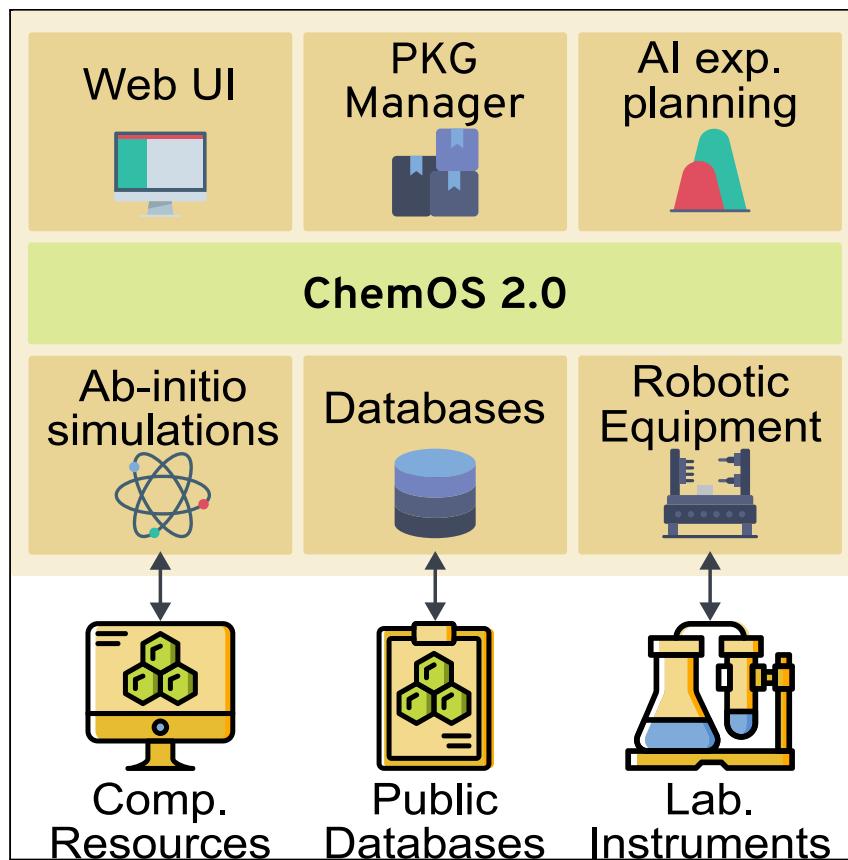


## Article

# ChemOS 2.0: An orchestration architecture for chemical self-driving laboratories



Malcolm Sim, Mohammad Ghazi Vakili, Felix Strieth-Kalthoff, ..., Santiago Miret, Sergio Pablo-García, Alán Aspuru-Guzik

spgarcia@gmail.com (S.P.-G.)  
alan@aspuru.com (A.A.-G.)

### Highlights

A modular strategy for building a self-driving lab for chemical research

Demonstrative workflows based on real-world research in materials discovery

High- and low-level implementation of laboratory hardware/software

Automated experiment planning and execution as well as automated data collection

ChemOS 2.0 is a comprehensive laboratory architecture for transforming the modern chemistry lab into one that accelerates the pace of chemical research. This new kind of laboratory, known as a self-driving lab (SDL), uses automated experimental tools, as well as computational experiment planners, to create fully automated workflows that require minimal human intervention. ChemOS 2.0 presents a modular and versatile approach to building one's own SDL, including real-life implementations of this framework.

### Demonstrate

Proof-of-concept of performance with intended application/response

4

Sim et al., Matter 7, 2959–2977  
September 4, 2024 © 2024 Elsevier Inc. All rights reserved.  
<https://doi.org/10.1016/j.matt.2024.04.022>



## Article

# ChemOS 2.0: An orchestration architecture for chemical self-driving laboratories

Malcolm Sim,<sup>1,2</sup> Mohammad Ghazi Vakili,<sup>1,2</sup> Felix Strieth-Kalthoff,<sup>1,2</sup> Han Hao,<sup>1,2</sup> Riley J. Hickman,<sup>1,2,3</sup> Santiago Miret,<sup>4</sup> Sergio Pablo-García,<sup>1,2,3,9,\*</sup> and Alán Aspuru-Guzik<sup>1,2,3,5,6,7,8,\*</sup>

## SUMMARY

**Self-driving laboratories (SDLs), which combine automated experimental hardware with computational experiment planning, have emerged as powerful tools for accelerating materials discovery. The intrinsic complexity created by their multitude of components requires an effective orchestration platform to ensure the correct operation of diverse experimental setups. Existing orchestration frameworks, however, are either tailored to specific setups or have not been implemented for real-world synthesis.** To address these issues, we introduce ChemOS 2.0, an orchestration architecture that efficiently coordinates communication, data exchange, and instruction management among modular laboratory components. By treating the laboratory as an “operating system,” ChemOS 2.0 combines *ab initio* calculations, experimental orchestration, and statistical algorithms to guide closed-loop operations. To demonstrate its capabilities, we showcase ChemOS 2.0 in a case study focused on discovering organic laser molecules. The results confirm ChemOS 2.0’s prowess in accelerating materials research and demonstrate its potential as a valuable design for future SDL platforms.

## INTRODUCTION

Global humanitarian and ecological challenges have sparked an unprecedented demand for novel functional materials across diverse industries, including clean energy technologies (renewable energy conversion and energy storage)<sup>1</sup> as well as medicine and health care.<sup>2</sup> Motivated by the urgency of these crises, researchers have realized the need to accelerate the often laborious and empirical discovery process of designing, fabricating, and testing new materials.

Given this need, recent research efforts have shown remarkable progress in automated experimentation for various steps of the materials discovery cycle, including synthesis, formulation and device fabrication, functional characterization, and computational simulations. Notable advances in this regard range from early examples of automated biomolecule synthesis,<sup>3</sup> flow chemistry and microfluidics for solution-phase synthesis,<sup>4,5</sup> high-throughput experimentation for biological assays or reaction screening,<sup>6,7</sup> to automated systems for fabricating solid-state or thin-film materials,<sup>8,9</sup> as well as automated computational tools<sup>10–12</sup> and the application of big data<sup>13</sup> for virtual screening. The merger of such automation platforms with advances from artificial intelligence (AI) has given rise to the concept of self-driving laboratories (SDLs): the closed-loop integration of data-driven experiment planning with automated experiment execution. Such autonomous experimentation systems

## PROGRESS AND POTENTIAL

Materials discovery and research in the broader field of chemistry is a long and arduous process that is currently dominated by traditional trial-and-error methodologies. One solution for accelerating chemical research is the pairing of automated experimental workflows with automated experiment planning. This combination forms an uninterrupted loop and is known as a self-driving lab (SDL). Many SDLs have been built across the world and have shown tremendous promise in helping chemists with their research. However, acceptance of SDLs by the scientific community remains hindered by a lack of resources, a lack of expertise, and significantly for this work, a lack of a generalized framework for building them (importantly, a framework with concrete, real-world examples). We present a general strategy for creating SDLs that are tailored to most applications in chemistry. We demonstrate the capabilities of this framework via examples that combine computational and experimental tools.



have been successfully demonstrated to address diverse optimization problems in a sample-efficient (and thereby time-efficient) manner, significantly reducing the required experimental resources.<sup>14</sup> While early examples have mainly focused on reaction condition optimization, recent SDLs have shown the potential to discover new materials compositions, e.g., for quantum dot synthesis,<sup>15</sup> thin-film devices,<sup>8</sup> nanoparticles,<sup>16</sup> or solid-state materials.<sup>17</sup>

Whereas automated systems have enabled increased throughput and enhanced reproducibility compared with human experimentation, their flexibility and reconfigurability has remained limited. While this can be partly attributed to the often human-centric hardware design, software integration and the dynamic orchestration of automated workflows have remained a major challenge in automated laboratories. To address this challenge, a range of software solutions such as ChemOS,<sup>18</sup> AresOS,<sup>19</sup> and NIMS-OS,<sup>20</sup> among others,<sup>21–27</sup> have been proposed for workflow management and hardware integration, with more ambitious approaches implementing complete frameworks combining hardware and software,<sup>28–30</sup> and standardizing experimental protocols.<sup>31</sup> However, existing frameworks overlook how researchers automate hardware, and do not provide flexible human-robotic integrations or a proper universal framework. Moreover, current frameworks do not incorporate Internet of Things (IoT) breakthroughs like fog computing, which offer strong data management and storage without relying on remote servers.<sup>32,33</sup> Failing to embrace these cutting-edge technologies is a missed opportunity for optimizing SDL performance and making them more usable for experimental scientists. Additionally, the disconnect between experimental and computational simulations limits many laboratories' research potential, as it fails to integrate valuable theoretical insights from computational studies. An ideal orchestration framework should tackle all of the aforementioned challenges to promote widespread adoption and continuous improvements of diverse kinds of SDLs.

