Autonomous experimentation for molecular discovery applications

by

Richard Benjamin Canty

January 8th, 3:00pm in 66-360

Autonomous automated platforms provide a means to accelerate scientific research. Automation has seen marked advances in the discovery and optimization of chemical compounds through machine-learning–driven design and through accelerated serendipity; many automated experimental platforms, however, are often highly specialized and lack the ability to pivot to new research areas. These factors have driven a shift in design to create generalized systems that incorporate decision logic to have more autonomy. The incorporation of autonomy beyond experimental selection, however, has been challenging. The agency to adapt process and workflow details in response to observations requires data management and control architectures which can, at times, be at odds with pure automation.

In this thesis, operational and workflow autonomy were integrated into an automated platform for the discovery of small organic dye-like molecules. This integration required adapting control architectures to handle goal-oriented commands—for which automated agents use expert-encoded rules to select a particular implementation of their task. Moreover, this integration required providing platform agents with up-to-date information about the platform and samples and the ability to make modifications to workflows during execution. Finally, concurrent execution of multiple, dynamic workflows on the platform required additional scheduling strategies previously not needed for non-autonomous automated research platforms. The incorporation of more platform autonomy presents exciting challenges for data management, systems integration, and platform orchestration.

For control, a central controller orchestrated platform tasks between multiple computers with varied laboratory instruments. In tandem, a separate database provided platform instruments with current information on samples, workflows, and platform resources. By separating the orchestration of tasks from the contextual details required to determine operation specifics, workflows could be described as a sequence of terse, high-level goals that each agent could then translate into concrete actions at run-time. In addition to the database, each agent synchronized their state information with the orchestrator and with collaborators to enable agent–agent control when necessary. This awareness of platform state and available functionality further enabled the stable monitoring of problems on the platform.

Workflows were designed with modular, high-level goals and were self-contained to allow platform agents to edit workflows. The use of goals facilitate workflow mutations as each agent does not need to know implementation-level details of other agents' operations. These goals also provided a sensible fundamental unit for scheduling. Moreover, by keeping workflows focused on a single experiment (a synthesis step or characterization assay) and defining prerequisites for execution, the platform could avoid race conditions and other logical hazards as it was impossible for two workflows to operate on the same well plate or sample (and thus database entries) at the same time.

To increase the autonomy of the platform further, agents require choices and robust logic to make good decisions. To provide expressive power and to implement solutions to goals, methods and contextual logic were defined near the hardware level—where available functionality is best understood. This approach lent itself to an inverted design pattern whereby each goal an agent

could accomplish is injected up to their parent controller (in contrast to defining a standardized set of operations at the top level of control). While this design choice facilitated giving each agent more choices for accomplishing its goals, it limited the transferability of the modules to other automated experimental platforms.

Context-aware agents and mutable workflows elevate the autonomy of the platform but challenge orchestration. The modularization and decomposition of workflow operations, the time-sensitive integrity of samples, and safety considerations required live scheduling with temporal constraints between operations. A greedy algorithm was developed to handle the simultaneous execution of multiple, mutable workflows. In addition, the movement of platform resources was simulated to inform scheduling decisions and avoid traffic jams. Scheduling provides a way to accommodate the adaptability and agency of the platform's systems, ensure sample integrity, and prevent resource and operational conflicts.

In efforts to further improve the autonomy of the platform, additional experimental controls were implemented to assist the platform when operating with unknowns. For example, when measuring UV–visible spectra of samples with unknown concentrations and extinction coefficients, the plate reader and liquid handler were made to collaborate to adjust sample loadings rather than discard experiments with bad signals. In addition, to address instances where reaction yield was detected but insufficient for subsequent operations, a reaction condition recommendation algorithm was developed to propose alternative conditions to increase (not optimize) yield under constraints for the plateability of samples with respect to available platform resources. This data-driven approach imitates a chemist by considering conditions from related reactions then evaluating trust in those conditions by the number and quality of successful reactions similar to those related reactions. On a per well basis, the approach performed comparably to using higher temperatures to drive yield; however, scaled better with additional conditions as the algorithm can provide a series of conditions to attempt in a single well plate whereas a temperature series requires a well plate for each temperature tested.

Thesis supervisor: Klavs F. Jensen
Title: Warren K. Lewis Professor of Chemical Engineering
Professor of Materials Science and Engineering