3D-QSAR studies of Checkpoint Kinase 1 inhibitors based on molecular docking and CoMFA

Rong-Wei Wang[a], Lu Zhou[a,*], Zhi-Li Zuo[b], Xiang Ma[a], Min Yang[a]

[a]College of Chemical Engineering, Sichuan University, Sichuan, Chengdu 610065,China. [b]Centre for Biomedical & Life Sciences, Singapore Polytechnic, 139651, Singapore.

Three-dimensional quantitative structure-activity relationship (3D-QSAR) studies were performed on a series of substituted 1,4-dihydroindeno[1,2-c]pyrazoles inhibitors, based on molecular docking scores obtained by using GOLD 3.01, with comparative molecular field analysis (CoMFA). The docking results provided a reliable conformational alignment scheme for the 3D-QSAR model. Based on the docking conformations and alignments, highly predictive CoMFA model was obtained with cross-validated q^2 value of 0.534 and non-cross-validated partial least-squares (PLS) analysis with the optimum components of six showed a conventional r^2 of 0.911. The predictive ability of this model was validated by the testing set with a conventional r^2 value of 0.812. The information obtained from the CoMFA 3D contour maps enables the interpretation of their structure-activity relationship and was also used to the design of several new inhibitors with improved activity.

Keywords: CoMFA, 3D-QSAR, Checkpoint kinase 1 (CHK1), Molecular docking, Substituted1,4-dihydroindeno[1,2-c]pyrazoles

1. Introduction

DNA-damaging anticancer agents is the mainstay of cancer treatment and has produced significant increases in the survival of cancer patients when used in combination with drugs that have different mechanisms of actions[1] .However, the clinical use of DNA-damaging anticancer agents is limited by their severe toxicity and by resistance from tumor cells. So, it is important to develop highly efficacious and minimally toxic treatments for cancer. Recently, the development of adjuvant therapeutics has been aggressively pursued. Such treatments may either sensitize tumor tissue or protect normal tissue from DNA damage [2].

With DNA damage, cells are arrested (G1, S, G2) to initiate the DNA repair process [3-5]. Most tumor cells distinguish themselves from normal cells by lacking the G1 checkpoint due to loss of p53, and therefore they are selectively arrested at the S or G2 checkpoint after DNA damage. If the S and G2 checkpoints are abrogated, G1-deficient cancer cells will lead cell death. In contrast, normal cells are still arrested in the G1 phase and are less affected by S and G2 checkpoint abrogation, suggesting that a favorable therapeutic window may be achieved for G2 and/or S abrogators [6].

Checkpoint kinase 1 (CHK1) is a serine/threonine protein kinase and a key mediator in the DNA damage-induced checkpoint network, and it is activated via activation of the upstream ATM/ATR pathway[7-10].Activation of CHK1 results in phosphorylation of Cdc25A at Ser123 and several other serine residues and Cdc25C at Ser216. The downstream event is the inhibition of cyclin E/Cdk2 or

*Corresponding author. Email: zhoulu@scu.edu.cn

cyclin B/Cdc2 kinases, which ultimately causes cell cycle arrest at S or G2 phase [11] .The inhibitors of CHK1 abrogate the S and G2 checkpoints, thereby preferentially sensitizing tumor cells, especially p53-null cells. Consequently, CHK1 has emerged as an attractive chemosensitization target, and the inhibitors of CHK1 may significantly improve the efficacy and selectivity of DNA-damaging agents in the clinic [6] .

In recent years, a number of CHK1 inhibitors have appeared in the primary literatures. Among them, Yunsong Tong et al. found that compound 1 (Figure 1) was a potent inhibitor of CHK1 (IC_{50} 510 nM). Moreover, they synthesized a series of 1,4-dihydroindeno[1,2-c]pyrazoles derivatives as potent, selective CHK1 inhibitors. The binding modes of this series of CHK1 inhibitors have been determined by X-ray crystallography, which provided not only insights into the interaction mechanisms of CHK1 with the inhibitors, but also valuable clues for designing new inhibitors. In this paper, with the molecular docking and Three-Dimensional(3-D) QSAR (CoMFA) analyses. It is possible to get new insights into the relationship between the structural information of a series of substituted 1,4-dihydroindeno[1,2-c]pyrazoles inhibitors and the inhibitory potency, aimed at identifying structural features in CHK1 that can be used to design new inhibitors.

2. COMPUTATIONAL DETAILS

2.1 Biological Data and Molecular Structures

We gathered a panel of 151 structurally and pharmacologically diverse compounds for the 3D QSAR analysis from four publications reported by one laboratory [2，11，14-15].Because of the similar experimental procedures applied for affinity determination in each publication, the biological data (represented as IC_{50} values) were considered comparable and thus merged into our study. The IC_{50} values, in nM, were converted to pIC_{50} (-log IC_{50}) values, which were used as dependent variables in the QSAR analyses. The biological data were divided into a training set and a testing set as shown in Table 1, 2 and 3. The training set which was selected randomly consists of 121 compounds and the testing set is comprised of 30 compounds.

The 3D structures of these compounds were carried out on a personal computer with the RHEL 4.0 operating system using SYBYL 8.0. The 3D structures of the training and test set of compounds were constructed using the Sketch Molecule function in SYBYL. Partial atomic charges were calculated by the Gasteiger-Hückel method and energy minimizations were performed using the Tripos force field with a distance-dependent dielectric and the Powell conjugate gradient algorithm (convergence criterion of 0.005 kcal/mol/Å).

2.2 Docking Studies

For the docking of ligands to protein active sites and for estimating the binding affinities of docked compounds, an advanced molecular docking program GOLD version 3.01 , with a powerful genetic algorithm (GA) method for conformational search and docking programs, was used in this study.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Atomic coordinates for the CHK1 complex with CHK2759M41, used for our modeling, have been deposited in the Brookhaven Protein Data Bank (PDB ID: 2E9N).The original ligand was removed from the coordinated set. The genetic operators were 100 for the population size, 1.1 for the selection, 5 for the number of subpopulations, 100000 for the maximum number of genetic applications, and 2 for the size of the niche used to increase population diversity. The weights were chosen so that crossover mutations were applied with equal probability (95/95 for the values), and migration was applied 5% of the time. ChemScoring function encoded in GOLD was applied to predict binding positions between CHK1 and 150 inhibitors. The fitness score is taken as the negative of the sum of the component energy terms, so that larger fitness scores are better [12].

2.3. Structural alignment

Molecular alignment is the most sensitive parameter in 3D-QSAR analysis. This renders the spatial alignment of molecules under study as one of the most sensitive and determining factors in obtaining robust and meaningful models. Ten conformations were obtained using GOLD for each ligand, in which the training set contained the conformations with the top-ranked ChemScores while the test set contained the conformations with the lowest residues between actual and predicted pIC_{50} predicted by the CoMFA model of the training set. 121 inhibitors were selected randomly as the training set, and these inhibitors were aligned together for CoMFA study to explore the specific contributions of electrostatic and steric effects of the molecular bioactivies.

2.4 CoMFA

Steric and electrostatic interactions were calculated using the Tripos force field with a distance-dependent dielectric constant at all intersections in a regular space (2 Å) grid taking a sp^3 carbon atom as steric probe and all charge as electrostatic probe. The cutoff was set to 30 kcal/mol. With standard options for scaling of variables, the regression analysis was carried out using the full cross-validated partial least-squares (PLS) method of LOO (leave-one-out). The minimum-sigma (column filtering) was set to 2.0 kcal/mol to improve the signal-to-noise ratio by omitting those lattice points whose energy variation was below this threshold. The final model, a non-cross-validated conventional analysis, was developed with the optimum number of components to yield a non-cross-validated r^2 value [13].

3 Results and Discussion

3.1. Docking study

In order to determine the probable binding conformations of these substituted 1,4-dihydroindeno[1,2-c]pyrazoles , GOLD was used to dock all compounds into the active sites of CHK1. The docking reliability was validated using the known X-ray structure of CHK1 in complex with a small molecular ligand CHK2759M41. The ligand CHK2759M41 was redocked to the binding

sites of CHK1 and the docked conformation corresponding to the lowest free energies was selected as the most probable binding conformation. The root-mean-square deviation (RMSD) between the conformations of cocrystallized CHK2759M41 and redocked CHK2759M41 was equal to 1.118 Å, suggesting that a high docking reliability of GOLD in reproducing the experimentally observed binding mode for CHK1 inhibitors and the parameters set for the GOLD simulation was reasonable to reproduce the X-ray structure. Just as shown in Figure 2, the cocrystallized CHK2759M41 and redocked CHK2759M41 are almost at the same position in the active sites of CHK1. Therefore, the GOLD method and the parameters set could be extended to search the CHK1 binding conformations for other inhibitors.

The predicted binding free energies (ChemScore) for all the inhibitors are listed in Table 2 and 3, and the pIC_{50} of the inhibitors are shown in Tables 1. Figure 3 represents the interaction model of the docked inhibitor CHK2759M41 with CHK1. Inhibitor CHK2759M41 binds to the active site and makes several interactions with the hinge-binding region of the enzyme.

