Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information

Discovery of Potential RIPK1 Inhibitors by Machine Learning and Molecular Dynamics Simulations

Ji-xiang Liu[a], Ri-song Na[b], Lian-juan Yang[c], Xu-ri Huang[a], Xi Zhao[*a]

a, Institute of Theoretical Chemistry, College of Chemistry, Jilin University, Liutiao Road #2, Changchun 130021, China.

b, Collaborative Innovation Center of Henan Grain Crops, National Key Laboratory of Wheat and Maize Crop Science, College of Plant Protection, Henan Agricultural University, Zhengzhou 450002, China.

c, Department of Medical Mycology, Shanghai Skin Disease Hospital, Tongji University School of Medicine, Shanghai, 200443, China.

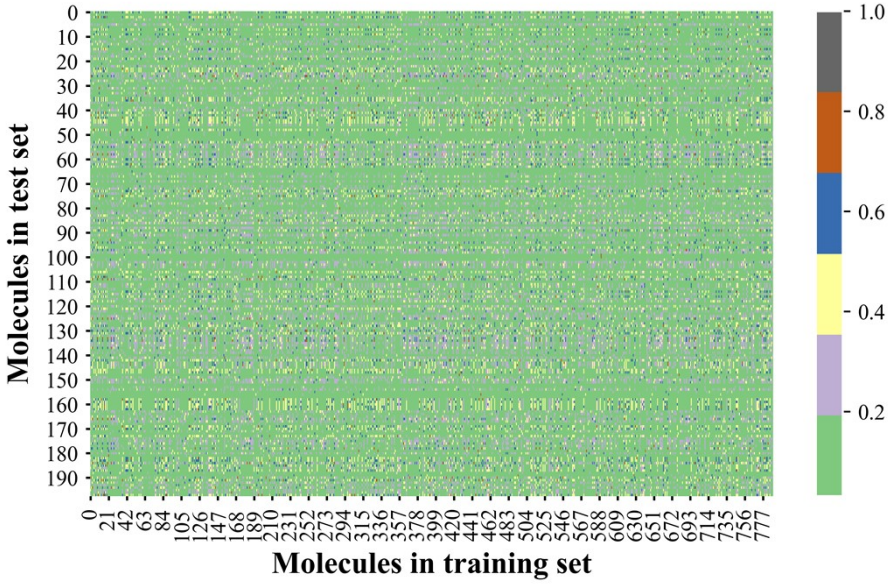

Figure S1. Comparison of molecular similarity between the training set and the test set and the closer to 1, the higher the similarity.