Remarkably, the current state of the automated laboratory landscape shows notable parallels to the early era of computer development in the 1950s and 1960s, where a multitude of specific devices, processing units, and workflows were developed with little flexibility and interoperability. Historically, this has inspired the development of standardized, portable operating systems (OSs), which form the backbone of modern computer technology, and have enabled the widespread adoption of computer systems. The UNIX OSs,<sup>34</sup> arguably the most prominent class of operating systems, are based on three simple design concepts: (1) the development of small functional units for specific tasks, (2) the robust interoperation of these units, and (3) their intercommunication through plain text streams.<sup>35</sup>

Based on this UNIX philosophy, we propose ChemOS 2.0 (Figure 1), a versatile, adaptable and portable “operating system” for automated laboratories, fostering a holistic approach to operate and orchestrate various experimental and computational units. Analogous to computational OSs, standardized plain-text communication protocols not only allow for the flexible integration of new software and hardware units but also enable seamless interaction with human researchers. At the heart of our architecture is the integration of a fog computing device, offering improved efficiency and faster decision-making within the laboratory. Much like a traditional OS kernel, the fog computing device manages data, coordinates components, and empowers the lab to operate autonomously. We first describe the general architecture of ChemOS 2.0 along with its main design elements. Second, we demonstrate the applicability of ChemOS 2.0 in our own laboratory, showcasing

<sup>1</sup>Department of Chemistry, University of Toronto, Lash Miller Chemical Laboratories, 80 St. George Street, Toronto, ON M5S 3H6, Canada

<sup>2</sup>Department of Computer Science, University of Toronto, Sandford Fleming Building, 40 St. George Street, Toronto, ON M5S 2E4, Canada

<sup>3</sup>Vector Institute for Artificial Intelligence, 661 University Avenue Suite 710, Toronto, ON M5G 1M1, Canada

<sup>4</sup>Intel Labs, 2200 Mission College Boulevard, Santa Clara, CA 95054, USA

<sup>5</sup>Department of Materials Science & Engineering, University of Toronto, 184 College St., Toronto, ON M5S 3E4, Canada

<sup>6</sup>Department of Chemical Engineering & Applied Chemistry, University of Toronto, 200 College St., Toronto, ON M5S 3E5, Canada

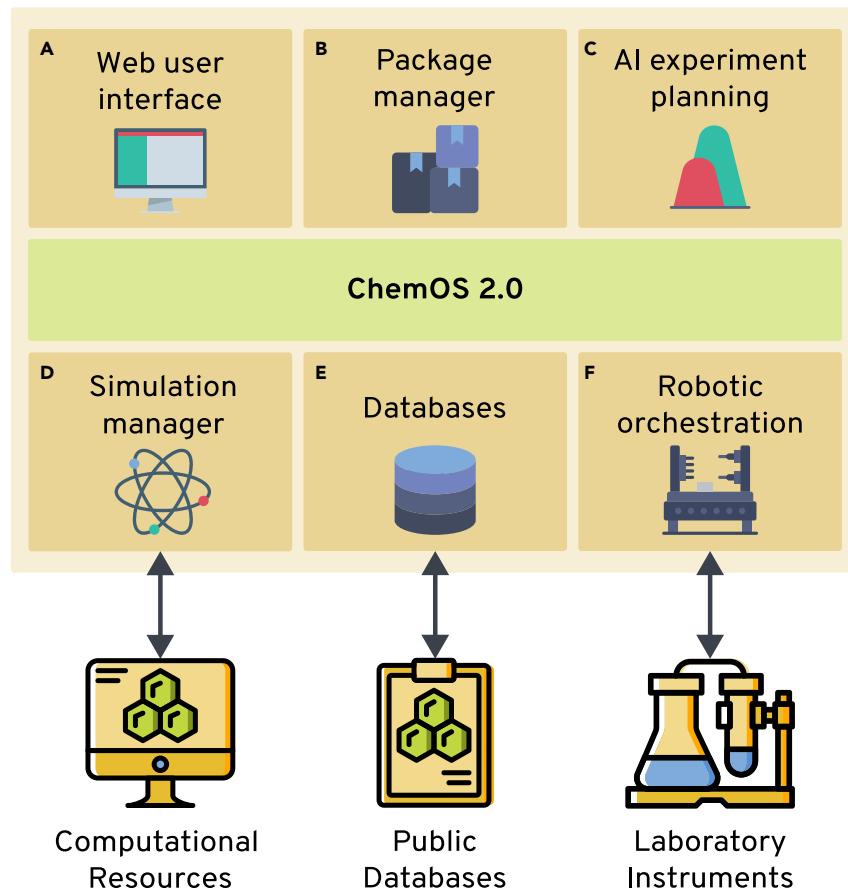
<sup>7</sup>Canadian Institute for Advanced Research (CIFAR), 661 University Avenue, Toronto, ON M5G 1M1, Canada

<sup>8</sup>Acceleration Consortium, 80 St. George Street, Toronto, ON M5S 3H6, Canada

<sup>9</sup>Lead contact

\*Correspondence:  
[srgarcia@gmail.com](mailto:srgarcia@gmail.com) (S.P.-G.),  
[alan@aspuru.com](mailto:alan@aspuru.com) (A.A.-G.)

<https://doi.org/10.1016/j.matt.2024.04.022>



**Figure 1. Features and capabilities of ChemOS 2.0**

- (A) Web graphical interface to ease the user interaction.
  - (B) Package to ensure full software reproducibility.
  - (C) Bayesian optimizer platform for experimental planning.
  - (D) DFT workflow manager connected to our high-performance computer cluster to orchestrate *ab initio* DFT experiment.
  - (E) SQL database server hosting both experimental and simulation databases.
  - (F) Communication protocol to control laboratory instruments.
- Modules (A) and (D)–(F) enable the communication layer with external agents.

the full orchestration of a complex materials discovery workflow toward novel gain materials for organic solid-state lasing devices.

## RESULTS

### Software management

Meticulous control over the laboratory's software ecosystem is imperative to enhance experimental reproducibility, increase transparency, and mitigate production failures. As such, achieving complete transparency necessitates stringent management of the software state within a laboratory with dependency conflicts due to incompatible software versions posing a well-recognized challenge in this regard.<sup>36</sup> To address these concerns, the core design of ChemOS 2.0 incorporates an orchestration fog device that runs the necessary software layers for laboratory operations while keeping each laboratory modular for streamlined integration. To ensure reproducibility, robustness, and seamless deployment, the fog orchestration platform is equipped with NixOS, a declarative package-manager-based (Nix) operating system known for its ability to provide precise control over system state and software

versions.<sup>37</sup> Nix provides precise control over the system state by enabling an accurate selection of software versions and dependencies. Defining local and global system states in NixOS revolves around the use of .nix files, containing a declarative system state recipe. Researchers can use configuration files to share digital experimental environments and deploy pre-configured OSs on different platforms. Thus, .nix files are pluggable modules whose addition or removal will change the state of the system. While some of the code may be hardware-specific, these files serve as a foundation for building unique SDL architectures. In our work, we have prepared a collection of .nix files containing the environmental state used for our experiments. The configuration code and installation instructions can be found in the ChemOS 2.0 repository.<sup>38</sup>

### Data management

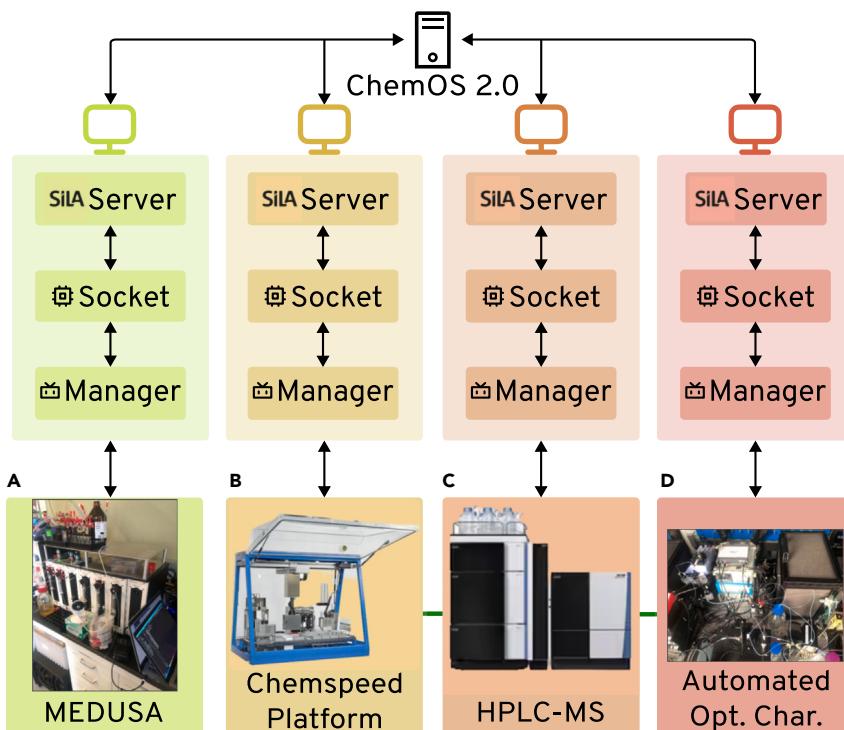
Ameliorating experimental reproducibility requires a proper data management plan involving the collection of experimental procedures, outcomes, and environment metadata from an SDL. Notably, we draw inspiration from the remarkable advancements in computational chemistry, where sophisticated data solutions have successfully ensured reproducibility in chemical simulations.<sup>13,39–43</sup>

To streamline data management, our orchestration device is equipped with two independent relational databases: one for experimental and another for simulation data. The first is designed to store information about the laboratory state, including raw and processed data and metadata of the instruments, as well as the input and output of experimental processes. We implement this desired modularity in the database layout by defining key universal sub-tables, device, job, and devicelog, and more specialized ones tailored to the needs of each instrument. This concise layout ensures maximum flexibility and a robust architecture for experimental data collection that is readily transferable to any SDL. A detailed layout of the experimental database can be found in Note S1 and Figure S1. On the other hand, the simulation database is managed by AiiDA, which provides a well-tested and robust design specifically tailored for simulation orchestration.<sup>12</sup> These databases are not only used to keep track of all the data collected in the laboratory, but they also serve as a broker to share data between different devices and software running in the laboratory. ChemOS 2.0's databases can be easily exported and published into specialized databases to report complete experimental information. As such, we gain a more comprehensive snapshot of the laboratory, encompassing all instructions executed by experimental devices and their corresponding metadata.