As show in figure 3, inhibitor CHK2759M41 forms six hydrogen bonds with Glu55, Phe149, Asn59, Cys87, Glu85 and Leu15. The phenyl ring interacts with the hydrophobic surface of the side chains of Leu84, Ile56, Val56 and Phe149. Also, the cyclohexane ring interacts with the hydrophobic surface of the side chains of Tyr86, Ser88, Thr14, Asp94 and Glu17. Figure 4 illustrates the probable binding conformational alignment for the the 151 substituted 1, 4-dihydroindeno[1,2-c]pyrazoles inhibitors chosen from the docked conformations.

3.2 CoMFA model

The CoMFA analysis was performed to explore the structure-activity relationship of substituted 1,4-dihydroindeno[1,2-c]pyrazoles inhibitors of CHK1. PLS analysis was carried out for the 121 training set, and the result is listed in Table 4, which showed that a CoMFA model with a cross-validated q^2 of 0.534 for 6 components was obtained. The non-cross-validated PLS analysis with the optimum components of 6 revealed a conventional r^2 value of 0.911, F =187.106, and an estimated standard error of 0.352. The steric field descriptors explain 43.0% of the variance, while the electrostatic descriptors explain 57.0%. The predicted activities for the 151 inhibitors versus their experimental activities with their residues are listed in Tables 2 and 3, and the correlation between the predicted activities and the experimental activities are depicted in Figure 5 and 6. Those values indicate a good statistical correlation and reasonable predictability of the CoMFA model.

The steric interaction is represented by green and yellow contours, in which green-colored regions indicate areas where increased steric bulk is associated with enhanced activity, and yellow regions suggest areas where increased steric bulk is unfavourable. Electrostatic interaction is indicated by red and blue contours, among which blue-colored regions show areas where more positively charged groups are favourable, and red regions highlight areas where groups with more negative partial charges are favourable.

The CoMFA steric contours of the most active molecule CHK2759M41 is displayed in Figure 7. A large green region (G1) near the 6-position of CHK2759M41 indicates that a bulky substituent is

preferred in the position to produce higher inhibitory activity. This is in agreement with the fact that the inhibitory activities of compounds CHK2759M38 and CHK2759M42 with more bulky substituents (with aliphatic hydrocarbon)in G1 are higher than compound CHK3618M5（-H）. Also, compounds CHK2759M39 and CHK2759M41, which are cyclohexane analogues, showed higher activity than compound CHK3618M5（-H）.And that cyclohexane ring interacts with the hydrophobic surface of the side chains of Tyr86, Ser88, Thr14, Asp94 and Glu17. So this volume (G1) is in agreement with the receptor structure.

Two yellow contours (Y1, Y2) were observed which suggest that compounds with substituents entering these contours are less active than those with substituents that do not entering these contours. Figure 3 shows that residues Phe149, Glu55, Lys38 and Leu84 in the binding pocket of the CHK1 are in the distance of less than 3.0 Å to Y1 and Y2 of these inhibitors. Therefore, any larger substitutes may lead to steric collision with those residues in the pocket. One finding lend further support to this opinion, the activity of compound CHK2759M41 decreases quickly after the -OH groups are replaced by bulky groups, just like CHK2759M31 which has benzene ring group in Y1 and Y2 .

Due to the steric collision with these residues in Y1 and Y2, those compounds move downward and the hydroxybenzenes of those compounds entering into the large green contour map(G3).This large region of green contour (G3) around the 3-position of the CHK2759M41 suggests that steric bulk is favourable there. For example, the activity of compounds CHK5944M23 and CHK5944M24 (with abenzene ring) are higher than compound CHK5944M22 (with a acetylene).It can be see in Figure 3 that the benzene ring interacts with the hydrophobic surface of the side chains of Leu84, Ile56, Val56 and Phe149. So this region (G3) is in agreement with the receptor structure.

The electrostatic contour map of the CoMFA model is shown in Figure 8. To facilitate the visualization, the most potent compound CHK2759M41 is overlaid on the map.

R1 suggests that electronegative groups at this position are favourable for inhibitory activity. For example, the high active compounds CHK2758M22，CHK2758M23，CHK2758M24，CHK2758M25，CHK2758M26，CHK2758M27 and CHK2758M28 have electronegative groups (carboxy) located in this red contour. As indicated in Figure 3, Lys38 which containing positive groups in this region (R1), has a strong electrostatic interaction with the negative charged groups (carboxy) of those inhibitors.

Two red contours (R2, R3) observed next to the plane of the benzene ring are contributed by the nitrile group in compounds CHK5944M40 to CHK5944M50 and are taken as an indication of the role of electronegative groups at these positions in increasing the inhibitory activity. Nitrile groups have an overall negative-charge and hence show reasonable activity. Furthermore, Asn59 containing positive groups rather near this region (R2, R3), has a strong electrostatic interaction with the negative charged substitute groups of the inhibitors. Also, the hydrogen in the -NH group is electrophilic, further docking study showed that the -NH groups of Asn59 or Phe149 might form H-bonds with the -OH groups of CHK4308M5-CHK4308M52.

Two large blue (B1 and B2) regions near Glu55, Gly150 and Asp148 suggest that electropositive substituents would increase the inhibitory activity. Those residues have a hydroxyl. The docking result of the potent inhibitor CHK2759M41 showed that the hydroxyl groups of CHK4308M5 to CHK4308M52 donated a hydrogen bond to Glu55.

A large blue contour (B3) is noticed in 6-position of CHK2759M41. As show in Figure 9, a small

area of blue (B4) contour around the 7-position of CHK2759M41 indicates that electropositive substituents are favourable there. As the experimental data shown, compounds CHK4308M7、CHK4308M9、CHK4308M26、CHK4308M27 and CHK4308M28 which have a methyl group on that position have higher pIC_{50} values from 0.1 to 1 nM.

In Figure 10, the blue contour region (B5) shows that electropositive groups are favourable in this region. Inhibitors 42 and CHK3618M24 which contain partially positive charged groups (methyl) in this region show higher pIC_{50} values.

3.3 Validation of the 3-D QSAR Models

The thirty randomly selected compounds (Table 3) were used as the testing set to verify the constructed CoMFA model. The calculated results are listed in Table 3 and displayed in Figure 7. The predicted pIC_{50} values were in good agreement with the experimental data within a statistically tolerable error range, with a correlation coefficient of $r^2=0.812$ for CoMFA model. The test results indicated that the CoMFA model would be reliable for new CHK1 inhibitors designing and developing drug leads against cancer.

4 Conclusions

In this study, CoMFA models have been built to explain the structure-activity relationship of a series of 1, 4-dihydroindeno[1,2-c]pyrazoles derivatives. The developed CoMFA model showed a good predictive power with low residuals for test set molecules. In addition, the 3D-QSAR results suggested that the hydrophobic group is the favourable group in th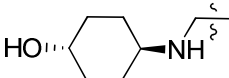e G3 region of those 1,4-dihydroindeno[1,2-c]pyrazoles derivatives. Also, the hydrophobic group is the favourable group in the G1 region of the 1, 4-dihydroindeno[1,2-c]pyrazoles derivatives. And the docking study revealed that a lot of inhibitors establish hydrogen bonds with Glu55 and Phe149, Asn59, Cys87, Glu85, Leu15. These results demonstrated the power of a combined docking/QSAR approach to explore the probable binding conformations of compounds at the active site of the protein target, and further provided useful information in understanding the structural and chemical features of 1,4-dihydroindeno[1,2-c]pyrazoles derivatives in designing and finding new potential inhibitors.

References

[1] Hurley. L. H. *DNA and its associated processes as targets for cancer therapy*, Nat. Rev. Cancer. *2*（2002）, p.188.

[2]Tao.ZF, Li.GQ, Tong.YS, et al. *Discovery of 4'-(1,4-dihydro-indeno[1,2-c]pyrazol-3-yl)-benzonitriles and 4'-(1,4-dihydro-indeno[1,2-c]pyrazol-3-yl)-pyridine-2'-carbonitriles as potent checkpoint kinase 1 (Chk1) inhibitors*, Bioorganic & Medicinal Chemistry Letters**.** 17（2007）**,** p.5944.

[3] Sancar.A, Lindsey-Boltz, L. A, Unsal-Kacmaz, K, Linn, S. *Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints*, Annu. ReV. Biochem.*73*（2004）, p.39

[4] Zhou, B.-B.; Bartek, J. *Targeting the checkpoint kinases: chemosensitization versus chemoprotection,* Nat. ReV. Cancer .*4*（2004）, p.216.

[5] Kastan, M. B, Bartek, J. *Cell-cycle checkpoints and cance,* .Nature .*432*（2004）, p.316

[6] Zhi-Fu Tao. *Structure-Based Design, Synthesis, and Biological Evaluation of Potent and Selective Macrocyclic Checkpoint*

*Kinase 1 Inhibitors***,** J. Med. Chem**.** *50*（2007）, p.1514

[7] Walwroth. N, Davery. S. Beach, D. *Fission yeast Chk1 protein kinase links the rad checkpoint pathway to cdc2*, Nature. *363* （1993）, p.368.

