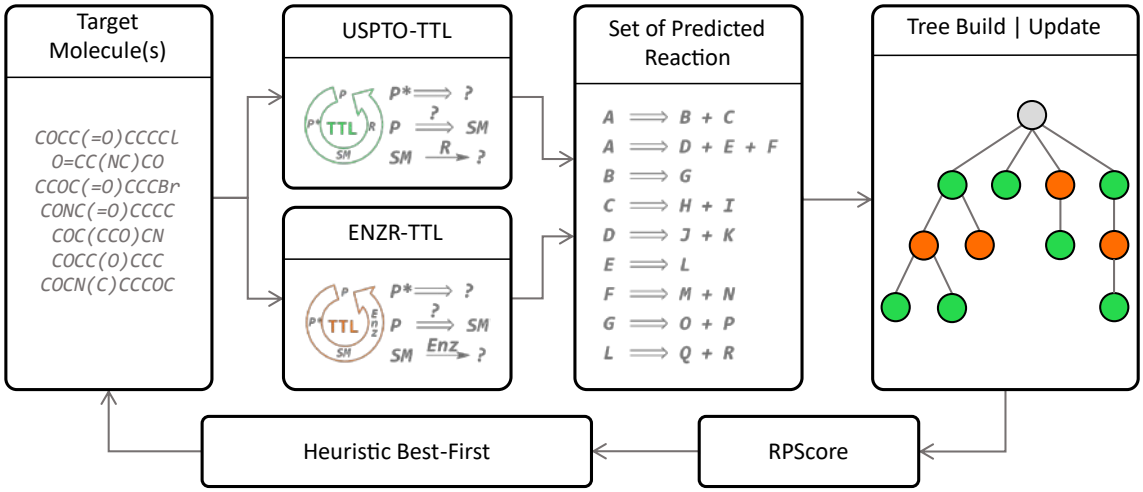
Chemoenzymatic Multistep Retrosynthesis with Transformer Loops

David Kreutter, Jean-Louis Rexmond

Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland
david.kreutter@unibe.ch



Computer-aided synthesis planning (CASP) is pivotal in automating retrosynthetic analyses for unseen molecules, leveraging organic reactivity knowledge from literature. To tackle challenges in proposing realistic disconnections while ensuring reaction novelty and diversity, and in exploring efficient short synthetic sequences, we introduce an innovative open-source CASP tool.[1] Our work extends to enzymatic reactions alongside chemocatalysis, proposing chemoenzymatic routes.

Our approach uses triple transformer loops (TTL) that separately predict starting materials (T1), reagents/enzymes (T2), and products (T3) separately.[2] It explores multiple disconnection sites through a combination of exhaustive, template-based, and transformer-based tagging procedures before T1, allowing extensive chemical space exploration.

We integrate the single-step TTL into a multistep tree search algorithm for biocatalysis (TTLAB) that prioritizes sequences and chemistry based on a route penalty score (RPScore). The RPScore considers factors such as the number of steps, confidence scores, and the simplicity of intermediates along the route. This scoring scheme enables TTLAB to prioritize shorter synthetic routes to readily available commercial starting materials.

Overall, our open-source multistep retrosynthesis tool provides a broader chemical space exploration in synthesis planning and can predict short chemoenzymatic synthetic routes for drug-like molecules.

[1] D. Kreutter, J.-L. Reymond, *Chem. Sci.* **2023**, *14*, 9959–9969.
[2] D. Kreutter, P. Schwaller, J.-L. Reymond, *Chem. Sci.* **2021**, *12*, 8648–8659.