

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.

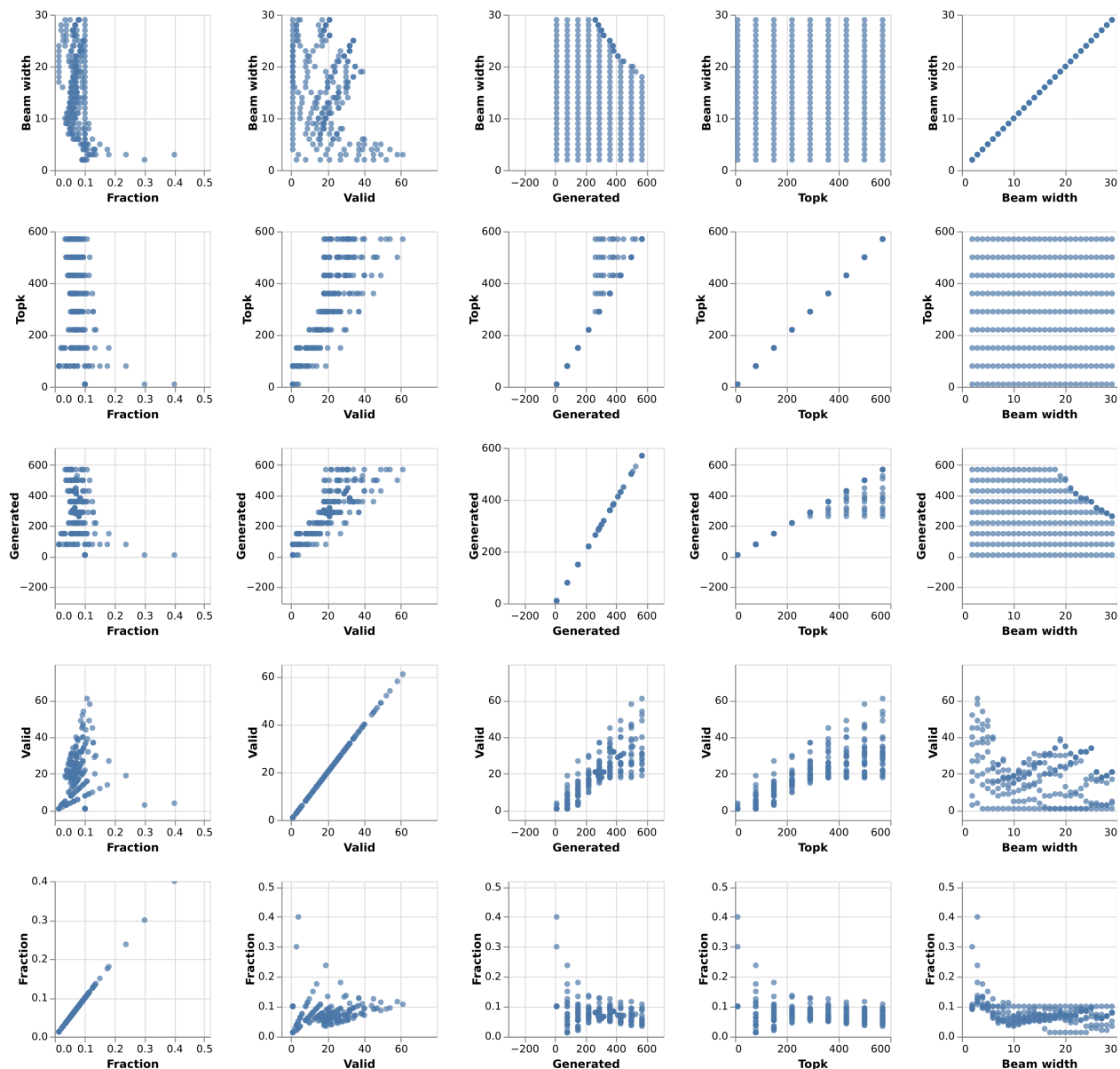
| Filter | Constraints |
|--|---|
| Lipinski rule of 5 ¹ (at most 1 violation is allowed) | Molecular weight ≤ 500 LogP ≤ 5.0 H-bond donor count ≤ 5 H-bond acceptor count ≤ 10 |
| Ghose ² | Molecular weight between 160 and 480 LogP between -0.4 and +5.6 Atom count between 20 and 70 Molar refractivity 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 be 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

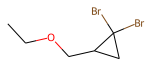
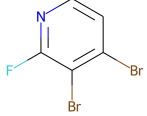
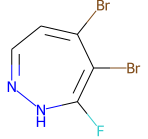
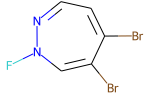
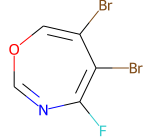
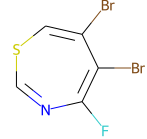
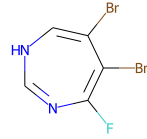
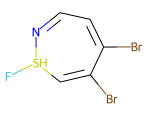
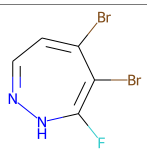
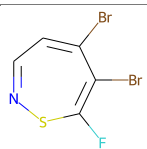
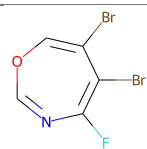
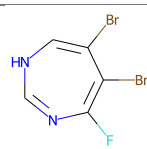
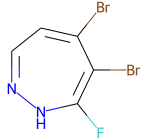
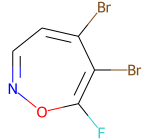
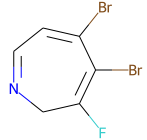
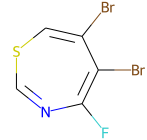
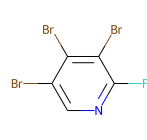
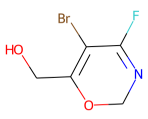
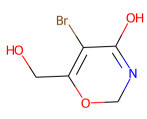
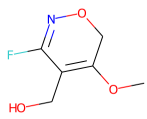
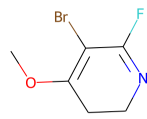
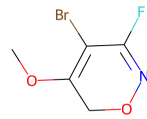
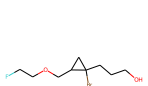
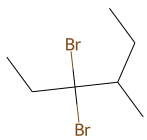
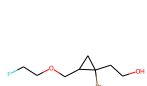
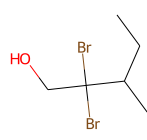
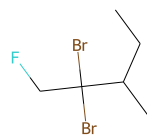
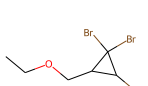
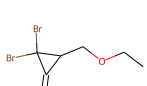
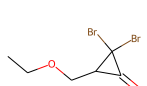
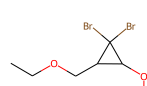
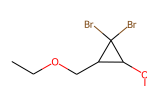
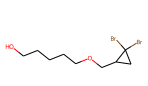

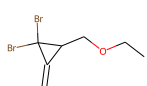

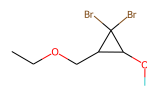
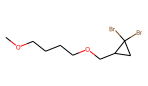
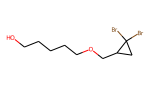
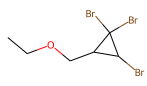
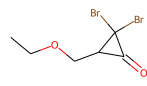
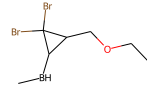
| Seed molecule 1 | | | Seed molecule 2 | | | | |
|---|----------|----------|---|---|--|---|---|
|  | | |  | | | | |
| No | Weight 1 | Weight 2 | Molecule 1 | Molecule 2 | Molecule 3 | Molecule 4 | Molecule 5 |
| 1 | 0.15 | 0.85 |  |  |  |  |  |
| 2 | 0.25 | 0.75 |  |  |  |  |  |
| 3 | 0.35 | 0.65 |  |  |  |  |  |
| 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

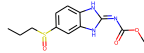
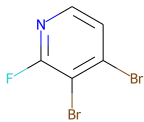
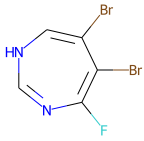
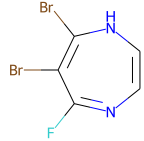
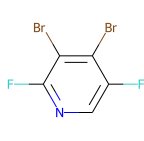
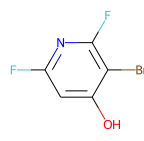
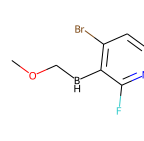
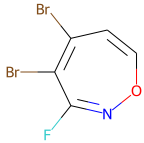
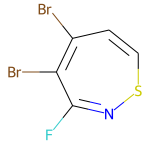
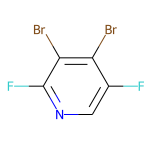
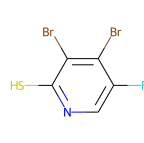
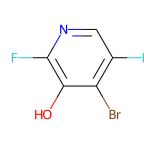
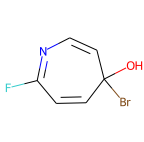
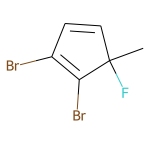
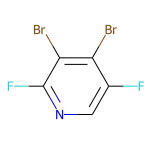
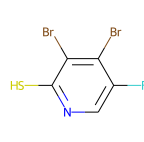
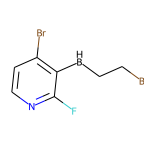
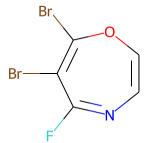
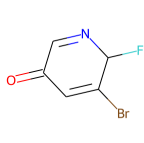
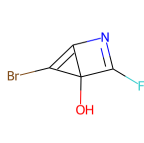
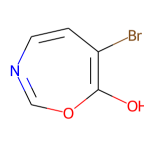
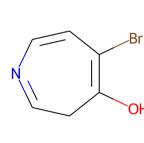
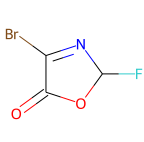
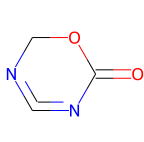
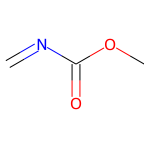
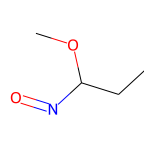
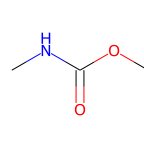
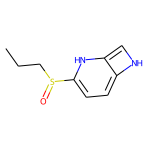
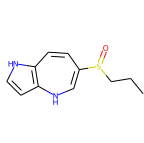
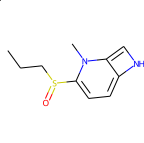
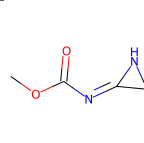
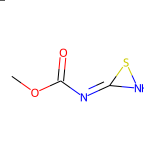
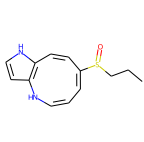
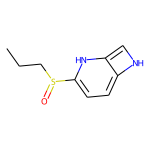
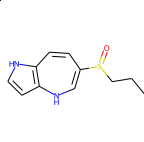
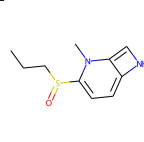
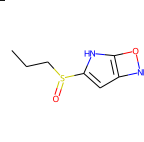
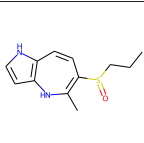
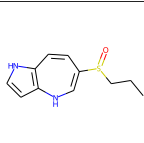
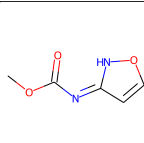
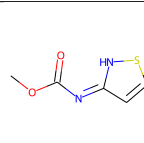
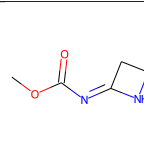
| Seed molecule 1 | | | Seed molecule 2 | | | | |
|---|----------|----------|---|---|--|---|---|
|  | | |  | | | | |
| No | Weight 1 | Weight 2 | Molecule 1 | Molecule 2 | Molecule 3 | Molecule 4 | Molecule 5 |
| 1 | 0.15 | 0.85 |  |  |  |  |  |
| 2 | 0.25 | 0.75 |  |  |  |  |  |
| 3 | 0.35 | 0.65 |  |  |  |  |  |
| 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(OCC(O)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(Cc2cncn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
CC(CCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(C=CCC(O)C12CCC3CC(C(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCn1c(C(=O)NC(O)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(=CCO)C1O
C=C1C(=CC=C2CCC(C)CCCC(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(NC3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc2cncn2)C1O
N=c1scc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCOC4C3)CCCC21C
COc1ccc(C(Nc2ccccc2)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)CCO)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(Nc3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(C(C4)C2)C3)CC1O
C=C1C(O)CC(=CC=C2CCCC3C(C)CCCC(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)NCCCC(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(NC1ccccc1)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(O)CCO)n1C
C=C1C(O)CC(=CC=CCC(C)(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
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(OC)c(O)CC(O)CC1O
CCC(CC)(c1ccc(OC(=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]c2cccc2c1C(Nc1cccc1)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=CC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
CCC(CC)(c1ccc(OC(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(=CCO)C1O
CC[Si](CC)(c1ccc(OC(=O)C(C)(C)C)c(C)c1)c1ccc(OC2C02)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C0CC(=C)C2(C)CCCC2)CC(O)CC1O
Cc1[nH]c2cccc2c1C(Nc1ccc(N(C)C)cc1)c1cccc1Cl
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
CC(CCCC(=O)Nc1cccc1N)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)=CC23)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=CC2C(=CC=C3CC(O)C(OC(O)C(O)C3)CCCC12C
Cc1ccc(CNC(c2cccc2Cl)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)=CC23)CC1O
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3cccc3Cl)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(c2cccc2Cl)c2c(C)[nH]c3cccc23)cc1
COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3cccc23)cc1OC
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
C=C1C(O)CC(=CC=CCCC(C)CCCC(C)(C)O)CC1O
CCC(CC)(c1ccc(OC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CO)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)N2Cc2cccc2)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)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=CCCC1(C)C
CCC(CC)(c1ccc(OC(O)CO)c(C)c1)c1ccc(OC(O)C(C)(C)CCc2cccc2)c(C)c1
Cc1ccc2c(C(NC3cccc3)c3cccc3Cl)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)COCc2cccc2)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(Cc2n2ncm2)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=CC2C(=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)O)c2ccc(C(C)(C)O)c2)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(CCCC(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(Cc2c2c2)C1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc2c2c2)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(=CCO)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)c1c1cc(Cc2ccc(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(Cc2c2c2)C1O
 C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(O)CC1O
 C=C1CC(C(C)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(C(C4)C2)C3)CC1O
 CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)C(O)C)C(C)c1
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(C(C4)C2)C3)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
 CCCc1cc(C(O)(CC)CC)c1c1cc(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(C(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)c1c1cc(C(C)(C)C)c(C)c1
 C=C1C(=CC=C2CCC(C)CCCC(C)(C)O)C3(C)C2CCC3C(C)CCCC)CC(O)CC1O
 C=C1CC(C(C)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(C(C)C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
 C=C1C(=O)OC(C(C)C)C2CCC3C(=CC=C4CC(O)C(O)C4=C)CCCC32C)C1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(O)C(O)C1O
 C=C1C(=O)OC(C(C)C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCCC)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(O)C(O)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(O)C(O)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)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(O)C(O)C1O
 CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC21C
 CC(CCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CC(CCCNS(=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)c1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COC#CC(C)(C)C)CC(O)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OC=C2C1O
 C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC1O

