Electronic Supplementary Information (ESI)

Using Natural Language Processing (NLP)-Inspired Molecular Embedding Approach to Predict Hansen Solubility Parameters

Jiayun Pang,*[a] Alexander Pine [a] and Abdulai Sulemana [a]

a. *School of Science, Faculty of Engineering and Science, University of Greenwich Medway Campus, Central Avenue, Chatham Maritime, ME4 3RL, United Kingdom*
Email: j.pang@gre.ac.uk

Table of Contents

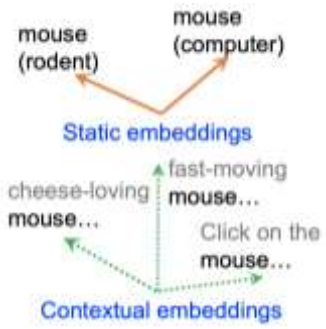


Figure S1. Illustration of static and contextual word embeddings using the word “mouse”. Static embeddings have one representation per word sense, while contextual embeddings are more context specific (Figure adapted from https://ai.stanford.edu/blog/contextual/).

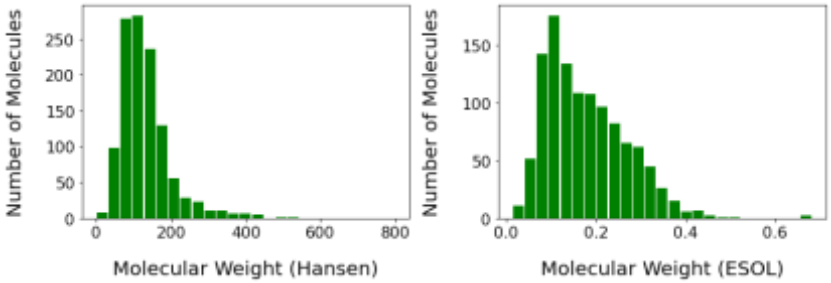


Figure S2: Molecular weight distribution of the Hansen dataset and ESOL dataset.

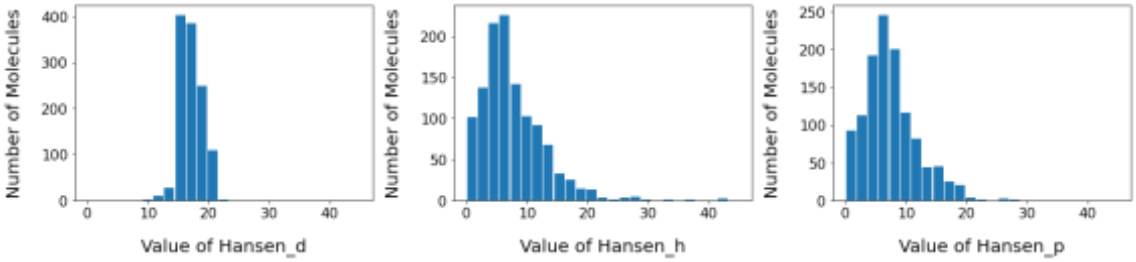


Figure S3. Distribution of the values of δd, δh and δp in the Hansen dataset.

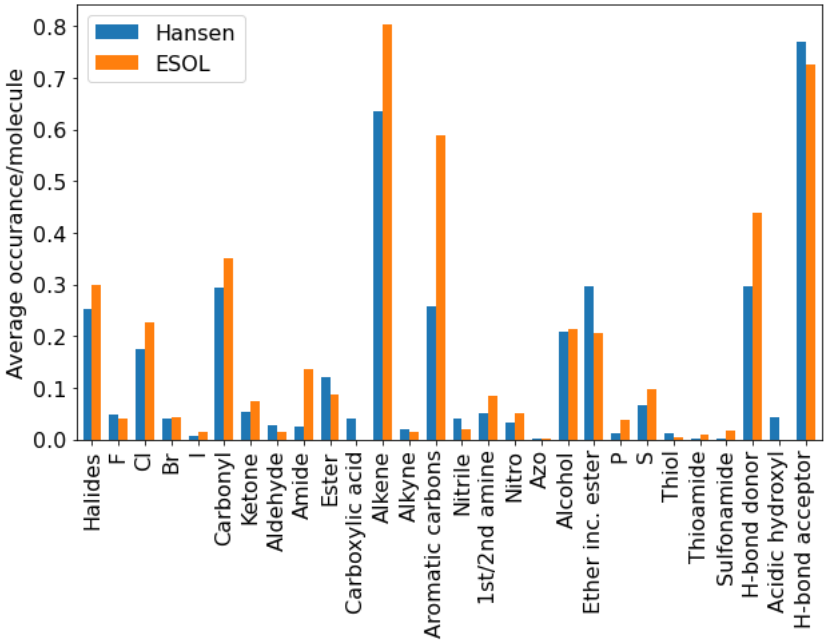


Figure S4. Functional group analysis of the Hansen and ESOL datasets.

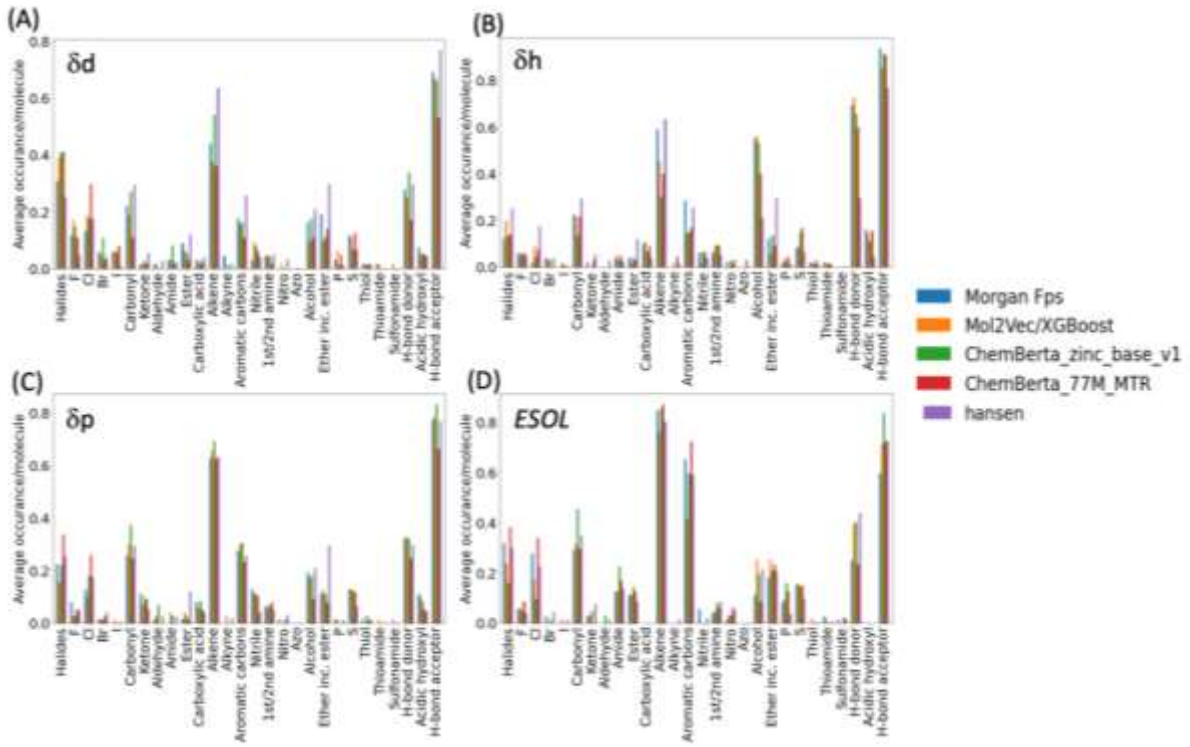


Figure S5. Functional group analysis of the outliers for (A) δd, (B) δh, (C) δp and (D) ESOL for four models and the full Hansen dataset. Functional groups were counted by matching their SMARTS codes to the SMILES strings using pybel/OpenBabel. The total number of each functional group was then divided by the total number of molecules in the dataset to derive the average occurrence per molecule.

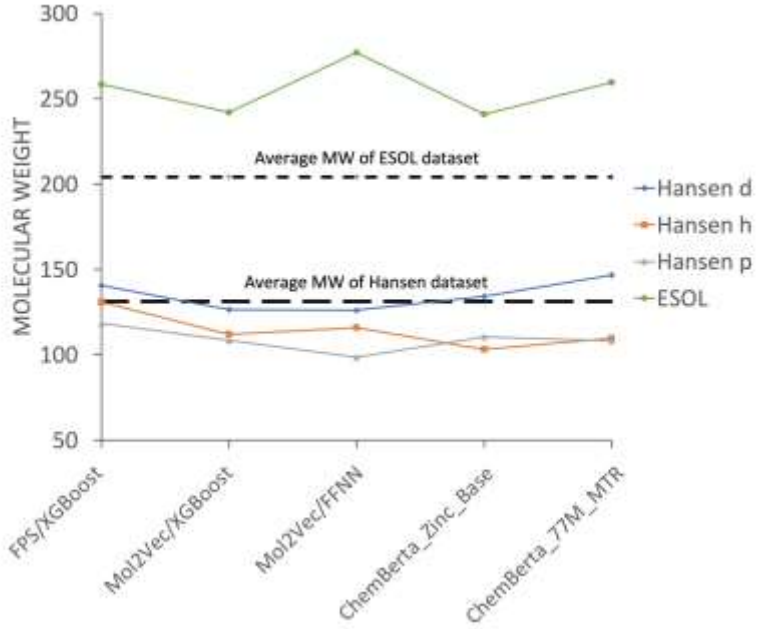


Figure S6. The average molecular weight of the outliers.