In concluding our data management approach, the choice to implement a relational database management system (RDBMS) was guided by its capability for precise data handling and executing complex queries, crucial for ensuring data integrity in laboratory environments. This system's universal-specialized schema design offers both broad lab compatibility and specific customization options. Although the flexibility and scalability of NoSQL databases make them appealing, especially for managing diverse data types in environments where multiple laboratories operate in parallel, they were set aside in favor of RDBMS's current suitability. However, as we look to the future where rapid response and scalability become paramount in multi-laboratory settings, the integration of NoSQL remains a considered option.

### Device communication

Enabling seamless communication between hardware and applications is a crucial role of an operating system, particularly in the context of SDLs, where applications can be likened to chemical procedures with multiple steps. In ChemOS 2.0, the



**Figure 2. Laboratory distribution**

The self-driving lab is orchestrated using ChemOS 2.0, which leverages the SiLA2 client/server protocol to control laboratory devices. The client interacts with the servers on the network to access available actions. Each server operates as a daemon on the respective computer connected to the laboratory instrument. In our configuration, device-specific managers also run as daemons, providing a GUI for user interaction with the instrument. Communication between the SiLA2 servers and managers is facilitated through a computational socket, enabling command transmission via ChemOS 2.0 or the GUI. This setup empowers users to command various laboratory instruments: (A) MEDUSA, a robotic organic synthesis platform<sup>45</sup>; (B) Chemspeed, a robotic general synthesis platform; (C) an HPLC-MS module for characterization; and (D) our custom-built automated optical characterization able to characterize lasing properties.

Instruments (B)–(D) are physically connected, enabling direct transfer of materials.

interaction between the system and hardware is facilitated through a core middleware software unit, operating on the terminals physically connected to the devices and the local network, as depicted in Figure 2 with the SiLA2 client/server standard managing the communications.<sup>27</sup> SiLA2 is specifically developed for communication within laboratories and is a highly flexible option. By contrast, other standards, such as MQTT or Redis, are primarily designed for lower-data-size communication and are more suited for the IoT rather than laboratory environments. Therefore, SiLA2's tailored features for laboratory settings give it a distinct advantage over these alternatives. In our implementation, the instruments themselves are treated as SiLA2 servers, exposing all their valid actions to the local network, while our orchestration device acts as a client capable of querying the devices to execute their operations. The SiLA2 package comes with a suite of commands that generate, deploy, and maintain the available actions of the SiLA2 servers by defining XML files. Details of using SiLA2 in Python can be found in Note S2 and Figures S2, S3, and S4. An example XML file can be found in Note S3. In adherence with the third UNIX principle, job files to the SiLA2 servers are all in a human-readable JSON format for all of the instruments (see Note S4). The manager component often serves as a device driver and commonly includes highly abstracted code capable of executing complex

actions. Commercial solutions usually consist of graphical user interfaces (GUIs) tailored for individual operations and usually lack an application programming interface (API) for direct communication.<sup>44</sup> Consequently, custom software modifications are often required to exert control over these devices. The manager therefore provides a “workaround” for some instruments that are limited by their provided user interfaces for automation, or a device driver for more bare-bones/simplistic instruments. To bridge the communication gap between the (standardized) SiLA2 server and the (device-specific) manager, we implemented a socket server. Although the decoupling introduced by this connection adds complexity to the code architecture, the socket communication layer acts as a safeguard in case of a loss of connection during the execution of sensitive/hazardous operations. The socket layer also addresses the practical need for our researchers to sometimes operate instruments without ChemOS 2.0, and critically, ensure the continued operation of the instrument in the event of a loss of connection between ChemOS 2.0 and the instrument. A deeper description of the SiLA2 server implementation and the socket server can be found in [Notes S5, S6, and S7](#) and [Figures S2, S3, and S4](#), respectively. A code example of interacting with a SiLA2 server can be found in [Note S8](#).

Last, we have successfully deployed the aforementioned middleware unit in our experimental laboratory. [Figure 2](#) illustrates our current setup, which includes (1) MEDUSA ([Note S9](#)), a robotic pump system capable of parallelized synthesis in solution<sup>45</sup>; (2) Chemspeed platform, an automated robotic framework for parallelized synthesis and characterization workflows ([Note S10](#) and [Figure S4](#))<sup>46</sup>; (3) a high-performance liquid chromatography/mass spectrometry (HPLC-MS) device ([Note S11](#)) for compound separation and identification; and (4) the custom-built automated optical characterization: a flow setup for automated spectroscopy<sup>46</sup> ([Note S12](#)). The available commands for each of the experimental units can be found in [Tables S1, S2, S3, S4, and S5](#). To test the SiLA2 server before deployment, we used “simulation” versions of the workflow managers ([Note S5](#)), ensuring minimal interruptions of operations while the server is in development. The operations of each instrument are controlled by an independent computer connected to the local network.

We have also explored additional capabilities of our system by successfully integrating a low-level sensor connected to a Raspberry Pi for real-time monitoring of laboratory conditions such as humidity and temperature. This integration exemplifies the adaptability of ChemOS 2.0 to incorporate custom devices, as detailed in [Note S13](#) and [Figure S5](#). Furthermore, to showcase the system’s parallelization/asynchronous potential shown in other SDL setups, we have conducted a simulation consisting of two HPLC-MS emulators run concurrently. This setup, configured to work in tandem with the Chemspeed platform, demonstrates efficient asynchronous processing of samples, highlighting the flexibility and scalability of the SiLA2 protocol ([Note S14](#) and [Figure S6](#)). Together, these tests not only validate the robustness and versatility of our communication protocol but also reinforce the capability of ChemOS 2.0 to adapt to diverse and evolving laboratory environments.

In addition to facilitating communication between devices, a critical aspect of an orchestrator is error handling. Our communication layer, SiLA2, enhances this safety aspect by allowing users to define and manage custom exceptions. These exceptions are raised by the server and communicated to the client in the event of an issue, enabling precise and tailored responses to specific errors. For instance, in the event of physical instrument failures, such as a sudden controller crash or communication interruptions, the SiLA2 server rapidly communicates the problem to the client. This triggers predefined protocols, including alerting laboratory managers and, where

possible, executing coded recovery actions without restarting the entire workflow. This approach not only ensures operational continuity but also maintains safety standards, especially during sensitive or hazardous operations. Additionally, in scenarios where data validity is in question, such as readings from HPLC-MS or optical tables, the system is capable of raising alerts for suspected measurement errors. By programmatically integrating lab workflows, ChemOS 2.0 unlocks the ability to automatically detect and troubleshoot experimental data. To validate and demonstrate these error-handling capabilities, we have included a detailed simulation case in [Note S15](#) and [Figure S7](#), showing how ChemOS 2.0 effectively manages and responds to various error scenarios. These simulators are faulty and will crash at random, allowing ChemOS 2.0 to demonstrate its error-handling capabilities.

### DFT integration

While experimental laboratories contribute significantly to scientific research, the forefront of chemical exploration relies on *ab initio* simulations to attain profound insights into complex chemical processes. Many of these simulations, however, come with substantial computational demands and require specialized external computational facilities for execution. The need for simulations in chemical exploration to improve the quality and efficiency of research makes it crucial to equip ChemOS 2.0 with a dedicated unit capable of handling such chemical simulations.

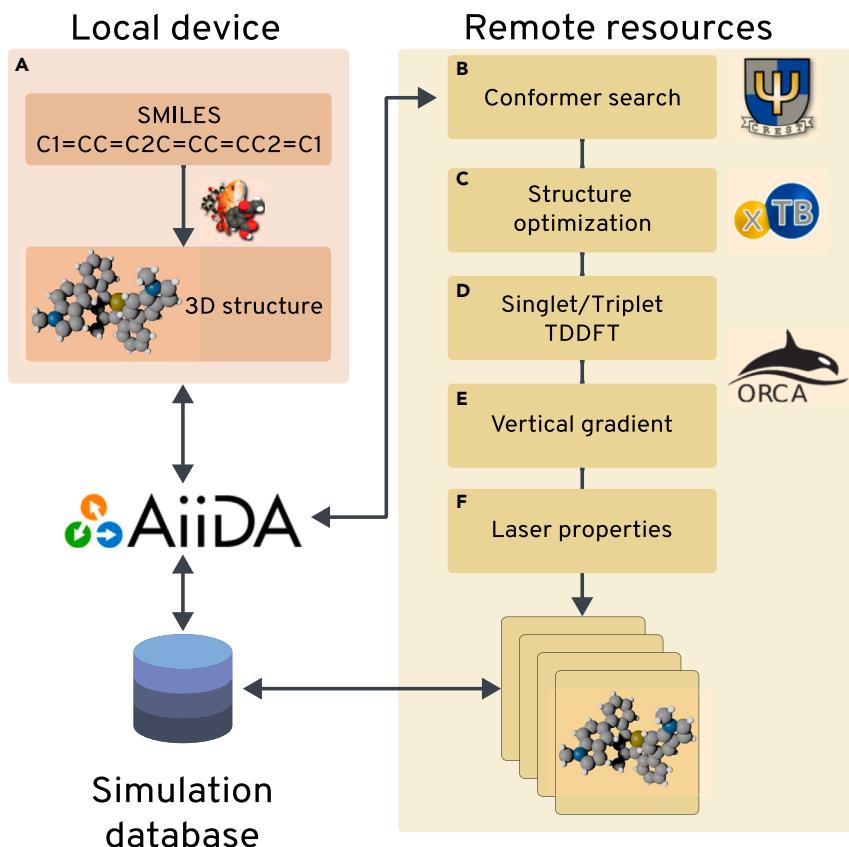
To tackle the computational challenges and ensure reproducibility, ChemOS 2.0 embraces the integration of the AiiDA software package.<sup>12</sup> AiiDA plays a critical role in automating data transfers between the user's local environment and the high-performance supercomputing cluster. By utilizing the SSH protocol, AiiDA submits calculations to an SLURM queue, retrieves outputs, and stores the results in a centralized database. It also resolves relevant dependencies between the different steps, helping to properly schedule their execution. To ensure reproducibility, AiiDA stores the data and metadata for each step of the workflow in the simulation database and provides Python bindings to ease the query. Additionally, AiiDA is prepared for distributed computing, enabling efficient utilization of computational resources. As a result, lightweight operations such as 3D molecular structure generation can be performed locally, while computationally intensive tasks utilizing packages like ORCA<sup>47</sup> and xtb<sup>48</sup> can be executed on our high-performance supercomputing cluster.