[8] Chen. Y, Sanchez.Y, *Chk1 in the DNA damage response: conserved roles from yeasts to mammals*, DNA Repair.*3*（2004）, p.1025.

[9] Liu. Q, Guntuku. S, Cui. X, Matsuoka. S, Cortez. D, Tamai. K, Luo.G, Carattini-Rivera. S, DeMayo. F. Bradley. A, Donehower. L.A., Elledge. S. J, *Chk1 is an essential kinase that is regulated by Atr and required for the G2/M DNA damage checkpoint,* Genes DeV.*14*（2000）, p.1448.

[10] Abraham. R, T, *Cell cycle checkpoint signaling through the ATM and ATR kinases*, Genes DeV.*15*（2001）, p.2177.

[11] Yunsong Tong, *1,4-Dihydroindeno[1,2-c]pyrazoles as potent checkpoint kinase 1 inhibitors: Extended exploration on phenyl ring substitutions and preliminary ADME/PK studies*, Bioorganic & Medicinal Chemistry Letters .17 (2007) , p.3618.

[12] Ma X, Zhou L, Zuo ZL. *Molecular docking and 3-D QSAR studies of substituted 2,2-bisaryl-bicycloheptanes as human 5-Lipoxygenase-Activating Protein (FLAP) inhibitors*, Qsar & Combinatorial Science.27（2008）, p.1083 .

[13] Yi P, Fang X, Qiu MH , *3D-QSAR studies of Checkpoint Kinase Weel inhibitors based on molecular docking, CoMFA and CoMSIA*, European Journal Of Medicinal Chemistry . 43（2008）, p.925.

[14] Tong YS, Claiborne A, Stewart KD. *Discovery of 1,4-dihydroindeno[1,2-c]pyrazoles as a novel class of potent and selective checkpoint kinase 1 inhibitors*, Bioorganic & Medicinal Chemistry Letters **.**15（2007）**,** p.2759 .

 [15] Tao. ZF, Li. GQ, Tong. Y, *Synthesis and biological evaluation of 4 '-(6,7-disubstituted-2, 4-dihydro-indenol]1,2-c] pyrazol-3-yl)-biphenyl-4-ol as potent Chk1 inhibitors,* Bioorganic & Medicinal Chemistry Letters.15（2007）**,** p.4308.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
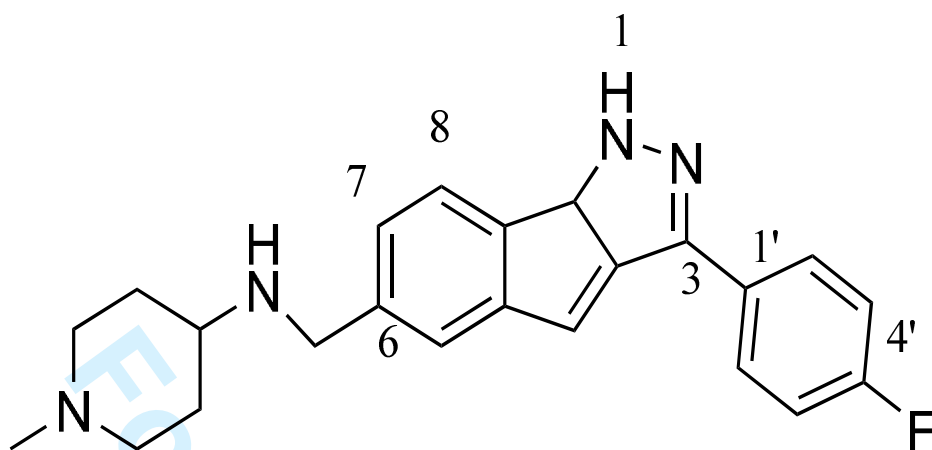
Figure 1. The structure of compound 1
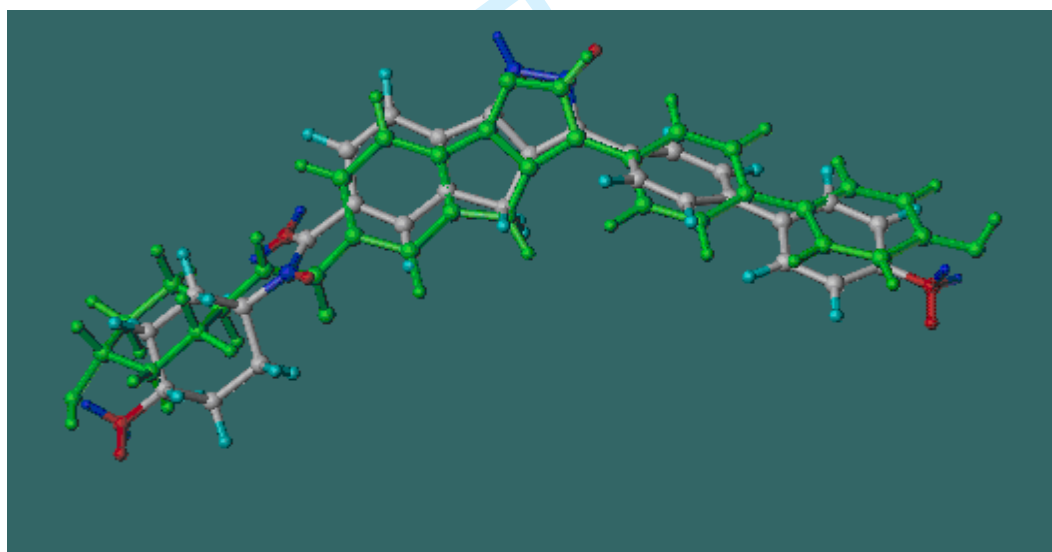
Figure 2.   Binding conformations of the cocrystallized CHK2759M41 and redocked CHK2759M41 at the active sites of CHK1.