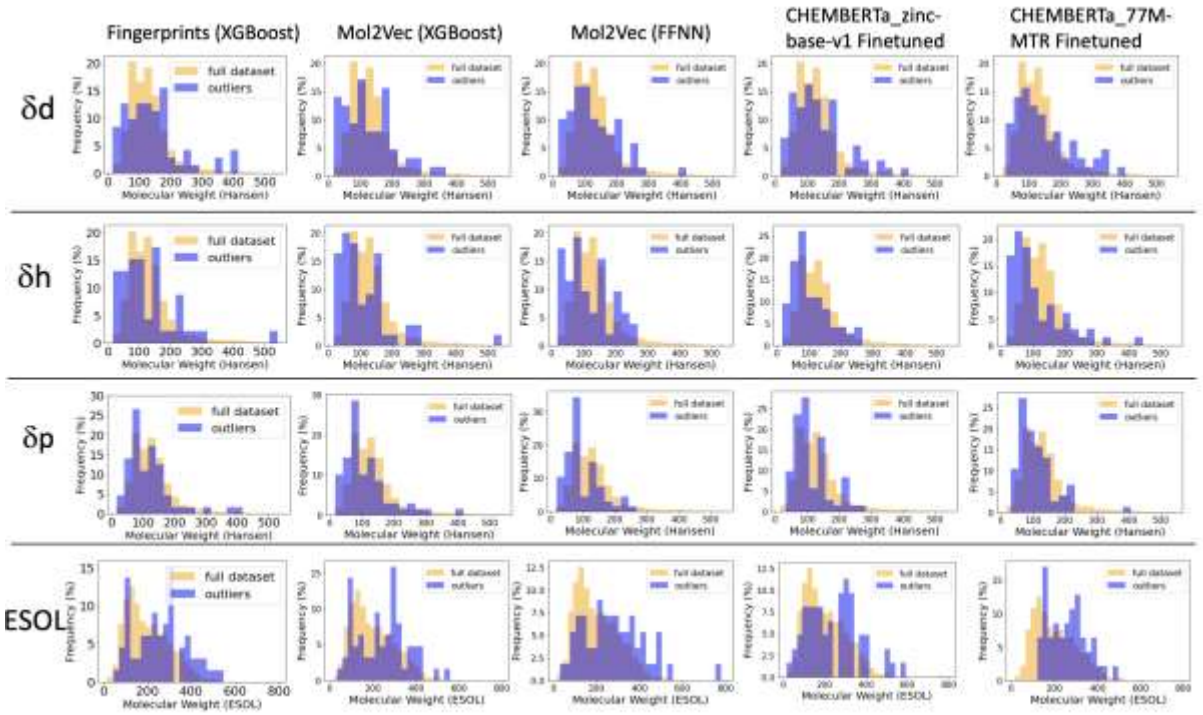
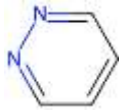
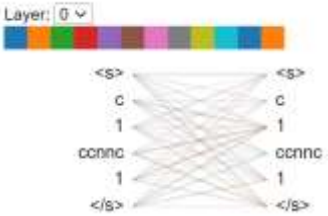
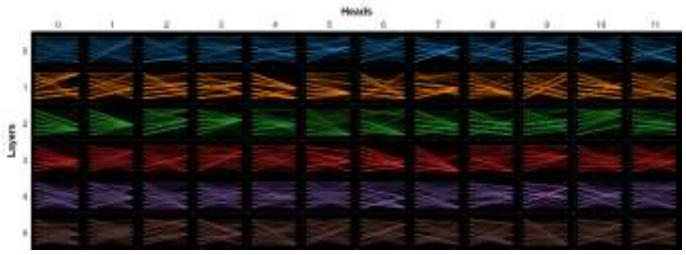
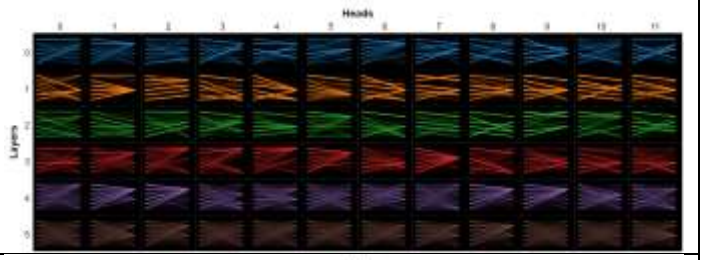
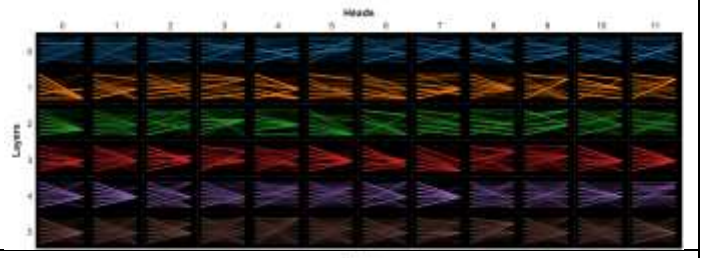
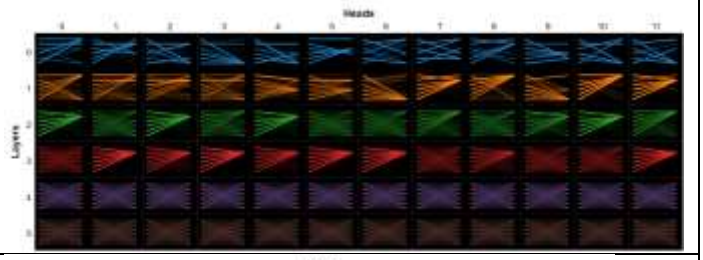
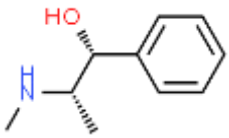
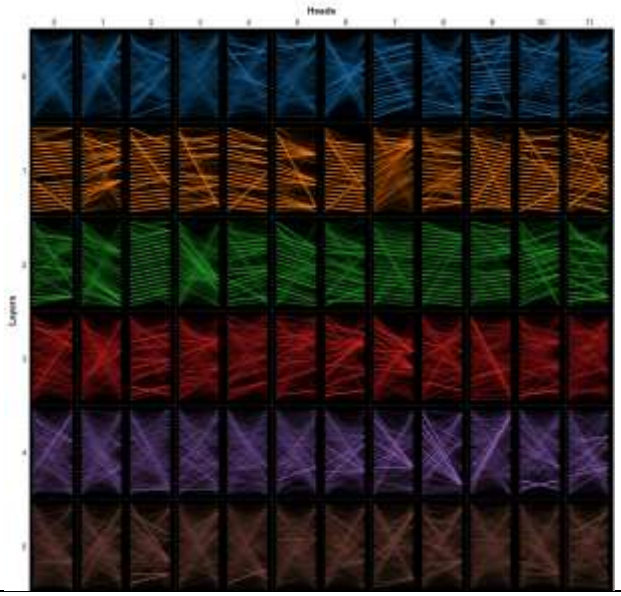


Figure S7: The molecular weight distribution of the outliers in comparison to the full dataset.

Table S1: Visualisation of all attentions for Pyridazine and L-(-)-Ephedrine.

Molecule	Model	Attention
Pyridazine  c1ccnnc1 	Original model	
	δd	
	δh	
	δp	
L-(-)-Ephedrine  CN[C@@H](C)[C@H](O)c1ccccc1	Original model	

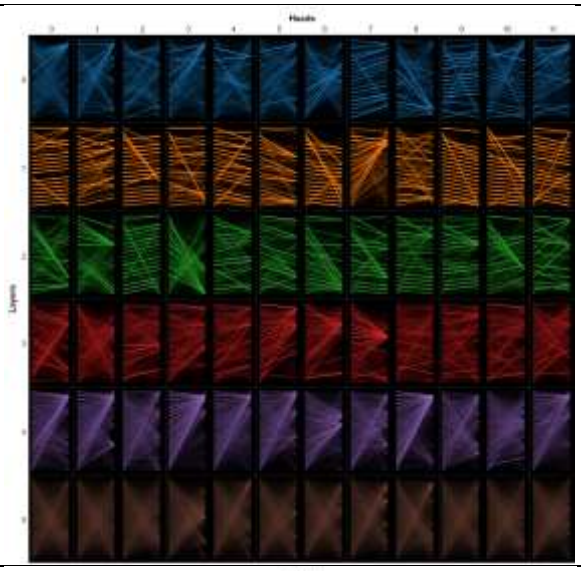
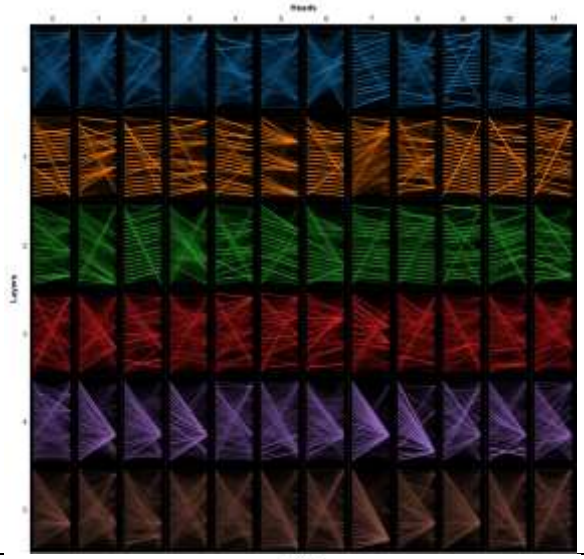
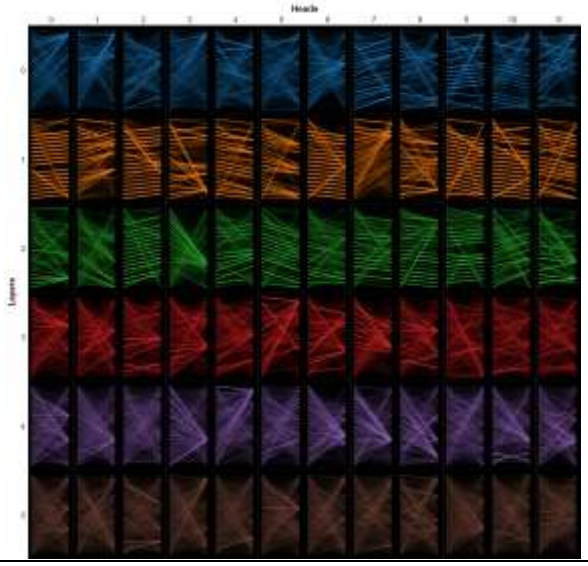
<table>
<tr>
<td rowspan="3"> 
</td>
<td>δd</td>
<td>  </td>
</tr>
<tr>
<td>δh</td>
<td>  </td>
</tr>
<tr>
<td>δp</td>
<td>  </td>
</tr>
</table>

Table S2. The predicted HSP values that are two standardised residual deviation (SRD) away from the experimental values are selected as the outliers and are presented here.

