Accelerating Molecular Design with Machine Learning and AI

Progress in autonomous and high-throughput experimentation and the ongoing maturation of machine learning (ML) and generative AI has led to a drastic increase in the pace of chemical and materials research. In this talk, we will overview two directions that emphasize how machine learning and experiments/simulations can be combined.

First, we will present a generative AI method for de-novo design of molecular compounds with desired biological properties. The general workflow of the proposed method integrates two deep neural networks – generative and predictive – that are initially trained separately but then trained jointly to generate novel chemical structures with desired properties. In the proof-of-concept study, we have employed this integrative strategy to design chemical libraries biased toward compounds with specific ranges of physical properties (like solubility and hydrophobicity) as well as to develop novel kinase inhibitors of JAK2, EGFR, and CDK1. This new approach can find general use for generating targeted chemical libraries optimized for a single desired property or polypharmacology.

Second, we proposed a novel ML-guided materials discovery platform that combines synergistic innovations in automated flow synthesis and automated machine learning (AutoML) method development. A software-controlled, continuous polymer synthesis platform enables rapid iterative experimental–computational cycles that result in the synthesis of hundreds of unique copolymer compositions within a multi-variable compositional space. The non-intuitive design criteria identified by ML, accomplished by exploring less than 1% of overall compositional space, upended conventional wisdom in the materials design. We are currently enhancing this platform with Reinforcement learning (RL) agents. By training via value-based or policy-based iterative schemes, RL agents can learn to suggest synthesis protocols, potential reactants, and experimental conditions.