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C0CC(=C)C(C)(C)CC(0)CC10
 CC(C=CC(=0)N0)=CC(C)C1CCC2C(=CC=C3CC(0)CC(0)C3)CCCC21C
 CCC(CC)(c1ccc(OCC(=0)C(C)(C)C)c(C)c1)c1ccc(OCC(0)C0)c(C)c1
CCC(CC)(c1ccc(OCC(0)C(C)(C)C)c(C)c1)c1ccc(C(=0)NCCCC)n1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(0)C(=0)N2Cc2cc(OCCC)cc(OCCC)c2)CC(0)CC10
 CC(CCCCC(=0)Nc1cccc1N)C1CCC2C(=CC=C3CC(0)CC(0)C3)CCCC21C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(0)C(=0)N2Cc2ccccc2)CC(0)CC10
 CC(CCCNS(C)(=0)=0)C1CCC2C(=CC=C3CC(0)CC(0)C3)CCCC21C
 CC(CCCCC(=0)N0)C1CCC2C(=CC=C3CC(0)CC(0)C3)CCCC21C
 CC(CCCNS(N)(=0)=0)C1CCC2C(=CC=C3CC(0)CC(0)C3)CCCC21C
 CC(CCCNS(C)(=0)=0)C1CCC2C(=CC=C3CC(0)CC(0)C3)CCCC21C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(0)C(=0)N2Cc2cc(OCCC)cc(OCCC)c2)CC(0)CC10
 C0c1ccc(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)(0)C(=0)N2Cc2ccc(OC)cc2)CC(0)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(0)C(=0)N2CCc2cc(OC)cc(OC)c2)CC(0)CC10
 CCC(CC)(c1ccc(OCC(=0)C(C)(C)C)c(C)c1)c1ccc(C(=0)NCCC(=0)OC)n1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(0)C(=0)N2CCc2cc(OC)cc(OC)c2)CC(0)CC10
 Cc1ccc(-c2csc(=N)n2CC(=0)N2CCN(c3ccccc3)CC2)cc1
 CCC(CC)(c1ccc(OCC(0)C(C)(C)C)c(C)c1)c1ccc(C(=0)NCCC(=0)O)n1C
 CCn1c(C(=0)NCC(=0)OC)ccc1C(CC)(CC)c1ccc(OCC(=0)C(C)(C)C)c(C)c1
 Nc1nc2ccc(S(=0)(=0)N3CCCC3)cc2s1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(0)C(=0)N2CCc2ccccc2)CC(0)CC10
 CC(C)S(=0)(=0)c1oc(-c2cccs2)nc1S(=0)(=0)c1ccccc1
 CCC(CC)(c1ccc(OCC(=0)C(C)(C)C)c(C)c1)c1ccc(C(=0)NC(C)(=0)O)n1C
C0c1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc10C
 Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccnc1
 CC(C)CC(N)C(=0)NC(CC(C)C)C(=0)NC1C0CC(0)C(0)C0CCC(C(=0)O)NC(=0)C(CC(C)C)NC(=0)C(CC(C)C)NC(=0)C(C)(C)NC1=0
 C0C(=0)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
 CCC(CC)(c1ccc(OCC(0)C0)c(C)c1)c1ccc(OCC(0)C(C)(C)CCCC2ccccc2)c(C)c1
 Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccnc1
 Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1nccs1
 Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
 C0c1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc10C
 CC(=0)N1CCN(C(=0)Cn2c(-c3ccccc3)csc2=N)CC1
 CCS(=0)(=0)c1oc(-c2cccs2)nc1S(=0)(=0)c1ccccc1
 Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
 C0c1ncccc1NC(c1ccnc1)c1c(C)[nH]c2cc(C)ccc12
 CCn1c(C(=0)NCCO)ccc1C(CC)(CC)c1ccc(OCC(0)C(C)(C)C)c(C)c1
 CC0c1ccc(C=CC2=[N+](C)c3ccc(C(=0)OC)cc3C2(C)C)cc1
 Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
 CC(C)CS(=0)(=0)c1oc(-c2cccs2)nc1S(=0)(=0)c1ccc(F)cc1
 Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
 C0C(=0)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(0)cc1)=[N+]2C
 Cc1[nH]c2ccccc2c1C(Nc1cccn1)c1ccc(C(C)C)cc1

S4.1.3 Seeds

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)0)CC(0)C(CcN2cncn2)C10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=0)C(=C)C2CCCC)CC(0)C(C)C10
 C=C1C(=CC=C2CCC(C)CCCC(C)(C)0)C3(C)C2CCC3C(C)C0)CC(0)CC10
 CCn1c(C(=0)NC(C)C0)ccc1C(CC)(CC)c1ccc(OCC(0)C(C)(C)C)c(C)c1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)0)CC(0)CC10
 CCC(CC)(c1ccc(OCC(=0)C(C)(C)C)c(C)c1)c1ccc(C(=0)NCC(=0)OC)n1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(0)C(=CCCC)C10
 CC(CCCNS(N)(=0)=0)C1CCC2C(=CC=C3CC(0)CC(0)C3)CCCC21C
 CN1CCN(C(=0)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1

CCC(=CC=CC(O)(CC)CC)c1cccc(OCc2ccc(CO)c(CO)c2)c1
 Cc1[nH]c2cccc2c1C(Nc1cccc1)cc1
 CC(CCCC(=O)Nc1cccc1N)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CCC(CC)(c1ccc(OC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
 CCC(CC)(c1ccc(OC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
 Cc1[nH]c2cccc2c1C(Nc1ccc(N(C)C)cc1)c1cccc1C1
 COc1ccc(NC(c2cccc2C1)c2c(C)[nH]c3cccc23)cc1
 CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(OCc2ccc(CO)c(CO)c2)c1
 Cc1ccc(CNC(c2cccc2C1)c2c(C)[nH]c3cc(C)ccc23)cc1
 Cc1[nH]c2cccc2c1C(NC1cccc1)c1cccc1C1
 Cc1ccc2c(C(NC3ccc(C(F)(F)F)cc3)c3cccc3C1)c(C)[nH]c2c1
 Cc1ccc2c(C(Nc3ccnc3)c3cccc3C1)c(C)[nH]c2c1
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC10
 CS(=O)(=O)c1ccc(C(=O)NCCCC(=O)N2CCN(c3ccc(C(F)(F)F)cc3)CC2)cc1[N+](=O)[O-]
 COc1ccc(C(Nc2cccn2)c2c(C)[nH]c3cccc23)cc1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C10
 CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(OC(O)C(O)C3)CCCC12C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cccc2)CC(O)CC10
 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=C2C10
 Cc1ccc2c(C(NC3cccc3)c3cccc3C1)c(C)[nH]c2c1
 C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)CC(O)C4=C)CCCC32C)OC1=O
 Cc1ccc(NC(c2cccc2C1)c2c(C)[nH]c3cccc23)cc1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC10
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC10
 CCC(CC)(c1ccc(OC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCC)n1C
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC10
 CCC(CC)(c1ccc(OC(O)C(C)(C)C)c(C)c1)c1ccc(OC(O)C(O)CO)c(C)c1
 CC(CCCC(=O)Nc1cccc1N)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C1)cc3)csc2=N)CC1
 N=c1scc(-c2ccc(C1)cc2)n1CC(=O)N1CCOCC1
 CCn1c(C(=O)NC(C)CO)ccc1C(CC)(CC)c1ccc(OC(O)C(C)(C)C)c(C)c1
 CCC(CC)(c1ccc(OC(O)C(C)(C)C)c(C)c1)c1ccc(OC(O)CCO)c(C)c1
 COc1ccc(NC(c2ccc(OC(F)(F)F)cc2)c2c(C)[nH]c3cccc23)cc10C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC10
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC10
 COc1ccc(NC(c2cccc2C1)c2c(C)[nH]c3cccc23)cc10C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)CC10
 CC[Si](CC)(c1ccc(OC(=O)C(C)(C)C)c(C)c1)c1ccc(OC2CO2)c(C)c1
 CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1cccc1
 C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)C(OC(O)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)c1cccc1
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC10
 COCOCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CN2cnm2)C10
 CCC(CC)(c1ccc(OC(=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)C10
 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)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12

