3D-QSAR and Docking Studies on Pyrimidine Derivatives of Second-Generation ALK Inhibitors

Gang-Long Jiang[1,2]   Lian-Hua Song[3]   Yong-Fu Qiu[1,2]   Yu Liu[1,2,*]

[1] Novel Technology Center of Pharmaceutical Chemistry, Shanghai Institute of Pharmaceutical Industry Co., Ltd., China State institute of Pharmaceutical Industry, Shanghai, People's Republic of China

[2] Shanghai Engineering Research Center of Pharmaceutical Process, Shanghai Institute of Pharmaceutical Industry Co., Ltd., China State institute of Pharmaceutical Industry, Shanghai, People's Republic of China

[3] Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, Zhejiang University of Technology, Hangzhou, People's Republic of China

Address for correspondence Yu Liu, PhD, Novel Technology Center of Pharmaceutical Chemistry, Shanghai Institute of Pharmaceutical Industry Co., Ltd., 285 Gebaini Road, Shanghai 201203, People's Republic of China (e-mail: liuyu_tianjin@126.com).

Abstract

Anaplastic lymphoma kinase (ALK) is a promising target for the treatment of non-small cell lung cancer. Under crizotinib treatment, drug resistance and progressive disease appeared after the point mutations arising in the kinase domain of ALK. Second-generation ALK inhibitors can solve the deficiencies of the first generation, especially the drug resistance in cancer chemotherapy. Ceritinib (LDK378), a pyrimidine derivative, for example, can inhibit the activity of ALK with an IC_{50} value of 40.7 nmol/L, and can experience disease progression after initial treatment with crizotinib. Unfortunately, clear structure–activity relationships have not been identified to date, impeding the rational design of future compounds possessing ALK inhibition activity. To explore interesting insights into the structures of pyrimidine derivatives that influence the activities of the second-generation ALK inhibitors, three-dimensional quantitative structure–activity relationship (3D-QSAR) and molecular docking were performed on a total of 45 derivatives of pyrimidine. Comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) techniques were used to generate 3D-QSAR models. CoMFA and CoMSIA were performed using the Sybyl X 2.0 package. Molecular docking analysis was performed using the Surflex-Dock module in SYBYL-X 2.0 package. We found in the CoMFA model that the non-cross-validated r^2 value was 0.998, the cross-validated q^2 value was 0.663, and the F statistic value was 2,401.970, while the r^2 value was 0.988; q^2 value was 0.730, and F value was 542.933 in CoMSIA models, suggesting the good predictability of the CoMFA and CoMSIA models. 3D contour maps and docking results suggested that different groups on the core parts of the compounds could enhance the biological activities. Based on these results, the established 3D-QSAR models and the binding structures of ALK inhibitors obtained favor the prediction of the activity of new inhibitors and will be helpful in the reasonable design of ALK inhibitors in the future.

Keywords
► 3D-QSAR
► docking
► second-generation ALK inhibitors
► CoMFA
► CoMSIA

Introduction

Anaplastic lymphoma kinase (ALK), a receptor tyrosine kinase that belongs to an insulin receptor superfamily, is responsible for the development of different tumor types. According to the current study, ALK has not been found to have an essential physiological role in the normal body, and ALK fusion proteins have been connected with a wide spectrum of human cancers.[1–3] Evidence suggested that ALK fusion gene has been mainly expressed in the central nervous system and other parts of the brain and its expression levels sharply decline after birth.[4] Although not widely expressed in adult tissue, ALK is implicated in neuronal development, differentiation, and basal dopaminergic signaling.[5] Interestingly, ALK fusion gene also plays a key role in the development of various cancers, such as non-small cell lung cancer (NSCLC), breast cancer, ovarian cancer, and inflammatory myofibroblastic tumors.[6–8] ALK participates in the regulation of cell proliferation associated the PLC-γ and Ras/Raf/MEK/ERK1/2 pathways, and cell survival associated the JAK/STAT and PI3K/Akt (PKB) pathways.[9–11] These different pathways can be activated by ALK fusion proteins, which are interconnected and overlapping. Thus, many novel potent ALK inhibitors have been reported just like mushrooming in recent years.[12]

The echinoderm microtubule-associated protein-like 4 (EML4)-ALK fusion-type tyrosine kinase is an oncoprotein found in NSCLC, and has been validated as a novel therapeutic target by Pfizer's first-generation ALK inhibitor crizotinib (PF2341066, Xalkori), which was approved by Food and Drug Administration in 2011 as the standard treatment for ALK-positive NSCLCs. However, despite the good clinical effect of crizotinib, subsequent treatment also resulted in mutation L1196M,[13–17] which is similar to gatekeeper mutations in the EGFR-T790M and ABL-T315I kinase domains.[18,19] Therefore, there is an urgent need for a new generation of ALK inhibitors that are able to overcome drug-resistant mutants such as L1196M. Recently, the potent and high selectivity of the second-generation ALK inhibitors including X-396, alectinib (CH-5424802), ASP3026, AP26113, and ceritinib (LDK378) have been found. These inhibitors exhibit good biological activity, but are not enough for the possible resistance mutations,[20,21] Given above, we can learn from both of the drug resistance of the first-generation ALK inhibitors of crizotinib and the better treatment effect of the second-generation ones, by means of computational techniques, to in-depth study the structure–activity relationship of the second-generation ALK inhibitors.

