Electronic Supplementary Material (ESI) for RSC Advances.
This journal is © The Royal Society of Chemistry 2021

Supplementary Information

Transmol: Repurposing a language model for molecular generation

R. Zhumagambetov, F. Molnár, V. A. Peshkov and S. Fazli

S1 Filters

In this section we briefly describe filters used in the analysis of the output molecules. As can be seen in Table S1, there are 5 different medical filters have been used. All filters rely on the empirical rules that certain molecules are highly likely to be useful. These rules are listed in the Constraints columns.

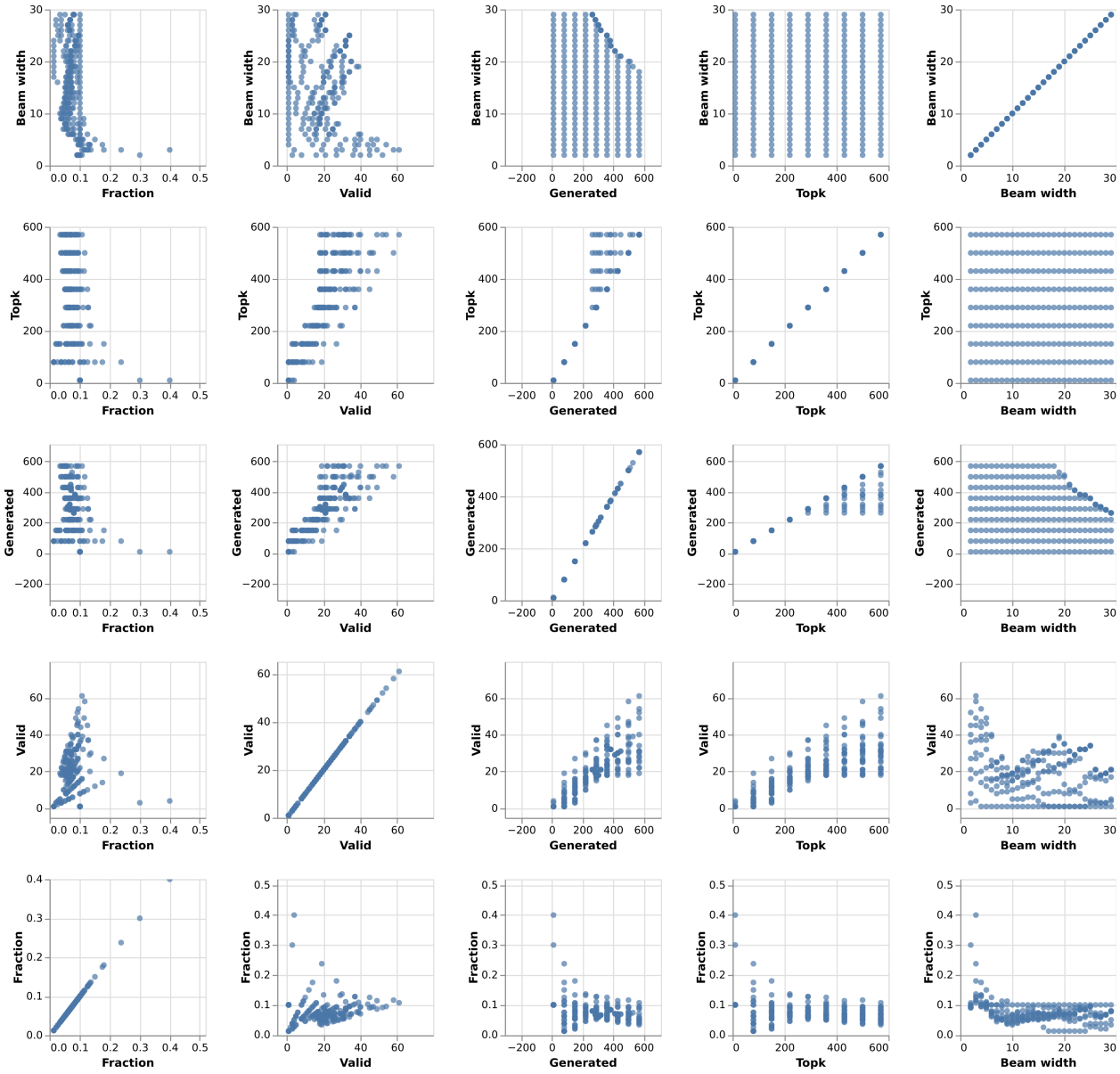
Filter	Constraints
Lipinski rule of 5[1] (at most 1 violation is allowed)	Molecular weight <= 500 LogP <= 5.0 H-bond donor count <= 5 H-bond acceptor count <= 10
Ghose[2]	Molecular weight between 160 and 480 LogP between -0.4 and +5.6 Atom count between 20 and 70 Molar refractivity between 40 and 130
Veber[3]	Rotatable bonds <= 10 Topological polar surface area <= 140
Rule of 3[4]	Molecular weight <= 300 LogP <= 3 H-bond donor <= 3 H-bond acceptor count <= 3 Rotatable bond count <= 3
REOS[5]	Molecular weight between 200 and 500 LogP between -5.0 and +5.0 H-bond donor count between 0 and 5 H-bond acceptor count between 0 and 10 Formal charge between -2 and +2 Rotatable bond count between 0 and 8 Heavy atom count between 15 and 50

Table S1: Description of filters used

S2 Adjusting beam search

S2.1 Adjusting beam search

To find the optimal parameters for the beam search a grid search has been conducted. A random molecule from the test set has been selected for the testing. Figure S1 shows the relation between 3 parameters (*Beam width*, *Topk* (the desired number of generated molecules), *Generated* (actual number of generated molecules), and 2 metrics (number of valid molecules and the fraction of valid molecules in the generated set). The chart demonstrates a dependency between *Beam width* and *Topk* (number of generated molecules)). In addition, for small numbers of *Topk*, the fraction of valid molecules is high. In this paper we have used a reward function that is reliant only on the input provided by the model. Further work could investigate other possible functions. For instance, the reward function could incorporate expert knowledge.



Supplemental Material, Figure S1: The scatter plot of the grid search with the following parameters: beam width, number of generation requests, actual number of generated SMILES strings, number of valid molecules, and fraction of valid molecules

S3 Example molecules

To generate a molecule that resembles both seeds we use the special latent vector z_{12} that was created using the weighted mean of individual vectors z_1 and z_2. One way to adjust the hybrid molecular output is to adjust the weights applied to individual vectors. Here we present the results of such adjustments using 3 exemplary pairs of molecules. Tables S5 to S7 contain the results of weight adjustments. In Tables S2 to S4 weights for both seeds are equal but α, a beam search parameter that influences the length, is perturbed. To demonstrate that the algorithm can be adjusted to produce heavier molecules we also generated molecular output with sampling reward 1.9 and 2.5 which can seen in Tables S8 and S9.

Seed molecule 1																		Seed molecule 2				
No	Sampling reward	Length				Tanimoto similarity 1				Tanimoto similarity 2				Tanimoto similarity (1+2)				Molecule 1	Molecule 2	Molecule 3	Molecule 4	Molecule 5
		min	max	mean	median	min	max	mean	median	min	max	mean	median	min	max	mean	median					
1	4	13	20	16.2	16	0.223	0.486	0.376	0.369	0.189	0.521	0.275	0.256	0.46	0.885	0.651	0.645					
2	3.7	13	20	16.2	16	0.223	0.486	0.371	0.367	0.189	0.521	0.265	0.25	0.46	0.885	0.635	0.638					
3	3.4	13	20	15.9	16	0.223	0.486	0.368	0.369	0.168	0.521	0.262	0.242	0.46	0.916	0.63	0.633					
4	3.1	11	21	16.2	16	0.223	0.486	0.374	0.378	0.168	0.487	0.255	0.242	0.46	0.916	0.629	0.637					
5	2.8	10	21	15.9	16	0.223	1	0.39	0.377	0.166	0.487	0.262	0.246	0.46	1.3	0.652	0.635					
6	2.5	10	23	16.2	16	0.223	1	0.389	0.37	0.166	0.521	0.262	0.246	0.46	1.3	0.651	0.628					
7	2.2	10	22	16	16	0.223	1	0.386	0.368	0.162	0.521	0.26	0.241	0.46	1.3	0.645	0.618					
8	1.9	5	22	15.6	16	0.175	1	0.381	0.372	0.159	0.521	0.265	0.241	0.334	1.3	0.646	0.617					
9	1.6	1	22	14.7	15	0	1	0.37	0.365	0	0.521	0.271	0.248	0	1.3	0.641	0.616					
10	1.35	1	19	14	15	0	1	0.364	0.358	0	0.702	0.28	0.255	0	1.3	0.644	0.63					
11	1.3	1	20	13.6	15	0	1	0.363	0.363	0	0.516	0.288	0.272	0	1.3	0.651	0.628					
12	1.2	1	19	13.3	14	0	1	0.356	0.356	0	0.702	0.284	0.265	0	1.3	0.64	0.633					
13	1.05	1	19	12.6	13	0	1	0.349	0.354	0	0.516	0.289	0.283	0	1.3	0.639	0.639					
14	0.9	1	19	11.7	12	0	1	0.339	0.353	0	0.516	0.291	0.294	0	1.3	0.629	0.642					
15	0.75	1	19	10.9	11	0	0.893	0.328	0.352	0	0.524	0.289	0.29	0	1.18	0.617	0.634					
16	0.6	1	19	9.53	10	0	0.893	0.31	0.333	0	0.516	0.278	0.287	0	1.18	0.588	0.633					
17	0.45	1	17	8.23	8	0	0.706	0.287	0.314	0	0.516	0.268	0.294	0	1.12	0.555	0.627					
18	0.3	1	16	7.39	7	0	0.694	0.263	0.305	0	0.516	0.253	0.286	0	1.12	0.516	0.596					
19	0.15	1	14	6.45	6	0	0.694	0.236	0.246	0	0.516	0.233	0.258	0	1.12	0.469	0.502					
20	0	1	14	6.03	6	0	0.492	0.222	0.22	0	0.516	0.223	0.238	0	1.01	0.445	0.468					