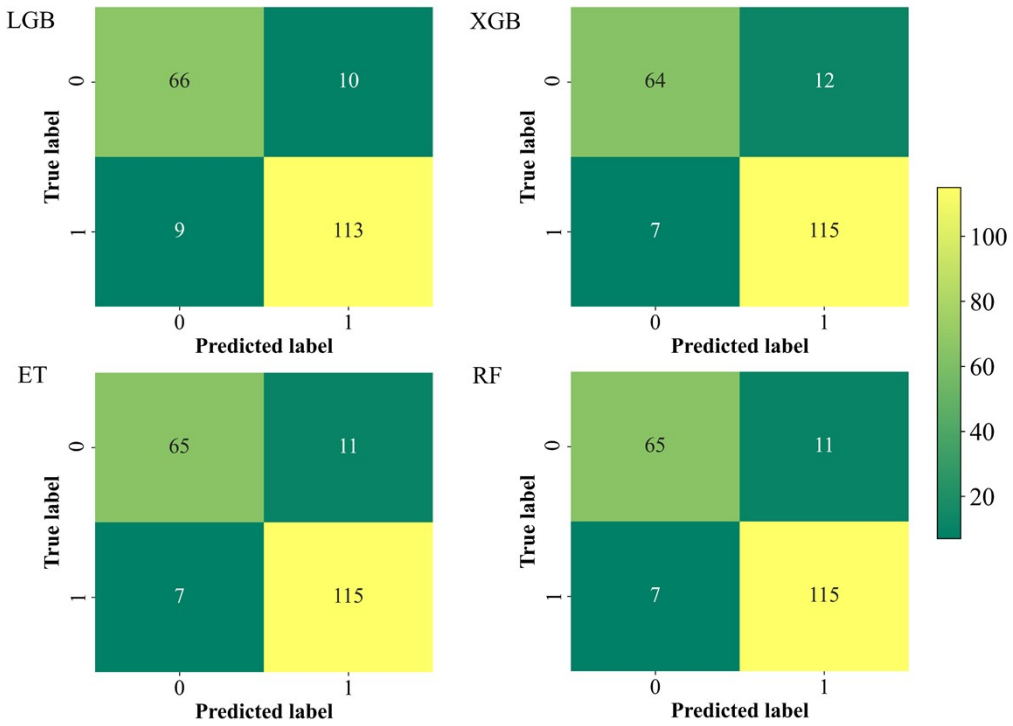

Figure S2. Confusion matrix of the test set of the four models.

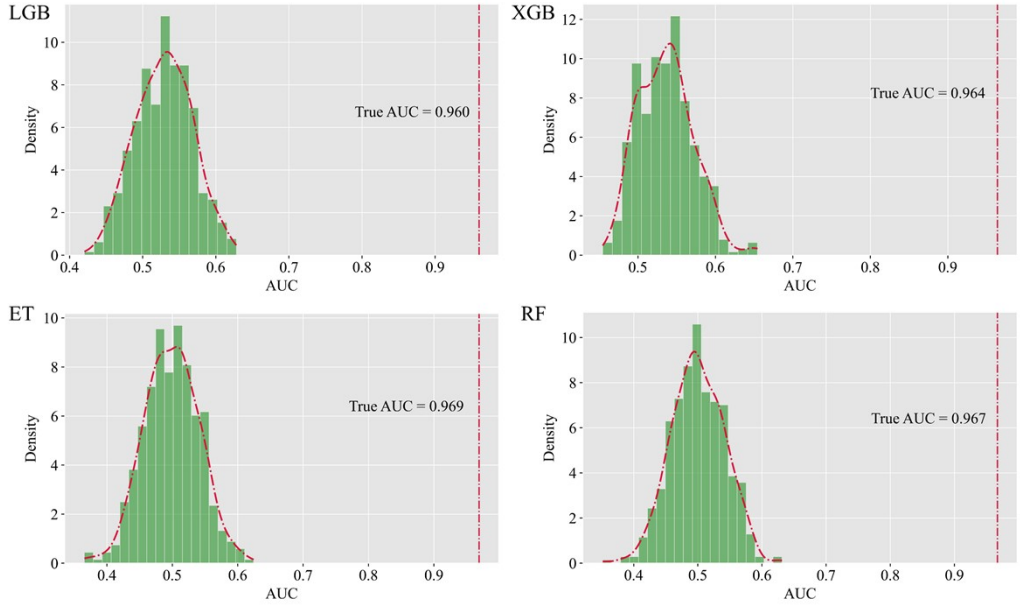

Figure S3. Distribution of AUC for the four randomized models compared with the true model in the Y-randomization test. The red vertical dashed line on the right represents the AUC of the true model, and the distribution on the left side represents the distribution of AUC values of randomized models.

A total of 500 random prediction models were constructed and a series of AUC values were obtained. The distributions of Y-randomized AUC of the predictive models and the fitted curves for RIPK1 activity prediction are shown in **Figure S3**. Compared

with the AUC of the original model, the AUC values of randomly shuffled models varied from 0.3656 to 0.6543 there is a statistically significant difference.

Table S1. Optimization Parameters

Models	Parameters	Meaning	Values
LGB	max_depth	Tree depth	9
	eta	Learning rate	0.1
	objective	Loss function	binary
	metric	Evaluation metrics	auc
	feature_fraction	Sampling ratio	0.7267
	lambda_l1	Regularization	1
	num_leaves	Number of leaves	34
	num_round	Max. number of iterations	1000
XGB	max_depth	Tree depth	6
	eta	Learning rate	0.1
	objective	Loss function	binary:logistic
	eval_metric	Evaluation metrics	auc
	lambda	Regularization	1
	colsample_bytree	Column sampling	0.8
	subsample	Sampling ratio	0.8
	gamma	Node splitting	0.1
	min_child_weight	Minimum leaf weights	1
	num_round	Max. number of iterations	1000
ET	max_depth	Tree depth	19
	max_features	Sampling ratio	0.0997
	min_samples_split	Minimum sample split	59
	n_estimators	Number of decision trees	104
RF	max_depth	Tree depth	10
	max_features	Sampling ratio	0.1
	min_samples_split	Minimum sample split	27
	n_estimators	Number of decision trees	72