	Morgan fingerprint/XGBOOST		
	SMILES (60 molecules)	**δd (exp)**	**δd (pred)**
1	C=CCN=C=S	17	19.12078
2	FC(F)(F)Br	14.3	17.88144
3	CCOC(=O)CCCC(=O)OCC	16.3	12.389435
4	CF	13.4	15.394183
5	C=O	12.8	15.971532
6	CCCCCCCCCCCCOC(=O)C(C)=C	14.4	16.672253
7	CN(C)CCOC(=O)COc1ccc(Cl)cc1	16	19.62697
8	C[P](F)(F)=O	14	16.214167
9	C#CC#N	15.5	13.476004
10	BrC(Br)C(Br)Br	21	18.23203
11	CC(Br)Br	18.5	15.310307
12	N	13.7	16.582228
13	C1CO1	15.6	17.543549
14	NN	14.2	16.532194
15	IC(=C)C=C	19.9	16.946806
16	IC(I)I	20.2	17.987347
17	CCCCCCCCCC(C)(O)Oc1ccccc1	16.7	18.81298
18	[O-][O+]=O	19.8	16.88974
19	ClCC(Cl)(Cl)Cl	18	15.49792
20	FC(F)(Cl)C(F)(Cl)Cl	14.7	12.033137
21	OC(=O)C(Cl)=C	19.1	16.077759
22	CC(C)S	16.3	13.730936
23	OCCCCl	17.5	15.38078
24	C=CCCC=O	15.5	17.99248
25	NC(=O)NC(N)=O	20	16.65098
26	FC(F)Cl	12.3	15.7675705
27	C#N	12.3	16.055342
28	CC1=CC(=O)CC(C)(C)C1	17	19.63967
29	CC=C	13.3	15.23298
30	s1ccnc1	20.5	17.588928
31	FC(F)(Cl)C(F)(F)Cl	12.6	14.901509
32	Cn1ccnc1	19.7	17.37856
33	CC1=CCC2CC1C2(C)C	16.9	18.82562
34	C#C.C#C	14.4	16.811045
35	CCCCCCCCCCCCCCCCCC(=O)OCCCC	14.5	17.18572
36	CCCCCCOCCCCCC	16	18.055115
37	OCCO.O[S](O)=O	20	16.513426
38	CSc1ccccc1	19.6	17.317583

39	CC1CCC[S]1(=O)=O	19	16.766928
40	O[N+]([O-])=O	13.5	16.808695
41	C[Si](C)(C)O[Si](C)(C)O[Si](C)(C)C	12.2	14.418021
42	FC(F)(F)C1(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C1(F)F	12.4	14.429077
43	NCCc1c[nH]c2ccc(O)cc12	18	20.42724
44	CCCCCC(CC)(CC)OC(=O)c1ccccc1C([O-])=O	16.6	18.92403
45	C=C1CC(=O)O1	16.2	18.128946
46	C	14	16.592916
47	C[N+]1([O-])CCOCC1	19	16.707497
48	C[Si](C)(C)O[Si](C)(C)C	12.6	16.35119
49	C#Cc1ccccc1	18.8	16.36547
50	O[P](O)(O)=O	14.7	17.428555
51	CCC	13.1	15.44409
52	CCO[Si](OCC)(OCC)OCC	13.9	16.911234
53	NC(N)=S	20	18.001728
54	Cl[SiH](Cl)Cl	14.2	17.201168
55	NC(=O)c1ccccc1	21.2	19.24533
56	[Br]	18.2	16.266197
57	CC(CCl)OC(C)CCl	19	16.155499
58	C[S](C)(=O)=O	19	16.660654
59	C[S](C)=O	18.4	16.312098
60	O=C1CCCCCN1	19.4	17.227629
61	OC=O	14.6	16.66062
62	OC(F)(C(F)F)C(F)(F)F	17.2	11.573037
63	Cc1c(C)c(C)c(C)c(C)c1C	19.2	16.717941
64	IC(=C)C=C	17.2	19.152853
65	CC1=CN([C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O)C(=O)NC1=O	17	20.89644
66	ICl	22	17.053907
67	CCCCCCCCCc1ccccc1O	16.5	18.689426
68	CC1COC(=O)O1	20	17.753138

	Morgan Fingerprint/XGBOOST		
No.	**SMILES (60 molecules)**	**δd (exp)**	**δd (pred)**
1	C=CCN=C=S	17	19.1
2	FC(F)(F)Br	14.3	17.9
3	CCOC(=O)CCCC(=O)OCC	16.3	12.4
4	CF	13.4	15.4
5	C=O	12.8	16.0
6	CCCCCCCCCCCCOC(=O)C(C)=C	14.4	16.7
7	CN(C)CCOC(=O)COc1ccc(Cl)cc1	16	19.6
8	C[P](F)(F)=O	14	16.2
9	C#CC#N	15.5	13.5

10	BrC(Br)C(Br)Br	21	18.2
11	CC(Br)Br	18.5	15.3
12	N	13.7	16.6
13	C1CO1	15.6	17.5
14	NN	14.2	16.5
15	IC(=C)C=C	19.9	16.9
16	IC(I)I	20.2	18.0
17	CCCCCCCCCC(C)(O)Oc1ccccc1	16.7	18.8
18	[O-][O+]=O	19.8	16.9
19	ClCC(Cl)(Cl)Cl	18	15.5
20	FC(F)(Cl)C(F)(Cl)Cl	14.7	12.0
21	OC(=O)C(Cl)=C	19.1	16.1
22	CC(C)S	16.3	13.7
23	OCCCCl	17.5	15.4
24	C=CCCC=O	15.5	18.0
25	NC(=O)NC(N)=O	20	16.7
26	FC(F)Cl	12.3	15.8
27	C#N	12.3	16.1
28	CC1=CC(=O)CC(C)(C)C1	17	19.6
29	CC=C	13.3	15.2
30	s1ccnc1	20.5	17.6
31	FC(F)(Cl)C(F)(F)Cl	12.6	14.9
32	Cn1ccnc1	19.7	17.4
33	CC1=CCC2CC1C2(C)C	16.9	18.8
34	C#C.C#C	14.4	16.8
35	CCCCCCCCCCCCCCCCCC(=O)OCCCC	14.5	17.2
36	CCCCCCOCCCCCC	16	18.1
37	OCCO.O[S](O)=O	20	16.5
38	CSc1ccccc1	19.6	17.3
39	CC1CCC[S]1(=O)=O	19	16.8
40	O[N+]([O-])=O	13.5	16.8
41	C[Si](C)(C)O[Si](C)(C)O[Si](C)(C)C	12.2	14.4
42	FC(F)(F)C1(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C1(F)F	12.4	14.4
43	NCCc1c[nH]c2ccc(O)cc12	18	20.4
44	CCCCCC(CC)(CC)OC(=O)c1ccccc1C([O-])=O	16.6	18.9
45	C=C1CC(=O)O1	16.2	18.1
46	C	14	16.6
47	C[N+]1([O-])CCOCC1	19	16.7
48	C[Si](C)(C)O[Si](C)(C)C	12.6	16.4
49	C#Cc1ccccc1	18.8	16.4
50	O[P](O)(O)=O	14.7	17.4
51	CCC	13.1	15.4

52	CCO[Si](OCC)(OCC)OCC	13.9	16.9
53	NC(N)=S	20	18.0
54	Cl[SiH](Cl)Cl	14.2	17.2
55	NC(=O)c1ccccc1	21.2	19.2
56	[Br]	18.2	16.3
57	CC(CCl)OC(C)CCl	19	16.2
58	C[S](C)(=O)=O	19	16.7
59	C[S](C)=O	18.4	16.3
60	O=C1CCCCCN1	19.4	17.2
61	OC=O	14.6	16.7
62	OC(F)(C(F)F)C(F)(F)F	17.2	11.6
63	Cc1c(C)c(C)c(C)c(C)c1C	19.2	16.7
64	IC(=C)C=C	17.2	19.2
65	CC1=CN([C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O)C(=O)NC1=O	17	20.9
66	ICl	22	17.1
67	CCCCCCCCCc1ccccc1O	16.5	18.7
68	CC1COC(=O)O1	20	17.8

	Morgan Fingerprint/XGBOOST		
No.	**SMILES (60 molecules)**	**δh (exp)**	**δh (pred)**
1	OCCC#N	17.6	10.3
2	CF	9.5	2.0
3	C=O	15.4	7.2
4	COO	30	11.1
5	CN[C@@H](C)[C@H](O)c1ccccc1	24.1	15.5
6	Oc1c(cc(cc1[N+]([O-])=O)[N+]([O-])=O)[N+]([O-])=O	6	15.9
7	[nH]1cnc2ncncc12	14.2	6.1
8	c1ccncc1	5.9	13.4
9	CC(C)(c1ccc(O)cc1)c2c(Br)cc(OBr)c(Br)c2Br	13.8	27.8
10	Oc1ccc(O)cc1	27.2	15.8
11	Cc1ccc(O)c(c1)C(C)(C)C	10.5	23.5
12	CC=NO	20.2	8.5
13	N	18.8	6.5
14	OO	42.7	13.1
15	CO	22.3	9.9
16	[O-][O+]=O	0	7.9
17	FC(F)=C(F)F	0	12.6
18	OC(=O)c1[nH]c2cc(O)c(O)cc2c1	10.6	19.2
19	BrC#N	0	13.2
20	COC(=O)CCC(C)C(=O)OC	5	12.7
21	COC(C)(C)C.OCCO	8.2	15.4

