In Silico Generation of Synthesizable Inhibitors of Monoglyceride Lipase

Martin Šícho[1,2], Alan Kai Hassen[3,4], Yorick van Aalst[1,5], Sohvi Luukkonen[1], Antonius P.A. Janssen[5], Djork-Arné Clevert[4], Gerard van Westen[1], Mike Preuss[3]

[1] *Leiden Academic Centre of Drug Research, Leiden University, The Netherlands*
[2] *CZ-OPENSCREEN: National Infrastructure for Chemical Biology, Department of Informatics and Chemistry, University of Chemistry and Technology Prague, Czech Republic*
[3] *Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands*
[4] *Machine Learning Research, Pfizer Worldwide Research Development and Medical, Berlin, Germany*
[5] *Leiden Institute of Chemistry, Leiden University, The Netherlands*

Monoglyceride Lipase (MGLL)

MGLL is an enzyme responsible for hydrolysis of monoesters of long-chain fatty acids. Inactivation of MGLL causes downregulation of cannabinoid receptor 1 (CB_1) in selective areas of the brain, which can cause behavioral effects [1]. Therefore, MGLL inhibitors have been sought as potential drug candidates for neurodegenerative diseases and similar ailments. In particular, covalent inhibitors of MGLL have been suggested as the most prospective leads (**Figure 1** and **2**).

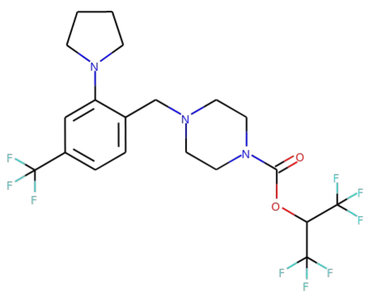

Figure 1: The structure of ABX-1431 (Elcubragistat), a phase II clinical candidate and a potent MGLL covalent inhibitor , IC50 = 14.0 nM.

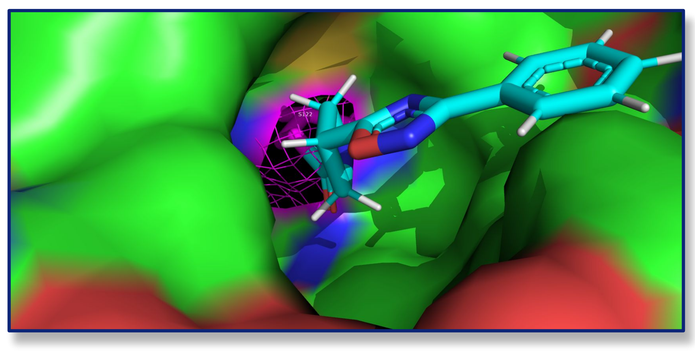

Figure 2: Covalent inhibitor analogous to ABX-1431 covalently bound to SER146 (purple) of MGLL enzyme (PDB: 6AX1).

DrugEx

DrugEx [2] is an open source software package providing a selection of deep learning models for molecular generation. It employs a unique reinforcement learning strategy with support for multiple objectives and an exploration parameter to tune structural diversity (**Figure 3**).

