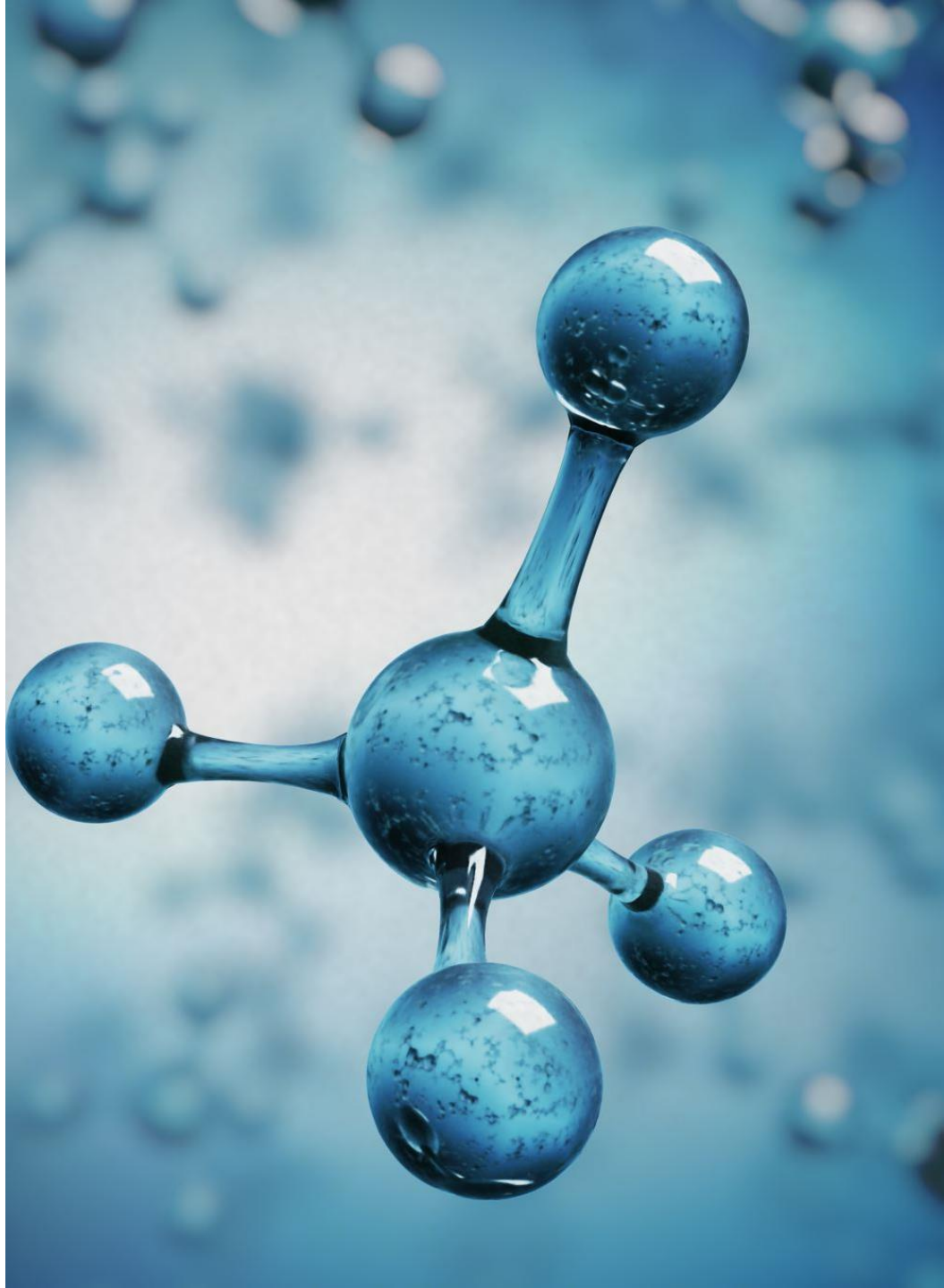
Merging enzymatic and synthetic chemistry with computational synthesis planning

An article written by Levin et al. (2022)