CC(CCCCN(C)=O)C1CCCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cccc2)CC(O)CC10
 Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3cccc3Cl)c(C)[nH]c2c1
 CCC(CC)(c1ccc(OC(O)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)CC10
 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)C10
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC)CC10
 CCC(CC)(c1ccc(OC(O)CO)c(C)c1)c1ccc(OC(O)C(C)(C)c(C)c1
 Cc1[nH]c2cccc2c1C(Nc1cccc1)c1ccc(Cl)cc1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC10
 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)CC10
 CC(CCCCC(=O)Nc1cccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CC(CCCCN(C)=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)C10
 C=C1C(O)CC(=CC=CCC(C)=CC(C)CCC(C)(C)O)CC10
 Cc1[nH]c2cccc2c1C(Nc1ccc([N+](=O)[O-])cc1)c1cccc1Cl
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C10
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC10
 C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(C)(C)O)CCC23C)CC(O)CC10
 COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3cccc23)cc1
 Cc1ccc2c(C(Nc3cccnc3)c3cccc3Cl)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(c2cccc2Cl)c2c(C)[nH]c3cccc23)cc1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cccc2)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCO)C10
 Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC10
 Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3cccc3Cl)c(C)[nH]c2c1
 COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C)[nH]c3cccc23)cc1
 CCC(=CC=CC(O)(CC)CC)c1ccc(OCc2ccc(O)c(O)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(=CCO)C10
 COc1ccc(NC(c2nccs2)c2c(C)[nH]c3cccc23)cc10C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCO)C10
 COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
 COc1ccc(NC(c2cccc2Cl)c2c(C)[nH]c3cccc23)cc1
 CC(C=CC(=O)NO)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CCC(CC)(c1ccc(OC(O)CO)c(C)c1)c1ccc(OC(O)C(C)(C)CCCc2cccc2)c(C)c1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cccc2)CC(O)CC10
 COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
 CCC(CC)(c1ccc(OC(O)C(C)(C)c(C)c1)c1ccc(C(=O)NC(O)CO)n1C
 CC(CCCCN(N)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3cccc23)cc10C
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC10
 Cc1[nH]c2cccc2c1C(Nc1cccc1)c1ccc(C(C)C)cc1
 CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCC(O)(CC)CC)=CCC23)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C10
 C=C1C(=O)OC(CC(C)C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
 N=c1sc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2cccc2)CC1
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCC)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C2(C)CCCC2)CC(O)CC10
 CCC(CC)(c1ccc(OC(=O)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)C(CCN2cncn2)C10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCO)C10
 CC(C=CC(=O)Nc1cccc1N)=CC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CC(=O)N1CCN(C(=O)Cn2c(-c3cccc3)csc2=N)CC1

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=O)C(=C)C2CCC)CC(0)C(0CCCC)C10
C=C1C(=CC=C2CCCC3(C)C(C4(CCC(=O)C(C)(C)O)CC4)=CCC23)CC(0)CC10
CC(CCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C0CC(=C)C(C)(C)C)CC(0)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OC)cc(OC)c2)CC(0)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(0)CC10
CCOc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OC)cc3C2(C)C)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C0c2ccc(C(C)(C)O)cc2)CC(0)CC10
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1C1

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(0)C(CcN2ncnm2)C10
CCC(0)(CC)CSC(C)C1=CC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(0)CC10
C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)(C)CCC(C)(C)O)CC10
CCC(0)(CC)CSC(C)C1=CC2C(=CC=C3CC(O)C(0CCO)C(O)C3)CCCC12C
CCC(0)(CC)CSC(C)C1=CC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(0)CC10
CCC(CC)(c1ccc(OC(O)CO)c(C)c1)c1ccc(OC(O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(0)C(CcN2ncn2)C10
C=C1C(O)CC(=CC=C2CCCC3C(C)CCCC(C)(C)O)CCC23)CC10
CCC(CC)(c1ccc(OC(O)C(C)(C)C)c(C)c1)c1ccc(OC(O)CO)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)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(0)CC10
CCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)OCC(C)C(C)(C)O)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(0)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCC(C)(C)O)CCC23)CC10
C=C1CCC(O)CC1=CC=C1CCC2(C)C1CCC2C(C)CCCC(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(0)C(CcN2cncm2)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCO)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2CC2)CC(0)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CC(O)(CC)CC)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(0)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC10
CCc1cc(C(O)(CC)CC)ccc1-c1cc(Cc2ccc(CO)c(CO)c2)ccc1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(0)C(CcN2cnc2)C10
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)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(0)CC10
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(Cc2ccc(CO)c(CO)c2)c1
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC10
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCc2ccc(CO)c(CO)c2)ccc1CC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC10
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC10
CcN1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(OC(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=C2CCCC(C)C)C3(C)C2CCC3C(C)CCO)CC(0)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=0)C(=C)C2(C)C)CC(0)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=0)C(=C)C2CCC)CC(0)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=0)C(=C)C2CCC3)CC(0)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=0)C(=C)C2CCC)CC(0)C(CCCO)C10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=0)C(=C)C2CCC)CC(0)C(C)C10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=0)C(=C)C2CC)CC(0)C(C)C10
 C=C1C(=0)OC(C)C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=0)C(=C)C2C)CC(0)CC10
 C=C1C(=CC=C2CCCC(C)C)C3(C)C2CCC3C(C)CO)CC(0)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=0)C(=C)C2CC)CC(0)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=0)C(=C)C2CCC)CC(0)C(OCCC)C10
 C=C1C(=CC=C2CCC(C)C)C3(C)C2CCC3C(C)CO)CC(0)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)C(O)C(=O)N2C)CC(0)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(C)C)CC(0)CC10
 CC(OCCC(C)C)C1OCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC21C
 CC(CCCNS(=O)(=O)C(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCC(O)CC)CC(0)CC1C(F)F
 CCC(CC)(c1ccc(OCC(=O)C(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
 CCCS(=O)(=O)NCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CC(OCCC(C)C)C1OCC2C(=CC=C3CC(O)C4=CCOC4C3)CCCC21C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(C)C)O)CC(0)CC1C(F)F
 CCC(CC)(c1ccc(OCC(=O)C(C)C)c(C)c1)c1ccc(OCC(=O)C(C)C)c(C)c1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC#CC(C)C)CC(0)CC10
 CC(C=CC(O)C1CC3CC(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(=CC23)CC(0)CC1C(F)F
 C=C1C(O)CC(=CC=C3CC(C)CCC(C)C)O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COc2cccc2)CC(0)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)O)C(=O)N2Cc2cc(OCC)cc(OCC)c2)CC(0)CC10
 CCC(CC)(c1ccc(OCC(=O)C(C)C)c(C)c1)c1ccc(OCC(O)CO)c(C)c1
 CCn1c(C(=O)NC(O)CC)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)C)c(C)c1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(0)CC10
 CC(OCCNS(=O)(=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CC(OCCC(=O)O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCC1(C)C
 CC(OCCNS(C)C(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CC(OCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)O)C(=O)N2CCCCc2cccc2)CC(0)CC10
 CCC(CC)(c1ccc(OCC(=O)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(OCCC)cc(OCCC)c2)CC(0)CC10
 CCC(CC)(c1ccc(OCC(O)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(O)cc2)CC(0)CC10
 Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3cccc3)CC2)cc1
 CCC(CC)(c1ccc(OCC(O)C(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
N=c1sc(-c2cccc2)n1CC(=O)N1CCN(c2cccc2)CC1
 CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OCC(=O)C(C)C)c(C)c1
 Nc1nc2ccc(S(=O)(=O)N3CCCC3)cc2s1
 CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOC(O)C(O)COCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(C)C(C)NC(=O)C(C)C(C)NC1=O
 Cc1ccc(CNC(c2cccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
 CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
 Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
 Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1ncs1
 COc1ccc(C=CC2=[N+](C)c3ccc(O)cc3C2(C)C)cc1
 Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1cccc1C1
 Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3cccc3Cl)c(C)[nH]c2c1
 Cc1[nH]c2cccc2c1C(NCc1cccc1)c1cccc1C1
 COc1ncccc1NC(c1ccnc1)c1c(C)[nH]c2cc(C)ccc12
 CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OCC(O)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(NC3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
 COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC
 COc1ccc(C(Nc2ccccc2)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)CCc2ccccc2)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)c1ccccc1
 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)N2CCc2cc(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)c1ccccc(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(C)C(C)C)CC(O)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCC)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=c1sc(-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(Cc2ccccc2)C1O
 CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1ccccc(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(C)C)CC(O)CC1O
 CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1ccccc(OCc2ccc(CO)c(CO)c2)c1
 COc1ccc(NC(c2ccccc2)c2c(C)[nH]c3ccccc23)cc1OC
 Nc1nc2ccc(S(=O)(=O)N3CCCC3)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(OCCCC)C1O
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)CC(O)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(C)C)CC1O
 C=C1C(O)CC(=CC=C2CCCC3(C)C(C)C)CC1O
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)CC1O
 C=C1C(=CC=C2CCC(C)C)C3(C)C2CCC3C(C)C)CC1O
 N=C1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)CC1O
 CC(C)C(C)C(=O)C1CCC2C(=CC=C3CC(O)C(C)C)CCCC21C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)CC1O
 CCC(O)C(C)C(C)C1=CC=C2C(=CC=C3CC(O)C(C)C)C3)CCCC12C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C2CC(C)C)C(=O)N2CCc2cccc2)CC(O)CC1O
 CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
 C=C1C(=O)OC(C)C2CCC3(=CC=C4CC(O)C(C)C)C4=C)CCCC32C)C1C
 CC(C)C(C)C(N)C(=O)C1CCC2C(=CC=C3CC(O)C(C)C)CCCC21C
 Cc1[nH]c2cccc2c1C(Nc1cccc1)c1cccc1C1
 Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1ncs1
 COc1ccc(C(Nc2ccc(C(F)F)cc2)c2c(C)[nH]c3cccc23)cc1
 CCC(C)C(c1ccc(OCC(O)C)cc1)c1ccc(OCC(O)C(C)C)c1
 COc1ccc(NC(c2cccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC
 COCCOCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)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)C#C(O)C23CC4CC(C)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)CC(C)C)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)CC(O)C(Cn2cncn2)C1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)C)C(=O)N2Cc2cc(O)CC(C)C)CC(O)CC1O
 CC(C)C(C)C(=O)C1CCC2C(=CC=C3CC(O)C(C)C)CCCC21C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C1O
 CCC(C)C(c1ccc(OCC(=O)C(C)C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
 N=C1scc(-c2cccc2)n1CC(=O)N1CCN(c2cccc2)CC1
 CC(C=CC(=O)Nc1cccc1N)=CC(C)C1CCC2C(=CC=C3CC(O)C(C)C)CCCC21C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)CC(O)C(Cn2nccn2)C1O
 CCC(=CC=CC(O)C)C1cccc(OCc2ccc(C)C)C2)c1
 Cc1ccc2c(C(Nc3cccnc3)c3cccc3Cl)c(C)[nH]c2c1
 COc1ccc(NC(c2ccc(OC(F)F)cc2)c2c(C)[nH]c3cccc23)cc1OC
 CCn1c(C(=O)NC(C)CO)ccc1C(C)C)c1ccc(OCC(O)C(C)C)c1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)CC(O)C2=CCOC12
 CCC(C)C(c1ccc(OCC(O)C(C)C)c1)c1ccc(C(=O)NCCC)O)n1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
 C=C1C(=O)OC(C)C2CCC3(=CC=C4CC(O)C(C)C)C4=C)CCCC32C)C1C
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)CC(C)C)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)C)C(=O)N2Cc2cc(O)CC(C)C)CC(O)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)CC(O)CC1O
 CCn1c(C(=O)NC(C)CO)ccc1C(C)C)c1ccc(OCC(O)C(C)C)c1
 C=C1C(O)CC(=CC=CCC(C)C)CC(C)CCCC(C)C)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)CC(O)C(C)C)CC(O)CC1O
 COc1ccc2c(C(Nc3ccc(Cl)cc3)c3cccc3Cl)c(C)[nH]c2c1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
 CC(C)C(C)C(=O)N1C1CCC2C(=CC=C3CC(O)C(C)C)CCCC21C
 CC(C)C(C)C(N)C(=O)C1CCC2C(=CC=C3CC(O)C(C)C)CCCC21C
 Cc1ccc(NC(c2cccc2Cl)c2c(C)[nH]c3cccc23)cc1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)CC(O)CC1O
 CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C)C)SCC(O)C)CC)C)CC23)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)C)C(=O)N2CCc2cccc2)CC(O)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)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)cc3)c3ccccc3Cl)c(C)[nH]c2c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)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(OC(=O)C)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(C(C4)C2)C3)CC1O
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OC(=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(=CCO)C(O)C3)CCCC21C
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
CCC(O)(CC)CSC(C)C1=CC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COC(=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=C2CCCC3(C)CCCC(C)(C)O)CCC23)CC1O
CC(CCCNS(C)(=O)O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+](C)2C
C=C1C(=CC=C2CCC(C)CCCC(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(CCCNS(N)(=O)O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OC(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(Cc1n2cnm2)C1O
CCC(O)(CC)CSC(C)C1=CC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)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(OC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CO)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccc(C(C)(C)O)cc2)CC(O)CC1O
CCOC1C(O)CC(=CC=C2CCCC3(C)C(C)SCCC(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)CCC23)CC(O)CC1O
C=C1CCC(O)CC1=CC=C1CCC2(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)(C)CC)CC(O)C(=CCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OC)CC2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)OCCC)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(Cc1ncnc2)C1O
C=C1CC(CC(C)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)CCO)CC1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)C(O)C)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(O)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(CC)CC(C)(C)O)CC1O
CCc1cc(C(O)(CC)CC)ccc1-c1cc(OCC2ccc(O)c(O)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(C)CCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1CC(CC(C)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(OCCC)C1O
C=C1C(=O)OC(CC(C)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)C(OCCC)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCC)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(C)O)C(=O)N2C)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)C)CC(O)CC1O
CC(OCCNS(=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(OCC(O)CCO)c(C)c1
CCC(CC)(c1ccc(OCC(=O)C(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
CCCCS(=O)(=O)NCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(OCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCOC4C3)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=C2CCCC3(C)CCCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C1O
C=C1C(O)CC(=CC=C2CCCC3(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)C)C(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCC(=C)C2(C)CCCC2)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(OCC)cc(OCC)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)O)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccc(OCC)cc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCCc2cc(OCC)cc(OCC)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] c2cccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
 CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)C(O)C(O)C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(C)C(NC1=O)C
 COc1ccc(C(Nc2ccc(OC)c(OC)c2)c2c(C) [nH] c3cccc23)cc1
 COC(=O)c1ccc2c(c1)C(C)C(C)C(C=Cc1ccc(OC)cc1)=[N+]2C
 Cc1ccc(CNC(c2cccc2Cl)c2c(C) [nH] c3cc(C)ccc23)cc1
 CCC(CC)(c1ccc(OC(O)CO)c(C)c1)c1ccc(OC(O)C(C)C)CCCc2cccc2)c(C)c1
 Cc1 [nH] c2cccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
 Cc1 [nH] c2cccc2c1C(Nc1ccc(Cl)cc1)c1cccc1Cl
 Cc1 [nH] c2cccc2c1C(Nc1ccc([N+](=O)[O-])cc1)c1cccc1Cl
 COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C) [nH] c3cccc23)cc1OC
 CC(=O)N1CCN(C(=O)Cn2c(-c3cccc3)csc2=N)CC1
 Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3cccc3Cl)c(C) [nH] c2c1
 CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1cccc1
 Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3cccc3Cl)c(C) [nH] c2c1
 Cc1 [nH] c2cccc2c1C(NC1cccc1)c1cccc1Cl
 COc1ncccc1NC(c1cccnc1)c1c(C) [nH] c2cc(C)ccc12
 CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OC(O)C(C)C)C(C)c(C)c1
 COc1ccc(NC(c2cccc2Cl)c2c(C) [nH] c3cccc23)cc1
 Cc1ccc2c(C(NC3cccc3) c3cccc3Cl)c(C) [nH] c2c1
 CCC(CC)(c1ccc(OC(O)CO)c(C)c1)c1ccc(OC(O)C(C)C)CCc2cccc2)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] c2cccc2c1C(Nc1cccc1)c1ccc(Cl)cc1
 COC(=O)c1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
 COc1cccc1NC(c1cccc1Cl)c1c(C) [nH] c2cccc12
 Cc1 [nH] c2cccc2c1C(Nc1ccc(N(C)C)cc1)c1cccc1Cl
 Cc1 [nH] c2cccc2c1C(Nc1cccc1)c1ccc(C(C)C)cc1
 Cc1ccc2c(C(NC3ccc(C(F)(F)F)cc3)c3cccc3Cl)c(C) [nH] c2c1
 COc1ccc(C(Nc2ccccn2)c2c(C) [nH] c3cccc23)cc1

