Optimization Approaches to Algorithmic Decision Making in Molecular Discovery

by

Jenna C. Fromer

Submitted to the Department of Chemical Engineering
on April 15, 2026 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

ABSTRACT

Drug discovery is an optimization problem that seeks to identify molecules that satisfy multiple design criteria. The discovery of small molecule drugs often involves computationally screening a molecular library to identify molecules predicted to bind to a target protein, followed by experimental validation and additional structural optimization of top-performing molecules. Each of these stages involves time-consuming physical and computational experiments. This dissertation involves the development and application of optimization tools to support hit identification, design-make-test cycles, and chemical space exploration.

This thesis begins with the development of Bayesian optimization methods to accelerate virtual screening. A novel acquisition function, multipoint Probability of Optimality, is introduced to improve sample efficiency in model-guided virtual screening. Further, while model-guided active learning has proven effective for optimizing individual predicted properties, single-objective optimization defers critical trade-offs to later stages of design. This thesis applies Pareto optimization to retrospective multi-objective virtual screening tasks.

Next, this thesis introduces an optimization framework, termed SPARROW, for simultaneous compound and synthetic route selection in design-make-test cycles. The selection of which compounds to synthesize and test from all possibilities is a complex decision making process that must consider both experimental resource constraints and the perceived value of candidates. SPARROW formalizes this decision making process by jointly optimizing compound utility and synthetic cost, accommodating flexible definitions of both objectives.

Compound downselection in design-make-test cycles is inherently limited to the chemical space known to be synthetically accessible; however, compounds with enhanced property profiles may exist in yet-unexplored chemical space. This thesis proposes a cheminformatics pipeline to aid in the systematic prioritization of unprecedented scaffolds for synthesis by assessing novelty, stability, and synthetic tractability. This approach aims to guide the exploration of new chemical space in a targeted and experimentally actionable manner.

Finally, this thesis concludes by describing two efforts to apply algorithmic design and downselection methods to prospective molecular discovery projects.

Altogether, this thesis demonstrates how optimization methods can support virtual screening, compound selection, and scaffold exploration in small molecule drug discovery.

Thesis supervisor: Connor W. Coley
Title: Associate Professor in Chemical Engineering