The ALK inhibitors developed recently are shown in ►Fig. 1. In the present study, a computer-aided drug design was used to set three-dimensional quantitative structure–activity relationship (3D-QSAR) models using 45 derivative molecules with anti-ALK ability.[22–24] Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) models based on ligand alignment were made. Both the contour maps of CoMFA and CoMSIA revealed the key factors affecting the activities of the inhibitors, and guided us to design new potent ALK inhibitors. This study may provide a theoretical basis for the design of ALK inhibitors in the future.
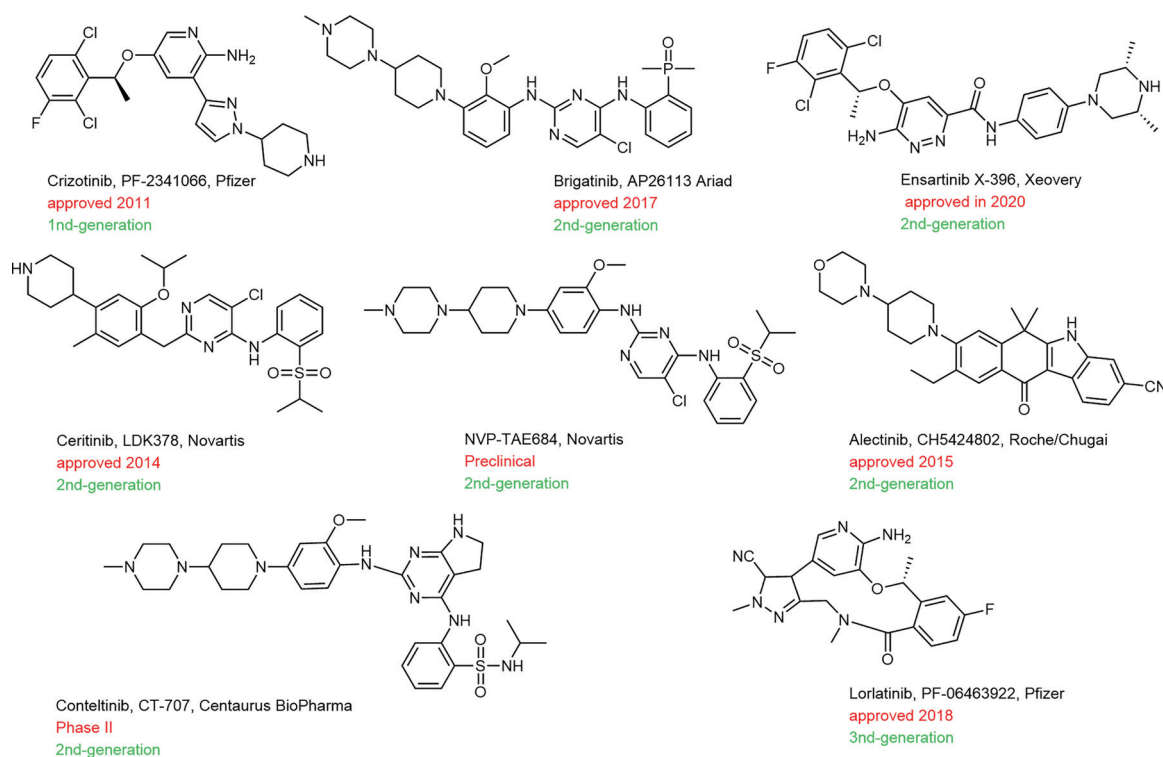

Fig. 1 Representative ALK inhibitors. ALK, anaplastic lymphoma kinase.

Materials and Methods

Data Sets

A total of 45 molecules of pyrimidine derivatives taken from the literature were used to build 3D-QSAR models. All the data were divided into a training set comprising 39 molecules (87%; compounds **1**–**4**, **6**–**15**, **17**–**34**, **36**–**38**, **40**, and **42**–**45**) for generating a 3D-QSAR model and a test set comprising 6 (13%; compounds **5**^{T}, **13**^{T}, **16**^{T}, **35**^{T}, **39**^{T}, and **41**^{T}) molecules for validating the QSAR models' quality.

The 3D structures of the compounds were constructed using the sketch module in Sybyl X 2.0 package. Partial atomic charges were calculated by the Gasteiger–Huckel method before energy minimization using the Tripos force field with a convergent threshold of a maximum deviation of 0.01 kcal/(mol Å).[25,26] Molecular alignment is the most sensitive factor for building reliable 3D-QSAR models. The biological activities were measured by calculating the concentration of the compound responsible for 50% of the maximum inhibition (IC_{50}) and were converted into pIC_{50} ($-logIC_{50}$) according to a reported study.[27] The chemical structures of compounds **1**–**45** along with their biological activities are given in ►**Table 1**.

Molecular Docking Studies

The diversity of the datasets containing different backbones and the molecular alignment, which are obtained through molecular docking, is crucial for developing CoMFA and CoMSIA models. Molecular docking was manipulated using the Surflex-Dock module in SYBYL-X 2.0 to explore the interactions between the binding sites of ALK (PDB code: 4MKC from the RCSB protein data bank) and molecules. After deleting the ligands and water molecules in the ALK-binding site, the ALK inhibitors (compounds **1**–**45**) with electric charges and hydrogen atoms were butt jointed with the ALK-binding pocket. The partial atomic charges of all inhibitors were calculated using Gasteiger–Huckel charge before docking. The minimization of each structure was also performed using the Tripos force Field. Other parameters were set to default and each docked molecule produced 20 conformations. The docking total scores of compounds **1**–**45** are shown in ►**Table 2**.

The docking results of compounds **1**–**45** and the binding pocket of ALK (PDB code: 4MKC) are shown in ►**Figs. 2** and **3**, respectively. Template molecule **21** (yellow) was docked into the binding pocket of ALK. Hydrogen bonding interactions are shown with dotted yellow lines. These images were generated using the Sybyl-x 2.0 program.

Next, according to the maximum similarity and the direction of stretching in the pocket, the best conformations of compounds **1**–**45** for molecular modeling alignment were selected for generating the best QSAR model.

Molecular Modeling Alignment

Molecules of the training sets and text sets were aligned onto compound **21** which possessed the highest activity (►**Fig. 2**). The alignments will be used for the further 3D-QSAR study. The 3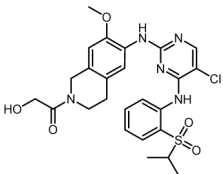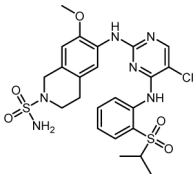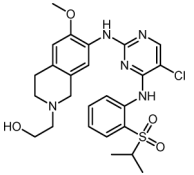9 molecule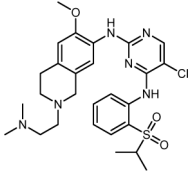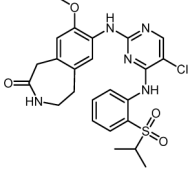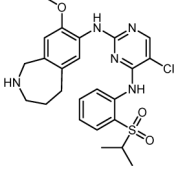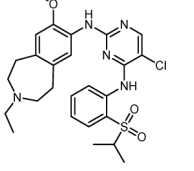s of the training set were aligned onto the high potent template molecule **21** (IC_{50} = 0.07 nm, pIC_{50} = 10.15). The alignment results and common substructures are shown in ►**Fig. 3**. The substructure had the largest basic structure and kept the molecules' space orientation consistent with the protein pocket of 4MKC. As shown in ►**Fig. 3**, the 39 molecules of the training set are mainly different in the piperidine ring direction, which is shown in previous reports that the ligands make hydrogen bonds at the hinge area via the pyrimidine and amino nitrogen atoms onto the backbone nitrogen and oxygen of Met1199, respectively.[28] The central pyrimidine ring of these 39 training set molecules is sandwiched between Ala1148 and Leu1256, and its chlorine substituent is directed toward the back of the pocket, making hydrophobic contacts with gatekeeper Leu1196. The piperidine ring extends to the solvent and is engaged in a salt bridge with Glu1210. The sulfonyl makes an intramolecular H-bond.[28,29] The 39 training set molecules aligned onto the high potent template molecule **21** extend almost outside the pocket, which can core likely to study the effect of the substituents at this position on the molecular activity, and the terminal piperidine fits closely to the protein surface and makes a salt bridge with E1210. As discussed earlier, the alignment of compound structures plays a key role in developing successful 3D-QSAR models. Our effort is to study the effect of different substituents on the piperidine ring and the orientation of the substituents on the activity of the ALK inhibitor. Hence, the docked poses of the ligands were used to develop receptor-based 3D-QSAR models.

