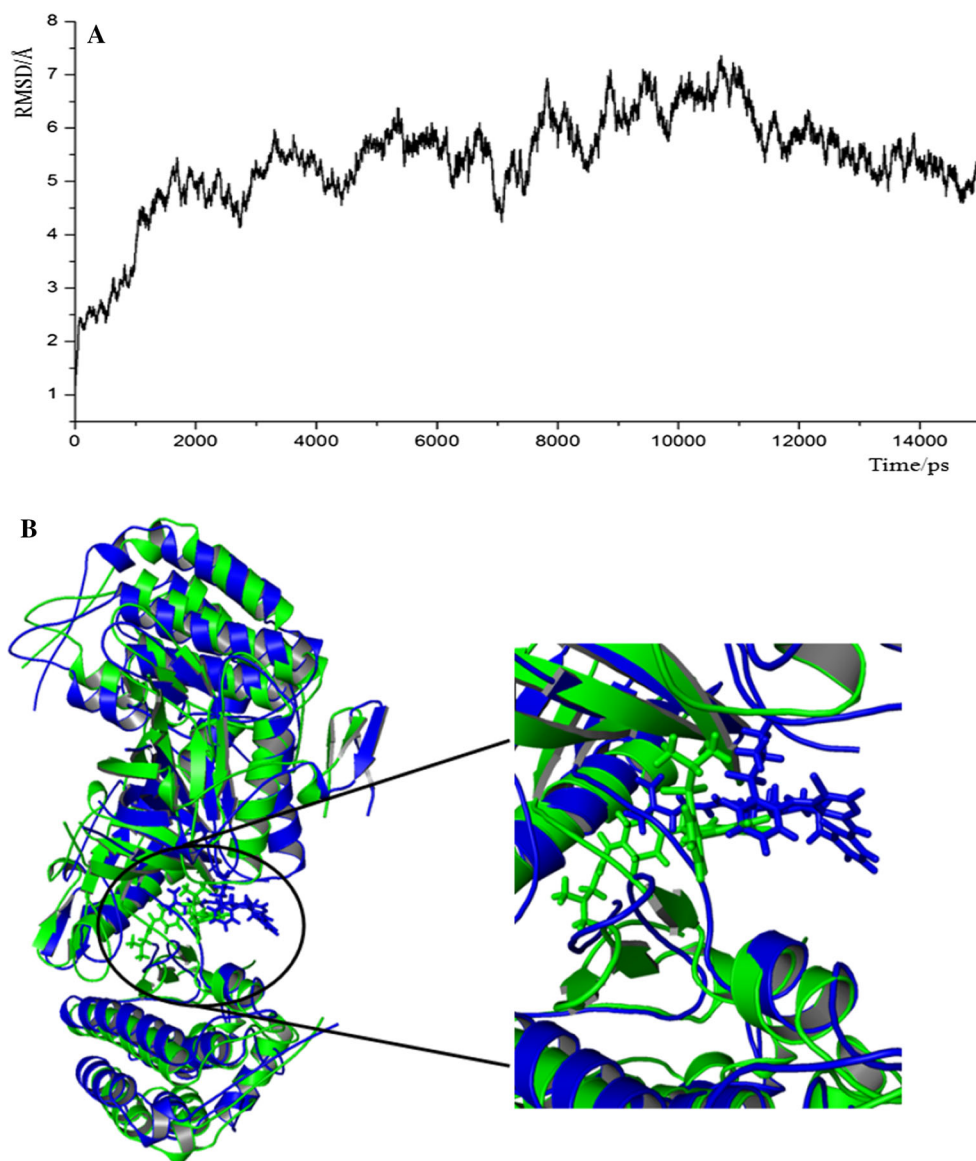


depicted in Fig. 3e. There is a large yellow contour around the benzene ring indicating that the introduction of hydrophobic groups in this region is beneficial for the inhibitory activity. According to the docking results, the large pocket composed of Asp170, Gly169, Lys50 as displayed in Fig. 4 is sufficient for medium bulky substituents. The green contour generated by the COMFA model at this position also verified this conclusion. However, large bulky substituents in this position would lead to steric hindrance with the surrounding amino acids, which would lead to a reduction in activity. Therefore, the introduction of too large groups in this position is detrimental, which is consistent with the steric contour map in the COMSIA model of Fig. 2b. The results of molecular docking show that most of the inhibitors have a similar binding pattern at the active site of ALK.

