

Generative AI for Synthesizable Antibiotic Design

Kyle Swanson, Gary Liu, Denise Catacutan, Autumn Arnold,
James Zou, Jonathan Stokes

Nature Machine Intelligence, 2024

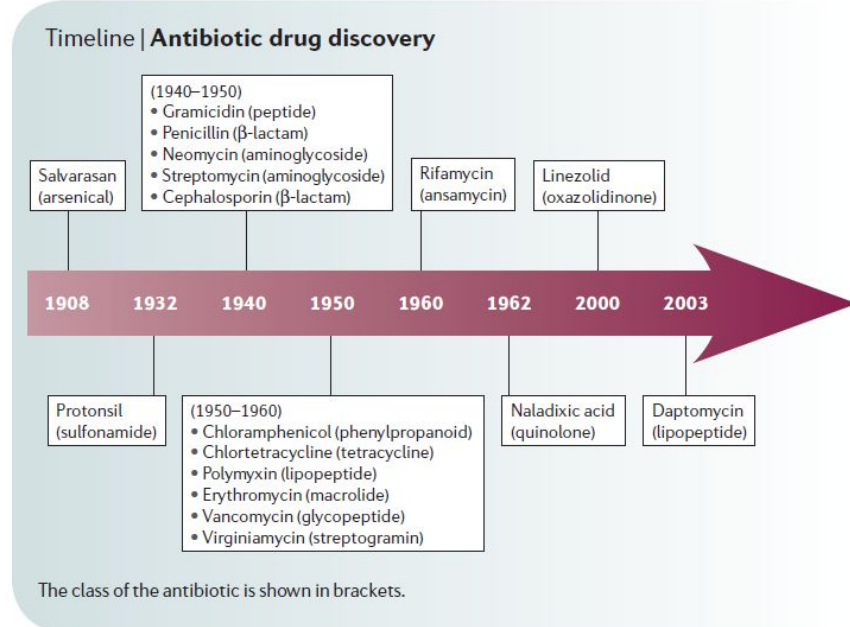
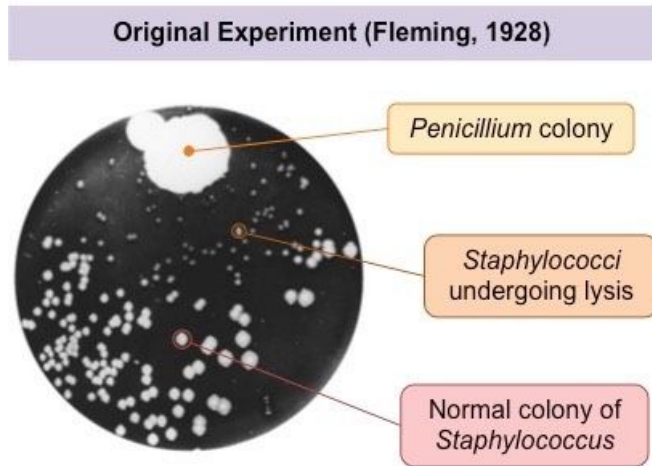


Brief history of antibiotics

1928: Alexander Fleming discovers penicillin

1940-60: Many new antibiotics discovered

1960-now: Few structurally novel antibiotics



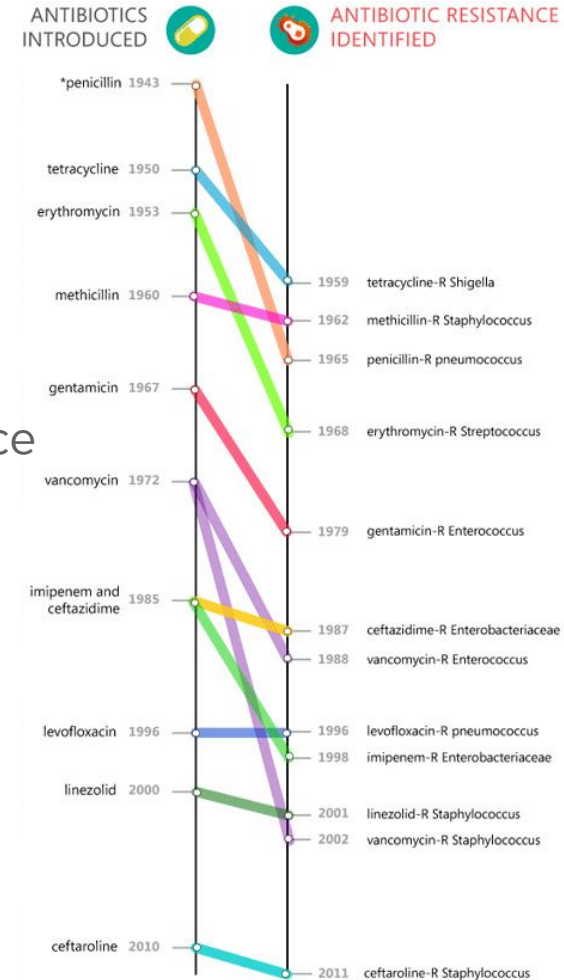
Drug resistant bacteria

Bacteria develop resistance to antibiotics

2019: 1.27 million people likely died from antibiotic resistance

2050: 10 million people may die from resistance annually

Takeaway: We need new antibiotics!

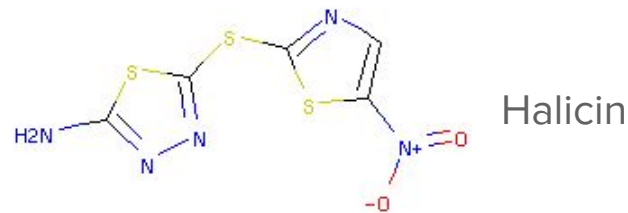


Property prediction

[A Deep Learning Approach to Antibiotic Discovery](#), Stokes, et al., *Cell*, 2020

Train GNN on 2.5K molecules with known *E. coli* inhibition (5% active)

Property prediction

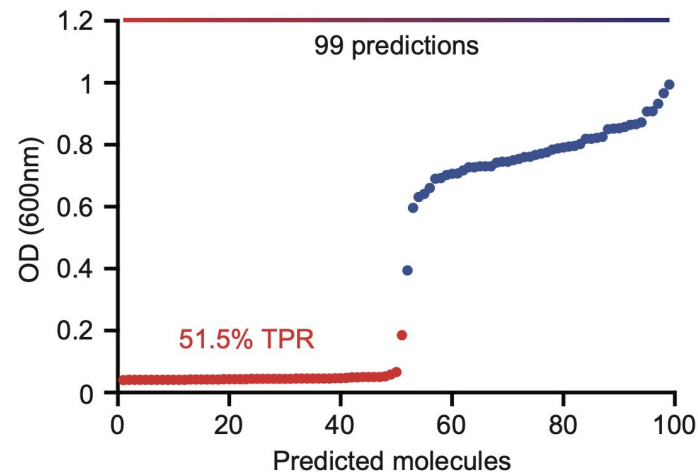


[A Deep Learning Approach to Antibiotic Discovery](#), Stokes, et al., *Cell*, 2020

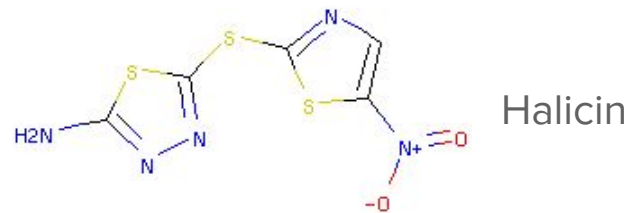
Train GNN on 2.5K molecules with known *E. coli* inhibition (5% active)

Predict on 6K molecules

- 51.5% active among top 99 predictions
- Halicin targets multi-drug resistant bacteria



Property prediction



[A Deep Learning Approach to Antibiotic Discovery](#), Stokes, et al., *Cell*, 2020

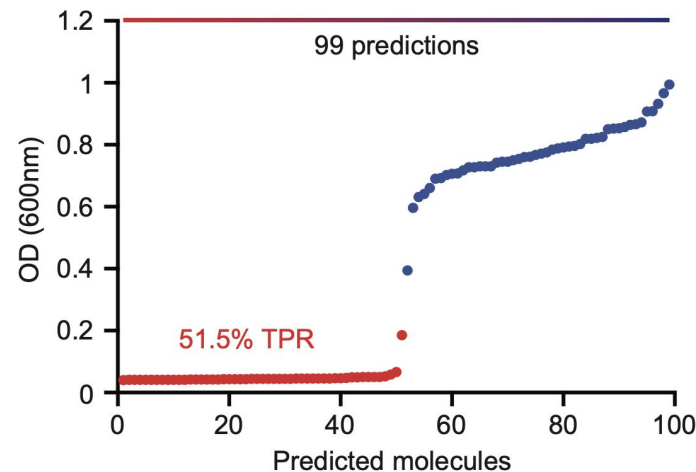
Train GNN on 2.5K molecules with known *E. coli* inhibition (5% active)

