Optimization of binding affinities in chemical space with transformer and deep reinforcement learning

Xiaopeng Xu[1,2], Juexiao Zhou[1,2], Chen Zhu[3], Qing Zhan[1,2], Zhongxiao Li[1,2], Ruochi Zhang[4], Yu Wang[4], Xingyu Liao[1,2], and Xin Gao[*,1,2]

[1] *Computer Science Program, Computer, Electrical and Mathematical Science and Engineering (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia*
[2] *Computational Bioscience Research Center (CBRC), KAUST, Thuwal, 23955-6900, Saudi Arabia*
[3] *KAUST Catalysis Center (KCC), KAUST, Thuwal, 23955-6900, Saudi Arabia*
[4] *Syneron Technology, Guangzhou, China*
[*] *Corresponding author: xin.gao@kaust.edu.sa.*

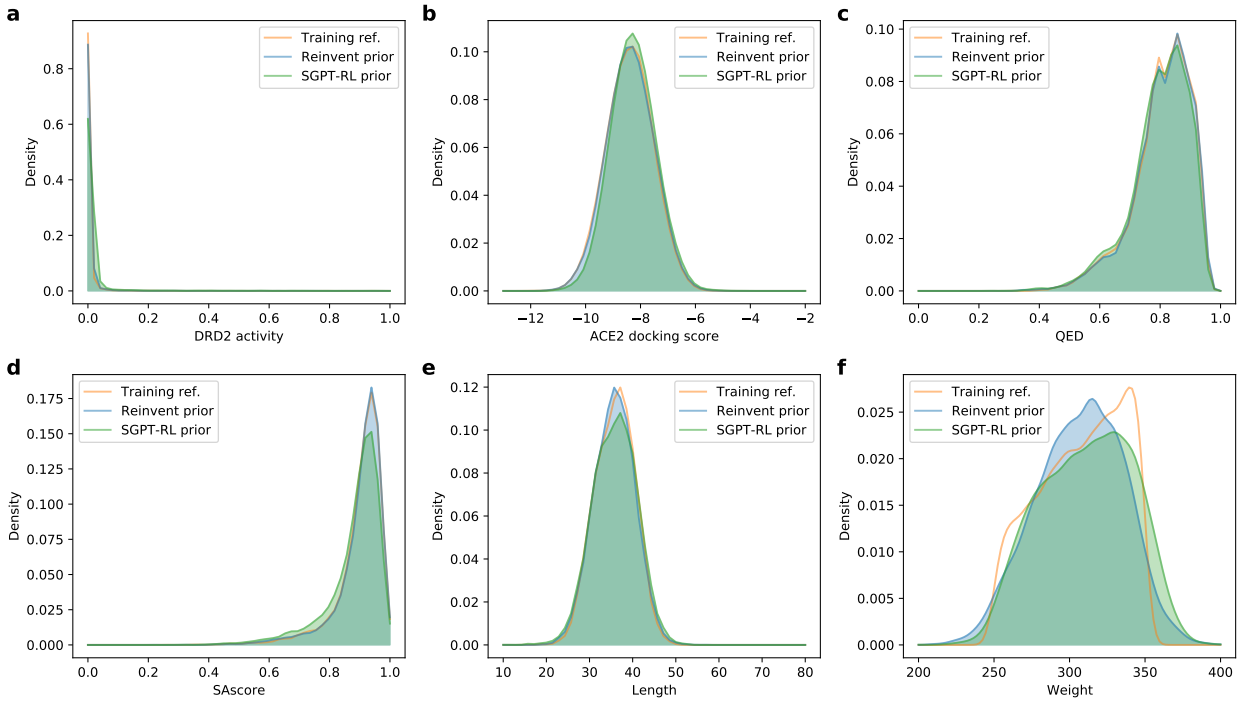
1 Evaluation of SGPT-RL on distribution learning

SGPT-RL and Reinvent prior models were evaluated on the Moses Benchmark. The results of the Moses metrics are shown in Supplementary Table 1. We can see that the SGPT-RL prior model learned a good validity and novelty on Moses benchmark.