22	C1N2CN3CN1CN(C2)C3	16	5.4
23	CC(N)=S	20.2	10.9
24	OCCOCCOCCOC=C	6.6	15.0
25	O	42.3	6.3
26	OCCCCO.OC(=O)C=C.OC(=O)C=C	4.2	15.0
27	Cn1cnc2N(C)C(=O)N(C)C(=O)c12	13	5.4
28	CCCCCCOCCO	11.4	3.6
29	OCCO.O[S](O)=O	5.1	20.5
30	CC(O)C(O)=O	28.4	12.8
31	OC(=O)C(O)=O	26	14.3
32	NCCc1c[nH]c2ccc(O)cc12	14.4	27.3
33	OC(=O)CS	20	12.4
34	Oc1ccccc1O	21.8	14.5
35	Cn1cnc2nc(Cl)nc(Cl)c12	14.2	3.8
36	CC(N)=O	19.3	11.9
37	N=C=O	13.6	6.0
38	O[P](O)(O)=O	28.4	10.3
39	O=C1CCC(=O)O1	16	7.6
40	CC(O)=S	8.9	26.3
41	c1ccc2nc3ccccc3cc2c1	2	11.5
42	OCC(O)C(O)C(O)CO	36	22.7
43	N#CC#N	0	15.8
44	O=C1CCCCCN1	3.9	11.8
45	OCCO	26	12.6
46	CO	10	17.2
47	CC1=CN([C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O)C(=O)NC1=O	32.8	9.0
48	FC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	0	19.1
49	C[C@@H]1CCC(=C(C)C)C(=O)C1	5.5	14.6

	Morgan Fingerprint/XGBOOST		
No.	**SMILES (60 molecules)**	**δp (exp)**	**δp (pred)**
1	CCc1ccccc1CC	0.1	9.8
2	ClC(Cl)(Cl)C=C	15.5	2.9
3	Cc1ccon1	14.8	4.1
4	C[S](O)(=O)=O.C=C	9.3	15.5
5	OCC(O)CO.OC(O)=O	25.5	13.9
6	CC(=O)C=O	16.1	9.3
7	COO	15	6.3
8	CN[C@@H](C)[C@H](O)c1ccccc1	10.7	4.4
9	CC(F)(F)F	10	2.9

10	Clc1ccc(cc1)C#N	8	14.6
11	O=C1C(=O)c2ccccc2c3ccccc13	17.1	9.6
12	C=CC[N+]#[C-]	13	6.7
13	N	16.7	6.0
14	S=C=S	0	6.3
15	NC=O	26.2	11.9
16	OO	12.2	6.1
17	CNC=O	18.8	11.1
18	[O-][O+]=O	4.2	15.6
19	CCc1ccc(CC)cc1	0	7.4
20	FC(F)=C(F)F	0	8.8
21	CC(=O)C#N	18.9	12.3
22	Cl[S](Cl)=O	6.4	13.1
23	CC(Cl)=C(Cl)Cl	15.7	6.5
24	NC1(CC1)C(O)=O	6.3	13.5
25	CC(C)=CC(C)=O	7.2	14.7
26	CC(=C)C#N	9.5	17.2
27	Cl[S](Cl)(=O)=O	7.2	15.1
28	s1ccnc1	18.8	10.9
29	CC1(C)C2CC3C(C2)C13C	0	6.3
30	O	16	8.2
31	[nH]1nnc2ccccc12	15.6	8.7
32	CCCCCCOC(=O)c1ccccc1C(=O)OCC	6.2	12.5
33	C1OC1C2CO2	14.4	4.4
34	O=C=S	3.7	11.6
35	[Ca++].NC#N	27.6	13.7
36	NCCO	15.5	9.0
37	CC1CCC[S]1(=O)=O	17.4	10.8
38	OC(=O)C(O)=O	17	8.5
39	CC1=CNC(=O)NC1=O	20.5	11.2
40	C/C=C/C#N	18.8	11.9
41	Cc1ccc(c(c1[N+]([O-])=O)[N+]([O-])=O)[N+]([O-])=O	3.5	11.5
42	Cc1cccc2ccccc12	0.8	7.9
43	CCCC=C	1.4	7.7
44	CC(=O)c1ccc(cc1)C(F)(F)F	6.1	12.3
45	COc1ccc(cc1)C#N	16.7	10.0
46	O=C(C#N)C#N	6.3	14.9
47	CCN=[CH+]	15.2	6.8
48	CCOCC.OCC(O)CO.OC(O)=O	18.5	10.7
49	Fc1c(F)c(F)c(F)c(F)c1F	0	7.2
50	CN=C=S	16.2	5.7
51	C[N+]1([O-])CCOCC1	16.1	6.0

52	O[P](O)(O)=O	18.6	11.8
53	c1ccnnc1	17.4	9.9
54	CCCCCCCC\C=C/CCCCCCCCCOCCOCCOCCO	3.1	15.5
55	O=C1c2ccccc2C(=O)c3ccccc13	7.6	14.9
56	[Br]	14.9	5.8
57	ClC(Cl)(Cl)Cl	0	6.3
58	COC(Cl)Cl	12.9	6.4
59	CC(=C)C(O)=O	2.8	11.7
60	ClC(=O)C(Cl)=O	3.8	10.9
61	FC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	0	6.1
62	SC#N	8.9	15.5

	Mol2Vec/FFNN		
No.	**SMILES**	**δd (exp)**	**δd (pred)**
1	CCSC#N	15.4	17.9
2	FC(F)Cl	12.3	16.1
3	ClP(Cl)Cl	18.4	15.5
4	ClC(Cl)=C(Cl)Cl	18.3	15.7
5	C[N+]1([O-])CCOCC1	19	16.2
6	NC(=O)NC(N)=O	20	16.1
7	Cc1c(C)c(C)c(C)c(C)c1C	19.2	16.1
8	[O-][O+]=O	19.8	15.9
9	C=C	15	10.9
10	C[S](C)(=O)=O	19	15.4
11	IC(=C)C=C	19.9	15.9
12	O=[S]1CCCC1	18.2	12.6
13	CN(C)CCOC(=O)COc1ccc(Cl)cc1	16	18.9
14	N#CCCC#N	18.2	14.0
15	[Cl]	17.3	12.0
16	CSSC	17.6	14.0
17	S1C=CSC1=C2SC=CS2	21	14.7
18	C1OC1C2CO2	18.3	15.8
19	N=C=O	15.8	18.8
20	C#C.C#C	14.4	11.1
21	BrCCBr	19.2	16.7
22	CC=CC	14.6	12.1
23	OCCO.O[S](O)=O	20	16.2
24	CC=CC	14.7	12.1
25	ClC(Cl)(Cl)C(Cl)(Cl)Cl	18.6	14.2
26	[nH]1c2ccccc2c3ccccc13	21.7	19.1

27	ClC(Cl)(Cl)C(=O)C(Cl)(Cl)Cl	18.3	15.6
28	CC1CCC[S]1(=O)=O	19	16.3
29	CC	15.5	12.6
30	O[N+]([O-])=O	13.5	16.2
31	C[S](C)=O	18.4	12.9
32	C#N	12.3	15.7
33	c1cncnc1	20.5	17.7
34	CC1=CN([C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O)C(=O)NC1=O	17	20.0
35	ICl	22	17.5
36	ClC1C(Cl)(Cl)C1(Cl)Cl	18.5	14.9
37	FC(F)=C(F)F	15.1	11.8
38	S	17.9	13.3
39	ClCBr	17.3	19.9
40	CC#CC	15.1	12.1
41	COc1c(Cl)cc(Cl)cc1Cl	21	18.3
42	FC(F)(F)c1ccc(Cl)c(Cl)c1	20	17.2
43	IC(I)I	20.2	15.6
44	OC(F)(C(F)F)C(F)(F)F	17.2	13.8
45	ClC=CCl	17	14.2
46	Oc1c(Cl)c(Cl)c(Cl)c(Cl)c1Cl	21.5	18.3
47	NC(N)=S	20	16.5
48	CC1COC(=O)O1	20	16.2
49	S=C=S	19.9	15.0
50	FC(F)(F)c1cc(Cl)ccc1Cl	20	17.3
51	Oc1cccc(O)c1O	20.7	18.1
52	CC(C)=CC(C)=O	16.4	13.6
53	CN(C)[P](=O)(N(C)C)N(C)C	18.5	15.3
54	S=C=S	20.2	15.0
55	OCC(O)C(O)C(O)CO	18	15.3
56	N	13.7	9.9
57	SS	17.3	14.2
58	s1ccnc1	20.5	17.5
59	[Br]	18.2	13.1
60	ClCCCl	18	15.1
61	C[P](F)(F)=O	14	16.5
62	Cl[SiH](Cl)Cl	14.2	17.1
63	C1N2CN3CN1CN(C2)C3	19.4	13.2
64	ClC(Cl)C(Cl)(Cl)Cl	18.2	15.7
65	ClC(Cl)C(Cl)Cl	18.8	14.4
66	O	15.5	10.6
67	C1c2ccccc2c3ccccc13	21.3	18.5
68	CC1OC(=O)OC1C	18	15.2

69	Clc1cc(Cl)c(Cl)cc1Cl	21.2	18.4

	Mol2Vec/FFNN		
No.	**SMILES**	**δh (exp)**	**δh (pred)**
1	N=C=O	13.6	7.0
2	FC=C	1	8.2
3	OO	42.7	25.9
4	COC(C)(C)C.OCCOCCO	7.2	13.9
5	CC(O)=O.CC(O)=O.OCCO	9.8	20.2
6	CN(C)C	1.8	8.5
7	OC(=O)c1[nH]c2cc(O)c(O)cc2c1	10.6	20.2
8	COO	30	14.8
9	CCCCOC.OCCO	4.9	14.3
10	CC(O)=O.CC(O)=O.OCC(O)CO	14.2	21.5
11	OCCCCO.OC(=O)C=C.OC(=O)C=C	4.2	17.3
12	Cn1cnc2nc(Cl)nc(Cl)c12	14.2	8.1
13	CC(O)C(O)=O	28.4	19.4
14	CC1(C)C2CC3C(C2)C13C	0	6.9
15	CC(=O)Nc1ccccc1	13.5	6.1
16	Oc1ccc(O)cc1	27.2	17.2
17	CO	10	19.8
18	CCCCOCC.OCCO	4.6	12.7
19	OC(=O)C(O)=O	26	19.1
20	CC1=CN([C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O)C(=O)NC1=O	32.8	19.3
21	C1N2CN3CN1CN(C2)C3	16	3.8
22	CC(O)=O.CC(O)=O.OCCCCCCO	7.2	16.4
23	CCOCC.OCC(O)CO.OC(O)=O	8.7	19.8
24	CC(N)=S	20.2	9.8
25	OCC(O)CO.OC(O)=O	17.4	26.0
26	O=[S]1CCCC1	9.1	2.9
27	O	42.3	13.6
28	CN	16	9.7
29	CC(C)(C)O	14.7	8.5
30	C[S](C)=O	10.2	4.0
31	CN[C@@H](C)[C@H](O)c1ccccc1	24.1	11.7
32	CF	9.5	1.9
33	C=O	15.4	4.9
34	CC(C)(O)C#N	15.5	8.6
35	CC(O)=S	8.9	15.1
36	CC(N)=O	19.3	13.0
37	NN	8.9	20.0

