

Supporting Information for

Augmented Hill-Climb increases reinforcement learning efficiency for language-based *de novo* molecule generation

Morgan Thomas¹, Noel M. O'Boyle², Andreas Bender^{1*} and Chris de Graaf^{2*}

¹ Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, UK

² Computational Chemistry, Sosei Heptares, Steinmetz Building, Granta Park, Great Abington, Cambridge, CB21 6DG, UK

E-mail: ab454@cam.ac.uk, chris.degraaf@soseiheptares.com

Supplementary Tables

Table S1: Intrinsic metrics of 10,000 sampled *de novo* molecules from the recurrent neural network (RNN), Transformer (Tr) and Gated Transformer (GTr) used in this work when trained on the GuacaMol training dataset. These metrics include the ratio of valid, unique, scaffold unique and novel molecules, as well as sphere exclusion diverse (ECFP4 of length 1,204 with sphere radius equal to Tanimoto distance of 0.65), functional group diverse (FG) and ring system diverse (RS). Lastly, the ratio of molecules passing MOSES filters.

Model	Valid (↑)	Unique (↑)	Novel (↑)	SEDiv@1k (↑)	Scaffold Unique (↑)	FG (↑)	RS (↑)	MOSES Filters (↑)
RNN	0.96	0.99	0.96	0.88	0.84	0.20	0.18	0.52
Tr	0.96	1.00	0.97	0.85	0.86	0.18	0.17	0.52
GTr	0.94	0.99	0.97	0.89	0.87	0.21	0.18	0.50

Table S2: Extrinsic metrics of 10,000 sampled *de novo* molecules from the recurrent neural network (RNN), Transformer (Tr) and Gated Transformer (GTr) used in this work when trained on the GuacaMol training dataset. These metrics measure the similarity of *de novo* molecules to the GuacaMol test dataset. Including the analogue similarity (ratio of *de novo* molecules that have test set analogues), analogue coverage (ratio of test set with *de novo* analogues), average single nearest neighbour similarity (SNN), cosine similarity of functional groups (FG) and ring systems (RS), fragments (Frag) and scaffolds (Scaff), and Wasserstein distance of LogP, natural product likeness (NP), synthetic accessibility (SA), quantitative estimate of drug-likeness (QED) and molecular weight distributions. Note that fingerprints similarities are based on ECFP4 fingerprints of length 1,024 and analogues are defined as having a Tanimoto similarity greater than 0.4.

Model	Analogue Similarity (↑)	Analogue Coverage (↑)	SNN (↑)	FG (↑)	RS (↑)	Frag (↑)	Scaff (↑)	logP (↓)	NP (↓)	SA (↓)	QED (↓)	Weight (↓)
RNN	0.83	0.43	0.52	1.00	1.00	1.00	0.64	0.02	0.04	0.08	0.01	13.01
Tr	0.82	0.40	0.51	1.00	1.00	1.00	0.59	0.07	0.08	0.04	0.01	2.90
GTr	0.76	0.36	0.49	1.00	1.00	1.00	0.55	0.07	0.07	0.03	0.00	3.47

Table S3: Default hyperparameters used for reinforcement learning strategies benchmarked in this work. Additional configurations with regularization and disclosed in square brackets.

RL strategy	Training steps	Batch size	σ	k	λ_{KL}	lr	α
REINFORCE [+reg]	500	64	-	-	[10]	1×10^{-4}	-
REINVENT	500	64	60	-	-	5×10^{-4}	-
REINVENT 2.0	250	128	120	-	-	5×10^{-4}	-
BAR	500	64	60	-	-	5×10^{-4}	0.5
Augmented Hill-Climb	500	64	60	50%	-	5×10^{-4}	-
Hill-Climb [+reg]	32	1024	-	50%	[10]	5×10^{-4}	-
Hill-Climb* [+reg]	500	64	-	50%	[10]	5×10^{-4}	-

Table S4: Diversity filter configurations used in this work.

	DF1	DF2	DF3
Topology simplification	BM scaffold	BM scaffold	BM scaffold
Fingerprint type	ECFP (Morgan)	ECFP (Morgan)	ECFP (Morgan)
Fingerprint radius	2	2	2
Fingerprint bits	1024	1024	1024
Similarity metric	Tanimoto	Tanimoto	Tanimoto
Minimum score threshold	0.8	0.5	0.0
Bin size	25	50	50
Output mode	Binary	Linear	Linear

Table S5: Number of molecules downloaded from ExCAPE-DB and those docked against targets used in this work to assess retrospective performance.

	ExCAPE-DB		Docked	
	Active	Inactive	Active	Inactive
DRD2	4613	343076	3734	9538
OPRM1	3128	2786	3125	2573
AGTR1	671	558	270	516
OX1R	681	322795	564	9860

Supplementary Figures

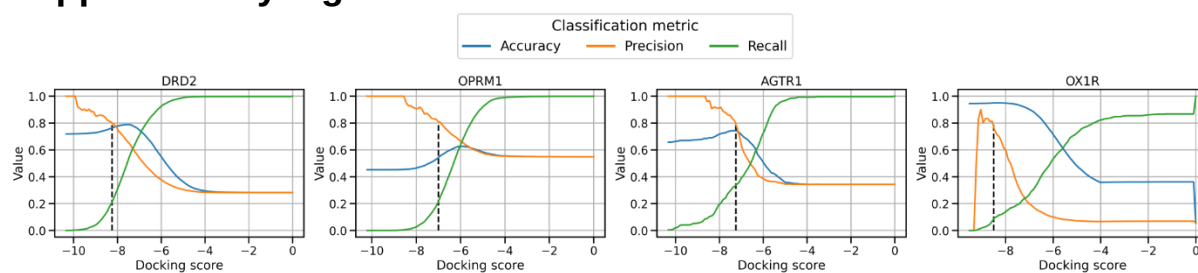


Figure S1: Retrospective classification performance of docking protocol on the four targets investigated in this work. In each case, the accuracy, precision and recall are traced with varying docking score decision thresholds. Thresholds corresponding to ~ 80% precision are annotated by black dashed lines.

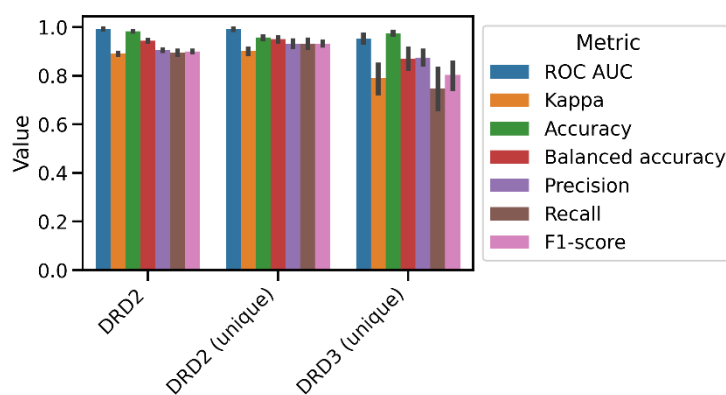


Figure S2: QSAR model performance of RF models trained on DRD2 and DRD3 active and inactive molecules, based on 5-fold stratified, clustered cross-validation with GHOST [85] decision threshold identification.