S4.3.3 Seeds

CC(CCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CCC(CC)(c1ccc(OC(O)C(C)C)C(C)c1)c1ccc(C(=O)NCCCC)N1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OC(O)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=c1sc(-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(C)CCO)CC1O
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC1O
 C=C1C(=CC=C2CCC(C)CCCC(C)C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=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(C)CCCC(C)C)O)CCC23)CC1O
 CCC(CC)(c1ccc(OC(O)C(C)C)C(C)c1)c1ccc(C(=O)NCC(=O)O)N1C
 CC(CCCNS(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)COc2cccc2)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(CCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 Cc1 [nH] c2cccc2c1C(Nc1ccc(Cl)cc1)c1nccs1

C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
C0c1ccc(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(=CCO)C(O)C3)CCCC21C
CC(C=CC(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(Cc2cncn2)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(CCCNS(=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)COC(=C)C2(C)CCCC2)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(Cc2cncn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCO)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
C0c1ccc(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
C0c1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCO)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)CCO)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
Cc1[nH]c2ccccc2c1C(NCc1ccccc1)c1ccccc1Cl
CC(CCCC(=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(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(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(C)CCCC(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)CC10
C=C1C(=CC=C2CCC(CCCCC(C)(C)O)C3(C)C2CCC3C(C)O)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC10
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C0CC(=C)C(C)(C)CC(O)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC10
C=C1C(O)CC(=CC=CCCC(C)CCCC(C)(C)O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCC)C10
CCC(CC)(c1ccc(OC(C)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)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)CCCC(C)(C)O)CCC23)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc2nccn2)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC10
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(OC(C)O)c(C)c1)c1ccc(OC(C)C(C)C)CCCc2ccccc2)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCO)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C10
Cc1[nH]c2ccccc2c1C(Nc1cccn1)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(C)CCCC(C)(C)O)CC4)=CCC23)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc2cnc2)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC)CCC(C)(C)O)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC10
CCC(CC)(c1ccc(OC(=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)CC10
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
CCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C(C)CCCC(C)(C)O)CCC23)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC10
Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
CC(O)CCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCOC4C3)CCCC21C
CC(O)CC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COc2ccc(C(C)(C)O)cc2)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC10
Cc1ccc(GNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CC)CC(O)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C10
CCC(CC)(c1ccc(OC(=O)C(C)(C)C)c(C)c1)c1ccc(OC(=O)C(C)(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C10
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)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc2nccn2)C10

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(OCC(O)CO)c(C)c1
 CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCF)C(O)C3)CCCC21C
 CCOCC1C(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(CCn2cmm2)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)OCCC)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)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(OCC(O)C(O)CO)c(C)c1
 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)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(C)CCCC(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
 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(C)CCCC(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
 C=C1C(=O)OC(CC(C)C)C2CCC3C(=CC=C4CC(O)C(O)CCO)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)C2CCC)CC(O)C(CCCO)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)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(O)C(CCCO)C1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=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(CCCNS(=O)(=O)c1cccc1)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CCCC(=O)(=O)NCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CC(CCCNC(=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

C0C0CCC=C1C(0)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)0)CC10
 C=C1C(0)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)0)CC10
 CC(C=CC(=0)N0)=CC(C)C1CCC2C(=CC=C3CC(0)CC(0)C3)CCCC21C
 CC(C=CC(=0)Nc1cccc1N)=CC(C)C1CCC2C(=CC=C3CC(0)CC(0)C3)CCCC21C
 CCn1c(C(=0)NC(0)CC0)ccc1C(CC)(CC)c1ccc(0CC(0)C(C)(C)C)c(C)c1
 CCC(CC)(c1ccc(0CC(0)C(C)(C)C)c(C)c1)c1ccc(C(=0)NCCC0)n1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(0)C(=0)N2Cc2cc(0CCC)cc(0CCC)c2)CC(0)CC10
 CC(CCCCCNS(N)(=0)=0)C1CCC2C(=CC=C3CC(0)CC(0)C3)CCCC21C
 CC(CCCCC(=0)N0)C1CCC2C(=CC=C3CC(0)CC(0)C3)CCCC21C
 CC(CCCCCNS(N)(=0)=0)C1CCC2C(=CC=C3CC(0)CC(0)C3)CCCC21C
 C=C(C0CC(C)C1CCC2C(=CC=C3CC(0)CC(0)C3=C)CCCC21C)C1=CCCC1(C)C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(0)C(=0)N2CCCCc2cccc2)CC(0)CC10
 CCC(CC)(c1ccc(0CC(=0)C(C)(C)C)c(C)c1)c1ccc(C(=0)NC(CC(=0)0)C(=0)0)n1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(0)C(=0)N2Cc2cc(0CCC)cc(0CCC)c2)CC(0)CC10
 C0c1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3cccc23)cc1
Cc1ccc(NC(c2cccc2Cl)c2c(C)[nH]c3cccc23)cc1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(0)C(=0)N2Cc2ccc(0C)cc2)CC(0)CC10
 CCC(CC)(c1ccc(0CC(=0)C(C)(C)C)c(C)c1)c1ccc(C(=0)NCCC(=0)0C)n1C
 Cc1ccc(-c2csc(=N)n2CC(=0)N2CCN(c3cccc3)CC2)cc1
 CCC(CC)(c1ccc(0CC(0)C(C)(C)C)c(C)c1)c1ccc(C(=0)NCCC(=0)0)n1C
 Nc1nc2ccc(S(=0)(=0)N3CCCC3)cc2s1
 C0c1ccc(NC(c2ccc(0C(F)(F)F)cc2)c2c(C)[nH]c3cccc23)cc10C
 CC(=0)N1CCN(C(=0)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
 CCC(CC)(c1ccc(0CC(=0)C(C)(C)C)c(C)c1)c1ccc(C(=0)NC(C)C(=0)0)n1C
 N=c1sc(-c2ccc(Cl)cc2)n1CC(=0)N1CCN(c2cccc2)CC1
 Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
 CC(C)CC(N)C(=0)NC(CC(C)C)C(=0)NC1C0CC(0)C(0)C0CCC(C(=0)0)NC(=0)C(CC(C)C)NC(=0)C(CC(C)C)NC(=0)C(C)(C)NC1=0
 C0c1ccc(C(Nc2ccc(0C)c(0C)c2)c2c(C)[nH]c3cccc23)cc1
 C0C(=0)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(0C)cc1)=[N+]2C
 CN1CCN(C(=0)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
 Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1ncs1
 C0c1ccc(C=CC2=[N+](C)c3ccc(0C)cc3C2(C)C)cc1
 Cc1[nH]c2cccc2c1C(Nc1ccc([N+](=0)[O-])cc1)c1cccc1C1
 CC(=0)N1CCN(C(=0)Cn2c(-c3cccc3)csc2=N)CC1
 CCS(=0)(=0)c1oc(-c2cccs2)nc1S(=0)(=0)c1cccc1
 C0c1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
 C0c1ccc(NC(c2cccc2Cl)c2c(C)[nH]c3cccc23)cc1
 Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3cccc3Cl)c(C)[nH]c2c1
 Cc1ccc2c(C(NC3cccc3)c3cccc3Cl)c(C)[nH]c2c1
 CC(C)CS(=0)(=0)c1oc(-c2cccs2)nc1S(=0)(=0)c1ccc(F)cc1
 C0C(=0)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(0)cc1)=[N+]2C
 Cc1ccc2c(C(Nc3cccnc3)c3cccc3Cl)c(C)[nH]c2c1
 CCC(CC)(c1ccc(0CC(0)0)c(C)c1)c1ccc(0CC(0)C(C)C)CCc2cccc2)c(C)c1
 C0c1ccc2c(C(Nc3ccc(Cl)cc3)c3cccc3Cl)c(C)[nH]c2c1
 Cc1[nH]c2cccc2c1C(Nc1cccc1)c1ccc(Cl)cc1
 C0C(=0)c1ccc(C=CC2=[N+](C)c3ccc(0C)cc3C2(C)C)cc1
Cc1[nH]c2cccc2c1C(Nc1cccc1)c1cccc1C1
 C0c1ccc(NC(c2cccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc10C
 C0c1ccc(C(Nc2cccn2)c2c(C)[nH]c3cccc23)cc1