38	OCCO.O[S](O)=O	5.1	23.8
39	OCC(O)C(O)C(O)CO	36	25.8
40	C#C.C#C	11.9	5.9
41	CC(O)=O.OCC(O)CO.OC(O)=O	9.2	25.0
42	Oc1c(Cl)c(Cl)c(Cl)c(Cl)c1Cl	12.8	6.9
43	C[S](C)(=O)=O	12.3	5.2
44	OC(F)(C(F)F)C(F)(F)F	14.7	4.5
45	CC=NO	20.2	13.8
46	CCN=C=S	9	2.0
47	O=C1CCC(=O)O1	16	8.5
48	CN=C=S	10.1	2.8
49	S	10.2	1.9
50	Cl[P](Cl)(Cl)=O	0	5.9
51	ClB(Cl)Cl	0	6.5
52	N#CSC#N	0	6.7

	Mol2Vec/FFNN		
No.	**SMILES**	**δp (exp)**	**δp (pred)**
1	OCC(O)CO.OC(O)=O	25.5	12.5
2	[nH]1cnc2ccccc12	14.9	7.0
3	Nc1ccncc1	16.1	10.1
4	S=C=S	0	7.3
5	[Ca++].NC#N	27.6	18.0
6	O=C1C(=O)c2ccccc2c3ccccc13	17.1	7.9
7	C=CC[N+]#[C-]	13	5.4
8	CF	10.6	4.6
9	C1COCCO1	1.8	9.6
10	CCCCCCCCCC(C)(O)Oc1ccccc1	10.2	3.1
11	NC=O	26.2	17.0
12	CC(N)=O	18.7	12.8
13	CC(=O)C(C)=O	5.1	11.2
14	CC(F)(F)F	10	2.9
15	Cc1ccc(c(c1[N+]([O-])=O)[N+]([O-])=O)[N+]([O-])=O	3.5	13.9
16	CC1=CNC(=O)NC1=O	20.5	12.7
17	Oc1c(cc(cc1[N+]([O-])=O)[N+]([O-])=O)[N+]([O-])=O	7	13.8
18	[Cl]	10	1.2
19	C/C=C/C#N	18.8	9.1
20	CC(F)=O	14	8.1
21	CC(Cl)=C(Cl)Cl	15.7	5.9
22	OC(=O)C(O)=O	17	10.3

23	C/C=C/C=O	14.9	7.8
24	C[S](O)(=O)=O.C=C	9.3	17.3
25	C#N	17.6	7.2
26	o1cccc1	1.8	9.8
27	CC1OC(=O)OC1C	16.8	9.7
28	[Br]	14.9	1.2
29	N#CC1CC1	16.2	10.0
30	CCN=[CH+]	15.2	8.4
31	NC=C	7.2	14.2
32	CN=C=S	16.2	5.2
33	O=C1OCC=C1	19.8	12.5
34	CCN=C=S	14.7	5.3
35	O	16	8.4
36	CC1(C)C2CC3C(C2)C13C	0	6.1
37	O=C1C=C2NC=CC2=CC1=O	9.3	16.7
38	CC(=O)Nc1ccccc1	14.4	7.9
39	C[N+]1([O-])CCOCC1	16.1	2.8
40	COc1ccc(cc1)C#N	16.7	10.0
41	COO	15	7.4
42	O=C1OC=CO1	18.1	9.6
43	CC(N)=S	20.6	13.2
44	s1ccnc1	18.8	11.0
45	c1ccnnc1	17.4	9.7
46	CCNC=O	10	16.2
47	C=O	14.4	6.0
48	CC(O)=O.CC(O)=O.OCCO	4.7	12.9
49	O=C1OCCO1	21.7	14.3
50	ClC=C(Cl)Cl	3.1	9.1
51	[O-][N+](=O)OCCO[N+]([O-])=O	18	11.8
52	OC=O	10	18.9
53	O=C(C#N)C#N	6.3	18.3
54	Cl[S](Cl)(=O)=O	7.2	13.7
55	ClC(=O)C(Cl)=O	3.8	11.0
56	CCOCC.OCC(O)CO.OC(O)=O	18.5	10.4
57	CC(=C)C(O)=O	2.8	8.7
58	ClC(Cl)(Cl)C=C	15.5	5.1
59	ClC#C	2.1	8.4
60	Cc1ccon1	14.8	8.0
61	[nH]1nnc2ccccc12	15.6	9.5
62	[O-][O+]=O	4.2	11.8
63	CN	7	13.9
64	NN	8.3	19.8

65	C1OC1C2CO2	14.4	6.4
66	Fc1c(F)c(F)c(F)c(F)c1F	0	6.9
67	CNC(C)=O	17	8.5

	Mol2Vec/XGBOOST		
No.	**SMILES**	**δd (exp)**	**δd (pred)**
1	SCCS	17.9	16.1
2	FC(F)(F)c1cccc(c1)C(F)(F)F	17	14.5
3	FC(F)(F)c1cc(Cl)ccc1Cl	20	17.0
4	Cc1ccon1	19.4	16.4
5	CCCCCCCCCCCCCCCCCCCC	16.5	14.6
6	CF	13.4	15.8
7	C=O	12.8	15.3
8	O=CC=O	15	17.1
9	CN(C)CCOC(=O)COc1ccc(Cl)cc1	16	18.2
10	C[P](F)(F)=O	14	16.7
11	N#CCCC#N	18.2	16.3
12	[O-][N+](=O)C([N+]([O-])=O)[N+]([O-])=O	15.5	18.9
13	BrC(Br)C(Br)Br	21	17.9
14	Oc1ccc(O)cc1	21	18.9
15	N	13.7	15.9
16	C=[N+]=[N-]	14.7	17.9
17	FC(F)(Cl)Cl	14.9	16.9
18	NN	14.2	16.4
19	IC(=C)C=C	19.9	17.2
20	[O-][O+]=O	19.8	15.3
21	CC1CO1	15.2	17.3
22	FC(F)(Cl)C(F)(Cl)Cl	14.7	16.9
23	OC(=O)C(Cl)=C	19.1	15.9
24	CNCC(O)c1ccc(O)c(O)c1	20.5	17.9
25	NC(=O)NC(N)=O	20	17.0
26	FC(F)Cl	12.3	16.2
27	BrC#N	18.3	15.6
28	ClC(Cl)(Cl)C(=O)C(Cl)(Cl)Cl	18.3	16.0
29	C#N	12.3	16.1
30	Cl[P](Cl)(Cl)=O	18.1	15.8
31	Cc1ccc(cc1N=C=O)N=C=O	19.3	17.4
32	ClC(Cl)C(Cl)Cl	18.8	16.4
33	FC(F)(Cl)C(F)(F)Cl	12.6	14.8
34	CCCCCCCCCCCCCCCCCC(=O)OCCCC	14.5	16.5

35	C1C=C1	17.2	14.1
36	N#CC1CC1	18.6	16.5
37	CSSC	17.6	15.2
38	CC	15.5	13.6
39	O[N+]([O-])=O	13.5	17.3
40	FC(F)(F)C(F)(F)C(F)(F)C(F)(F)F	10.6	14.1
41	O=[S]1CCCC1	18.2	16.1
42	[nH]1c2ccccc2c3ccccc13	21.7	19.0
43	COC(C)COC(C)COC(C)=O	16.3	14.5
44	C	14	17.7
45	O[P](O)(O)=O	14.7	17.1
46	CCC	13.1	15.2
47	CCO[Si](OCC)(OCC)OCC	13.9	16.1
48	NC(N)=S	20	17.0
49	Cl[SiH](Cl)Cl	14.2	16.9
50	CN(C)C	14.6	16.7
51	CO[P](=O)(OC)OC	15.7	17.5
52	FC(F)(F)C(=O)OC=C	13.9	16.5
53	NC(=O)c1ccccc1	21.2	19.3
54	[Br]	18.2	15.7
55	N#CC#N	15.1	17.3
56	C[S](C)(=O)=O	19	14.7
57	CCI	17.5	15.7
58	OC(F)(C(F)F)C(F)(F)F	17.2	13.0
59	S	17.9	15.0
60	IC(=C)C=C	17.2	19.9
61	ICl	22	18.3
62	CC1COC(=O)O1	20	18.0
63	O=C1N[S](=O)(=O)c2ccccc12	21.1	19.1
64	ClC(Cl)(Cl)C#N	16.4	18.5

	Mol2Vec/XGBOOST		
No.	**SMILES**	**δh (exp)**	**δh (pred)**
1	CC1(O)CCCCC1	12	4.8
2	OC1CCC=C1	15.6	5.4
3	CC(=O)Nc1ccccc1	13.5	4.8
4	C1CC1	0	6.7
5	C=O	15.4	8.5
6	OCC(O)CO.OC(O)=O	17.4	31.4
7	CC(O)=O.OCC(O)CO.OC(O)=O	9.2	21.2
8	COO	30	10.1
9	CN[C@@H](C)[C@H](O)c1ccccc1	24.1	10.6