Construction of the 3D-QSAR Model

QSAR and 3D-QSAR had a profound impact on medicinal chemistry.[30] CoMFA is a promising new approach to structure/activity correlation, its characteristic features are a representation of ligand molecules by their steric and electrostatic fields, combining the method sampling at the intersections of a 3D lattice, optimal mutual alignment within a series, minimizing the root mean square field differences between molecules, analyzing data by partial least squares (PLSs), using cross-validation to maximize the likelihood that the results have predictive validity, and graphic representation of results, as contoured 3D coefficient plots. However, comparative molecular similarity index analysis (CoMSIA) is more comprehensive than CoMFA because of other two fields' analysis, hydrogen bond receptor and hydrogen bond donor.[31] Residual of actual and predicted pIC_{50} values is an important factor for CoMFA, CoMSIA, and 3D-QSAR models.

An ideal way to access the robustness and predictive ability of 3D-QSAR models is to estimate the performance of the models on a validation set of compounds which were not used in model constructions.[32] The observed activity and calculated activity are shown in ►**Table 3** based on the prediction of the models, which aligned with the 39 molecules of the training set. The best CoMSIA model was developed using four descriptor fields, which were steric, electrostatic, hydrophobic, and hydrogen-bond acceptor fields.

Table 1 Compounds **1–45** and their biological activities for ALK target

Compound	Structure	IC$_{50}$ (nmol/L)	pIC$_{50}$
1		1	9.00
2		7	8.16
3		2	8.70
4		16	7.80
5 ^{T}		5	8.30
6		5.2	8.28
7		1.7	8.77
8		1000	6.00

(*Continued*)

Table 1 (*Continued*)
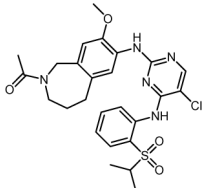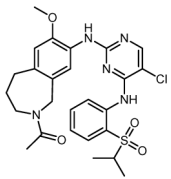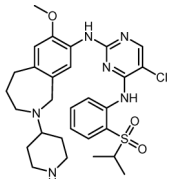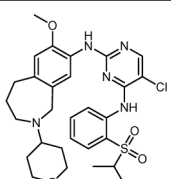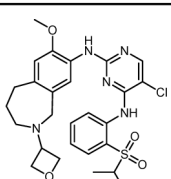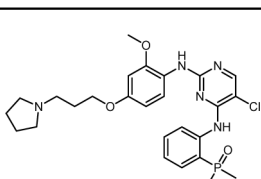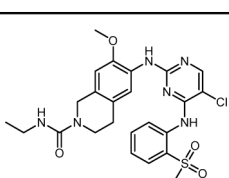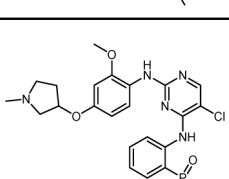
Compound	Structure	IC$_{50}$ (nmol/L)	pIC$_{50}$
9		1900	5.72
10		800	6.10
11		2500	5.60
12		1500	5.82
13[T]		3600	5.44
14		0.22	9.66
15		2	8.70
16[T]		0.33	9.48

Table 1 (*Continued*)

Compound	Structure	IC$_{50}$ (nmol/L)	pIC$_{50}$
17		3.1	8.51
18		0.18	9.74
19		0.72	9.14
20		0.33	9.48
21		0.07	10.15
22		0.28	9.55
23		0.49	9.31
24		0.37	9.43

(*Continued*)

Table 1 (*Continued*)
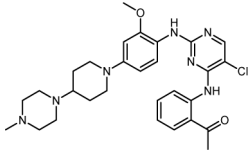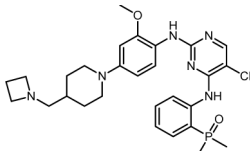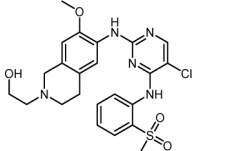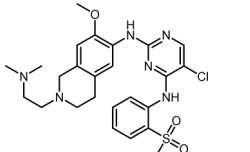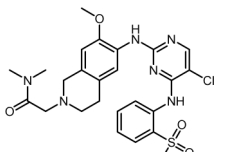
Compound	Structure	IC$_{50}$ (nmol/L)	pIC$_{50}$
25		0.49	9.31
26		2	8.70
27		0.76	9.12
28		0.59	9.23
29		0.38	9.42
30		1.19	8.92
31		0.12	9.92
32		0.86	9.07

Table 1 (*Continued*)

Compound	Structure	IC$_{50}$ (nmol/L)	pIC$_{50}$
33		0.75	9.12
34		0.18	9.74
35[T]		0.68	9.17
36		1.36	8.87
37		1.86	8.73
38		1.88	8.73
39[T]		4.77	8.32
40		2000	5.70

(*Continued*)

Table 1 (Continued)

Compound	Structure	IC$_{50}$ (nmol/L)	pIC$_{50}$
41^T		1.59	8.80
42		0.88	9.06
43		1	9.00
44		18	7.74
45		3	8.52

Note: T = test set. Red section is the common substructure of compounds **1–45**.

The 3D-QSAR methods rely on the principle that the 3D geometric and electronic features of molecules correlate with their biological activities.[33] Furthermore, 3D-QSAR also plays an important role in the optimization of pharmacologically active compounds and in the prediction of the biological activities of newly designed compounds.[34] The predicted pIC$_{50}$ values for the training and test set compounds are listed in ►**Table 3** and depicted graphically in ►**Fig. 4**, from which it can be seen that all points are located near.

