Supplementary Material for New Journal of Chemistry

Drug Design and Molecular Docking Simulations of Polo-like Kinase 1 Inhibitors Based on QSAR Study

——Background Information with Discussion and Asumptions

Jianbo TONG[a, b①] Tianhao WANG[a, b] Yi FENG[a, b]

[a] (*College of Chemistry and Chemical Engineering,*

Shaanxi University of Science and Technology, Xi'an 710021, *China*)

[b] (*Key Laboratory of Chemical Additives for China National Light Industry, Xi'an* 710021, *China*)

1. Supplementary figure

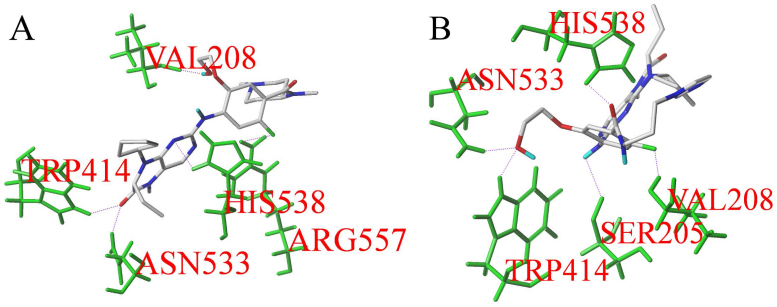


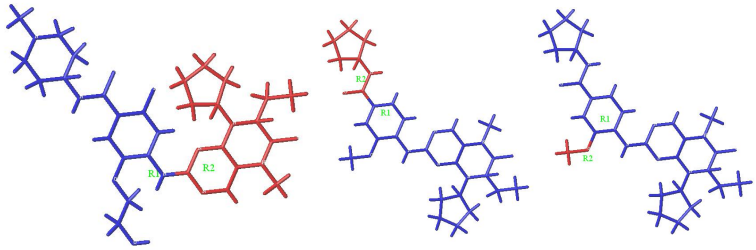
Figure S1. The hydrogen-bond interaction between 6GY2 and compounds, (A) N3 and (B) N8

(the purple dotted line represented hydrogen bonding)

2. The variables of QSAR model

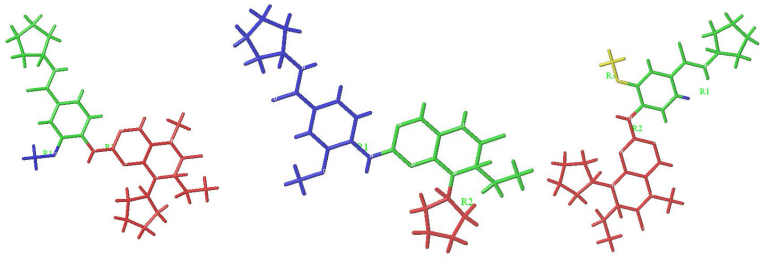
2.1 Different cutting styles

We have tried many different cutting styles, and then choose the best one. The cutting styles and results we tried are shown in Table S1. It can be seen that the cutting style in the paper can get the best result. These data are as follows,



1 2 3

①Corresponding author. Professor. E-mail: jianbotong@aliyun.com



4 5 6

Figure S2. Different cutting styles of template molecule

Table S1. Comparison of results of Topomer CoMFA model

Number of R groups	Cutting style	LOO		Non-cross-validation		
		q^2	N	r^2	SEE	F
R=1	1	0.271	2	0.716	0.374	37.764
	2	0.445	5	0.931	0.195	72.794
	3	0.356	4	0.901	0.229	63.374
R=2	4	0.374	4	0.758	0.358	21.927
	5	0.331	5	0.753	0.368	16.467
R=3	6	0.366	3	0.761	0.349	30.778

2.2 Different training sets and test sets

We have tried a lot of models, such as changing different training sets and test sets or using different cutting styles, and finally selected the best model. The modeling results of different training sets and test sets are as follows，

Table S2. The modeling results of different training sets and test sets

Cutting mode	Number of molecules in the test set	Molecules the test set	LOO		Non-cross-validation		
			q^2	N	r^2	SEE	F
Split in two	5	6, 13, 20, 27, 34	0.438	7	0.953	0.159	76.136
	7	5, 10, 15, 20, 25, 30, 35	0.432	5	0.914	0.219	55.173
	6	6, 12, 18, 24, 30, 36	0.501	7	0.977	0.118	148.680

2.3 The chemical diversity of the training set and test set

We have considered using molecules containing an iso-propyl unity as test set. For example, we try to replace two molecules (25 and 39) containing an iso-propyl unity instead of molecules 24 and 36 as the test set molecules, and the rest are the training set. The molecules of the new test set are 6, 12, 18, 25, 30 and 39, and the results are shown in Table S3.

Table S3. Comparison of results of Topomer CoMFA model (Different test set molecules)

Test set molecules	LOO		Non-cross-validation		
	q^2	N	r^2	SEE	F
6, 12, 18, **24**, 30, **36**	0.501	7	0.977	0.118	148.680
6, 12, 18, **25**, 30, **39**	0.446	10	0.974	0.129	83.842

The q^2 and r^2 of the new model have reduced and the SEE (standard error of estimate) value has increased compared with the model in our paper, which is detrimental to the stability and predictability of the model. So the division of date set in the original paper can get a better QSAR model.

3. Superimposition of the reference ligand

To validate the docking reliability, the crystal structure of protein (6GY2) with the cognate ligand was re-docked. As a reference ligand, the cognate ligand was taken out of its protein-ligand complex (6GY2) and re-docked back into its binding site. In Fig the redocked and reference ligands are almost completely superimposed together. Their rotational tendency is basically similar. The result shows that the docking method is reasonable and reliable.

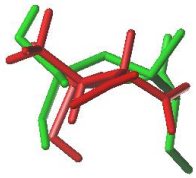


Figure S3. Superimposition of the reference ligand

(The green stick represents the re-docked ligand, and the red stick shows the reference ligand)