10	CC(C)(c1ccc(O)cc1)c2c(Br)cc(OBr)c(Br)c2Br	13.8	6.0
11	OCC(F)(F)F	16.4	9.8
12	OCC#C	18.8	8.1
13	CC=NO	20.2	11.8
14	N	18.8	9.3
15	COc1cc(CC=C)ccc1O	13	5.9
16	FC=C	1	7.8
17	NN	8.9	16.5
18	OO	42.7	15.0
19	CO	22.3	10.0
20	OCCN1CCCC1=O	15.7	8.3
21	CNC=O	15.9	8.3
22	ClP(Cl)Cl	0	6.7
23	O=C1OC(=O)c2ccccc12	4.1	11.0
24	CC(C)CC(C)(C)C	0	6.9
25	OCC(Cl)=C	16.4	5.2
26	CC1CCC(O)CC1	12.5	5.7
27	C1N2CN3CN1CN(C2)C3	16	8.2
28	Cl[P](Cl)(Cl)=O	0	7.8
29	CC(N)=S	20.2	10.7
30	O	42.3	16.7
31	CC(C)O	16.4	8.5
32	[Ca++].NC#N	16.8	9.8
33	C1CC=CC1	1.7	11.0
34	CC	0	9.8
35	OCCO.O[S](O)=O	5.1	26.3
36	CC(O)C(O)=O	28.4	16.8
37	O[N+]([O-])=O	18.6	6.3
38	OC(=O)C(O)=O	26	14.0
39	OC(=O)CS	20	12.1
40	OCC(Br)=C	16.2	7.9
41	CC(N)=O	19.3	11.6
42	Oc1ccc(cc1)[S](=O)(=O)c2ccc(O)cc2	16.3	7.3
43	Clc1c(Cl)c(Cl)c(Cl)c(Cl)c1Cl	0	6.9
44	CN	16	5.6
45	CN=C=S	10.1	2.8
46	CCC	0	7.5
47	CC(C)(C)O	14.7	5.5
48	OC(=O)C(Cl)(Cl)Cl	13	5.6
49	CN(C)C	1.8	11.3
50	OCC(O)C(O)C(O)CO	36	17.0
51	OC(F)(C(F)F)C(F)(F)F	14.7	5.2

52	CO	10	22.3
53	CC1=CN([C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O)C(=O)NC1=O	32.8	12.6
54	ICI	5.5	14.6
55	OCCN(CCO)CCO	21	13.3

	Mol2Vec/XGBOOST		
No.	**SMILES**	**δp (exp)**	**δp (pred)**
1	N#Cc1occc1	15	8.8
2	ClC(Cl)(Cl)C=C	15.5	3.7
3	Cc1ccon1	14.8	6.9
4	[nH]1cnc2ccccc12	14.9	6.6
5	CCCCOC(=O)c1ccccc1C(=O)OCc2ccccc2	11.2	4.7
6	C=O	14.4	7.8
7	OCC(O)CO.OC(O)=O	25.5	12.9
8	COO	15	6.8
9	C=CC#C	1.7	8.4
10	O=C1OC=CO1	18.1	6.3
11	S=C1CCCO1	6.9	13.4
12	O=C1C(=O)c2ccccc2c3ccccc13	17.1	7.6
13	C=CC[N+]#[C-]	13	5.8
14	ClB(Cl)Cl	2.5	12.0
15	NC=O	26.2	11.6
16	OCCOC(=O)C=C	13.2	6.9
17	IC(I)I	3.6	10.9
18	CO	12.3	5.0
19	CN(C)C(C)=O	11.5	4.2
20	CNC=O	18.8	12.3
21	[O-][O+]=O	4.2	12.4
22	CC(Cl)=C(Cl)Cl	15.7	6.0
23	CC1OC(=O)OC1C	16.8	9.8
24	O=C=Nc1ccc(Cc2ccc(cc2)N=C=O)cc1	4.1	11.7
25	ClC(Cl)(Cl)Cl	8.3	0.0
26	C#N	17.6	9.5
27	Cl[S](Cl)(=O)=O	7.2	14.1
28	s1ccnc1	18.8	6.5
29	CC(N)=S	20.6	11.9
30	O	16	7.4
31	ClC#C	2.1	9.7
32	[Ca++].NC#N	27.6	12.3

33	N#CC1CC1	16.2	9.2
34	CC	0	9.2
35	CC1(C)CO1	4.8	11.3
36	O[N+]([O-])=O	18	11.4
37	OC(=O)C(O)=O	17	8.4
38	CC1=CNC(=O)NC1=O	20.5	12.0
39	C/C=C/C#N	18.8	6.8
40	Nc1ccncc1	16.1	8.9
41	COc1ccc(cc1)C#N	16.7	10.4
42	Oc1ccc(cc1)[S](=O)(=O)c2ccc(O)cc2	14.6	8.1
43	O=C(C#N)C#N	6.3	14.8
44	C/C=C/C=O	14.9	6.2
45	CCN=[CH+]	15.2	6.1
46	CCOC(=O)C(=C)C#N	10.3	3.9
47	CCOCC.OCC(O)CO.OC(O)=O	18.5	9.2
48	Fc1c(F)c(F)c(F)c(F)c1F	0	8.5
49	C	0	12.4
50	CNC(C)=O	17	10.3
51	C[N+]1([O-])CCOCC1	16.1	9.6
52	Fc1c(F)c(F)c(c(F)c1F)C(=O)c2ccccc2	8.1	1.6
53	O[P](O)(O)=O	18.6	12.0
54	CCC	0	7.6
55	c1ccnnc1	17.4	10.2
56	O=[S]1(=O)CCCC1	17.4	9.9
57	CC(O)=S	6.7	13.2
58	s1cccc1	2.4	9.7
59	C[CH]=[C]=[CH2]	1.7	8.0
60	O=C1c2ccccc2C(=O)c3ccccc13	7.6	17.1
61	NC(=O)c1ccccc1	14.7	5.1
62	[Br]	14.9	5.0
63	CC(=O)C(C)=O	5.1	13.9
64	ClC(Cl)(Cl)Cl	0	8.3
65	C[S](C)(=O)=O	19.4	12.6
66	O=Cc1occc1	14.9	8.3
67	CC(=C)C(O)=O	2.8	10.2
68	CO	5	12.3
69	C[S](=O)(=O)c1ccccc1	16.9	10.0
70	Nc1ccc(cc1)[S](N)(=O)=O	19.5	12.0

	CHEMBERTa_zinc-base-v1		
No.	**SMILES (74 molecules)**	**δd (exp)**	**δd (pred)**
1	COC(=O)C(Cl)=C	15.9	17.7
2	CC1COC(=O)O1	20	17.7
3	BrC(Br)C(Br)Br	21	18.2
4	OC=O	14.6	16.9
5	[SiH3]C=C	15.5	17.7
6	BrC(Br)Br	20	17.9
7	CC(CCl)OC(C)CCl	19	16.2
8	N	13.7	15.6
9	Fc1c(F)c(F)c(F)c(F)c1F	16	18.1
10	Oc1cccc(O)c1O	20.7	18.9
11	BrC#C	15.7	17.9
12	BrC=C	15.9	17.8
13	C#N	12.3	15.4
14	C[C@H]1C[C@H](C)C(=O)[C@@H](C1)[C@H](O)CC2CC(=O)NC(=O)C2	18.3	16.2
15	CC\C(C)=N\O	14.7	16.6
16	Oc1ccccc1O	20	18.2
17	O=C1OC=CO1	17.3	19.0
18	[SiH3]C(Cl)(Cl)Cl	16.5	19.3
19	C1CS1	19.3	17.2
20	FCC=C	14.9	16.8
21	Cl[SiH](Cl)Cl	14.2	17.7
22	FC(F)=C(F)C(F)=C(F)F	13.8	16.4
23	CC(=O)Nc1ccccc1	20.6	18.1
24	NNc1ccccc1	20.4	18.5
25	NC(=O)C=C	15.8	17.5
26	NC(=O)c1ccccc1	21.2	19.4
27	FC(Cl)(Cl)Cl	15.3	17.0
28	Oc1ccc(O)cc1	21	18.7
29	CCSC#N	15.4	17.2
30	OC(F)(C(F)F)C(F)(F)F	17.2	13.6
31	CN(C)[P](=O)(N(C)C)N(C)C	18.5	16.7
32	ClC(Cl)(Cl)C#N	16.4	18.1
33	FC(F)F	14.4	12.6
34	NN	14.2	16.5
35	CN(C)C=O	17.4	15.2
36	NC(=O)NC(N)=O	20	18.2
37	O[N+]([O-])=O	13.5	16.3
38	BrC=C(Br)Br	18.3	20.1
39	SS	17.3	20.0

40	CC(F)F	14.9	13.1
41	NCCc1c[nH]c2ccc(O)cc12	18	20.3
42	[O-][O+]=O	19.8	16.6
43	FC(F)(F)C1(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C1(F)F	12.4	14.2
44	CN(C)CCOC(=O)COc1ccc(Cl)cc1	16	19.1
45	CCCCCCCCCCCCOC(=O)C(C)=C	14.4	16.6
46	ClB(Cl)Cl	16.6	18.3
47	N#CC#N	15.1	16.9
48	BrCC=C	16.5	18.2
49	CCCCCCCCCCCCCCCCCC(=O)OCCCC	14.5	17.0
50	C=O	12.8	15.1
51	FC(F)Cl	12.3	15.4
52	FC(F)(F)C1(C(F)(F)F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C1(F)F	12.4	14.2
53	Cl[S](Cl)=O	16.9	18.6
54	FC(F)(F)C(F)(F)C(F)(F)C(F)(F)F	10.6	12.3
55	IC(=C)C=C	17.2	20.2
56	OCCO.O[S](O)=O	20	17.2
57	OC(=O)C(Cl)=C	19.1	16.6
58	OC1CCOC1	18.9	17.2
59	ClC(=O)C(Cl)=O	16.1	18.4
60	CNC=O	17.4	15.6
61	CCC1COC(=O)O1	17.5	19.4
62	ClC(Cl)(Cl)Cl	16.1	18.0
63	IC(=C)C=C	19.9	17.6
64	CC(=O)Nc1ccc(O)cc1	17.8	20.0
65	BrC#N	18.3	16.6
66	C=C	15	12.6
67	ICC=C	18.3	15.9
68	o1ncc2ccccc12	20.6	18.9
69	CCC	13.1	15.2
70	O=C(C#N)C#N	15	16.7
71	C1CCCCCCCCCCC1	16.4	18.1
72	C=CN1CCCC1=O	16.4	18.3
73	BrC(Br)=C	16.2	18.8
74	ICl	22	16.9