Table S2: Comparison of sampling reward for the first pair of molecules

Seed molecule 1																		Seed molecule 2				
No	Sampling reward	Length min				Tanimoto mol1 min				Tanimoto mol2 min				Tanimoto sum min				Molecule 1	Molecule 2	Molecule 3	Molecule 4	Molecule 5
		min	max	mean	median	min	max	mean	median	min	max	mean	median	min	max	mean	median					
1	2.5	13	23	17.3	17	0.258	0.438	0.346	0.348	0.18	1	0.355	0.268	0.438	1.32	0.701	0.657					
2	2.1	13	26	15.7	15	0.229	0.5	0.352	0.327	0.176	1	0.343	0.317	0.476	1.32	0.695	0.67					
3	1.7	10	22	15.4	15	0.228	0.5	0.352	0.329	0.176	1	0.346	0.318	0.476	1.32	0.698	0.645					
4	1.35	1	26	14.8	15	0	0.498	0.348	0.374	0	1	0.321	0.288	0	1.32	0.669	0.646					
5	1.3	1	22	14.3	15	0	0.5	0.344	0.334	0	1	0.332	0.31	0	1.32	0.676	0.677					
6	1.2	1	20	14.4	15	0	0.498	0.341	0.373	0	1	0.321	0.305	0	1.32	0.662	0.652					
7	1.05	1	20	12.9	15	0	0.5	0.313	0.337	0	1	0.302	0.301	0	1.32	0.615	0.641					
8	0.9	1	20	12	14	0	0.5	0.305	0.329	0	1	0.295	0.293	0	1.32	0.599	0.634					
9	0.75	1	20	10.7	13	0	0.5	0.284	0.322	0	1	0.276	0.288	0	1.32	0.559	0.613					
10	0.6	1	20	9.92	10	0	0.5	0.278	0.322	0	1	0.273	0.291	0	1.32	0.551	0.625					
11	0.45	1	18	9.1	8	0	0.498	0.271	0.323	0	0.702	0.261	0.291	0	1.08	0.533	0.625					
12	0.3	1	18	8.4	8	0	0.498	0.257	0.312	0	0.702	0.249	0.281	0	1.08	0.506	0.598					
13	0.15	1	19	7.26	7	0	0.516	0.236	0.259	0	0.524	0.227	0.249	0	1.04	0.463	0.541					
14	0	1	17	6.71	7	0	0.547	0.226	0.203	0	0.556	0.22	0.206	0	1.1	0.446	0.409					

Table S3: Comparison of sampling reward for the second pair of molecules

Seed molecule 1

Seed molecule 2

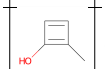
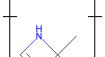
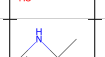
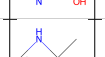
No	Sampling reward	Length min				Tanimoto mol1 min				Tanimoto mol2 min				Tanimoto sum min				Molecule 1	Molecule 2	Molecule 3	Molecule 4	Molecule 5
		min	max	mean	median	min	max	mean	median	min	max	mean	median	min	max	mean	median					
1	6.7	36	36	36	36	0.36	0.36	0.36	0.36	0.426	0.426	0.426	0.426	0.786	0.786	0.786	0.786					
2	6.1	36	36	36	36	0.36	0.36	0.36	0.36	0.426	0.426	0.426	0.426	0.786	0.786	0.786	0.786					
3	5.5	36	36	36	36	0.36	0.36	0.36	0.36	0.426	0.426	0.426	0.426	0.786	0.786	0.786	0.786					
4	4.9	36	36	36	36	0.36	0.36	0.36	0.36	0.426	0.426	0.426	0.426	0.786	0.786	0.786	0.786					
5	4.3	36	36	36	36	0.36	0.36	0.36	0.36	0.426	0.426	0.426	0.426	0.786	0.786	0.786	0.786					
6	3.7	36	36	36	36	0.36	0.36	0.36	0.36	0.426	0.426	0.426	0.426	0.786	0.786	0.786	0.786					
7	3.1	36	36	36	36	0.36	0.36	0.36	0.36	0.426	0.426	0.426	0.426	0.786	0.786	0.786	0.786					
8	2.5	36	36	36	36	0.36	0.372	0.365	0.362	0.426	0.5	0.453	0.435	0.786	0.862	0.818	0.807					
9	1.9	30	44	35.6	36	0.349	0.42	0.373	0.367	0.409	0.501	0.447	0.433	0.784	0.862	0.82	0.82					
10	1.35	1	32	9.55	9	0	0.531	0.322	0.362	0	0.531	0.322	0.363	0	1.06	0.644	0.724					
11	1.3	1	20	9.13	9	0	0.406	0.234	0.234	0	0.406	0.233	0.234	0	0.812	0.467	0.469					
12	1.2	1	32	9.07	7.5	0	0.547	0.297	0.348	0	0.547	0.297	0.352	0	1.09	0.594	0.704					
13	1.05	1	32	7.76	7	0	0.547	0.264	0.258	0	0.547	0.264	0.258	0	1.09	0.528	0.516					
14	0.9	1	13	6.33	6	0	0.547	0.244	0.219	0	0.547	0.244	0.219	0	1.09	0.487	0.438					
15	0.75	1	13	6.34	6	0	0.547	0.239	0.203	0	0.547	0.239	0.203	0	1.09	0.477	0.406					
16	0.6	1	13	6.25	6	0	0.547	0.228	0.203	0	0.547	0.228	0.203	0	1.09	0.456	0.406					
17	0.45	1	13	5.98	6	0	0.547	0.216	0.188	0	0.547	0.216	0.188	0	1.09	0.432	0.375					
18	0.3	1	13	5.73	6	0	0.547	0.204	0.156	0	0.547	0.204	0.156	0	1.09	0.408	0.312					
19	0.15	1	13	5.6	6	0	0.547	0.196	0.156	0	0.547	0.196	0.156	0	1.09	0.393	0.312					
20	0	1	12	5.48	6	0	0.547	0.188	0.156	0	0.547	0.188	0.156	0	1.09	0.377	0.312					

Table S4: Comparison of sampling reward for the third pair of molecules

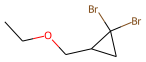
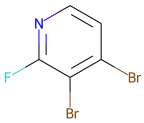
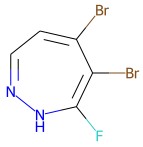
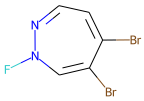
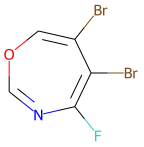
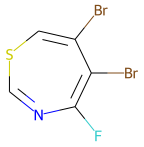
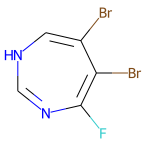
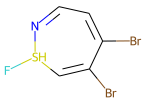
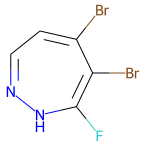
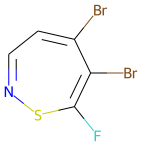
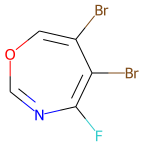
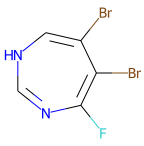
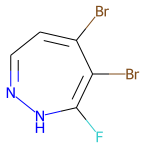
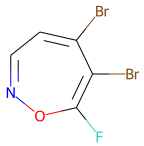
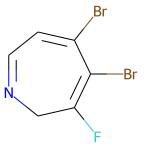
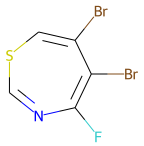
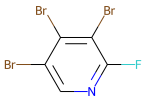
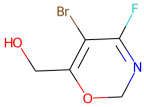
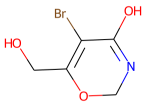
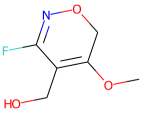
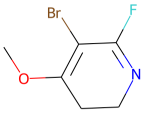
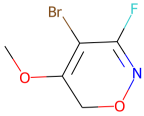
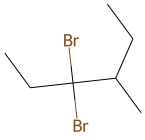
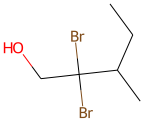
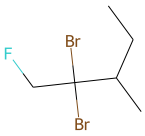
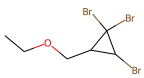
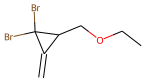
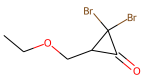
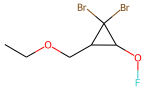
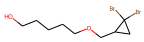
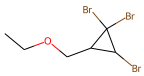
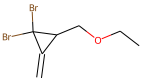
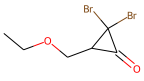
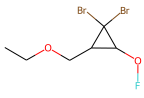
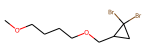
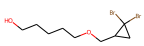
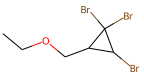
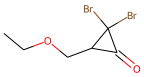
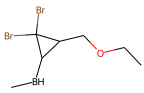
Seed molecule 1				Seed molecule 2			
							