Cecilia Maijala, Elizaveta Sidorova, Nea Möttönen, Carl-Alfons Antson

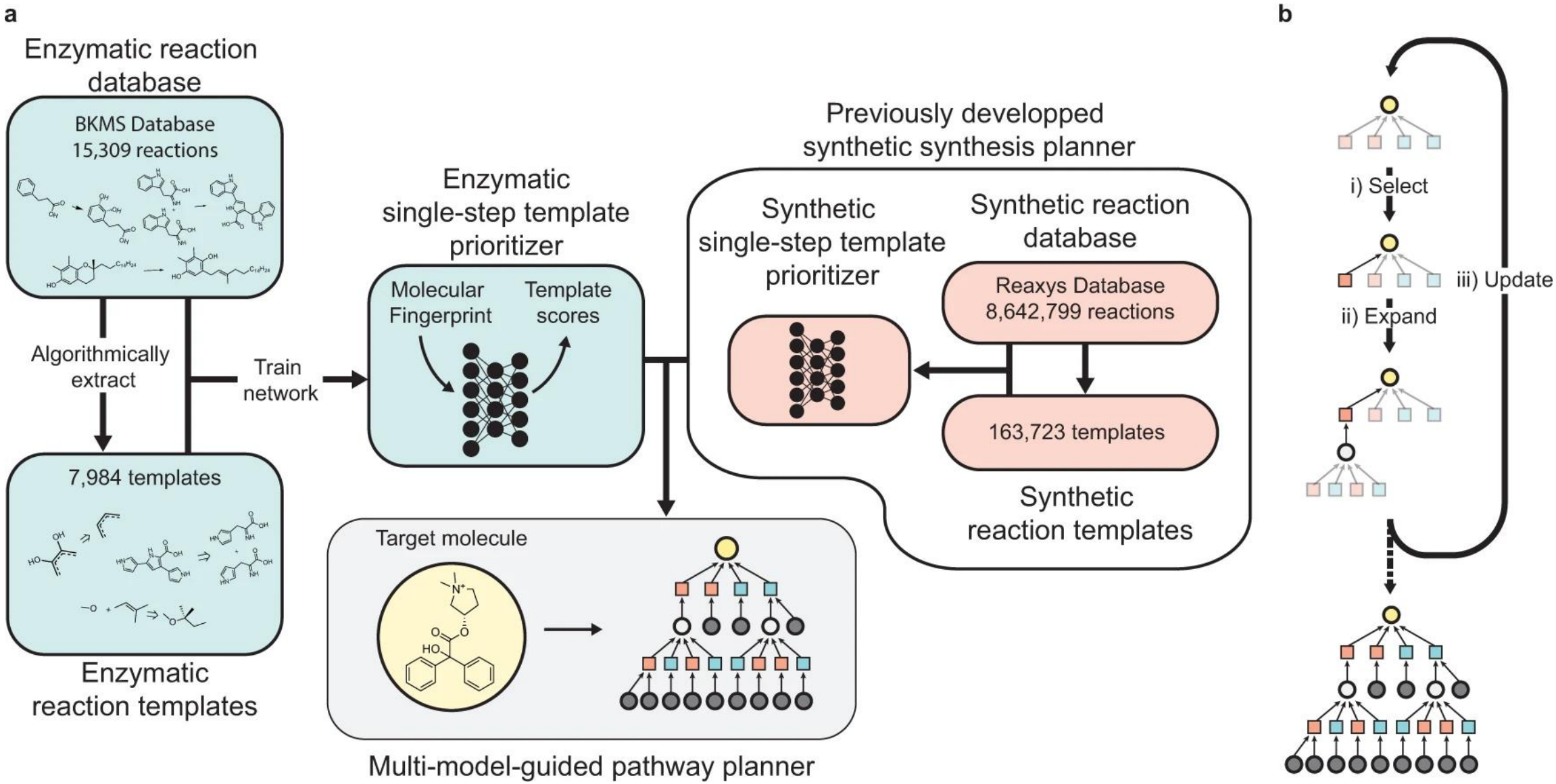
Introduction

- Synthesis planning: process of designing efficient routes to new molecules of interest
- Retrosynthesis: working backward from the target molecule
- A critical role in areas where complex molecules are important: e.g. drug discovery and materials science
- Current programs limited in their capacity to utilize rare chemical transformations
- Enzymatic reactions specific and sustainable but rare –> a major challenge



