

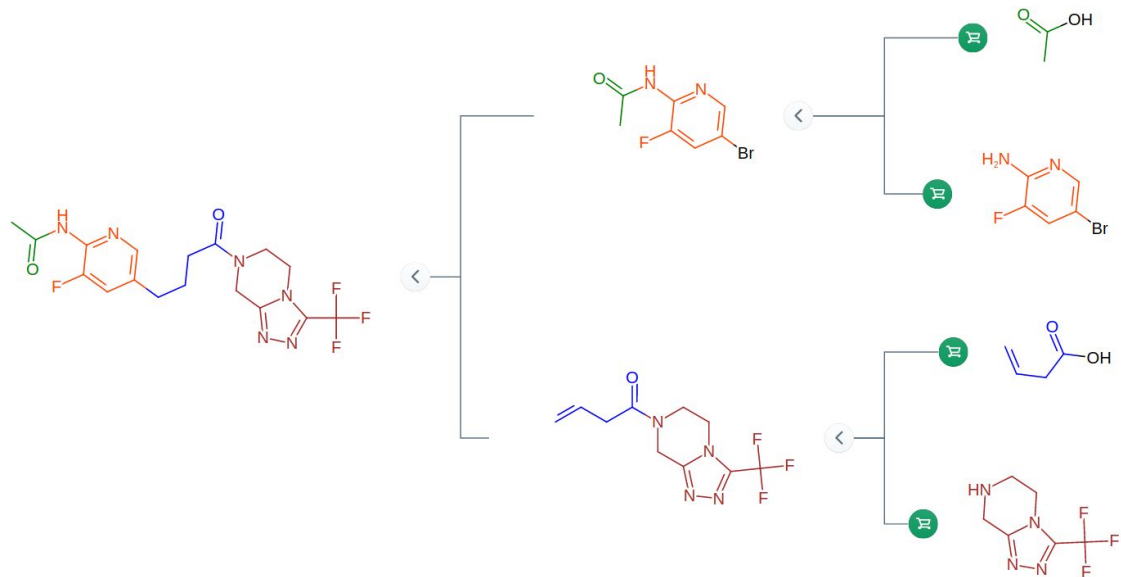
# M1 Synthetic Accessibility Score



Molecule.one

How Do We Compute It?

# Our system finds actual synthesis paths



... and transform the paths into the M1 synthetic accessibility score of target compounds

M1 score = 13.27

# Factors affecting the score

- Number of steps in synthesis path
- Certainty of each step
  - We are able to assess it with AI
- Cost of starting materials
  - Real prices from eMolecules database
- Shape of the synthesis tree (convergent/linear)
- Reactions application order
  - Risky reactions are preferred to be used as initial synthesis steps

# Results Overview

# Accessibility score for structurally similar compounds

- Seeing only compound structure, any heuristic score is unlikely to distinguish accessibility of analogous compounds.
- M1 (our) score in general agrees with Ertl<sup>1</sup> score, but knowing the whole synthesis path allows it to differentiate even almost identical targets.