No	Weight 1	Weight 2	Molecule 1	Molecule 2	Molecule 3	Molecule 4	Molecule 5
1	0.15	0.85					
2	0.25	0.75					
3	0.35	0.65					
4	0.45	0.55					
5	0.55	0.45					
6	0.65	0.35					
7	0.75	0.25					
8	0.85	0.15					

Table S5: Comparison of weights for the first pair of molecules with sampling reward 0.75

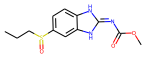
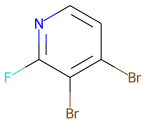
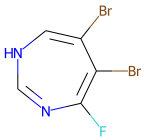
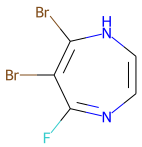
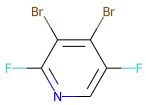
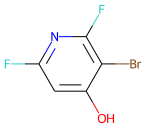
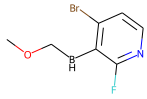
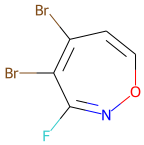
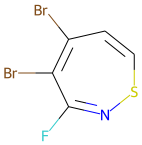
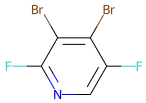
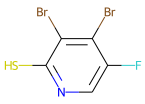
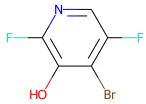
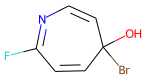
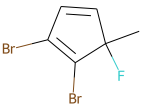
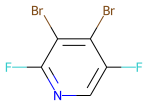
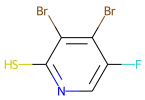
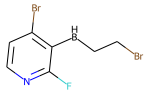
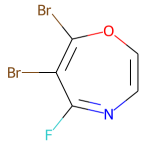
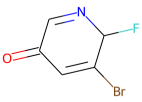
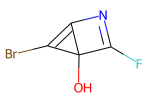
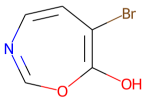
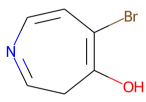
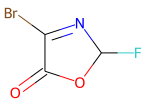
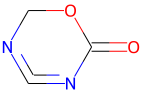
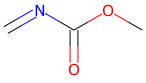
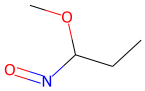
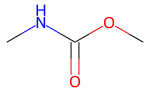
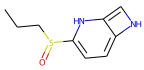
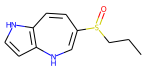
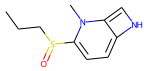
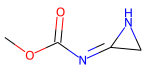
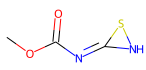
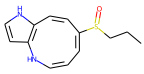
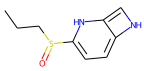
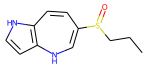
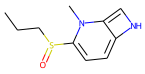
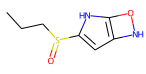
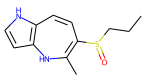
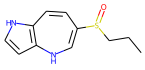
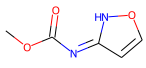
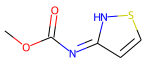
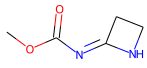
Seed molecule 1				Seed molecule 2			
							
No	Weight 1	Weight 2	Molecule 1	Molecule 2	Molecule 3	Molecule 4	Molecule 5
1	0.15	0.85					
2	0.25	0.75					
3	0.35	0.65					
4	0.45	0.55					
5	0.55	0.45					
6	0.65	0.35					
7	0.75	0.25					
8	0.85	0.15					

Table S6: Comparison of weights for the second pair of molecules with sampling reward 0.75

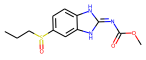
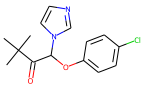
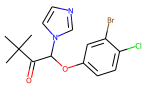
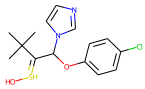
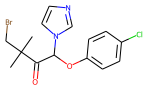
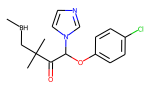
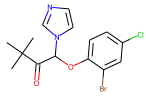
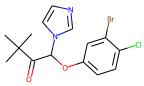
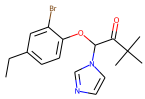
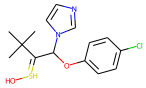
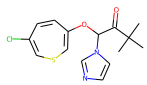
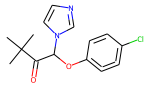
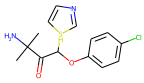
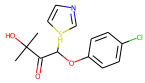
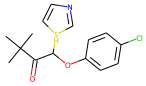
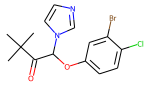
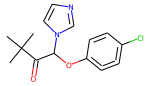
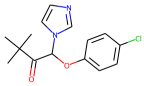
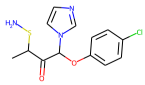
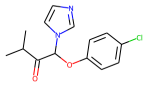
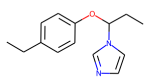
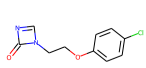
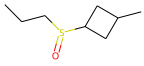
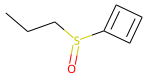
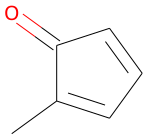
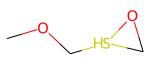
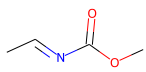
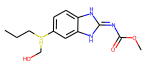
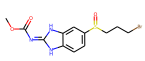
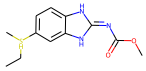
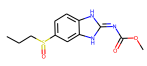
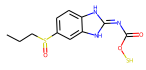
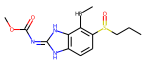
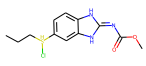
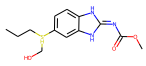
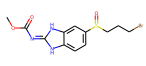
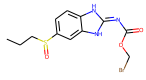
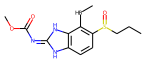
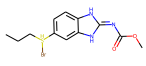
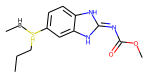
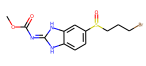
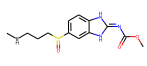
Seed molecule 1				Seed molecule 2			
							
No	Weight 1	Weight 2	Molecule 1	Molecule 2	Molecule 3	Molecule 4	Molecule 5
1	0.15	0.85					
2	0.25	0.75					
3	0.35	0.65					
4	0.45	0.55					
5	0.55	0.45					
6	0.65	0.35					
7	0.75	0.25					
8	0.85	0.15					

Table S7: Comparison of weights for the third pair of molecules with sampling reward 0.75

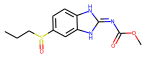
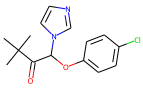
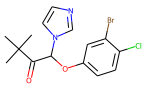
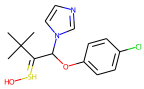
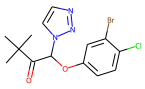
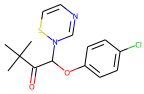
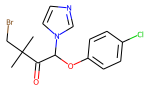
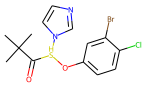
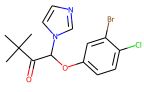
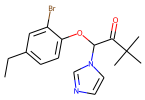
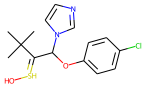
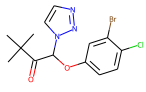
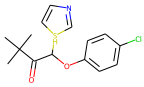
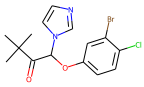
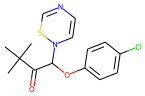
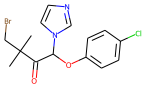
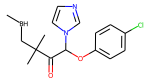
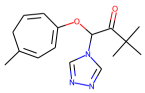
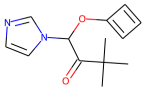
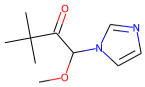
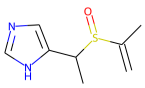
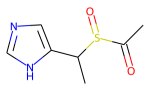
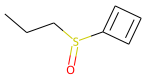
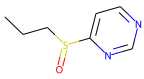
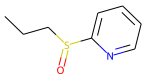
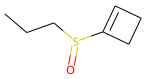
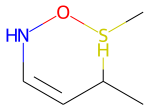
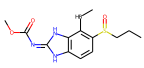
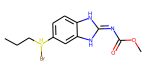
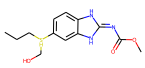
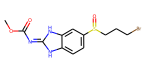
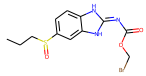
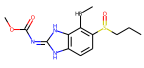
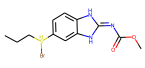
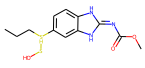
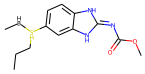
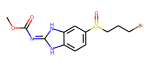
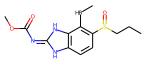
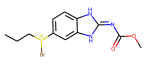
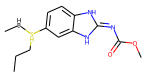
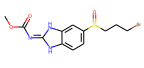
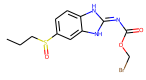
Seed molecule 1				Seed molecule 2			
							