Sigma	RL strategy	Top 1	Top 2	Top 3	Top 4	Top 5
30	REINVENT	 CS: 5 DS: -9.13 AvDS: -9.53	 CS: 4 DS: -9.16 AvDS: -9.26	 CS: 4 DS: -9.28 AvDS: -9.79	 CS: 3 DS: -9.30 AvDS: -9.41	 CS: 2 DS: -9.10 AvDS: -9.36
	Augmented Hill-Climb	 CS: 14 DS: -10.39 AvDS: -10.60	 CS: 7 DS: -10.34 AvDS: -10.57	 CS: 5 DS: -10.32 AvDS: -10.67	 CS: 4 DS: -10.40 AvDS: -10.50	 CS: 4 DS: -10.48 AvDS: -10.53
60	REINVENT	 CS: 8 DS: -9.10 AvDS: -9.28	 CS: 3 DS: -9.12 AvDS: -9.37	 CS: 3 DS: -9.42 AvDS: -9.53	 CS: 2 DS: -9.09 AvDS: -9.50	 CS: 2 DS: -9.12 AvDS: -9.50
	Augmented Hill-Climb	 CS: 19 DS: -11.48 AvDS: -11.64	 CS: 5 DS: -11.43 AvDS: -11.52	 CS: 3 DS: -11.16 AvDS: -11.25	 CS: 3 DS: -11.28 AvDS: -11.54	 CS: 3 DS: -11.43 AvDS: -11.45
90	REINVENT	 CS: 4 DS: -9.53 AvDS: -9.78	 CS: 4 DS: -9.56 AvDS: -9.84	 CS: 3 DS: -9.59 AvDS: -9.78	 CS: 3 DS: -9.66 AvDS: -9.87	 CS: 2 DS: -9.39 AvDS: -9.54
	Augmented Hill-Climb	 CS: 10 DS: -11.21 AvDS: -11.38	 CS: 7 DS: -11.23 AvDS: -11.51	 CS: 7 DS: -11.53 AvDS: -11.91	 CS: 5 DS: -11.24 AvDS: -11.36	 CS: 4 DS: -11.33 AvDS: -11.38
120	REINVENT	 CS: 5 DS: -9.95 AvDS: -9.84	 CS: 4 DS: -9.29 AvDS: -9.77	 CS: 2 DS: -9.28 AvDS: -9.40	 CS: 2 DS: -9.30 AvDS: -9.41	 CS: 2 DS: -9.34 AvDS: -9.42
	Augmented Hill-Climb	 CS: 21 DS: -9.28 AvDS: -9.64	 CS: 9 DS: -9.35 AvDS: -9.91	 CS: 7 DS: -9.28 AvDS: -9.58	 CS: 6 DS: -9.84 AvDS: -10.16	 CS: 5 DS: -9.28 AvDS: -9.52
180	REINVENT	 CS: 4 DS: -9.19 AvDS: -9.26	 CS: 4 DS: -9.35 AvDS: -9.57	 CS: 2 DS: -9.10 AvDS: -9.15	 CS: 2 DS: -9.12 AvDS: -9.25	 CS: 2 DS: -9.16 AvDS: -9.64
	Augmented Hill-Climb	 CS: 8 DS: -12.80 AvDS: -12.94	 CS: 7 DS: -12.77 AvDS: -12.83	 CS: 6 DS: -12.58 AvDS: -12.84	 CS: 5 DS: -12.76 AvDS: -13.03	 CS: 4 DS: -12.60 AvDS: -12.94
240	REINVENT	 CS: 3 DS: -8.75 AvDS: -8.81	 CS: 3 DS: -9.04 AvDS: -9.24	 CS: 2 DS: -8.74 AvDS: -8.86	 CS: 2 DS: -8.77 AvDS: -8.94	 CS: 2 DS: -8.80 AvDS: -8.96
	Augmented Hill-Climb	 CS: 28 DS: -13.46 AvDS: -13.77	 CS: 19 DS: -13.85 AvDS: -13.96	 CS: 5 DS: -13.43 AvDS: -13.92	 CS: 5 DS: -13.75 AvDS: -13.89	 CS: 4 DS: -13.48 AvDS: -13.69

Figure S3: Centroid of the top 5 largest clusters for the top 100 molecules generated by REINVENT and Augmented Hill-Climb (both with DF1) according to docking score against DRD2 receptor for varying values of sigma (σ). Cluster size (CS), centroid docking score (DS) and the average cluster docking score (AvDS) is annotated below.

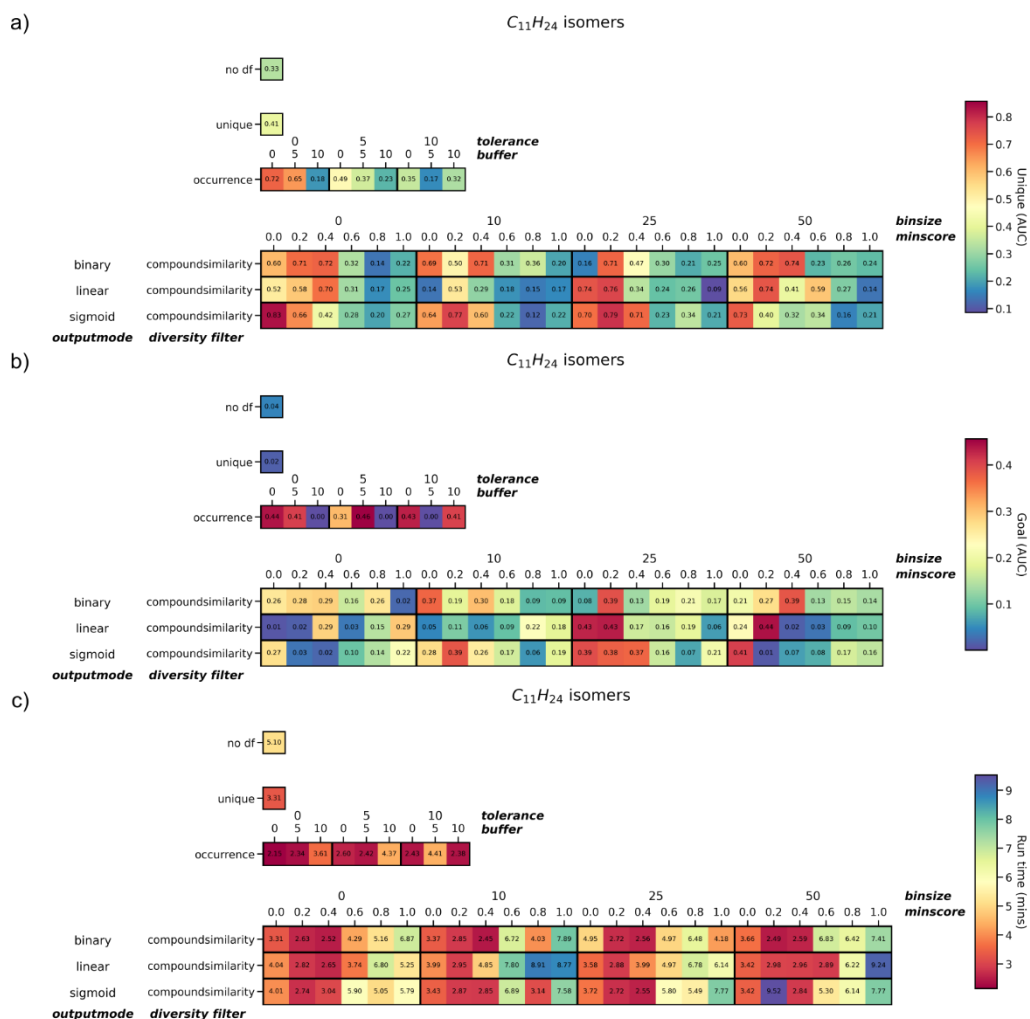


Figure S5: Optimization of $C_{11}H_{24}$ isomers task with different diversity filters and their parameters. Three endpoints are measured: (a) unique area under the curve (AUC), (b) goal AUC (c) and run time. Note: scaffold-based diversity filters aren't shown as $C_{11}H_{24}$ isomers cannot form rings and satisfy the molecular formula and so no scaffolds will be detected.

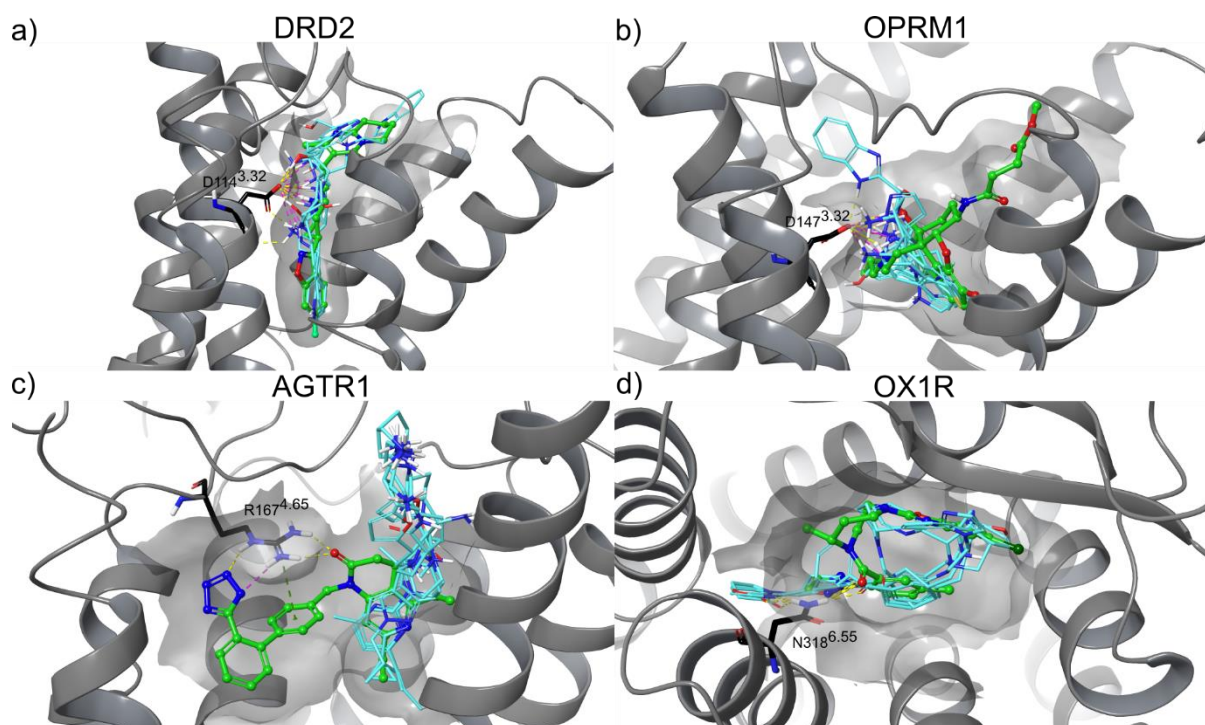


Figure S7: Docked poses of the centroid molecules (cyan) shown in Figure 5 generated by AHC + DF2 compared to the co-crystallized ligand (green) for each respective target. For (a) DRD2 and (b) OPRM1, RL and docking results in sensible pose generation satisfying crucial residue interactions with D114^{3.32}. While for (c) AGTR1 poses occupy only one relevant sub-pocket compared to the co-crystal ligand and form no interactions with R167^{4.65}. Only one pose for (d) OX1R mimics the horseshoe shape adopted by the co-crystal ligand, however, most form an interaction with N318^{6.55}.