Predict on 6K molecules

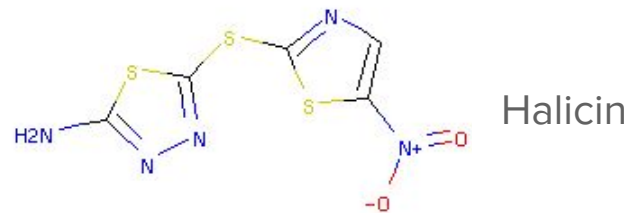
- 51.5% active among top 99 predictions
- Halicin targets multi-drug resistant bacteria

Predict on 107M molecules

- 4 days of computation
- 8 structurally novel antibiotics among top 23



Property prediction



[A Deep Learning Approach to Antibiotic Discovery](#), Stokes, et al., *Cell*, 2020

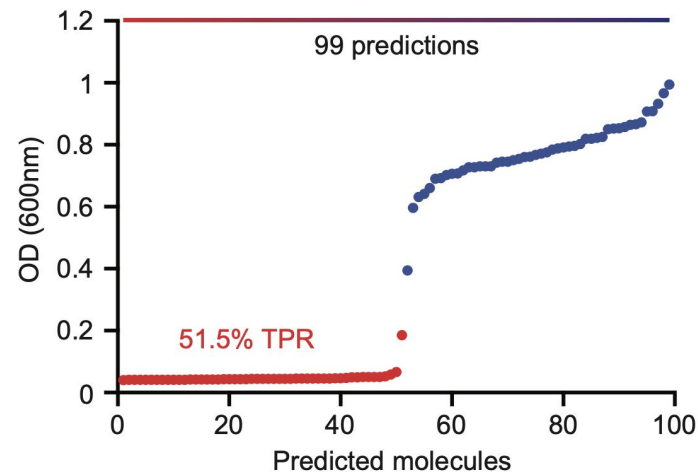
Train GNN on 2.5K molecules with known *E. coli* inhibition (5% active)

Predict on 6K molecules

- 51.5% active among top 99 predictions
- Halicin targets multi-drug resistant bacteria

Predict on 107M molecules

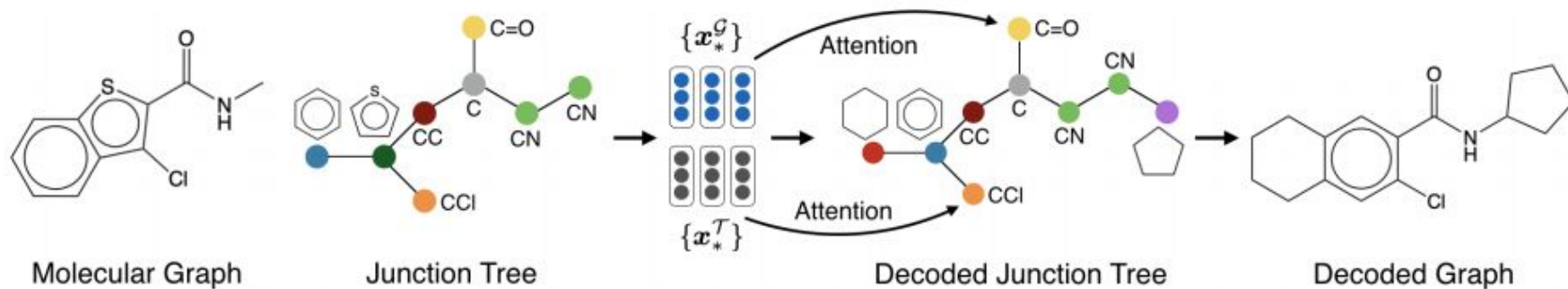
- 4 days of computation
- 8 structurally novel antibiotics among top 23



Limitation: No intelligent search $\Rightarrow$ scales poorly to larger chemical spaces

Generative models

Model: Generative models directly design molecules with desirable properties



Benefit: Rapid design of good molecules without slow search

Limitation: Generated molecules are difficult to synthesize $\Rightarrow$ practically useless

Synthesis-aware generative model

Goals

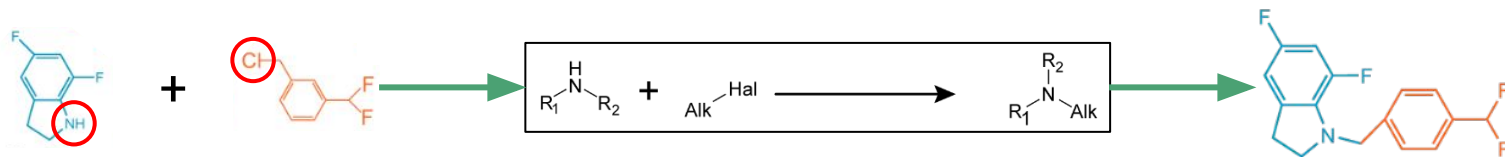
- 1) Build a generative model that guarantees synthesizability
- 2) Generate, synthesize, and experimentally validate generated molecules

Synthesis-aware generative model

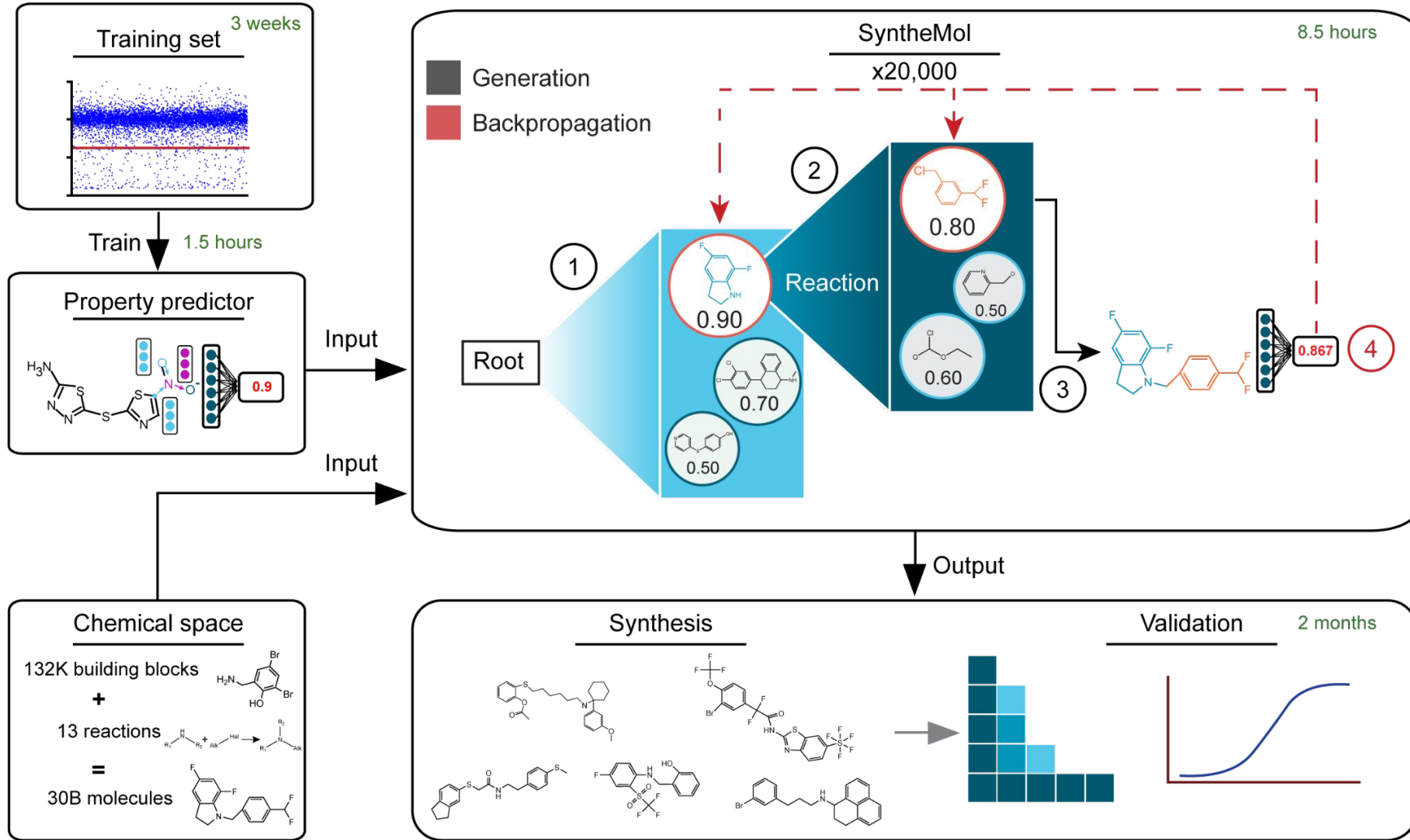
Goals

- 1) Build a generative model that guarantees synthesizability
- 2) Generate, synthesize, and experimentally validate generated molecules

Idea: Design molecules with off-the-shelf building blocks + easy chemical reactions