Supplementary Table 1: Comparison of models on the Moses benchmark

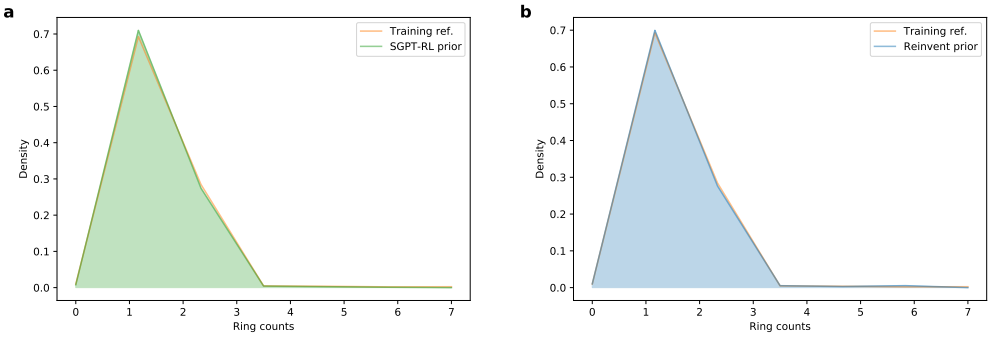
Model	Validity	Uniqueness	SNN	IntDiv	Novelty
Baseline	**1.000**	**1.000**	**0.642**	0.856	0.000
CharRNN	0.975	0.999	0.602	0.856	0.842
VAE	0.977	0.998	0.626	0.856	0.695
AAE	0.937	0.997	0.608	0.856	0.793
JT-VAE	**1.000**	**1.000**	0.548	0.855	0.914
LatentGAN	0.897	0.997	0.537	**0.857**	0.950
MCMG	0.886	-	0.427	0.835	**0.983**
MolGPT	**1.000**	**1.000**	0.529	0.871	0.931
Reinvent prior	0.986	1.000	0.619	0.856	0.783
SGPT-RL prior	0.936	0.997	0.563	0.856	0.946

The property distributions of molecules generated by the prior models and training references were plotted as shown in Supplementary Figure 1. The SGPT-RL and the Reinvent prior models follow the same distributions as the training reference.



Supplementary Figure 1: Property distribution of molecules generated by prior models. 10,000 molecules sampled from training dataset were used as a reference. Curves of SGPT-RL and Reinvent prior samples are shown with shading. Curves of training references are shown without shading. We can see that molecules from the three sources follow similar property distributions.

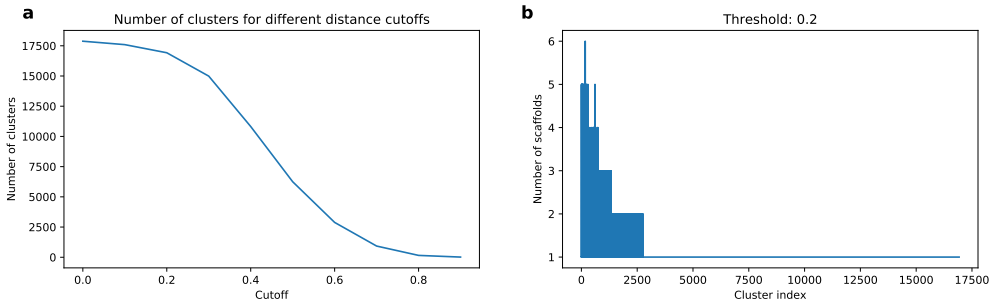
The distribution of ring counts were analyzed as shown in Supplementary Figure 2. The prior models follow the same distribution as the training reference.



Supplementary Figure 2: Ring count distribution of molecules generated by the SGPT-RL and Reinvent prior models. 10,000 molecules sampled from the training data were used as a reference. The molecules generated by SGPT-RL and the Reinvent prior model follow the same distributions as the training reference.

1.1 Scaffold clustering threshold selection

Analysis of different distance cutoffs for scaffold clustering is shown in Supplementary Figure 3. A cutoff of 0.2 can be a good threshold as the curve in Supplementary Subfigure 3a is becoming flat before 0.2 and most of the clusters contain only a single scaffold using a cutoff of 0.2.