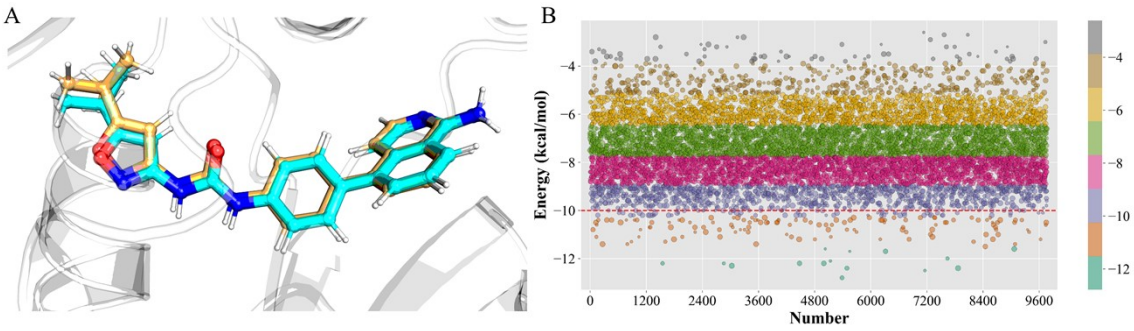

Figure S4. (A) Conformational overlay of the re-docked. Cyan is the re-docked molecule. (B) The distribution of the docking results for all molecules, where the red dashed line is energy = -10 (kcal/mol).

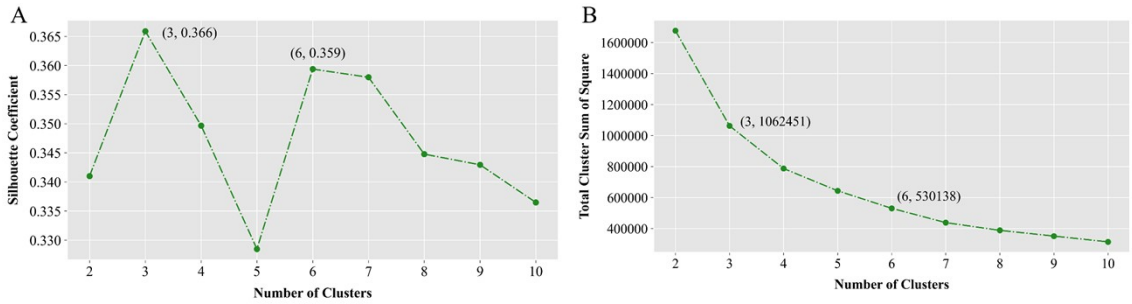

Figure S5. Evaluation metrics for K-means clustering. (A) Line plot of silhouette coefficients, (B) line plot of total cluster sum of square.

For clustering, the larger the value of the silhouette coefficient the better, and the smaller the total cluster sum of square the better, so according to **Figure S5** we consider six classes to be more reasonable.

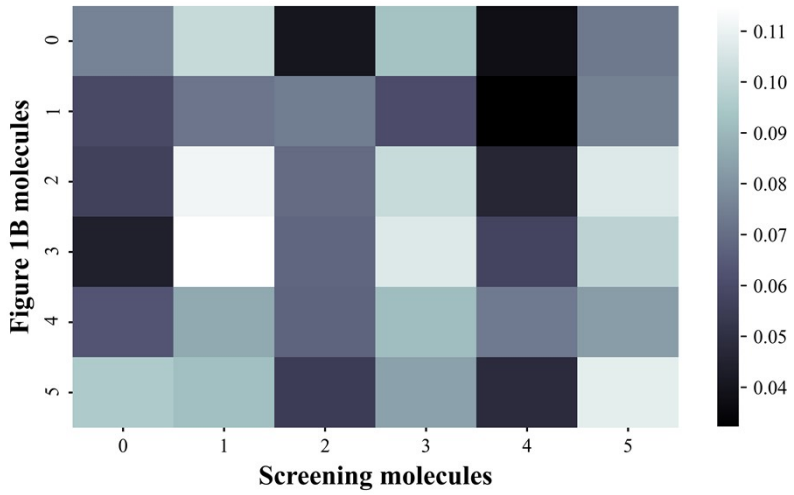

Figure S6. The similarity of the six selected molecules to the molecules listed in Figure 1B.

