

Automating reaction development: hardware and software for fully-automated high-fidelity navigation of high-dimensional chemical reaction space

Natalie Eyke

Thesis advisors: William H. Green and Klavs F. Jensen

Process development is one of the major hurdles that pharmaceutical companies face when bringing a new pharmaceutical compound to the market. Acceleration of process development minimizes costs and gets new compounds to the market sooner, which 1) gets medication to patients in need quickly and 2) maximizes revenue relative to a fixed patent expiration date. Incorporating automation into the process development workflow can accelerate the acquisition of high-quality data, which helps process development scientists and engineers develop safe and efficient processes rapidly. Automation may come in the form of software, such as experimental design algorithms, and hardware, such as platforms that prepare, run, and analyze reactions.

The field of statistics offers a number of approaches to optimal experimental design that have formed the basis for so-called "quality-by-design" development strategies in the pharmaceutical industry. More recently, iterative, algorithmic optimization routines have been adapted for reaction optimization as well. However, these strategies are most appropriate for reaction domains that consist primarily of continuous reaction variables. New techniques are needed that enable optimization over discrete and continuous reaction variables - in other words, high-dimensional chemical space - simultaneously. Discrete reaction variables include factors like the catalyst, ligand, solvent, and other reagents that are used to effect a chemical transformation. Given a tensor capturing relevant information about each discrete reaction variable, machine learning is well-suited to use data to identify patterns and relationships between the various settings of each discrete variable to accurately model reaction outcomes and enable optimization. We used machine learning in combination with an experimental design routine from the field of statistics, known as uncertainty sampling, to demonstrate that machine learning can be used to minimize the number of experiments that need to be performed to obtain an accurate model of a high-dimensional reaction domain defined by multiple discrete reaction variables.

On the hardware side, automated experimentation platforms have gained a foothold in process development organizations, but there remains a need for platforms that can reproduce the flexibility and accuracy of the bench chemist while achieving high throughput and using as little material as possible. Droplet microfluidics is attractive for reaction development because it uses small quantities of precious reaction material, and the high surface area to volume ratio enables efficient heat transfer and interphase mass transfer. We adapted an automated droplet reactor platform for high-fidelity, flexible, reproducible, high-throughput operation by upgrading the design and operating procedures, placing multiple reactor channels in parallel, and creating a scheduling algorithm that orchestrates all of the parallel hardware operations and ensures droplet integrity as well as overall efficiency. We designed and incorporated all of the necessary hardware and software to enable both thermal and photochemical reactions. We demonstrated both the single-channel and parallelized versions of the platform using a series of

model thermal and photochemical reactions, and demonstrated how the parallelized platform allows for rapid acquisition of the data necessary to determine reaction kinetics. The platform is flexible in terms of use case: through the integration of a variety of experimental design algorithms, it can be used for either screening or optimization over a wide range of chemical domains, and a fraction collector can be appended to the end of the platform to capture reacted droplets and thereby enable library synthesis.

The software and hardware developed in this thesis can together enable accelerated process development that minimizes delays between the discovery of transformative medicines and their delivery to patients in need.