Results and Discussion

CoMFA and CoMSIA Analysis
CoMFA and CoMSIA models were derived from a training set of 45 compounds with pIC50 values ranging from 5.44 to 10.15 (►**Table 3**). The PLS statistical parameters of CoMFA and CoMSIA models are listed in ►**Table 4**, which are useful tools to derive multilinear relationships between independent and dependent variables, together with a cross-validation test were used for the generation and internal validation of the CoMFA and CoMSIA models.[35] As ►**Table 4** shows, CoMFA and CoMSIA models were developed, and the final models were selected according to the statistical parameters.

PLS analysis on all of the compounds in the training set resulted in a CoMFA model with a cross-validated q^2 of 0.663. This model gave an optimal number of components of 8 and a conventional correlation coefficient r^2 of 0.998. The corresponding steric and electrostatic field descriptors explained 47 and 52% of the total variance. For CoMSIA analysis, five descriptor fields (steric, electrostatic, donor, acceptor, and hydrophobic) were considered, and r^2 (0.988) and q^2 (0.732) values were obtained with a satisfactory result.

CoMFA and CoMSIA Statistical Result
It is important to make an initial inspection of the inhibitor molecules before establishing the 3D-QSAR models. Statistically, the r^2 value > 0.3 of the predicted set is usually considered significant, while the r^2 value > 0.5 is statistically more significant in CoMFA and CoMSIA studies.[36] As listed in ►**Fig. 4(A, B)**, they show a good line fit. The reason for this outlier may be different structure or different binding

Table 2 Docking total scores of compounds **1–45** by Surflex-Dock using Sybyl-X 2.0

Compounds	Total score	Compounds	Total score
1	8.5401	24	8.6672
2	8.1256	25	9.1738
3	9.1248	26	8.7432
4	9.2192	27	9.1253
5[T]	8.0458	28	8.1218
6	7.9657	29	9.4808
7	8.2245	30	8.4022
8	7.8364	31	9.5286
9	7.5511	32	8.4212
10	8.0818	33	9.4346
11	8.3982	34	9.1064
12	7.6292	35[T]	10.2484
13[T]	8.6261	36	7.4582
14	9.4632	37	7.7205
15	8.4694	38	7.3739
16[T]	9.7901	39[T]	8.3825
17	8.9787	40	8.5255
18	9.4412	41[T]	8.5025
19	8.3258	42	8.6453
20	8.5517	43	8.3596
21	9.1950	44	6.9046
22	8.8529	45	8.6297
23	8.7688		

Fig. 2 Compounds 21 (template molecule), 35 (the highest total score), and 4MKC (ligand of 4MKC) were docked and superimposed in the pocket of crystalline complex of 4MKC, hydrogen bond (yellow).
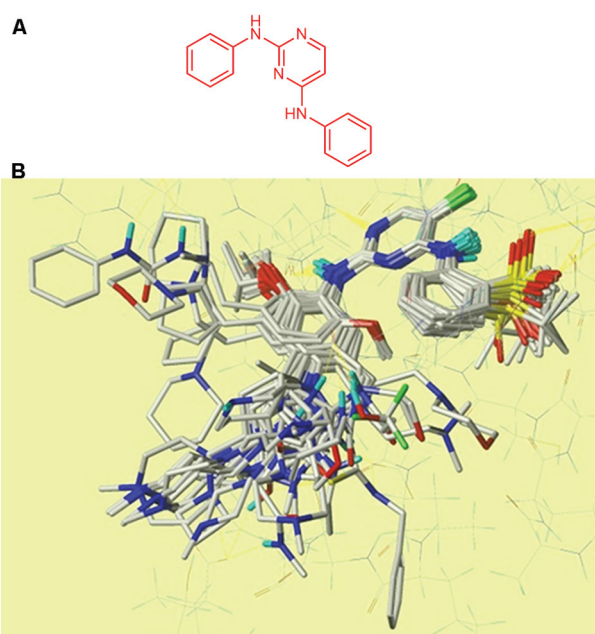
Fig.3 Molecular alignment. (A) Common sub-structure of compounds **1–45**. (B) 39 molecules of the training sets aligned on the template molecule.

conformations, as well as the larger deviation between the actual and predicted pIC50 values. As shown in ►**Table 4**, the optimal CoMFA model resulted in a cross-validated q^2 of 0.663, a non-cross-validated correlation coefficient r^2 of 0.998, a standard error of estimation (SEE) value of 0.057, and F statistic value (F) of 2401.97. For the CoMSIA analysis, the q^2 value of 0.732, r^2 value of 0.988, SEE value of 0.15, and the F value of 542.933 were calculated, respectively.

For the CoMFA model, the contributions of the steric and electrostatic fields were calculated to be 47 and 52%, respectively; thus, the electrostatic field has more influence in comparison 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 10.1, 30.3, 13.2, 20.9, and 24.5%. ►**Table 3** lists the actual and predicted pIC_{50} values of the training and test set as well as the residues between them.

CoMFA Contour Maps
In the 3D-QSAR study, the striking feature of CoMFA and CoMSIA modeling is the visualization of 3D plot. The CoMFA model based on the training set of 39 compounds was

employed to investigate the existence of any correlation between chemical structures and activity of ALK inhibitors. Here, to elaborate the possible effects of different groups on the core parts of compounds **1–45** on the biological activities of the compounds, we use the following two skeletons (Skeleton 1 and Skeleton 2) to elaborate with R_1, R_2, R_3, R_4, and R_5, as listed in ►**Fig. 5(A, B)**.