In this work, we leverage the capabilities of AiiDA to create a powerful DFT workflow for estimating the lasing performance of organic laser molecules. Taking a SMILES string as input, the AiiDA-enabled workchain efficiently executes the necessary calculations, monitors their status, and retrieves critical data related to the lasing performance. These data are subsequently used in a Bayesian optimization campaign, enabling researchers to make informed decisions and advance their chemical exploration efforts. The workflow is depicted in [Figure 3](#) and additional information can be found in the methods section.

### Web interface

ChemOS 2.0 operates as a server-based system, primarily relying on usage via command-line interactions. While this approach is well-suited for specialists, it may not be the preferred method for chemists who are more accustomed to GUIs.

To facilitate the adoption of our system, we developed a user-friendly web application interface, empowering any researcher to send experimental jobs easily, monitor their progress, and analyze the resulting data. This layout is shown in [Figure 4](#). The

**Figure 3. DFT workflow to compute the laser properties**

- (A) Building a 3D structure of the molecule from a SMILES string using Open Babel.<sup>69</sup>
- (B) Searching for the lowest-energy conformer using the crest package.<sup>64</sup>
- (C) Optimizing the ground state structure and calculating the corresponding Hessian of the lowest-energy conformer using the xTB-GFN2 semi-empirical Hamiltonian.<sup>48</sup>
- (D) Evaluating the ground and excited state energies and gradient of both singlet and triplet states at the time-dependent DFT level using ORCA.<sup>47</sup>
- (E) Determining the vibrationally coupled absorption and emission spectra using the vertical gradient approximation.
- (F) Computing a proxy for lasing properties.<sup>70</sup>

web application is built using the Streamlit package,<sup>49</sup> is hosted on the orchestration device, and includes a sidebar menu that enables toggling between the available devices of the laboratory. Once the user clicks into a tab, the web interface shows the actions of the device and the inputs needed for their execution. Users can then upload job files, visualize and download results, and control the Bayesian optimizer using the application. Workflows involving multiple steps need to be coded in Python using scripts that send commands to the instruments and modules. Moreover, the SiLA2-type system offers a promising approach for implementing a no-code platform, enabling users to define workflows in a more user-friendly manner.

Overall, the introduction of this web application interface significantly simplifies the usability of our system and enhances its accessibility for researchers across diverse skill levels. In this work, we have implemented functionality for all the available hardware presented in the device communication section. While a full-featured web interface would enhance the capability of ChemOS 2.0, its development would



**Figure 4. Web application for ChemOS 2.0**

- (A) Sidebar for toggling between instruments/features.
- (B) Job submission for the HPLC-MS.
- (C) Data visualization of the HPLC-MS's job results.

require substantial efforts from additional fields and is beyond the scope of this study.

**Experimental planner**

Experimental planners are the keystone of SDL, prompting us to equip ChemOS 2.0 with Atlas,<sup>50</sup> a package for Bayesian optimization (BO) designed explicitly for experimental sciences. Atlas is a general-purpose optimization framework for expensive-to-evaluate black-box problems, capable of mixed-parameter and multi-objective optimization. Importantly, Atlas incorporates BO concepts particularly relevant to chemistry, such as a molecular kernel function,<sup>51</sup> general experimental parameter optimization,<sup>52</sup> robust optimization,<sup>53</sup> and asynchronous experimental execution.<sup>54,55</sup> For more information, see Note S16 and Hickman et al.,<sup>50</sup> where we show an additional closed-loop electrochemical campaign orchestrated by ChemOS 2.0.

**Experimental campaign**

To demonstrate a proof-of-concept of a fully closed loop of automated experiment planning and execution, we designed a multi-objective campaign for the synthesis of Bis[(N-carbazole)styryl]biphenyl (BSBCz) derivatives based on previous work.<sup>46</sup> The campaign goals are to simultaneously maximize the experimental gain cross section and the simulated lasing gain factor (produced via DFT simulations), involving BO-based experiment planning, automated synthesis, purification, and functional characterization, as well as simultaneous computational simulation via the AiiDA module.

The search space of potential target molecules in this campaign is defined by enumerating all products of a double Suzuki-Miyaura coupling between the MIDA ester (MIDA = N-methyliminodiacetic acid) of (E)-2-(4-(N-Carbazoly)phenyl)-vinyl-1-boronic acid and 38 commercially available dihalides.<sup>46</sup> A detailed synthesis scheme can be found in Figure S8. This space is navigated by Atlas using a Gaussian process surrogate model<sup>56</sup> with a Tanimoto similarity kernel function.<sup>51</sup> Initial recommendations were generated by training the optimizer on two data points from the original work. The recommended target molecules are then synthesized and characterized experimentally, while spectroscopic properties are simulated computationally through the aforementioned workflow. We provide Atlas with the goal of maximizing both the simulated spectral gain factor (from the AiiDA workflow), as well as the experimental gain cross section via an equally weighted expected hypervolume improvement acquisition function that jointly maximizes all objectives.

**Testing hardware functionality with MEDUSA**

Before creating a fully closed loop, we evaluated the experimental capabilities of our setup by performing a human-in-the-loop orchestration test excluding an experimental planner or DFT simulations. The test involved the synthesis, characterization, and measurement of optical properties for a compound recommended by our experimental planner.

The synthesis was performed using MEDUSA. During the test, ChemOS 2.0 directed the synthesis of the suggested BSBCz derivative as recommended during the creation of the BO campaign by Atlas. The synthesis instructions were seamlessly

communicated to MEDUSA through its SiLA2 server, and real-time data were logged in the experimental database.

Following synthesis, the crude reaction mixture was manually filtered and transferred to the HPLC-MS device for characterization. Purification (HPLC-MS) and spectroscopic characterization (custom-built automated optical characterization platform) were controlled by ChemOS 2.0 through SiLA2 commands. The desired product was successfully detected by HPLC-MS, and the purified product underwent downstream optical property analysis. All data produced during this test can be found in [Note S17](#).

#### Building a closed-loop workflow with the Chemspeed platform

Last, we deployed and executed a closed-loop workflow for the synthesis of BSBCz derivatives. In addition to orchestrating the experimental component of the campaign, we conducted DFT simulations parallel to satisfy the two objectives required by the BO campaign (in the experimental planner section). Details can be found in [Note S18](#) and [Figure S9](#). For the closed-loop case study, we used Chemspeed to synthesize the BSBCz derivative. Creating the closed loop encompassed sending the SiLA2 commands to the SiLA2 servers.