S4.4.3 Seeds

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)0)CC(0)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)0)CC(0)C(Cc1nccn2)C10
 C=C1C(0)CC(=CC=CCCC(C)CCCC(C)(C)0)CC10
 Cc1ccc(NC(c2cccc2Cl)c2c(C)[nH]c3cccc23)cc1
 C0C1ccc(C=CC2=[N+](C)c3ccc(C(=0)0C)cc3C2(C)C)cc1

CCC(CC) (c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)CCCc2ccccc2)c(C)c1
 CC(OCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC21C
 Cc1ccc2c(C(Nc3ccc(O(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)CC10
 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)CC10
 C=C1C(O)CC(=CC=C2CCCC3C(C)CC(C)O)CCC23)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCO)C10
 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(C)CC(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C10
 Cc1[nH]c2ccccc2c1C(NC1CCCC1)c1ccccc1Cl
 Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cncnc1
 C=C1C(O)CC(=CC=C2CCCC3(C)C(C)CC(C)O)CCC23)CC10
 C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCO)C10
 CC(OCCC(=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)CC10
 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(OCCC(=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(C)CC(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC10
 Cc1ccc2c(C(Nc3ccc(O(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
 CC(OCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC21C
 C=C1C(O)CC(=CC=C2CCCC3(C)C(C)CC(C)O)CCC23)CC10
 COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)CC10
 CCC(CC) (c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC(=O)C(C)(C)C)c(C)c1
 CC(OCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 Cc1ccc2c(C(Nc3ccccc3)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)CCC23)CC(O)CC10
 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(Cn2cnm2)C10
 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)CC10
 CC(OCCNS(=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)C10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)C)CC(O)CC10
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)CCO)CC10
 C=C1C(=CC=C2CCCC3(C)C(C4(C)CC(C)(C)O)CC4)=CCC23)CC(O)CC10
 C=C1C(=CC=C2CCC(C)CC(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC10
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CC)CC10
 CCn1c(C(=O)NC(C)CO)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)c1ccccc1Cc2ccc(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)c(C)c1c1ccc(OCC(O)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(=CCO)C1O
 COc1ncccc1NC(c1cccnc1)c1c(C)[nH]c2cc(C)ccc12
 CCC(CC) (c1ccc(OCC(O)C(C)C)C(C)C)c(C)c1c1ccc(C(=O)NCC(=O)O)n1C
 COc1ccc(C(Nc2cccn2)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(OCCCC)C1O
 CC(CCCNC(=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)COC(=C)C(C)C)CC(O)CC1O
 CCC(CC) (c1ccc(OCC(=O)C(C)C)C(C)C)c(C)c1c1ccc(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)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)c1cnc1
 C=C1C(=O)OC(C)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(OCCCC)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)c(C)c1c1ccc(OCC(O)CCO)c(C)c1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC1O
 C=C1C(=O)OC(C)C)C2CCC3C(=CC=C4CC(O)C(OCCCC)C(O)C4=C)CCCC32C)C1C
 CCC(CC) (c1ccc(OCC(O)C(C)C)C(C)C)c(C)c1c1ccc(C(=O)NC(C)CO)n1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COC#CC(C)C)CC(O)CC1O
 C=C1C(O)CC(=CC=C2CCCC3C(C)C)O)CC23)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)c1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCC)CC(O)C(C)C)CC(O)CC1O
 COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)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(CCCNS(N) (=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CCC(CC) (c1ccc(OCC(=O)C(C)C)C(C)C)c(C)c1c1ccc(C(=O)NCC(=O)O)n1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCc2cc(OCC)cc(OCC)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(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(CCCNS(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(OCC(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(OCCCC)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)c(C)c1c1ccc(C(=O)NCCO)n1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(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(C)CCC(C)C)O)CC(O)CC1O

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(=O)C(=O)C1O
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCC1(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(O)CC)c(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(O)CC)c(O)c2)CC(O)CC1O
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1cccc(Cc2ccc(O)c(O)c2)c1
CCC(CC)(c1ccc(OCC(O)C(C)C)c(C)c1)c1ccc(OCC(O)CCO)c(C)c1
Cc1ccc2c(C(NC3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOC(O)C(O)C(O)C(O)C(=O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(C)NC1=O
CCOC1C(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=CC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
Nc1nc2ccc(S(=O)(=O)N3CCCC3)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(CcN2ncnm2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)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)C(C)(C)CCC(C)(C)O)CC1O
CCC(O)(CC)CSC(C)C1=CC2C(=CC=C3CC(O)C(O)C3)CCCC12C
CCC(O)(CC)CSC(C)C1=CC2C(=CC=C3CC(O)C(=CO)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(CcN2ncnm2)C1O
CC(O)CCC(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
CCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(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(CcN2ccnm2)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(=CCO)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(=O)O)CC2)CC(O)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(Cc2ccc(O)c(O)c2)ccc1C
CC(O)CCC(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)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(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(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(O)C2ccc(O)c(O)c2)c1

CC0c1ccc(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(Nc1ccccc1)c1ccc(C(C)C)cc1
 Cc1ccc2c(C(NC3ccc(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(C)CCC(C)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)COC(=C)C(C)(C)C)CC(O)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCC)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(Nc3ccnc3)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(OCCCC)C1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC1O
 CC(OCCC(=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)(O)CC1O
 C=C1C(=O)OC(CC(C)C)C2CCC3C(=CC=C4CC(O)C(OCCCC)C(O)C4=C)CCCC32C)C1C
 CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOC(C)C(O)C(OCCC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(C)(C)NC(=O)C(C)C)NC1=O
 CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CO)c(C)c1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(O)CC(O)C2=CCOC12
 Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1Cl
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COC#CC(C)(C)C)CC(O)CC1O
 CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCC)n1C
 CCn1c(C(=O)NC(CO)CO)ccc1C(CC)(CC)c1ccc(OCC(O)C(C)(C)C)c(C)c1
 C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCC(C)(C)O)CCC23)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(O)CC(O)C(Cc2cnm2)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(C)C)CC(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(OCCCC)C1O
 CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1

Cc1ccc2c(C(Nc3ccc(C(F)(F)F)cc3)c3ccccc3C1)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(OC(O)C(C)(C)c(C)c1
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
Cc1[nH]c2cccc2c1C(Nc1cccc1)c1cccc1C1
CCC(CC)(c1ccc(OC(=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)Nc1cccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
COc1ccc2c(C(Nc3ccc(C1)cc3)c3ccccc3C1)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)N2CCCCc2cccc2)CC(O)CC1O
CCC(CC)(c1ccc(OC(=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)OC)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(CCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
Nc1nc2ccc(S(=O)(=O)N3CCCC3)cc2s1
Cc1[nH]c2cccc2c1C(Nc1ccc(C1)cc1)c1ncc[nH]1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OC(O)C)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCO)CC1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC1O
CC(CCCNS(=O)(=O)c1cccc1)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)CC2OC(=O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC1O
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OC(=O)C(C)(C)c(C)c1
CC(CCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
CCC(CC)(c1ccc(OC(=O)C(C)(C)C)c(C)c1)c1ccc(OC(=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(C)CCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1CC(CC(C)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(OC(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(c2cccc2C1)c2c(C)[nH]c3cccc23)cc1OC
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COC#CC(C)(C)C)CC(O)CC1O
C=C1C(O)CC(=CC=CCCC(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(C)CCCC(C)(C)O)CCC23)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=CCCC4C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC1O
COc1cccc1NC(c1cccc1C1)c1c(C)[nH]c2cccc12
COc1ccc(NC(c2cccc2C1)c2c(C)[nH]c3cccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OC(O)C)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2cccc2)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)OCCC(O)(CC)CC)CC(O)CC1C(F)F
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)N2Cc2cc(OC)cc(OC)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(=CCO)C(O)C3)CCCC21C
CC(OCC(=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(Cc2n2ccnc2)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)NCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C)CC(C)CC(O)C(=CCO)C1O
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
C=C1C(=O)OC(C(C)C)C2CCC3C(=CC=C4CC(O)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)N2CCc2cc(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
CO(C(=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(NC1c1ccccc1)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(Cc2ccc(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
CO(C(=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(Cc2n2ccnm2)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(OC)cc(OC)c2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc2n2cnm2)C1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
COc1ncccc1NC(c1ccccc1)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(Cc2n2ccnm2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CC(C)(C)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(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
CCCCO1C(O)CC(=CC=C2CCCC3(C)C(C(C)S)CC(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(Cc2n2ccnc2)C1O
C=C1C(O)CC(=CC=C2CCCC3C(C)CCCC(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(Cn2cn2)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(=CCO)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(C)CC(O)(CC)CC)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCO)C1O
 CC(OCCC(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(C(C4)C2)C3)CC1O
 CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)C(O)C)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(C)CCO)CC1O
 C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O
 C=C1C(=CC=C2CCC(C)CCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
 C=C1CC(CC(C)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)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(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)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(C)CCO)CC1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
 C=C1C(=CC=C2CCC(C)CCCC(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(OCCC)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(OCC(O)CCO)c(C)c1
 CC(OCCNC(=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=CC(O)C12CC3CC(CC(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 C=C1C(=CC=C2CCCC3(C)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)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(C(=O)NCCO)n1C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCC)cc(OCCC)c2)CC(O)CC1O
 CC(OCCCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCC(=C)C2(C)CCCC2)CC(O)CC1O
 CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
 CC(OCCNS(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=CCCC1(C)C
 CC(OCCC(=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)C2CC3C(C)CC2CC(C) (O)C(=O)N2CCCC2ccccc2)CC(O)CC10
 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) c1ccnc1
 CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1COC(C)C(=O)C(C)C(C)C(C)C(=O)NC(=O)C(C)C(C)NC(=O)C(C)C(C)NC(=O)C(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)CCc2ccccc2)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(NC3ccccc3) c3ccccc3Cl)c(C) [nH] c2c1
 Cc1ccc2c(C(Nc3ccccc3) 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(Nc1ccccc1) c1ccc(C(C)C)cc1
 COc1ccc(NC(c2ccccc2Cl)c2c(C) [nH] c3cc(C)ccc23)cc1OC
 COc1ccc(C(Nc2ccccc2)c2c(C) [nH] c3ccccc23)cc1