Aim

- To propose a new approach to synthesis planning that combines enzymatic and synthetic chemistry with computational methods
  - ➢ A search algorithm that generates multi-step synthesis plans
- Demonstration of retrosynthetic analysis with dronabinol and arformotelor
  - Comparison of enzymatic, synthetic and hybrid search
- New pathways for elusive molecules, shorter routes for others, discovery of completely new molecules


a
Enzymatic reaction database
BKMS Database
15,309 reactions
Algorithmically extract
7,984 templates
Enzymatic reaction templates
Train network
Enzymatic single-step template prioritizer
Molecular Fingerprint
Template scores
Previously developped synthetic synthesis planner
Synthetic single-step template prioritizer
Synthetic reaction database
Reaxys Database
8,642,799 reactions
163,723 templates
Synthetic reaction templates
Target molecule
Multi-model-guided pathway planner
b
i) Select
ii) Expand
iii) Update

Methods

1. Database parsing
2. Template prioritizers
3. Multi-prioritizer guided tree search
    1. Hybrid route search
4. Validation experiments

Processing the BKMS database

Retrieving database as a flat file with 37 235 reactions →

Machine-readable format →

Reaction templates

A + B + … = C + … , where A, B, and C are unstandardized chemical names.

Reactions that led to invalid templates were removed →

7984 valid templates

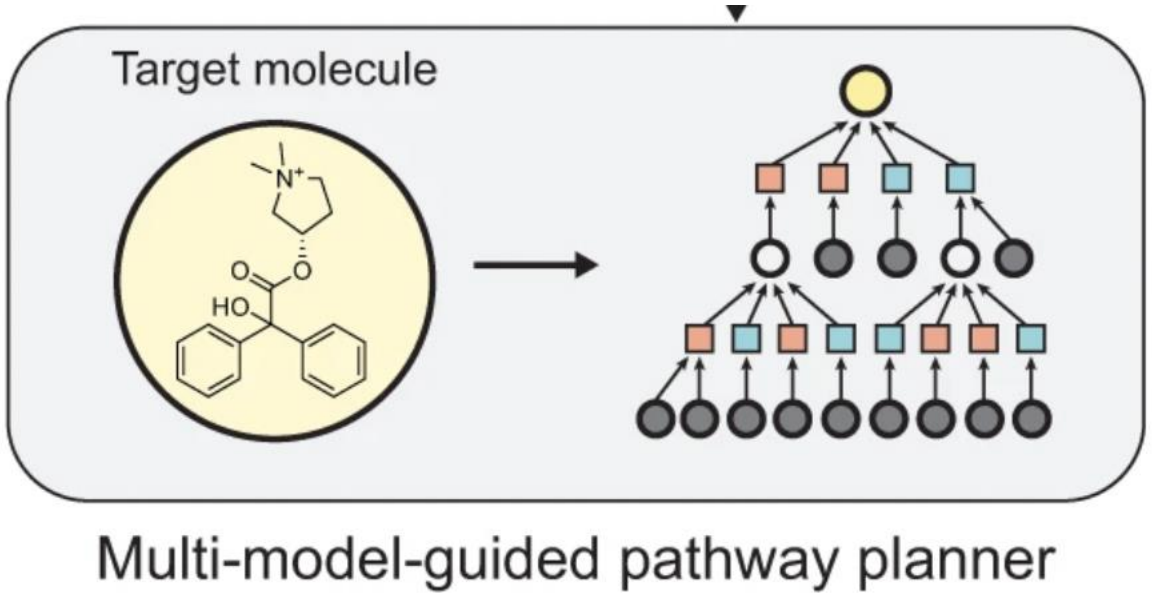
Training the template prioritizer

IN	OUT
2048-bit Morgan fingerprint	A vector of length 7984 (number of templates)

MLP machine learning model: multiple layers of interconnected nodes
→ learn complex patterns.

The multi-prioritizer guided tree search

1. Target molecule (a yellow circle).
2. Reaction templates (squares).
3. Template prioritizers.
4. Tree search.
5. Multi-prioritizer guidance.
6. Output synthesis plan.