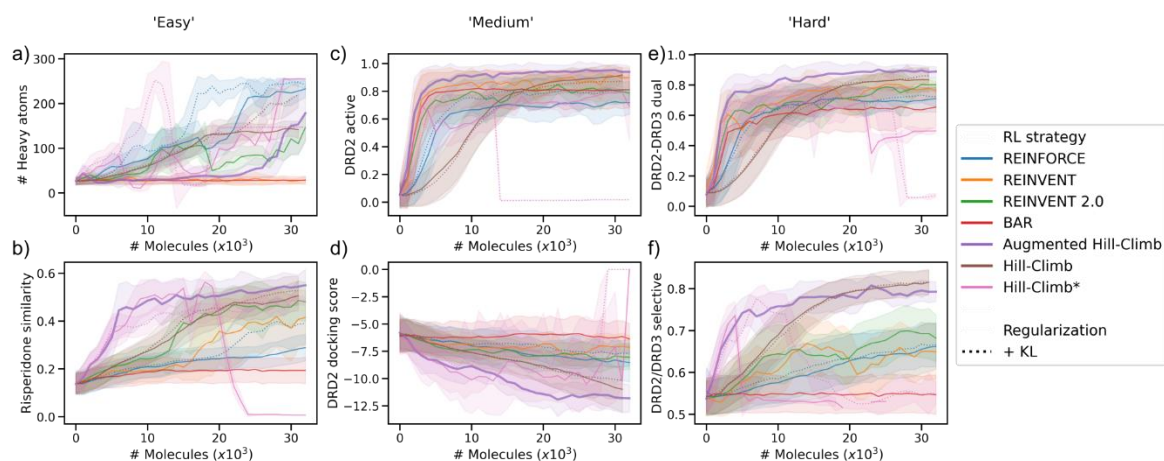


Figure S8: Per-molecule optimization for objective optimization using different RL strategies. In all cases, except the number of heavy atoms, AHC outperforms all other RL strategies with respect to objective optimization while maintaining validity and uniqueness. Only valid molecules are plotted, therefore gaps seen with HC* denote regions where no valid molecules were generated.

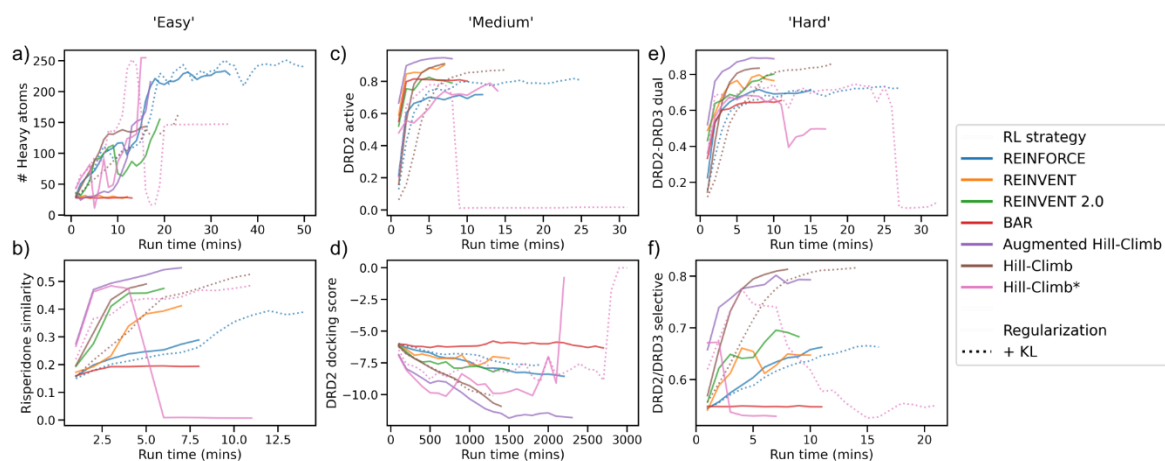


Figure S9: Wall time for objective optimization using different RL strategies. Run using an AMD Threadripper 1920x CPU and Nvidia GeForce RTX 2060 super GPU. The docking tasks were parallelized over 10 CPU cores while all other tasks used only 1 CPU core.

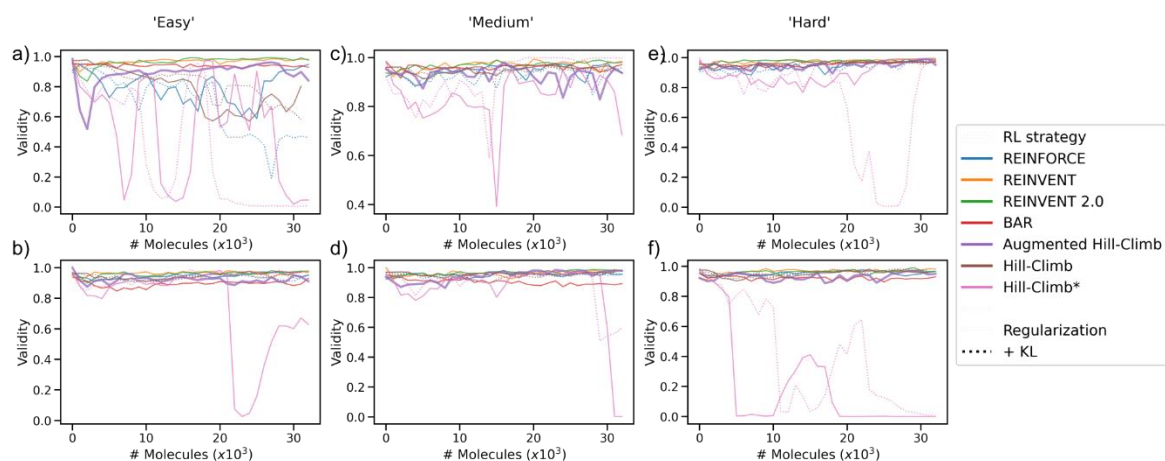


Figure S10: Validity for objective optimization using different RL strategies. HC* suffers from a drop in validity.

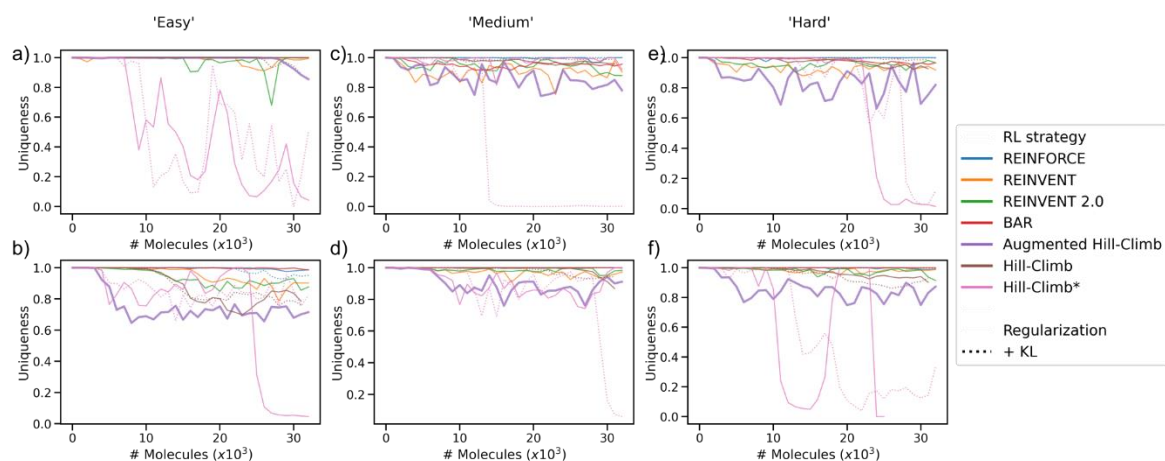


Figure S11: Uniqueness for objective optimization using different RL strategies. HC* suffers from a drop in uniqueness.