Molecular dynamics simulation

In order to further verify the models of 3D-QSAR and molecular docking, we applied MD simulations to establish a more reliable mechanism to illustrate interactions between ligand and receptor. The basic theory of MD simulations is given by the molecular system initial motion state and the natural motion of the molecules in the phase space. A 15 ns simulation was run to obtain a stable conformation of ligand–receptor complex in this study was shown in Fig. 5a. The RMSDs of the trajectory with respect to the initial structure ranged from 2.5 to 3.0 Å. After 2 ns, the RMSDs of the complex reached about 5.3 Å and maintained a similar value in the following simulation, which indicated that the docked complex could reach metastable conformation after 2 ns of simulation. A superposition of the lowest-energy structure extracted from the MD simulation (blue)

Fig. 5 **a** Plot of the root-mean-square deviation (RMSD) of docked complex versus the MD simulation time in the MD-simulated structures. **b** View of superimposed backbone atoms of the lowest-energy structure of the MD simulation (*blue*) and the initial structure (*green*) for the **23**/4DCE complex



and the initial structure (green) for the **23**-4DCE complex are shown in Fig. 5b. Through the analysis of the interactions between **23** and the receptor after MD simulation, we explored the similarities and differences between molecular docking and MD simulation. Figure 6 shows the lowest-energy structure extracted from the MD simulation, from which we can see that it mainly forms three hydrogen bonds between the ligand and the receptor. The amide oxygen at the *meta* position of the benzene ring forms a hydrogen bond with NH of His32 ($-C=O \cdots HN-$, 2.45 Å). The oxygen atom of the sulfonic group acts as a hydrogen bond acceptor to form a H-bond with NH of Gly31 ($-C=O \cdots HN-$, 1.77 Å), which is consistent with the H-bond acceptor contour depicted in Fig. 3e. There is a large magenta contour near the R¹ position, which indicates that H-bond acceptor groups in this area are favorable for the inhibitory activity. Residues Gly31, Val35, and His32 form hydrophobic

contacts with the ligands, which are beneficial for the inhibitory activity. On the other hand, the benzene ring at the R² forms a π -stacking bond with His32 which further strengthens the correlation between the inhibitory and the receptor.

Conclusion

QSAR and molecular dynamics were applied to analyze and explore characteristics of DAAP analogues as ALK inhibitors. The CoMFA and CoMSIA models nicely explained the intermolecular interactions between the inhibitors and the surrounding environment. Docking and molecular dynamics studies demonstrated that hydrogen bond formed between the inhibitors and ALK-4DCE protein play an important role in activity of the inhibitors. In

