

## Supplementary Materials

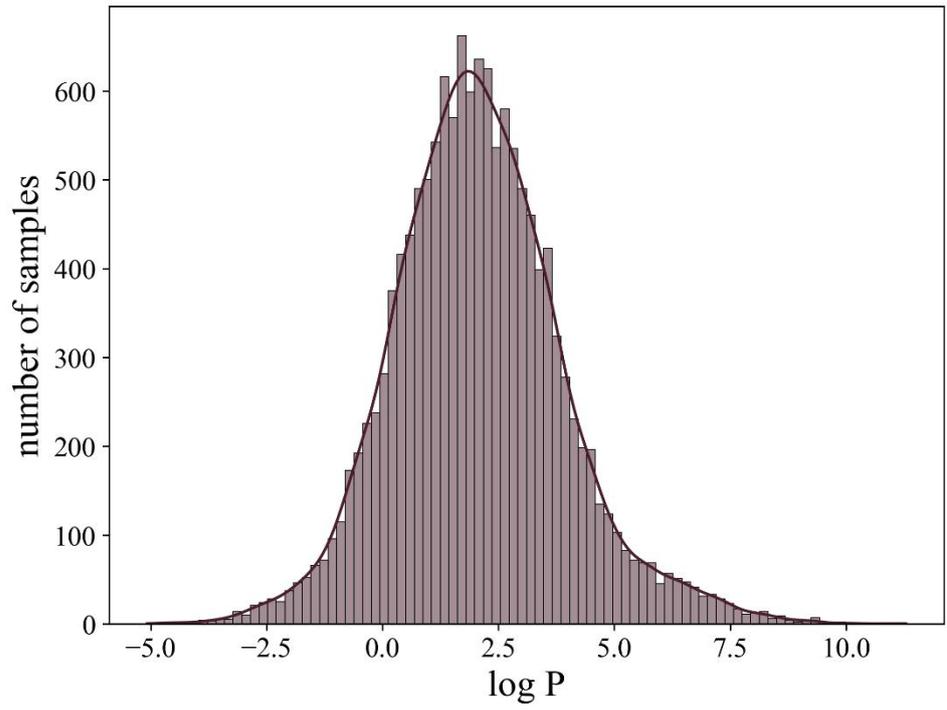
**A new simple and efficient molecular descriptor for the fast and accurate prediction of log P**

**Xiaojian Zeng<sup>#</sup>, Xin Ye<sup>#</sup>, Donghua Liu, Ningyi Cui, Xiaopeng Li, Yufan Bao, Yecheng Zhou<sup>\*</sup>**

Guangzhou Key Laboratory of Flexible Electronic Materials and Wearable Devices, School of Materials Science & Engineering, Sun Yat-sen University, Guangzhou 510006, Guangdong, China.

<sup>#</sup>Authors contributed equally.

**\*Correspondence to:** Prof. Yecheng Zhou, Guangzhou Key Laboratory of Flexible Electronic Materials and Wearable Devices, School of Materials Science & Engineering, Sun Yat-sen University, No.132 East Outer Ring Road, Guangzhou 510006, Guangdong, China. E-mail: zhouych29@mail.sysu.edu.cn



**Supplementary Figure 1.** Distribution of experimental log P values in the OPERA dataset.

**Supplementary Table 1.** The RMSEs of predictions for test sets of OPERA datasets with various algorithms and optimized 3D-MoRSE descriptors calculated by various  $s_L$ . The test size was 0.1. Feature selection was not considered.

RMSE	Linear regression	ridge	ARD Regression	Bayesian Ridge	Lasso	Elastic Net	Tweedie Regressor
$s_L=0.03$ , $N_s=100$	>10000	1.00	1.00	1.13	1.39	1.32	1.31
$s_L=0.05$ , $N_s=100$	>10000	0.93	0.92	0.94	1.39	1.3	1.28
$s_L=0.07$ , $N_s=100$	>10000	0.90	0.90	0.92	1.39	1.29	1.26
$s_L=0.1$ , $N_s=100$	>10000	0.85	0.95	0.99	1.39	1.29	1.26
$s_L=0.2$ , $N_s=100$	>10000	0.75	0.81	0.82	1.39	1.27	1.22
$s_L=0.5$ , $N_s=100$	0.77	0.73	0.71	0.7	1.4	1.28	1.2
$s_L=1.0$ , $N_s=100$	0.77	0.78	0.79	0.78	1.44	1.35	1.28
$s_L=1.5$ , $N_s=100$	0.78	0.81	0.79	0.79	1.48	1.38	1.29
$s_L=2$ , $N_s=100$	0.83	0.86	0.85	0.85	1.45	1.42	1.36
$s_L=2.5$ , $N_s=100$	0.85	0.86	0.85	0.86	1.48	1.47	1.37
$s_L=3.0$ , $N_s=100$	0.87	0.89	0.87	0.88	1.51	1.41	1.34