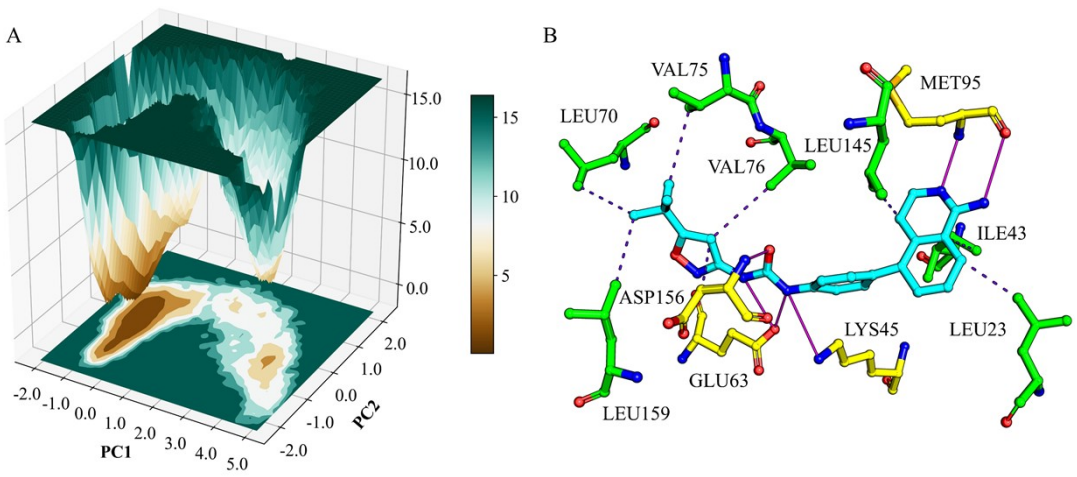

Figure S7. (A) The FEL of the co-crystal complex system. (B) The interaction of the co-crystal complex system.

Table S2. Key Interactions in the Seven Complexes During the Simulation.

Molecules	Residues	Occupancy (%)	Interaction	Average number of hydrogen bonds
Cpd-1	Ser25	21.8	H-bond	1.168
	Asn99	20.5	H-bond	
	Glu142	42.8	H-bond	
Cpd-2	Lys45	10.1	H-bond	2.132
	Glu63	69.6	Salt bridge	
	Asp156	110.0	H-bond	
Cpd-3	Asp24	80.6	H-bond	2.156
	Asn99	20.4	H-bond	
	Asp156	90.1	H-bond	
Cpd-4	Met95	59.9	H-bond	1.928
	Asp156	74.8	H-bond	
	Ala313	37.3	H-bond	
Cpd-5	Glu63	82.7	H-bond	1.763
	Met95	85.2	H-bond	
Cpd-6	Glu63	40.5	H-bond	1.600
	Val76	20.4	H-bond	
	Asp156	87.9	H-bond	