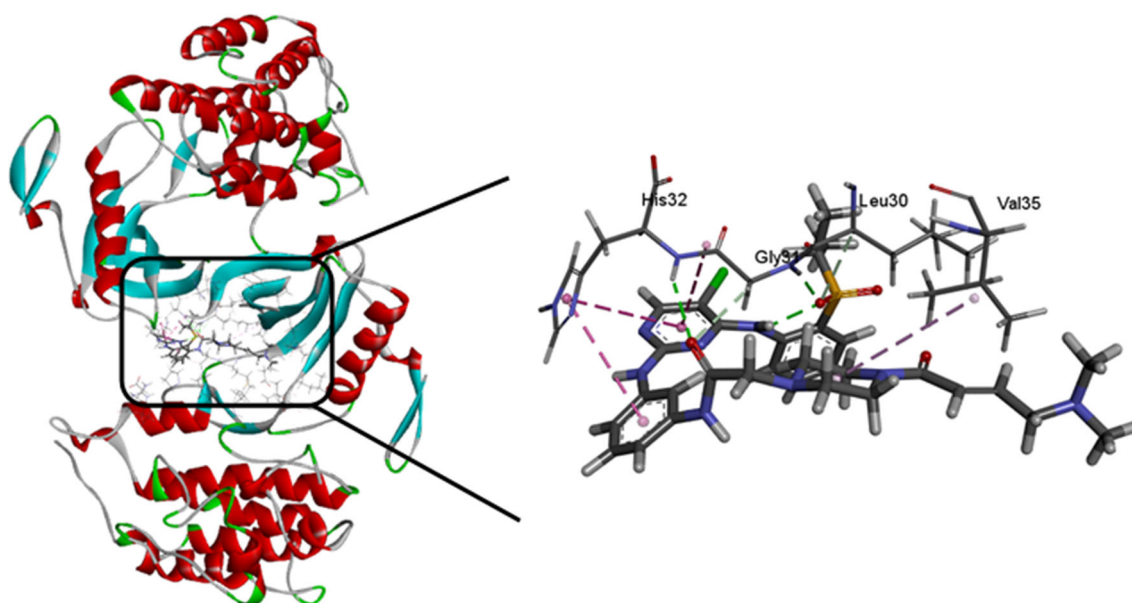


Fig. 6 Plot of the MD-simulated structure of the binding site with the ligand. Compound **23** in the complex is in the active site of the ALK enzyme. Active site amino acid residues are represented as *sticks*; the inhibitor is shown as *stick and ball* model

addition, the MD simulation results are consistent with the results of QSAR models and molecular docking in terms of the reliability and stability of the derived models. Some key residues (His32, Gly31, Gly169, Asp170, Val35, Ala100, Pro160, Lys50, and Leu30) and three hydrogen bonds (His32, Gly31, and Leu30) were discovered in the binding site, which indicated that the model could provide guidance for further research in the development of new ALK inhibitors.

Materials and methods

Dataset and biological activity

Sixty DAAP analogues involved in this work were reported by Ao Zhang and co-workers [16, 24, 28]. The range of IC_{50} values for these compounds was 0.7–775 nM. The bioactivities of the derivatives were expressed as pIC_{50} ($= -\log IC_{50}$) values. The samples were divided into a training set of 44 molecules for model generation and a test set of 15 molecules for model validation at a ratio of 3:1. The structures and activity values of each molecule used in the study are shown in Table 3. The test molecules were selected randomly such that the dataset showed high structural diversity and a wide range of activities [30].

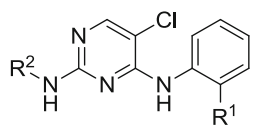
Molecular modeling and alignment procedure

CoMFA and CoMSIA models were all performed using the SYBYL-X 2.0 software. All molecules were loaded with

Gasteiger–Hückel charges and optimized by using the Tripos force field [31] with Powell energy gradient algorithms at a convergence criterion of 0.02 kJ/mol Å and a maximum of 1000 iterations [32]. Table 3 lists the common scaffold of the samples, various substituents, and the IC_{50} value of each molecule. Molecular alignment was the most critical step in the establishment of the CoMFA and CoMSIA models, which needed to analyze the three-dimensional structure of the samples to find a suitable conformational template for alignment [33]. Since the molecules share a common structure, it was assumed that each molecule binds into the active site of protein in a similar way. In this context, we adopted the rigid body alignment rule. Compound **23**, which had the highest pIC_{50} (9.155), was selected as the template molecule of DAAP derivatives. In the end, the program automatically superposed all the molecules and then the database was updated to a new molecular library with new orientation [32]. Alignment of training and test set compounds is shown in Fig. 7. The common substructure is depicted in bold.

