Supplementary Material: *Molecular Design in Synthetically Accessible Chemical Space via Deep Reinforcement Learning*

Julien Horwood[*,†,‡] and Emmanuel Noutahi[*,†]

†*InVivo AI*

‡*Mila, Université de Montréal*

E-mail: julien@invivoai.com; emmanuel@invivoai.com

Supplementary Material

Results

Table S1: Goal-Directed Molecule Design

Objective	Method	Mean Reward	Max Reward	Diversity	Scaff. Similarity	Uniqueness
cLogP	BLOCKS	-1.80 ± 0.08	1.80 ± 0.32	0.94 ± 0	N/A	100% ± 0
	Hill Climbing	7.14 ± 0.20	10.90 ± 0.04	0.73 ± 0.01	**0.13 ± 0.01**	**100% ± 0**
	ORGAN	-2.47	0.97	0.83	0.14	63%
	JTVAE	-1.48 ± 0.56	0.16 ± 0.14	0.54 ± 0.2	N/A	41% ± 34%
	GCPN	1.03 ± 0.28	8.51 ± 0.35	**0.90 ± 0**	0.20 ± 0.01	**100% ± 0**
	MolDQN	**12.84 ± 0.23**	**18.42 ± 0.37**	0.71 ± 0.01	0.72 ± 0.23	72% ± 3.6%
	REACTOR	8.01 ± 0.18	10.74 ± 0.28	0.69 ± 0.01	0.20 ± 0	99.7% ± 0.5%
QED	BLOCKS	0.523 ± 0.005	0.763 ± 0.009	0.94 ± 0	N/A	100% ± 0
	Hill Climbing	0.811 ± 0.007	0.943 ± 0.004	0.879 ± 0.003	0.20 ± 0.023	**100% ± 0**
	ORGAN	0.608	0.906	0.871	0.178	89.5%
	JTVAE	0.604 ± 0.017	0.876 ± 0.048	0.841 ± 0.018	0.638 ± 0.046	92.8% ± 5.5%
	GCPN	0.607 ± 0.012	0.916 ± 0.012	**0.91 ± 0.002**	**0.112 ± 0.004**	**100% ± 0**
	MolDQN	0.857 ± 0.026	0.936 ± 0.004	0.791 ± 0.007	0.620 ± 0.123	67% ± 5.8%
	REACTOR	**0.876 ± 0.007**	**0.947 ± 0.001**	0.878 ± 0.002	0.161 ± 0.021	**100% ± 0**