(Green compound represents cocrystallized CHK2759M41 and white compound represents redocked CHK2759M41)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
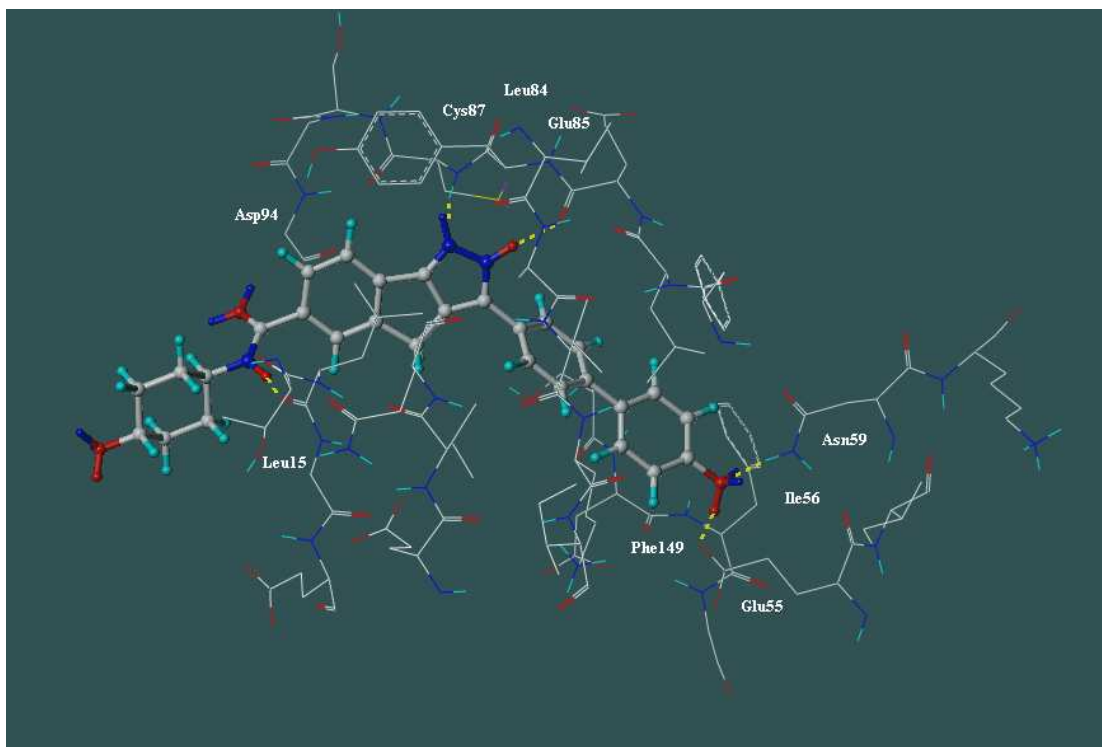
Figure 3.   Position of docked inhibitor CHK2759M41 in the binding pocket of CHK1. H-bonds and interactions between CHK2759M41 and side chain residues
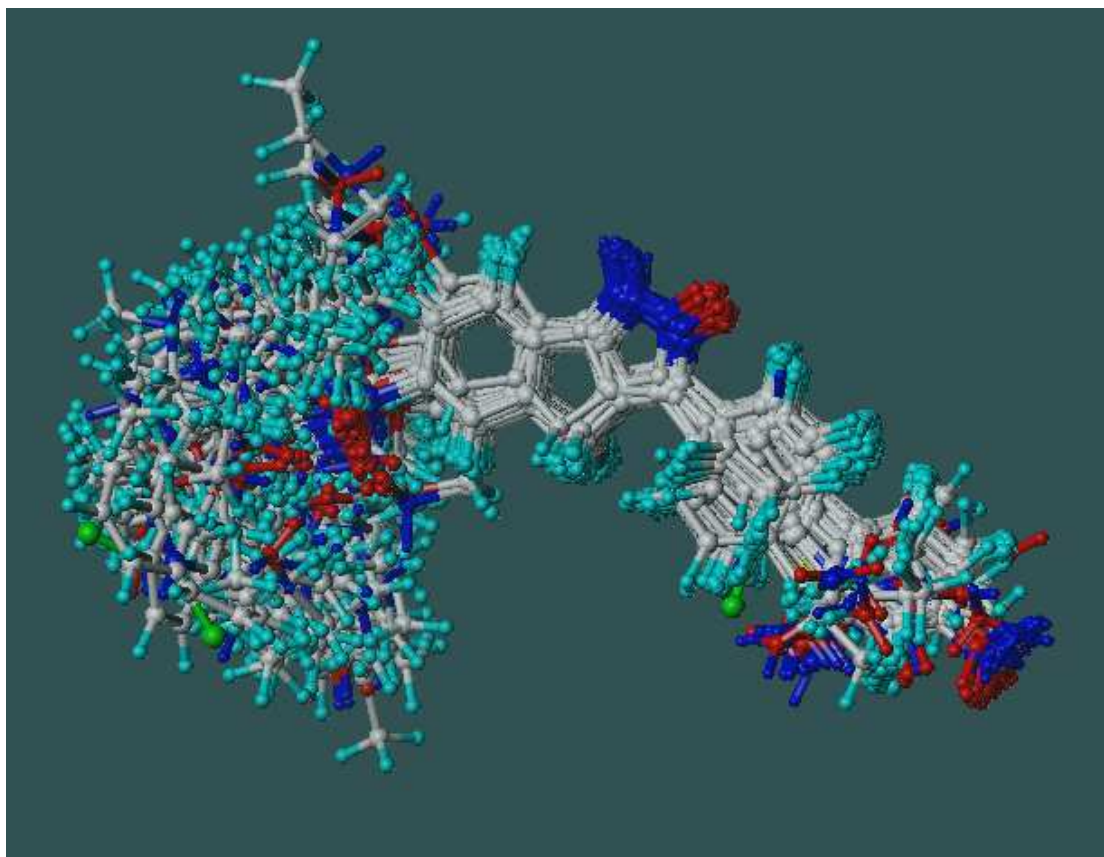
Figure 4.   Superimposition of 151 substituted 1,4-dihydroindeno[1,2-c]pyrazoles for 3D- QSAR studies.
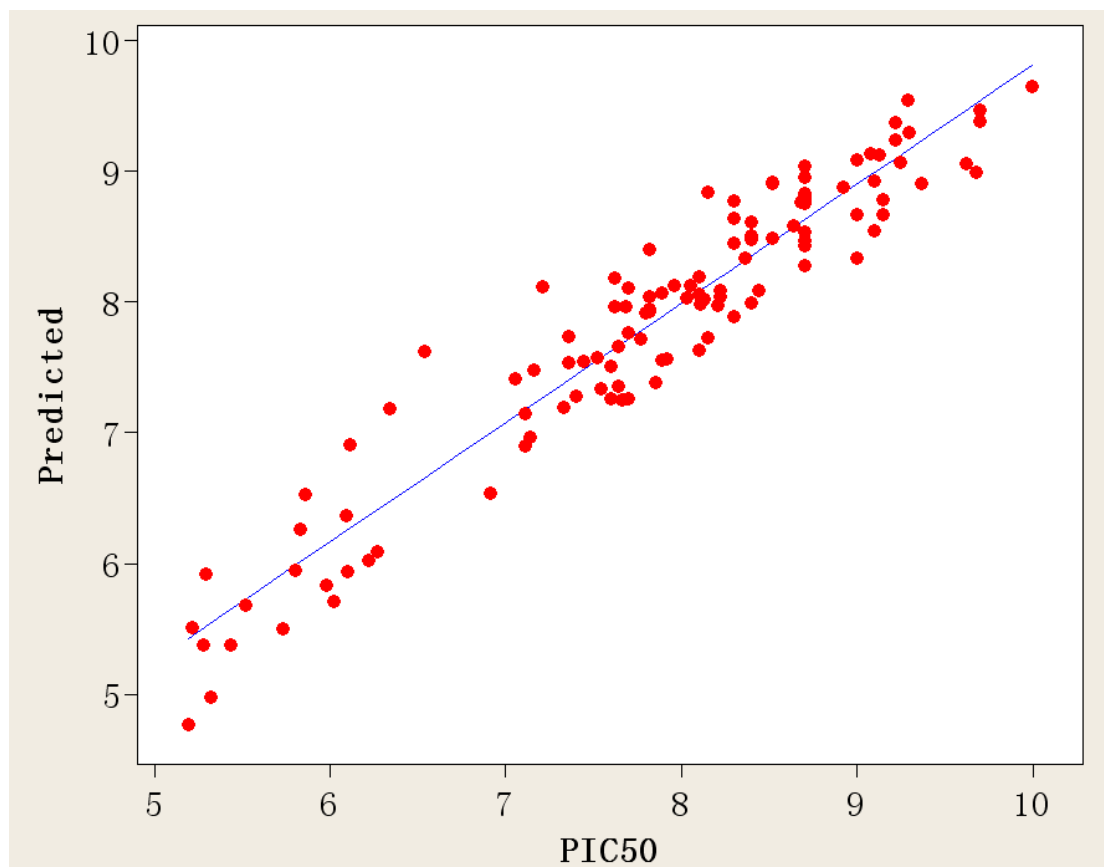
Figure 5.   Correlation between predicted activities by CoMFA model and the experimental pIC$_{50}$ values of training set
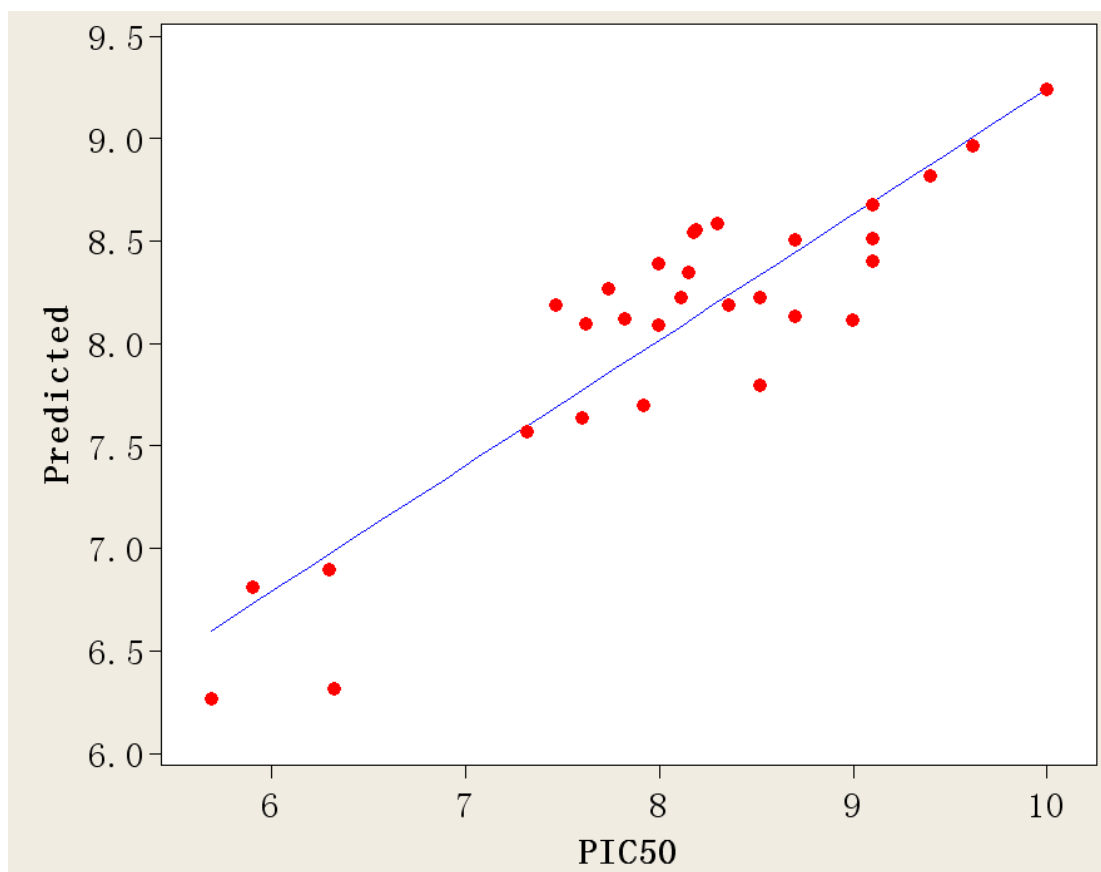
Figure 6.   Correlation between predicted activities by CoMFA model and the experimental pIC$_{50}$ values of testing set.
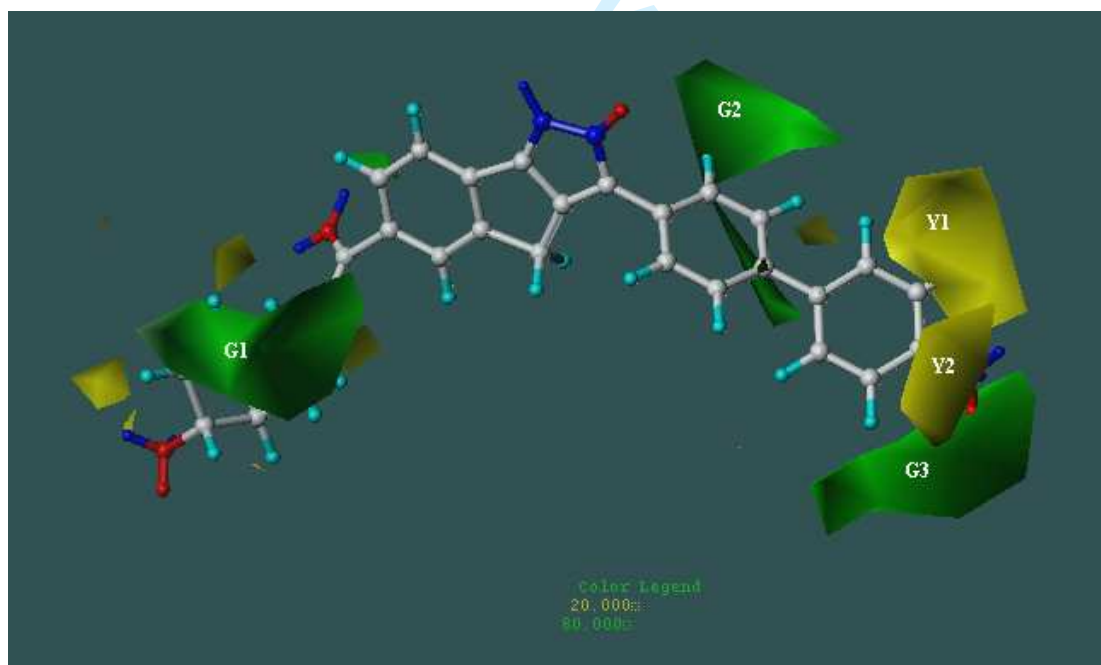
Figure 7.   CoMFA steric field distribution contour map in combination with inhibitor CHK2759M41.