CoMFA and CoMSIA

In order to build a reliable 3D-QSAR model, the partial least squares (PLS) method was carried out based on the above alignment of molecules [34]. To find the best models, we calculated various parameters which were used to evaluate and analyze the robustness and predictive ability of these models, including the internal validations of LOO cross-validated q^2 , non-cross-validated coefficient r^2 , standard error of estimate (SEE), and F statistic values.

Table 3 Structures and activity values of the DAAP molecules [16, 24, 28]

Compound	Substituent		IC_{50}/nM
	R^1	R^2	
1			280
2			80.3
3			32.8
4 ^{TS}			2.7
5			144
6			94.5
7			7.7
8			1.8
9 ^{TS}			14.9
10			50.8
11 ^{TS}			5.5
12			28.0
13			3.1
14 ^{TS}			19.1

Table 3 continued

15			44.2
16			20.7
17			0.8
18			17.8
19 ^{TS}			25.5
20			2.5
21			1.7
22			4.9
23			0.7
24 ^{TS}			9.0
25			2.3
26			51.3
27			3.0
28			2.3
29 ^{TS}			5.5

Table 3 continued

30			31.9
31 ^{TS}			95.3
32			9.0
33 ^{TS}			7.9
34			38
35			50
36			53
37			3.5
38			7.1
39 ^{TS}			1.3
40			14
41			7.7
42			19
43 ^{TS}			60
44			17

Table 3 continued


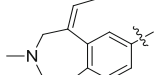

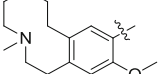

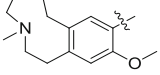

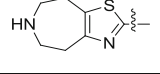

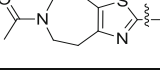

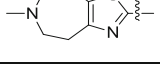

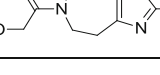

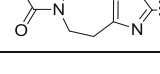

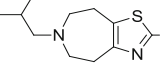

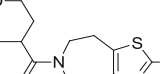

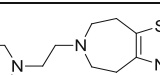

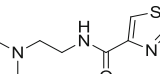

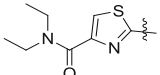

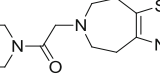

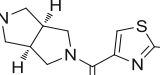

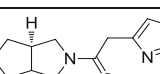
45 ^{TS}			13
46			17
47			36
48 ^{TS}			61.8
49			99
^a 50 ^{TS}			773
51			171
52			178
53 ^{TS}			18.0
54			12.4
55 ^{TS}			67.8
56			121
57			471
58			775
59			134
60			243

Fig. 7 Molecular alignments of all compounds in the dataset. Compound **23** was used as the template for alignment