	CHEMBERTa_zinc-base-v1		
No.	**SMILES**	**δh (exp)**	**δh (pred)**
1	OCCC#N	17.6	6.0
2	NCCO	21	6.5

3	CC=NO	20.2	8.3
4	S=C=S	0.6	9.5
5	CCCCO	15.8	7.0
6	O=C=S	0	11.3
7	OC[CH]=[C]=[CH2]	16.8	10.3
8	COO	30	13.0
9	COCCO	15	5.5
10	CN[C@@H](C)[C@H](O)c1ccccc1	24.1	11.8
11	O=C=Nc1ccc(Cc2ccc(cc2)N=C=O)cc1	1.7	9.1
12	CO	22.3	10.9
13	OC(=O)c1[nH]c2cc(O)c(O)cc2c1	10.6	18.9
14	COCO	18.5	7.9
15	O=C1CCC(=O)O1	16	9.5
16	CC(=O)Nc1ccccc1	13.5	6.5
17	NNc1ccccc1	14	6.6
18	CNN	14.8	6.4
19	S=C=S	0.6	9.9
20	Oc1ccc(O)cc1	27.2	18.2
21	O=C1OC(=O)c2ccccc12	4.1	11.4
22	CC(N)=S	20.2	12.4
23	CCNCC	6.1	14.9
24	CC1=CN([C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O)C(=O)NC1=O	32.8	17.1
25	OC(F)(C(F)F)C(F)(F)F	14.7	5.0
26	COCCOC	6	16.0
27	OC(=O)C(O)=O	26	18.5
28	OCC(F)(F)F	16.4	7.9
29	OCCSCCO	19.8	12.4
30	[Cl]	0	10.6
31	O	42.3	20.5
32	CC(O)C(O)=O	28.4	18.3
33	CI	5.3	12.5
34	OCC(O)C(O)C(O)CO	36	20.4
35	CN=C=S	10.1	-0.2
36	O[N+]([O-])=O	18.6	5.1
37	COCCOC	6	16.0
38	CC(CO)[N+]([O-])=O	15.4	7.6
39	CCCCCCCCCCCCCCCCO	8.1	1.5
40	N=C=O	13.6	7.1
41	CCO[Si](OCC)(OCC)OCC	0.6	7.0
42	CO[N+]([O-])=O	4.8	14.0
43	[Ca++].NC#N	16.8	6.8

44	[O-][O+]=O	0	8.7
45	Cl[S](Cl)(=O)=O	0	8.5
46	OCCCCO.OC(=O)C=C.OC(=O)C=C	4.2	10.7
47	CO	10	21.8
48	N#CC#N	0	9.1
49	Oc1c(cc(cc1[N+]([O-])=O)[N+]([O-])=O)[N+]([O-])=O	6	13.5
50	Cl[P](Cl)(Cl)=O	0	7.1
51	O[P](O)(O)=O	28.4	21.5
52	CCCC(F)(F)C(F)(F)C(O)(F)F	9.9	2.5
53	OO	42.7	21.8
54	OCCCO	23.2	7.6
55	CCCO	17.4	5.5
56	N#CSC#N	0	9.0
57	OCCOc1ccccc1	14.3	6.0
58	CCO	19.4	9.5
59	OCCc1ccccc1	11.2	3.8
60	CCCCCCCO	11.7	4.8
61	CF	9.5	2.8
62	OCCOC(=O)C=C	13.4	6.4
63	OCCO.O[S](O)=O	5.1	15.8
64	CN=C=O	2.5	9.7
65	[Br]	0	13.1
66	C[SiH3]	0	8.1
67	C	0	10.8
68	OC(=O)CS	20	12.9
69	OCCN1CCCC1=O	15.7	8.9
70	BrC#N	0	6.5
71	C1CCCCCCCCCCC1	0	7.3
72	CCN=C=S	9	1.5
73	OCCOCCOCCOC=C	6.6	15.3

	CHEMBERTa_zinc-base-v1		
No.	**SMILES**	**δp (exp)**	**δp (pred)**
1	SC#N	8.9	15.4
2	CC1COC(=O)O1	18	10.1
3	NCCO	15.5	6.5
4	Cc1ccon1	14.8	4.6
5	O=C1OCC=C1	19.8	12.2
6	O=C1c2ccccc2C(=O)c3ccccc13	7.6	13.7
7	CC(=C)C(Cl)=O	10.6	4.0
8	c1ccnnc1	17.4	7.7

9	OC=O	10	16.4
10	Brc1ccc(cc1)C#N	9.3	16.5
11	CCCCCCCCCCC	0	6.0
12	Fc1c(F)c(F)c(F)c(F)c1F	0	6.9
13	S=C=S	0	11.6
14	CC1OC(=O)OC1C	16.8	8.9
15	ClC(Cl)(Cl)Cl	0	8.9
16	O=C=S	3.7	11.5
17	CC\C(C)=N\O	4.9	11.9
18	Cc1cnccn1	12.3	4.2
19	Clc1cc(Cl)c(Cl)cc1Cl	10.7	2.8
20	Clc1ccc(cc1)C#N	8	15.2
21	CC(=C)C(C)=O	12.1	4.6
22	C1OC1C2CO2	14.4	7.6
23	CCCCOC(=O)/C=C/C(=O)OCCCC	3	9.0
24	CO	12.3	5.4
25	NC1(CC1)C(O)=O	6.3	12.4
26	CCOCC.OCC(O)CO.OC(O)=O	18.5	11.0
27	CCN=[CH+]	15.2	9.3
28	ClC(Cl)(Cl)C=O	7.4	14.4
29	O=[S]=O	8.4	17.0
30	CC(=O)Nc1ccccc1	14.4	7.8
31	CC(Cl)CCl	7.1	13.9
32	s1cnc2ccccc12	5.2	12.6
33	ClC1(Cl)CCCC(Cl)(Cl)C1=O	14	6.8
34	OC(=O)C(O)=O	17	9.5
35	Cc1occc1	2.8	9.8
36	Sc1cc(Cl)c(Cl)cc1Cl	4.5	13.4
37	Cn1ccnc1	15.6	7.3
38	CC(=O)C(C)=O	5.1	13.1
39	o1cccn1	13.4	7.0
40	O=CC=O	17	11.0
41	CN=C=S	16.2	4.8
42	ClC(Cl)(Cl)C=C	15.5	4.5
43	FC(F)=C(F)F	0	6.4
44	[Ca++].NC#N	27.6	13.5
45	C/C=C/C=O	14.9	8.9
46	O=C\C=C\c1ccccc1	12.4	6.2
47	[O-][O+]=O	4.2	14.7
48	Cl[S](Cl)(=O)=O	7.2	14.7
49	CO	5	13.0
50	CC(=O)C#N	18.9	10.3

51	N#CC#N	11.8	17.8
52	NC=O	26.2	15.4
53	Nc1ccccn1	8.1	15.9
54	OCCCO	13.5	6.7
55	C/C=C/C#N	18.8	11.3
56	CC(=C)C(O)=O	2.8	9.9
57	Cl[S](Cl)=O	6.4	14.3
58	[nH]1ccc2ccccc12	7.5	13.6
59	O=C1C(=O)c2ccccc2c3ccccc13	17.1	9.1
60	CNC=O	18.8	13.1
61	ClC(Cl)(Cl)Cl	8.3	0.4
62	OCC(O)CO.OC(O)=O	25.5	13.7
63	Nc1ccncc1	16.1	8.8
64	s1ccnc1	18.8	5.4
65	CNC(C)=O	17	6.4
66	O=C(C#N)C#N	6.3	14.6
67	C1CCCCCCCCCCC1	0	5.9
68	CC(Cl)=C(Cl)Cl	15.7	7.8
69	OCCOCCOCCO	12.5	6.7
70	O=Cc1occc1	14.9	9.0
71	CCCCCCCCCC(C)(O)Oc1ccccc1	10.2	4.3
72	CCN=C=S	14.7	4.5

	CHEMBERTa_77M-MTR		
No.	**SMILES**	**δd (exp)**	**δd (pred)**
1	CC1COC(=O)O1	20	17.2
2	CC(CCl)OC(C)CCl	19	14.6
3	CC(C)(C)Cl	15.6	17.8
4	C[Si](C)(C)O[Si](C)(C)O[Si](C)(C)C	12.2	9.6
5	C#N	12.3	14.6
6	CC\C(C)=N\O	14.7	16.7
7	ClCCCCl	18	15.9
8	C[Si](C)(C)O[Si](C)(C)O[Si](C)(C)O[Si](C)(C)C	11.7	14.1
9	C[N+]1([O-])CCOCC1	19	16.8
10	Cl[SiH](Cl)Cl	14.2	16.7
11	ClCCCl	18	15.7
12	Clc1ccc(cc1)C(Cl)(Cl)Cl	20.3	17.8
13	CC=C(C)C	14.7	16.8
14	ClC1(Cl)CCCC(Cl)(Cl)C1=O	19.5	17.1
15	NC(=O)C=C	15.8	18.3
16	COc1ccc2nccc([C@@H](O)[C@@H] 3C[C@H]4CCN3C[C@@H]4C=C)c2c1	19	16.4