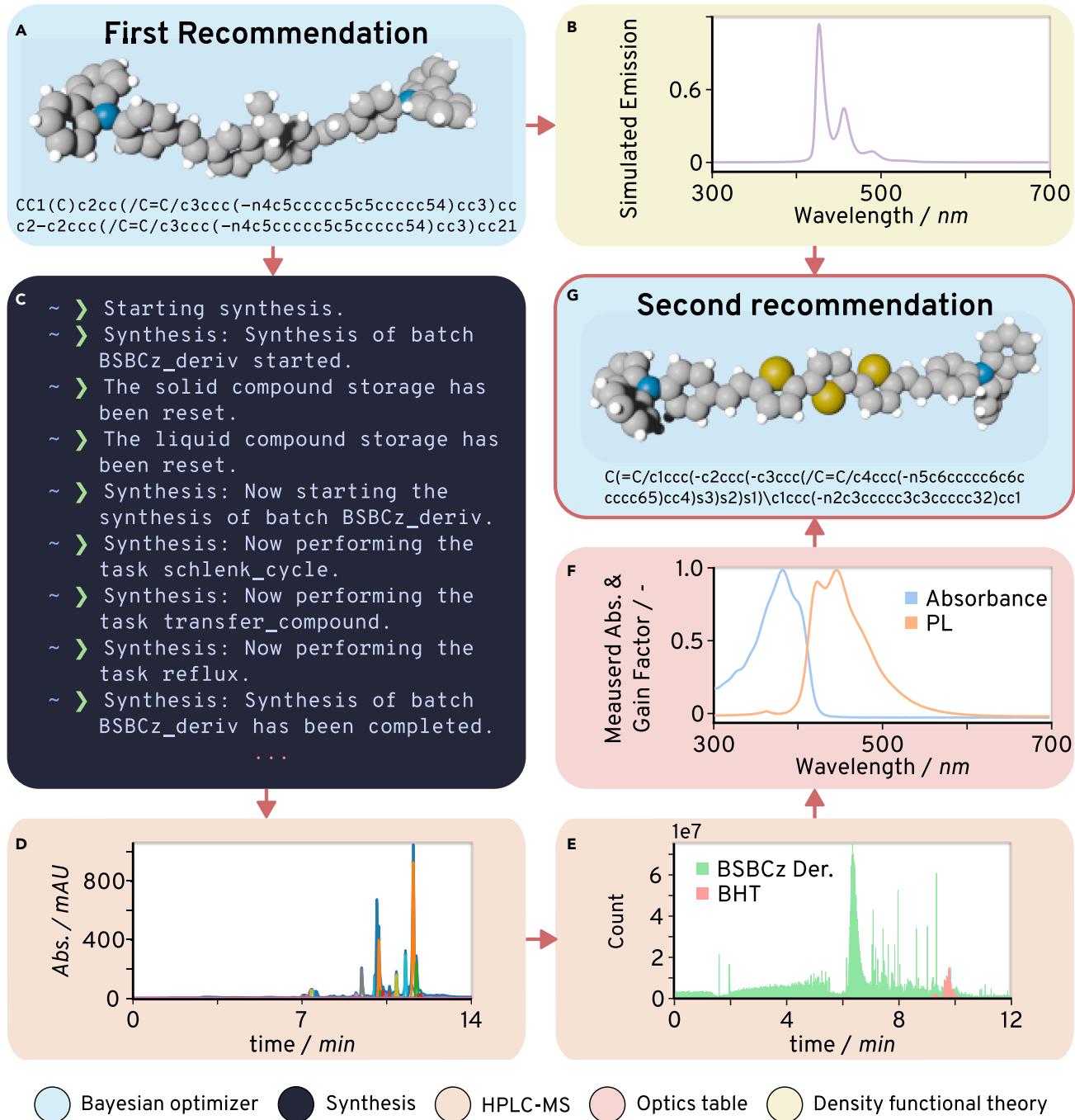
Additional information about the design of the campaign can be found in [Note S19](#).

The results and workflow are illustrated in [Figure 5](#) showing the following steps: (1) ChemOS 2.0 queries the Atlas Bayesian optimizer with two literature-known “seed” observations and receives a recommendation for which BSBCz derivative to synthesize. (2) Then, it instructs Chemspeed to execute the synthesis. (3) When the reaction is complete, ChemOS 2.0 sends the compounds to the HPLC-MS to detect the product in the crude mixture and transfers the purified BSBCz derivative to the custom-built automated optical characterization platform. (4) Finally, the optical characterization platform measures sample fluorescence and absorbance spectra and the emission lifetime to compute the gain cross section. (5) In parallel to the experimental loop, the computational job is sent to the AiiDA daemon to obtain the simulated spectral gain factor. (6) The results from the two previous steps are used to retrain the Bayesian optimizer and query another candidate. All logging information was stored in the databases using the database architecture described in the data management section. The experimental gain cross-sectional value was estimated experimentally using fluorescence spectra, relative quantum yields, and emission rates of the BSBCz derivatives using our automated optical characterization platform. Details of the calculation can be found in previous work.<sup>46</sup>

In summary, ChemOS 2.0 was able to successfully orchestrate a full SDL loop, including manipulation of different instruments, an AI experiment planner, and stored all output data in an internal database.

#### DISCUSSION

We demonstrate that more general, transferable, and multifaceted SDLs are possible thanks to central lab orchestration. Building upon its predecessor, ChemOS 2.0 showcases significant advancements in automated laboratory systems. Tested in real-world laboratory environments using rigorous state-of-the-art experimental methods, it addresses key challenges in orchestrating SDLs. Its enhancements include *ab initio* simulations integration, a versatile communication protocol, a middleware layer for instrument control, a modular database, delocalized



(○) Bayesian optimizer (●) Synthesis (○) HPLC-MS (○) Optics table (○) Density functional theory

**Figure 5. Results saved on ChemOS 2.0 for the synthesis of BSBCz derivatives**

- (A) First recommendation candidate made by the experimental planner.
- (B) Simulated fluorescence spectrum of BSBCz derivative produced by DFT workflows on ChemOS 2.0.
- (C) Logging information for organic synthesis on Chemspeed.
- (D) Diode array detector (DAD) chromatogram and (E) extracted ion chromatogram (XIC) plots for the HPLC-MS.
- (F) Time-correlated single photon counting and absorbance/fluorescence spectra of the custom-built automated optical characterization platform analysis.
- (G) Second recommendation made by the experimental planner based on the results of (B) and (F). All results and raw data files are stored in the database during the execution of experiments.

computational resources, a package manager for reproducibility, and a prototype web interface for laboratory management. The modular nature of ChemOS 2.0 enables laboratories from different fields to configure it according to their domain-specific requirements.

Following the UNIX philosophy, ChemOS 2.0 acts an “operative system,” with tools and instruments working together to create workflows and present data in a human-readable format. Unlike monolithic designs in other orchestration configurations, our approach reflects an operating system’s design philosophy, promoting flexibility and allowing users to easily modify layers and incorporate new features into the setup, such as new instruments and experiment planners. The database design follows the same principles, allowing users to rapidly add new devices to automated workflows and data collection schemes using a universal-specialized table scheme as well as device-specific additions. Thus, rather than building a framework specific to our setup, we have designed an architecture amenable to this experimental lab along with a diverse set of laboratory settings. Moreover, by employing a central fog orchestration device as the core of our architecture, we ensure greater robustness and adaptability, allowing an automated management of laboratory services and ensuring software reproducibility.

One of the remarkable advantages of our modular architecture is its support for flexible human/robot integration of automated tools. Instead of demanding a complete overhaul of laboratory layout and hardware, ChemOS 2.0 allows the automation of existing tools and minimally interrupts existing workflows. A drastic transformation of existing labs is often impractical, as it disrupts the research activity of academics and industry professionals. This can discourage them from transitioning to digitization. ChemOS 2.0 has been designed with a hybrid integration in mind, allowing researchers to gradually introduce automation into their workflows. Initially, the Nix configuration can be installed on a workstation in the laboratory and later expanded upon. Users can begin by integrating simpler devices, before adding instruments one by one to their ChemOS 2.0 ecosystem. While testing connectivity between ChemOS 2.0 and the instrument servers, simulators were used instead of the actual lab hardware. As such, the deployment of the SiLA2 servers and testing of the actual hardware was therefore minimized, and regular operations of the laboratory were minimally interrupted.

Digitization, including the implementation of tools like ChemOS 2.0, represents a considerable challenge for many laboratories, as it requires significant experience in the efficient use and integration of computational resources. We believe this creates the need for a new, emerging role in research laboratories: a “digital manager,” responsible for orchestration, workflow assembly, and the integration of new (digital) tools, which we believe will be crucial to the proliferation of SDLs. Moreover, while it is a financial barrier for automating many laboratories, we remain optimistic for the future given that the cost of automated tools is steadily decreasing each year.<sup>57</sup> Following digital managers and cost, another hurdle may be less straightforward to address: the API-type access to automated tools. To aid the proliferation of SDLs, manufacturers should comprehensively support the automation of their hardware, and recognize that in doing so, they can actually encourage more adoption of their products by academia, startups, and industry.

Importantly, our work represents a prototype implementation of the discussed design principles, and leaves room for structural refinements and further software development. As SDLs become more prevalent, the connection between

orchestration devices will become a significant consideration, presenting challenges related to scalability and cybersecurity. Furthermore, the increasing complexity of laboratory environments will necessitate the development of more powerful and user-friendly interfaces. Graphical interfaces and natural language processing<sup>58</sup> enable users to communicate with the laboratory and coupled with advanced data analysis methods, will allow programming-averse users to leverage the features of the ChemOS 2.0 platform effectively.

### Conclusions

In conclusion, this work represents an advancement in promoting the adoption of innovative digital tools by chemists in their laboratories. The integration of BO, experimental techniques, *ab initio* simulations, and data management within ChemOS 2.0 unifies automated laboratories to conduct sophisticated workflows. Our hybrid approach not only relies on experimental results but also incorporates in-depth sample analysis through DFT calculations, enabling comprehensive data-driven research. The automation of data collection opens up exciting opportunities for machine learning and data science applications in chemistry, fueling accelerated research progress and discoveries. Looking ahead, we envision even greater possibilities through enhanced communication and collaboration among multiple ChemOS 2.0 platforms, fostering seamless cooperation between different laboratories. ChemOS 2.0 lays a solid foundation for designing novel SDLs, making them more modular and robust and making their units transferable. This advancement supposes a step forward in the quest to revolutionize materials research through streamlined and efficient SDLs.