Application: Generate structurally novel antibiotics



Training set

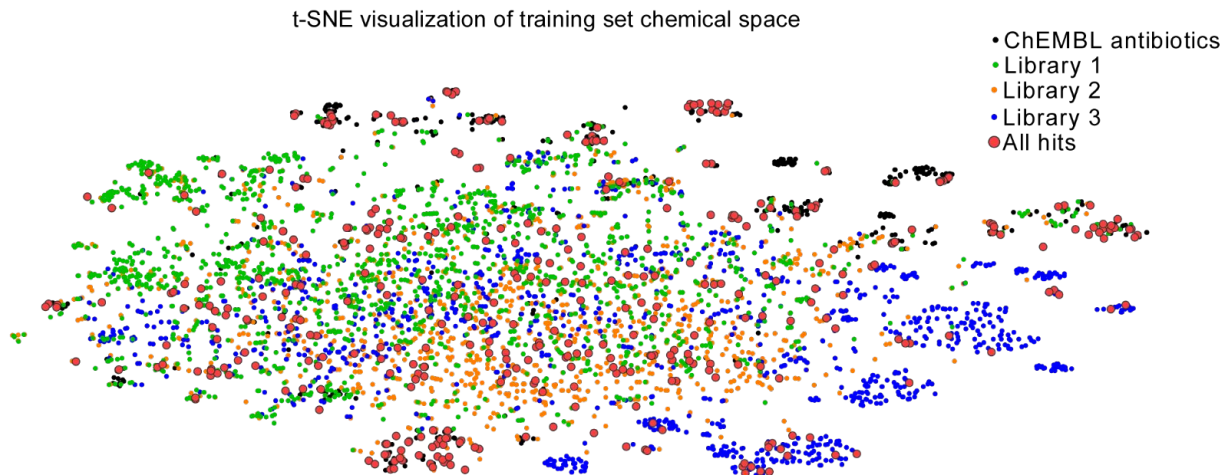
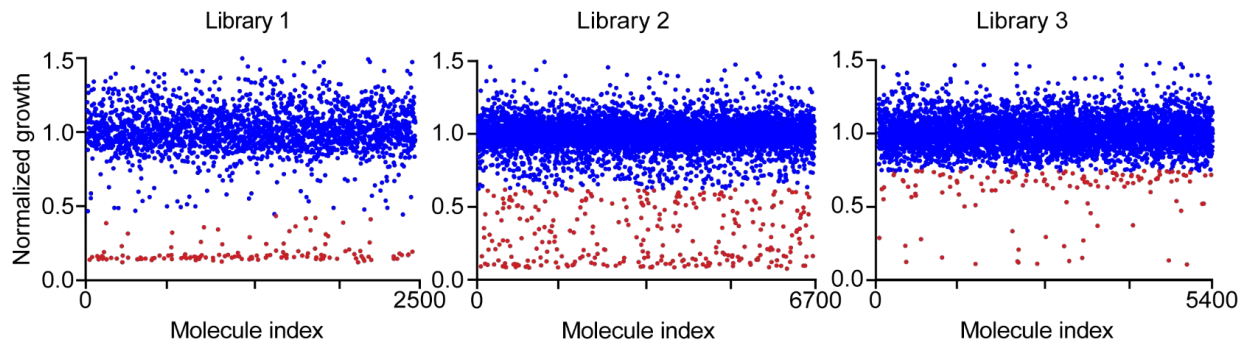
Target: *A. baumannii*

3 libraries

- 2 drug repurposing
- 1 synthetic compounds

13,524 molecules

- 470 active
- 13,054 inactive



Property predictor

Three models

1) Chemprop

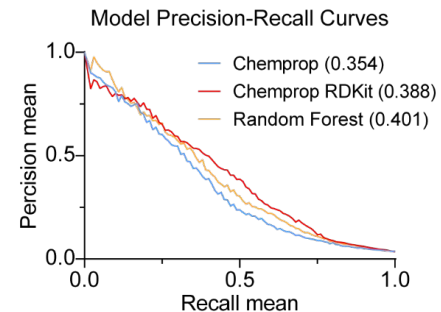
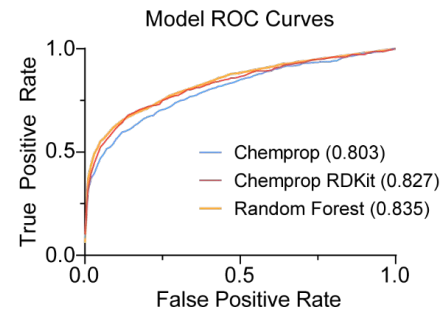
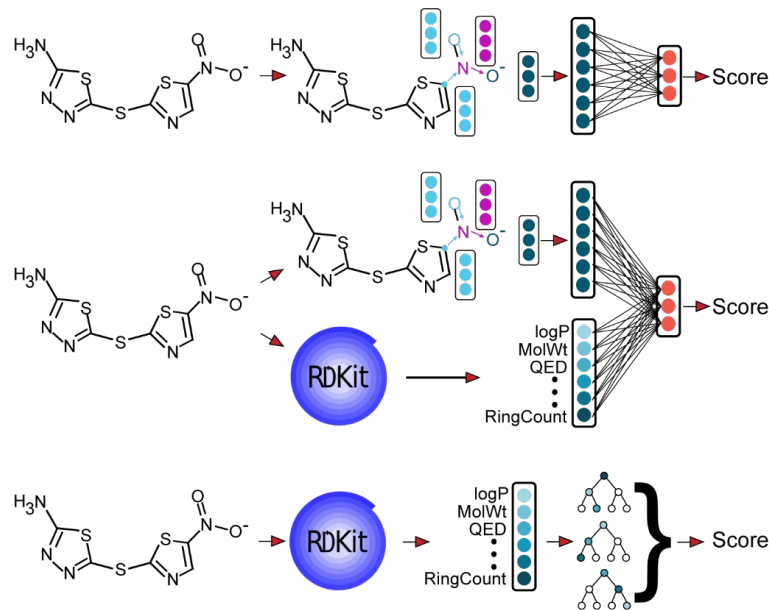
a) GNN

2) Chemprop RDKit

a) GNN + 200 features

3) Random Forest

a) RF on 200 features

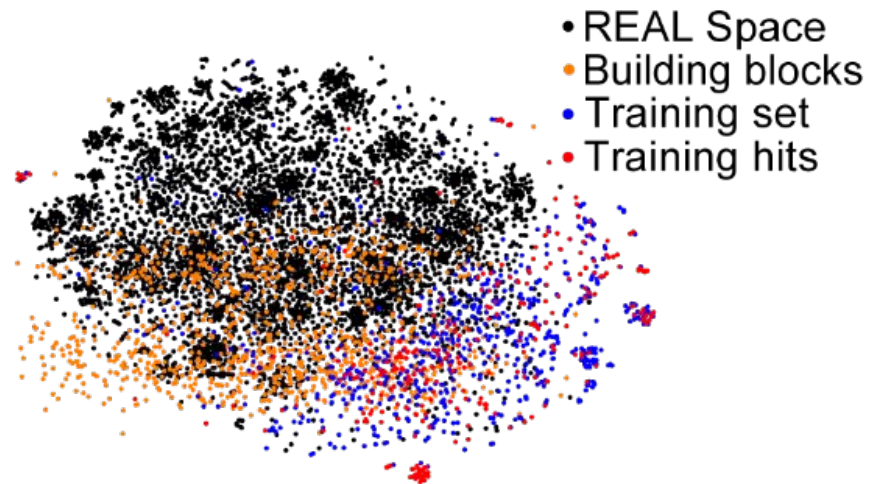


Performance: ROC-AUC = 0.80–0.84 and PRC-AUC = 0.35–0.40 on 10-fold CV

Chemical space

Enamine REAL Space: 31 billion molecules

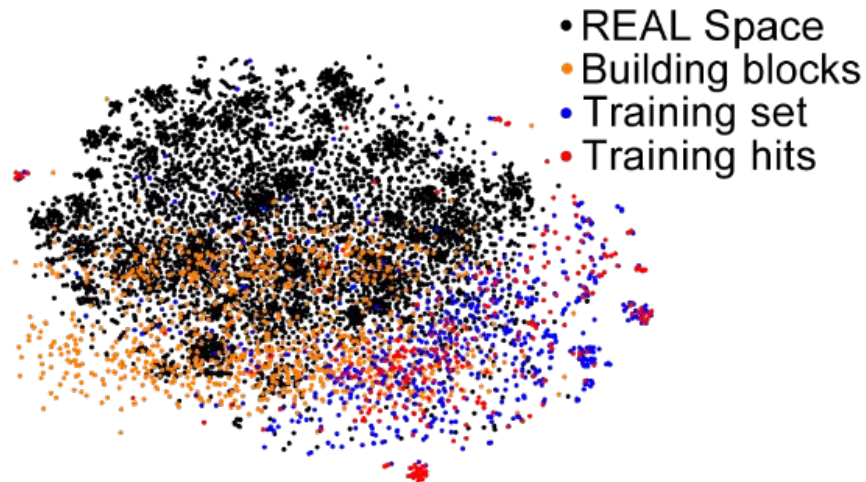
- 138,000 building blocks
- 169 chemical reactions