The contour map for the final COMFA model is shown in ►**Fig. 6**. Results from the prediction of the steric field contour map of CoMFA (►**Fig. 6A**) and the biological activities of

Table 3 The actual and predicted pIC$_{50}$ values of compounds **1–45**

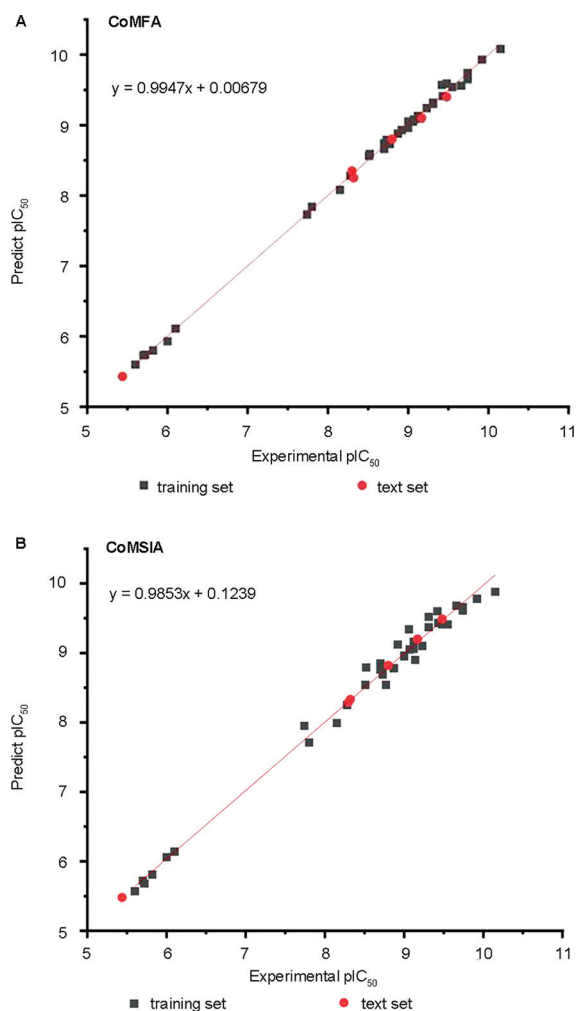| Compound	pIC$_{50}$	CoMFA		CoMSIA	
		Predicted	Residuals	Predicted	Residuals
1	9.00	9.05	−0.05	8.86	0.04
2	8.15	8.08	0.07	7.99	0.16
3	8.70	8.66	0.04	8.82	−0.12
4	7.80	7.84	−0.04	7.71	0.09
5^T	8.30	8.35	0.05	8.29	−0.01
6	8.28	8.28	0.00	8.25	0.03
7	8.77	8.73	0.04	8.54	0.16
8	6.00	5.93	0.07	6.06	−0.06
9	5.72	5.74	−0.02	5.68	0.04
10	6.10	6.11	−0.01	6.14	−0.04
11	5.60	5.60	−0.00	5.57	0.03
12	5.82	5.80	0.02	5.81	0.01
13^T	5.44	5.43	0.01	5.48	0.04
14	9.66	9.56	0.10	9.68	−0.02
15	8.70	8.69	0.01	8.85	−0.15
16^T	9.48	9.40	−0.08	9.49	0.01
17	8.51	8.56	−0.05	8.54	−0.03
18	9.74	9.74	0.00	9.61	0.13
19	9.14	9.09	0.05	8.90	0.24
20	9.48	9.59	−0.11	9.41	0.07
21	10.15	10.08	0.07	9.88	0.07
22	9.55	9.54	0.01	9.41	0.14
23	9.31	9.30	0.01	9.52	−0.21
24	9.43	9.41	0.02	9.43	0.00
25	9.31	9.32	−0.01	9.37	−0.06
26	8.70	8.74	−0.04	8.76	−0.06
27	9.12	9.13	−0.01	9.16	−0.04
28	9.23	9.24	−0.01	9.10	0.13
29	9.42	9.57	−0.15	9.60	−0.18
30	8.92	8.93	−0.01	9.12	−0.20
31	9.92	9.93	−0.01	9.78	0.14
32	9.07	9.08	−0.01	9.05	0.02
33	9.12	9.13	−0.01	9.06	0.06
34	9.74	9.65	0.09	9.66	0.08
35^T	9.17	9.10	−0.07	9.20	0.03
36	8.87	8.88	−0.01	8.78	0.09
37	8.73	8.77	0.04	8.69	0.04
38	8.73	8.79	−0.06	8.80	−0.07
39^T	8.32	8.25	−0.07	8.33	0.01
40	5.70	5.73	−0.03	5.72	−0.02
41^T	8.80	8.80	0.00	8.82	0.02
42	9.06	9.05	0.01	9.34	−0.28
43	9.00	8.96	0.04	8.95	−0.05
44	7.74	7.73	0.01	7.95	−0.21
45	8.52	8.59	−0.07	8.79	−0.27

Fig. 4 **(A)** Plot of experimental versus predicted pIC_{50} for training and test set by CoMFA. **(B)** Plot of experimental versus predicted pIC_{50} for the training and test set by CoMSIA.

compounds (►**Table 1**) showed that substituents of 1,2,3,4-tetrahydro-isoquinoline with bulk substituents at this site would possess better biological activity. Compounds **1** (pIC_{50} = 9.00) and **43** (pIC_{50} = 9.00) with 1-hydroxy-propane-2-one group and propan-1-ol group, respectively, possessed better biological activity than the methane sulfonamide group, carbonyl group, and 3-methyl-oxetane group-substituted compounds **2** (pIC_{50} = 8.16), 26 (pIC_{50} = 8.79), and **40** (pIC_{50} = 5.70). In addition, compound **12** (pIC_{50} = 5.82) with a bulky substituent of tetrahydropyran group in the position of 2,3,4,5-tetrahydro-1*H*-benzoazepine possessed better biological activity than the compound13^T (pIC_{50} = 5.44) with the tetrahydropyran group at the same position (►**Table 1**). Compounds with bulky groups at the R_4 position such as compound **19** (pIC_{50} = 9.42) exhibit higher activity. On the contrary, compounds **28** (pIC_{50} = 9.23) and **30** (pIC_{50} = 8.92) exhibit lesser inhibitory activities. The R_3 position showed that the first atom to which it is connected shows a yellow contour, however, at the end of R_3 position shows green. For instance, compound **39** (pIC_{50} = 8.32) with a bulky group displayed lower activity than compounds **36** (pIC_{50} = 8.87), **37** (pIC_{50} = 8.73), **38** (pIC_{50} = 8.73), and **41** (pIC_{50} = 8.80). Similarly, at the end of R_3 position, compound **31** (pIC_{50} = 9.92) with bulky substituent possessed better biological activity than the compound **22** (pIC_{50} = 9.55) and compound **28** (pIC_{50} = 9.23).