No	Weight 1	Weight 2	Molecule 1	Molecule 2	Molecule 3	Molecule 4	Molecule 5
1	0.15	0.85					
2	0.25	0.75					
3	0.35	0.65					
4	0.45	0.55					
5	0.55	0.45					
6	0.65	0.35					
7	0.75	0.25					
8	0.85	0.15					

Table S8: Comparison of weights for the third pair of molecules with sampling reward 1.9

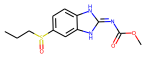
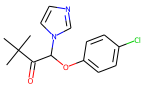
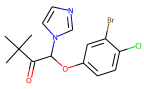
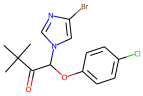
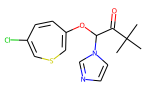
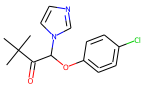
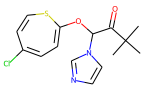
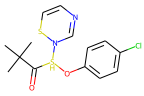
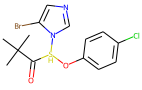
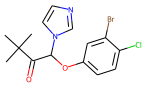
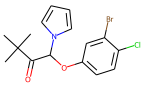
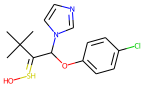
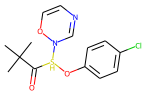
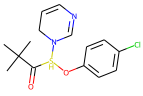
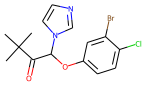
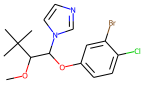
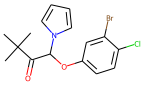
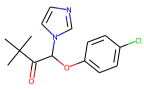
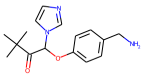
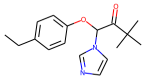
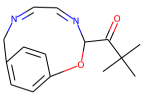
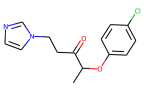
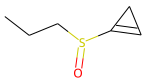
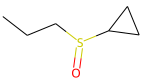
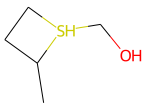
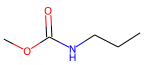
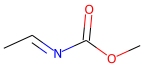
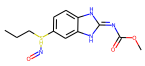
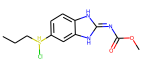
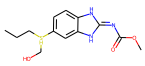
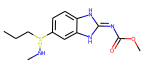
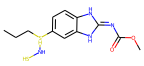
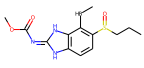
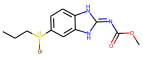
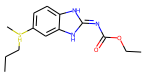
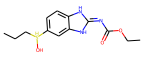
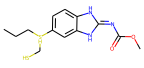
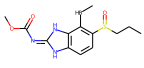
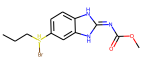
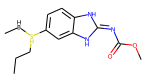
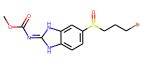
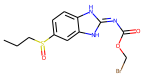
Seed molecule 1				Seed molecule 2			
							
No	Weight 1	Weight 2	Molecule 1	Molecule 2	Molecule 3	Molecule 4	Molecule 5
1	0.15	0.85					
2	0.25	0.75					
3	0.35	0.65					
4	0.45	0.55					
5	0.55	0.45					
6	0.65	0.35					
7	0.75	0.25					
8	0.85	0.15					

Table S9: Comparison of weights for the third pair of molecules with sampling reward 2.5

S4 List of vitamin D analogues

S4.1 Sample 1

S4.1.1 Training set

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CCO)CCO)c(C)c1
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cncn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
CC(CCCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(C=CCC(O)C12CC3CC(CC(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
COc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCCc2ccccc2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC1O
CC(C=CC(=O)Nc1ccccc1N)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
Cc1ccc2c(C(NCc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C1O
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCCOC4C3)CCCC21C
COc1ccc(C(Nc2ccccn2)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CC(CCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC1O
Cc1ccc2c(C(Nc3cccnc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)(C)CCC(C)(C)O)CC1O
CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C)[nH]c3ccccc23)cc1
COc1ccc(NC(c2nccs2)c2c(C)[nH]c3ccccc23)cc1OC
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(CCc2ccc(CO)c(CO)c2)c1
COc1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2ccccc2)CC1
Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl
Cc1[nH]c2ccccc2c1C(Nc1ccc([N+](=O)[O-])cc1)c1ccccc1Cl
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(O)(CC)CC)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
CCC(CC)(c1ccc(OCC(CO)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC

C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC1O
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCC)cc(OCC)c2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccc(Cl)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(CCCO)C1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCCO)C1O
CC[Si](CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC2CO2)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C2(C)CCCCC2)CC(O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(F)(F)C(O)(CC)CC)=CCC23)CC(O)CC1C(F)F
CS(=O)(=O)c1ccc(C(=O)NCCCC(=O)N2CCN(c3ccc(C(F)(F)F)cc3)CC2)cc1[N+](=O)[O-]
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(OCCO)C(O)C3)CCCC12C
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
COC(=O)c1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
C=C1C(O)CC(=CC=CCCCC(C)CCCC(C)(C)O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2C)CC(O)CC1O
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCCC1(C)C
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCc2ccccc2)c(C)c1
Cc1ccc2c(C(NCc3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC1O
CC(CCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C

S4.1.2 Testing set

The SMILES strings of the recreated VDR ligands are highlighted in yellow.

C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ncnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O
CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC1O

CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCF)C(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COc2ccc(C(C)(C)O)cc2)CC(O)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)OCC(C)C(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cnnn2)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2CC2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(CO)c(CO)c2)ccc1C
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)CC(O)C3=O)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C4(CCCC(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnc2)C1O
C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)C(=C)C(O)C4)CCCC32C)OC1=O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)C(O)CO)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCc2ccc(CO)c(CO)c2)ccc1CC
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCCO)CC(O)CC1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)CC(O)C4=C)CCCC32C)OC1=O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OCCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCO)C1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCCO)C1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCCO)C(O)C3)CCCC21C
CC(CCCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNS(=O)(=O)c1ccccc1)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1C(F)F
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)(C)C)CC(O)CC1O
COCOCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C1O
C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC1O

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
CC(C=CC(=O)NO)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CO)c(C)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(O)CC1O
CC(CCCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
CC(CCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCCC)cc(OCCCC)c2)CC(O)CC1O
COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccc(OC)cc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(OC)cc(OC)c2)CC(O)CC1O
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
Nc1nc2ccc(S(=O)(=O)N3CCCCC3)cc2s1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC1O
CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)C(=O)O)n1C
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(
    C)NC1=O
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1nccs1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CC(=O)N1CCN(C(=O)Cn2c(-c3ccccc3)csc2=N)CC1
CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COc1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
Cc1[nH]c2ccccc2c1C(Nc1ccccn1)c1ccc(C(C)C)cc1

S4.1.3 Seeds

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cncn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C1O
CC(CCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1

```
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccc(Cl)cc1
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl
Cc1ccc2c(C(NCc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1ccc2c(C(Nc3cccnc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC1O
CS(=O)(=O)c1ccc(C(=O)NCCCC(=O)N2CCN(c3ccc(C(F)(F)F)cc3)CC2)cc1[N+](=O)[O-]
COc1ccc(C(Nc2ccccn2)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(OCCO)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
```

S4.2 Sample 2

S4.2.1 Training set

```
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C1O
Cc1ccc2c(C(NCc3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)CC(O)C4=C)CCCC32C)OC1=O
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)C(O)CO)c(C)c1
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CCO)CCO)c(C)c1
COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC1O
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)CC1O
CC[Si](CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC2CO2)c(C)c1
CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OCCCO)C(O)C4=C)CCCC32C)C1C
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O
COCOCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cnnn2)C1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)C(=O)O)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C4(CCCC(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
```

CC(CCCCNS(=O)(=O)c1ccccc1)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC1O
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O
CCC(CC)(c1ccc(OCC(CO)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccc(Cl)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC1O
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(OC)cc(OC)c2)CC(O)CC1O
CC(CCCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(CCCO)C1O
C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccc([N+](=O)[O-])cc1)c1ccccc1Cl
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC1O
COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
Cc1ccc2c(C(Nc3cccnc3)c3ccccc3Cl)c(C)[nH]c2c1
CS(=O)(=O)c1ccc(C(=O)NCCCC(=O)N2CCN(c3ccc(C(F)(F)F)cc3)CC2)cc1[N+](=O)[O-]
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCCO)C1O
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C)[nH]c3ccccc23)cc1
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)CC(O)C3=O)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C1O
COc1ccc(NC(c2nccs2)c2c(C)[nH]c3ccccc23)cc1OC
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C1O
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
CC(C=CC(=O)NO)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccccn1)c1ccc(C(C)C)cc1
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2ccccc2)CC1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C2(C)CCCCC2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cncn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCCO)C1O
CC(C=CC(=O)Nc1ccccc1N)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(=O)N1CCN(C(=O)Cn2c(-c3ccccc3)csc2=N)CC1

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
CC(CCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC1O
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COc2ccc(C(C)(C)O)cc2)CC(O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl

S4.2.2 Testing set

The SMILES strings of the recreated VDR ligands are highlighted in yellow.