**Supplementary Table 2.** The RMSEs of predictions for test sets of OPERA datasets with various algorithms and optimized 3D-MoRSE descriptors calculated by various  $s_L$ . The test size was 0.1. The estimator of SelectFromModel module was ARD regression, and the threshold was 0.001.

RMSE	Linear regression	ridge	ARD Regression	Bayesian Ridge	Lasso	Elastic Net	Tweedie Regressor
$s_L=0.03,$ $N_s=100$	>10000	1.07	1.01	1.00	1.39	1.31	1.30
$s_L=0.05,$ $N_s=100$	>10000	0.93	0.93	0.93	1.39	1.29	1.27
$s_L=0.07,$ $N_s=100$	186.89	0.98	0.93	0.94	1.40	1.29	1.22
$s_L=0.1,$ $N_s=100$	1.03	0.91	0.91	0.90	1.40	1.31	1.20
$s_L=0.2,$ $N_s=100$	0.80	0.80	0.80	0.80	1.41	1.31	1.07
$s_L=0.5,$ $N_s=100$	0.71	0.71	0.71	0.70	1.51	1.35	0.97
$s_L=1.0,$ $N_s=100$	0.85	0.86	0.85	0.86	1.68	1.55	1.13
$s_L=1.5,$ $N_s=100$	0.89	0.89	0.89	0.89	1.77	1.70	1.25
$s_L=2,$ $N_s=100$	0.91	0.93	0.92	0.91	1.77	1.63	1.33
$s_L=2.5,$ $N_s=100$	0.94	0.96	0.95	0.95	1.83	1.73	1.42
$s_L=3.0,$ $N_s=100$	0.95	0.97	0.96	0.95	1.77	1.70	1.39

**Supplementary Table 3.** The RMSEs of models using descriptors selected with Threshold  $10^{-5}$  and dimensions of the descriptors. The descriptors were selected using 20 random states. The RMSEs shown in the table were the average of 20 predictions.

Random state	ARD Regression		Ridge		Bayesian Ridge	
	RMSE	Dimension	RMSE	Dimension	RMSE	Dimension
0	0.70	3515	0.69	3515	0.69	3515
1	0.70	3514	0.69	3514	0.69	3514
2	0.69	3498	0.69	3498	0.69	3498
3	0.70	3557	0.69	3557	0.70	3557
4	0.69	3507	0.67	3507	0.67	3507
5	0.70	3558	0.69	3558	0.69	3558
6	0.70	3537	0.69	3537	0.69	3537
7	0.69	3501	0.69	3501	0.69	3501
8	0.69	3541	0.68	3541	0.68	3541
9	0.69	3577	0.68	3577	0.68	3577
10	0.70	3534	0.69	3534	0.69	3534
11	0.69	3535	0.68	3535	0.68	3535
12	0.70	3539	0.68	3539	0.68	3539
13	0.70	3557	0.69	3557	0.69	3557
14	0.69	3557	0.68	3557	0.68	3557
15	0.70	3485	0.69	3485	0.69	3485
16	0.69	3533	0.68	3533	0.68	3533
17	0.70	3502	0.69	3502	0.69	3502
18	0.69	3504	0.69	3504	0.69	3504
19	0.69	3535	0.69	3535	0.69	3535

**Supplementary Table 4.** The RMSEs of models using descriptors selected with Threshold  $10^{-4}$  and dimensions of the descriptors. The descriptors were selected using 20 random states. The RMSEs shown in the table were the average of 20 predictions.