In the CoMFA electrostatic field map, blue contours represented regions where electropositive substitutions increased the activity, while red contours indicated regions where electronegative substitutions increased the activity. The electrostatic field map is shown in ►**Fig. 6B**. Blue contour maps surrounding R_3 positions suggested that electropositive groups were favorable at these positions for improved inhibitory potency. Position R_3 with electropositive

Table 4 The best results of the CoMFA and CoMSIA PLS statistical results

Descriptors	ONC	q^2	r^2	SEE	F	Field contributions				
						S	E	H	A	D
COMFA										
S + E	8	0.663	0.998	0.057	2401.97	0.47	0.52	–	–	–
COMSIA										
S + E	3	0.57	0.889	0.444	93.332	0.272	0.728	–	–	–
S + E + H	3	0.609	0.9	0.421	104.898	0.219	0.564	0.217	–	–
S + E + A	3	0.509	0.878	0.466	83.597	0.191	0.482	–	0.327	–
S + E + D	5	0.738	0.985	0.169	430.668	0.146	0.512	–	–	0.342
S + E + H + A	5	0.57	0.975	0.219	252.912	0.145	0.389	0.191	0.275	–
S + E + H + D	**10**	**0.77**	**0.999**	**0.034**	**5, 243.544**	**0.118**	**0.432**	**0.166**	**–**	**0.284**
S + E + D + A	5	0.7	0.985	0.168	434.205	0.117	0.365	–	0.233	0.286
S + E + H + A + D	5	0.732	0.988	0.15	542.933	0.101	0.313	0.132	0.209	0.245

Abbreviations: A, acceptor; D, donor; E, electrostatic; F value, F-test value; H, hydrophobic; S, steric; q^2, cross-validated correlation coefficient; NOC, optimum number of components; r^2, non-cross-validated correlation coefficient; SEE, standard error of estimation.
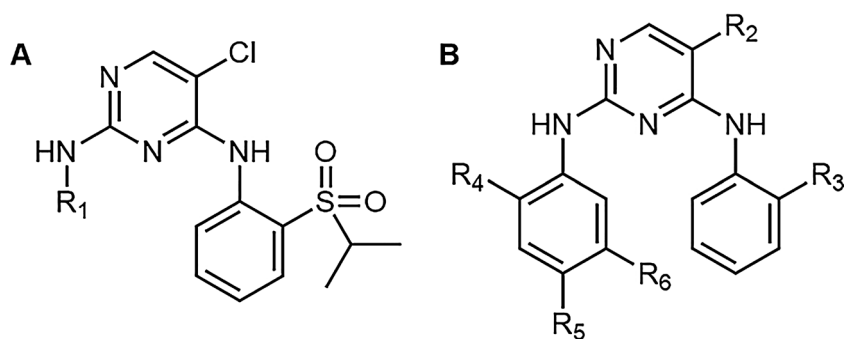Note: Final chosen model for COMSIA analysis is indicated in bold font.

Fig. 5 Skeleton for all molecules of dataset in CoMFA and CoMSIA analysis. **(A)** Skeleton 1 represented the common skeleton of these compounds: **1–12**, **13^T**, **15**, **26**, **40**, **43**, **44**, and **45**. **(B)** Skeleton 2 represented the common skeleton of compounds **1–45** except skeleton 1 represented.
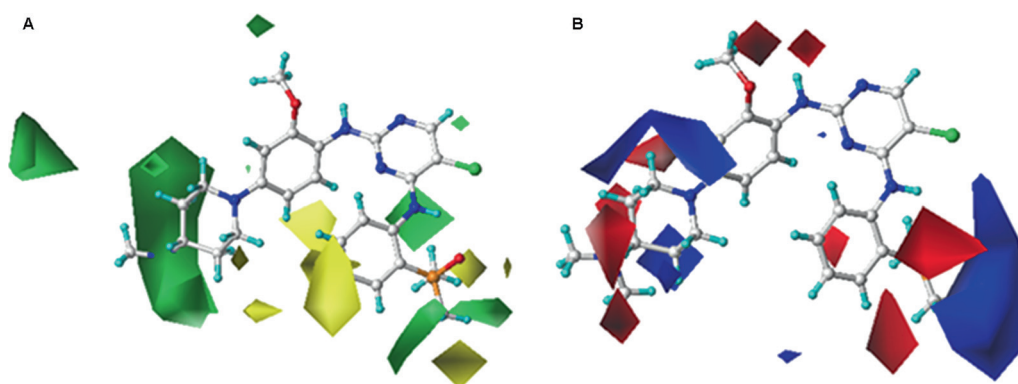

Fig. 6 CoMFA contour maps illustrating **(A)** steric and **(B)** electrostatic features in combination. The green color signifies regions that favor sterically bulky groups, and the yellow color signifies the opposite. Similarly, the blue color signifies positive charge and red color signifies negative charge that favored to increase the potency.
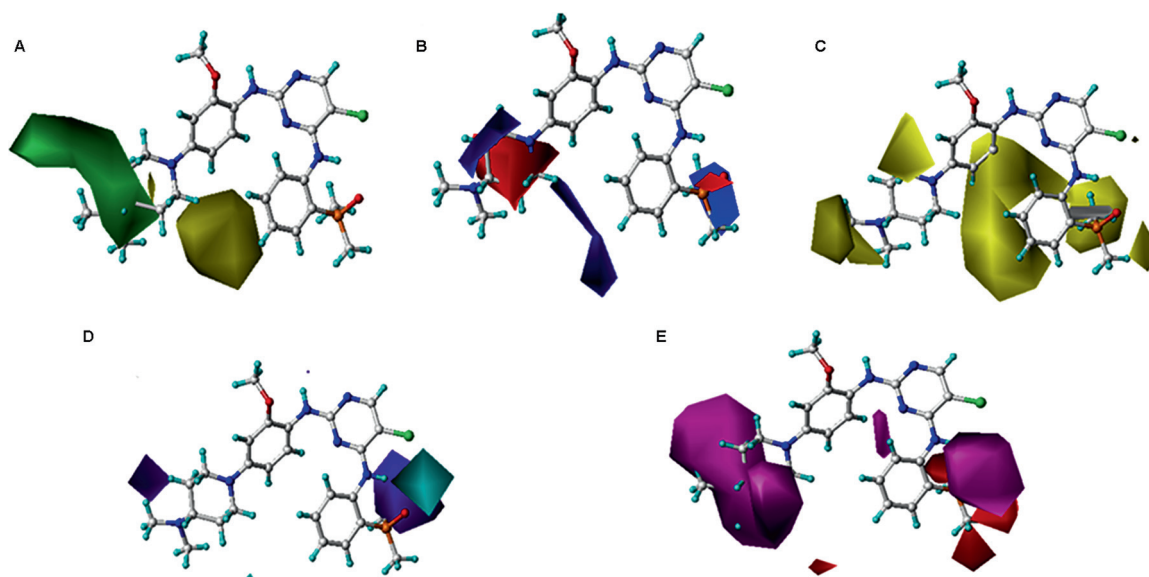