Figure 8. CoMFA electrostatic field distribution contour map in combination with inhibitor CHK2759M41
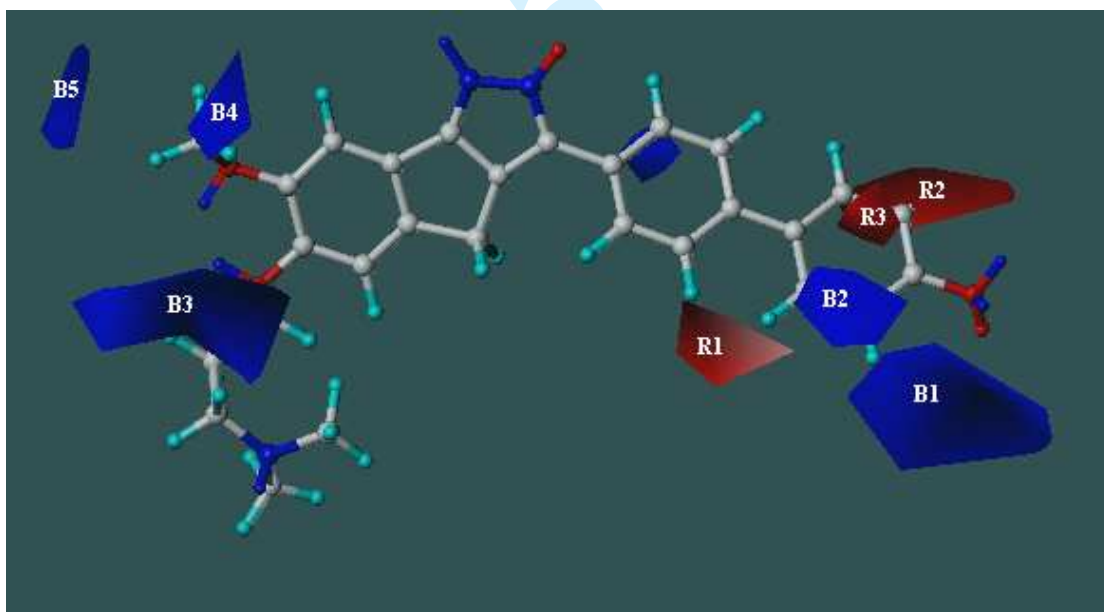
Figure 9. CoMFA electrostatic field distribution contour map in combination with inhibitor CHK4308M7
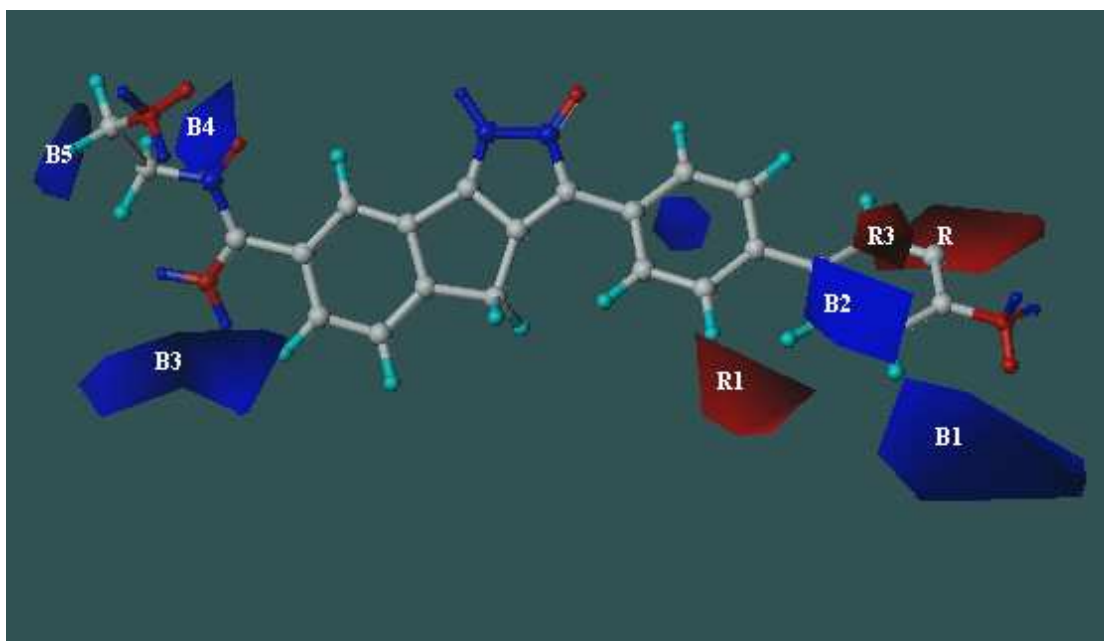
Figure 10.   CoMFA electrostatic field distribution contour map in combination with inhibitor CHK3618M24

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
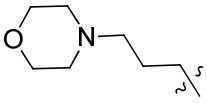41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

3D-QSAR studies of Checkpoint Kinase 1 inhibitors based on molecular docking and CoMFA

Rong-Wei Wang[a], Lu Zhou[a], Zhi-Li Zuo[b], Xiang Ma[a], Min Yang[a]

[a]College of Chemical Engineering, Sichuan University, Sichuan, Chengdu 610065,China, E-mail: zhoulu@scu.edu.cn

[b] Centre for Biomedical & Life Sciences, Singapore Polytechnic, 139651, Singapore.