S4.6.3 Seeds

C=C1C(O)CC(=CC=C2CCCC3(C)C2CC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)COCc2ccccc2)CC(O)CC10
 Cc1ccc2c(C(Nc3ccc(Cl)cc3) c3ccccc3Cl)c(C) [nH] c2c1
 COc1nccccc1NC(c1ccccc1) c1c(C) [nH] c2cc(C)ccc12
 Cc1ccc2c(C(Nc3ccc(OC(F) (F)F)cc3) c3ccccc3Cl)c(C) [nH] c2c1
 C=C1C(=CC=C2CCC(CCCCC(C) (C)O)C3(C)C2CC3C(C)CCO)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC10
 Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3ccccc3)CC2)cc1
 C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C) (C)O)CCC23)CC10
 CCC(CC) (c1ccc(OCC(O)CO)c(C) c1) c1ccc(OCC(O)C(C) (C)C)c(C) c1
 CCCc1cc(C(O) (CC)CC)ccc1-c1cc(CCc2ccc(O)c(O)c2)ccc1C
 COc1ccc(C=CC2=[N+] (C) c3ccc(C(=O)OC)cc3C2(C)C)cc1
 CC(OCCNS(C) (=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCC21C
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)COC#CC(C) (C)C)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(OCCO)C10
 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)C2CC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2CC(C) (O)C(=O)N2CCCC2cc(OC)cc(OC)c2)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2CC(C) (O)C(=O)N2CCc2cc(OC)cc(OC)c2)CC(O)CC10
 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(=CCO)C1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(=O)C2(C=CC(=O)OCC)CC2)CC(=O)CC1O
 CC(CCCNS(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)OCCC)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)N2Cc2cccc2)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=CCCC(C)CCCC(C)(C)O)CC1O
 CC(CCCNS(=O)(=O)c1cccc1)C1CCC2C(=CC=C3CC(=O)CC(=O)C3)CCCC21C
 Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1ncs1
 CCC(=CC=CC(=O)(C(F)(F)F)C(F)(F)F)c1cccc(Cc2ccc(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(NC3ccc(C(F)(F)F)cc3)c3cccc3Cl)c(C)[nH]c2c1
 COc1ccc2c(C(Nc3ccc(Cl)cc3)c3cccc3Cl)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)CCO)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)=CC23)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(C)CCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(=O)CC1O
 Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3cccc3)CC2)cc1
 C=C1CC(C(C)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(C(=O)O)C(=O)O)n1C
 CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(=O)C(=CCO)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=CCOC4C3)CCCC21C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)OCCCC(O)(CC)CC)CC(O)CC1C(F)F
 COc1ccc(NC(c2cccc2Cl)c2c(C)[nH]c3cccc23)cc1OC
 CC(CCCCC(=O)Nc1cccc1N)C1CCC2C(=CC=C3CC(=O)CC(=O)C3)CCCC21C
 CC(CCCNS(=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(Nc2cccc2)c2c(C)[nH]c3cccc23)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(CCO)C1O
 C=C1C(=O)CC(=CC=C2CCCC3(C)CCCC(C)(C)O)CCC23)CC1O
 CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CO)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)CCC2cccc2)c(C)c1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(=O)C(C)C1O
 CN1CCN(C(=O)Gn2c(-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)CS2ccc(C(C)(C)O)cc2)CC(=O)CC1O
 Cc1ccc(CNC(c2cccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
 C=C1C(=O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CO)CCO)CC1O

Cc1 [nH] c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccnc1
 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(Nc1cccc1)c1ccc(Cl)cc1
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC10
 Cc1ccc2c(C(Nc3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2CC(C)(O)C(=O)N2CCCc2ccccc2)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(O)cc(O)c2)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CCCC(C)(C)O)CC(O)C2=CCOC12
 C=C1C(O)CC(=CC=C2CCCC3C2CC3C(C)CCCC(C)(C)O)CC10
 COc1ccc(NC(c2ccc(O)F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
 CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
 COCCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CC3C(C)CCCC(C)(C)O)CC10
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC10
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CCCC(C)(C)O)CC(O)C(Cc2n2cnc2)C10
 CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC)nc1C
 CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)nc1C
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)C(C)CCC(C)(C)O)CC(O)CC10
 Cc1 [nH] c2ccccc2c1C(Nc1cccc1)c1ccccc1Cl
 CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
 C=C1C(=CC=C2CCC(CCCCC(C)(C)O)C3(C)C2CC3C(C)CO)CC(O)CC10
 N=c1sc(-c2ccccc2)nc1CC(=O)N1CCN(c2ccccc2)CC1
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CC3C(C)C(C)CCC(C)(C)O)CC10
 CCn1c(C(=O)NCC(=O)O)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)C2CC3C(C)C=CC(O)C2(C(=O)OCCCC)CC2)CC(O)CC10
 C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCC1(C)C
 Cc1ccc2c(C(Nc3ccc(C(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(O)cc(O)c2)CC(O)CC10
 CC(CCC(=O)O)C1CCC2C3CCC4CC(O)CCC4(C)C3CCC12C
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CC3C(C)C(C)CCC(C)(C)O)CC10
 Cc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
 C=C1C(=O)OC(CC(C)C2CC3C(=CC=C4CC(O)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)C2CC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
 CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCC(O)(CC)CC)=CCC23)CC10
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC10
 CC(CCCC(=O)Nc1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2CC(C)(O)C(=O)N2C)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC)CC(O)C(=CCO)C10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C10
 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)C2CC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C10
 CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)nc1C
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)C=CC(O)C2(C=CC(=O)O)CC2)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CCC(O)C(C)C)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CC3C(C)CCCC(C)(C)C)CC(O)CC10

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(Cc2ncnm2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(O)CCO)C(O)C3)CCCC12C
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(=CO)C(O)C3)CCCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc2ncnm2)C1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)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)c2)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(CCCC(C)(C)O)CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc2c2nm2)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(Cc2c2nm2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCO)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(C)CC(O)(CC)CC)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCO)C1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(O)c(O)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(Cc2c2nc2)C1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)C(O)CO)c(C)c1
CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)C)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cccc2)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(O)c(O)c2)ccc1CC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)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(O)c(O)c2)c1
C=C1CC(CC(C)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)C)C2CCC3C(=CC=C4CC(O)C(CCO)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(CCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCC)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
C=C1C(=CC=C2CCC(C)CCCC(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(OCCCC)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)C2CCC)CC(O)C(OCCCC)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(CCO)C1O
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CCO)c(C)c1
CCCCS(=O)(=O)NCCCC(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(CCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C

CC[Si] (CC) (c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC2C02)c(C)c1
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C0CC#CC(C)(C)C)CC(O)CC10
 CC(C=CCC(O)C12CC3CC(C(C3)C1)C2)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 C=C1C(=CC=C2CCCC3(C)C(C(C)C)CC(F)(F)C(O)(CC)CC)=CCG23)CC(O)CC1C(F)F
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C10
 C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C0Cc2ccccc2)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C0CC(=C)C(C)(C)C)CC(O)CC10
 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)CC10
 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(O)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)N2Cc2ccccc2)CC(O)CC10
 CC(CCCCN(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C0CC(=C)C2(C)CCCC2)CC(O)CC10
 CC(CCCN(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CC(CCCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CC(CCCCN(C)(=O)=O)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)CC10
 CC(CCCCN(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)CC10
 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(OCC)cc2)CC(O)CC10
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(OCC)cc(OCC)c2)CC(O)CC10
 CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
 Nc1nc2ccc(S(=O)(=O)N3CCCC3)cc2s1
 N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2ccccc2)CC1
 CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1C0CC(O)C(O)C0CC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(C)C(C)NC(=O)C(C)C(NC1=O)

COc1ccc(C(Nc2ccc(OCC)cc(OCC)c2)c2c(C)[nH]c3ccccc23)cc1
 COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(OCC)cc1)=[N+]2C
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
 COc1ccc(C=CC2=[N+](C)c3ccc(OCC)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(OCC(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
 CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
 Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl
 COc1ncccc1NC(c1ccccc1)c1c(C)[nH]c2cc(C)ccc12
 COc1ccc(C=CC2=[N+](C)c3ccc(C(=O)OCC)cc3C2(C)C)cc1
 COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
 Cc1ccc2c(C(NC3ccccc3)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(Nc1ccccc1)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)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=O)OC(CC(C)C2CCCC3(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCC(C)CCCC(C)(C)O)C3(C)C2CCC3C(C)CCCC(O)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(C)CCCC)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)NCCC)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
COCOC=CC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O
N=c1sc(-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=CCCC1(C)C
C=C1C(=CC=C2CCCC3(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)c1cccn1
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(C)CCCC(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)C1CC2C(=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)NC1COC(O)C(O)COC(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(NC3ccc(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(C)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(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(OCC(O)CCO)c(C)c1

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C1O
CCCCC1C(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(Cc1n2cnn2)C1O
CCC(O)(CC)CSC(C)C1=CC2C(=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(OC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC0)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc1n2cnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COC#CC(C)(C)O)CC(O)CC1O
COc1ccc(C(Nc2cccn2)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(Cc1n2cnn2)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(OC)C)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OC)C1O
C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC1O
Cc1[nH]c2ccccc2c1C(Nc1cccn1)c1ccc(C(C)C)cc1
CCC(CC)(c1ccc(OC(=O)C(C)(C)C)c(C)c1)c1ccc(OC(=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)O)CC(O)CC1O
CCCCC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC1O
C=C1C(O)CC(=CC=C2CCCC3C(C)CCCC(C)(C)O)CCC23)CC1O
CCn1c(C(=O)NCC(=O)OC)ccc1C(CC)(CC)c1ccc(OC(=O)C(C)(C)C)c(C)c1
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCOC4C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC1O
CCC(O)(CC)CSC(C)C1=CC2C(=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(CCCNS(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)CC2OC=C2C1O
COc1ccc(NC(c2ccc(OC(F)F)cc2)c2c(C)[nH]c3ccccc23)cc1OC
CCn1c(C(=O)NCCO)ccc1C(CC)(CC)c1ccc(OC(O)C(C)(C)C)c(C)c1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cnc1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCO)CC1O
N=c1sc(-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(=CCO)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC1O
C=C1C(=CC=C2CCC(C)CCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COC(=C)C(C)(C)O)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]c2cccc2c1C(NCc1cccc1)c1cccc1C1
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=CC2C(=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(=CCO)C(O)C3)CCCC21C
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C1)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)NCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
Cc1ccc2c(C(NC3cccc3)c3cccc3C1)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(OCC(O)C(O)O)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(CCCC)CCC(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(C1)cc3)c3cccc3C1)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(=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)O)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(OCC(O)O)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(Cc2cncn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc2c2cnn2)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCC)CC(O)C(=CCO)C1O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CC(O)CC(O)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(=CCO)C1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(CCc2ccc(O)c(O)c2)ccc1C
CC(OCCC(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(O)O)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)N2Cc2cccc2)CC(O)CC1O
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCC2ccc(O)c(O)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(=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(O)c(O)c2)c1
C=C1CC(CC(C)C2CCC3C(=CC=C4CC(O)CC(O)C4=C)CCCC32C)OC1=O

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=O)C(=C)C2(C)C)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=O)C(=C)C23CCCC3)CC(O)CC10
C=C1C(=O)OC(CC(C)C)C2CCC3C(=CC=C4CC(O)C(CCCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=O)C(=C)C2CCC)CC(O)C(CCCO)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=O)C(=C)C2CCC)CC(O)C(C)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=O)C(=C)C2CC)CC(O)C(C)C10
C=C1C(=O)OC(CC(C)C)C2CCC3C(=CC=C4CC(O)C(C)C(O)C4=C)CCCC32C)C1C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC10
C=C1C(=CC=C2CCC(C)CCCC(C)C(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=O)C(=C)C2CC)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=O)C(=C)C2CCCC)CC(O)C(CCCO)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=O)C(=C)C2CCCC)CC(O)C(OCCC)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=O)C(=C)C2CC)CC(O)C(C)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)O)C(=O)N2C)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC20C(=O)C(=C)C2CC(C)C)CC(O)C(C)C10
CC(CCCNS(=O)(=O)C(F)F)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCNS(=O)(=O)C1cccc1)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(=O)C(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)OC)n1C
CC(CCCNC(=O)CS)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC[Si](CC)(c1ccc(OCC(=O)C(C)C)c(C)c1)c1ccc(OCC2CO2)c(C)c1
CCC(=CC=CC(O)CC)CC1cccc(OCc2ccc(CO)c(CO)c2)c1
C=C1C(=CC=C2CCCC3(C)C(C(C)CCC(F)F)C(O)CC)CC(=CC23)CC(O)CC1C(F)F
C=C1C(O)CC(=CC=C2CCCC(C)CCCC(C)C)O)CC10
COCOC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)C)O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2cccc2)CC(O)CC10
CC(C=CC(=O)NO)=CC(C)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(C=CC(=O)Nc1cccc1N)=CC(C)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)O)C(=O)N2Cc2cc(OCC)cc(OCCC)c2)CC(O)CC10
CC(CCCNS(N)(=O)=O)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COC(=C)C2(C)CCCC2)CC(O)CC10
CC(CCCNS(C)(=O)=O)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCC(=O)NO)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C(COCC(C)C1CC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCC1(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC10
CC(CCCNS(C)(=O)=O)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCC(=O)O)C1CC2C3CC4CC(O)CCC4(C)C3CCC12C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)O)C(=O)N2CCCCc2cccc2)CC(O)CC10
CC(CCCC(=O)Nc1cccc1N)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CCC(CC)(c1ccc(OCC(O)C(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C
COC1ccc(C(Nc2ccc(C(F)F)cc2)c2c(C)[nH]c3cccc23)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)O)C(=O)N2Cc2ccc(OC)cc2)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)O)C(=O)N2CCCCc2cc(OC)cc(OC)c2)CC(O)CC10
CCC(CC)(c1ccc(OCC(=O)C(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)OC)n1C
Cc1ccc(-c2csc(=N)n2CC(=O)N2CCN(c3cccc3)CC2)cc1
CCCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
CCC(CC)(c1ccc(OCC(O)C(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)n1C
Nc1nc2ccc(S(=O)(=O)N3CCCC3)cc2s1
CC(C)S(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1cccc1
CC(=O)N1CCN(C(=O)Cn2c(-c3ccc(C)cc3)csc2=N)CC1
N=c1sc(-c2ccc(Cl)cc2)n1CC(=O)N1CCN(c2cccc2)CC1
Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1cccnc1
COC(=O)c1ccc2c(c1)C(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)CCCc2cccc2)c(C)c1
CN1CCN(C(=O)Cn2c(-c3ccc(Cl)cc3)csc2=N)CC1
Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1ncs1

