

Supplementary Material

QSAR Studies on Hepatitis C Virus NS5A Protein Tetracyclic Inhibitors in Wild Type and Mutants by CoMFA and CoMSIA

Zijian Qin, Aixia Yan*

The Supplementary Material containing the following files:

1. Four dataset files, same information with different file formats (SDF and CSV):
 - a) “1_dataset196_aligned.sdf”: 3D molecular structures of all 196 inhibitors after molecules alignment;
 - b) “2a_dataset192_GT1a_traintest.sdf” and “2a_dataset192_GT1a_traintest.csv”: the name of the compounds, their molecular structures (3D molecular structures in SDF, SMILES in CSV), the experimental EC₉₀ values (with unit), the experimental pEC₉₀ values (with unit), and the training/test labels of three random splitting of dataset GT1a;
 - c) “2b_dataset180_GT1aY93H_traintest.sdf” and “2b_dataset180_GT1aY93H_traintest.csv”: the name of the compounds, their molecular structures (3D molecular structures in SDF, SMILES in CSV), the experimental EC₉₀ values (with unit), the experimental pEC₉₀ values (with unit), and the training/test labels of three random splitting of dataset GT1a Y93H;
 - d) “2c_dataset159_GT1aL31V_traintest.sdf” and “2c_dataset159_GT1aL31V_traintest.csv”: the name of the compounds, their molecular structures (3D molecular structures in SDF, SMILES in CSV), the experimental EC₉₀ values (with unit), the experimental pEC₉₀ values (with unit), and the training/test labels of three random splitting of dataset GT1a L31V.
2. Figure S1 and Table S1 (PDF)
 - a) Figure S1 (corresponding to the sections entitled “Datasets and Bioactivities” and “Datasets and Molecules Alignment Analyses” in the main text): the histogram of the number of inhibitors versus the pEC₉₀ distribution of the datasets GT1a, GT1a Y93H, and GT1a L31V.
 - b) Table S1 (corresponding to the section entitled “Application domain analyses” in the main

text): molecular weight (MolWt), Lop P (MolLogP), shape similarity (ShapeTanimoto), and chemical feature similarity (ColorTanimoto) of all 196 inhibitors in the dataset.

3. a models' performance file (SI_all_model_performances.xlsx)
 - a) There are 12 sheets in the xlsx file.
 - b) Three sheets "1a_ran01", "1a_ran02", and "1a_ran03" include the descriptor type (TYPE), the number of components (n_comp), the number of inhibitors (all_N, tr_N, and te_N), the models' performance of the leave-one-out training set (loo_r2, loo_q2, loo_rmse, loo_sd, and loo_se), the models' performance of the training set and the test set, and the Over-Training index (se: (loo-tr)/tr) of all models of the dataset GT1a;
 - c) Three sheets "1aY93H_ran01", "1aY93H_ran02", and "1aY93H_ran03" include same columns as the sheets "1a_ran01", "1a_ran02", and "1a_ran03" of all models of the dataset GT1a Y93H;
 - d) Three sheets "1aL31V_ran01", "1aL31V_ran02", and "1aL31V_ran03" include same columns as the sheets "1a_ran01", "1a_ran02", and "1a_ran03" of all models of the dataset GT1a L31V;
 - e) Three sheets "summary_1a", "summary_1aY93H", and "summary_1aL31V" summarize the average and the standard deviation of all models of three random splitting.