Random state	ARD Regression		Ridge		Bayesian Ridge	
	RMSE	Dimension	RMSE	Dimension	RMSE	Dimension
0	0.70	3310	0.69	3310	0.69	3310
1	0.69	3293	0.69	3293	0.69	3293
2	0.69	3334	0.69	3334	0.69	3334
3	0.71	3352	0.69	3352	0.70	3352
4	0.68	3276	0.67	3276	0.67	3276
5	0.70	3348	0.68	3348	0.69	3348
6	0.69	3314	0.69	3314	0.69	3314
7	0.69	3273	0.69	3273	0.68	3273
8	0.68	3339	0.67	3339	0.68	3339
9	0.68	3326	0.67	3326	0.68	3326
10	0.70	3317	0.69	3317	0.69	3317
11	0.69	3322	0.68	3322	0.68	3322
12	0.70	3340	0.68	3340	0.68	3340
13	0.70	3358	0.68	3358	0.69	3358
14	0.68	3340	0.67	3340	0.68	3340
15	0.70	3280	0.69	3280	0.69	3280
16	0.69	3334	0.68	3334	0.68	3334
17	0.70	3271	0.69	3271	0.69	3271
18	0.69	3304	0.69	3304	0.69	3304
19	0.69	3325	0.68	3325	0.68	3325

**Supplementary Table 5.** The RMSEs of models using descriptors selected with Threshold  $10^{-3}$  and dimensions of the descriptors. The descriptors were selected using 20 random states. The RMSEs shown in the table were the average of 20 predictions.

Random state	ARD Regression		Ridge		Bayesian Ridge	
	RMSE	Dimension	RMSE	Dimension	RMSE	Dimension
0	0.68	1724	0.67	1724	0.68	1724
1	0.67	1661	0.66	1661	0.67	1661
2	0.67	1732	0.66	1732	0.67	1732
3	0.68	1752	0.67	1752	0.68	1752
4	0.66	1738	0.65	1738	0.66	1738
5	0.67	1747	0.66	1747	0.67	1747
6	0.67	1789	0.66	1789	0.67	1789
7	0.67	1683	0.66	1683	0.67	1683
8	0.66	1769	0.66	1769	0.66	1769
9	0.66	1796	0.65	1796	0.66	1796
10	0.67	1745	0.67	1745	0.67	1745
11	0.66	1697	0.65	1697	0.66	1697
12	0.67	1746	0.66	1746	0.67	1746
13	0.67	1793	0.67	1793	0.67	1793
14	0.66	1765	0.65	1765	0.66	1765
15	0.68	1693	0.66	1693	0.67	1693
16	0.67	1773	0.65	1773	0.66	1773
17	0.68	1752	0.67	1752	0.68	1752
18	0.67	1723	0.67	1723	0.67	1723
19	0.67	1778	0.66	1778	0.67	1778

**Supplementary Table 6.** The RMSEs of models using descriptors selected with Threshold  $10^{-2}$  and dimensions of the descriptors. The descriptors were selected using 20 random states. The RMSEs shown in the table were the average of 20 predictions.

Random state	ARD Regression		Ridge		Bayesian Ridge	
	RMSE	Dimension	RMSE	Dimension	RMSE	Dimension
0	0.66	965	0.66	965	0.66	965
1	0.65	939	0.65	939	0.64	939
2	0.65	989	0.65	989	0.64	989
3	0.66	999	0.66	999	0.65	999
4	0.64	954	0.63	954	0.63	954
5	0.65	986	0.65	986	0.64	986
6	0.65	977	0.65	977	0.65	977
7	0.65	950	0.65	950	0.65	950
8	0.64	989	0.65	989	0.64	989
9	0.65	995	0.64	995	0.63	995
10	0.66	971	0.65	971	0.65	971
11	0.65	997	0.64	997	0.64	997
12	0.65	987	0.65	987	0.64	987
13	0.65	992	0.65	992	0.65	992
14	0.65	974	0.64	974	0.64	974
15	0.66	964	0.65	964	0.65	964
16	0.65	982	0.64	982	0.64	982
17	0.66	969	0.65	969	0.65	969
18	0.66	970	0.65	970	0.65	970
19	0.65	986	0.65	986	0.65	986

**Supplementary Table 7.** The RMSEs of models using descriptors selected with Threshold  $5 \times 10^{-2}$  and dimensions of the descriptors. The descriptors were selected using 20 random states. The RMSEs shown in the table were the average of 20 predictions.

