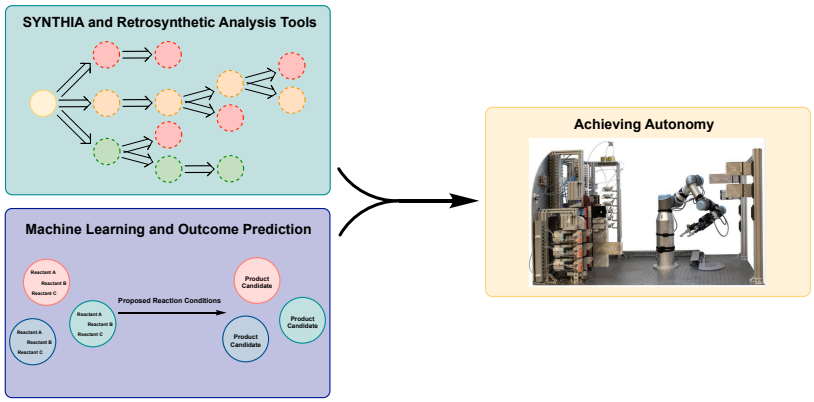
Machine Learning in Computer-Aided Synthetic Planning: The Machine is Here to Stay

Reported by Jacob Weaver March 1, 2022

Introduction

While computer-assisted synthetic planning tools have been under development since the 1960s[1], recent advancements in machine learning have greatly improved their efficacy in retrosynthetic development, reaction prediction, and automation. Significant improvements in computational power, as well as improved interpretation of vast chemical databases like Reaxys and Sci-Finder have greatly improved the ability of modern synthetic planning programs to evaluate and interpret large sets of reaction data. This enables advanced programs like Synthia to perform rigorous retrosynthetic analysis and *De novo* synthesis. Likewise, machine learning techniques have been successfully applied towards reaction prediction in the forward direction, a critical component in the validation of retrosynthetic analysis. Incorporation of these programs shows promising applications in reaction automation, such as in flow chemistry (Scheme 1).[2]



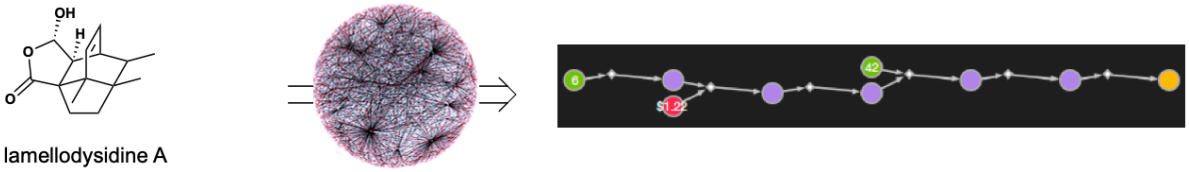
Scheme 1. Merging Retrosynthetic Analysis and Forward Reaction Prediction towards autonomous synthesis

This literature seminar aims to evaluate both historical and recent approaches to incorporate computers into synthetic planning, as well as addressing the various challenges that accompany such an endeavor.

Synthia and Retrosynthetic Analysis Tools

While many retrosynthetic planning tools like Reaxys and Sci-Finder have already become a staple in the day-to-day activities of the organic chemist, there exist many limitations of these programs to efficiently filter and prioritize large amounts of reaction data. Additionally, these programs are incapable of *De novo* synthesis, or the ability to propose **novel** synthetic routes. The use of techniques such as network representations as well as advancements in prioritization of chemical reaction data have enabled programs like Synthia to greatly improve analysis and generation of intelligent retrosynthetic routes.[3] Through the use of manually

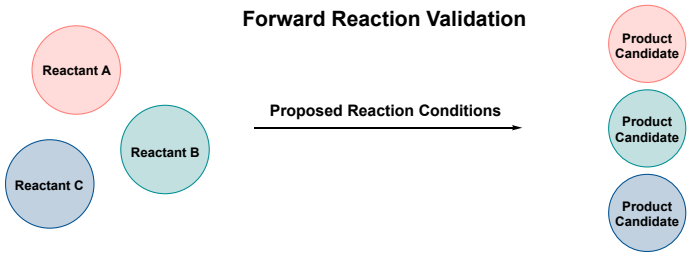
encoded reaction rules and templates, programs like Synthia are also now capable of *De novo* synthesis. The novel routes generated by Synthia have been successfully validated in the synthesis of both pharmaceuticals and natural products and are seemingly indistinguishable from human-generated routes (Scheme 2).[4]



Scheme 2. Successful *De novo* synthesis of lamellodysidine A generated by Synthia

Advancements in Forward Reaction Prediction

While retrosynthetic analysis is a crucial component in the design of synthetic routes, validation of these proposed routes through forward reaction prediction is equally necessary. When given a set of reaction conditions, machine learning techniques have been applied to successfully predict and rank potential reaction products (Scheme 3).[5] A common approach to forward reaction prediction involves the use of reaction templates which aim to generalize reactions to specific atomistic changes on the molecule which can be mapped onto the substrate under study. Through analysis of readily calculable molecular descriptors, machine learning has enabled modern programs to rank product candidates with high levels of confidence. While this approach has shown reasonable success in basic reaction prediction, there exist many limitations due to the inability to evaluate subtle molecular characteristics like conformation, electronics, and sterics. The incorporation of improved and easily calculable molecular descriptors which encapsulate these characteristics remains a major area of study in this field.



Scheme 3. Generation of Product Candidates in Forward Reaction Prediction

References

1. Corey, E.J. *et al. Science*, **1969**, *166*, 178-192
2. Coley, C. *et al. Science.* **2019**, *365*
3. Grzybowski, B.A. *et. al. Angew. Chem. Int. Ed.* **2016**, *55*, 5904-5937
4. Grzybowski, B.A. *et. al. Nature.* **2020**, *588*, 83-88
5. Jensen, K.F. *et al. ACS Cent. Sci.* **2017**, *3*, 434-443