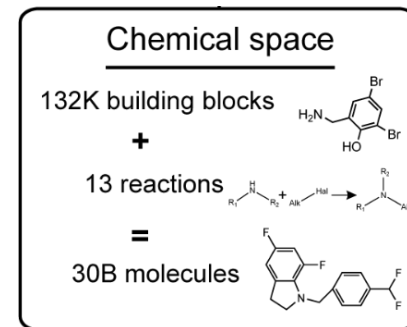
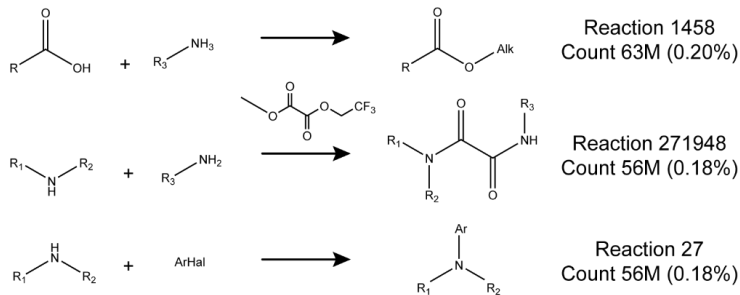
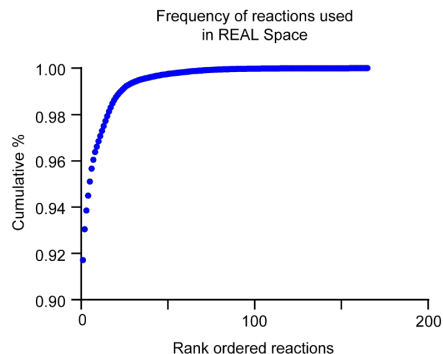
Chemical space

Enamine REAL Space: 31 billion molecules

- 138,000 building blocks
- 169 chemical reactions



Simplification: 96.6% of molecules with 13 reactions

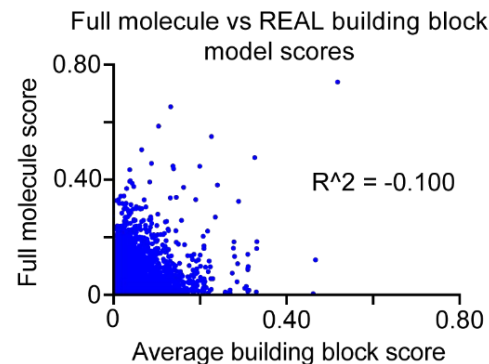
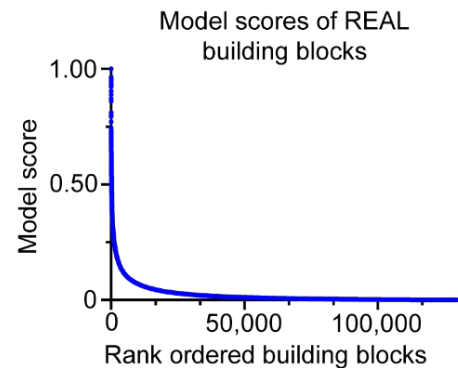


Generative model

Greedy: build molecules with highest scoring building blocks

Problems

- Few building blocks have high scores $\Rightarrow$ low diversity
- Building block scores not correlated with full molecule



Generative model

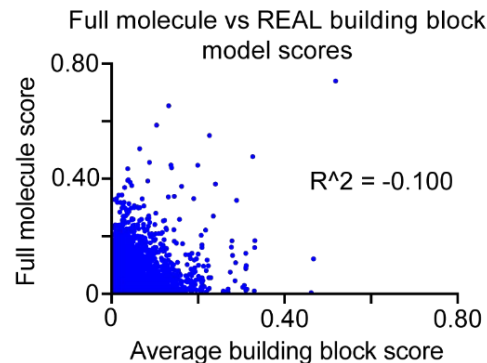
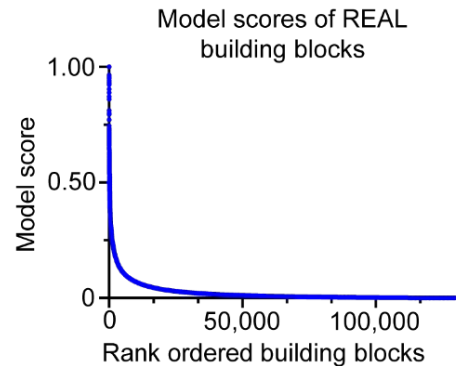
Greedy: build molecules with highest scoring building blocks

Problems

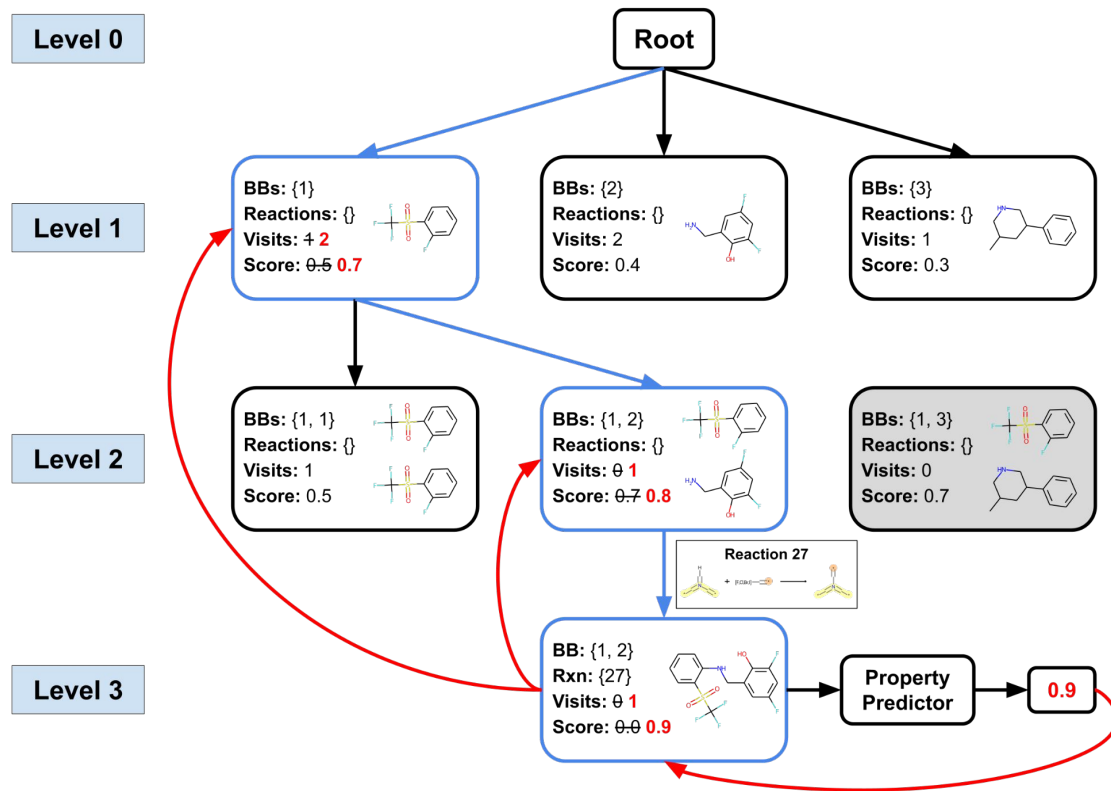
- Few building blocks have high scores $\Rightarrow$ low diversity
- Building block scores not correlated with full molecule

Monte Carlo tree search (MCTS)

- **Exploration:** construct diverse molecules
- **Exploitation:** use full molecule scores to guide search



SyntheMol: MCTS guided by property predictor



$$S(N) = \frac{Q(N) + P(N) \cdot U(N)}{D(N)}$$

$Q(N)$ **Exploit:** average full molecule score

$P(N)$ **Property:** average building block score

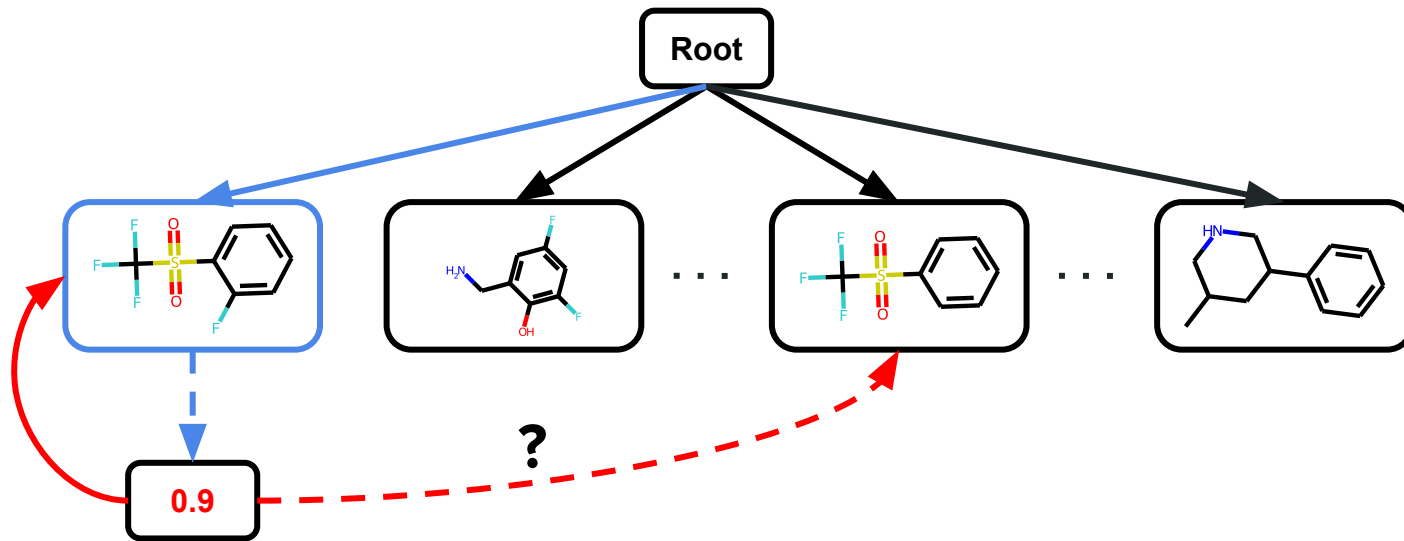
$U(N)$ **Explore:** visit count vs sibling nodes