Random state	ARD Regression		Ridge		Bayesian Ridge	
	RMSE	Dimension	RMSE	Dimension	RMSE	Dimension
0	0.69	838	0.69	838	0.69	838
1	0.68	797	0.68	797	0.67	797
2	0.68	847	0.68	847	0.67	847
3	0.68	847	0.68	847	0.67	847
4	0.69	802	0.69	802	0.69	802
5	0.68	827	0.67	827	0.67	827
6	0.68	840	0.68	840	0.67	840
7	0.67	814	0.68	814	0.67	814
8	0.68	843	0.69	843	0.68	843
9	0.69	837	0.68	837	0.68	837
10	0.69	821	0.69	821	0.68	821
11	0.68	855	0.68	855	0.67	855
12	0.67	827	0.67	827	0.66	827
13	0.68	829	0.68	829	0.67	829
14	0.69	823	0.69	823	0.69	823
15	0.68	820	0.69	820	0.68	820
16	0.68	846	0.68	846	0.68	846
17	0.68	831	0.68	831	0.67	831
18	0.68	800	0.69	800	0.68	800
19	0.70	830	0.70	830	0.69	830

**Supplementary Table 8.** The RMSEs of models using descriptors selected with Threshold  $10^{-1}$  and dimensions of the descriptors. The descriptors were selected using 20 random states. The RMSEs shown in the table were the average of 20 predictions.

Random state	ARD Regression		Ridge		Bayesian Ridge	
	RMSE	Dimension	RMSE	Dimension	RMSE	Dimension
0	0.72	698	0.73	698	0.72	698
1	0.73	673	0.74	673	0.73	673
2	0.72	700	0.73	700	0.72	700
3	0.73	705	0.73	705	0.72	705
4	0.73	680	0.73	680	0.73	680
5	0.71	698	0.72	698	0.71	698
6	0.72	695	0.73	695	0.72	695
7	0.72	686	0.73	686	0.72	686
8	0.72	709	0.73	709	0.72	709
9	0.73	704	0.73	704	0.72	704
10	0.72	698	0.73	698	0.72	698
11	0.73	703	0.73	703	0.72	703
12	0.72	701	0.72	701	0.71	701
13	0.72	684	0.73	684	0.72	684
14	0.73	673	0.74	673	0.73	673
15	0.73	683	0.74	683	0.73	683
16	0.73	687	0.73	687	0.73	687
17	0.72	697	0.72	697	0.71	697
18	0.72	664	0.73	664	0.72	664
19	0.74	690	0.74	690	0.73	690

**Supplementary Table 9.** The MAEs, RMSEs, and  $R^2$  of prediction on test sets of OPERA dataset with various random states for splitting training and test sets. The test size was 0.25.

Random state	ARD Regression			Ridge			Bayesian Ridge		
	MAE	RMSE	$R^2$	MAE	RMSE	$R^2$	MAE	RMSE	$R^2$
0	0.49	0.68	0.87	0.50	0.69	0.86	0.49	0.68	0.87
1	0.50	0.68	0.85	0.50	0.68	0.86	0.49	0.67	0.86
2	0.49	0.67	0.86	0.49	0.67	0.86	0.48	0.66	0.86
3	0.49	0.68	0.87	0.50	0.68	0.86	0.49	0.67	0.87
4	0.49	0.66	0.87	0.49	0.66	0.87	0.49	0.66	0.87
5	0.50	0.69	0.86	0.51	0.69	0.86	0.50	0.68	0.87
6	0.49	0.68	0.87	0.50	0.68	0.87	0.49	0.68	0.87
7	0.50	0.68	0.86	0.50	0.69	0.86	0.49	0.68	0.86
8	0.49	0.68	0.86	0.49	0.68	0.86	0.49	0.68	0.86
9	0.49	0.68	0.86	0.49	0.68	0.86	0.49	0.68	0.86
10	0.51	0.71	0.86	0.51	0.70	0.86	0.50	0.69	0.86
11	0.50	0.67	0.86	0.50	0.67	0.87	0.50	0.66	0.87
12	0.52	0.73	0.84	0.52	0.72	0.85	0.51	0.72	0.85
13	0.50	0.68	0.87	0.51	0.69	0.86	0.50	0.68	0.87
14	0.49	0.67	0.87	0.49	0.67	0.87	0.49	0.66	0.87
15	0.49	0.66	0.87	0.49	0.66	0.87	0.49	0.66	0.87
16	0.49	0.68	0.86	0.49	0.68	0.86	0.49	0.67	0.87
17	0.50	0.68	0.86	0.50	0.68	0.86	0.50	0.67	0.86
18	0.49	0.67	0.87	0.50	0.68	0.86	0.49	0.67	0.87
19	0.50	0.68	0.86	0.51	0.69	0.86	0.51	0.69	0.86
average	0.50	0.68	0.86	0.50	0.68	0.86	0.49	0.68	0.86

