Computational design of molecular templates for porous material synthesis

by

Mingrou Xie

Submitted to the Department of Chemical Engineering
on Dec 7, 2026 in partial fulfillment of the requirements for the degree of

DOCTORATE OF PHILOSOPHY IN CHEMICAL ENGINEERING

ABSTRACT

Computation-driven materials discovery has accelerated over the past decade. However, candidate structures are useful only if they can be made. Designing the synthesis recipe for a given material bottlenecks current materials discovery efforts. In particular, zeolites are a class of industrial materials that require specific organic molecule templates (OSDAs) to synthesize. These templates are extremely challenging to design and predict synthesis outcomes for. This thesis builds experiment-informed computational tools and workflows for high throughput screening of these templates for zeolite synthesis.

An end-to-end workflow ZeoBind was developed to screen OSDAs. A library of close to 2 million template-like, synthetically accessible molecules was curated, prediction tasks and surrogate models were developed, and domain-informed heuristics and phase competition knowledge was incorporated in the workflow to intelligently select candidates. Two campaigns that screened zeolite-molecule space orders of magnitude larger than previous efforts produced experimentally validated OSDAs that synthesized zeolites with novel compositions.

Next, the thesis extends the screening to hypothetical zeolite space, which is 10^4 times larger than the known zeolite space. To do so, an in-depth investigation of model architectures and behaviour under sparse data conditions in unseen zeolite-molecule spaces was carried out, and experimental gel compositions were used for efficient consideration of phase competition. Hypothetical zeolites of potential industrial interest and OSDA candidates for them were then identified in targeted screening campaigns.

This thesis also delves into the improvement of existing stochastic docking methods for challenging cases with complex pore networks and flexible molecules. A benchmark is designed to test different docking procedures, and revised protocols are used to improve data quality and resolve misalignment between simulation and experiment in two case studies.

The last part of this thesis presents a first-of-its-kind flow matching model, MatDock, for generating 3D systems of periodic materials with multiple molecules in one pass. MatDock not only serves as a solution for bypassing the costly computational validation step in ZeoBind, but also as a framework for other material-molecule generative tasks.

Thesis supervisor: Rafael Gómez-Bombarelli
Title: Professor of Materials Science

Thesis supervisor: Connor Coley
Title: Professor of Chemical Engineering