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ncnn2)C1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)(C)CCC(C)(C)O)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(OCCO)C(O)C3)CCCC12C
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C1O
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCF)C(O)C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)OCC(C)C(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2CC2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(O)(CC)CC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(CO)c(CO)c2)ccc1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnc2)C1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)C(=C)C(O)C4)CCCC32C)OC1=O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(CCc2ccc(CO)c(CO)c2)c1
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCc2ccc(CO)c(CO)c2)ccc1CC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1

C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCCO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCCO)C(O)C3)CCCC21C
CC(CCCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCCOC4C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1C(F)F
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)(C)C)CC(O)CC1O
CC(C=CCC(O)C12CC3CC(CC(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(F)(F)C(O)(CC)CC)=CCC23)CC(O)CC1C(F)F
C=C1C(O)CC(=CC=CCCCC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCC)cc(OCC)c2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CO)c(C)c1
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(O)CC1O
CC(CCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCCC1(C)C
CC(CCCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCCc2ccccc2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCCC)cc(OCCCC)c2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccc(OC)cc2)CC(O)CC1O
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
Nc1nc2ccc(S(=O)(=O)N3CCCCC3)cc2s1
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(
    C)NC1=O
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1nccs1
COc1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl
COc1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1

CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCc2ccccc2)c(C)c1
COc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COC(=O)c1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl
COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
Cc1ccc2c(C(NCc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC
COc1ccc(C(Nc2ccccn2)c2c(C)[nH]c3ccccc23)cc1

S4.2.3 Seeds

C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(O)CC1O
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CC(CCCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
CC(=O)N1CCN(C(=O)Cn2c(-c3ccccc3)csc2=N)CC1
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccc(Cl)cc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C1O
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C)[nH]c3ccccc23)cc1
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2ccccc2)CC1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)CC(O)C3=O)CCCC21C

S4.3 Sample 3

S4.3.1 Training set

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnn2)C1O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC1O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(CCc2ccc(CO)c(CO)c2)c1
COc1ccc(NC(c2nccs2)c2c(C)[nH]c3ccccc23)cc1OC
Nc1nc2ccc(S(=O)(=O)N3CCCCC3)cc2s1
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCCO)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O

C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)OCC(C)C(C)(C)O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cnnn2)C1O
CC(CCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(OCCO)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC1O
CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OCCCO)C(O)C4=C)CCCC32C)C1C
CC(CCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1nccs1
COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC
COCOCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O
COc1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2CC2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)(C)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cncn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCCC)cc(OCCCC)c2)CC(O)CC1O
CC(CCCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1
CC(C=CC(=O)Nc1ccccc1N)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C1O
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
Cc1ccc2c(C(Nc3cccnc3)c3ccccc3Cl)c(C)[nH]c2c1
COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
COc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(CCCO)C1O
CC(CCCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNS(=O)(=O)c1ccccc1)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O
CC(CCCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)(C)C)CC(O)CC1O

CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(CCCO)C1O
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(O)(CC)CC)CC1O
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1C(F)F
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCCO)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCF)C(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCCO)C(O)C3)CCCC21C
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
CC(CCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
CC(CCCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC1O

S4.3.2 Testing set

The SMILES strings of the recreated VDR ligands are highlighted in yellow.

C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ncnn2)C1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O
CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COc2ccc(C(C)(C)O)cc2)CC(O)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC1O
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(CO)c(CO)c2)ccc1C
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)CC(O)C3=O)CCCC21C

C=C1C(=CC=C2CCCC3(C)C(C4(CCCC(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnc2)C1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)C(=C)C(O)C4)CCCC32C)OC1=O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)C(O)CO)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
CCC(CC)(c1ccc(OCC(CO)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCc2ccc(CO)c(CO)c2)ccc1CC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCCO)CC(O)CC1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)CC(O)C4=C)CCCC32C)OC1=O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCO)C1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O
CC(CCCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CCO)CCO)c(C)c1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCCOC4C3)CCCC21C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
CC[Si](CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC2CO2)c(C)c1
CC(C=CCC(O)C12CC3CC(CC(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(F)(F)C(O)(CC)CC)=CCC23)CC(O)CC1C(F)F
C=C1C(O)CC(=CC=CCCCC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C1O
C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC1O
CC(C=CC(=O)NO)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCC)cc(OCC)c2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CO)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C2(C)CCCCC2)CC(O)CC1O
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCCC1(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCCc2ccccc2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccc(OC)cc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(OC)cc(OC)c2)CC(O)CC1O
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)C(=O)O)n1C
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2ccccc2)CC1

```
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(
    C)NC1=O
COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C)[nH]c3ccccc23)cc1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
Cc1[nH]c2ccccc2c1C(Nc1ccc([N+](=O)[O-])cc1)c1ccccc1Cl
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CC(=O)N1CCN(C(=O)Cn2c(-c3ccccc3)csc2=N)CC1
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl
COc1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
Cc1ccc2c(C(NCc3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCc2ccccc2)c(C)c1
CS(=O)(=O)c1ccc(C(=O)NCCCC(=O)N2CCN(c3ccc(C(F)(F)F)cc3)CC2)cc1[N+](=O)[O-]
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccc(Cl)cc1
COC(=O)c1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
Cc1[nH]c2ccccc2c1C(Nc1ccccn1)c1ccc(C(C)C)cc1
Cc1ccc2c(C(NCc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COc1ccc(C(Nc2ccccn2)c2c(C)[nH]c3ccccc23)cc1
```

S4.3.3 Seeds

```
CC(CCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
CC(CCCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
COc1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2CC2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1C(F)F
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
CC(CCCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1nccs1
```

C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1

S4.4 Sample 4

S4.4.1 Training set

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCC)cc(OCC)c2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCCO)C(O)C3)CCCC21C
CC(C=CCC(O)C12CC3CC(CC(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cncn2)C1O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1
C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)CC(O)C4=C)CCCC32C)OC1=O
CC(CCCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C2(C)CCCCC2)CC(O)CC1O
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cnnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(O)CC1O
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC1O
COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
Cc1ccc2c(C(NCc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl
CC(CCCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)OCC(C)C(C)(C)O)CC1O
CS(=O)(=O)c1ccc(C(=O)NCCCC(=O)N2CCN(c3ccc(C(F)(F)F)cc3)CC2)cc1[N+](=O)[O-]
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
CC(CCCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CO)c(C)c1
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
CC(CCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=CCCCC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCO)C1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CCO)CCO)c(C)c1
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC
CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
Cc1[nH]c2ccccc2c1C(Nc1ccccn1)c1ccc(C(C)C)cc1
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(OCCO)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C(C4(CCCC(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnc2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1C(F)F
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC1O
Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCCOC4C3)CCCC21C
CC(CCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COc2ccc(C(C)(C)O)cc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC1O
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(O)(CC)CC)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(CO)c(CO)c2)ccc1C
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
COc1ccc(NC(c2nccs2)c2c(C)[nH]c3ccccc23)cc1OC

S4.4.2 Testing set

The SMILES strings of the recreated VDR ligands are highlighted in yellow.

C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ncnn2)C1O

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O
CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)(C)CCC(C)(C)O)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC1O
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCF)C(O)C3)CCCC21C
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnn2)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)CC(O)C3=O)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)C(=C)C(O)C4)CCCC32C)OC1=O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)C(O)CO)c(C)c1
CCC(CC)(c1ccc(OCC(CO)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC1O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(CCc2ccc(CO)c(CO)c2)c1
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCc2ccc(CO)c(CO)c2)ccc1CC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCCO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OCCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
CC(CCCCNS(=O)(=O)c1ccccc1)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC[Si](CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC2CO2)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)(C)C)CC(O)CC1O
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(F)(F)C(O)(CC)CC)=CCC23)CC(O)CC1C(F)F

```
COCOCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC1O
CC(C=CC(=O)NO)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(C=CC(=O)Nc1ccccc1N)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(O)CC1O
CC(CCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCCC1(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCCc2ccccc2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCCC)cc(OCCCC)c2)CC(O)CC1O
COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccc(OC)cc2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
Nc1nc2ccc(S(=O)(=O)N3CCCCC3)cc2s1
COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)C(=O)O)n1C
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2ccccc2)CC1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(
    C)NC1=O
COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C)[nH]c3ccccc23)cc1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1nccs1
COc1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc([N+](=O)[O-])cc1)c1ccccc1Cl
CC(=O)N1CCN(C(=O)Cn2c(-c3ccccc3)csc2=N)CC1
CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
COc1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1ccc2c(C(NCc3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
Cc1ccc2c(C(Nc3cccnc3)c3ccccc3Cl)c(C)[nH]c2c1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCc2ccccc2)c(C)c1
COc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccc(Cl)cc1
COC(=O)c1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC
COc1ccc(C(Nc2ccccn2)c2c(C)[nH]c3ccccc23)cc1
```

S4.4.3 Seeds

```
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C1O
C=C1C(O)CC(=CC=CCCCC(C)CCCC(C)(C)O)CC1O
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
```

CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCCO)C(O)C3)CCCC21C
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(CO)c(CO)c2)ccc1C
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C1O
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
CC(CCCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C