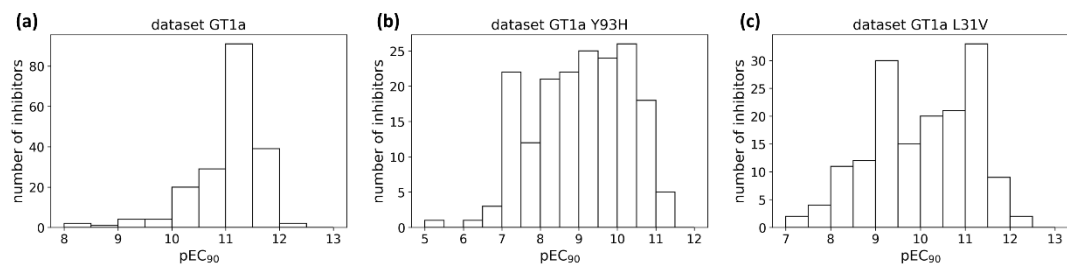


Figure S1. The histogram of the number of inhibitors versus the pEC_{90} distribution of the dataset GT1a (a), the dataset GT1a Y93H (b), and the dataset GT1a L31V (c).

Table S1. Molecular weight (MolWt), Lop P (MolLogP), shape similarity (ShapeTanimoto), and chemical feature similarity (ColorTanimoto) of all 196 inhibitors in the dataset.

Molecule Name	MolWt	MolLogP	ShapeTanimoto	ColorTanimoto
ns5a_indore_001_0	882.035	8.116	0.72	1
ns5a_indore_002_25	998.126	8.3507	0.688	0.757
ns5a_indore_003_17	998.126	8.3507	0.688	0.757
ns5a_indore_004_27	1013.141	7.8543	0.687	0.757
ns5a_indore_005_51	962.118	8.39562	0.685	0.81
ns5a_indore_006_111	976.145	8.6496	0.72	0.85
ns5a_indore_007_21	990.172	9.0397	0.716	0.85
ns5a_indore_008_1	990.172	9.2106	0.682	0.782
ns5a_indore_009_9	988.156	8.9646	0.719	0.882
ns5a_indore_010_7	998.098	9.0248	0.72	0.882
ns5a_indore_011_32	1016.088	9.106	0.723	0.882
ns5a_indore_012_1	998.151	9.2404	0.688	0.782
ns5a_indore_013_66	988.156	8.5759	0.724	0.801
ns5a_indore_014_45	998.126	8.3507	0.686	0.753
ns5a_indore_015_178	1013.141	7.8543	0.689	0.748
ns5a_indore_016_10	1002.183	9.3531	0.691	0.81
ns5a_indore_017_44	1002.183	9.3531	0.714	0.882
ns5a_indore_018_78	1016.21	9.7416	0.692	0.81
ns5a_indore_019_96	1016.21	9.7416	0.696	0.827
ns5a_indore_020_393	1044.264	10.5218	0.477	0.457
ns5a_indore_021_38	1016.21	9.7432	0.696	0.797
ns5a_indore_022_120	1016.21	9.7432	0.712	0.85
ns5a_indore_023_76	1014.194	9.4972	0.694	0.827
ns5a_indore_024_134	1014.194	9.4956	0.71	0.744
ns5a_indore_025_177	1044.264	10.5218	0.687	0.766
ns5a_indore_026_106	1026.18	9.1293	0.72	0.793
ns5a_indore_027_281	1040.207	9.5178	0.714	0.793
ns5a_indore_028_162	1004.199	9.4282	0.717	0.819
ns5a_indore_029_130	1018.226	9.8183	0.713	0.819
ns5a_indore_030_136	1018.226	9.9892	0.715	0.819
ns5a_indore_031_140	1016.21	9.3545	0.723	0.82
ns5a_indore_032_151	1005.187	8.8232	0.718	0.791
ns5a_indore_033_108	1019.214	9.2133	0.714	0.791
ns5a_indore_034_77	1019.214	9.3842	0.716	0.791
ns5a_indore_035_88	1017.198	9.1382	0.714	0.819
ns5a_indore_036_134	1017.198	8.7495	0.731	0.791
ns5a_indore_037_151	1004.199	9.4282	0.718	0.819
ns5a_indore_038_119	1016.21	9.7432	0.717	0.849
ns5a_indore_039_4	922.1	8.9934	0.71	1
ns5a_indore_040_10	926.088	8.5147	0.705	0.96
ns5a_indore_041_148	944.078	8.6538	0.71	0.96