	Glu63	196.3	H-bond	
4NEU	Met95	176.9	H-bond	4.619
	Asp156	88.5	H-bond	

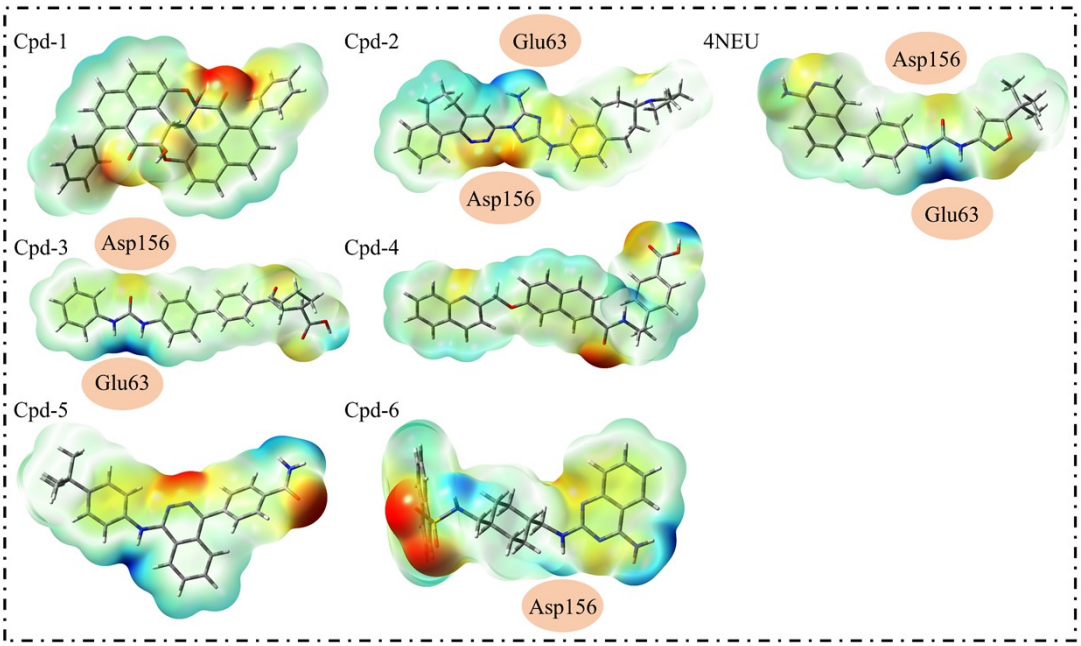

Figure S8. Electrostatic potential graph of seven small molecules. Blue represents the positive potential, red represents the negative potential, the darker the color the greater the potential, light pink ovals represent acidic amino acids.

Table S3. The various decomposition energies of the important residues in the seven complexes (kJ/mol)

Residues	Cpd-1	Cpd-2	Cpd-3	Cpd-4	Cpd-5	Cpd-6	4NEU
Leu23	0.219[a]	-0.010	-0.239	0.477	0.164	0.133	0.248
	-6.530[b]	-3.105	-12.669	-4.092	-2.544	-8.829	-3.419
	0.662[c]	0.352	6.924	0.011	0.126	0.732	0.070
	-0.994[d]	-0.527	-1.708	-0.887	-0.543	-1.602	-0.543
	-6.643[e]	-3.382	-7.692	-4.491	-2.797	-9.566	-3.644
Val31	-0.098	-0.144	0.055	-0.031	0.290	0.040	-0.157
	-3.756	-3.590	-5.706	-2.571	-4.483	-2.740	-4.168
	0.570	0.978	0.882	0.307	0.590	0.507	0.671
	-0.408	-0.311	-0.525	-0.320	-0.553	-0.213	-0.426
	-3.692	-3.066	-5.294	-2.615	-4.155	-2.406	-4.080
Ile43	0.081	-0.209	-0.270	-0.382	-0.334	0.049	-0.298
	-4.230	-6.026	-3.139	-8.132	-7.368	-4.969	-7.183
	0.383	0.546	0.945	0.099	1.022	0.348	0.487
	-0.668	-0.851	-0.124	-0.690	-0.576	-0.703	-0.455
	-4.433	-6.540	-2.588	-9.105	-7.255	-5.275	-7.448
Glu63	1.916	-7.496	-9.479	-1.770	-16.386	0.516	-25.845
	-0.378	-1.406	-3.444	-0.695	-3.406	-0.804	-1.741
	1.723	-1.648	24.723	3.522	29.561	-0.518	30.434
	-0.038	-0.070	-0.246	-0.002	-0.277	-0.013	-0.306
	3.223	-10.621	11.554	1.055	9.492	-0.818	2.541
Met67	-0.012	-1.001	-1.563	-0.624	0.880	-0.987	-0.070
	-1.169	-8.268	-7.591	-7.423	-7.038	-7.356	-8.679
	0.531	2.580	2.843	3.087	1.709	3.453	0.761
	-0.203	-0.644	-0.556	-0.497	-0.457	-0.699	-0.406
	-0.852	-7.333	-6.866	-5.458	-4.906	-5.588	-8.395
Val75	-0.107	0.388	-0.201	-0.506	-0.060	-0.330	0.127
	-0.180	-2.298	-1.757	-3.479	-2.901	-2.774	-3.411
	0.091	0.047	0.433	-0.624	0.834	-0.139	0.225
	-0.000	-0.135	-0.136	-0.327	-0.110	-0.343	-0.194
	-0.196	-1.999	-1.661	-4.936	-2.237	-3.586	-3.253
Val76	-0.131	-0.604	-0.426	-1.827	-0.641	-4.169	-0.677
	-3.180	-3.835	-5.449	-8.800	-8.043	-6.539	-6.568
	0.414	0.303	1.291	8.672	0.511	8.598	2.915
	-0.263	-0.133	-0.200	-0.520	-0.368	-0.447	-0.230
	-3.160	-4.270	-4.783	-2.476	-8.540	-2.558	-4.560
Met92	-0.239	0.201	-1.287	0.095	0.309	0.262	-0.673
	-3.303	-5.482	-4.029	-6.462	-9.285	-5.943	-5.537
	1.463	0.830	2.487	1.504	1.350	2.166	1.941
	-0.349	-0.458	-0.323	-0.446	-0.580	-0.580	-0.410
	-2.428	-4.909	-3.152	-5.310	-8.207	-4.095	-4.679