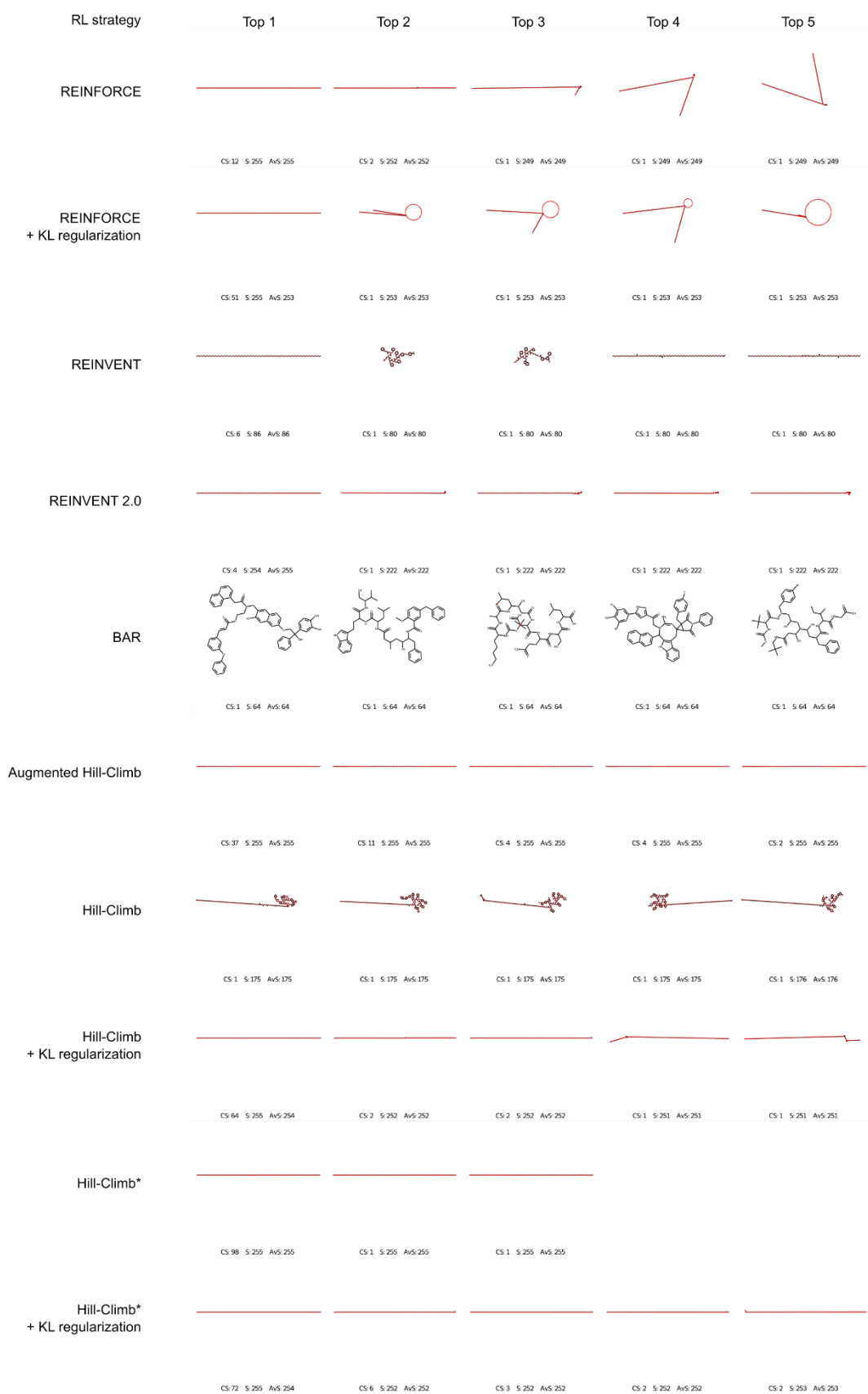


Figure S12: Centroid of the top 5 largest clusters for the top 100 molecules in the heavy atom task for different RL strategies. Cluster size (CS), centroid score (S) and the average cluster score (AvS) is annotated below.

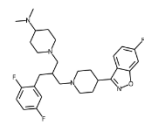
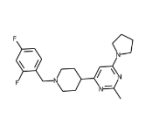
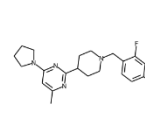
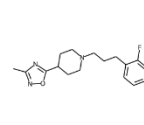
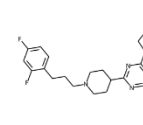
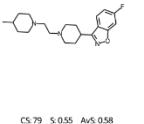
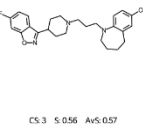
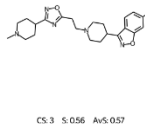
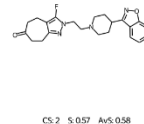
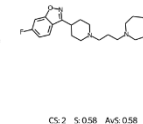
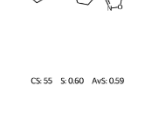
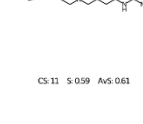
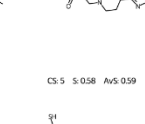
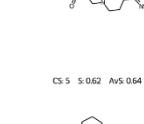
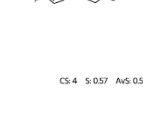
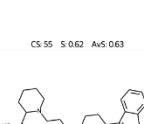
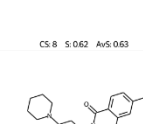
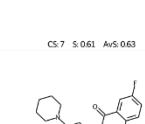
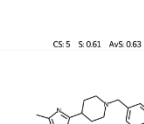
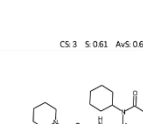
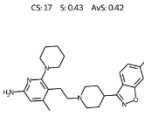
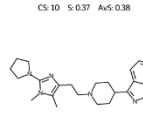
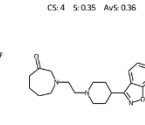
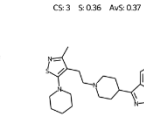
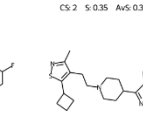
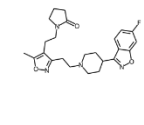
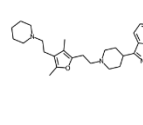
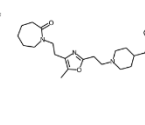
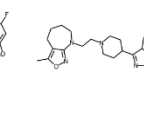
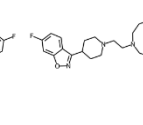
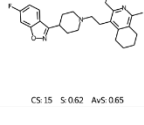
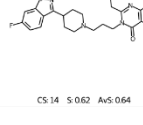
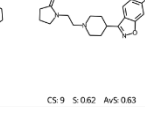
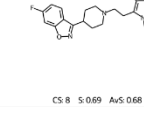
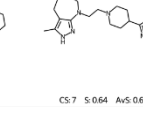
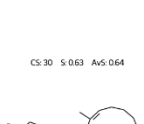
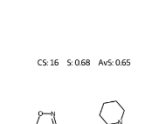
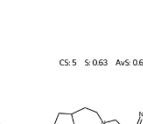
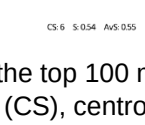
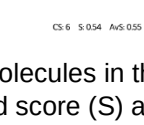
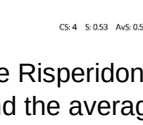
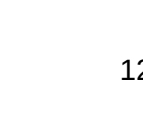
RL strategy	Top 1	Top 2	Top 3	Top 4	Top 5
REINFORCE	 CS: 9 S: 0.43 AvS: 0.47	 CS: 5 S: 0.41 AvS: 0.43	 CS: 4 S: 0.41 AvS: 0.42	 CS: 4 S: 0.42 AvS: 0.43	 CS: 3 S: 0.41 AvS: 0.43
REINFORCE + KL regularization	 CS: 79 S: 0.55 AvS: 0.58	 CS: 3 S: 0.56 AvS: 0.57	 CS: 3 S: 0.56 AvS: 0.57	 CS: 2 S: 0.57 AvS: 0.58	 CS: 2 S: 0.58 AvS: 0.58
REINVENT	 CS: 55 S: 0.60 AvS: 0.59	 CS: 11 S: 0.59 AvS: 0.61	 CS: 5 S: 0.58 AvS: 0.59	 CS: 5 S: 0.62 AvS: 0.64	 CS: 4 S: 0.57 AvS: 0.58
REINVENT 2.0	 CS: 55 S: 0.62 AvS: 0.63	 CS: 8 S: 0.62 AvS: 0.63	 CS: 7 S: 0.61 AvS: 0.63	 CS: 5 S: 0.61 AvS: 0.63	 CS: 3 S: 0.61 AvS: 0.61
BAR	 CS: 17 S: 0.43 AvS: 0.42	 CS: 10 S: 0.37 AvS: 0.38	 CS: 4 S: 0.35 AvS: 0.36	 CS: 3 S: 0.36 AvS: 0.37	 CS: 2 S: 0.35 AvS: 0.35
Augmented Hill-Climb	 CS: 32 S: 0.65 AvS: 0.66	 CS: 21 S: 0.64 AvS: 0.64	 CS: 19 S: 0.65 AvS: 0.65	 CS: 6 S: 0.67 AvS: 0.67	 CS: 5 S: 0.63 AvS: 0.64
Hill-Climb	 CS: 40 S: 0.64 AvS: 0.63	 CS: 13 S: 0.61 AvS: 0.63	 CS: 7 S: 0.65 AvS: 0.63	 CS: 5 S: 0.64 AvS: 0.63	 CS: 4 S: 0.60 AvS: 0.61
Hill-Climb + KL regularization	 CS: 15 S: 0.62 AvS: 0.65	 CS: 14 S: 0.62 AvS: 0.64	 CS: 9 S: 0.62 AvS: 0.63	 CS: 8 S: 0.69 AvS: 0.68	 CS: 7 S: 0.64 AvS: 0.63
Hill-Climb*	 CS: 30 S: 0.63 AvS: 0.64	 CS: 16 S: 0.68 AvS: 0.65	 CS: 15 S: 0.62 AvS: 0.63	 CS: 7 S: 0.63 AvS: 0.64	 CS: 5 S: 0.63 AvS: 0.64
Hill-Climb* + KL regularization	 CS: 42 S: 0.54 AvS: 0.55	 CS: 26 S: 0.54 AvS: 0.54	 CS: 6 S: 0.54 AvS: 0.55	 CS: 6 S: 0.54 AvS: 0.55	 CS: 4 S: 0.53 AvS: 0.54