Table 1

Inhibitory activity of 1,4-dihydroindeno[1,2-c]pyrazoles derivatives

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
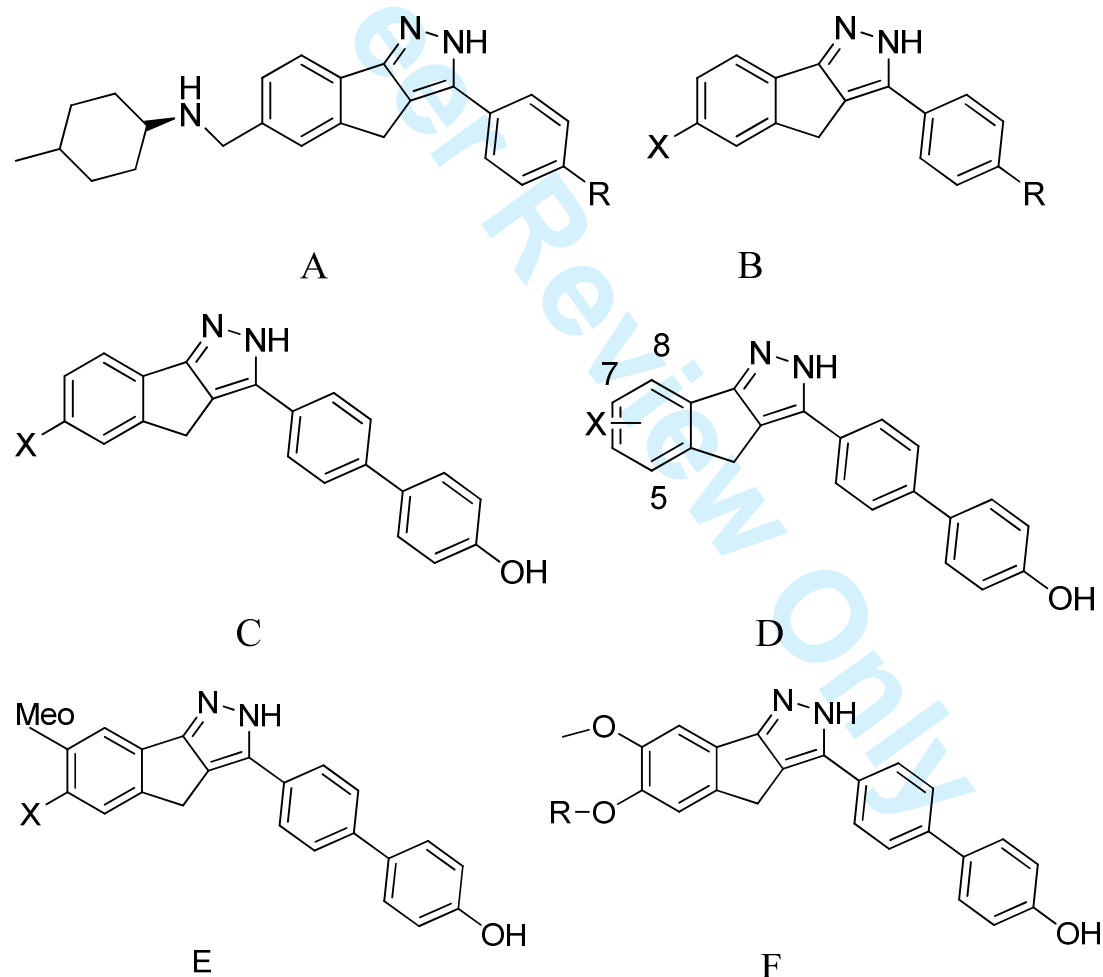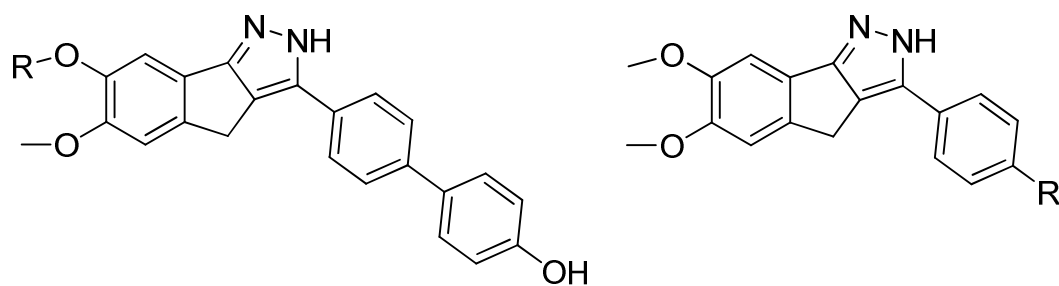
G

H
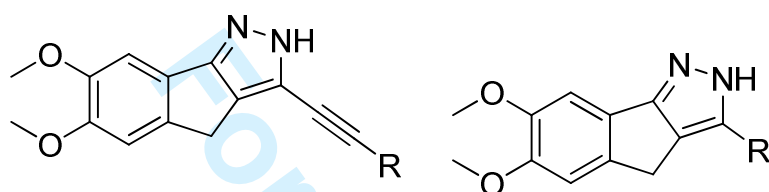
I
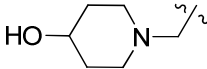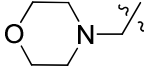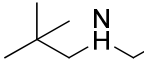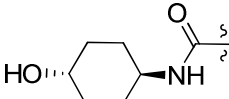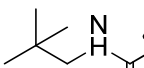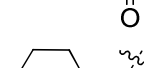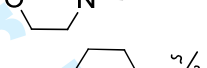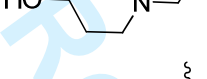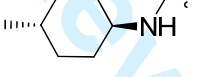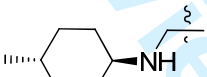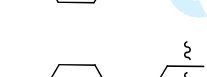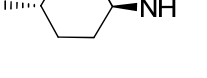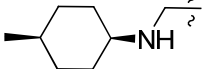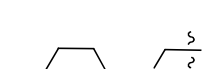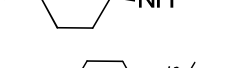
J
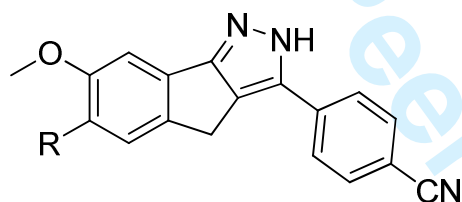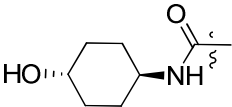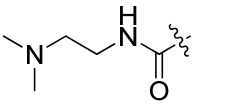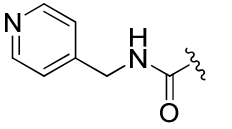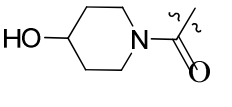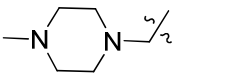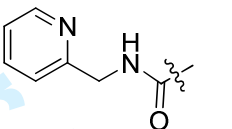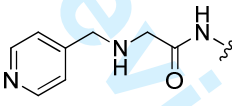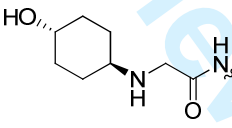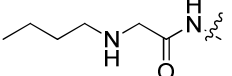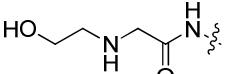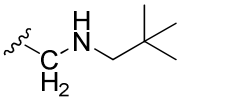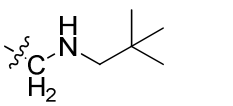
K
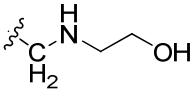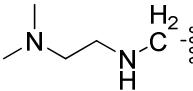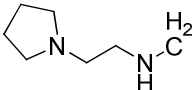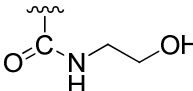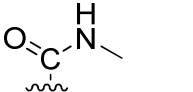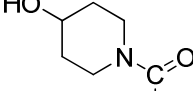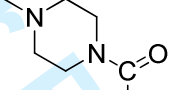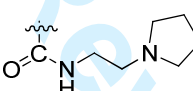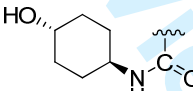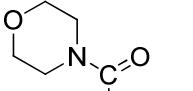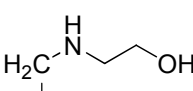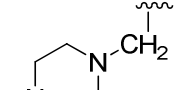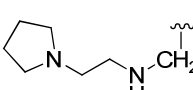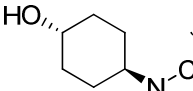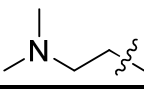
NO.	Compounds	Struc-ture	Stereoche-mistry of cyclohexyl moiety	X Posit-ion	X	R	IC$_{50}$(nM)	pIC$_{50}$
1	CHK2759M7	A	cis			F	459	6.34
2	CHK2759M8	A	trans			F	814	6.09
3	CHK2759M9	A	trans			H	1473	5.83
4	CHK2759M10	A	trans			Cl	481	6.32
5	CHK2759M11	A	trans			NH$_2$	1042	5.98
6	CHK2759M12	A	trans			CH$_2$COMe	497	6.3
7	CHK2759M13	A	trans			CH$_2$OH	790	6.1
8	CHK2759M14	A	trans			Me	966	6.02
9	CHK2759M15	A	trans			OEt	1600	5.8
10	CHK2759M16	A	trans			OCF3	2044	5.69
12	CHK2759M18	A	trans			COOMe	602	6.22
13	CHK2759M19	A	trans			OH	124	6.91