## EXPERIMENTAL PROCEDURES

### Resource availability

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sergio Pablo-García ([spgarcica@gmail.com](mailto:spgarcica@gmail.com)).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

As a showcase, the results of the DFT calculations have been uploaded to ioChem-BD.<sup>39,59</sup>

All the code needed for this work can be found in our GitHub repository.<sup>38</sup> Frozen copies of the major ChemOS 2.0 versions can be found on Zenodo: <https://www.doi.org/10.5281/zenodo.11089654>. Tutorials can also be found to replicate the more generalizable parts of this work. The experimental data can also be found in this repository.

### Closed-loop workflow

The synthesis of BSBCz derivatives was performed using the Chemspeed SWING XL automated platform, available at the University of Toronto. For more details of the automation manager for the Chemspeed platform, see Note S3. Automated Suzuki-Miyaura couplings were performed following the general slow-release conditions reported by Grzybowski, Burke, and coworkers.<sup>45</sup> Synthesis instructions were received from the SiLA2 client on ChemOS 2.0. Reactions were performed in the ISYNTH reactor under a nitrogen atmosphere, established by five evacuate-backfill cycles. Reactants and reagents were transferred as stock solutions in 1,4-Dioxane or water, respectively, under a constant stream of nitrogen, using the 4-Needle liquid

dispensing tool. The reaction mixture was heated to 100°C under vortex stirring for 12 h. The reaction mixture was cooled down to room temperature for 1 h, and the crude reaction mixture was filtered over a short plug of Celite (using the Solid Phase Extraction module of the Chemspeed SWING XL platform), eluting with tetrahydrofuran and collecting the eluate. Following communication with the HPLC-MS device through the respective SiLA2 client on ChemOS (see details below), the eluate is automatically injected to the HPLC-MS using the on-deck injection valve equipped with a sample loop.

### Separation and identification in HPLC-MS

Analysis of crude reaction mixtures and product isolation was performed by HPLC-MS on a ThermoFisher system, equipped with a Vanquish HPLC and a QExactive mass spectrometer. Injection to the HPLC-MS system was realized through a six-port injection valve on the deck of the Chemspeed SWING XL platform. Job submission and orchestration of operations between the two devices was enabled by ChemOS 2.0. Upon submission of a job from ChemOS 2.0 to the HPLC-MS device, an HPLC-MS run is automatically started. After the equilibration stage, ChemOS 2.0 can read the "ready" status from the HPLC-MS device and commence the analyte transfer into the sample loop on the Chemspeed SWING XL. Data acquisition on the HPLC-MS device is started through an electronic trigger sent from the injection valve driver on the Chemspeed SWING XL, indicating the exact sample injection time. Upon completion of the first analysis run, the target compounds are identified by matching peaks in the wavelength-averaged UV/vis chromatogram and their extracted ion chromatograms, respectively. The detailed workflow for peak identification is described in previous work.<sup>46</sup> All analysis results are stored on ChemOS 2.0. In case of successful detection of the desired reaction product, a second HPLC-MS run is initialized by ChemOS 2.0, following the workflow detailed above. In this case, the fraction containing the desired product is automatically collected using a downstream selection valve and is used for spectroscopic characterization (see below).

### Custom-built automated optical characterization

The collected sample is automatically characterized in the spectroscopic workflow described in Note S5, orchestrated by ChemOS 2.0. Upon successful completion of the HPLC-MS run and automated fraction collection, spectroscopic analysis can be initiated from ChemOS 2.0. On the custom-built device, the sample is transferred to different flow cuvettes using a syringe pump. Steady-state absorption spectra are recorded using a white-light lamp (Ocean Insight DH-MINI) combined with a spectrometer (Ocean Insight, QEPro XR). Steady-state emission spectra were recorded on the same spectrometer using a 365-nm solid-state LED (Thorlabs, M365FP1 with DC4100HUB) as the excitation source. Quantum yields were determined by integration of the emission spectrum, while simultaneously recording the excitation transmission using a photodetector (Thorlabs, S120VC with PM100D). Quantum yields were referenced to a pre-measured sample of 4,4'-Bis[(N-carbazole)styryl] biphenyl ( $\phi = 0.89$  in toluene).<sup>60</sup> Time-resolved emission spectra were recorded using time-correlated single-photon counting (PicoQuant TimeHarp 260 board) after excitation with a picosecond-pulsed laser (PicoQuant, LDH-D-C-375 with PL800-D), and the corresponding lifetimes were obtained by exponential fitting. Details on the data analysis workflow are described in Wu et al.<sup>46</sup> Upon completion of the analysis workflow, all analysis results are automatically stored on ChemOS 2.0.

### MEDUSA

MEDUSA is a synthetic platform developed from "The Machine,"<sup>45</sup> consisting of syringe pumps and selection valves connected by tubing and further to solution and

reaction vials and N<sub>2</sub> supply. The atmosphere handling was performed by inserting drawing gas from head-space of vials solution consistently refilling N<sub>2</sub> and purge syringe connect to exhaust. The automated ligand transfer was performed by washing the syringe with desired solution by pumping from the source and dispensing to waste, followed by drawing a certain amount and dispensing to the destination vial. Finally, flushing required drawing N<sub>2</sub> and flushing tubing content to destination. Each hardware execution was recorded and streamed as standard output. A final exit message was sent at the completion of the experiment.

### Computational procedures

All SiLA2 servers were constructed using the Python implementation of the SiLA2 0.10.1 communication standard. The SiLA2 servers operate in a separate thread on the computers that control the driver for the instruments in the lab. Communication between the SiLA2 server and the instrument on the local computer is accomplished using either socket connections or via status folders and files. Meanwhile, all communication between the local instrument and other devices on the network is only done via the SiLA2 server.

ChemOS 2.0 uses NixOS 22.11 configuration files. Using this configuration file will produce an identical shell on any Linux device that has NixOS installed. The experimental database was constructed using SQLAlchemy 1.4.41,<sup>61</sup> and the simulation database was built by AiiDA.<sup>12</sup> Both databases use PostgreSQL 14.6 as the backend. Interaction with the database on ChemOS 2.0 is accomplished through the use of psycopg2 2.9.5.<sup>62</sup> All storage of data in the database is done automatically through Python scripts executed on ChemOS 2.0. Connecting ChemOS 2.0 to the other devices (SiLA2 servers) in the lab is easy and straightforward. One has to create a client Python object and initialize it using the IP address and port of the targeted SiLA2 server. Afterward, one can send any of the available commands to the SiLA2 server (the instrument), which are represented as methods of the SilaClient object. Calling a method of the client object returns a SiLA2 command instance. If a command is observable, it is possible to subscribe to the command and receive a stream of information as the command is executing. It is through this mechanism that output data from the instruments is streamed to the fog server.

Tracking the progress of experimental campaigns with Atlas 0.0.0 is accomplished by using Olympus 0.0.1b0,<sup>63</sup> a benchmarking framework for BO. Using Olympus' Campaign object, we can keep track of the search space and previous observations over the course of the optimization. To use the Atlas SiLA2 server, The Olympus Campaign object is passed to the optimizer and ultimately to Atlas, which uses the observations attribute of the Campaign as input for suggesting new parameters. Saving the campaign state as a .pkl file is one way to track each successive step of the experiments. The Atlas SiLA2 server returns a parameter vector, which can then be added to the Campaign and used to automate the next round of experiments.

### Density functional theory

Generation of molecular conformers was performed using crest 2.11.1.<sup>64</sup> Semi-empirical geometry optimization and frequency calculations of S0 and T1 structures were performed at gfn2-xtb level of theory using xtb 6.4.1.<sup>48</sup> DFT geometry optimization and frequency calculations of S0 structure were performed with the ORCA 5.0.4 package,<sup>47</sup> at PBE0-D3/Def2-SV(P) level of theory. TD-DFT single-point and gradient calculations were performed with the ORCA 5.0.4 package, at PBE0-D3/Def2-SV(P) level of theory, with Tamm-Danoff approximation.<sup>65–67</sup> The total

electronic energies of S0, S1-S5, and T1-T5 and the gradients of S0, S1-S3, and T1-T3 were computed at the DFT-optimized S0 structure. AiiDA 2.4.0 and AiiDA-shell 0.5.3 were used for workflow orchestration.<sup>12,68</sup>

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.matt.2024.04.022>.

## ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, subaward by University of Minnesota, project title "Development of Machine Learning and Molecular Simulation Approaches to Accelerate the Discovery of Porous Materials for Energy-Relevant Applications," under award number DE-SC0023454 (UMN subaward A010026303) A.A.-G. acknowledges support from the Canada 150 Research Chairs Program as well as Anders G. Frøseth. Computations were performed on the Niagara supercomputer at the SciNet HPC Consortium. SciNet is funded by the Canada Foundation for Innovation. M.S. acknowledges tremendous support from the Ontario Student Assistance Program's Ontario Graduate Scholarship, the Government of Ontario, Ontario Research Fund - Research Excellence, and the University of Toronto.

## AUTHOR CONTRIBUTIONS

M.S., software, methodology, investigation, writing original draft. M.G.V., methodology, writing, conceptualization, architecture design. F.S.-K., preparation of experimental workflows, writing review and editing. H.H., preparation of simulation and experimental workflows. R.J.H., implementation of the experiment planner, writing review and editing. S.M., conceptualization, writing review and editing. S.P.-G., methodology, writing, conceptualization, supervision. A.A.-G., conceptualization, funding acquisition, supervision, project administration, writing review and editing.

## DECLARATION OF INTERESTS

A.A.-G is a founder of Kebotix, Inc., and Intrepid Labs, Inc.

