

Computer Assistance in Organic Synthesis Planning and Execution

Connor W. Coley

May 15, 2019

The identification and synthesis of molecules that exhibit a desired function is an essential part of addressing contemporary problems in science and technology. Small molecules are the predominant solution to challenges in the development of medicines, chemical probes, specialty polymers, and organocatalysts, among others. The typical discovery paradigm is an iterative process of designing candidate compounds, synthesizing those compounds, and testing their performance. The rate at which this process yields successful compounds can be limited by bottlenecks and mispredictions at all three stages and is plagued by inefficiencies, not the least of which is the predominantly-manual nature of synthesis planning and execution.

This thesis describes techniques to streamline the synthesis of small molecules in the context of pharmaceutical discovery from two perspectives: one experimental and the other using techniques in data science and machine learning.

The first part of this thesis focuses on the time-, material-, and experimental-efficiency of data collection. It describes the development of an automated microfluidic reactor platform for studying physical and chemical processes at the micromole scale. Synthesis and purification of small molecule compound libraries are performed without human intervention at a scale suitable for a medicinal chemistry setting. Integration of online analytics enables efficient, closed-loop self-optimization using an optimal design of experiments algorithm to identify reaction conditions suitable for production-scale flow synthesis.

To complement the generation of new data through automated experimentation, the second part of this thesis is motivated by the goal of applying existing reaction data to problems in synthesis and synthesis design. This includes the development of data-driven methodologies for the design and validation of small molecule synthetic routes. An enabling factor in ensuring the feasibility of computationally-proposed reactions is the use of models to predict organic reaction outcomes *in silico* that leverage the flexibility in pattern recognition afforded by neural networks to understand chemical reactivity in the same way we might by reading the literature. Several predictive models are integrated into an overall framework for computer-aided synthesis planning that can rapidly propose routes to new molecules with the complexity of modern active pharmaceutical ingredients.

As a final demonstration, machine learning assisted synthesis planning is brought together with laboratory automation to illustrate an accelerated approach to target-oriented flow synthesis. This is a proof-of-concept for how chemical development might one day occur with less human intervention.

Thesis Supervisor: Klavs F. Jensen
Warren K. Lewis Professor of Chemical Engineering
Professor of Materials Science and Engineering

Thesis Supervisor: William H. Green
Hoyt C. Hottel Professor in Chemical Engineering