$D(N)$ **Diversity:** frequency of building block use

Limitations of MCTS

Independence: MCTS treats nodes independently, ignoring chemical similarity

Coverage: First level alone has 132k nodes $\Rightarrow$ cannot test all with 20k rollouts

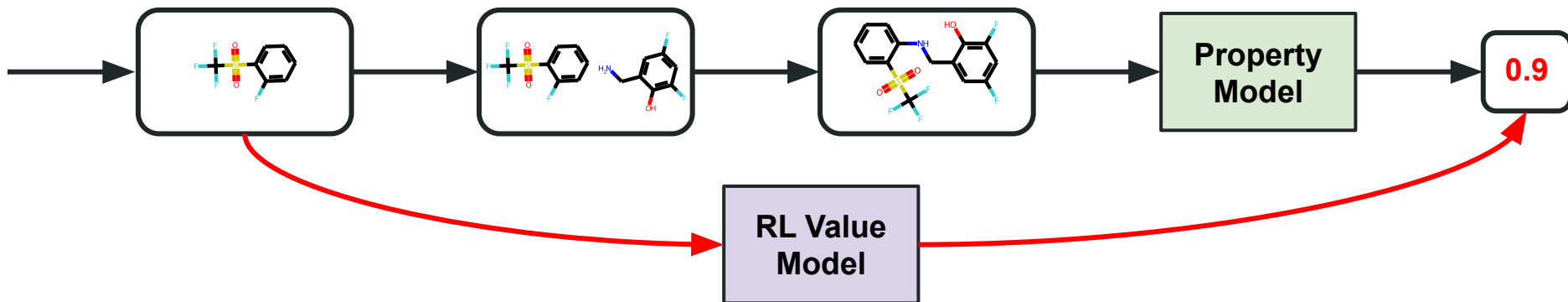


Result: MCTS is not an efficient value function for building blocks

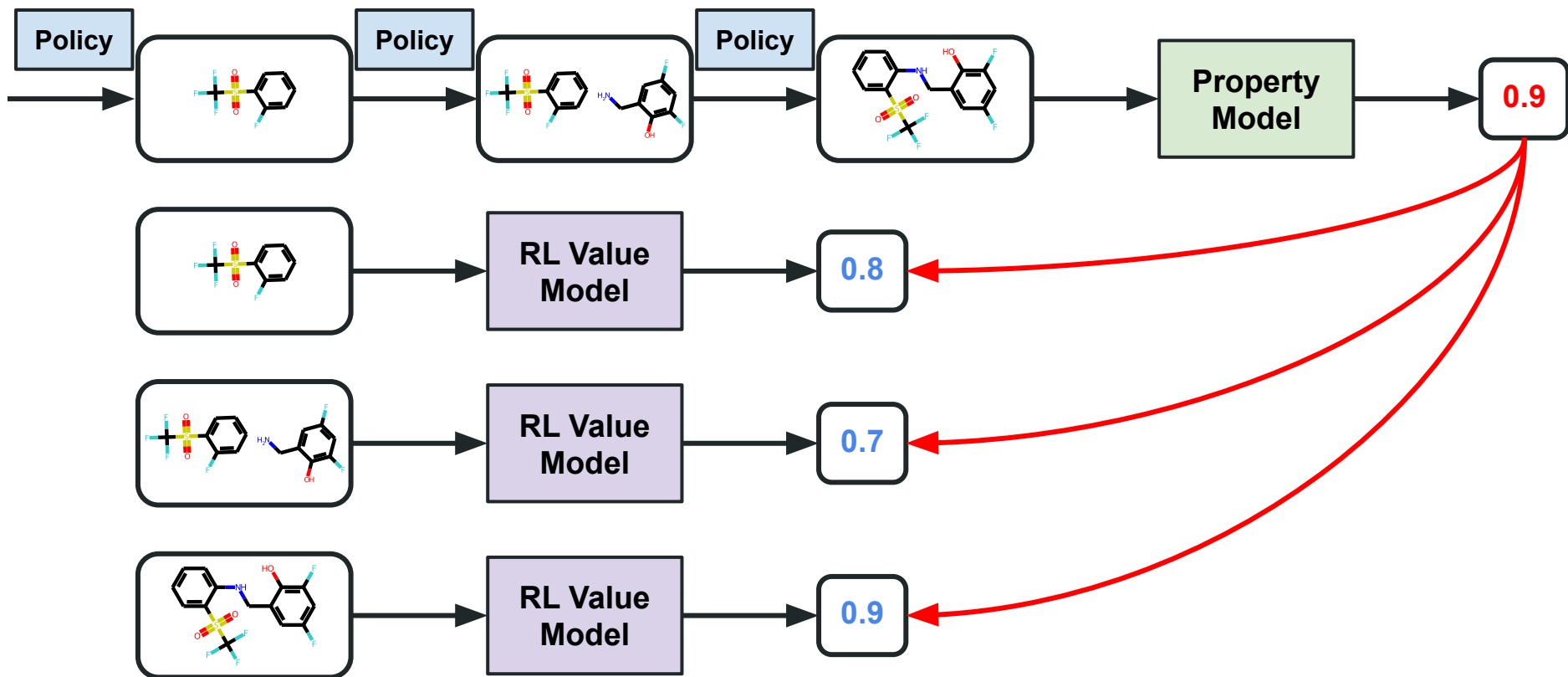
MCTS $\Rightarrow$ RL

Idea: Use reinforcement learning (RL) in place of MCTS

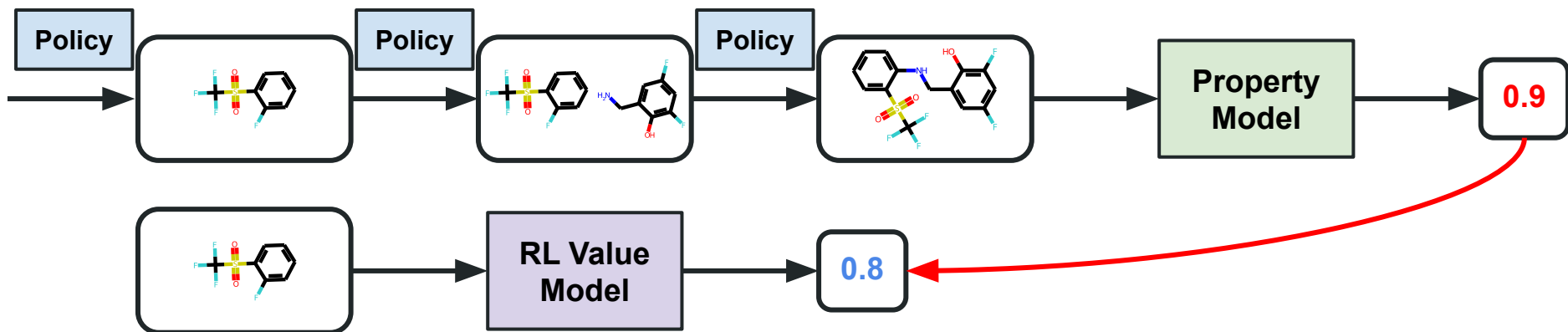
- **MCTS** *computes* a value for each BB separately
- **RL** *learns* a value function that generalizes to chemically similar BBs



RL method



RL method



**Property
Model**

GNN pretrained to predict
molecular property (fixed)

**RL Value
Model**

GNN trained to predict full
molecule score from
building block(s)