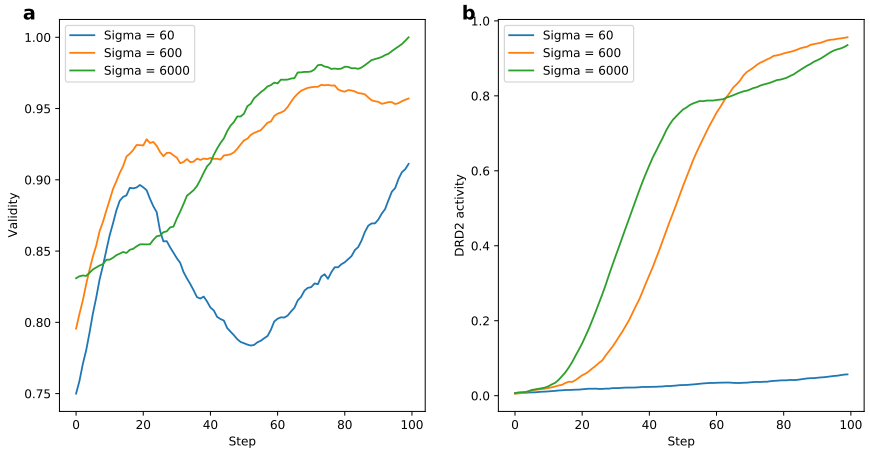


Supplementary Figure 3: Cutoffs for scaffold clustering. a) The number of clusters given a cutoff range from 0 to 0.9. A cutoff of 0.2 can be a good threshold as the curve is becoming flat before 0.2. b) The number of scaffolds for each cluster when using a cutoff of 0.2. Most of the clusters contain only a single scaffold using a cutoff of 0.2.

2 Hyper-parameter tuning of SGPT-RL

The hyper-parameter of SGPT-RL was fine-tuned to find the suitable value for goal-directed generation tasks. The main hyper-parameter, sigma, was evaluated as described below.

The step curves of different sigma values are shown in Supplementary Figure 4. From Supplementary Figure 4a, we can see that with larger sigma, the agents achieved a better validity after 100 steps. Besides, from Supplementary Figure 4b, we can see that larger sigma value also enabled generating more active molecules in the initial steps.



Supplementary Figure 4: Learning curves of different sigma values in SGPT-RL on the DRD2 task

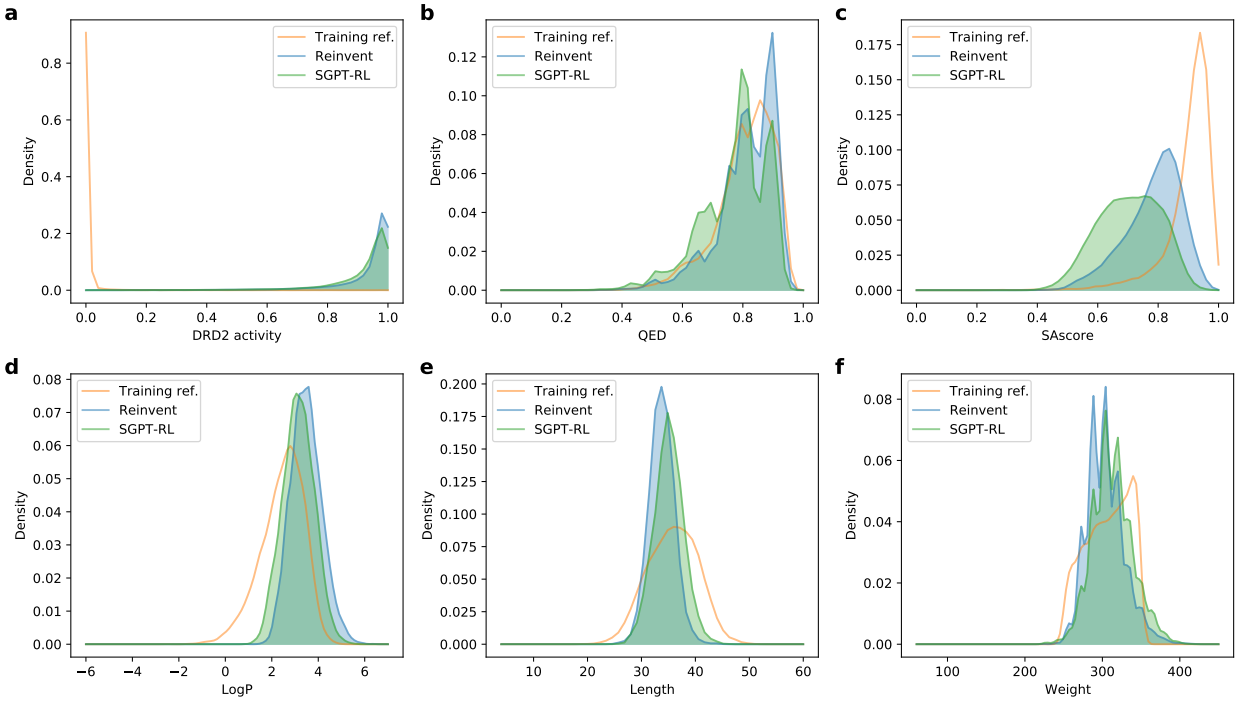
However, when comparing the agents trained after 3000 steps on Moses metrics, we found larger sigma values lead to poorer uniqueness and internal diversity of generated molecules, as shown in Supplementary Table 2. This was probably caused by mode collapse during training the agent. As good diversity was preferred in our study, a sigma value of 60 was chosen for the SGPT-RL agent.