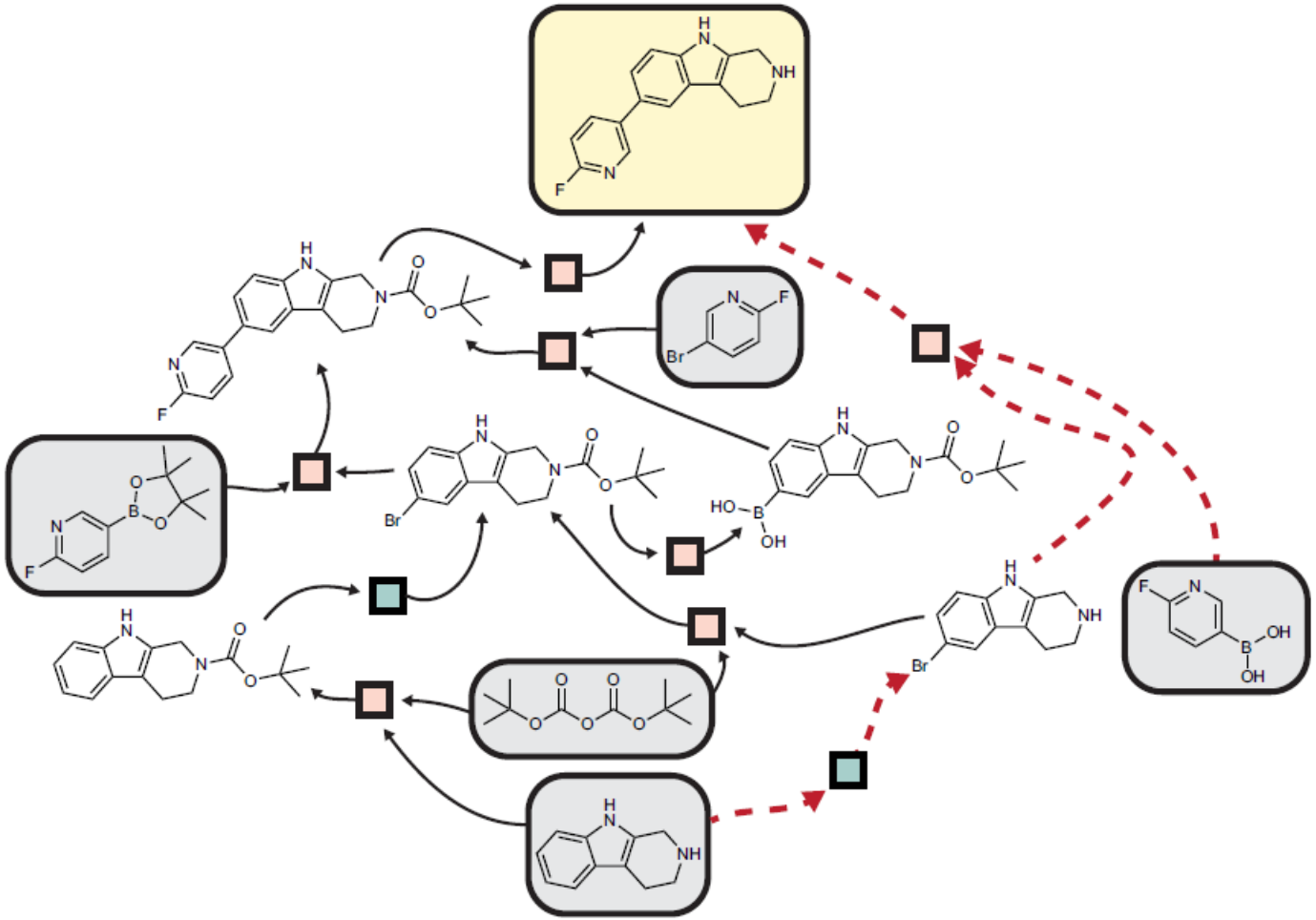


Validation experiments

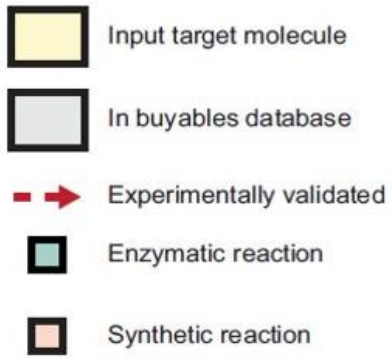
The synthesis plans for fluoropyridinyl tryptoline, dronabinol, and arformoterol were automatically generated, using the same parameters as the hybrid search.

After comparing the results to those obtained using traditional methods was found that the approach was able to identify new routes that were not previously known.