Asn143	-1.010	0.027	-0.803	0.020	0.605	0.103	0.060
	-7.017	-0.970	-9.804	-0.687	-0.805	-2.074	-0.843
	3.372	-0.307	8.605	0.224	-0.962	0.724	0.356
	-0.192	-0.002	-0.800	-0.003	0.000	-0.183	-0.009
	-4.847	-1.252	-2.803	-0.446	-1.162	-1.431	-0.437
Leu145	-0.140	-0.118	-0.053	-0.609	-0.474	-0.283	-0.444
	-8.565	-5.039	-3.697	-8.086	-6.745	-6.312	-6.640
	0.805	0.514	0.591	0.471	0.652	1.038	0.745
	-1.114	-1.064	-0.265	-1.034	-0.999	-1.031	-0.973
	-9.013	-5.708	-3.424	-9.258	-7.566	-6.589	-7.312
Ala155	-1.407	-2.250	-4.154	-2.379	-3.280	-1.963	-2.393
	-3.909	-7.775	-6.344	-6.615	-7.231	-7.271	-6.635
	1.464	3.759	2.798	3.313	4.691	3.526	3.077
	-0.472	-0.272	-0.345	-0.516	-0.440	-0.690	-0.302
	-4.324	-6.502	-8.044	-6.197	-6.259	-6.399	-6.253
Asp156	3.091	-4.649	0.648	-2.091	0.233	-1.138	-1.384
	-2.915	-9.977	-6.366	-5.333	-9.992	-6.904	-11.022
	1.911	17.543	2.664	7.430	13.097	6.141	17.369
	-0.288	-1.126	-0.441	-0.285	-0.626	-0.541	-0.750
	1.799	1.791	-3.496	-0.279	2.712	-2.441	4.214
Leu157	0.156	0.997	0.129	0.169	-0.063	0.082	-0.004
	-8.561	-9.451	-4.223	-3.705	-1.592	-6.161	-4.001
	0.723	0.078	0.706	0.696	0.170	0.389	0.467
	-1.039	-0.570	-0.113	-0.535	-0.095	-0.876	-0.413
	-8.720	-8.947	-3.502	-3.375	-1.580	-6.565	-3.951
Leu159	-0.081	0.577	-1.377	-0.135	0.581	-0.266	-0.324
	-0.242	-4.711	-7.269	-4.808	-4.341	-1.723	-2.183
	-0.180	0.393	2.072	-0.420	-0.645	0.353	-0.347
	-0.006	-0.489	-0.522	-0.341	-0.399	-0.249	-0.155
	-0.509	-4.230	-7.096	-5.703	-4.804	-1.886	-3.010

(a: ΔE_{ele} b: ΔE_{vdW} c: ΔG_{PB} d: ΔG_{NP} e: ΔG_{bind})