Policy

Sample building blocks
proportional to their score
with temperature control

$$P(BB) \propto e^{\text{value}(BB)/T}$$

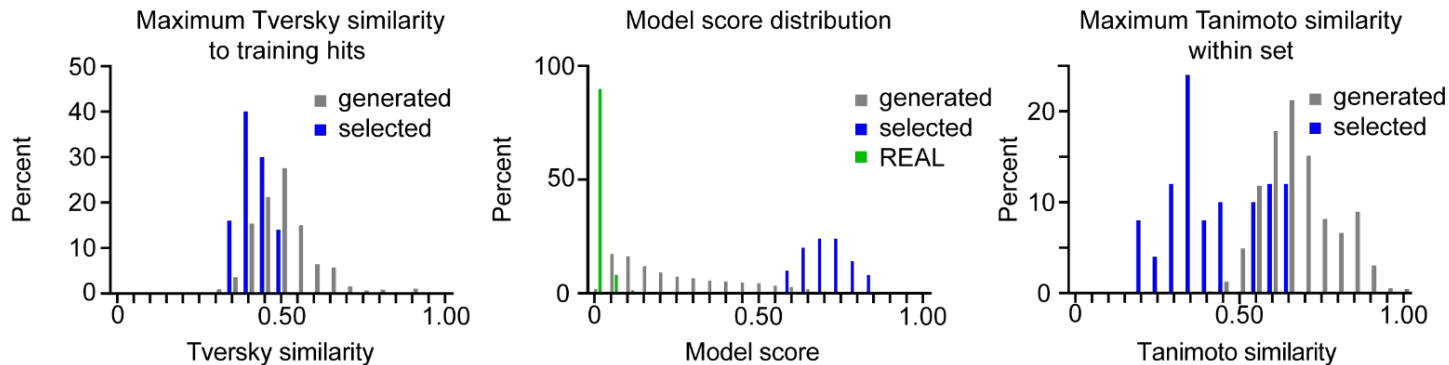
SyntheMol: antibiotic design

Generations: SyntheMol-MCTS for 20,000 rollouts guided by 3 property predictors

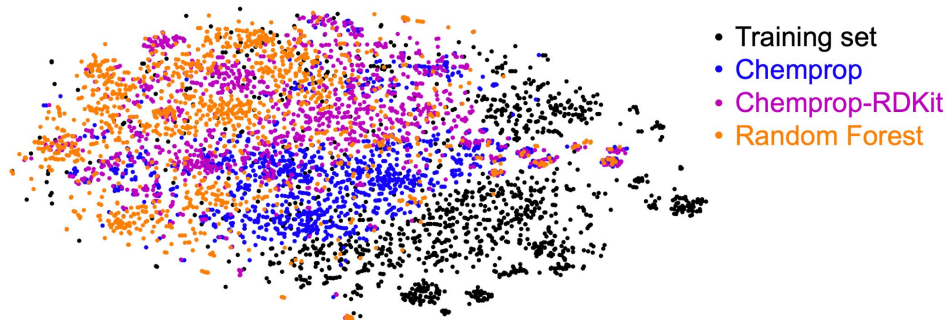
Filters to select optimal molecules

- 1) **Novel:** avoid analogs of known antibiotics
 - a) Tversky similarity(generated, antibiotic) ≤ 0.5
- 2) **Effective:** high property prediction score
 - a) Top 20% of molecules by score
- 3) **Diverse:** avoid analogs of the same compound
 - a) K-means clustering with Tanimoto similarity

SyntheMol: antibiotic generations



t-SNE visualization of training and generated sets



Synthesis

Selected: 150 molecules (50 each from three models)

Requested: 70 molecules

- Not all 150 molecules are available from Enamine
- Reaction templates are overly simple $\Rightarrow$ not all matches are synthesizable

Synthesized: 58 molecules (83% success) in four weeks

- 26 Chemprop, 22 Chemprop RDKit, 10 random forest

Experimental validation

Experiment: Test generated molecules against *A. baumannii*

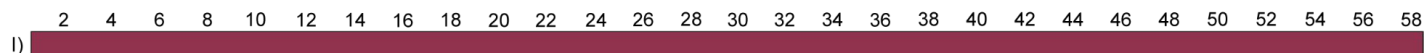
- Same growth inhibition assay as training set creation

Experimental validation

Experiment: Test generated molecules against *A. baumannii*

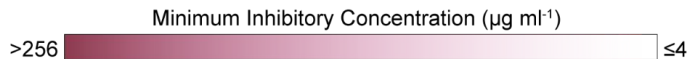
- Same growth inhibition assay as training set creation

Result: No molecules worked



LEGEND

1) *A. baumannii* ATCC 17978



Experimental validation

Challenge: Killing Gram-negative bacteria like *A. baumannii* requires **two** abilities

- 1) **Permeability:** Pass through double cell wall
- 2) **Activity:** Inhibit an essential component (e.g., protein)

Idea: What if our molecules have activity but lack permeability?