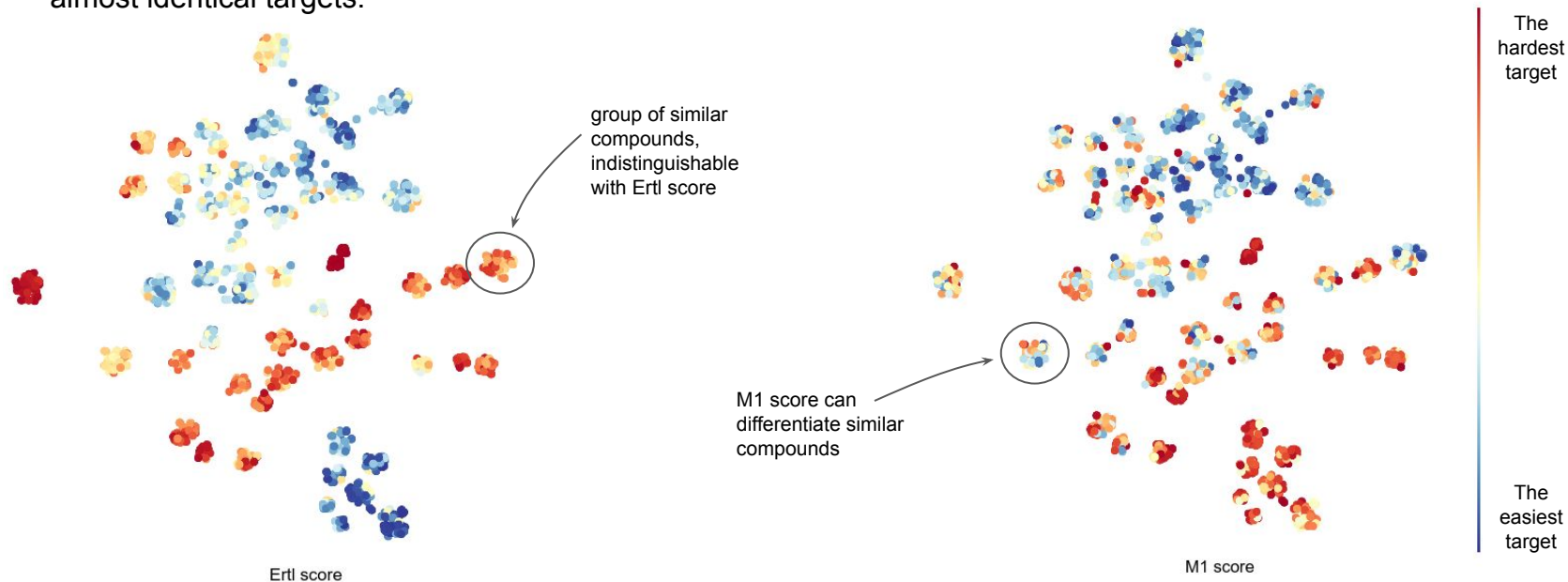


Fig. Groups of similar compounds, arranged spatially by Morgan fingerprint similarity. Colouring corresponds to target difficulty, assessed by Ertl (left) and M1 (right) score.

# Accessibility of compounds with certain synthesis paths

M1 score can distinguish between compounds with particularly certain synthesis paths (blue) and compounds with regular synthesis paths (orange) better than a heuristic score.

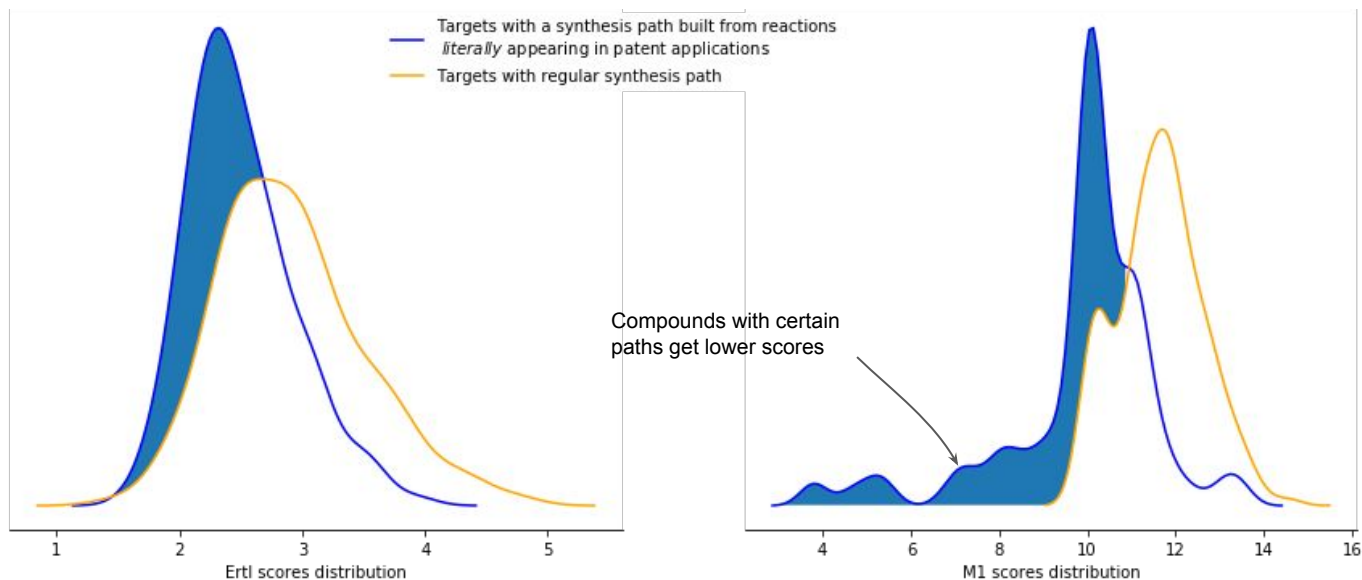


Fig. Histograms of Ertl (left) and M1 (right) scores computed for target compounds appearing in patent applications. Both scores are increasing with increased target difficulty.