**Supplementary Table 10.** Comparison of the performance of our models and other approaches on the SAMPL6 dataset.

RMSE (best, worst)	MAE (best, worst)	R2 (best, worst)	Method	Category	ref
0.309 (0.29,0.34)	0.27 (0.25,0.30)	0.78 (0.82,0.75)	opt3DM(ARD +ARD)	Simple Machine Learning	This work
0.31( $\pm$ 0.06)			DNNmono	Deep Learning	1
0.326 (0.31,0.36)	0.28 (0.26,0.30)	0.76 (0.78,0.71)	opt3DM(ARD +Bayesian Ridge)	Simple Machine Learning	This work
0.329 (0.32,0.37)	0.28 (0.25,0.30)	0.76 (0.77,0.70)	opt3DM(ARD +Ridge)	Simple Machine Learning	This work
0.33 ( $\pm$ 0.05)			DNNtaut	Deep Learning	1
0.38 (0.23,0.55)	0.31 (0.19,0.46)	0.77 (0.34,0.94)	COSMO-RS	Physical (QM)	2
0.39			Jchem		3
0.47 (0.17,0.75)	0.31 (0.15,0.55)	0.74 (0.33,0.97)	EC-RISM	Physical (QM)	4
0.49 (0.34,0.62)	0.43 (0.29,0.58)	0.54 (0.14,0.87)	SM12-Solvatio n-Trained	QM/MM	5
0.49 (0.37,0.61)	0.44 (0.32,0.57)	0.74 (0.40,0.94)	RayLogP-II	Simple Machine Learning	6
0.49 (0.33,0.62)	0.42 (0.26,0.57)	0.69 (0.37,0.91)	LogP_SMD_S olvation_DFT	Physical	7
0.53 (0.34,0.68)	0.43 (0.25,0.62)	0.62 (0.13,0.90)	MDFP-ML	QM/MM	8
0.62 (0.41,0.80)	0.51 (0.31,0.72)	0.66 (0.17,0.90)	NHLBI-NN-5 HL	Deep Learning	9
0.74 (0.55,0.93)	0.67 (0.48,0.87)	0.62 (0.16,0.88)	Molecular-Dyna mics-Expanded- Ensembles	Physical (MM)	10
0.82 (0.55,1.06)	0.67 (0.39,0.96)	0.33 (0.01,0.80)	Alchemical-C GenFF	Physical (MM)	11

**Supplementary Table 11.** The MAEs, RMSEs, and  $R^2$ s of prediction for the SAMPL6 challenge. Models were trained with various random states for splitting training and test sets on the OPERA dataset. The test size was 0.1.

Random state	ARD Regression			Ridge			Bayesian Ridge		
	MAE	RMSE	$R^2$	MAE	RMSE	$R^2$	MAE	RMSE	$R^2$
0	0.26	0.29	0.81	0.28	0.32	0.77	0.27	0.31	0.78
1	0.27	0.31	0.78	0.28	0.33	0.76	0.28	0.33	0.76
2	0.28	0.32	0.77	0.27	0.32	0.77	0.27	0.32	0.77
3	0.27	0.31	0.78	0.27	0.33	0.75	0.27	0.33	0.76
4	0.28	0.32	0.77	0.28	0.33	0.75	0.28	0.33	0.75
5	0.25	0.29	0.80	0.27	0.33	0.76	0.27	0.32	0.77
6	0.27	0.31	0.79	0.28	0.34	0.74	0.28	0.33	0.75
7	0.26	0.30	0.79	0.27	0.32	0.77	0.27	0.32	0.77
8	0.26	0.31	0.79	0.29	0.34	0.74	0.29	0.34	0.75
9	0.28	0.32	0.78	0.28	0.32	0.76	0.28	0.32	0.76
10	0.27	0.31	0.79	0.27	0.32	0.77	0.26	0.32	0.77
11	0.26	0.31	0.79	0.27	0.33	0.76	0.27	0.32	0.77
12	0.24	0.29	0.82	0.25	0.32	0.77	0.26	0.31	0.78
13	0.28	0.31	0.79	0.28	0.33	0.75	0.29	0.33	0.75
14	0.26	0.30	0.80	0.27	0.33	0.76	0.28	0.33	0.76
15	0.27	0.32	0.77	0.28	0.33	0.75	0.28	0.33	0.76
16	0.30	0.34	0.75	0.31	0.37	0.70	0.30	0.36	0.71
17	0.29	0.33	0.75	0.28	0.33	0.75	0.28	0.33	0.76
18	0.28	0.31	0.79	0.28	0.33	0.76	0.28	0.32	0.77
19	0.26	0.30	0.80	0.26	0.32	0.77	0.26	0.32	0.77
average	0.27	0.31	0.78	0.28	0.33	0.76	0.28	0.33	0.76