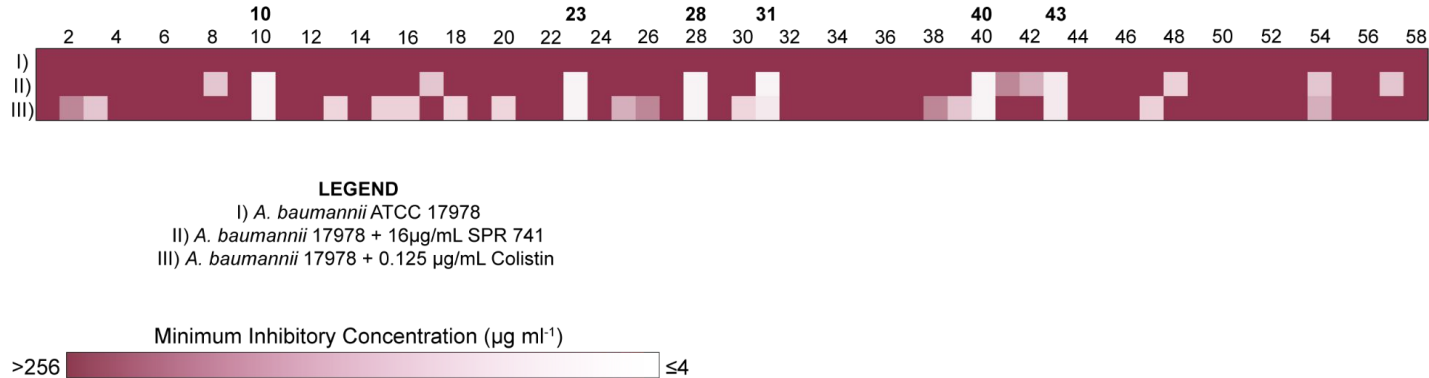
Experimental validation

Experiment: Couple generated molecules with a permeabilizer

Experimental validation

Experiment: Couple generated molecules with a permeabilizer

Result: Six of the molecules are extremely potent



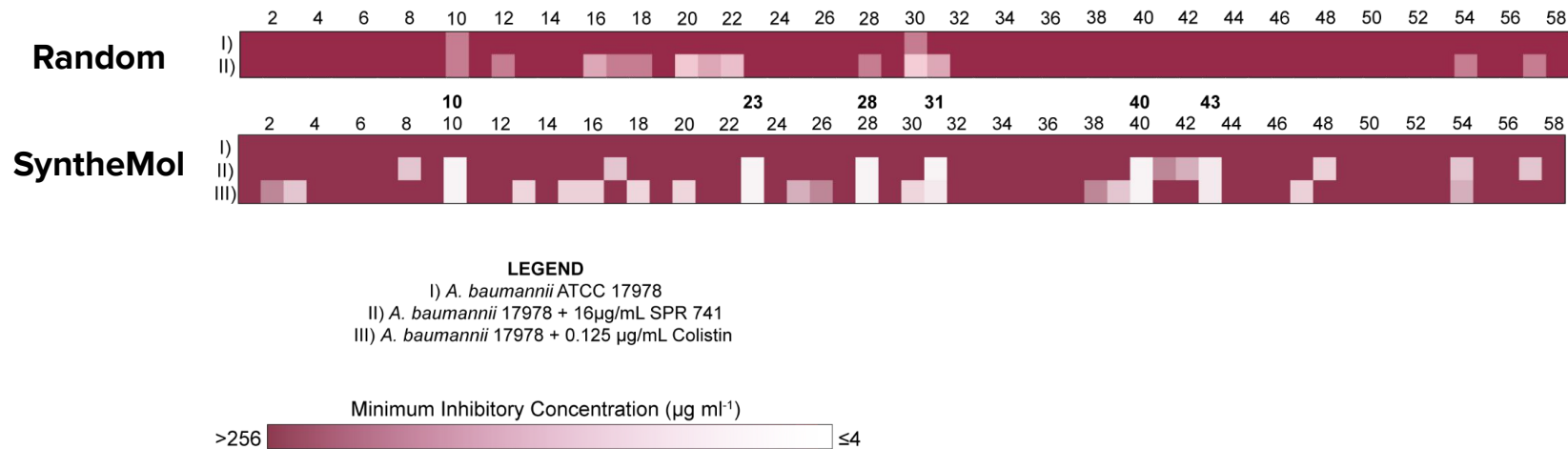
Experimental validation

Experiment: Test randomly selected molecules for comparison

Experimental validation

Experiment: Test randomly selected molecules for comparison

Result: Generated compounds are more effective than random ones



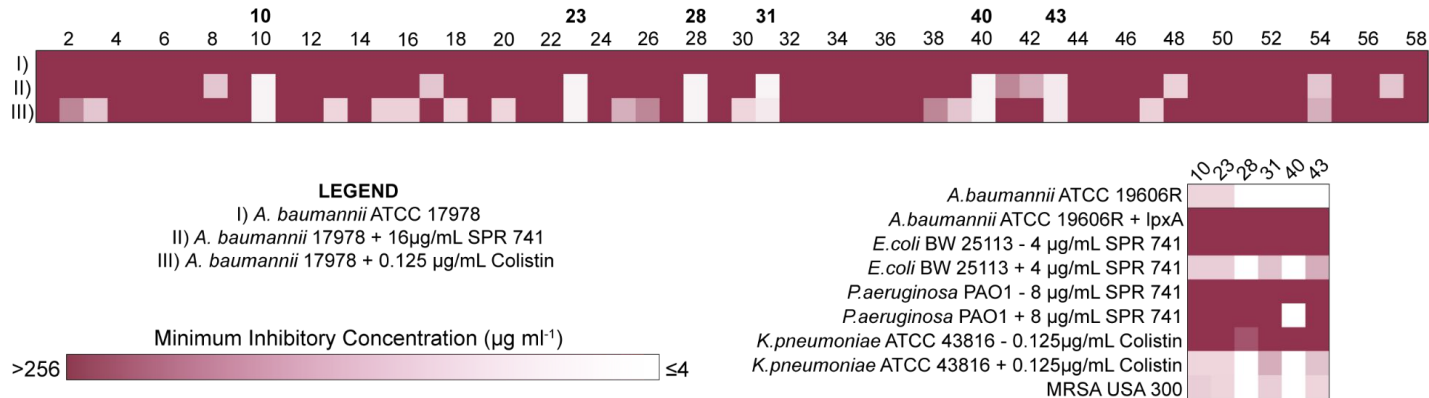
Experimental validation

Experiment: Test generated molecules against other bacterial species

Experimental validation

Experiment: Test generated molecules against other bacterial species

Result: Six potent molecules are broad spectrum (with permeabilizer)



GFlowNet comparison

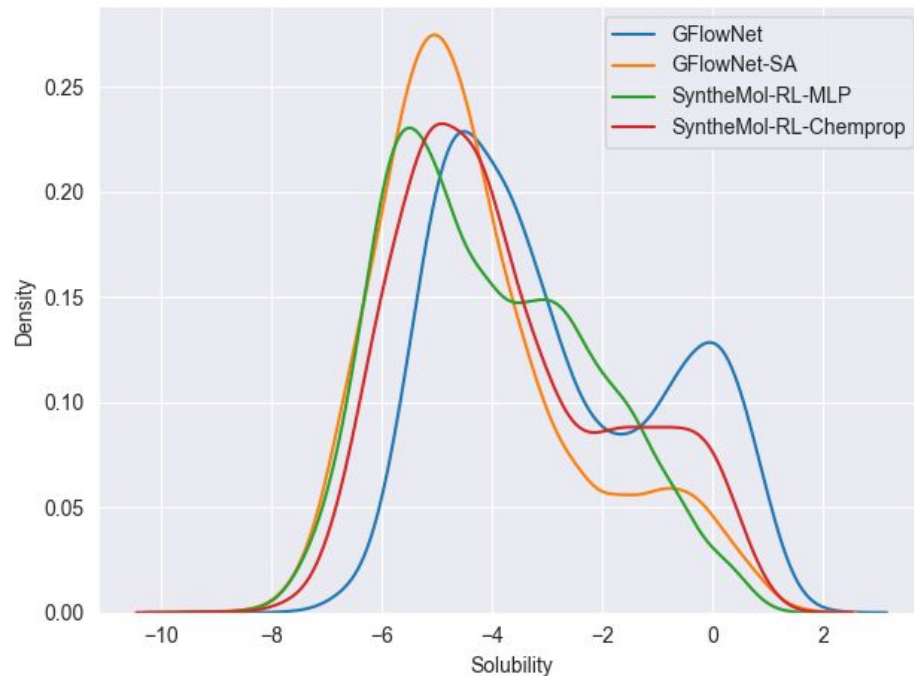
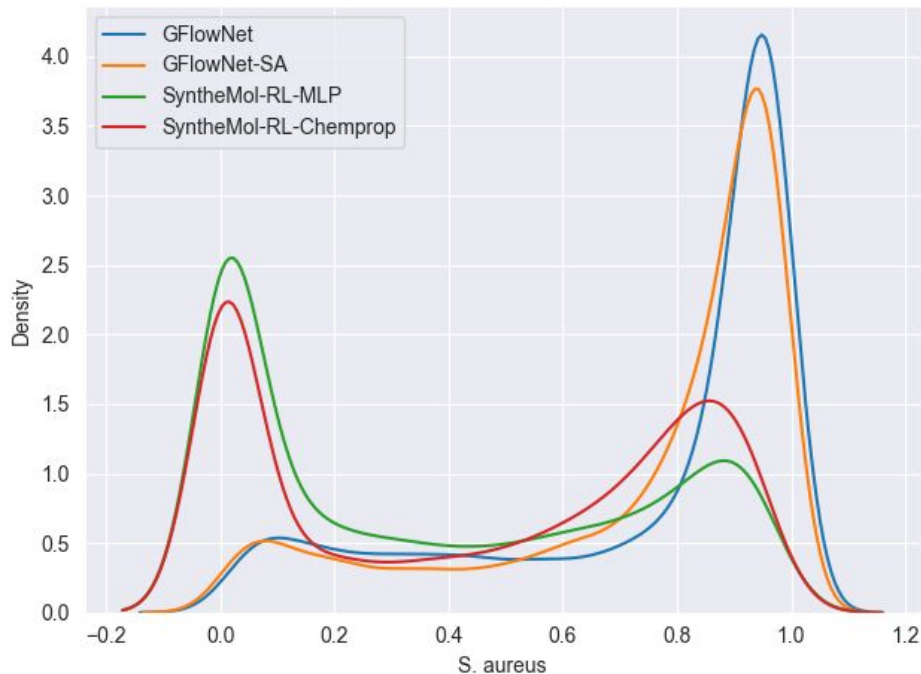
Question: How does SyntheMol compare to other generative models?

Multi-Objective GFlowNets (ICML, 2023)

- GFlowNets use RL + temperature scaling for diverse molecule generation
- Uses arbitrary molecular fragments **not** from known synthetic routes

Experiment: Modified GFlowNet to optimize for antibiotic efficacy (*S. aureus*), solubility, and optionally SAScore (synthesizability)

GFlowNet vs SyntheMol: generated



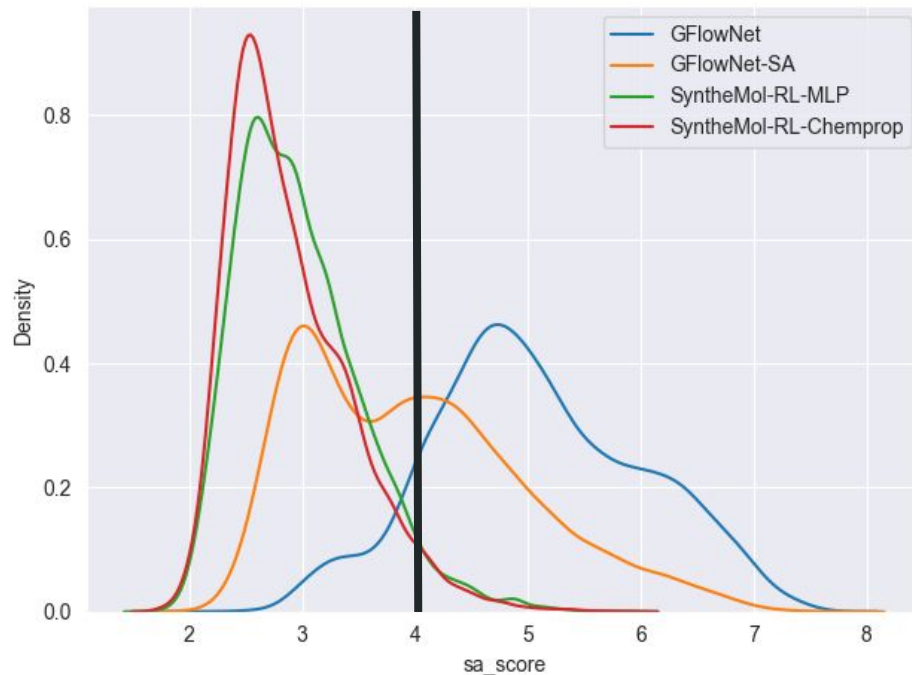
Takeaway: Appears that GFlowNet >> SyntheMol for generating antibiotics

