Incorporating In-House Reaction Knowledge into Template-Based Retrosynthesis via Priority Rules

Almaz Gilmullin[1], Guillaume Schmitz[1], Tagir Akhmetshin[1], Olga Klimchuk[1], Dragos Horvath[1], Alexandre Varnek[1,2]

[1] *Laboratory of Chemoinformatics, UMR 7140 CNRS, University of Strasbourg, 4, rue B. Pascal, 67000 Strasbourg, France*

[2] *Institute of Chemical Reaction Design and Development, Hokkaido University, Japan*

Keywords: retrosynthesis; computer-aided synthesis planning; reaction templates

In retrosynthetic analysis, key bond disconnections are identified to break down a target molecule into simpler, more accessible precursors [1]. In template-based approaches, bond disconnections are detected using reaction rules (templates), typically extracted from reaction databases such as USPTO. Such rules encode the reaction center (i.e., ensemble of atoms at the bonds formed, broken, or reorganized during the course of a reaction) and their nearest environment. Quantitative estimation of a rule's relevance to a target molecule is defined by the special neural network model ("policy network"). Although such approach remains attractive due to its interpretability and deterministic applicability, it is limited by the reaction rules coverage. Most existing retrosynthetic tools use rules extracted from the database of reactions extracted from the patents USPTO, which covers a limited number of types of chemical transformations. In this case, application of these tools to out-of-domain transformations is rather problematic.

Here, we propose a simple priority rule injection strategy to incorporate *in-house* reaction knowledge into template-based retrosynthesis without retraining the base model. The new rules can be extracted from publications in two steps: (1) transformation of a chemical reaction from PNG or PDF file into a SMILES string and (2) a rule extraction by the dedicated SynPlanner module. During retrosynthetic procedure, a set of project-specific reaction rules is queried first through substructure matching. When applicable, these priority rules are used to generate candidate precursors before the planner falls back to the policy network. In such a way, this design both preserves the original policy network model and enables rapid deployment of proprietary chemistry knowledge. The approach was implemented in SynPlanner - a multi-step retrosynthetic planning platform developed in the Laboratory of Chemoinformatics [2]. In this work, we demonstrate the priority-rule injection strategy using the Ugi reaction as a representative rare, multicomponent transformation that is poorly captured by standard USPTO-derived template policies but is synthetically valuable for modular route design.

References.

[1] Corey, E. J.; Wipke, W. T. Science 1969, 166 (3902), 178–192. https://doi.org/10.1126/science.166.3902.178.
[2] Akhmetshin, T. et al., J. Chem. Inf. Model. 2025, 65 (1), 15–21. https://doi.org/10.1021/acs.jcim.4c02004.