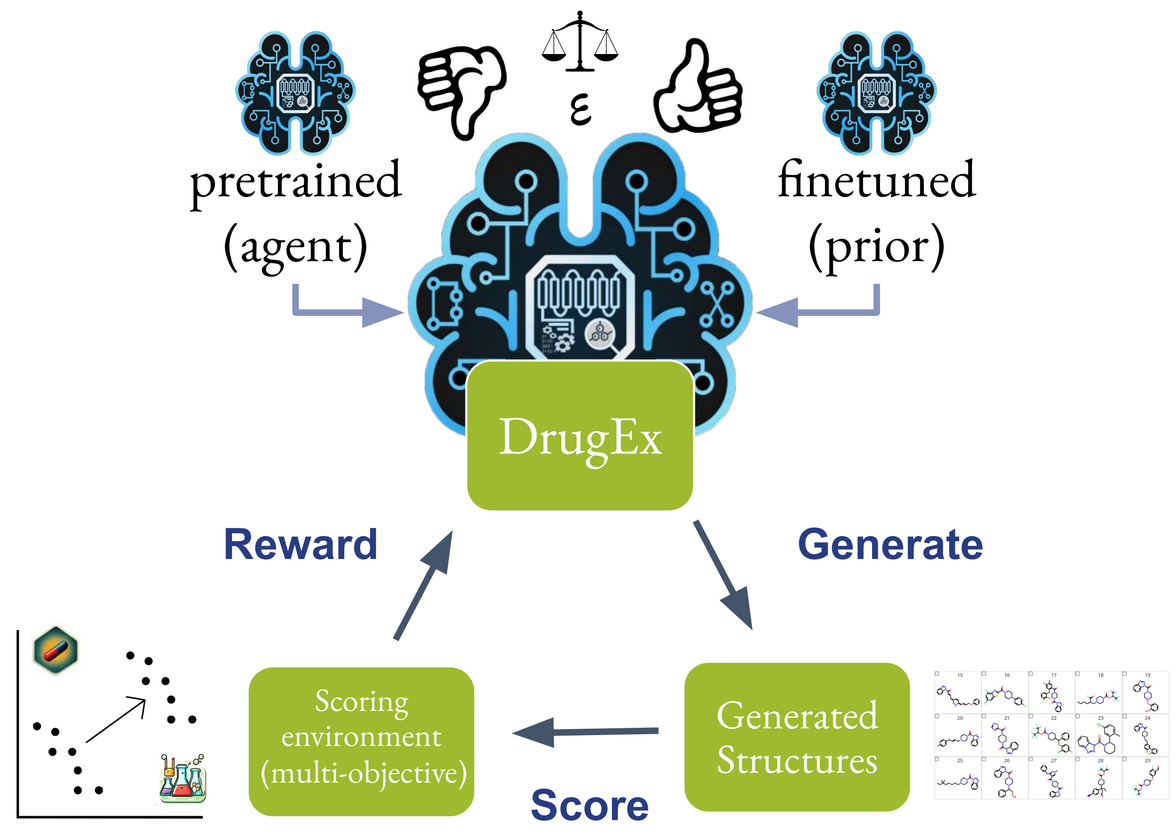

Figure 3: Reinforcement learning loop of the DrugEx models. In each epoch, the exploration parameter (ε) determines probability that a vocabulary token is sampled from the *agent* or the *prior*. The prior is set to a model finetuned on MGLL-specific data while the agent is a general chemistry model. Increasing the value of ε, results in a distribution more similar to training data.

Objective 1: QSAR Model

The QSPRPred [3] package was used to train an XGBoost classification model on 700 examples of both active and inactive compounds. The model was evaluated using two test set selection strategies (**Figure 4**).

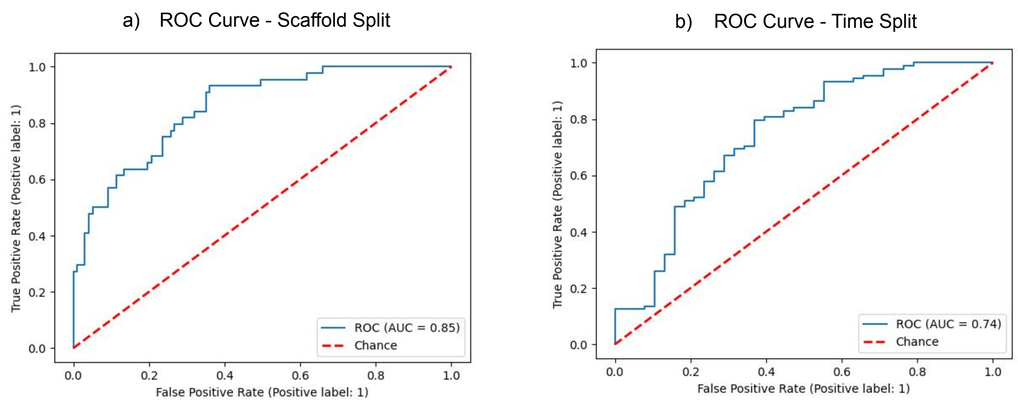

Figure 4: Evaluation of the QSAR classification model with (a) scaffold split and (b) time split strategy. Each time, the test set contained ~20% of the original data.