COc1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1cccc1Cl
N=c1scc(-c2ccc(Cl)cc2)n1CC(=O)N1CCOCC1
COc1ccc(NC(c2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3cccc23)cc10C
CCC(CC)(c1ccc(OC(C)C(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
Cc1ccc2c(C(Nc3ccc(OC(F)(F)F)cc3)c3cccc3Cl)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(c2cccc2Cl)c2c(C)[nH]c3cccc23)cc1
CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
Cc1[nH]c2cccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
Cc1ccc2c(C(Nc3cccnc3)c3cccc3Cl)c(C)[nH]c2c1
CCC(CC)(c1ccc(OC(C)C(C)C)c(C)c1)c1ccc(OC(C)C(C)CCc2cccc2)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]c2cccc2c1C(Nc1cccc1)c1ccc(Cl)cc1
COC(=O)c1ccc(C=CC2=[N+](C)c3ccc(OC)cc3C2(C)C)cc1
Cc1[nH]c2cccc2c1C(Nc1cccc1)c1cccc1Cl
COc1ccc(NC(c2nccs2)c2c(C)[nH]c3cccc23)cc10C
COc1ccc(NC(c2cccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc10C

S4.8.3 Seeds

CCC(CC)(c1ccc(OC(C)C(C)C)c(C)c1)c1ccc(OC(C)C(C)C)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C(C)C)CCC23)CC1O
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C)C)SCCCC(O)(CC)CC=CCC23)CC1O
CCC(CC)(c1ccc(OC(=O)C(C)C)c(C)c1)c1ccc(OC(C)C(C)C)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C(C)C)CCC(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(c2cccc2Cl)c2c(C)[nH]c3cccc23)cc1
CCC(CC)(c1ccc(OC(=O)C(C)C)c(C)c1)c1ccc(C(=O)NC(CC(=O)O)C(=O)O)n1C
CCC(CC)(c1ccc(OC(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)c3cccc3Cl)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)CCC(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(c2cccc2Cl)c2c(C)[nH]c3cccc23)cc1C
C=C1C(O)CC(=CC=CCC(C)(C)CC(C)CCCC(C)(C)O)CC1O
C=C1C(=O)OC(C(C)C)C2CCC3C(=CC=C4CC(O)C(O)C4=C)CCCC32C)C1C
CCC(=CC=CC(O)(C(F)(F)F)C(F)(F)F)c1ccc(CCCc2ccc(CO)c(CO)c2)c1
Cc1[nH]c2cccc2c1C(Nc1ccccn1)c1ccc(C(C)C)cc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2C)CC(O)CC1O
CCC(CC)(c1ccc(OC(C)C(C)C)c(C)c1)c1ccc(C(=O)NCCC)O)n1C
CCC(CC)(c1ccc(OC(=O)C(C)C)c(C)c1)c1ccc(OC(=O)C(C)(C)C)c(C)c1
CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(OC(C)C(C)C)c(C)c1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cccc2)CC(O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)C)CC(=C)C(C)C)CC(O)CC1O

S4.9 Sample 9

S4.9.1 Training set

CDc1ccc(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(OCCNS(C)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CDc1ccc(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(OCCO)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(=CCO)C1O
 CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)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)OCCO)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(OCCO)C1O
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OCCO)cc(OCCO)c2)CC(O)CC1O
 CC(OCCNS(=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(=CCO)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(OCCC(=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(OCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCOC4C3)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(Cc1n2ccn2)C1O
 C=C1C(=CC=C2CCC(CCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(OCCC)CCO)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(Cc1n2ccn2)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)N3CCCC3)cc2s1
 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)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(OCCO)C1O
 CDc1ccc(NC(c2nccs2)c2c(C)[nH]c3ccccc23)cc1OC
 C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCO)C1O
 C=C1C(O)CC(=CC=CCCC(C)CCCC(C)(C)O)CC1O
 Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1cccn1

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)C2CCCC)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(O)c2ccc(CO)c(CO)c2)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C(C)CCCC(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(C)CCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC1O
Cc1ccc2c(C(Nc3ccnc3)c3ccccc3Cl)c(C)[nH]c2c1
CC[Si](CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(OCC2CO2)c(C)c1
CC(C)CCCCNS(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(O)CC1O
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(C)C(C)C(C)O)CC4)=CCC23)CC(O)CC1O
CCCCS(=O)(=O)NCCCC(C)C1CC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCN2cnmn2)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(C)CCC(C)C(C)O)CC(O)CC1O
COc1ccc2c(C(Nc3ccc(Cl)cc3)c3ccccc3Cl)c(C)[nH]c2c1
COc1cccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
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(CCN2cncn2)C1O
C=C1C(=O)OC(CC(C)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)c1cncnc1
C=C1C(O)CC(=CC=C2CCCC3C(C)CCCC(C)(C)O)CCC23)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCCC2ccccc2)CC(O)CC1O
Cc1[nH]c2ccccc2c1C(NC1ccccc1)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
CCOC1C(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)C)C(C)c1)c1ccc(OCC(O)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(OCC(O)C)C(C)c1
CC(C)CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCF)C(O)C3)CCCC21C

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C0c2ccc(C(C)(C)O)c2)CC(O)CC10
CCCC0C1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCCC(O)(CC)CC)=CCC23)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(C)(C)O)CC10
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC10
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCO)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C=CC(=O)OCC)CC2)CC(O)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C(C)CC(O)(CC)CC)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC10
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)CC10
C=C1CC(C(C)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(C(C4)C2)C3)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC(O)CC10
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)CC10
CCCc1cc(C(O)(CC)CC)ccc1-c1cc(OCCc2ccc(CO)c(CO)c2)ccc1CC
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CC(C)(C)O)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(C(C4)C2)C3)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)C(C)C)CC(O)CC10
C=C1CC(C(C)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)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC10
C=C1C(=O)OC(C(C)C)C2CCC3C(=CC=C4CC(O)C(CCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=O)OC(C(C)C)C2CCC3C(=CC=C4CC(O)C(CCO)C(O)C4=C)CCCC32C)C1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCO)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(OCCO)C10
C=C1C(=CC=C2CCC(C)CCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C10
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C(=CCO)C(O)C3)CCCC21C
CC(CCCNS(=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)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(CCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COC#CC(C)(C)C)CC(O)CC10
CCC(=CC=CC(O)(CC)CC)c1cccc(OCCc2ccc(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
C0C0CCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC2OCC=C2C10
C=C1C(O)CC(=CC=CCC(C)=CC(C)CCCC(C)(C)O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C0Cc2ccccc2)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C0CC(=C)C(C)(C)CC(O)CC10
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)NCCO)n1C
CC(CCCCC(=O)N)c1ccccc1N)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2ccccc2)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COC(=C)C2(C)CCCC2)CC(O)CC10
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(CO)CCO)n1C
CC(CCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCNS(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=CCCC1(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(OC)cc(OC)c2)CC(O)CC10
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)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)N2Cc2cc(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)COCC(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(NC3ccccc3)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
 CCS(=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(NC3ccc(C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
 COc1ccc(C(Nc2ccccc2)c2c(C)[nH]c3ccccc23)cc1

S4.9.3 Seeds

C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2ncnm2)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)CCO)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(C)CCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC1O
 C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC1O
 CCCCS(=O)(=O)NCCCC(C)C1CC2C(=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)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(O)C2ccc(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(CCN2cnm2)C1O
 CCn1c(C(=O)NC(C)C(=O)O)ccc1C(CC)(CC)c1ccc(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(O)CC(C)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(O)C(C)(C)C)c(C)c1
 CCC(CC)(c1ccc(O)C(CO)C)c(C)c1)c1ccc(O)C(C)(C)C)c(C)c1

S4.10 Sample 10

S4.10.1 Training set

Cc1[nH]c2cccc2c1C(Nc1cccc1)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)COc2cccc2)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)CCO)CC1O
 Nc1nc2ccc(S(=O)(=O)N3CCCC3)cc2s1
 CCC(CC)(c1ccc(O)C(CO)C)c(C)c1)c1ccc(O)C(C)(C)CCc2cccc2)c(C)c1
 CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1cccc1
 CCC(CC)(c1ccc(O)C(C)(C)C)c(C)c1)c1ccc(O)C(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)CCCC2)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)N2CCc2cc(O)cc(O)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(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(CCN2cnm2)C1O
 CC(CCCNC(=O)CS)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
 CCC(CC)(c1ccc(O)C(CO)C)c(C)c1)c1ccc(O)C(C)(C)CCc2cccc2)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(O)C(CO)C1O
 Cc1ccc(NC(c2cccc2Cl)c2c(C)[nH]c3cccc23)cc1
 Cc1[nH]c2cccc2c1C(Nc1cccc1)c1ccc(Cl)cc1
 CC[Si](CC)(c1ccc(O)C(C)(C)C)c(C)c1)c1ccc(O)C2CO2)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(c1ccnc1)c1c(C)[nH]c2cc(C)ccc12
 CC(CCCNS(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(O)C(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCC(=O)O)n1C
 COc1ccc(C(Nc2ccc(C(F)(F)F)cc2)c2c(C)[nH]c3cccc23)cc1
 CC(C)CS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccc(F)cc1
 CCn1c(C(=O)NC(CO)CCO)ccc1C(CC)(CC)c1ccc(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)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C23CCCC3)CC(O)CC10
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)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2ccccc2)CC(O)CC10
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(CCCC(C)(C)O)CCC23)CC10
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)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)CC10
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)N2Cc2ccccc2)CC(O)CC10
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOCC(O)C(O)COCC(C(=O)O)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)CC10
CC(CCCC(=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(CCn2cnmn2)C10
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(=CCO)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)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)C23CC4CC(CC(C4)C2)C3)CC10
C=C1C(=CC=C2CCC(CCCCCC(C)(C)O)C3(C)C2CCC3C(C)CCO)CC(O)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CC#CC(O)CC23CC4CC(CC(C4)C2)C3)CC10
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(OCCO)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2CCc2cc(O)cc(O)C2)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(C)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCC(C)(C)O)CC(O)CC10
C=C1CCC(O)CC1=CC=C1CCCC2(C)C1CCC2C(C)CCCC(C)C
Cc1[nH]c2ccccc2c1C(Nc1ccc(N(C)C)cc1)c1ccccc1C1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(O)CC(O)CC10
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)CC10
CCCCS(=O)(=O)NCCCC(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)NCC(=O)O)n1C
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC)O)n1C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(CCn2nccn2)C10
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)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)C(C)C10
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)CC10
CC(CCCC(C)(C)O)C1CCC2C(=CC=C3CC(O)C4=CCOC4C3)CCCC21C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)CC23CC4CC(CC(C4)C2)C3)CC10
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)CC10
C=C1C(=O)OC(CC(C)C2CCC3C(=CC=C4CC(O)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)c3ccccc3C1)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)C0CC#CC(C)(C)C)CC(O)CC10