Example: Synthesis pathways for fluoropyridinyl tryptoline generated by the hybrid model



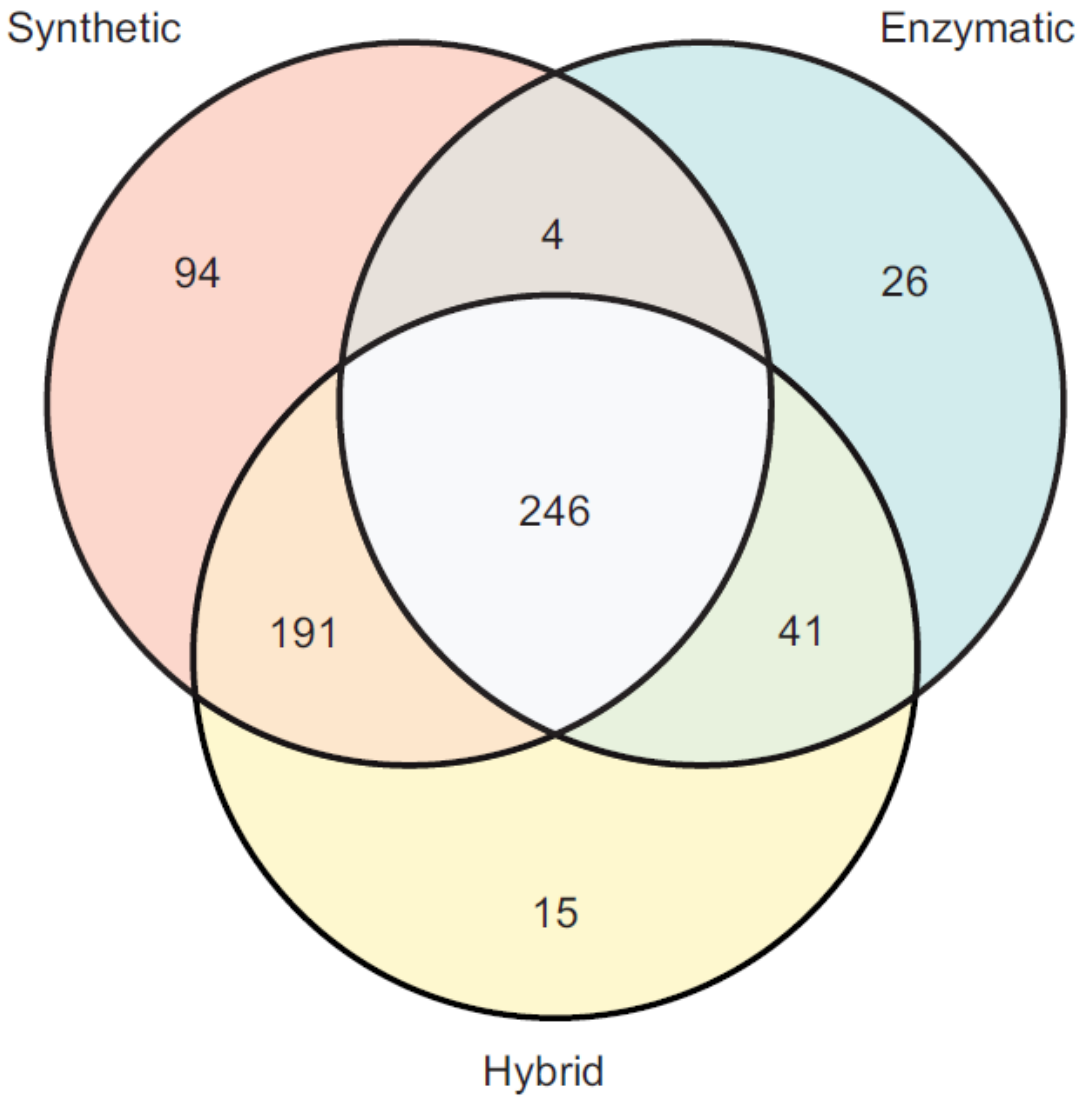
- Uses only buyable starting materials
- Experimentally validated route founded
- Enzymatic reactions not available in the training data found
  - ➢ was able to generalize from the extracted enzymatic templates



Comparison of the models

- The models were given 1000 randomly selected molecules from a biogenic (natural products) molecule catalog
- Models identified synthesis pathways to the molecules, when using a set of buyable compounds as a starting material