Figure S13: Centroid of the top 5 largest clusters for the top 100 molecules in the Risperidone similarity task for different RL strategies. Cluster size (CS), centroid score (S) and the average cluster score (AvS) is annotated below.

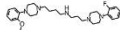
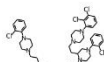
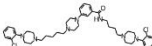
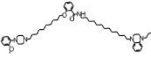
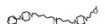
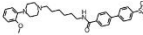
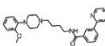
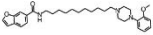
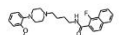
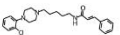
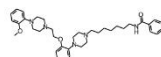
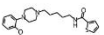
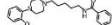
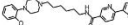
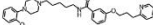
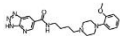
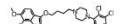
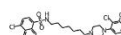
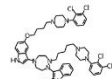
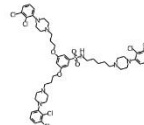
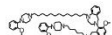
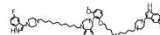
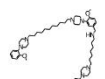
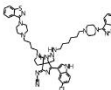
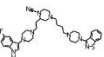
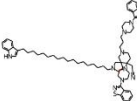
RL strategy	Top 1	Top 2	Top 3	Top 4	Top 5
REINFORCE	 CS: 30 S: 0.88 AvS: 0.91	 CS: 4 S: 0.87 AvS: 0.88	 CS: 4 S: 0.87 AvS: 0.89	 CS: 4 S: 0.91 AvS: 0.90	 CS: 2 S: 0.87 AvS: 0.88
REINFORCE + KL regularization	 CS: 14 S: 0.94 AvS: 0.95	 CS: 8 S: 0.94 AvS: 0.94	 CS: 8 S: 0.94 AvS: 0.94	 CS: 7 S: 0.94 AvS: 0.94	 CS: 6 S: 0.94 AvS: 0.95
REINVENT	 CS: 12 S: 0.98 AvS: 0.98	 CS: 9 S: 0.98 AvS: 0.98	 CS: 8 S: 0.98 AvS: 0.98	 CS: 5 S: 0.98 AvS: 0.98	 CS: 5 S: 0.98 AvS: 0.98
REINVENT 2.0	 CS: 19 S: 0.94 AvS: 0.97	 CS: 14 S: 0.94 AvS: 0.94	 CS: 8 S: 0.94 AvS: 0.94	 CS: 8 S: 0.95 AvS: 0.94	 CS: 6 S: 0.94 AvS: 0.94
BAR	 CS: 26 S: 0.97 AvS: 0.98	 CS: 14 S: 0.97 AvS: 0.97	 CS: 10 S: 0.98 AvS: 0.98	 CS: 4 S: 0.97 AvS: 0.97	 CS: 4 S: 0.98 AvS: 0.98
Augmented Hill-Climb	 CS: 12 S: 0.98 AvS: 0.98	 CS: 8 S: 0.98 AvS: 0.98	 CS: 8 S: 0.99 AvS: 0.98	 CS: 7 S: 0.98 AvS: 0.98	 CS: 5 S: 0.98 AvS: 0.98
Hill-Climb	 CS: 16 S: 0.97 AvS: 0.97	 CS: 9 S: 0.97 AvS: 0.97	 CS: 6 S: 0.96 AvS: 0.97	 CS: 4 S: 0.96 AvS: 0.97	 CS: 4 S: 0.96 AvS: 0.97
Hill-Climb + KL regularization	 CS: 19 S: 0.94 AvS: 0.94	 CS: 16 S: 0.94 AvS: 0.94	 CS: 10 S: 0.94 AvS: 0.94	 CS: 3 S: 0.94 AvS: 0.94	 CS: 3 S: 0.94 AvS: 0.94
Hill-Climb*	 CS: 32 S: 0.85 AvS: 0.85	 CS: 20 S: 0.85 AvS: 0.87	 CS: 11 S: 0.85 AvS: 0.88	 CS: 8 S: 0.87 AvS: 0.87	 CS: 4 S: 0.86 AvS: 0.86
Hill-Climb* + KL regularization	 CS: 97 S: 0.85 AvS: 0.85	 CS: 1 S: 0.85 AvS: 0.85	 CS: 1 S: 0.85 AvS: 0.85	 CS: 1 S: 0.86 AvS: 0.86	

Figure S14: Centroid of the top 5 largest clusters for the top 100 molecules in the DRD2 activity task for different RL strategies. Cluster size (CS), centroid score (S) and the average cluster score (AvS) is annotated below.