Received: August 24, 2023

Revised: January 31, 2024

Accepted: April 16, 2024

Published: May 14, 2024

## REFERENCES

1. Helm, D. (2012). The Kyoto approach has failed. *Nature* 491, 663–665. <https://doi.org/10.1038/491663a>.
2. Wouters, O.J., McKee, M., and Luyten, J. (2020). Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009–2018. *JAMA* 323, 844–853. <https://doi.org/10.1001/jama.2020.1166>.
3. Merrifield, R.B., and Stewart, J.M. (1965). Automated peptide synthesis. *Nature* 207, 522–523. <https://doi.org/10.1038/207522a0>.
4. Bédard, A.C., Adamo, A., Aroh, K.C., Russell, M.G., Bedermann, A.A., Torosian, J., Yue, B., Jensen, K.F., and Jamison, T.F. (2018). Reconfigurable system for automated optimization of diverse chemical reactions. *Science* 361, 1220–1225. <https://doi.org/10.1126/science.aat0650>.
5. Hartrampf, N., Saebi, A., Poskus, M., Gates, Z.P., Callahan, A.J., Cowfer, A.E., Hanna, S., Antilla, S., Schissel, C.K., Quartararo, A.J., et al. (2020). Synthesis of proteins by automated flow chemistry. *Science* 368, 980–987. <https://doi.org/10.1126/science.abb2491>.
6. Macarron, R., Banks, M.N., Bojanic, D., Burns, D.J., Cirovic, D.A., Garyantes, T., Green, D.V.S., Hertzberg, R.P., Janzen, W.P., Paslay, J.W., et al. (2011). Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* 10, 188–195. <https://doi.org/10.1038/nrd3368>.
7. Santanilla, A.B., Regalado, E.L., Pereira, T., Shevlin, M., Bateman, K., Campeau, L.-C., Schneeweis, J., Berritt, S., Shi, Z.-C., et al. (2015). Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science* 347, 49–53. <https://doi.org/10.1126/science.1259203>.
8. MacLeod, B.P., Parlane, F.G.L., Morrissey, T.D., Häse, F., Roch, L.M., Dettelbach, K.E., Moreira, R., Yunker, L.P.E., Rooney, M.B., Deeth, J.R., et al. (2020). Self-driving laboratory for accelerated discovery of thin-film materials. *Sci. Adv.* 6, eaaz8867. <https://doi.org/10.1126/sciadv.aaz8867>.
9. Kusne, A.G., Yu, H., Wu, C., Zhang, H., Hattrick-Simpers, J., DeCost, B., Sarker, S., Oses, C., Toher, C., Curtarolo, S., et al. (2020). On-the-fly

- closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* 11, 5966. <https://doi.org/10.1038/s41467-020-19597-w>.
10. Jain, A., Ong, S.P., Chen, W., Medasani, B., Qu, X., Kocher, M., Brafman, M., Petretto, G., Rignanese, G., Hautier, G., et al. (2015). Fireworks: a dynamic workflow system designed for high-throughput applications. *Concurr. Comput.* 27, 5037–5059. <https://doi.org/10.1002/cpe.3505>.
  11. Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., et al. (2021). Sustainable data analysis with Snakemake. *F1000Res.* 10, 33. <https://doi.org/10.12688/f1000research.29032.1>.
  12. Huber, S.P., Zoupanos, S., Uhrin, M., Talirz, L., Kahle, L., Häuselmann, R., Gresch, D., Müller, T., Yakutovich, A.V., Andersen, C.W., et al. (2020). AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Sci. Data* 7, 300. <https://doi.org/10.1038/s41597-020-00638-4>.
  13. Bo, C., Maseras, F., and López, N. (2018). The role of computational results databases in accelerating the discovery of catalysts. *Nat. Catal.* 1, 809–810. <https://doi.org/10.1038/s41929-018-0176-4>.
  14. Stach, E., DeCost, B., Kusne, A.G., Hattrick-Simpers, J., Brown, K.A., Reyes, K.G., Schrier, J., Billinge, S., Buonassisi, T., Foster, I., et al. (2021). Autonomous experimentation systems for materials development: a community perspective. *Matter* 4, 2702–2726. <https://doi.org/10.1016/j.matt.2021.06.036>.
  15. Abdel-Latif, K., Epps, R.W., Bateni, F., Han, S., Reyes, K.G., and Abolhasani, M. (2021). Self-Driven Multistep Quantum Dot Synthesis Enabled by Autonomous Robotic Experimentation in Flow. *Adv. Intell. Syst.* 3, 2000245. <https://doi.org/10.1002/adfm.202000245>.
  16. Tao, H., Wu, T., Kheiri, S., Aldeghi, M., Aspuru-Guzik, A., and Kumacheva, E. (2021). Self-Driving Platform for Metal Nanoparticle Synthesis: Combining Microfluidics and Machine Learning. *Adv. Funct. Mater.* 31, 2106725. <https://doi.org/10.1002/adfm.202106725>.
  17. Li, J., Li, J., Liu, R., Tu, Y., Li, Y., Cheng, J., He, T., and Zhu, X. (2020). Autonomous discovery of optically active chiral inorganic perovskite nanocrystals through an intelligent cloud lab. *Nat. Commun.* 11, 2046. <https://doi.org/10.1038/s41467-020-15728-5>.
  18. Roch, L.M., Häse, F., Kreisbeck, C., Tamayo-Mendoza, T., Yunker, L.P.E., Hein, J.E., and Aspuru-Guzik, A. (2020). ChemOS: An orchestration software to democratize autonomous discovery. *PLoS One* 15, e0229862. <https://doi.org/10.1371/journal.pone.0229862>.
  19. Deneault, J.R., Chang, J., Myung, J., Hooper, D., Armstrong, A., Pitt, M., and Maruyama, B. (2021). Toward autonomous additive manufacturing: Bayesian optimization on a 3D printer. *MRS Bull.* 46, 566–575. <https://doi.org/10.1557/s43577-021-00051-1>.
  20. Tamura, R., Tsuda, K., and Matsuda, S. (2023). NIMS-OS: an automation software to implement a closed loop between artificial intelligence and robotic experiments in materials science. *Sci. Technol. Adv. Mater.* 3, 2232297. <https://doi.org/10.1080/27660400.2023.2232297>.
  21. Kusne, A.G., and McDannald, A. (2023). Scalable multi-agent lab framework for lab optimization. *Matter* 6, 1880–1893. <https://doi.org/10.1016/j.matt.2023.03.022>.
  22. Rahamanian, F., Flowers, J., Guevarra, D., Richter, M., Fichtner, M., Donnelly, P., Gregoire, J.M., and Stein, H.S. (2022). Enabling modular autonomous feedback-loops in materials science through hierarchical experimental laboratory automation and orchestration. *Adv. Mater. Interfaces* 9, 2101987. <https://doi.org/10.1002/admi.202101987>.
  23. Maffettone, P.M., Campbell, S., Hanwell, M.D., Wilkins, S., and Olds, D. (2022). Delivering real-time multi-modal materials analysis with enterprise beamlines. *Cell Rep.* 3, 101112. <https://doi.org/10.1016/j.xcrp.2022.101112>.
  24. Fitzpatrick, D.E., Maujean, T., Evans, A.C., and Ley, S.V. (2018). Across-the-world automated optimization and continuous-flow synthesis of pharmaceutical agents operating through a cloud-based server. *Angew. Chem. Int. Ed.* 57, 15128–15132. <https://doi.org/10.1002/anie.201809080>.
  25. Fitzpatrick, D.E., Battilocchio, C., and Ley, S.V. (2016). A Novel Internet-Based Reaction Monitoring, Control and Autonomous Self-Optimization Platform for Chemical Synthesis. *Org. Process. Res. Dev.* 20, 386–394. <https://doi.org/10.1021/acs.oprd.5b00313>.
  26. Leong, C.J., Low, K.Y.A., Recatala-Gomez, J., Quijano Velasco, P., Vissol-Gaudin, E., Tan, J.D., Ramalingam, B., Made, R.I., Pethe, S.D., Sebastian, S., et al. (2022). An object-oriented framework to enable workflow evolution across materials acceleration platforms. *Matter* 5, 3124–3134. <https://doi.org/10.1016/j.matt.2022.08.017>.
  27. Bromig, L., Leiter, D., Mardale, A.-V., von den Eichen, N., Bieringer, E., and Weuster-Botz, D. (2022). The SiLA 2 Manager for rapid device integration and workflow automation. *SoftwareX* 17, 100991. <https://doi.org/10.1016/j.softx.2022.100991>.
  28. Manzano, J.S., Hou, W., Zalesskiy, S.S., Frei, P., Wang, H., Kitson, P.J., and Cronin, L. (2022). An autonomous portable platform for universal chemical synthesis. *Nat. Chem.* 14, 1311–1318. <https://doi.org/10.1038/s41557-022-01016-w>.
  29. Rohrbach, S., Šiaučiulis, M., Chisholm, G., Pirvan, P.-A., Saleeb, M., Mehr, S.H.M., Trushina, E., Leonov, A.I., Keenan, G., Khan, A., et al. (2022). Digitization and validation of a chemical synthesis literature database in the ChemPU. *Science* 377, 172–180. <https://doi.org/10.1126/science.abo0058>.
  30. Fakhruldeen, H., Pizzuto, G., Glowacki, J., and Cooper, A.I. (2022). ARChemist: Autonomous Robotic Chemistry System Architecture. In 2022 International Conference on Robotics and Automation (ICRA) (IEEE), pp. 6013–6019. <https://doi.org/10.1109/ICRA46639.2022.9811996>.
  31. Steiner, S., Wolf, J., Glatzel, S., Andreou, A., Granda, J.M., Keenan, G., Hinkley, T., Aragon-Camarasa, G., Kitson, P.J., Angelone, D., and Cronin, L. (2019). Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* 363, eaav2211. <https://doi.org/10.1126/science.aav2211>.
  32. Shi, W., Cao, J., Zhang, Q., Li, Y., and Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet Things J.* 3, 637–646. <https://doi.org/10.1109/JIOT.2016.2579198>.
  33. Yi, S., Hao, Z., Qin, Z., and Li, Q. (2015). Fog Computing: Platform and Applications. In 2015 Third IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb) (IEEE), pp. 73–78. <https://doi.org/10.1109/HotWeb.2015.22>.
  34. Ritchie, D.M., and Thompson, K. (1974). The UNIX time-sharing system. *Commun. ACM* 17, 365–375. <https://doi.org/10.1145/361011.361061>.
  35. Raymond, E.S. (2003). *The Art of UNIX Programming* (Addison-Wesley Educational).
  36. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Sci. Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.
  37. Dolstra, E., and Löh, A. (2008). NixOS: a purely functional Linux distribution. In Proceedings of the 13th ACM SIGPLAN international conference on Functional programming (ICFP '08) (Association for Computing Machinery), pp. 367–378. <https://doi.org/10.1145/1411204.1411255>.
  38. Sim, M. (2023). ChemOS 2.0 (University of Toronto). <https://github.com/malcolm sim/ChemOS2.0>.
  39. Álvarez-Moreno, M., de Graaf, C., López, N., Maseras, F., Poblet, J.M., and Bo, C. (2015). Managing the computational chemistry big data problem: the ioChem-BD platform. *J. Chem. Inf. Model.* 55, 95–103. <https://doi.org/10.1021/ci500593j>.
  40. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. <https://doi.org/10.1093/nar/28.1.235>.
  41. Jain, A., Ong, S.P., Hautier, G., Chen, W., Richards, W.D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., and Persson, K.A. (2013). Commentary: the materials project: a materials genome approach to accelerating materials innovation. *Apl. Mater.* 1, 011002. <https://doi.org/10.1063/1.4812323>.
  42. Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., Tran, K., Heras-Domingo, J., Ho, C., Hu, W., et al. (2021). Open catalyst 2020 (OC20) dataset and community challenges. *ACS Catal.* 11, 6059–6072. <https://doi.org/10.1021/acscatal.0c04525>.
  43. Miret, S., Lee, K.L.K., Gonzales, C., Nassar, M., and Spellings, M. (2023). The open MatSci ML toolkit: A flexible framework for machine learning in materials science. Preprint at