S4.5 Sample 5

S4.5.1 Training set

CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCCO)CC(O)CC1O
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCCO)C(O)C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
Cc1ccc2c(C(Nc3cccnc3)c3ccccc3Cl)c(C)[nH]c2c1
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC1O
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cnnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1C(F)F
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O
CC(CCCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O
C=C1C(=CC=C2CCCC3(C)C(C4(CCCC(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC1O
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1

CC(CCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)C(O)CO)c(C)c1
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C1O
COc1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
COc1ccc(C(Nc2ccccn2)c2c(C)[nH]c3ccccc23)cc1
CC(=O)N1CCN(C(=O)Cn2c(-c3ccccc3)csc2=N)CC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCCO)C1O
CC(CCCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCc2ccc(CO)c(CO)c2)ccc1CC
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC1O
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCC)cc(OCC)c2)CC(O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCO)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OCCCO)C(O)C4=C)CCCC32C)C1C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)(C)C)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(O)(CC)CC)CC1O
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
CC(CCCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCCC)cc(OCCCC)c2)CC(O)CC1O
COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccc(Cl)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCCO)C1O
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCCC1(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(OC)cc(OC)c2)CC(O)CC1O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(CCc2ccc(CO)c(CO)c2)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CCO)CCO)c(C)c1
Cc1ccc2c(C(NCc3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(
    C)NC1=O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
Nc1nc2ccc(S(=O)(=O)N3CCCCC3)cc2s1

S4.5.2 Testing set

The SMILES strings of the recreated VDR ligands are highlighted in yellow.

C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ncnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)(C)CCC(C)(C)O)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(OCCO)C(O)C3)CCCC12C
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC1O
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCF)C(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COc2ccc(C(C)(C)O)cc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)OCC(C)C(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cncn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnn2)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(CO)c(CO)c2)ccc1C
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)CC(O)C3=O)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnc2)C1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)C(=C)C(O)C4)CCCC32C)OC1=O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(CC)(c1ccc(OCC(CO)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC1O
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1

C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)CC(O)C4=C)CCCC32C)OC1=O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2C)CC(O)CC1O
CC(CCCCNS(=O)(=O)c1ccccc1)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCCOC4C3)CCCC21C
CC[Si](CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC2CO2)c(C)c1
CC(C=CCC(O)C12CC3CC(CC(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(F)(F)C(O)(CC)CC)=CCC23)CC(O)CC1C(F)F
C=C1C(O)CC(=CC=CCCCC(C)CCCC(C)(C)O)CC1O
COCOCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C1O
CC(C=CC(=O)NO)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CO)c(C)c1
CC(C=CC(=O)Nc1ccccc1N)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
CC(CCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C2(C)CCCCC2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
CC(CCCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccc(OC)cc2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC1O
CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)C(=O)O)n1C
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2ccccc2)CC1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C)[nH]c3ccccc23)cc1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1nccs1
COc1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
Cc1[nH]c2ccccc2c1C(Nc1ccc([N+](=O)[O-])cc1)c1ccccc1Cl
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl

```
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCc2ccccc2)c(C)c1
CS(=O)(=O)c1ccc(C(=O)NCCCC(=O)N2CCN(c3ccc(C(F)(F)F)cc3)CC2)cc1[N+](=O)[O-]
COc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
COC(=O)c1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl
COc1ccc(NC(c2nccs2)c2c(C)[nH]c3ccccc23)cc1OC
Cc1[nH]c2ccccc2c1C(Nc1ccccn1)c1ccc(C(C)C)cc1
Cc1ccc2c(C(NCc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
```

S4.5.3 Seeds

```
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCO)C1O
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C(C4(CCCC(C)(C)O)CC4)=CCC23)CC(O)CC1O
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC1O
Cc1ccc2c(C(Nc3cccnc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OCCCO)C(O)C4=C)CCCC32C)C1C
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(
    C)NC1=O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)(C)C)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cnnn2)C1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
```

S4.6 Sample 6

S4.6.1 Training set

```
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC1O
CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
CCC(CC)(c1ccc(OCC(CO)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
```

```
Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1C(F)F
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCF)C(O)C3)CCCC21C
CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(CO)c(CO)c2)ccc1C
CC(CCCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
COc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCCc2ccccc2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C1O
CC(CCCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
Nc1nc2ccc(S(=O)(=O)N3CCCCC3)cc2s1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCO)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC1O
CC(CCCCNS(=O)(=O)c1ccccc1)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2C)CC(O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)C(=C)C(O)C4)CCCC32C)OC1=O
CC(CCCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC1O
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)(C)C)CC(O)CC1O
C=C1C(O)CC(=CC=CCCCC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23C)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCCOC4C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F
CC(CCCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCCC)cc(OCCCC)c2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC1O
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
```

CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCCO)C(O)C3)CCCC21C
CC(CCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnc2)C1O
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
Cc1ccc2c(C(NCc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C1O
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OCCCO)C(O)C4=C)CCCC32C)C1C
CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccc(OC)cc2)CC(O)CC1O
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC1O
Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(CCc2ccc(CO)c(CO)c2)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COC(=O)c1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCC)cc(OCC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cnnn2)C1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
COc1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C

S4.6.2 Testing set

The SMILES strings of the recreated VDR ligands are highlighted in yellow.

C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ncnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)(C)CCC(C)(C)O)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(OCCO)C(O)C3)CCCC12C
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C1O
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COc2ccc(C(C)(C)O)cc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC1O

CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)OCC(C)C(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cncn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2CC2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(O)(CC)CC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)CC(O)C3=O)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C4(CCCC(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)C(O)CO)c(C)c1
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCc2ccc(CO)c(CO)c2)ccc1CC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCCO)CC(O)CC1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)CC(O)C4=C)CCCC32C)OC1=O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CCO)CCO)c(C)c1
CC(CCCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC[Si](CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC2CO2)c(C)c1
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
CC(C=CCC(O)C12CC3CC(CC(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(F)(F)C(O)(CC)CC)=CCC23)CC(O)CC1C(F)F
COCOCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C1O
C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
CC(C=CC(=O)NO)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CO)c(C)c1
CC(C=CC(=O)Nc1ccccc1N)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(O)CC1O
CC(CCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C2(C)CCCCC2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
CC(CCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCCC1(C)C
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C

```
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC1O
COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)C(=O)O)n1C
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2ccccc2)CC1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(
    C)NC1=O
COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C)[nH]c3ccccc23)cc1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1nccs1
COc1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
Cc1[nH]c2ccccc2c1C(Nc1ccc([N+](=O)[O-])cc1)c1ccccc1Cl
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CC(=O)N1CCN(C(=O)Cn2c(-c3ccccc3)csc2=N)CC1
Cc1ccc2c(C(NCc3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1ccc2c(C(Nc3cccnc3)c3ccccc3Cl)c(C)[nH]c2c1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCc2ccccc2)c(C)c1
CS(=O)(=O)c1ccc(C(=O)NCCCC(=O)N2CCN(c3ccc(C(F)(F)F)cc3)CC2)cc1[N+](=O)[O-]
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccc(Cl)cc1
COc1ccc(NC(c2nccs2)c2c(C)[nH]c3ccccc23)cc1OC
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
Cc1[nH]c2ccccc2c1C(Nc1ccccn1)c1ccc(C(C)C)cc1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC
COc1ccc(C(Nc2ccccn2)c2c(C)[nH]c3ccccc23)cc1
```

S4.6.3 Seeds

```
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COc1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
CCC(CC)(c1ccc(OCC(CO)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(CO)c(CO)c2)ccc1C
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
CC(CCCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(OC)cc(OC)c2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
```

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC1O
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1C(F)F
CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCC)cc(OCC)c2)CC(O)CC1O

S4.7 Sample 7

S4.7.1 Training set

C=C1C(O)CC(=CC=CCCCC(C)CCCC(C)(C)O)CC1O
CC(CCCCNS(=O)(=O)c1ccccc1)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1nccs1
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(CCc2ccc(CO)c(CO)c2)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)C(=O)O)n1C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC1O
Cc1ccc2c(C(NCc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)OCC(C)C(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)(C)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCCO)CC(O)CC1O
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)C(=C)C(O)C4)CCCC32C)OC1=O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCCO)C(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCCOC4C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC
CC(CCCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
COc1ccc(C(Nc2ccccn2)c2c(C)[nH]c3ccccc23)cc1
COC(=O)c1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCF)C(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(CCCO)C1O
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C1O
CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC1O
CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC1O
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O

Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccc(Cl)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC1O
Cc1ccc2c(C(Nc3cccnc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O
COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
COCOCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cncn2)C1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC1O
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC1O
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCCC1(C)C
Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(O)CC1O
CC(CCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OCCCO)C(O)C4=C)CCCC32C)C1C
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC1O
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)CC(O)C3=O)CCCC21C
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O

S4.7.2 Testing set

The SMILES strings of the recreated VDR ligands are highlighted in yellow.