ns5a_indore_042_50	956.089	8.5801	0.715	0.923
ns5a_indore_043_0	916.118	8.7399	0.708	0.958
ns5a_indore_044_0	934.108	8.879	0.713	0.958
ns5a_indore_045_0	928.129	9.0549	0.703	0.998
ns5a_indore_046_0	946.119	9.194	0.707	0.998
ns5a_indore_047_0	906.054	8.3166	0.709	0.997
ns5a_indore_048_12	946.119	9.194	0.71	0.997
ns5a_indore_049_10	946.119	9.194	0.705	0.998
ns5a_indore_050_0	946.119	9.194	0.709	0.996
ns5a_indore_051_1	948.135	9.44	0.711	0.958
ns5a_indore_052_1	948.135	9.2691	0.709	0.958
ns5a_indore_053_0	974.051	9.3354	0.709	0.998
ns5a_indore_054_0	956.061	9.2542	0.711	0.998
ns5a_indore_055_25	988.078	9.4214	0.702	0.998
ns5a_indore_056_1	960.146	9.2691	0.704	0.998
ns5a_indore_058_0	931.064	8.18828	0.712	0.958
ns5a_indore_059_5	977.133	8.0184	0.702	0.958
ns5a_indore_060_0	950.107	8.463	0.706	0.958
ns5a_indore_061_2	960.146	9.50242	0.711	0.998
ns5a_indore_062_0	964.109	9.3331	0.711	0.998
ns5a_indore_063_95	964.109	9.3331	0.719	0.998
ns5a_indore_064_26	974.104	9.4178	0.688	0.873
ns5a_indore_065_2	946.119	8.8053	0.719	0.959
ns5a_indore_066_56	960.146	9.1954	0.716	0.959
ns5a_indore_067_0	890.974	7.2431	0.708	0.955
ns5a_indore_068_160	931.039	8.1205	0.713	0.954
ns5a_indore_069_1	941.034	8.3963	0.721	0.918
ns5a_indore_070_23	933.055	8.3665	0.713	0.918
ns5a_indore_071_0	936.084	7.669	0.718	0.877
ns5a_indore_072_0	908.03	7.1066	0.71	0.91
ns5a_indore_073_29	947.107	8.589	0.716	0.956
ns5a_indore_074_1	947.107	8.589	0.714	0.956
ns5a_indore_075_0	947.107	8.589	0.71	0.958
ns5a_indore_076_0	907.042	7.7116	0.71	0.958
ns5a_indore_077_12	961.134	8.7631	0.708	0.958
ns5a_indore_078_0	935.096	8.274	0.714	0.922
ns5a_indore_079_0	949.123	8.6641	0.71	0.922
ns5a_indore_080_0	949.123	8.835	0.712	0.922
ns5a_indore_081_215	989.066	8.8164	0.704	0.958
ns5a_indore_082_0	957.049	8.6492	0.712	0.958
ns5a_indore_083_0	937.068	7.7202	0.717	0.921
ns5a_indore_084_0	951.095	7.858	0.707	0.922
ns5a_indore_085_21	963.15	9.14342	0.714	0.921
ns5a_indore_086_11	959.118	8.5874	0.71	0.893