# Accessibility of compounds with short synthesis paths

One of the factors influencing M1 score (right) is the number of reactions in the synthesis paths. Heuristic score (left) cannot directly take it into account.

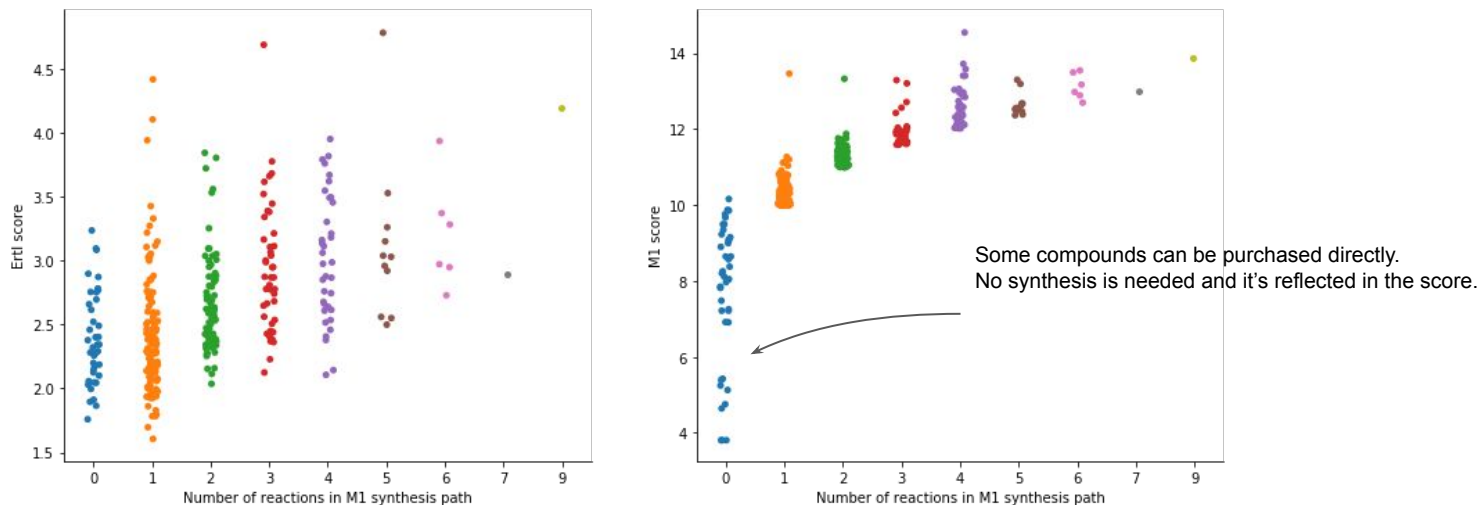


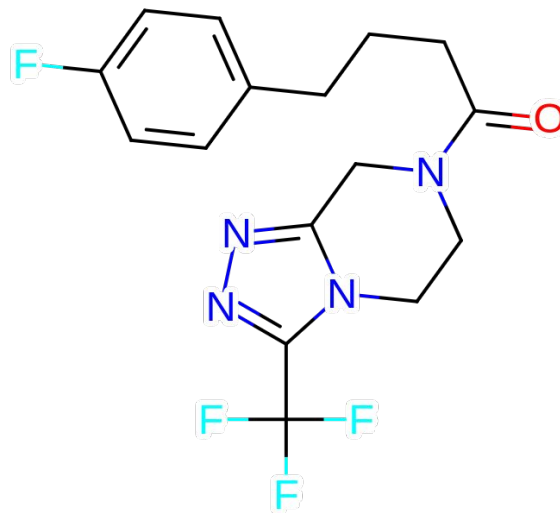
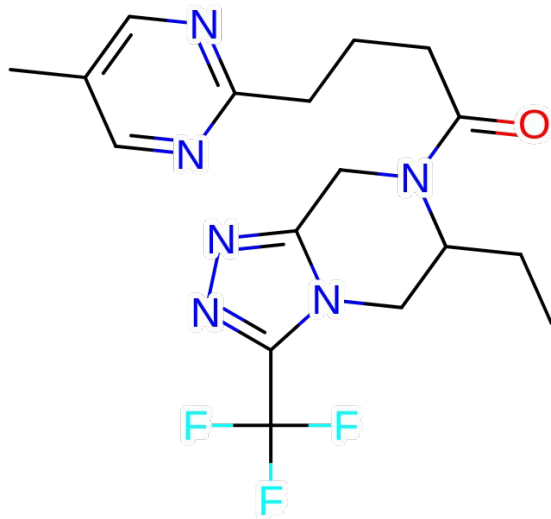
Fig. Relationship between synthesis path size and Ertl (left) and M1 (right) scores.