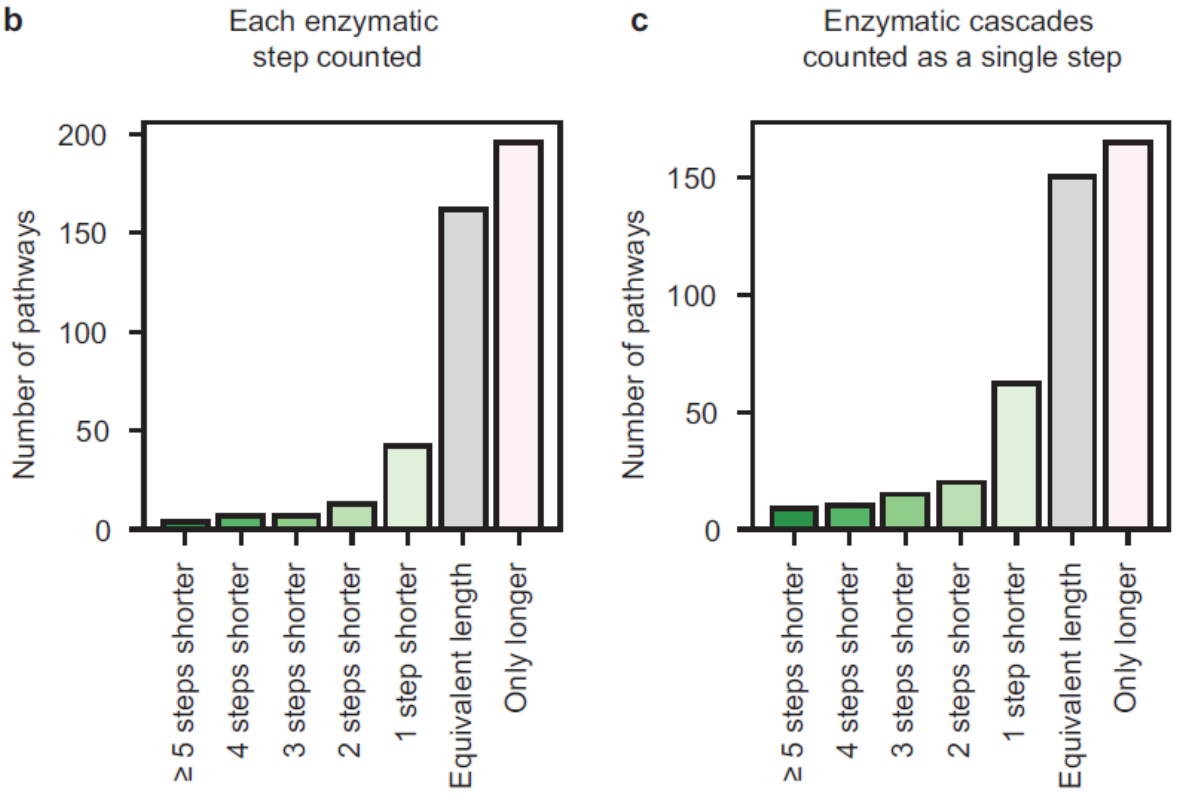
- Smallest number of pathways were found by the enzymatic model
  - Least number of reaction templates
- Synthetic model found the highest number of reactions
  - The search was time-limited, restricting the power of hybrid model
  - ➢ Another comparison method is the number of steps in the shortest synthesis pathways



Comparison of the models

- Hybrid and synthetic pathways are compared
  - Molecules chosen, for which both models found a pathway, and where the pathway given by the hybrid model contains at least one enzymatic reaction (431)