Figures

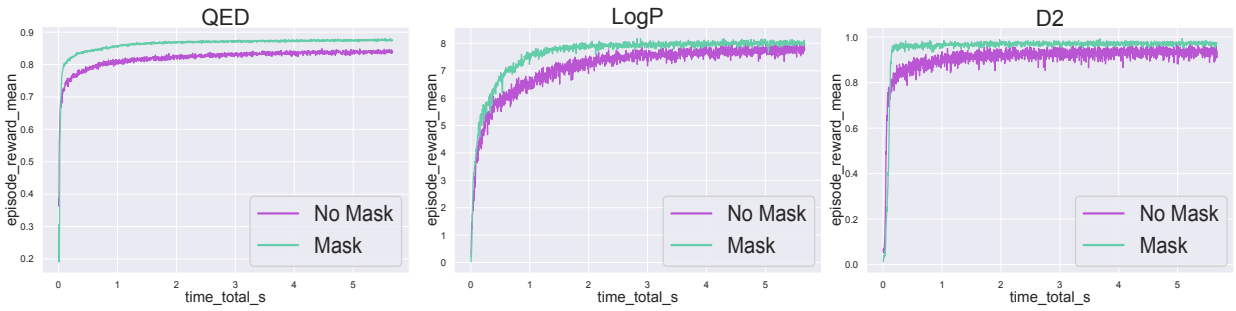
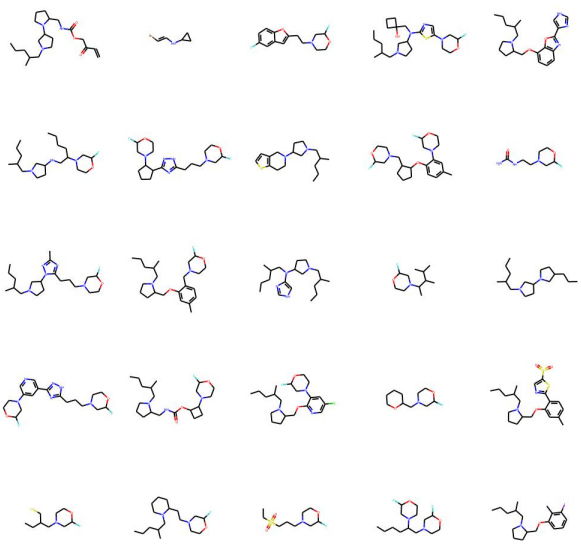
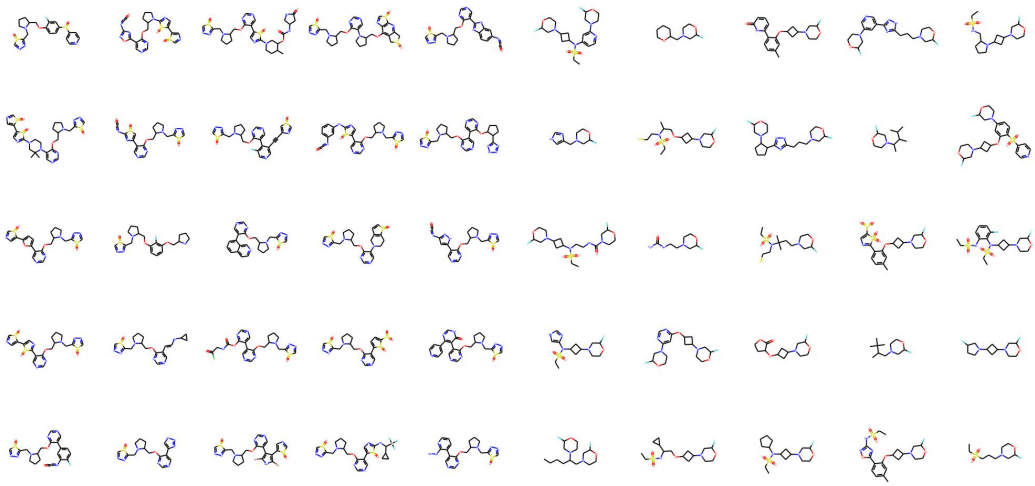


Figure S1: REACTOR convergence ablation when using a masked action space



(a) DRD2 Molecule Samples



(b) DRD2 with D1 selectivity

(c) DRD2 with D3 selectivity

Figure S2: Molecule samples for the various DRD2 optimization tasks

Reward Model Details

DRD2 Reward Model

The model for the DRD2 receptor was trained using data from ExCAPE-DB,[S1] with 8323 active and 343206 inactive compounds. Molecules were then sanitized and duplicate molecules were removed. We then performed a stratified split consisting of 90% training and 10%

test splits. 3-fold cross validation was performed over the training set in order to select a model. We compared Random Forest, Gradient Boosting, Support Vector Machines and Feed-Forward Neural Networks, using 2048 Morgan Fingerprints with radius 2 as molecular featurizations. The selected model is a 200 neuron single-layer neural network, with its classification performance on the held-out test set provided in Figure S3.