Supplementary Table 2: Moses metrics of different sigma values in SGPT-RL on the DRD2 task

Sigma	Validity	Uniqueness	SNN	IntDiv	Novelty
60	0.998	**0.933**	**0.515**	**0.683**	0.995
120	0.998	0.797	0.486	0.631	0.999
600	0.999	0.334	0.340	0.481	**1.000**
1920	**1.000**	0.257	0.392	0.474	**1.000**

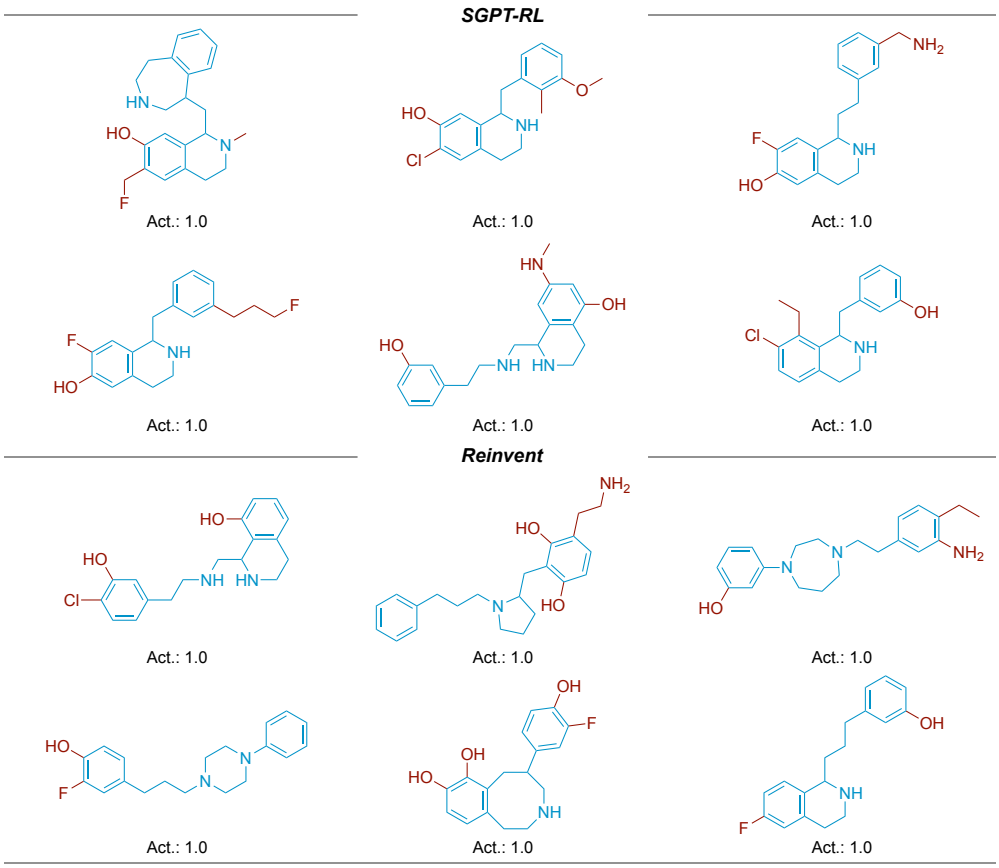
3 Comparison of SGPT-RL and Reinvent on the DRD2 task

The property distributions of agents trained after the final step are shown in Supplementary Figure 5. Both SGPT-RL and Reinvent agents were able to generate molecules with high DRD2 activities.



Supplementary Figure 5: Property distributions of molecules generated by agent models on the DRD2 task. 10,000 molecules are sampled from training dataset to be used as the reference. Both SGPT-RL and Reinvent agents were able to generate molecules with high DRD2 activities.