S4.10.2 Testing set

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

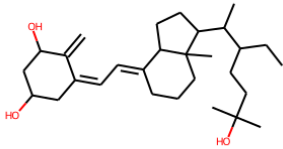
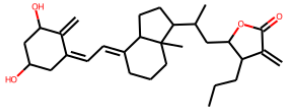
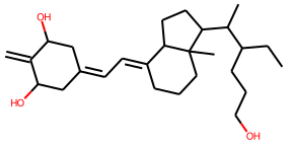
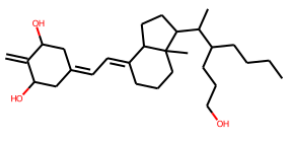
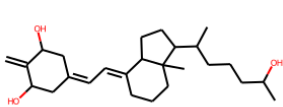
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc2n2cmm2)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C23CC4CC(CC(C4)C2)C3)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)(C)CCC(C)(C)O)CC10
CCC(O)(CC)CSC(C)C1=CC2C(=CC=C3CC(O)C(O)C3)CCCC12C
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)(CC)CC)CC10
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC10
C=C1C(O)CC(=CC=C2CCCC3C(C)CCCC(C)(C)O)CCC23)CC10
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)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CSc2ccc(C(C)(C)O)cc2)CC(O)CC10
CCCOC1C(O)CC(=CC=C2CCCC3(C)C(C(C)SCCC(O)(CC)CC)=CCC23)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)OCC(C)C(C)(C)O)CC10
C=C1C(=CC=C2CCCC3(C)C(C(C)CCCC(C)(C)O)CCC23C)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc2c2cmm2)C10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)(C)CC)CC(O)C(=CCCO)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C=CC(O)C2(C(=O)O)CCCC)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC(C)C)CC(O)C(=CCCO)C10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)(CC)CC)CC10
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)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)C(Cc2c2cnc2)C10
C=C1CC(CC(C)C)C2CCC3C(=CC=C4CC(O)C(=C)C(O)C4)CCCC32C)OC1=O
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CC(O)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C=CCC(O)C23CC4CC(CC(C4)C2)C3)CC10
Cc2n1c(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)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(C)(C)O)CC10
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)CC10
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)CC10
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CC(O)CC10
C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC10
C=C1C(=CC=C2CCC(C)CCCC(C)(C)O)C3(C)C2CCC3C(C)CC(O)CC10
C=C1CC(CC(C)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)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(CCCO)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)C(OCCCC)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(C)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(O)C(C)C)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCCC)CC(O)CC10
C=C1C(=CC=C2CCCC(C)CCCC(C)(C)O)C3(C)C2CCC3C(C)CO)CC(O)CC10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)C(CCCO)C10
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC)CC(O)CC10

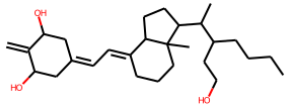
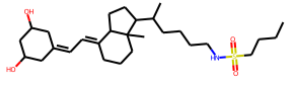
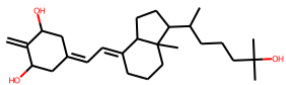
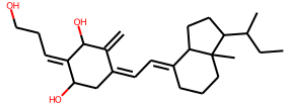
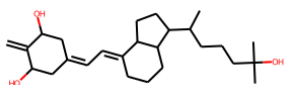
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(CCCO)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(OCCCC)C1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CC(C)C)CC(O)C(C)C1O
CC(CCCNS(=O)(=O)C(F)(F)F)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCNS(=O)(=O)c1ccccc1)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
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=C3CC(C)CCC(C)(C)O)CC1O
COOCCC=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCC(C)(C)O)CC1O
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCC(C)(C)O)CC2OCC=C2C1O
C=C1C(O)CC(=CC=C3CC(C)=CC(C)CCC(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(OCC)cc(OCC)c2)CC(O)CC1O
CC(CCCNS(N)(=O)=O)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(CCCC(=O)NO)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
C=C(COCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3=C)CCCC21C)C1=CCCC1(C)C
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2CC(C)(O)C(=O)N2Cc2cc(O)cc(O)C2)CC(O)CC1O
CC(CCCNS(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(C(=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(O)cc2)CC(O)CC1O
CCC(CC)(c1ccc(OCC(=O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCC(=O)O)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)N2CCCCc2ccccc2)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)(=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(O)C(O)C2)c2c(C)[nH]c3ccccc23)cc1
COC(=O)c1ccc2c(c1)C(C)(C)C(C=Cc1ccc(O)cc1)=[N+]2C
Cc1ccc(CNC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncs1
COc1ccc(C=CC2=[N+](C)c3ccc(O)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(O)C(F)(F)F)cc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1[nH]c2ccccc2c1C(NC1ccccc1)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)O)cc3C2(C)C)cc1
COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1
Cc1ccc2c(C(NC3ccccc3)c3ccccc3Cl)c(C)[nH]c2c1
Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ncc[nH]1
Cc1ccc2c(C(Nc3ccccc3)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(O)cc3C2(C)C)cc1
Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl
COc1ccccc1NC(c1ccccc1Cl)c1c(C)[nH]c2ccccc12
COc1ccc(NC(c2nccs2)c2c(C)[nH]c3ccccc23)cc1OC

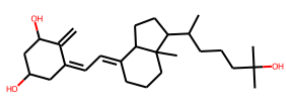
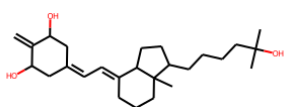
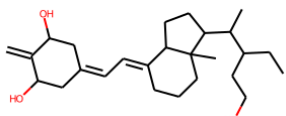
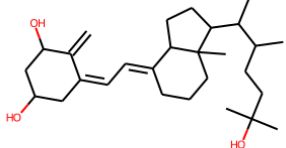
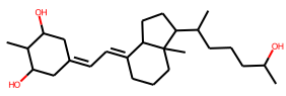
CC1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3cc(C)ccc23)cc1OC
CC1ccc(C(Nc2cccn2)c2c(C)[nH]c3ccccc23)cc1

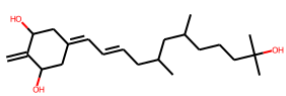
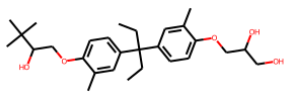
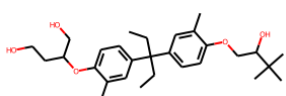
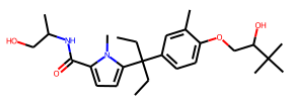
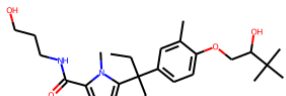
S4.10.3 Seeds

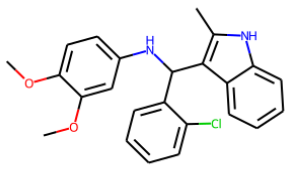
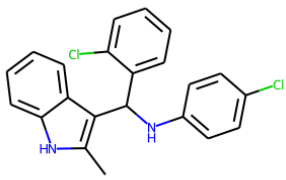
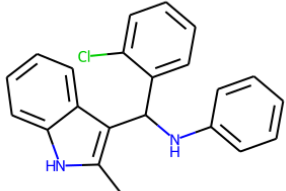
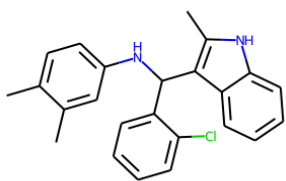
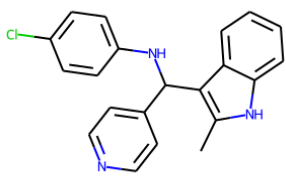
CCC(O)(CC)CSC(C)C1=CCC2C(=CC=C3CC(O)C(C)(O)C(O)C3)CCCC12C
CCS(=O)(=O)c1oc(-c2cccs2)nc1S(=O)(=O)c1ccccc1
C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)COCc2ccccc2)CC(O)CC1O
CCCCS(=O)(=O)NCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C
CC(C)CC(N)C(=O)NC(CC(C)C)C(=O)NC1CCOC(O)C(O)COC(C(=O)O)NC(=O)C(CC(C)C)NC(=O)C(CC(C)C)NC(=O)C(C)(C)NC1=O
Cc1ccc2c(C(NC3ccc(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)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
CC1ccc(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)COC(=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(Nc1cccn1)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)NCCC)nc1C
CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OCC(O)CCO)c(C)c1
C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O
Nc1nc2ccc(S(=O)(=O)N3CCCC3)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 |
|--|---|
| <chem>C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCC(C)(C)O)CC(O)CC1O</chem> |  |
| <chem>C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC2OC(=O)C(=C)C2CCC)CC(O)CC1O</chem> |  |
| <chem>C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCCO)CC1O</chem> |  |
| <chem>C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCCC)CCCO)CC1O</chem> |  |
| <chem>C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)O)CC1O</chem> |  |

| | |
|--|---|
| <chem>C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CCO)CCCC)CC1O</chem> |  |
| <chem>CCCCS(=O)(=O)NCCCC(C)C1CCC2C(=CC=C3CC(O)CC(O)C3)CCCC21C</chem> |  |
| <chem>C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC1O</chem> |  |
| <chem>C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CC)CC(O)C(=CCCO)C1O</chem> |  |
| <chem>C=C1C(O)CC(=CC=C2CCCC3C2CCC3C(C)CCCC(C)(C)O)CC1O</chem> |  |

| | |
|---|---|
| <chem>C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)CCCC(C)(C)O)CC(O)CC1O</chem> |  |
| <chem>C=C1C(O)CC(=CC=C2CCCC3(C)C(CCCCC(C)(C)O)CCC23)CC1O</chem> |  |
| <chem>C=C1C(O)CC(=CC=C2CCCC3(C)C2CCC3C(C)C(CC)CCO)CC1O</chem> |  |
| <chem>C=C1C(=CC=C2CCCC3(C)C2CCC3C(C)C(C)CCC(C)(C)O)CC(O)CC1O</chem> |  |
| <chem>CC(O)CCCC(C)C1CCC2C(=CC=C3CC(O)C(C)C(O)C3)CCCC21C</chem> |  |

| | |
|---|---|
| <chem>C=C1C(O)CC(=CC=CCC(C)CC(C)CCCC(C)(C)O)CC1O</chem> |  |
| <chem>CCC(CC)(c1ccc(OCC(O)CO)c(C)c1)c1ccc(OCC(O)C(C)(C)C)c(C)c1</chem> |  |
| <chem>CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(OC(CO)CCO)c(C)c1</chem> |  |
| <chem>CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NC(C)CO)n1C</chem> |  |
| <chem>CCC(CC)(c1ccc(OCC(O)C(C)(C)C)c(C)c1)c1ccc(C(=O)NCCCO)n1C</chem> |  |

| | |
|--|---|
| <chem>COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1OC</chem> |  |
| <chem>Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccccc1Cl</chem> |  |
| <chem>Cc1[nH]c2ccccc2c1C(Nc1ccccc1)c1ccccc1Cl</chem> |  |
| <chem>Cc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1C</chem> |  |
| <chem>Cc1[nH]c2ccccc2c1C(Nc1ccc(Cl)cc1)c1ccncc1</chem> |  |

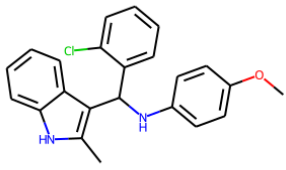
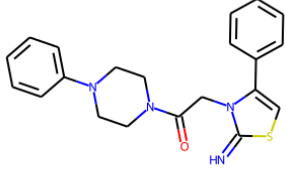
| | |
|--|---|
| <chem>COc1ccc(NC(c2ccccc2Cl)c2c(C)[nH]c3ccccc23)cc1</chem> |  |
| <chem>N=c1sc(-c2ccccc2)n1CC(=O)N1CCN(c2ccccc2)CC1</chem> |  |

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.