Objective 2: LED3Score [4]

Novel retrosynthetic accessibility score, which predicts, given a limited set of in house building blocks, the retrosynthetic solvability of a given structure.

Generated Structures

Out of 100,000 unique generated structures 25,044 had an obtainable retrosynthetic route and were classified as active by the QSAR model with probability higher than 0.8.
A shortlist of 2,739 structures with predicted routes of length 4 was assembled (**Figure 5**).

Three candidate structures from this shortlist are currently selected for synthesis and biological evaluation for MGLL inhibition.

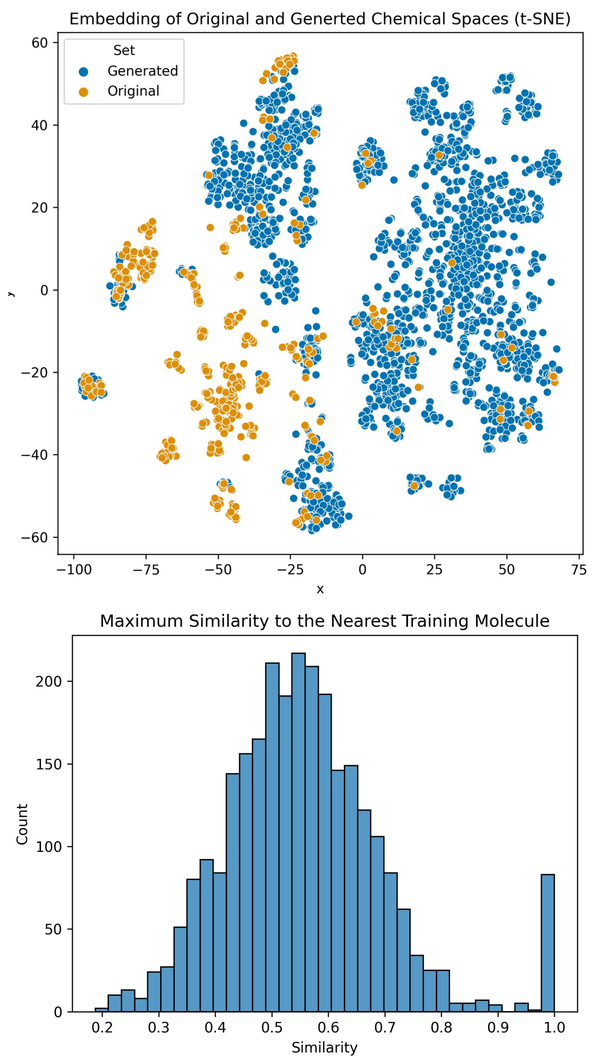

Figure 5: Evaluation of chemical space and novelty of the generated structures. Morgan fingerprint (length=2048, radius=3) was used.

References

1. Grabner et al., Pharmacol Ther., *2017*, DOI: 10.1016/j.pharmthera.2017.02.033
2. Sicho et al., J. Chem. Inf. Model, *2023*, DOI: 10.1021/acs.jcim.3c00434
3. van den Maagdenberg et al., github.com/CDDLeiden/QSPRPred
4. Hassen et. al (In preparation)

Acknowledgements & Funding

M.Š. is supported by Czech Science Foundation Grant No. 22-17367O and by the Ministry of Education, Youth and Sports of the Czech Republic (project number LM2023052).

Author Information:

@ m.sicho@lacdr.leidenuniv.nl
github.com/martin-sicho
@martin_sicho