- OpenReview. Trans. Mach. Learn. Res. <https://openreview.net/forum?id=QBMjyDZsPMd>.
44. Seifrid, M., Pollice, R., Aguilar-Granda, A., Morgan Chan, Z., Hotta, K., Ser, C.T., Vestfrid, J., Wu, T.C., and Aspuru-Guzik, A. (2022). Autonomous Chemical Experiments: Challenges and Perspectives on Establishing a Self-Driving Lab. *Acc. Chem. Res.* 55, 2454–2466. <https://doi.org/10.1021/acs.accounts.2c00220>.
  45. Li, J., Ballmer, S.G., Gillis, E.P., Fujii, S., Schmidt, M.J., Palazzolo, A.M.E., Lehmann, J.W., Morehouse, G.F., and Burke, M.D. (2015). Synthesis of many different types of organic small molecules using one automated process. *Science* 347, 1221–1226. <https://doi.org/10.1126/science.aaa5414>.
  46. Wu, T.C., Aguilar-Granda, A., Hotta, K., Yazdani, S.A., Pollice, R., Vestfrid, J., Hao, H., Lavigne, C., Seifrid, M., Angello, N., et al. (2023). A Materials Acceleration Platform for Organic Laser Discovery. *Adv. Mater.* 35, 2207070. <https://doi.org/10.1002/adma.202207070>.
  47. Neese, F., Wennmohs, F., Becker, U., and Riplinger, C. (2020). The ORCA quantum chemistry program package. *J. Chem. Phys.* 152, 224108. <https://doi.org/10.1063/5.0004608>.
  48. Bannwarth, C., Caldeweyher, E., Ehlert, S., Hansen, A., Pracht, P., Seibert, J., Spicher, S., and Grimme, S. (2020). Extended tight-binding quantum chemistry methods. *WIREs Comput. Mol. Sci.* 11, e1493. <https://doi.org/10.1002/wcms.1493>.
  49. Snowflake (2024). Streamlit (Snowflake Inc.). <https://streamlit.io>.
  50. Hickman, R., Sim, M., Pablo-García, S., Woolhouse, I., Hao, H., Bao, Z., Bannigan, P., Allen, C., Aldeghi, M., and Aspuru-Guzik, A. (2023). Atlas: A brain for self-driving laboratories. Preprint at ChemRxiv. <https://doi.org/10.26434/chemrxiv-2023-8nxx>.
  51. Griffiths, R.-R., Klarner, L., Moss, H.B., Ravuri, A., Truong, S., Stanton, S., Tom, G., Rankovic, B., Du, Y., Jamasb, A., et al. (2023). GAUCHE: A Library for Gaussian Processes in Chemistry. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2212.04450>.
  52. Angello, N.H., Rathore, V., Beker, W., Wołos, A., Jira, E.R., Roszak, R., Wu, T.C., Schroeder, C.M., Aspuru-Guzik, A., Grzybowski, B.A., and Burke, M.D. (2022). Closed-loop optimization of general reaction conditions for heteroaryl Suzuki-Miyaura coupling. *Science* 378, 399–405. <https://doi.org/10.1126/science.adc8743>.
  53. Aldeghi, M., Häse, F., Hickman, R.J., Tamblyn, I., and Aspuru-Guzik, A. (2021). Golem: an algorithm for robust experiment and process optimization. *Chem. Sci.* 12, 14792–14807. <https://doi.org/10.1039/D1SC01545A>.
  54. Ginsbourger, D., Janusevskis, J., and Riche, R.L. (2011). Dealing with asynchronicity in parallel Gaussian Process based global optimization. In Conference of the ERCIM WG on Computing and Statistics (HAL), pp. hal-00507632.
  55. Desautels, T., Krause, A., and Burdick, J.W. (2014). Parallelizing exploration-exploitation tradeoffs in Gaussian process bandit optimization. *J. Mach. Learn. Res.* 15, 4053–4103.
  56. Rasmussen, C.E., and Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning* (MIT Press).
  57. MacLeod, B.P., Parlane, F.G.L., Brown, A.K., Hein, J.E., and Berlinguet, C.P. (2022). Flexible automation accelerates materials discovery. *Nat. Mater.* 21, 722–726. <https://doi.org/10.1038/s41563-021-01156-3>.
  58. Bran, A.M., Cox, S., White, A.D., and Schwaller, P. (2023). Chemcrow: Augmenting large-language models with chemistry tools. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2304.05376>.
  59. Pablo-García, S. and Sim, M. ChemOS 2.0 DFT database, Version 1.0 (ioChem-BD). 10.19061/iochem-bd-10-1
  60. Oyama, Y., Mamada, M., Shukla, A., Moore, E.G., Lo, S.-C., Namdas, E.B., and Adachi, C. (2020). Design Strategy for Robust Organic Semiconductor Laser Dyes. *ACS Mater. Lett.* 2, 161–167. <https://doi.org/10.1021/acsmaterialslett.9b00536>.
  61. Bayer, M. (2012). SQLAlchemy. In *The Architecture of Open Source Applications* Volume II: Structure, Scale, and a Few More Fearless Hacks, A. Brown and G. Wilson, eds. (Brown & Wilson), pp. 291–315.
  62. Di Gregorio, F., and Varrazzo, D.; The Psycopg Team (2021). Psycopg2 – PostgreSQL database adapter for Python. <https://www.psycopg.org/docs/>.
  63. Häse, F., Aldeghi, M., Hickman, R.J., Roch, L.M., Christensen, M., Liles, E., Hein, J.E., and Aspuru-Guzik, A. (2021). Olympus: a benchmarking framework for noisy optimization and experiment planning. *Mach. Learn. Sci. Technol.* 2, 035021. <https://doi.org/10.1088/2632-2153/abedc8>.
  64. Pracht, P., Bohle, F., and Grimme, S. (2020). Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* 22, 7169–7192. <https://doi.org/10.1039/c9cp06869d>.
  65. Adamo, C., and Barone, V. (1999). Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* 110, 6158–6170. <https://doi.org/10.1063/1.478522>.
  66. Weigend, F., and Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 7, 3297–3305. <https://doi.org/10.1039/B508541A>.
  67. Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32, 1456–1465. <https://doi.org/10.1002/jcc.21759>.
  68. Huber, S.P. (2022). aiida-shell (École Polytechnique Fédérale de Lausanne). <https://github.com/sphuber/aiida-shell>.
  69. O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., and Hutchison, G.R. (2011). Open Babel: an open chemical toolbox. *J. Cheminform.* 3, 33. <https://doi.org/10.1186/1758-2946-3-33>.
  70. de Souza, B., Neese, F., and Izsák, R. (2018). On the theoretical prediction of fluorescence rates from first principles using the path integral approach. *J. Chem. Phys.* 148, 034104. <https://doi.org/10.1063/1.5010895>.