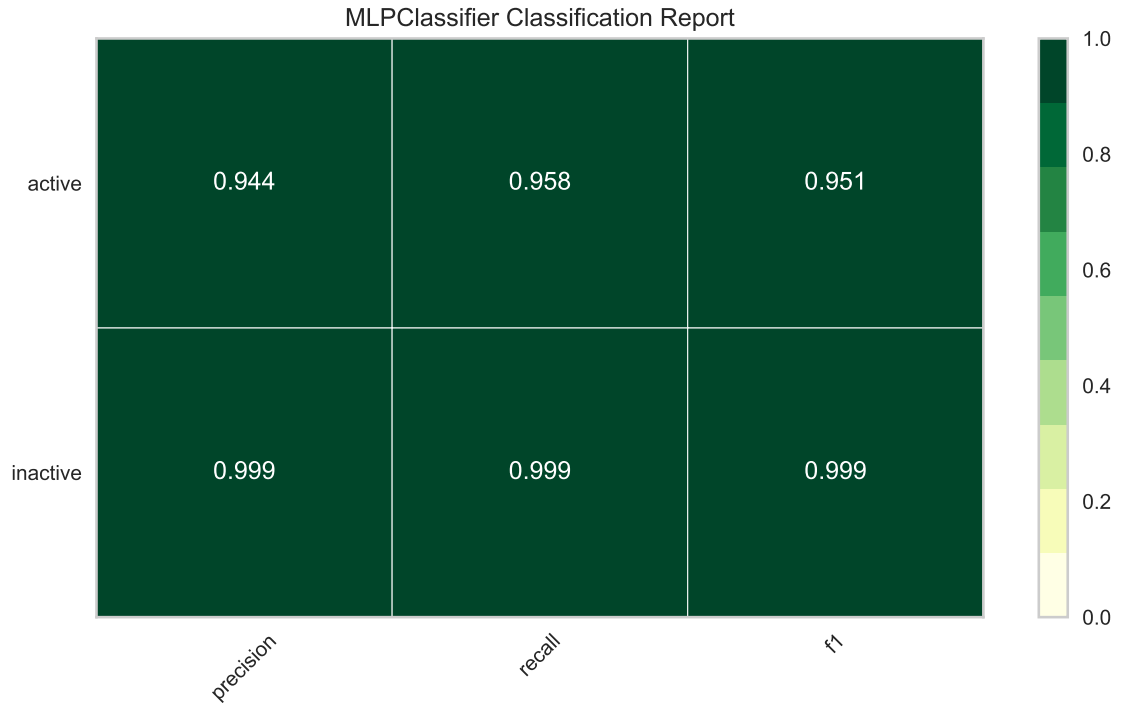


Figure S3: Model performance on test data for the selected DRD2 model

DRD1 and DRD3 Reward Models

DRD1 and DRD3 modulators were obtained from ExCAPEDB.[S1] Due to the high data imbalance, only structurally diverse inactive (pXC50 < 5) compounds with experimentally validated activity were retained. Each dataset was subsequently cleaned using the following procedure:

- All molecules are sanitized and standardized.

- Duplicate compounds were removed and only the largest fragment was retained for each molecule. This resulted in a dataset of 1753 actives vs 10317 inactives for DRD1 and 3498 vs 10074 inactives for DRD3. Each dataset was split into an 80% training and 20% test set, using a stratified split.

- The DRD1 and DRD3 activity models were trained using cross-validation (80-20) under various splits of the training set (random split, stratified activity split, structural-similarity based clustering split, scaffold split) and evaluated using balanced accuracy and f1-score. We considered various featurizations and their combinations, as well as several machine learning algorithms (Support Vector Machine, Random Forest, Gradient Boosting, Logistic Regression and a Multi-Layer Perceptron). The hyper-parameters, including molecular featurization, resulting in the best performances were selected for each algorithm, and the best performing model on the held out test set was retained.

For both datasets, the best model according to the F1-score/ROC-AUC/Balanced Accuracy was a Gradient Boosting Classifier.

Caco-2 Reward Model

Data for the Caco-2 cell permeability assay was obtained from Wang et al., with a measurement unit of $log(10^{-6})cm/s$. Model selection was performed using a 6-fold stratified split. Algorithms compared at this stage were Random Forest, Kernel Ridge, and Gaussian Process regression algorithms, with model selection additionally performed over various Fingerprint featurizations. The final model is a Kernel Ridge Regression model with a Laplacian kernel, with 512-bit Estate Fingerprints.

References

(S1) Sun, J.; Jeliazkova, N.; Chupakhin, V.; Golib-Dzib, J.-F.; Engkvist, O.; Carlsson, L.; Wegner, J.; Ceulemans, H.; Georgiev, I.; Jeliazkov, V.; et al. ExCAPE-DB: an integrated large scale dataset facilitating Big Data analysis in chemogenomics. *J. Cheminf.* **2017**, *9*, 17.

(S2) Wang, N.-N.; Dong, J.; Deng, Y.-H.; Zhu, M.-F.; Wen, M.; Yao, Z.-J.; Lu, A.-P.; Wang, J.-B.; Cao, D.-S. ADME Properties Evaluation in Drug Discovery: Prediction of Caco-2 Cell Permeability Using a Combination of NSGA-II and Boosting. *J. Chem. Inf. Model.* **2016**, *56*, 763–773.