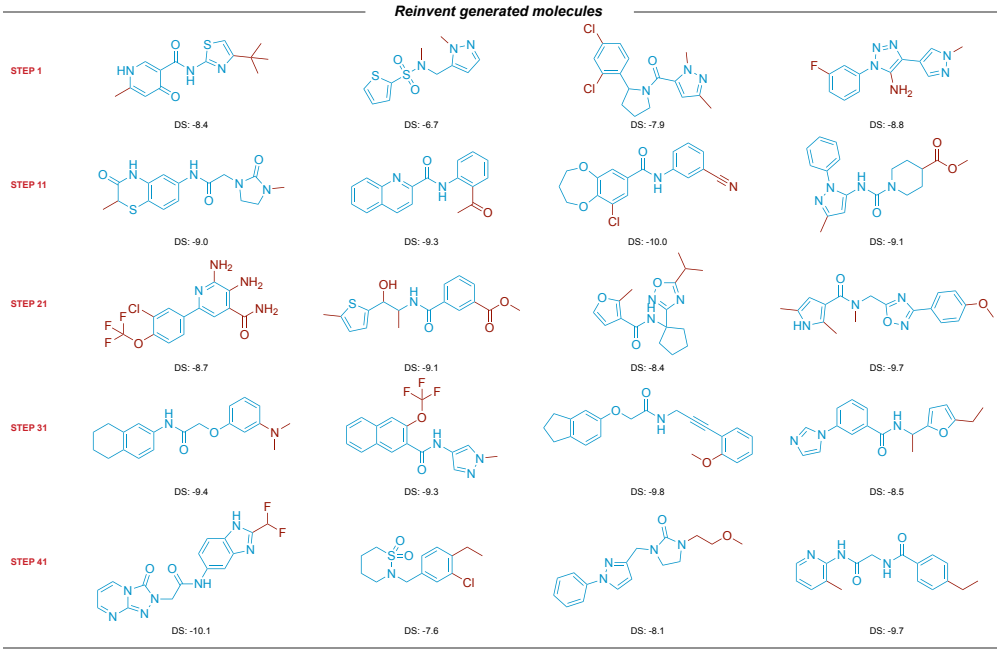
Supplementary Figure 6 shows the six top scoring molecules generated by the agents in the 1,000 th step. SGPT-RL generated molecules are more similar to each other comparing to Reinvent generated molecules.



Supplementary Figure 6: Top scoring molecules generated by the agents in the 1,000th step of the DRD2 task. SGPT-RL generated molecules are more similar to each other comparing to Reinvent generated molecules.

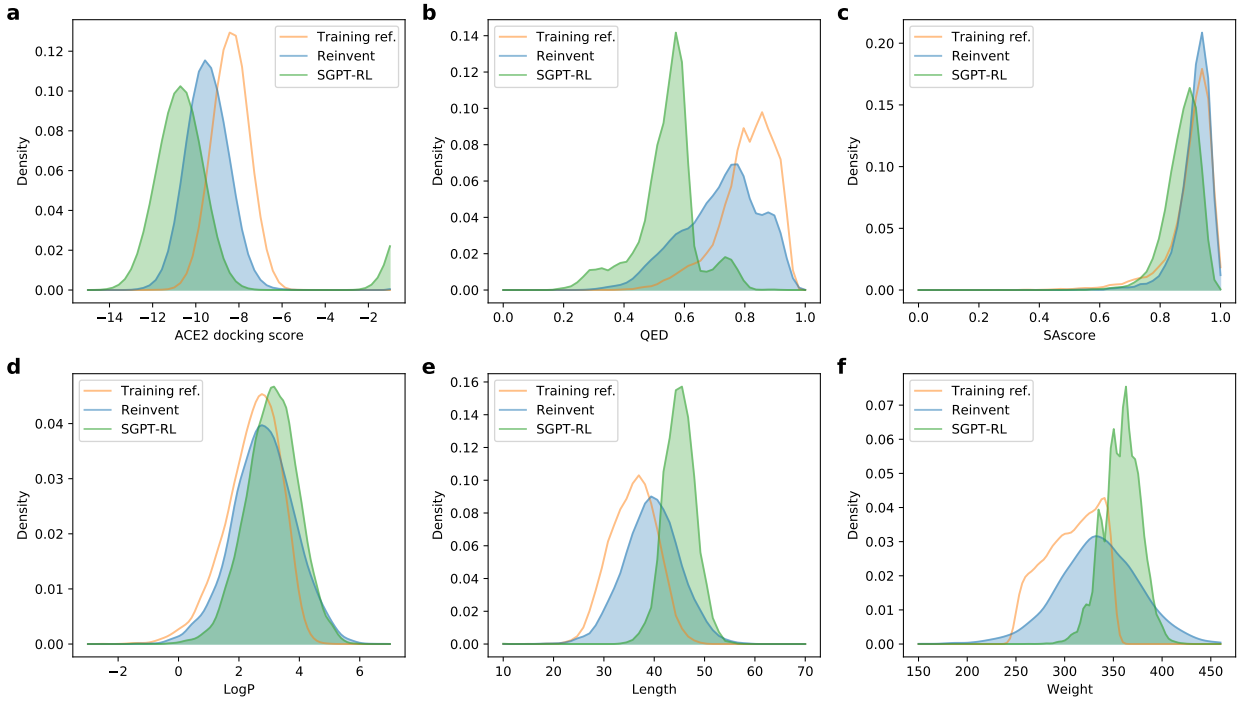
4 Structure-based molecular generation with ACE2 as the target