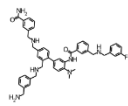
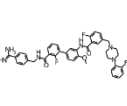
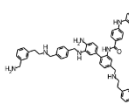
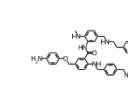
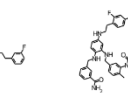
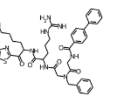
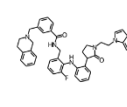
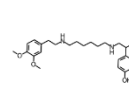
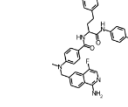
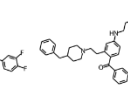
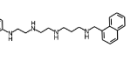
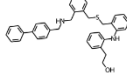
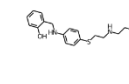
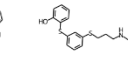
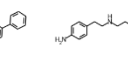
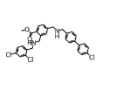
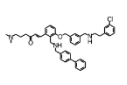
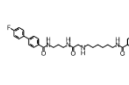
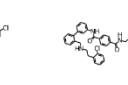
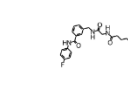
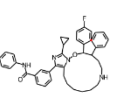
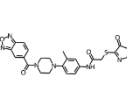
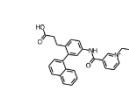
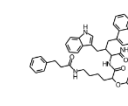
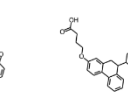
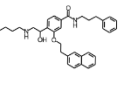
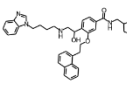
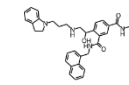
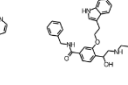
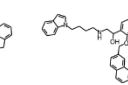
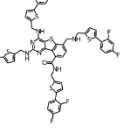
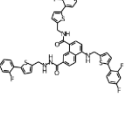
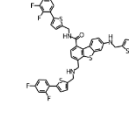
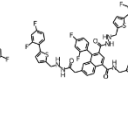
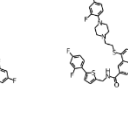
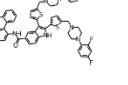
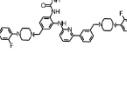
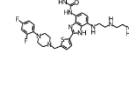
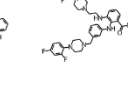
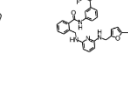
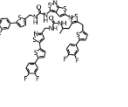
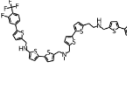
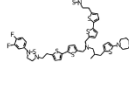
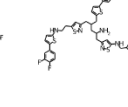
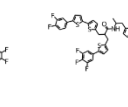
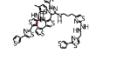
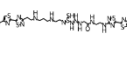
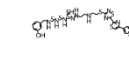
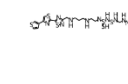
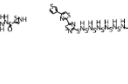
RL strategy	Top 1	Top 2	Top 3	Top 4	Top 5
REINFORCE	 CS: 3 S: -1197 AvS: -1261	 CS: 2 S: -1168 AvS: -1176	 CS: 2 S: -1190 AvS: -1200	 CS: 1 S: -1168 AvS: -1168	 CS: 1 S: -1168 AvS: -1168
REINFORCE + KL regularization	 CS: 1 S: -1058 AvS: -1058	 CS: 1 S: -1059 AvS: -1059	 CS: 1 S: -1059 AvS: -1059	 CS: 1 S: -1059 AvS: -1059	 CS: 1 S: -1061 AvS: -1061
REINVENT	 CS: 1 S: -987 AvS: -987	 CS: 1 S: -988 AvS: -988	 CS: 1 S: -988 AvS: -988	 CS: 1 S: -988 AvS: -988	 CS: 1 S: -989 AvS: -989
REINVENT 2.0	 CS: 3 S: -1071 AvS: -1081	 CS: 2 S: -1081 AvS: -1100	 CS: 1 S: -1065 AvS: -1065	 CS: 1 S: -1065 AvS: -1065	 CS: 1 S: -1065 AvS: -1065
BAR	 CS: 6 S: -934 AvS: -957	 CS: 1 S: -927 AvS: -927	 CS: 1 S: -927 AvS: -927	 CS: 1 S: -927 AvS: -927	 CS: 1 S: -927 AvS: -927
Augmented Hill-Climb	 CS: 14 S: -1398 AvS: -1423	 CS: 11 S: -1391 AvS: -1410	 CS: 9 S: -1383 AvS: -1389	 CS: 8 S: -1389 AvS: -1416	 CS: 4 S: -1383 AvS: -1422
Hill-Climb	 CS: 41 S: -1348 AvS: -1398	 CS: 13 S: -1395 AvS: -1383	 CS: 5 S: -1364 AvS: -1404	 CS: 4 S: -1343 AvS: -1386	 CS: 2 S: -1350 AvS: -1361
Hill-Climb + KL regularization	 CS: 68 S: -1279 AvS: -1314	 CS: 2 S: -1284 AvS: -1293	 CS: 2 S: -1287 AvS: -1292	 CS: 1 S: -1277 AvS: -1277	 CS: 1 S: -1279 AvS: -1279
Hill-Climb*	 CS: 91 S: -1351 AvS: -1384	 CS: 3 S: -1348 AvS: -1376	 CS: 2 S: -1342 AvS: -1357	 CS: 1 S: -1344 AvS: -1344	 CS: 1 S: -1358 AvS: -1358
Hill-Climb* + KL regularization	 CS: 68 S: -1280 AvS: -1301	 CS: 20 S: -1246 AvS: -1270	 CS: 3 S: -1252 AvS: -1245	 CS: 3 S: -1258 AvS: -1295	 CS: 2 S: -1251 AvS: -1254

Figure S15: Centroid of the top 5 largest clusters for the top 100 molecules in the DRD2 docking score task for different RL strategies. Cluster size (CS), centroid score (S) and the average cluster score (AvS) is annotated below.

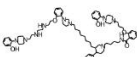
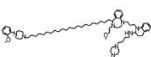
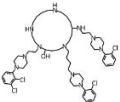
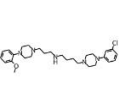
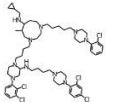
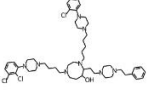
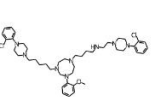
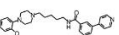
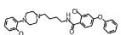
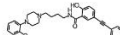
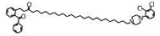
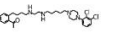
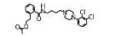
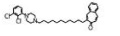
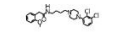
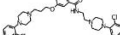
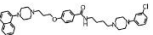
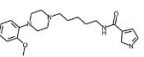
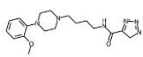
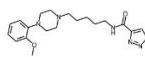
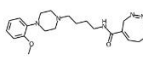
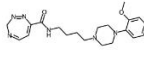
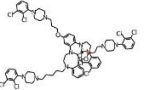
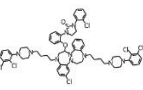
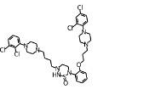
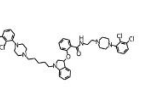
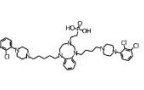
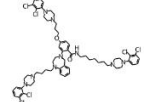
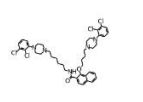
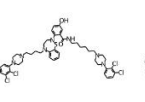
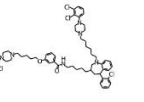
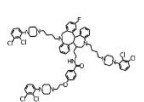
RL strategy	Top 1	Top 2	Top 3	Top 4	Top 5
REINFORCE					
	CS: 12 S: 0.81 AvS: 0.82	CS: 12 S: 0.82 AvS: 0.83	CS: 5 S: 0.84 AvS: 0.83	CS: 4 S: 0.81 AvS: 0.82	CS: 3 S: 0.81 AvS: 0.81
REINFORCE + KL regularization					
	CS: 10 S: 0.82 AvS: 0.82	CS: 4 S: 0.82 AvS: 0.82	CS: 4 S: 0.83 AvS: 0.83	CS: 4 S: 0.83 AvS: 0.83	CS: 3 S: 0.83 AvS: 0.84
REINVENT					
	CS: 43 S: 0.91 AvS: 0.91	CS: 13 S: 0.91 AvS: 0.91	CS: 8 S: 0.91 AvS: 0.91	CS: 6 S: 0.91 AvS: 0.91	CS: 5 S: 0.91 AvS: 0.91
REINVENT 2.0					
	CS: 5 S: 0.86 AvS: 0.86	CS: 4 S: 0.86 AvS: 0.86	CS: 4 S: 0.87 AvS: 0.89	CS: 4 S: 0.88 AvS: 0.87	CS: 3 S: 0.86 AvS: 0.87
BAR					
	CS: 29 S: 0.88 AvS: 0.88	CS: 10 S: 0.88 AvS: 0.89	CS: 5 S: 0.87 AvS: 0.87	CS: 3 S: 0.87 AvS: 0.89	CS: 3 S: 0.87 AvS: 0.89
Augmented Hill-Climb					
	CS: 15 S: 0.93 AvS: 0.93	CS: 10 S: 0.93 AvS: 0.93	CS: 5 S: 0.92 AvS: 0.92	CS: 5 S: 0.93 AvS: 0.93	CS: 4 S: 0.92 AvS: 0.92
Hill-Climb					
	CS: 28 S: 0.88 AvS: 0.88	CS: 10 S: 0.88 AvS: 0.88	CS: 3 S: 0.88 AvS: 0.88	CS: 3 S: 0.90 AvS: 0.90	CS: 2 S: 0.88 AvS: 0.89
Hill-Climb + KL regularization					
	CS: 21 S: 0.92 AvS: 0.92	CS: 18 S: 0.92 AvS: 0.92	CS: 10 S: 0.92 AvS: 0.92	CS: 6 S: 0.92 AvS: 0.92	CS: 5 S: 0.91 AvS: 0.92
Hill-Climb*					
	CS: 84 S: 0.75 AvS: 0.76	CS: 8 S: 0.76 AvS: 0.76	CS: 1 S: 0.75 AvS: 0.75	CS: 1 S: 0.75 AvS: 0.75	CS: 1 S: 0.76 AvS: 0.76
Hill-Climb* + KL regularization					
	CS: 70 S: 0.81 AvS: 0.81	CS: 11 S: 0.81 AvS: 0.81	CS: 8 S: 0.81 AvS: 0.81	CS: 7 S: 0.83 AvS: 0.82	CS: 3 S: 0.81 AvS: 0.81

Figure S16: Centroid of the top 5 largest clusters for the top 100 molecules in the DRD2-DRD3 QSAR dual predicted probability of activity task for different RL strategies. Cluster size (CS), centroid score (S) and the average cluster score (AvS) is annotated below. Note where AHC generates cationic species due to imperfections in the neutralization of training data resulting in charge symbols in the RNN vocabulary.