Fig. 7 CoMSIA contour maps of **(A)** steric, **(B)** electrostatic, **(C)** hydrophobic, **(D)** H-bond donor, and **(E)** H-bond acceptor. In the CoMSIA steric map, green contours represented sterically favorable regions where bulky substituents increased the activity. The yellow contours indicated sterically unfavorable region where bulky substituents decreased the activity.
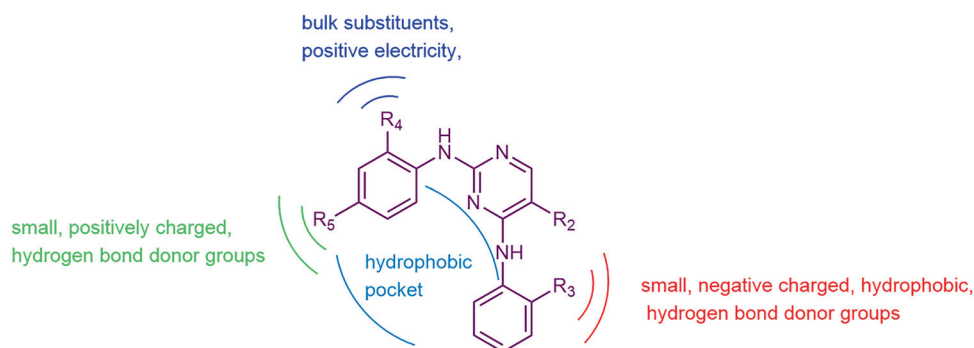
Fig. 8 3D-QSAR models of ALK piperidine-substituted analogues. ALK, anaplastic lymphoma kinase.

substitutions displayed better biological activity and position R_4 with electronegative substitutions can increase the activity.

CoMSIA Contour Maps

The five CoMSIA contour maps, namely, steric, electrostatic, hydrophobic, donor, and acceptor are shown in ►**Fig. 7**.

►**Fig. 7A** depicts the CoMSIA steric contour map of the optimal model with compound **21** overlaid. In this map, the green (sterically favorable) and yellow (sterically unfavorable) contours represented 80 and 20% of level contributions, respectively. The large green contour occurring at the end of R_1 position indicated that compounds with bulk substituents at this site would possess better biological activity. The activity of compound **6** ($pIC_{50} = 8.28$) and compound **8** ($pIC_{50} = 6.00$) increased after the secondary nitrogen substituents on the position were replaced with a small-size group, such as compound **10** ($pIC_{50} = 6.10$), compound **11** ($pIC_{50} = 5.60$), and compound **12** ($pIC_{50} = 5.82$). It is clear that compounds **22** ($pIC_{50} = 9.55$), **23** ($pIC_{50} = 9.31$), **24** ($pIC_{50} = 9.43$), and **25** ($pIC_{50} = 9.31$) with bulky substituents displayed higher biological activity than compound **27** ($pIC_{50} = 9.12$).

►**Fig. 7B** shows a large and a small red contour enclosing regions R_1 and R_3, respectively. By comparing the series of compounds **22** with **24**, it can be seen that as charge effects increased (N has electronegativity of 3.04 and C has electronegativity of 2.55), the inhibitory activity increased as well.

The CoMSIA contour map of hydrophobic contribution is described in ►**Fig. 7C**. In this figure, the yellow (hydrophobic favorable) and white (hydrophobic unfavorable) contours represent 80 and 20% of level contributions, respectively. Obviously, a large yellow and a white region that R_3 position with hydrophilic groups can increase biological activity. However, a large-sized yellow contour located at the para-site of the R_5 and R_6 aromatic ring represented regions where the compounds with more hydrophobicity at this site would possess better biological activity. Moreover, at the end of R_1 substituents, the compounds with hydrophobicity will possess better biological activity.

In ►**Fig. 7D**, the cyan and purple contour maps indicated favorable and unfavorable H-bond donor groups, representing 80 and 20% of level contributions, respectively. Light purple appears at the end of the R_1 position, indicating that H-bond donor groups enhanced biological activity.

As a validation of the CoMSIA model that H-bond acceptor in ►**Fig. 7E**, the hydrogen bond acceptor field of the CoMSIA model was represented by magenta (hydrogen bond acceptor favorable) and red (hydrogen bond acceptor disfavorable), representing 80 and 20% of contributions ratio, respectively. The large magenta contours near the R_1 position suggested that the introduction of hydrogen bond acceptor substituents to these regions would increase activity. In addition, the R_3 regions showed that the biological activity influenced by the H-bond acceptor depends on the nature of its substituents.

Conclusion

In this article, we developed a statistically significant 3D-QSAR model of a series of piperidine-substituted analogues through CoMFA, CoMSIA, and molecular docking. In this 3D-QSAR model, shown in ►**Fig. 8**, the derived leave-one-out (LOO) validated the PLS regression QSAR model and showed acceptable r^2 (0.988) and q^2 (0.730). As noted, this model demonstrates that R_1 and R_3 played a key role in increasing molecular activity, where the R_1 position may be better with small, negatively charged, hydrophobic, and hydrogen bond donor groups, and the R_3 position may be better with small, positively charged, and hydrogen bond donor groups.

As is shown in experimental data as well as CoMFA and CoMSIA contour maps, this model is credible for explaining the structure–activity relationship of the second-generation ALK inhibitors with steric, electrostatic, hydrophobic, donor, and acceptor. The results from the 3D-QSAR models and molecular docking would lead to a better understanding of the structural features needed to design and synthesize novel potential second-generation ALK inhibitors.

Conflict of Interest
The authors report no declarations of interest. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject

matter or materials discussed in the manuscript apart from those disclosed.