**Supplementary Table 12.** Comparison of the performance of our models and other approaches on the SAMPL9 dataset.

RMSE (best, worst)	MAE (best, worst)	R2 (best, worst)	Method	Category	ref
0.84	0.63	0.86	RFR	Simple Machine Learning	<sup>12</sup>
0.86	0.71	0.85	MLR	Simple Machine Learning	<sup>12</sup>
1.01 (0.97,1.04)	0.79 (0.75,0.82)	0.79 (0.81,0.78)	opt3DM (ARD+ARD)	Simple Machine Learning	This work
1.02		0.79	Multi-target D-MPNN	Deep Learning	<sup>13</sup>
1.06 (1.02,1.08)	0.84 (0.80,0.87)	0.78 (0.79,0.77)	opt3DM (ARD+ Bayesian Ridge)	Simple Machine Learning	This work
1.07 (1.03,1.09)	0.87 (0.83,0.89)	0.77 (0.79,0.76)	opt3DM (ARD+Ridge)	Simple Machine Learning	This work
1.07		0.77	Solvation GNN	Deep Learning	<sup>13</sup>
1.23		0.93	COSMO-RS	Physical (QM)	<sup>14</sup>
1.52	1.28		MM.PBSA	Physical (MM)	<sup>15</sup>
1.56		0.51	GDI-GNN	Deep Learning	<sup>13</sup>
1.64	1.35	0.75	IEF-PCM/MST	Physical (QM)	<sup>12</sup>
1.69		0.43	GH-GNN	Deep Learning	<sup>13</sup>
1.96	1.56		DLPNO.CCSD.T. def2.SVP	Physical (QM)	<sup>15</sup>
2.00	1.55		Non equilibrium fast growth	Physical (MM)	<sup>15</sup>
2.26	1.75		EE.Openff.2.0.TI P3P.MD.EE.WL	Physical (MM)	<sup>15</sup>
2.61	2.06		MD(GAFF/TIP3P )	Physical (MM)	<sup>15</sup>

**Supplementary Table 13.** The MAEs, RMSEs, and  $R^2$  of prediction for the SAMPL9 challenge. Models were trained with various random states for splitting training and test sets on the OPERA dataset. The test size was 0.1.

Random state	ARD Regression			Ridge(alpha=17)			Bayesian Ridge		
	MAE	RMSE	$R^2$	MAE	RMSE	$R^2$	MAE	RMSE	$R^2$
0	0.77	1.01	0.79	0.87	1.08	0.77	0.84	1.06	0.77
1	0.75	0.97	0.81	0.83	1.04	0.79	0.80	1.02	0.79
2	0.82	1.01	0.80	0.88	1.07	0.77	0.85	1.06	0.78
3	0.81	1.04	0.78	0.89	1.09	0.76	0.87	1.08	0.77
4	0.79	1.01	0.80	0.86	1.07	0.77	0.84	1.06	0.78
5	0.80	1.03	0.79	0.87	1.09	0.76	0.84	1.08	0.77
6	0.78	1.01	0.80	0.87	1.08	0.77	0.84	1.07	0.77
7	0.78	0.99	0.81	0.85	1.05	0.78	0.83	1.04	0.78
8	0.79	1.01	0.80	0.88	1.08	0.77	0.85	1.06	0.77
9	0.80	1.02	0.79	0.87	1.07	0.77	0.85	1.06	0.78
10	0.80	1.02	0.79	0.87	1.07	0.77	0.84	1.06	0.78
11	0.82	1.04	0.79	0.86	1.07	0.77	0.84	1.06	0.78
12	0.79	1.00	0.80	0.86	1.07	0.77	0.82	1.05	0.78
13	0.79	1.02	0.79	0.87	1.09	0.76	0.84	1.08	0.77
14	0.82	1.04	0.79	0.89	1.09	0.76	0.87	1.08	0.77
15	0.78	1.02	0.79	0.87	1.07	0.77	0.84	1.06	0.77
16	0.79	1.01	0.80	0.85	1.07	0.77	0.83	1.06	0.78
17	0.79	1.01	0.80	0.88	1.07	0.77	0.85	1.06	0.78
18	0.79	1.02	0.79	0.84	1.07	0.77	0.82	1.06	0.78
19	0.77	1.00	0.80	0.86	1.06	0.77	0.84	1.05	0.78
average	0.79	1.01	0.79	0.87	1.07	0.77	0.84	1.06	0.78