C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ncnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(OCCO)C(O)C3)CCCC12C
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COc2ccc(C(C)(C)O)cc2)CC(O)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnn2)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cnnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2CC2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(O)(CC)CC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(CO)c(CO)c2)ccc1C
C=C1C(=CC=C2CCCC3(C)C(C4(CCCC(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnc2)C1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)C(O)CO)c(C)c1
CCC(CC)(c1ccc(OCC(CO)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCc2ccc(CO)c(CO)c2)ccc1CC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)CC(O)C4=C)CCCC32C)OC1=O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCCO)C1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CCO)CCO)c(C)c1
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1C(F)F
CC(CCCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C

CC[Si](CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC2CO2)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)(C)C)CC(O)CC1O
CC(C=CCC(O)C12CC3CC(CC(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(F)(F)C(O)(CC)CC)=CCC23)CC(O)CC1C(F)F
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C1O
C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
CC(C=CC(=O)NO)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCC)cc(OCC)c2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CO)c(C)c1
CC(C=CC(=O)Nc1ccccc1N)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
CC(CCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C2(C)CCCCC2)CC(O)CC1O
CC(CCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
CC(CCCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCCC)cc(OCCCC)c2)CC(O)CC1O
COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccc(OC)cc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(O)CC1O
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
Nc1nc2ccc(S(=O)(=O)N3CCCCC3)cc2s1
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2ccccc2)CC1
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(
    C)NC1=O
COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C)[nH]c3ccccc23)cc1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
COc1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
Cc1[nH]c2ccccc2c1C(Nc1ccc([N+](=O)[O-])cc1)c1ccccc1Cl
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CC(=O)N1CCN(C(=O)Cn2c(-c3ccccc3)csc2=N)CC1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl
COc1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
Cc1ccc2c(C(NCc3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCc2ccccc2)c(C)c1
CS(=O)(=O)c1ccc(C(=O)NCCCC(=O)N2CCN(c3ccc(C(F)(F)F)cc3)CC2)cc1[N+](=O)[O-]
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
COc1ccc(NC(c2nccs2)c2c(C)[nH]c3ccccc23)cc1OC
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
Cc1[nH]c2ccccc2c1C(Nc1ccccn1)c1ccc(C(C)C)cc1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC

S4.7.3 Seeds

```
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cncn2)C1O
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCCO)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC1O
Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
COCOCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCCC1(C)C
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(CCc2ccc(CO)c(CO)c2)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O
```

S4.8 Sample 8

S4.8.1 Training set

```
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(
    C)NC1=O
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC1O
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O
Cc1ccc2c(C(NCc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CO)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC1O
Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)C(=C)C(O)C4)CCCC32C)OC1=O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(CCc2ccc(CO)c(CO)c2)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CCO)CCO)c(C)c1
```

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ncnn2)C1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC1O
COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C)[nH]c3ccccc23)cc1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)(C)C)CC(O)CC1O
COc1ccc(C(Nc2ccccn2)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnc2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCF)C(O)C3)CCCC21C
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OCCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCCO)C1O
C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccccn1)c1ccc(C(C)C)cc1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC1O
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1C(F)F
CC(CCCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCCOC4C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)(C)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
Cc1[nH]c2ccccc2c1C(Nc1ccc([N+](=O)[O-])cc1)c1ccccc1Cl
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C1O
COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCCO)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCCO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(OC)cc(OC)c2)CC(O)CC1O
CC(=O)N1CCN(C(=O)Cn2c(-c3ccccc3)csc2=N)CC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F

Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(OCCO)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCCO)C(O)C3)CCCC21C
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCCC)cc(OCCCC)c2)CC(O)CC1O
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
Cc1ccc2c(C(NCc3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
CC(C=CCC(O)C12CC3CC(CC(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)C(O)CO)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCC)cc(OCC)c2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
COc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)C(=O)O)n1C
C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O

S4.8.2 Testing set

The SMILES strings of the recreated VDR ligands are highlighted in yellow.

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COc2ccc(C(C)(C)O)cc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)OCC(C)C(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cncn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnn2)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cnnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(O)(CC)CC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(CO)c(CO)c2)ccc1C
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)CC(O)C3=O)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C4(CCCC(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
CCC(CC)(c1ccc(OCC(CO)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCc2ccc(CO)c(CO)c2)ccc1CC
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)CC(O)C4=C)CCCC32C)OC1=O

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
CC(CCCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNS(=O)(=O)c1ccccc1)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
CC(CCCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC[Si](CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC2CO2)c(C)c1
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(F)(F)C(O)(CC)CC)=CCC23)CC(O)CC1C(F)F
C=C1C(O)CC(=CC=CCCCC(C)CCCC(C)(C)O)CC1O
COCOCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
CC(C=CC(=O)NO)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(C=CC(=O)Nc1ccccc1N)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(O)CC1O
CC(CCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C2(C)CCCCC2)CC(O)CC1O
CC(CCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCCC1(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
CC(CCCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCCc2ccccc2)CC(O)CC1O
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccc(OC)cc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
Nc1nc2ccc(S(=O)(=O)N3CCCCC3)cc2s1
CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2ccccc2)CC1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1nccs1

COc1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COc1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
Cc1ccc2c(C(Nc3cccnc3)c3ccccc3Cl)c(C)[nH]c2c1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCc2ccccc2)c(C)c1
CS(=O)(=O)c1ccc(C(=O)NCCCC(=O)N2CCN(c3ccc(C(F)(F)F)cc3)CC2)cc1[N+](=O)[O-]
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccc(Cl)cc1
COC(=O)c1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl
COc1ccc(NC(c2nccs2)c2c(C)[nH]c3ccccc23)cc1OC
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC

S4.8.3 Seeds

CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)C(O)CO)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CO)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
COc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC1O
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OCCCO)C(O)C4=C)CCCC32C)C1C
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(CCc2ccc(CO)c(CO)c2)c1
Cc1[nH]c2ccccc2c1C(Nc1ccccn1)c1ccc(C(C)C)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O

S4.9 Sample 9

S4.9.1 Training set

COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC1O
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1nccs1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C1O
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(OCCO)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
CC(CCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2CC2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)C(=O)O)n1C
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1cccn1)c1ccc(C(C)C)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCCO)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)C(O)CO)c(C)c1
CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCC)cc(OCC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCCC)cc(OCCCC)c2)CC(O)CC1O
CC(CCCCNS(=O)(=O)c1ccccc1)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)OCC(C)C(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCCO)C1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(CO)c(CO)c2)ccc1C
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
CC(C=CCC(O)C12CC3CC(CC(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCCOC4C3)CCCC21C
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CO)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCCO)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O
CC(C=CC(=O)Nc1ccccc1N)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnn2)C1O
Cc1[nH]c2ccccc2c1C(Nc1ccc([N+](=O)[O-])cc1)c1ccccc1Cl
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
Nc1nc2ccc(S(=O)(=O)N3CCCCC3)cc2s1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(CO)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(CCCO)C1O
COc1ccc(NC(c2nccs2)c2c(C)[nH]c3ccccc23)cc1OC
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
C=C1C(O)CC(=CC=CCCCC(C)CCCC(C)(C)O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1

CC(CCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cncn2)C1O
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC1O
COc1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(CCCO)C1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
CC(C=CC(=O)NO)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC1O
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC1O
Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
Cc1ccc2c(C(Nc3cccnc3)c3ccccc3Cl)c(C)[nH]c2c1
CC[Si](CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC2CO2)c(C)c1
CC(CCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2(C)C)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cnnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
COc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnc2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ncnn2)C1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC1O
Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl

S4.9.2 Testing set

The SMILES strings of the recreated VDR ligands are highlighted in yellow.

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O
CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)(C)CCC(C)(C)O)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCF)C(O)C3)CCCC21C

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COc2ccc(C(C)(C)O)cc2)CC(O)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC1O
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(O)(CC)CC)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)CC(O)C3=O)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C4(CCCC(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)C(=C)C(O)C4)CCCC32C)OC1=O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(CCc2ccc(CO)c(CO)c2)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCc2ccc(CO)c(CO)c2)ccc1CC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)CC(O)C4=C)CCCC32C)OC1=O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OCCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCCO)C(O)C3)CCCC21C
CC(CCCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CCO)CCO)c(C)c1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1C(F)F
CC(CCCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)(C)C)CC(O)CC1O
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(F)(F)C(O)(CC)CC)=CCC23)CC(O)CC1C(F)F
COCOCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C1O
C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
CC(CCCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C2(C)CCCCC2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
CC(CCCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCCC1(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
CC(CCCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCCc2ccccc2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccc(OC)cc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(OC)cc(OC)c2)CC(O)CC1O
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2ccccc2)CC1
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(
    C)NC1=O
COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C)[nH]c3ccccc23)cc1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
COc1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
CC(=O)N1CCN(C(=O)Cn2c(-c3ccccc3)csc2=N)CC1
CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1ccc2c(C(NCc3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCc2ccccc2)c(C)c1
CS(=O)(=O)c1ccc(C(=O)NCCCC(=O)N2CCN(c3ccc(C(F)(F)F)cc3)CC2)cc1[N+](=O)[O-]
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccc(Cl)cc1
COC(=O)c1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
Cc1ccc2c(C(NCc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COc1ccc(C(Nc2ccccn2)c2c(C)[nH]c3ccccc23)cc1

S4.9.3 Seeds

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ncnn2)C1O
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C1O
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC
COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCCO)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC1O

```
CC(CCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cnnn2)C1O
CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C1O
CC(C=CC(=O)NO)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCCC)cc(OCCCC)c2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(CC)(c1ccc(OCC(CO)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
```