References

1 Shaw AT, Engelman JA. ALK in lung cancer: past, present, and future. J Clin Oncol 2013;31(08):1105–1111

2 Roskoski R Jr. Anaplastic lymphoma kinase (ALK): structure, oncogenic activation, and pharmacological inhibition. Pharmacol Res 2013;68(01):68–94

3 Katayama R, Lovly CM, Shaw AT. Therapeutic targeting of ana-plastic lymphoma kinase in lung cancer: a paradigm for precision cancer medicine. Clin Cancer Res 2015;21(10):2227–2235

4 Vernersson E, Khoo NK, Henriksson ML, Roos G, Palmer RH, Hallberg B. Characterization of the expression of the ALK receptor tyrosine kinase in mice. Gene Expr Patterns 2006;6(05):448–461

5 Tuma RS. ALK gene amplified in most inflammatory breast cancers. J Natl Cancer Inst 2012;104(02):87–88

6 Webb TR, Slavish J, George RE, et al. Anaplastic lymphoma kinase: role in cancer pathogenesis and small-molecule inhibitor devel-opment for therapy. Expert Rev Anticancer Ther 2009;9(03): 331–356

7 Ren H, Tan ZP, Zhu X, et al. Identification of anaplastic lymphoma kinase as a potential therapeutic target in ovarian cancer. Cancer Res 2012;72(13):3312–3323

8 Wang P, Cai J, Chen J, et al. 3D-QSAR and docking studies of piperidine carboxamide derivatives as ALK inhibitors. Med Chem Res 2013;23(05):2576–2583

9 Aubry A, Galiacy S, Allouche M. Targeting ALK in cancer: thera-peutic potential of proapoptotic peptides. Cancers (Basel) 2019; 11(03):275

10 Wu W, Haderk F, Bivona TG. Non-canonical thinking for targeting ALK-fusion onco-proteins in lung cancer. Cancers (Basel) 2017;9 (12):164

11 Hallberg B, Palmer RH. Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. Nat Rev Cancer 2013;13 (10):685–700

12 Ou SI, Zhu VW, Nagasaka M. Catalog of 5′ fusion partners in ALK-positive NSCLC Circa 2020. JTO Clin Res Rep 2020;1(01):100015

13 Shaw AT, Yeap BY, Solomon BJ, et al. Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: a retrospective analysis. Lancet Oncol 2011;12(11):1004–1012

14 Doebele RC, Pilling AB, Aisner DL, et al. Mechanisms of resistance to crizotinib in patients with ALK gene rearranged non-small cell lung cancer. Clin Cancer Res 2012;18(05):1472–1482

15 Katayama R, Shaw AT, Khan TM, et al. Mechanisms of acquired crizotinib resistance in ALK-rearranged lung cancers. Sci Transl Med 2012;4(120):120ra17

16 Zhang S, Wang F, Keats J, et al. Crizotinib-resistant mutants of EML4-ALK identified through an accelerated mutagenesis screen. Chem Biol Drug Des 2011;78(06):999–1005

17 Kinoshita K, Oikawa N, Tsukuda T. Chapter nineteen - anaplastic lymphoma kinase inhibitors for the treatment of ALK-positive cancers. Annu Rep Med Chem 2012;47:281–293

18 Choi YL, Soda M, Yamashita Y, et al. ALK Lung Cancer Study Group. EML4-ALK mutations in lung cancer that confer resistance to ALK inhibitors. N Engl J Med 2010;363(18):1734–1739

19 Sasaki T, Koivunen J, Ogino A, et al. A novel ALK secondary mutation and EGFR signaling cause resistance to ALK kinase inhibitors. Cancer Res 2011;71(18):6051–6060

20 Haratake N, Toyokawa G, Seto T, et al. The mechanisms of resis-tance to second- and third-generation ALK inhibitors and strate-gies to overcome such resistance. Expert Rev Anticancer Ther 2021;21(09):975–988

21 Smolle E, Taucher V, Lindenmann J, Jost PJ, Pichler M. Current knowledge about mechanisms of drug resistance against ALK inhibitors in non-small cell lung cancer. Cancers (Basel) 2021;13 (04):699

22 Huang WS, Liu S, Zou D, et al. Discovery of Brigatinib (AP26113), a phosphine oxide-containing, potent, orally active inhibitor of ana-plastic lymphoma kinase. J Med Chem 2016;59(10):4948–4964

23 Achary R, Yun JI, Park CM, et al. Discovery of novel tetrahydroi-soquinoline-containing pyrimidines as ALK inhibitors. Bioorg Med Chem 2016;24(02):207–219

24 Kang GA, Lee M, Song D, et al. Synthesis and evaluation of novel 2,4-diaminopyrimidines bearing bicyclic aminobenzazepines for anaplastic lymphoma kinase (ALK) inhibitor. Bioorg Med Chem Lett 2015;25(18):3992–3998

25 Clark M, Cramer RD III, Van Opdenbosch N. Validation of the general purpose Tripos 5.2 force field. J Comput Chem 1989;10 (08):982–1012

26 Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron 1980;36(22):3219–3228

27 Selvaraj C, Tripathi S, Reddy K, Singh SK. Tool development for prediction of pIC50 values from the IC$_{50}$ values - a pIC$_{50}$ value calculator. Curr Trends Biotechnol Pharm 2011;5(02):1104–1109

28 Marsilje TH, Pei W, Chen B, et al. Synthesis, structure-activity relationships, and in vivo efficacy of the novel potent and selective anaplastic lymphoma kinase (ALK) inhibitor 5-chloro-N2-(2-iso-propoxy-5-methyl-4-(piperidin-4-yl)phenyl)-N4-(2-(isopropyl-sulfonyl)phenyl)pyrimidine-2,4-diamine (LDK378) currently in phase 1 and phase 2 clinical trials. J Med Chem 2013;56(14): 5675–5690

29 Bryan MC, Whittington DA, Doherty EM, et al. Rapid development of piperidine carboxamides as potent and selective anaplastic lym-phoma kinase inhibitors. J Med Chem 2012;55(04):1698–1705

30 Robinson DD, Winn PJ, Lyne PD, Richards WG. Self-organizing molecular field analysis: a tool for structure-activity studies. J Med Chem 1999;42(04):573–583

31 Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 1988;110(18): 5959–5967

32 Shen M, LeTiran A, Xiao Y, Golbraikh A, Kohn H, Tropsha A. Quantitative structure-activity relationship analysis of function-alized amino acid anticonvulsant agents using k nearest neighbor and simulated annealing PLS methods. J Med Chem 2002;45(13): 2811–2823

33 Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in drug design–a review. Curr Top Med Chem 2010;10(01):95–115

34 Kubinyi H. QSAR and 3D QSAR in drug design Part 2: appli-cations and problems. Drug Discov Today 1997;2(12): 538–546

35 Liu X, Chen X, Zhang L, Zhan P, Liu X. 3D-QSAR and docking studies on piperidine-substituted diarylpyrimidine analogues as HIV-1 reverse transcriptase inhibitors. Med Chem Res 2015;24(08):3314–3326

36 Roe SM, Prodromou C, O'Brien R, Ladbury JE, Piper PW, Pearl LH. Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin. J Med Chem 1999;42(02):260–266