14	CHK2759M20	A	trans		COOH	22	7.66
15	CHK2759M21	A	cis		COOH	20	7.70
16	CHK2759M22	B			COOH	15	7.82
17	CHK2759M23	B			COOH	7.9	8.10
18	CHK2759M24	B			COOH	6.0	8.22
19	CHK2759M25	B			COOH	15	7.82
20	CHK2759M26	B			COOH	12	7.92
21	CHK2759M27	B			COOH	12	7.92
22	CHK2759M28	B			$CONH_2$	69	7.16
23	CHK2759M29	B			$CONH_2$	29	7.54
24	CHK2759M30	B			SO_2NH_2	535	6.27
25	CHK2759M31	B			$SO_2NHCH_2CH_2OH$	8742	5.06
26	CHK2759M32	B			NHCOMe	4825	5.32
27	CHK2759M35	B			CH_2CN	1880	5.73
28	CHK2759M36	B			CH_2COOH	6472	5.19
29	CHK2759M37	C				2.0	8.70
30	CHK2759M38	C				6.6	8.18
31	CHK2759M40	C				7.4	8.13

32	CHK2759M41	C		6.2	8.21
33	CHK2759M42	C		4.0	8.40
34	CHK2759M43	C		13	7.89
35	CHK2759M44	C		3.6	8.44
36	CHK3618M5	C	H	1373	5.86
37	CHK3618M6	C		24	7.62
38	CHK3618M7	C		9.3	8.03
39	CHK3618M8	C	OH	4.4	8.36
40	CHK3618M9	C	CH₂OH	7.7	8.11
41	CHK3618M10	C		6.4	8.19
42	CHK3618M11	C		1.2	8.92
43	CHK3618M12	C		2.1	8.68
44	CHK3618M13	C		0.74	9.13
45	CHK3618M15	D	5	CH₂OH	7.8	8.11
46	CHK3618M16	D	5		209	6.68
47	CHK3618M19	D	7	CH₂OH	4.4	8.36
48	CHK3618M20	D	7		34	7.47

49	CHK3618M21	D	7		12	7.92
50	CHK3618M22	D	7		5.2	7.82
51	CHK3618M23	D	7		7.1	8.15
52	CHK3618M24	D	7		5.0	8.300
53	CHK3618M25	D	7		2.3	8.64
54	CHK3618M33	E			0.24	9.62
55	CHK3618M34	E			2.3	8.64
56	CHK3618M35	E			0.83	9.08
57	CHK3618M36	E			0.24	9.62
58	CHK3618M37	E			0.55	9.26
59	CHK3618M38	E			0.43	9.37
60	CHK3618M39	E			0.51	9.29
61	CHK3618M40	E			0.21	9.68
62	CHK3618M41	E			0.56	9.25
63	5	F			2	8.70

64	6	F		2	8.70
65	7	F		0.7	9.15
66	8	F		2	8.70
67	9	F		0.8	9.10
68	10	F		1	9.00
69	11	F		4	8.40
70	12	F		2	8.70
71	13	F		6	8.22
72	14	F		25	7.60
73	15	F		8	8.10
74	16	F		8	8.10
75	17	F		18	7.74
76	18	F		20	7.70
77	19	F		5	8.30
78	20	F		13	7.89

79	21	F		30	7.52
80	22	F		23	7.64
81	23	F		62	7.21
82	24	F		5	8.30
83	26	F		0.1	10.00
84	27	F		0.1	10.00
85	28	F		0.8	9.10
86	29	F		0.2	9.70
87	30	F		0.4	9.40
88	31	F		0.7	9.15
89	32	F	H	1	9.00
90	33	G	H	5	8.30
91	34	G		0.8	9.10
92	35	G		2	8.70
93	36	G		0.6	9.22
94	37	G		2	8.70
95	38	G		0.8	9.10
96	39	G		0.2	9.70
97	40	G		0.6	9.22
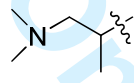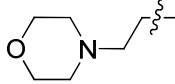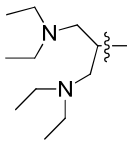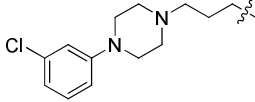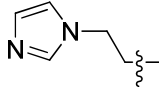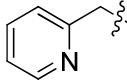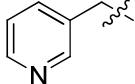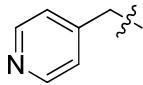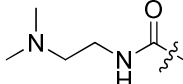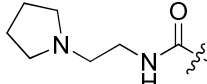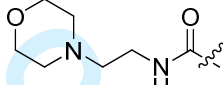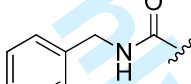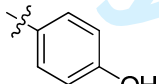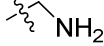
98	41	G		5	8.30
99	42	G		2	8.70
100	43	G		10	8.00
101	44	G		3	8.52
102	45	G		24	7.62
103	46	G		23	7.64
104	47	G		36	7.44
105	48	G		40	7.40
106	49	G		21	7.68
107	50	G		74	7.13
108	51	G		48	7.32
109	CHK5944M2	H		2	8.70
110	CHK5944M3	H	OH	25	7.60
111	CHK5944M4	H	CN	7	8.15
112	CHK5944M5	H	COOH	2	8.70
113	CHK5944M6	H		25	7.60
114	CHK5944M7	H		779	6.11

115	CHK5944M8	H		355	6.45
116	CHK5944M10	H	Br	73	7.14
117	CHK5944M13	H		6223	5.21
118	CHK5944M15	H		11	7.96
119	CHK5944M16	H		47	7.33
120	CHK5944M17	H		77	7.11
121	CHK5944M18	H		1260	5.90
122	CHK5944M19	H		1374	5.86
123	CHK5944M20	H		3709	5.43
124	CHK5944M21	H		3003	5.52
125	CHK5944M22	I	H	5140	5.29
126	CHK5944M24	I		5	8.30
127	CHK5944M28	J		0.5	9.30
128	CHK5944M29	J		0.8	9.10
129	CHK5944M30	J		15	7.82
130	CHK5944M31	J		4	8.40
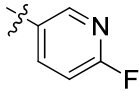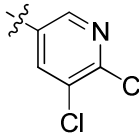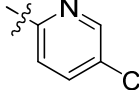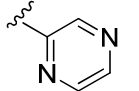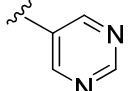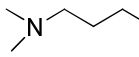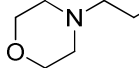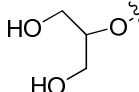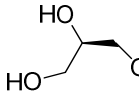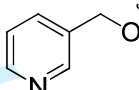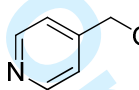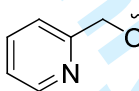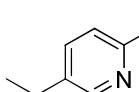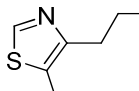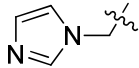
131	CHK5944M32	J		2	8.70
132	CHK5944M35	J		16	7.80
133	CHK5944M37	J		44	7.36
134	CHK5944M38	J		89	7.05
135	CHK5944M39	J		44	7.36
136	CHK5944M40	K		4	8.40
137	CHK5944M41	K		3	8.52
138	CHK5944M42	K		3	8.52
139	CHK5944M43	K		6	8.22
140	CHK5944M44	K		14	7.85
141	CHK5944M45	K		20	7.70
142	CHK5944M46	K		10	8.00
143	CHK5944M47	K		77	7.11
144	CHK5944M48	K		9	8.05
145	CHK5944M49	K		2	8.70
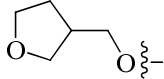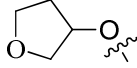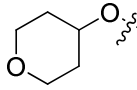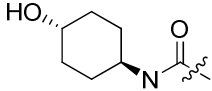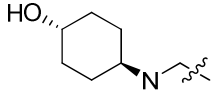
146	CHK5944M50	K		3	8.52
147	CHK5944M51	K		2	8.70
148	CHK5944M52	K		1	9.00
149	CHK5944M53	K		3	8.52
150	CHK5944M54	K		17	7.77
151	CHK5944M55	K	OH	2	8.70

Table 2

Actual and predicted inhibitory activities (pIC_{50}) and residuals of the training set molecules

NO	Compound numbers	ChemScore	pIC_{50}	Predicted	Residuals
1	10	43.39	9.000	8.822	0.178
2	11	43.56	8.398	8.961	-0.563
3	12	45.02	8.699	8.781	-0.082
4	13	45.24	8.222	8.323	-0.101
5	14	42.53	7.602	7.353	0.249
6	15	45.04	8.097	8.314	-0.217
7	16	45.82	8.097	8.098	-0.001
8	18	44.73	7.699	7.969	-0.270
9	19	43.50	8.301	8.425	-0.124
10	20	43.83	7.886	8.214	-0.328
11	21	43.31	7.523	7.385	0.138
12	22	44.77	7.638	7.691	-0.053
13	23	44.26	7.208	7.831	-0.623
14	24	43.63	8.301	8.154	0.147
15	27	43.21	10.000	9.689	0.311
16	28	45.98	9.097	9.006	0.091
17	29	42.83	9.699	9.316	0.383
18	31	44.35	9.155	8.839	0.316
19	32	45.62	9.000	9.034	-0.034