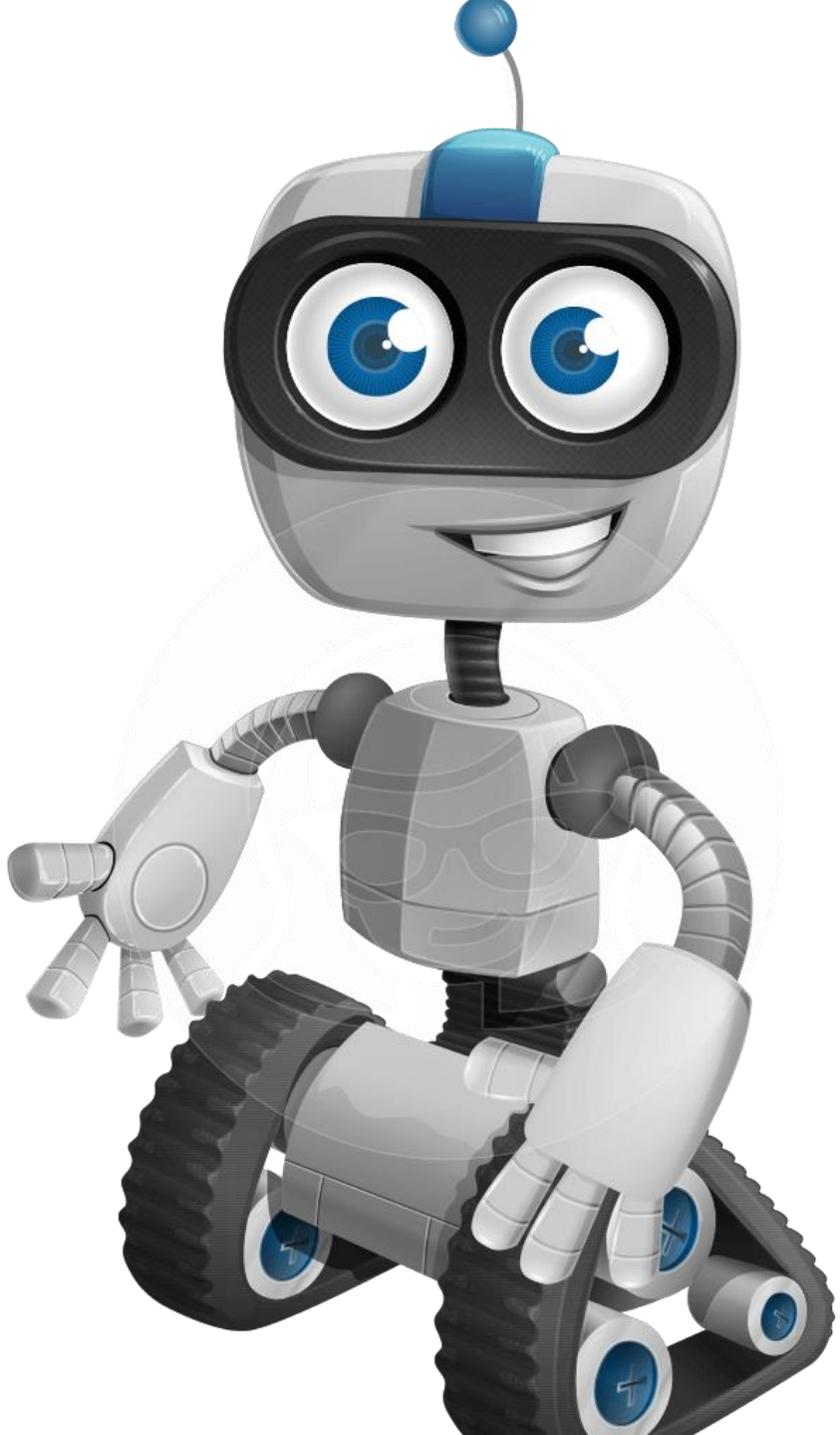
- Number of steps in the shortest pathway (b)
  - Hybrid synthesis gives smaller step count for 17% of the molecules, and step count of equal length for 38% of the molecules
- When counting consecutive enzyme reaction steps as a single step in the hybrid model (c)
  - Subsequent enzymatic reactions can be performed in one pot without purifying the intermediates in between
  - Hybrid synthesis gives smaller step count for 27% of the molecules, and step count of equal length for 35% of the molecules



Fewer reactions means fewer reagents and purification steps —> cheaper and more efficient syntheses!

Path forward

- As described before, a hybrid approach to retrosynthetic planning that generates promising synthesis plans with both enzymatic and synthetic steps to complex molecular targets is demonstrated
  - Molecular intermediates that would not be accessible with synthetic-, or enzymatic chemistry can now be explored
- Generalization of overly specific templates needs to be improved
  - Over-generalization, while improving accuracy, may remove chemical context –> Fewer experimentally implementable suggestions
  - Generalization overall is poor, even in relatively data-rich regimes



Path forward

- Case studies of dronabinol and arfomoterol demonstrated that you can unlock routes novel building blocks or intermediates to compounds of interest.
    - Models suggest enzymatic transformations that would require enzyme engineering -> Applications of enzymes to novel substrates that could expand the biocatalysis toolbox.
- With now proven concept, assessing whether an enzyme could be evolved to perform the desired reaction could be the next step.
    - Human brain likely won't do that calculation alone, so computational modeling combined with expert knowledge, intuition and experimentation is required

Thank you!

Questions?

Levin, I., Liu, M., Voigt, C., & Coley, C. 2022. Merging enzymatic and synthetic chemistry with computational synthesis planning. *Nature Communications.* 13(1). https://doi.org/10.1038/s41467-022-35422-y