17	ClC(Cl)C(Cl)Cl	18.8	16.4
18	CCSC#N	15.4	17.4
19	OC(F)(C(F)F)C(F)(F)F	17.2	13.0
20	CN(C)[P](=O)(N(C)C)N(C)C	18.5	16.4
21	COc1c(Cl)cc(Cl)cc1Cl	21	19.0
22	IC=C	18	16.0
23	[Cl]	17.3	15.3
24	NC(=O)NC(N)=O	20	17.0
25	CN=C=S	17.3	13.9
26	O[N+]([O-])=O	13.5	16.2
27	S	17.9	15.3
28	Cc1cccc(C)c1C	18	20.1
29	CCO[Si](OCC)(OCC)OCC	13.9	15.9
30	CNCC(O)c1ccc(O)c(O)c1	20.5	18.4
31	FC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	11.3	14.6
32	IC[CH]=[C]=[CH2]	17.4	19.6
33	[O-][O+]=O	19.8	14.6
34	FC(F)(F)C1(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C1(F)F	12.4	14.8
35	C[S](C)(=O)=O	19	16.7
36	FC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	12	14.9
37	CC(C)C=C	14	16.0
38	CN(C)CCOC(=O)COc1ccc(Cl)cc1	16	18.2
39	ClP(Cl)Cl	18.4	16.3
40	C1CO1	15.6	18.3
41	ClC=C(Cl)Cl	18	15.9
42	Cl[P](Cl)(Cl)=O	18.1	16.1
43	CCC=C	13.2	15.7
44	FC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	12.1	14.1
45	Cc1cc(C)cc(C)c1	18	20.0
46	CCCCCCCCCCCCCCCCCC(=O)OCCCC	14.5	16.5
47	N#CSC#N	18.1	8.3
48	FC(F)Cl	12.3	14.4
49	FC(F)(F)C(F)(F)C(F)(F)C(F)(F)F	10.6	13.8
50	IC(=C)C=C	17.2	19.3
51	ClCCCCCl	18.3	15.9
52	ClCCOCCCl	18.8	15.7
53	OCCO.O[S](O)=O	20	16.7
54	[Br]	18.2	16.1
55	CCCC	14.1	16.1
56	ClCC=C	17	15.0
57	IC(=C)C=C	19.9	16.7
58	OC(=O)CS	16	18.6

59	ClCCCCCCl	18	16.0
60	BrC#N	18.3	16.0
61	CCC	13.1	16.0
62	CCC(C)C	13.8	16.2
63	ICl	22	17.8
64	CCN=C=S	17.2	12.7

	CHEMBERTa_77M-MTR		
No.	**SMILES**	**δh (exp)**	**δh (pred)**
1	SC#N	10.9	5.0
2	C[P](F)(F)=O	8.4	-0.9
3	CC=NO	20.2	13.5
4	OC=O	14	22.8
5	N	18.8	12.7
6	S=C=S	0.6	8.3
7	CN(C)C	1.8	9.1
8	C#N	9	0.4
9	C[C@H]1C[C@H](C)C(=O)[C@@H](C1)[C@H](O)CC2CC(=O)NC(=O)C2	13.8	7.5
10	O=C=S	0	6.2
11	COO	30	16.6
12	OCC1COCO1.OC2COCOC2	16.5	10.4
13	Cn1cnc2nc(Cl)nc(Cl)c12	14.2	7.5
14	ClC#C	2.5	8.5
15	C1N2CN3CN1CN(C2)C3	16	5.4
16	CN[C@@H](C)[C@H](O)c1ccccc1	24.1	12.3
17	CO	22.3	14.5
18	OC(=O)c1[nH]c2cc(O)c(O)cc2c1	10.6	19.6
19	O=C1CCC(=O)O1	16	8.2
20	CC(=O)Nc1ccccc1	13.5	7.0
21	CNN	14.8	8.7
22	S=C=S	0.6	6.8
23	ClCCl	7.1	-0.4
24	COc1ccc2nccc([C@@H](O)[C@@H]3C[C@H]4CCN3C[C@@H]4C=C)c2c1	11	3.5
25	Oc1ccc(O)cc1	27.2	21.5
26	CC(N)=S	20.2	11.1
27	CC(O)=S	8.9	17.9
28	CC1=CN([C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O)C(=O)NC1=O	32.8	16.5
29	OC(F)(C(F)F)C(F)(F)F	14.7	5.8
30	OC(=O)C(O)=O	26	17.2
31	C=CCN=[N+]=[N-]	13.4	6.2

32	Fc1c(F)c(F)c(c(F)c1F)C(=O)c2ccccc2	5.4	-0.7
33	NN	8.9	19.8
34	O	42.3	22.8
35	CC(O)C(O)=O	28.4	20.5
36	OCC(O)C(O)C(O)CO	36	23.1
37	NC(=O)NC(N)=O	18.8	12.7
38	O[N+]([O-])=O	18.6	8.3
39	S	10.2	17.6
40	CCCCC(CC)COC(=O)CCCCCCCCCC(=O)OCC(CC)CCCC	9	3.2
41	N=C=O	13.6	5.3
42	CCO[Si](OCC)(OCC)OCC	0.6	6.2
43	[Ca++].NC#N	16.8	9.7
44	[O-][O+]=O	0	6.8
45	C[S](C)(=O)=O	12.3	6.0
46	Cl[S](Cl)(=O)=O	0	6.0
47	OCCCCO.OC(=O)C=C.OC(=O)C=C	4.2	13.2
48	CO	10	20.9
49	C1CNCCN1	6	14.1
50	N#CC#N	0	7.9
51	Oc1c(cc(cc1[N+]([O-])=O)[N+]([O-])=O)[N+]([O-])=O	6	11.7
52	Cl[P](Cl)(Cl)=O	0	7.4
53	O[P](O)(O)=O	28.4	20.6
54	OO	42.7	20.8
55	CC#CC	7.6	1.9
56	C=O	15.4	5.3
57	CF	9.5	1.0
58	CC#C	9.2	2.7
59	OCCO.O[S](O)=O	5.1	18.4
60	C[SiH3]	0	6.0
61	CC(=O)Nc1ccc(O)cc1	13.9	20.4
62	OC(=O)CS	20	13.9
63	C1C=Cc2ccccc12	9	3.1
64	OCCOCCOCCOC=C	6.6	13.0
65	CCN=[N+]=[N-]	12.9	6.1

	CHEMBERTa_77M-MTR		
No.	**SMILES**	**δp (exp)**	**δp (pred)**
1	SC#N	8.9	16.5
2	CC1COC(=O)O1	18	11.1
3	O=C1c2ccccc2C(=O)c3ccccc13	7.6	14.0
4	CC(=C)C(Cl)=O	10.6	4.2
5	CC1=CNC(=O)NC1=O	20.5	13.9

6	c1ccnnc1	17.4	7.0
7	OC=O	10	17.6
8	Brc1ccc(cc1)C#N	9.3	17.6
9	[SiH3]C=C	2.6	9.4
10	Nc1ncnc2nc[nH]c12	16	8.4
11	Fc1c(F)c(F)c(F)c(F)c1F	0	8.9
12	S=C=S	0	11.6
13	CC1OC(=O)OC1C	16.8	10.0
14	ClC(Cl)(Cl)Cl	0	6.8
15	COO	15	8.7
16	Clc1cc(Cl)c(Cl)cc1Cl	10.7	-0.3
17	Clc1ccc(cc1)C#N	8	17.3
18	[nH]1nnc2ccccc12	15.6	9.0
19	C[N+]1([O-])CCOCC1	16.1	9.3
20	CC(=C)C(C)=O	12.1	2.8
21	NC1(CC1)C(O)=O	6.3	14.1
22	CCN=[CH+]	15.2	7.6
23	COC(Cl)Cl	12.9	4.4
24	O=[S]=O	8.4	16.4
25	s1cnc2ccccc12	5.2	13.0
26	S=C=S	5.8	14.4
27	C[S](C)=O	16.4	6.2
28	OC(=O)C(O)=O	17	9.7
29	FC(F)F	8.7	1.7
30	CCCCCCl	6.9	0.5
31	CC(=O)C(C)=O	5.1	11.7
32	S	6	15.1
33	ClC(Cl)(Cl)C=C	15.5	5.0
34	ClC(Cl)=C(Cl)Cl	5.7	14.7
35	[Ca++].NC#N	27.6	14.1
36	[O-][O+]=O	4.2	15.8
37	ClCNCCl	7.5	1.1
38	CC(C)=CC(C)=O	7.2	14.1
39	Cl[S](Cl)(=O)=O	7.2	17.8
40	Cc1ccc(Cl)c(Cl)c1	9.8	1.5
41	COC(C)=C	4.2	10.6
42	CO	5	12.0
43	CC(=O)C#N	18.9	10.4
44	NC=O	26.2	17.7
45	Nc1ccccn1	8.1	17.6
46	CC=CCl	6.9	-0.5
47	CCC=C	1.3	7.6

48	Cc1cc(C)cc(C)c1	0.6	9.1
49	C/C=C/C#N	18.8	10.7
50	ClC1C(Cl)(Cl)C1(Cl)Cl	10.5	2.6
51	N#CSC#N	13.5	5.5
52	CCC(C)=C	1.8	9.2
53	CCCCCC	0	7.2
54	FC(F)(F)C1(C(F)(F)F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C1(F)F	0	8.1
55	Cl[S](Cl)=O	6.4	18.0
56	CC	0	6.5
57	CCc1ccc(CC)cc1	0	7.3
58	CC(F)=O	14	7.5
59	ClC=CCl	8	-0.5
60	OCCOC(=O)C=C	13.2	6.6
61	ClC1CC1	7.2	0.8
62	ClC(=O)C(Cl)=O	3.8	11.7
63	C[Si](C)(C)C=C	1	7.8
64	CCCC	0	8.0
65	O=C1C(=O)c2ccccc2c3ccccc13	17.1	8.0
66	CCNC=O	10	16.6
67	ClCCCCCCl	7.8	-0.4
68	OCC(O)CO.OC(O)=O	25.5	16.2
69	CCC	0	8.4
70	Nc1ccncc1	16.1	9.3
71	s1ccnc1	18.8	12.2
72	CCC(C)C	0	9.4
73	O=C(C#N)C#N	6.3	16.9
74	CC(Cl)=C(Cl)Cl	15.7	5.8
75	BrC(Br)=C	4.8	11.6
76	Cc1ccc(c(c1[N+]([O-])=O)[N+]([O-])=O)[N+]([O-])=O	3.5	11.2
77	CCCCCCCl	6.2	-0.4