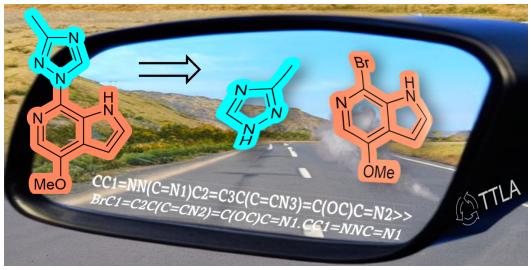
Disconnection-Aware Triple Transformer Loop with a Route-Penalty Score for Multistep Retrosynthesis

D. Kreutter[1], J. L. Reymond[1]*

[1]Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland

Computer-aided synthesis planning (CASP) plays a crucial role in automating retrosynthetic analyses of unseen molecules by learning organic reactivity from literature. To address the challenges of (1) proposing realistic disconnections while maintaining reaction novelty and diversity, and (2) exploring efficient short synthetic sequences, we present an innovative open-source CASP tool.



Our approach uses a triple transformer loop (TTL) that separately predicts starting materials (T1), reagents (T2), and products (T3). It explores multiple disconnections sites through a combination of exhaustive, template-based, and transformer-based tagging procedures prior to T1, allowing an extensive chemical space exploration.

Furthermore, we integrate the single-step TTL into a multistep tree search algorithm (TTLA) that prioritizes sequences 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 TTLA to prioritize shorter synthetic routes to readily available commercial starting materials during the tree search exploration. The effectiveness of our approach is demonstrated by showcasing retrosynthetic analyses of recently approved drugs.

Overall, our open-source multistep retrosynthesis tool provides a broader chemical space exploration in synthesis planning and can predict short synthetic routes for drug-like molecules. Moreover, separating the prediction of starting material and reagents might be adapted to more complex reaction types.[1]

[1] D. Kreutter, J.-L. Reymond, *ChemRxiv*, **2023**, DOI: 10.26434/chemrxiv-2022-8khth-v2