# Case Studies

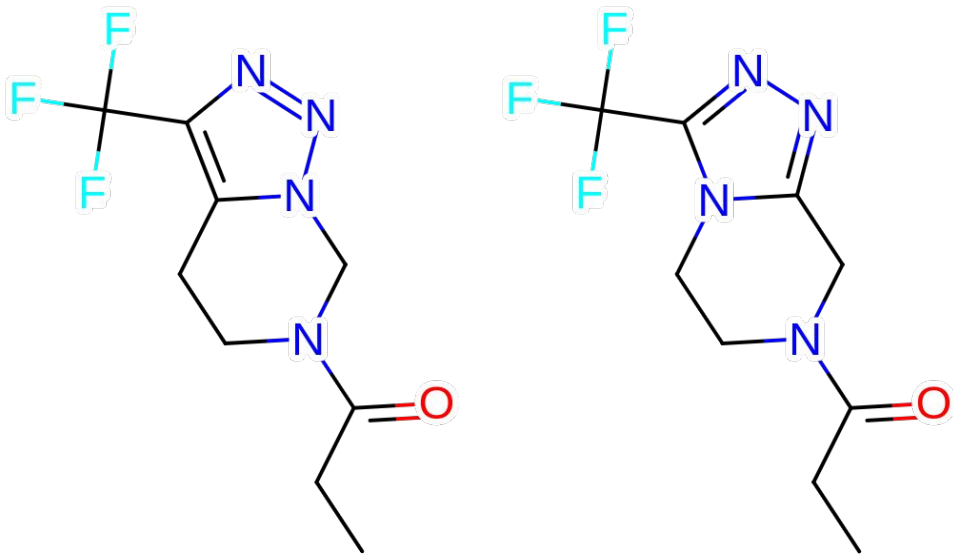
# Starting materials matter

These molecules are almost identical with regards to Ertl score, but one on the right is makeable in one step, whereas one on the left requires at least three steps more due to inavailability of starting materials.



# Sensitivity to small (but significant) modifications

Again, these molecules are almost identical w.r.t. Ertl score, but the one on the right has a straightforward synthesis, whereas the one on the left poses a significant challenge due to an unusual heterocycle.



# Supplementary Information

## Appendix A

# M1 SAScore agrees with Ertl score on “dataset level”

Average M1 scores agree with average Ertl scores for targets sampled from different datasets. Both scores reflect the expected increasing difficulty of targets from respective sources.

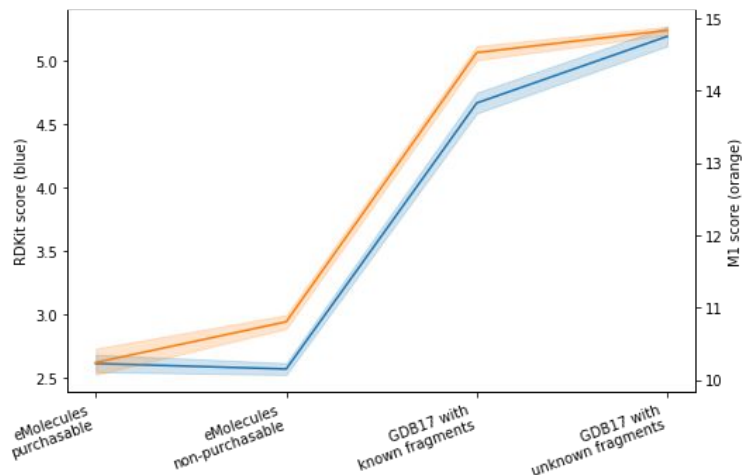


Fig. Average Ertl (orange) and M1 (blue) score computed on targets from different datasets.

Target sources (from left to right):

- Compounds from eMolecules database, directly purchasable
- Compounds from eMolecules database, with limited availability
- Compounds from GDB17<sup>1</sup> dataset, with compound fragments appearing in eMolecules
- Compounds from GDB17 dataset, with compound fragments not seen in any of eMolecules compounds