ns5a_indore_087_8	961.134	8.8334	0.712	0.861
ns5a_indore_088_7	947.107	8.2724	0.714	0.861
ns5a_indore_089_13	961.134	8.6625	0.71	0.861
ns5a_indore_090_7	973.145	8.7615	0.707	0.893
ns5a_indore_091_11	958.13	9.1924	0.708	0.927
ns5a_indore_092_20	958.13	8.8037	0.719	0.893
ns5a_indore_093_11	950.085	9.4083	0.715	0.96
ns5a_indore_094_0	950.085	9.4083	0.727	0.96
ns5a_indore_095_7	951.073	8.8033	0.715	0.923
ns5a_indore_096_29	940.09	8.7438	0.72	0.96
ns5a_indore_097_109	954.117	9.1339	0.715	0.96
ns5a_indore_098_6	942.062	8.2853	0.72	0.923
ns5a_indore_099_2	944.034	7.9838	0.721	0.889
ns5a_indore_100_23	980.014	8.5766	0.711	0.889
ns5a_indore_101_46	956.089	8.5801	0.716	0.923
ns5a_indore_102_20	956.089	8.3278	0.715	0.923
ns5a_indore_104_25	968.056	8.3615	0.71	0.923
ns5a_indore_105_131	971.06	7.586	0.71	0.857
ns5a_indore_106_27	985.087	7.6103	0.707	0.889
ns5a_indore_107_16	970.116	8.9686	0.712	0.923
ns5a_indore_108_31	970.116	8.9686	0.711	0.923
ns5a_indore_109_38	984.143	9.3587	0.707	0.889
ns5a_indore_110_23	970.116	8.9686	0.711	0.923
ns5a_indore_111_22	970.116	8.9686	0.711	0.923
ns5a_indore_112_19	984.143	9.3587	0.705	0.889
ns5a_indore_113_118	970.116	8.9686	0.712	0.889
ns5a_indore_114_13	970.116	9.1411	0.713	0.923
ns5a_indore_115_25	970.116	9.1411	0.713	0.923
ns5a_indore_116_23	982.127	9.0692	0.707	0.923
ns5a_indore_117_75	992.069	9.1294	0.71	0.923
ns5a_indore_118_158	930.051	8.2637	0.716	0.96
ns5a_indore_120_335	990.103	7.5135	0.52	0.594
ns5a_indore_121_36	990.103	7.5135	0.727	0.888
ns5a_indore_122_203	990.103	7.5135	0.724	0.881
ns5a_indore_123_27	956.089	8.6522	0.719	0.841
ns5a_indore_124_64	966.031	8.1597	0.727	0.955
ns5a_indore_125_273	1062.063	8.6991	0.507	0.594
ns5a_indore_126_421	982.127	9.0439	0.729	0.932
ns5a_indore_127_75	954.073	8.2605	0.718	0.84
ns5a_indore_128_59	1054.087	9.5311	0.718	0.764
ns5a_indore_129_294	1038.235	10.6011	0.703	0.803
ns5a_indore_130_420	948.041	8.2117	0.722	0.958
ns5a_indore_131_342	948.041	8.2117	0.723	0.958
ns5a_indore_132_341	942.062	8.2621	0.716	0.894