## References

1. Ulrich, N.; Goss, K.-U.; Ebert, A., Exploring the octanol–water partition coefficient dataset using deep learning techniques and data augmentation. *Communications Chemistry* **2021**, *4* (1), 90.
2. Loschen, C.; Reinisch, J.; Klamt, A., COSMO-RS based predictions for the SAMPL6 logP challenge. *Journal of Computer-Aided Molecular Design* **2020**, *34* (4), 385-392.
3. ChemAxon JChem for Excel v. 20.6.0.618.
4. Tielker, N.; Tomazic, D.; Eberlein, L.; Güssregen, S.; Kast, S. M., The SAMPL6 challenge on predicting octanol–water partition coefficients from EC-RISM theory. *Journal of Computer-Aided Molecular Design* **2020**, *34* (4), 453-461.
5. Ouimet, J. A.; Paluch, A. S., Predicting octanol/water partition coefficients for the SAMPL6 challenge using the SM12, SM8, and SMD solvation models. *Journal of Computer-Aided Molecular Design* **2020**, *34* (5), 575-588.
6. Lui, R.; Guan, D.; Matthews, S., A comparison of molecular representations for lipophilicity quantitative structure–property relationships with results from the SAMPL6 logP Prediction Challenge. *Journal of Computer-Aided Molecular Design* **2020**, *34* (5), 523-534.
7. Guan, D.; Lui, R.; Matthews, S., LogP prediction performance with the SMD solvation model and the M06 density functional family for SAMPL6 blind prediction challenge molecules. *Journal of Computer-Aided Molecular Design* **2020**, *34* (5), 511-522.
8. Wang, S.; Riniker, S., Use of molecular dynamics fingerprints (MDFPs) in SAMPL6 octanol–water log P blind challenge. *Journal of Computer-Aided Molecular Design* **2020**, *34* (4), 393-403.
9. Prasad, S.; Brooks, B. R., A deep learning approach for the blind logP prediction in SAMPL6 challenge. *Journal of Computer-Aided Molecular Design* **2020**, *34* (5), 535-542.
10. Nikitin, A., Non-zero Lennard-Jones parameters for the Toukan–Rahman water model: more accurate calculations of the solvation free energy of organic substances. *Journal of Computer-Aided Molecular Design* **2020**, *34* (4), 437-441.
11. Procacci, P.; Guarnieri, G., SAMPL6 blind predictions of water-octanol partition coefficients using nonequilibrium alchemical approaches. *Journal of Computer-Aided Molecular Design* **2020**, *34* (4), 371-384.
12. Zamora, W. J.; Viayna, A.; Pinheiro, S.; Curutchet, C.; Bisbal, L.; Ruiz, R.; Ràfols, C.; Luque, F. J., Prediction of toluene/water partition coefficients in the SAMPL9 blind challenge: assessment of machine learning and IEF-PCM/MST continuum solvation models. *Physical Chemistry Chemical Physics* **2023**, *25* (27), 17952-17965.
13. Nevolianis, T.; Rittig, J. G.; Mitsos, A.; Leonhard, K., Multi-fidelity graph neural networks for predicting toluene/water partition coefficients. **2024**.
14. Nevolianis, T.; Ahmed, R. A.; Hellweg, A.; Diedenhofen, M.; Leonhard, K., Blind prediction of toluene/water partition coefficients using COSMO-RS: results from the SAMPL9 challenge. *Physical Chemistry Chemical Physics* **2023**, *25* (46), 31683-31691.
15. Martin Amezcua; L, D.; Mobley; Bergazin, T. D. samplchallenges/SAMPL9: 0.8.