GFlowNet vs SyntheMol: filtering

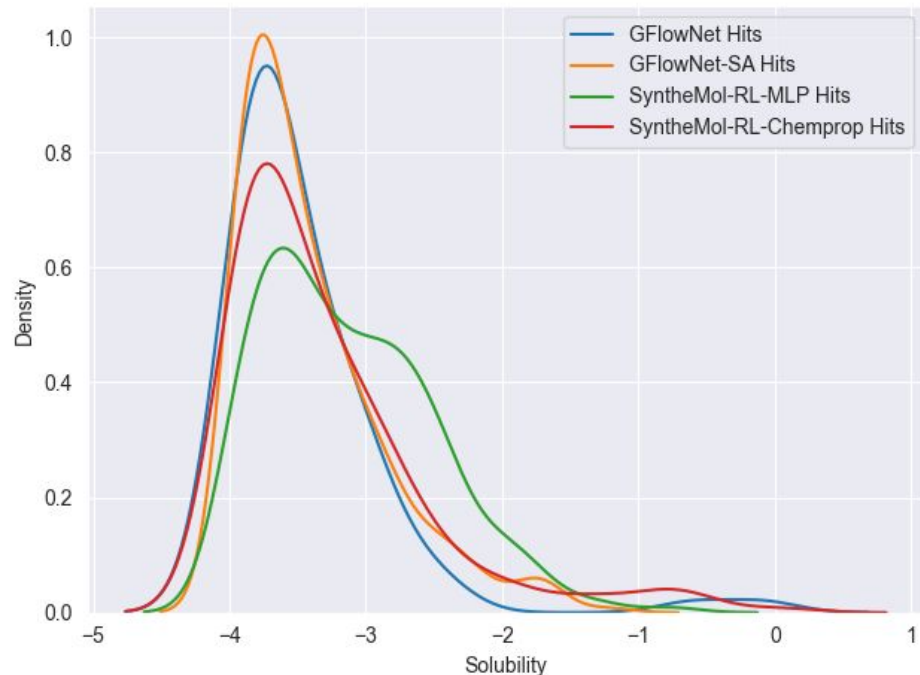
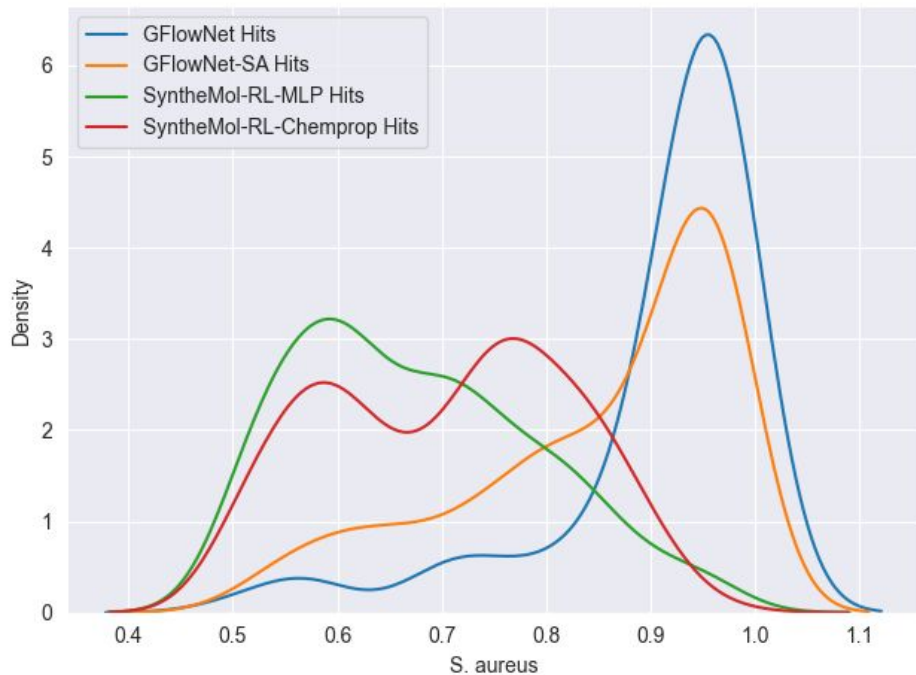
Limitation: GFlowNet molecules are less synthesizability based on SAScore

Synthesis filter: Need synthesizable compounds so filter by $\text{SAScore} \leq 4$

Selection: Then, apply typical filters for hits, novelty, and diversity



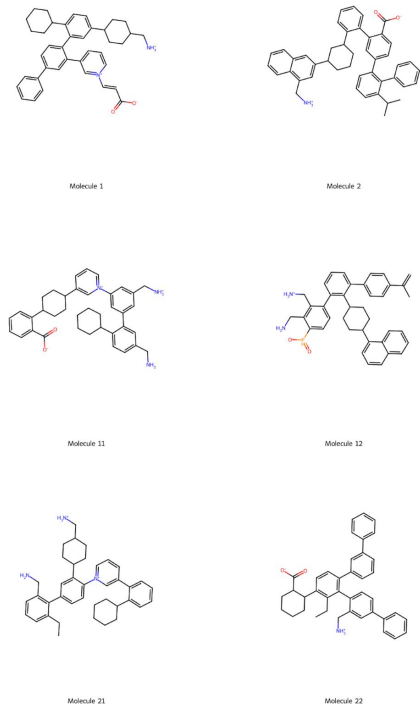
GFlowNet vs SyntheMol: selected



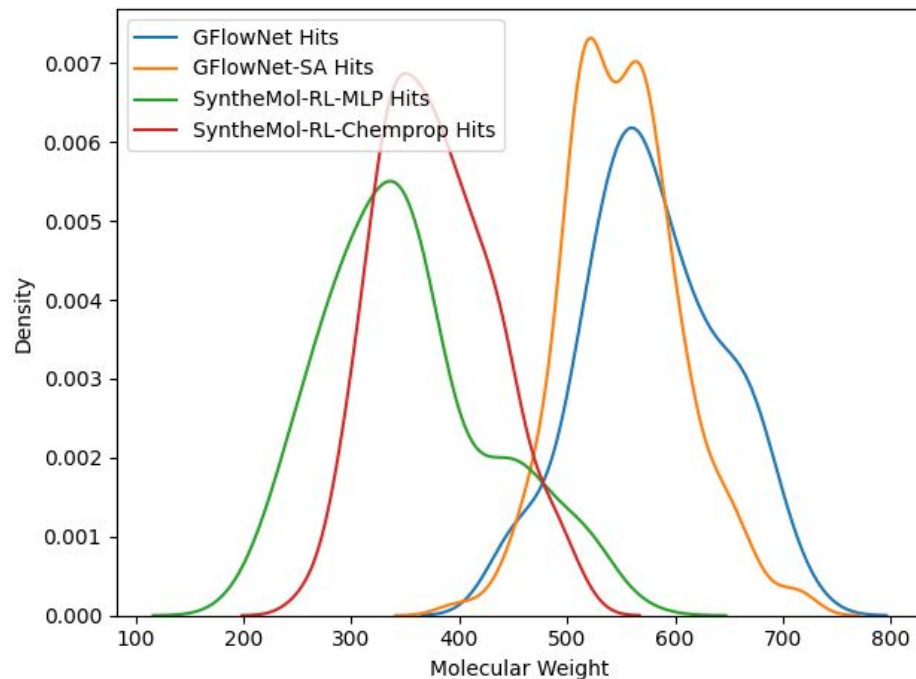
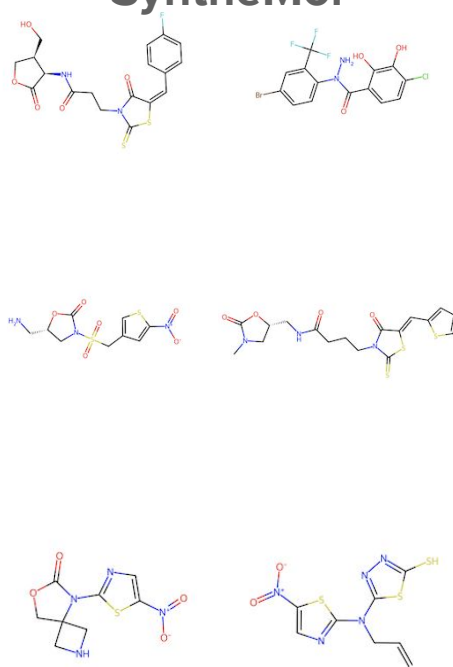
Takeaway: Appears that GFlowNet >> SyntheMol for generating *synthesizable* antibiotics

GFlowNet vs SyntheMol: selected

GFlowNet



SyntheMol



Takeaway: GFlowNet molecules are bulky, not drug-like, and look difficult to synthesize

GFlowNet: synthesis

Test: Sent 300 GFlowNet compounds to Enamine

Enamine: “...our chemistry group has reviewed list of cpds...and, unfortunately, we are not able to propose a synthesis. Our apologies for the inconvenience caused.”

Takeaway: GFlowNet designs impressive molecules according to ML-based objectives, but they are **not easily synthesizable** $\Rightarrow$ need SyntheMol!

Conclusion

SyntheMol is a **synthesis-aware generative model** for drug design

⇒ property predictor + MCTS/RL to explore vast chemical spaces

Filters select for **novel**, **effective**, and **diverse** generated molecules

We **synthesized** and **experimentally validated** 58 generated molecules

We discovered **six highly potent** and **structurally novel** antibiotic candidates

Questions?

Kyle Swanson

swansonk@stanford.edu

Generative AI for designing and validating easily synthesizable and structurally novel antibiotics, *Nature Machine Intelligence*, 2024.

github.com/swansonk14/SyntheMol

