Data-Driven Chemistry, Small Molecule Synthesis, And Design

The typical molecular discovery paradigm is an iterative process of designing candidate compounds, synthesizing those compounds, and testing their performance, where each repeat of this cycle can require weeks or months, requires extensive manual effort, and relies on expert intuition. This talk will describe our efforts to accelerate the process of molecular discovery from two interrelated angles. The first involves facilitating the chemical synthesis of new structures through data-driven computer-aided synthesis planning, where we can learn patterns of chemical reactivity directly from tabulated organic reaction data. The second involves prioritizing candidate molecules to test using machine learning models for molecular property prediction. Together, these studies illustrate how data science and statistical learning offer unprecedented opportunities to systematize and streamline the process by which new functional small molecules are designed and synthesized.

Bio: Connor W. Coley is an Assistant Professor at MIT in the Department of Chemical Engineering and the Department of Electrical Engineering and Computer Science. His work in computer assistance and automation for organic synthesis has included the development of a data-driven synthesis planning program and *in silico* strategies for predicting the outcomes of organic reactions. His continuing research interests are in how data science, statistical learning, and laboratory automation can be used to streamline discovery in the chemical sciences. Connor has been named one of C&EN's "Talented Twelve" and one of Forbes Magazine's "30 Under 30" for Healthcare. He received his B.S. and Ph.D. in Chemical Engineering from Caltech and MIT, respectively, and did his postdoctoral training at the Broad Institute.