S4.10 Sample 10

S4.10.1 Training set

```
Cc1[nH]c2ccccc2c1C(Nc1ccccn1)c1ccc(C(C)C)cc1
CC(C=CCC(O)C12CC3CC(CC(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1C(F)F
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2CC2)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O
Nc1nc2ccc(S(=O)(=O)N3CCCCC3)cc2s1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCc2ccccc2)c(C)c1
CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C2(C)CCCCC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(F)(F)C(O)(CC)CC)=CCC23)CC(O)CC1C(F)F
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cncn2)C1O
CC(CCCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(O)(CC)CC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccc(Cl)cc1
CC[Si](CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC2CO2)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2(C)C)CC(O)CC1O
COc1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
CC(CCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCCO)C1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
```

```
Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC1O
CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)C(O)CO)c(C)c1
CC(C=CC(=O)NO)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC1O
Cc1ccc2c(C(NCc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(
    C)NC1=O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O
CC(CCCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2cnnn2)C1O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCCO)C(O)C3)CCCC21C
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC1O
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCCC)cc(OCCCC)c2)CC(O)CC1O
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2ccccc2)CC1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CCO)CCO)c(C)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C1O
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCCOC4C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OCCCO)C(O)C4=C)CCCC32C)C1C
```

CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C
Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CC(C=CC(=O)Nc1ccccc1N)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)(C)C)CC(O)CC1O

S4.10.2 Testing set

The SMILES strings of the recreated VDR ligands are highlighted in yellow.

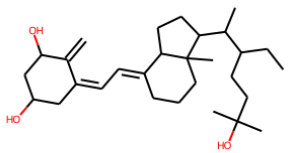
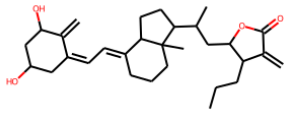
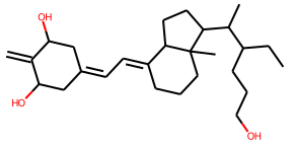
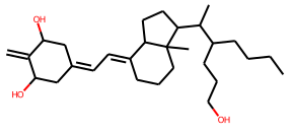
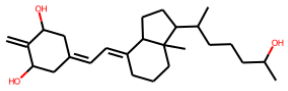
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ncnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)(C)CCC(C)(C)O)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(OCCO)C(O)C3)CCCC12C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3C(CCCCC(C)(C)O)CCC23)CC1O
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCF)C(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COc2ccc(C(C)(C)O)cc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)OCC(C)C(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnn2)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(CO)c(CO)c2)ccc1C
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)CC(O)C3=O)CCCC21C
C=C1C(=CC=C2CCCC3(C)C(C4(CCCC(C)(C)O)CC4)=CCC23)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ccnc2)C1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)C(=C)C(O)C4)CCCC32C)OC1=O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(CC)(c1ccc(OCC(CO)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC1O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(CCc2ccc(CO)c(CO)c2)c1
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCc2ccc(CO)c(CO)c2)ccc1CC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCCO)CC(O)CC1O
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)CC(O)C4=C)CCCC32C)OC1=O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC1O
C=C1C(=CC=C2CCC(CCCCCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC1O

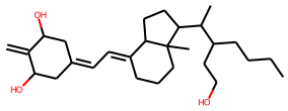
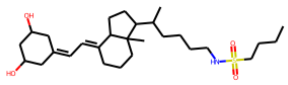
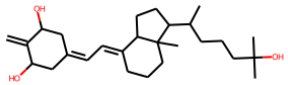
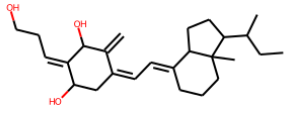
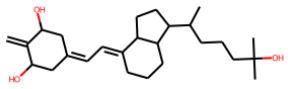
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
CC(CCCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCNS(=O)(=O)c1ccccc1)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
C=C1C(O)CC(=CC=CCCCC(C)CCCC(C)(C)O)CC1O
COCOCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C1O
C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCC)cc(OCC)c2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CO)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(O)CC1O
CC(CCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCCC1(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
CC(CCCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCCc2ccccc2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccc(OC)cc2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC1O
CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)C(=O)O)n1C
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC
COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C)[nH]c3ccccc23)cc1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1nccs1
COc1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
Cc1[nH]c2ccccc2c1C(Nc1ccc([N+](=O)[O-])cc1)c1ccccc1Cl
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CC(=O)N1CCN(C(=O)Cn2c(-c3ccccc3)csc2=N)CC1
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl
CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
Cc1ccc2c(C(NCc3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
Cc1ccc2c(C(Nc3cccnc3)c3ccccc3Cl)c(C)[nH]c2c1
CS(=O)(=O)c1ccc(C(=O)NCCCC(=O)N2CCN(c3ccc(C(F)(F)F)cc3)CC2)cc1[N+](=O)[O-]
COc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COC(=O)c1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl
COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
COc1ccc(NC(c2nccs2)c2c(C)[nH]c3ccccc23)cc1OC

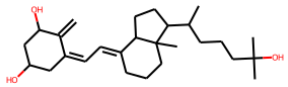
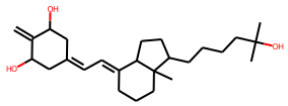
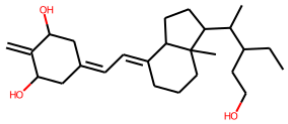
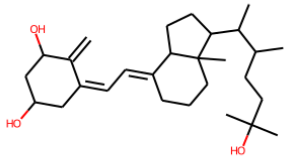
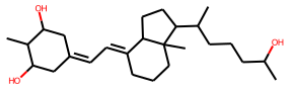
```
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC
COc1ccc(C(Nc2ccccn2)c2c(C)[nH]c3ccccc23)cc1
```

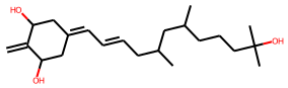
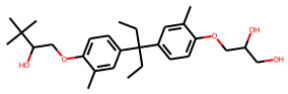
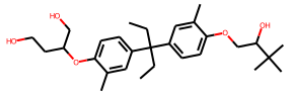
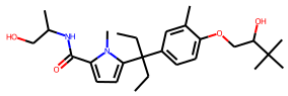
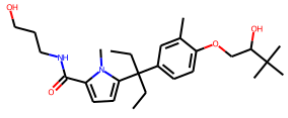
S4.10.3 Seeds

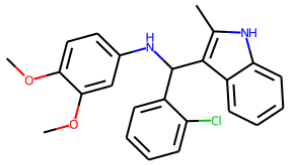
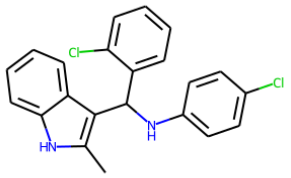
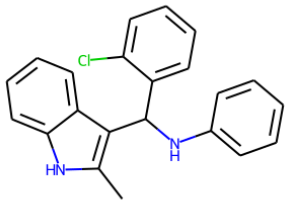
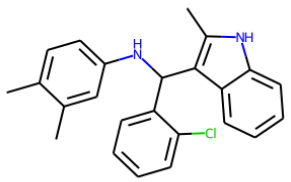
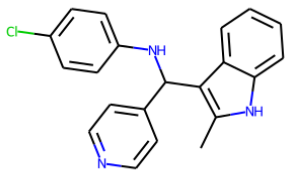
```
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(
    C)NC1=O
Cc1ccc2c(C(NCc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)C(O)CO)c(C)c1
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCCO)C1O
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(F)(F)C(O)(CC)CC)=CCC23)CC(O)CC1C(F)F
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C(C)(C)C)CC(O)CC1O
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCCO)C1O
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1
Cc1[nH]c2ccccc2c1C(Nc1ccccn1)c1ccc(C(C)C)cc1
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CCO)CCO)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O
Nc1nc2ccc(S(=O)(=O)N3CCCCC3)cc2s1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
```

SMILES	Structure
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O	
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O	
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O	
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O	
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O	

C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC1O	
CCCCS(=O)(=O)NCCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C	
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O	
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCCO)C1O	
C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O	

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O	
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O	
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O	
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O	
CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C	

C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O	
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1	
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1	
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C	
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C	

COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC	
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl	
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl	
Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C	
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1	

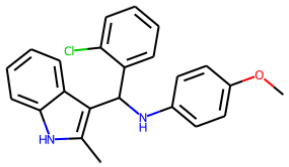
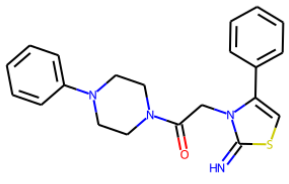
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1	
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1	

Table S10: A list of all recreated VDR ligands

References

[1] C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug. Deliver. Rev.*, 1997, **23**, 3.

[2] A. K. Ghose, V. N. Viswanadhan and J. J. Wendoloski, *J. Comb. Chem.*, 1999, **1**, 55.

[3] D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward and K. D. Kopple, *J. Med. Chem.*, 2002, **45**, 2615.

[4] M. Congreve, R. Carr, C. Murray and H. Jhoti, *Drug Discov. Today*, 2003, **8**, 876.

[5] W. P. Walters and M. Namchuk, *Nat. Rev. Drug Discovery*, 2003, **2**, 259.