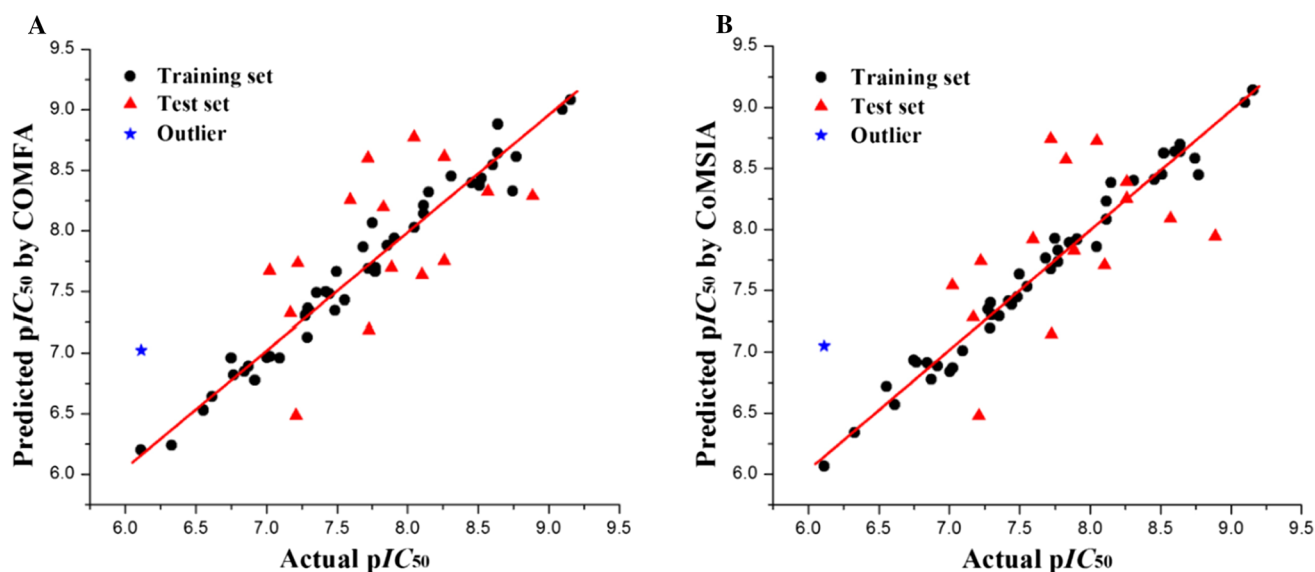
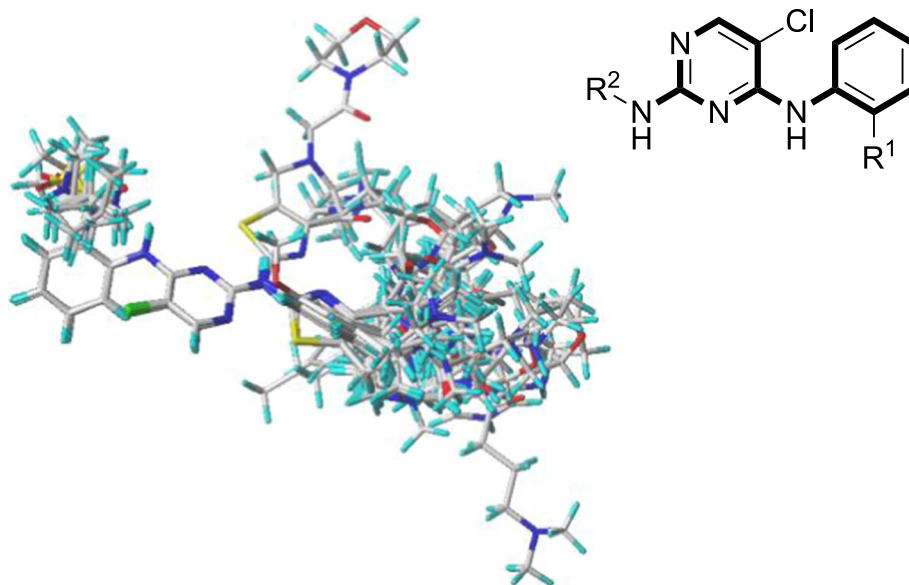


Fig. 8 Plots of predicted versus actual $pI_{C_{50}}$ values for all the molecules based on CoMFA (a) and CoMSIA models (b)

According to these statistical results, the final models were established. The statistical results of the CoMFA and CoMSIA models are summarized in Table 1. We used the default settings in SYBYL in the optimization process of the CoMFA and CoMSIA descriptors [35]. Figure 8 showed a linear relationship between the predicted and true values calculated by the CoMFA and CoMSIA models.

Molecular docking

Molecular docking is an important computational chemistry tool with clear and intuitive definition. The structure and binding energy of protein–ligand complexes can be found in the case of known protein and ligand space

structures [36]. We applied the SYBYL-X 2.0, which is based on a prototype to explore more information on the binding mode of ligand and ALK protein. The crystal structure of the ALK protein complex was obtained from the RCSB Protein Data Bank (PDB entry code: 4DCE). After the extraction of ligands, removal of water molecules, and hydrogenation, a prototype was generated by using the ligand extraction method for molecular docking. The energy minimization of the protein structure was performed by applying the Tripos force field, and partial atomic charges were calculated by means of the Gasteiger–Hückel method. The protein interaction of the ligand was visualized by using Discovery Studio Visualizer 2.5 (Accelrys Software Inc.), which provided a molecular

modeling environment for both small molecule and macromolecule.