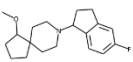
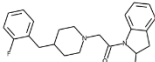
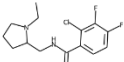
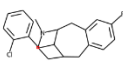
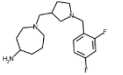
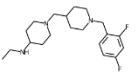
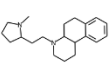
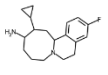
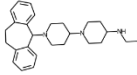
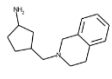
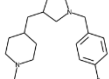
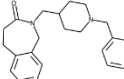
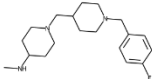
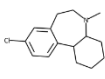
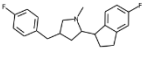
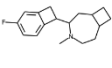
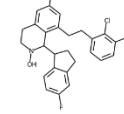
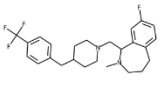
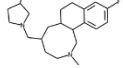
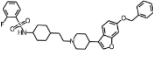
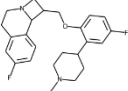
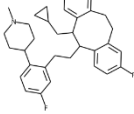
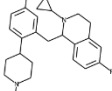
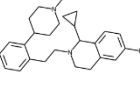
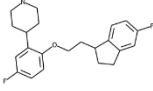
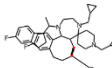
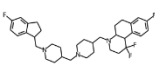
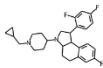
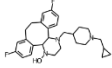
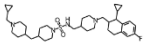
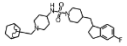
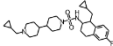
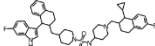
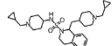
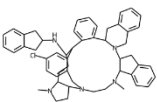
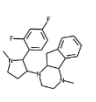
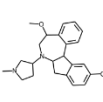
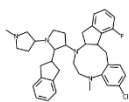
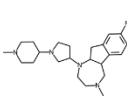
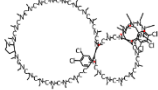
RL strategy	Top 1	Top 2	Top 3	Top 4	Top 5
REINFORCE					
	CS: 6 S: 0.78 AvS: 0.80	CS: 5 S: 0.81 AvS: 0.83	CS: 4 S: 0.83 AvS: 0.83	CS: 3 S: 0.78 AvS: 0.80	CS: 3 S: 0.79 AvS: 0.82
REINFORCE + KL regularization					
	CS: 5 S: 0.79 AvS: 0.80	CS: 4 S: 0.78 AvS: 0.80	CS: 4 S: 0.79 AvS: 0.80	CS: 3 S: 0.79 AvS: 0.80	CS: 2 S: 0.78 AvS: 0.79
REINVENT					
	CS: 10 S: 0.79 AvS: 0.80	CS: 6 S: 0.80 AvS: 0.81	CS: 3 S: 0.80 AvS: 0.80	CS: 3 S: 0.80 AvS: 0.81	CS: 3 S: 0.80 AvS: 0.80
REINVENT 2.0					
	CS: 9 S: 0.81 AvS: 0.83	CS: 6 S: 0.81 AvS: 0.83	CS: 6 S: 0.82 AvS: 0.83	CS: 5 S: 0.82 AvS: 0.82	CS: 3 S: 0.82 AvS: 0.82
BAR					
	CS: 4 S: 0.77 AvS: 0.78	CS: 3 S: 0.76 AvS: 0.77	CS: 2 S: 0.74 AvS: 0.76	CS: 2 S: 0.74 AvS: 0.80	CS: 2 S: 0.75 AvS: 0.75
Augmented Hill-Climb					
	CS: 8 S: 0.86 AvS: 0.86	CS: 7 S: 0.85 AvS: 0.86	CS: 5 S: 0.85 AvS: 0.86	CS: 5 S: 0.85 AvS: 0.86	CS: 4 S: 0.85 AvS: 0.85
Hill-Climb					
	CS: 17 S: 0.86 AvS: 0.86	CS: 11 S: 0.86 AvS: 0.86	CS: 5 S: 0.86 AvS: 0.86	CS: 5 S: 0.86 AvS: 0.86	CS: 5 S: 0.86 AvS: 0.86
Hill-Climb + KL regularization					
	CS: 32 S: 0.88 AvS: 0.87	CS: 10 S: 0.87 AvS: 0.87	CS: 8 S: 0.86 AvS: 0.87	CS: 8 S: 0.86 AvS: 0.86	CS: 5 S: 0.88 AvS: 0.87
Hill-Climb*					
	CS: 10 S: 0.78 AvS: 0.78	CS: 9 S: 0.78 AvS: 0.78	CS: 8 S: 0.78 AvS: 0.78	CS: 5 S: 0.77 AvS: 0.78	CS: 5 S: 0.78 AvS: 0.78
Hill-Climb* + KL regularization					
	CS: 68 S: 0.81 AvS: 0.81	CS: 7 S: 0.81 AvS: 0.82	CS: 4 S: 0.81 AvS: 0.81	CS: 2 S: 0.81 AvS: 0.81	CS: 2 S: 0.81 AvS: 0.81

Figure S17: Centroid of the top 5 largest clusters for the top 100 molecules in the DRD2/DRD3 QSAR predicted probability of selective activity task for different RL strategies. Cluster size (CS), centroid score (S) and the average cluster score (AvS) is annotated below.

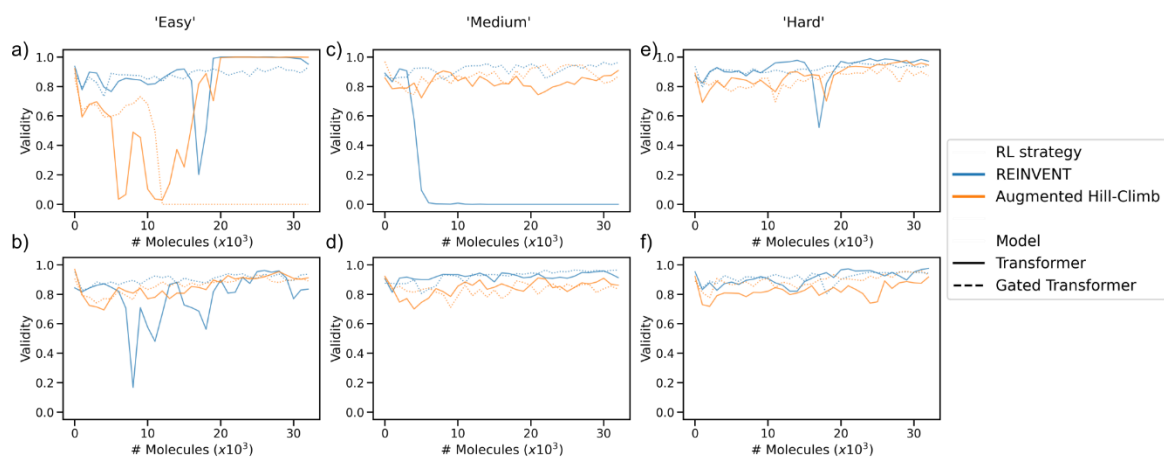


Figure S18: Validity for objective optimization using REINVENT and Augmented Hill-Climb with a transformer or gated transformer model.

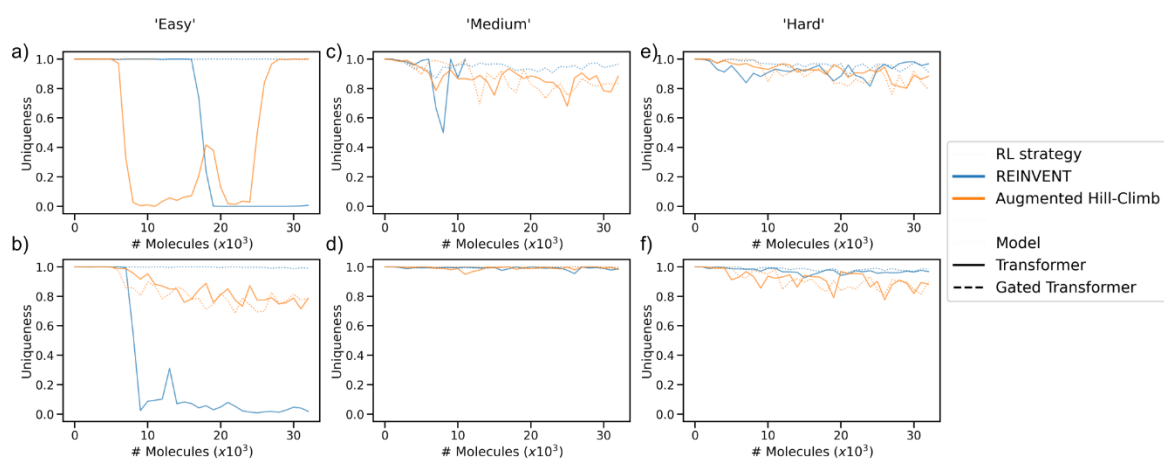


Figure S19: Uniqueness for objective optimization using REINVENT and Augmented Hill-Climb with a transformer or gated transformer model. Transformer model is more prone to undergoing a drop in uniqueness.