20	33	45.48	8.301	8.765	-0.464
21	36	42.49	9.222	9.066	0.156
22	37	44.29	8.699	8.778	-0.079
23	39	43.96	9.699	9.421	0.278
24	40	41.33	9.222	9.391	-0.169
25	42	39.12	8.699	8.522	0.177
26	44	42.31	8.522	8.773	-0.251
27	46	44.00	7.638	7.271	0.367
28	47	44.18	7.444	7.288	0.156
29	48	44.80	7.398	7.112	0.286
30	49	43.99	7.678	7.911	-0.233
31	5	43.12	8.699	8.781	-0.082
32	50	44.29	7.131	7.673	-0.542
33	7	43.86	9.155	8.893	0.262
34	8	46.22	8.699	8.717	-0.018
35	9	43.40	9.099	8.398	0.701
36	CHK2759M11	37.51	5.982	5.739	0.243
37	CHK2759M13	39.45	6.102	6.124	-0.022
38	CHK2759M14	38.91	6.015	5.662	0.353
39	CHK2759M15	40.31	5.796	6.002	-0.206
40	CHK2759M17	40.52	5.279	5.321	-0.042
41	CHK2759M18	40.42	6.220	6.127	0.093
42	CHK2759M19	35.72	6.907	6.491	0.416
43	CHK2759M20	37.69	7.658	7.601	0.057
44	CHK2759M21	39.25	7.699	7.454	0.245
45	CHK2759M22	35.38	7.823	7.91	-0.087
46	CHK2759M23	36.94	8.102	7.737	0.365
47	CHK2759M24	37.05	8.222	7.96	0.262
48	CHK2759M26	38.37	7.921	8.524	-0.603
49	CHK2759M28	36.08	7.161	7.541	-0.380
50	CHK2759M29	35.60	7.538	7.501	0.037
51	CHK2759M30	38.20	6.272	6.254	0.018
52	CHK2759M31	40.43	5.058	5.137	-0.079
53	CHK2759M32	40.03	5.317	5.384	-0.067
54	CHK2759M35	41.73	5.726	5.802	-0.076
55	CHK2759M36	39.07	5.189	4.842	0.347
56	CHK2759M37	41.12	8.699	8.531	0.168
57	CHK2759M40	47.68	8.131	8.081	0.050
58	CHK2759M41	48.52	8.208	8.099	0.109
59	CHK2759M42	44.21	8.398	8.548	-0.150
60	CHK2759M43	44.35	7.886	7.634	0.252
61	CHK2759M44	40.62	8.444	8.421	0.023
62	CHK2759M7	38.46	6.338	6.925	-0.587
63	CHK2759M8	37.26	6.089	6.192	-0.103

64	CHK2759M9	37.74	5.832	5.977	-0.145
65	CHK3618M11	46.11	8.921	8.981	-0.060
66	CHK3618M12	45.50	8.678	8.711	-0.033
67	CHK3618M13	46.36	9.131	9.167	-0.036
68	CHK3618M16	46.18	6.680	7.246	-0.566
69	CHK3618M19	44.35	8.357	8.412	-0.055
70	CHK3618M21	41.66	7.921	7.436	0.485
71	CHK3618M22	43.65	7.824	7.633	0.191
72	CHK3618M24	39.99	8.301	8.615	-0.314
73	CHK3618M25	41.79	8.638	8.593	0.045
74	CHK3618M33	43.40	9.619	9.066	0.553
75	CHK3618M34	42.42	8.638	8.702	-0.064
76	CHK3618M35	44.99	9.081	8.908	0.173
77	CHK3618M37	43.87	9.259	9.063	0.196
78	CHK3618M38	45.76	9.367	8.817	0.550
79	CHK3618M39	40.95	9.292	9.599	-0.307
80	CHK3618M40	46.44	9.678	9.022	0.656
81	CHK3618M41	47.03	9.252	9.141	0.111
82	CHK3618M5	43.65	5.862	6.021	-0.159
83	CHK3618M6	42.84	7.620	7.936	-0.316
84	CHK3618M7	44.76	8.032	8.165	-0.133
85	CHK3618M9	44.00	8.113	7.905	0.208
86	CHK5944M10	36.01	7.137	6.899	0.238
87	CHK5944M13	39.24	5.206	5.190	0.016
88	CHK5944M15	40.73	7.959	8.139	-0.180
89	CHK5944M16	42.84	7.328	7.233	0.095
90	CHK5944M17	40.42	7.114	6.802	0.312
91	CHK5944M19	37.50	5.862	6.385	-0.523
92	CHK5944M2	43.02	8.699	8.902	-0.203
93	CHK5944M20	38.68	5.431	5.164	0.267
94	CHK5944M21	39.59	5.522	5.651	-0.129
95	CHK5944M22	38.12	5.289	5.579	-0.290
96	CHK5944M24	34.22	8.301	8.125	0.176
97	CHK5944M28	40.54	9.301	9.472	-0.171
98	CHK5944M30	32.86	7.824	7.752	0.072
99	CHK5944M31	34.64	8.398	8.003	0.395
100	CHK5944M32	32.69	8.699	8.351	0.348
101	CHK5944M35	35.13	7.796	7.870	-0.074
102	CHK5944M37	32.86	7.357	7.629	-0.272
103	CHK5944M38	32.11	7.051	7.285	-0.234
104	CHK5944M39	29.26	7.357	7.315	0.042
105	CHK5944M4	34.32	8.155	7.742	0.413
106	CHK5944M40	34.03	8.398	8.615	-0.217
107	CHK5944M42	36.51	8.523	8.816	-0.293

108	CHK5944M43	38.50	8.222	8.383	-0.161
109	CHK5944M44	36.64	7.854	7.367	0.487
110	CHK5944M45	35.62	7.699	7.773	-0.074
111	CHK5944M47	36.19	7.114	7.302	-0.188
112	CHK5944M48	36.37	8.046	8.445	-0.399
113	CHK5944M49	35.47	8.699	8.705	-0.006
114	CHK5944M5	34.87	8.699	8.571	0.128
115	CHK5944M51	34.65	8.699	8.417	0.282
116	CHK5944M53	35.03	8.523	8.421	0.102
117	CHK5944M54	36.83	7.770	7.947	-0.177
118	CHK5944M55	36.39	8.699	8.624	0.075
119	CHK5944M6	37.32	7.602	7.279	0.323
120	CHK5944M7	35.30	6.108	6.811	-0.703
121	CHK5944M8	34.26	6.450	6.235	0.215

Table 3

Actual and predicted inhibitory activities (pIC_{50}) and residuals of the test set molecules

NO	Compound numbers	ChemScore	pIC_{50}	Predicted	Residuals
1	17	46.45	7.745	8.265	-0.52
2	26	45.89	10.000	9.239	0.761
3	30	44.59	9.398	8.818	0.58
4	34	43.10	9.097	8.519	0.578
5	35	42.07	8.699	8.505	0.194
6	38	41.92	9.097	8.676	0.421
7	41	42.01	8.301	8.585	-0.284
8	43	45.35	8.000	8.387	-0.387
9	45	43.44	7.620	8.096	-0.476
10	51	45.01	7.319	7.569	-0.25
11	6	42.63	8.699	8.135	0.564
12	CHK2759M10	39.28	6.318	6.311	0.007
13	CHK2759M12	42.81	6.304	6.897	-0.593
14	CHK2759M16	36.23	5.690	6.265	-0.575
15	CHK2759M25	38.37	7.824	8.122	-0.298
16	CHK2759M27	39.12	7.921	7.697	0.224
17	CHK2759M38	45.28	8.180	8.542	-0.362
18	CHK3618M10	47.64	8.194	8.554	-0.36
19	CHK3618M15	44.39	8.108	8.225	-0.117
20	CHK3618M20	42.75	7.469	8.187	-0.718

21	CHK3618M23	42.42	8.149	8.348	-0.199
22	CHK3618M36	43.40	9.620	8.965	0.655
23	CHK3618M8	44.67	8.357	8.199	0.158
24	CHK5944M18	41.29	5.900	6.813	-0.913
25	CHK5944M29	32.61	9.097	8.496	0.601
26	CHK5944M3	33.18	7.602	7.637	-0.035
27	CHK5944M41	36.79	8.523	8.225	0.298
28	CHK5944M46	35.77	8.000	8.089	-0.089
29	CHK5944M50	36.97	8.523	7.794	0.729
30	CHK5944M52	34.22	9.000	8.116	0.884

Table 4. Summary of CoMFA analysis.

Parameters	CoMFA
Training set	-
PLS statistics	-
LOO	-
q^2 (CV correlation coefficient)	0.534
N (number of components)	6
Uncross-validated	-
r2 (correlation coefficient)	0.911
SEE (standard error of estimate)	0.352
F (F-ratio)	187.106
Field distribution	-
Steric	43.0%
Electrostatic	57.0%
Testing set	-
r^2 (correlation coefficient)	0.812
S (standard error of prediction)	0.313