Acknowledgements We are grateful for financial support from the National Natural Science Foundation of China (21172148, 21472126, and 21502117) and Shanghai Municipal Education Commission (Plateau Discipline Construction Program).

References

1. Achary R, Yun JI, Park CM (2016) *Bioorg Med Chem* 24:207
2. Le Beau MM, Bitter MA, Larson RA (1989) *Leukemia* 3:866
3. Gleason BC, Hornick JL (2008) *J Clin Pathol* 61:428
4. Lee HW, Kim K, Kim W, Ko YH (2008) *Hematol Oncol* 26:108
5. Debelenko LV, Raimondi SC, Daw N (2011) *Mod Pathol* 24:430
6. Soda M, Choi YL, Enomoto M (2007) *Nature* 448:561
7. Caren H, Abel F, Kogner P, Martinsson T (2008) *Biochem J* 416:153
8. Mosse YP, Laudenslager M, Longo L (2008) *Nature* 455:930
9. Ren H, Tan ZP, Zhu X (2012) *Cancer Res* 72:3312
10. Katayama R, Friboulet L, Koike S (2014) *Clin Cancer Res* 20:5686
11. Choudhari R, Minero VG, Menotti M (2016) *Blood* 127:1297
12. Moritake H, Shimonodan H, Marutsuka K (2011) *Am J Hematol* 86:75
13. Wynne C, Lazzari E, Smith S (2014) *PLoS One* 9:e101503
14. Tort F, Pinyol M, Pulford K (2001) *Lab Invest* 81:419
15. Vyas VK, Patel A, Gupta N, Ghate M (2014) *Med Chem Res* 23:603
16. Liu ZQ, Yue XH, Song ZL (2014) *Eur J Med Chem* 86:438
17. Marsilje TH, Pei W, Chen B (2013) *J Med Chem* 56:5675
18. Christensen JG, Zou HY, Arango ME (2007) *Mol Cancer Ther* 6:3314
19. Kinoshita K, Asoh K, Furuichi N (2012) *Bioorg Med Chem* 20:1271
20. Katayama R, Khan TM, Benes C (2011) *Proc Natl Acad Sci USA* 108:7535
21. Galkin AV, Melnick JS, Kim S (2007) *Proc Natl Acad Sci USA* 104:270
22. Lovly CM, Heuckmann JM, Thomas RK (2011) *Cancer Res* 71:4920
23. Ke ZP, Lu T, Liu HC (2014) *J Mol Struct* 1067:127
24. Song ZL, Yang YH, Liu ZQ (2015) *J Med Chem* 58:197
25. Cheng LP, Huang XY, Wang Z, Kai ZP, Wu FH (2014) *Monatsh Chem* 145:1213
26. Wang P, Cai J, Chen JQ (2014) *Med Chem Res* 23:2576
27. Lv QQ, Wang ZH, Wang ZH, Wu FH (2015) *Lett Drug Des Discov* 12:219
28. Liu ZQ, Ai J, Peng X, Song ZL (2014) *ACS Med Chem Lett* 5:304
29. Boehm M, Stuerzebecher J, Klebe G (1999) *J Med Chem* 42:458
30. Liu J, Li Y, Zhang HX, Zhang SW (2012) *J Mol Model* 18:991
31. Clark M, Cramer RD III, Van ON (1989) *J Comput Chem* 10:982
32. Wang FF, Yang W, Shi YH, Le GW (2015) *J Biomol Struct Dyn* 33:1929
33. Thaimattam R, Daga PR, Banerjee R, Iqbal J (2005) *Bioorg Med Chem* 13:4704
34. Liu JL, Wang FF, Ma Z, Wang X (2011) *Int J Mol Sci* 12:946
35. Xu C, Ren YJ (2015) *Bioorg Med Chem Lett* 25:4522
36. Kirkpatrick P (2004) *Nat Rev Drug Discov* 3:299