Model	RL strategy	Top 1	Top 2	Top 3	Top 4	Top 5
Transformer	REINVENT	 CS: 1 S: 47 AvS: 47	 CS: 1 S: 47 AvS: 47	 CS: 1 S: 47 AvS: 47	 CS: 1 S: 47 AvS: 47	 CS: 1 S: 47 AvS: 47
	Augmented Hill-Climb					
Gated Transformer	REINVENT	 CS: 1 S: 54 AvS: 54	 CS: 1 S: 54 AvS: 54	 CS: 1 S: 54 AvS: 54	 CS: 1 S: 54 AvS: 54	 CS: 1 S: 54 AvS: 54
	Augmented Hill-Climb	 CS: 1 S: 89 AvS: 89	 CS: 1 S: 89 AvS: 89	 CS: 1 S: 89 AvS: 89	 CS: 1 S: 89 AvS: 89	 CS: 1 S: 89 AvS: 89

Figure S20: Centroid of the top 5 largest clusters for the top 100 molecules in the heavy atom task for transformer models. Cluster size (CS), centroid score (S) and the average cluster score (AvS) is annotated below.

Model	RL strategy	Top 1	Top 2	Top 3	Top 4	Top 5
Transformer	REINVENT	 CS: 2 S: 0.28 AvS: 0.28	 CS: 2 S: 0.30 AvS: 0.31	 CS: 1 S: 0.27 AvS: 0.27	 CS: 1 S: 0.27 AvS: 0.27	 CS: 1 S: 0.27 AvS: 0.27
	Augmented Hill-Climb	 CS: 70 S: 0.64 AvS: 0.65	 CS: 14 S: 0.65 AvS: 0.65	 CS: 6 S: 0.65 AvS: 0.64	 CS: 2 S: 0.63 AvS: 0.64	 CS: 2 S: 0.63 AvS: 0.64
Gated Transformer	REINVENT	 CS: 6 S: 0.35 AvS: 0.37	 CS: 3 S: 0.36 AvS: 0.36	 CS: 3 S: 0.37 AvS: 0.38	 CS: 2 S: 0.35 AvS: 0.38	 CS: 2 S: 0.36 AvS: 0.37
	Augmented Hill-Climb	 CS: 39 S: 0.68 AvS: 0.67	 CS: 10 S: 0.68 AvS: 0.68	 CS: 7 S: 0.65 AvS: 0.67	 CS: 5 S: 0.67 AvS: 0.67	 CS: 5 S: 0.67 AvS: 0.67

Figure S21: Centroid of the top 5 largest clusters for the top 100 molecules in the Risperidone similarity task for transformer models. Cluster size (CS), centroid score (S) and the average cluster score (AvS) is annotated below.

Model	RL strategy	Top 1	Top 2	Top 3	Top 4	Top 5
Transformer	REINVENT					
		CS: 22 S: 0.84 AvS: 0.88	CS: 16 S: 0.84 AvS: 0.88	CS: 11 S: 0.84 AvS: 0.88	CS: 5 S: 0.85 AvS: 0.87	CS: 4 S: 0.85 AvS: 0.87
	Augmented Hill-Climb					
		CS: 15 S: 0.98 AvS: 0.98	CS: 10 S: 0.98 AvS: 0.98	CS: 7 S: 0.98 AvS: 0.98	CS: 7 S: 0.98 AvS: 0.98	CS: 6 S: 0.98 AvS: 0.98
Gated Transformer	REINVENT					
		CS: 14 S: 0.98 AvS: 0.98	CS: 9 S: 0.98 AvS: 0.98	CS: 6 S: 0.98 AvS: 0.98	CS: 6 S: 0.98 AvS: 0.98	CS: 5 S: 0.98 AvS: 0.98
	Augmented Hill-Climb					
		CS: 11 S: 0.98 AvS: 0.98	CS: 9 S: 0.98 AvS: 0.98	CS: 6 S: 0.98 AvS: 0.98	CS: 6 S: 0.98 AvS: 0.98	CS: 5 S: 0.98 AvS: 0.98

Figure S22: Centroid of the top 5 largest clusters for the top 100 molecules in the DRD2 activity task for transformer models. Cluster size (CS), centroid score (S) and the average cluster score (AvS) is annotated below.

Model	RL strategy	Top 1	Top 2	Top 3	Top 4	Top 5
Transformer	REINVENT					
		CS: 2 S: -9.79 AvS: -10.09	CS: 1 S: -9.69 AvS: -9.69	CS: 1 S: -9.69 AvS: -9.69	CS: 1 S: -9.70 AvS: -9.70	CS: 1 S: -9.70 AvS: -9.70
	Augmented Hill-Climb					
		CS: 7 S: -14.89 AvS: -15.10	CS: 6 S: -14.79 AvS: -15.02	CS: 5 S: -15.26 AvS: -15.05	CS: 2 S: -14.75 AvS: -14.81	CS: 2 S: -14.76 AvS: -14.78
Gated Transformer	REINVENT					
		CS: 1 S: -10.09 AvS: -10.09	CS: 1 S: -10.09 AvS: -10.09	CS: 1 S: -10.09 AvS: -10.09	CS: 1 S: -10.09 AvS: -10.09	CS: 1 S: -10.10 AvS: -10.10
	Augmented Hill-Climb					
		CS: 4 S: -14.98 AvS: -15.05	CS: 4 S: -15.48 AvS: -15.12	CS: 3 S: -14.85 AvS: -15.46	CS: 3 S: -14.99 AvS: -14.94	CS: 2 S: -14.95 AvS: -15.42

Figure S23: Centroid of the top 5 largest clusters for the top 100 molecules in the DRD2 docking score task for transformer models. Cluster size (CS), centroid score (S) and the average cluster score (AvS) is annotated below.

Model	RL strategy	Top 1	Top 2	Top 3	Top 4	Top 5
Transformer	REINVENT					
		CS: 14 S: 0.85 AvS: 0.85	CS: 7 S: 0.84 AvS: 0.84	CS: 7 S: 0.84 AvS: 0.84	CS: 7 S: 0.85 AvS: 0.84	CS: 6 S: 0.85 AvS: 0.85
	Augmented Hill-Climb					
		CS: 46 S: 0.92 AvS: 0.92	CS: 9 S: 0.89 AvS: 0.91	CS: 7 S: 0.89 AvS: 0.89	CS: 6 S: 0.89 AvS: 0.91	CS: 5 S: 0.91 AvS: 0.91
Gated Transformer	REINVENT					
		CS: 43 S: 0.92 AvS: 0.92	CS: 11 S: 0.91 AvS: 0.91	CS: 6 S: 0.91 AvS: 0.91	CS: 5 S: 0.91 AvS: 0.92	CS: 5 S: 0.91 AvS: 0.91
	Augmented Hill-Climb					
		CS: 15 S: 0.94 AvS: 0.94	CS: 11 S: 0.95 AvS: 0.94	CS: 10 S: 0.94 AvS: 0.94	CS: 4 S: 0.94 AvS: 0.94	CS: 4 S: 0.94 AvS: 0.94

Figure S24: Centroid of the top 5 largest clusters for the top 100 molecules in the DRD2-DRD3 QSAR dual predicted probability of activity task for transformer models. Cluster size (CS), centroid score (S) and the average cluster score (AvS) is annotated below.

Model	RL strategy	Top 1	Top 2	Top 3	Top 4	Top 5
Transformer	REINVENT					
		CS: 6 S: 0.74 AvS: 0.76	CS: 4 S: 0.74 AvS: 0.75	CS: 3 S: 0.74 AvS: 0.75	CS: 3 S: 0.74 AvS: 0.75	CS: 3 S: 0.74 AvS: 0.76
	Augmented Hill-Climb					
		CS: 10 S: 0.85 AvS: 0.85	CS: 10 S: 0.86 AvS: 0.85	CS: 9 S: 0.85 AvS: 0.85	CS: 5 S: 0.85 AvS: 0.85	CS: 4 S: 0.85 AvS: 0.85
Gated Transformer	REINVENT					
		CS: 5 S: 0.79 AvS: 0.81	CS: 4 S: 0.78 AvS: 0.81	CS: 3 S: 0.79 AvS: 0.79	CS: 3 S: 0.79 AvS: 0.81	CS: 3 S: 0.79 AvS: 0.80
	Augmented Hill-Climb					
		CS: 10 S: 0.87 AvS: 0.87	CS: 9 S: 0.87 AvS: 0.87	CS: 6 S: 0.86 AvS: 0.86	CS: 5 S: 0.87 AvS: 0.87	CS: 4 S: 0.86 AvS: 0.86

Figure S25: Centroid of the top 5 largest clusters for the top 100 molecules in the DRD2/DRD3 QSAR predicted probability of selective activity task for different transformer models. Cluster size (CS), centroid score (S) and the average cluster score (AvS) is annotated below.