Supplementary Figure 7 shows the molecules generated by the Reinvent agent in the initial steps on the ACE2 task. This agent was randomly exploring the sequences, with no clear patterns observed.



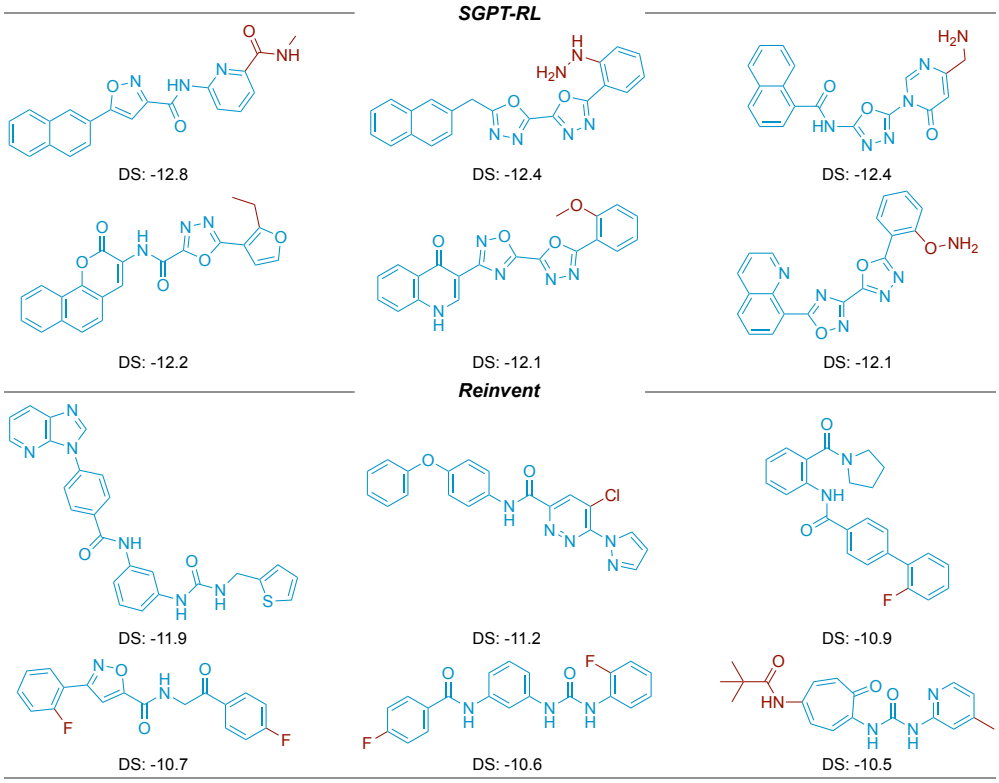
Supplementary Figure 7: Reinvent generated molecules in the initial steps of the ACE2 task. This model was randomly exploring the sequences, with no clear patterns observed.

The property distributions of molecules generated by the agents and the training reference on the ACE2 task are shown in Supplementary Figure 8. SGPT-RL shifted the distribution towards better docking scores.



Supplementary Figure 8: Property distributions of molecules generated by the agent models on the ACE2 task. 10,000 molecules from the training dataset were evaluated as a reference. SGPT-RL shifted the distribution towards better docking scores.

Supplementary Figure 9 shows the six top scoring molecules generated by the agents in the last step of the ACE2 task. SGPT-RL generated molecules are more similar to each other comparing to Reinvent generated ones.



Supplementary Figure 9: Top scoring molecules generated in the last step of the ACE2 task. SGPT-RL generated molecules are more similar to each other comparing to Reinvent generated ones.