ns5a_indore_133_297	942.062	8.2621	0.716	0.894
ns5a_indore_134_140	960.052	8.2101	0.691	0.824
ns5a_indore_135_151	992.17	8.6791	0.626	0.68
ns5a_indore_136_6	912.036	8.2535	0.724	0.928
ns5a_indore_137_339	924.072	8.123	0.712	0.894
ns5a_indore_138_15	918.015	8.2031	0.695	0.91
ns5a_indore_139_342	930.051	8.0726	0.718	0.958
ns5a_indore_141_59	968.1	8.5785	0.716	0.863
ns5a_indore_142_15	805.937	6.5266	0.643	0.93
ns5a_indore_143_0	900.025	8.2551	0.724	1
ns5a_indore_144_1	900.025	8.2551	0.724	1
ns5a_indore_145_0	900.025	8.2551	0.725	1
ns5a_indore_146_93	823.927	6.6657	0.648	0.929
ns5a_indore_147_106	823.927	6.6657	0.648	0.929
ns5a_indore_148_0	900.025	8.2551	0.724	1
ns5a_indore_149_5	840.382	7.18	0.573	0.728
ns5a_indore_150_0	934.47	8.9085	0.727	1
ns5a_indore_151_29	846.002	7.404	0.567	0.728
ns5a_indore_152_54	882.035	8.1936	0.628	0.893
ns5a_indore_153_0	846.002	7.404	0.641	0.93
ns5a_indore_154_21	882.035	8.1936	0.627	0.893
ns5a_indore_155_521	896.062	8.1174	0.603	0.797
ns5a_indore_156_13	918.015	8.2031	0.695	0.91
ns5a_indore_158_412	980.058	8.5498	0.689	0.904
ns5a_indore_159_230	994.085	8.9383	0.669	0.782
ns5a_indore_160_103	1008.112	9.33	0.715	0.919
ns5a_indore_161_89	1020.001	9.0497	0.658	0.811
ns5a_indore_162_197	1020.001	9.0497	0.664	0.811
ns5a_indore_164_563	962.068	8.6018	0.686	0.902
ns5a_indore_165_191	976.095	8.9903	0.692	0.853
ns5a_indore_166_457	1002.011	9.1017	0.688	0.886
ns5a_indore_167_379	984.021	8.8045	0.711	0.958
ns5a_indore_168_168	1002.011	9.1017	0.687	0.886
ns5a_indore_170_152	912.061	8.1246	0.711	0.96
ns5a_indore_173_34	958.105	9.0423	0.709	0.923
ns5a_indore_174_35	940.115	8.9032	0.704	0.923
ns5a_indore_175_2	930.051	8.2637	0.72	0.96
ns5a_indore_176_2	912.061	8.1246	0.715	0.96
ns5a_indore_177_12	960.077	8.2723	0.715	0.923
ns5a_indore_178_13	942.087	8.1332	0.71	0.923
ns5a_indore_179_86	964.496	8.9171	0.717	0.96
ns5a_indore_180_69	825.927	6.8438	0.643	0.816
ns5a_indore_181_27	853.981	7.624	0.671	0.78
ns5a_indore_182_0	910.089	8.8962	0.727	0.808

ns5a_indore_183_455	878.003	7.624	0.686	0.816
ns5a_indore_184_419	934.111	9.1844	0.689	0.728
ns5a_indore_185_8	962.165	9.9646	0.739	0.845
ns5a_indore_186_13	1034.125	10.455	0.72	0.845
ns5a_indore_187_141	966.109	7.6572	0.712	0.69
ns5a_indore_188_89	914.033	6.8738	0.718	0.753
ns5a_indore_189_181	914.033	6.877	0.683	0.754
ns5a_indore_190_99	994.163	8.4374	0.675	0.703
ns5a_indore_191_138	763.899	7.7128	0.593	0.643
ns5a_indore_192_427	938.055	7.1652	0.392	0.362
ns5a_indore_193_186	966.109	7.6572	0.669	0.574
ns5a_indore_194_592	890.105	9.9606	0.662	0.643
ns5a_indore_195_599	926.085	10.2388	0.668	0.643
ns5a_indore_196_436	946.213	11.521	0.51	0.438
ns5a_indore_197_117	942.181	11.029	0.584	0.458
ns5a_indore_198_0	924.072	7.8866	0.691	0.813
ns5a_indore_199_34	898.034	7.4949	0.718	0.868
ns5a_indore_200_33	958.08	9.2855	0.719	0.919
ns5a_indore_201_0	938.099	8.2767	0.696	0.836
ns5a_indore_202_0	910.045	7.6406	0.676	0.77
ns5a_indore_203_67	924.072	7.8866	0.716	0.831
ns5a_indore_204_44	898.034	7.4949	0.718	0.868
ns5a_indore_205_0	896.062	8.5061	0.723	0.897
mean^a	956.053	8.669	0.697	0.871
std^b	47.714	0.810	0.045	0.110
less than (mean-std)^c	26	31	18	24
greater than (mean+std)^d	31	21	0	20
coverage^e	0.709	0.735	0.908	0.776

^a the mean (average) value of all 196 inhibitors

^b the standard deviation of all 196 inhibitors

^c number of inhibitors less than (mean-std)

^d number of inhibitors greater than (mean+std)

^e the coverage of inhibitors in the range of [mean-std, mean+std]