

Autonomous Chemical Experiments: Challenges and Perspectives on Establishing a Self-Driving Lab

Martin Seifrid,¹ Robert Pollice,¹ Andrés Aguilar-Granda,[†] Zamyla Morgan Chan,[†] Kazuhiro Hotta,[†] Cher Tian Ser,[†] Jenya Vestfrid,[†] Tony C. Wu,[†] and Alán Aspuru-Guzik*



Cite This: *Acc. Chem. Res.* 2022, 55, 2454–2466



Read Online

ACCESS |

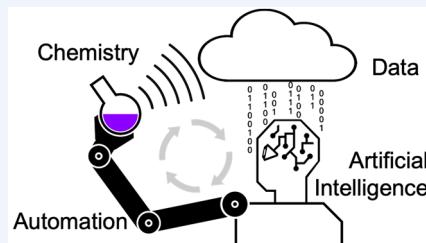
Metrics & More

Article Recommendations

CONSPPECTUS: We must accelerate the pace at which we make technological advancements to address climate change and disease risks worldwide. This swifter pace of discovery requires faster research and development cycles enabled by better integration between hypothesis generation, design, experimentation, and data analysis. Typical research cycles take months to years. However, data-driven automated laboratories, or self-driving laboratories, can significantly accelerate molecular and materials discovery. Recently, substantial advancements have been made in the areas of machine learning and optimization algorithms that have allowed researchers to extract valuable knowledge from multidimensional data sets. Machine learning models can be trained on large data sets from the literature or databases, but their performance can often be hampered by a lack of negative results or metadata. In contrast, data generated by self-driving laboratories can be information-rich, containing precise details of the experimental conditions and metadata. Consequently, much larger amounts of high-quality data are gathered in self-driving laboratories. When placed in open repositories, this data can be used by the research community to reproduce experiments, for more in-depth analysis, or as the basis for further investigation. Accordingly, high-quality open data sets will increase the accessibility and reproducibility of science, which is sorely needed.

In this Account, we describe our efforts to build a self-driving lab for the development of a new class of materials: organic semiconductor lasers (OSLs). Since they have only recently been demonstrated, little is known about the molecular and material design rules for thin-film, electrically-pumped OSL devices as compared to other technologies such as organic light-emitting diodes or organic photovoltaics. To realize high-performing OSL materials, we are developing a flexible system for automated synthesis via iterative Suzuki–Miyaura cross-coupling reactions. This automated synthesis platform is directly coupled to the analysis and purification capabilities. Subsequently, the molecules of interest can be transferred to an optical characterization setup. We are currently limited to optical measurements of the OSL molecules in solution. However, material properties are ultimately most important in the solid state (e.g., as a thin-film device). To that end and for a different scientific goal, we are developing a self-driving lab for inorganic thin-film materials focused on the oxygen evolution reaction.

While the future of self-driving laboratories is very promising, numerous challenges still need to be overcome. These challenges can be split into cognition and motor function. Generally, the cognitive challenges are related to optimization with constraints or unexpected outcomes for which general algorithmic solutions have yet to be developed. A more practical challenge that could be resolved in the near future is that of software control and integration because few instrument manufacturers design their products with self-driving laboratories in mind. Challenges in motor function are largely related to handling heterogeneous systems, such as dispensing solids or performing extractions. As a result, it is critical to understand that adapting experimental procedures that were designed for human experimenters is not as simple as transferring those same actions to an automated system, and there may be more efficient ways to achieve the same goal in an automated fashion. Accordingly, for self-driving laboratories, we need to carefully rethink the translation of manual experimental protocols.

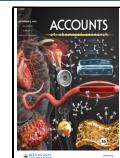


KEY REFERENCES

- Roch, L. M.; Häse, F.; Kreisbeck, C.; Tamayo-Mendoza, T.; Yunker, L. P. E.; Hein, J. E.; Aspuru-Guzik, A. ChemOS: An Orchestration Software to Democratize Autonomous Discovery. *PLoS One* **2020**, 15 (4), e0229862.¹ The “ChemOS” orchestration software for autonomous laboratories, featuring machine learning algorithms, online analysis, and interaction with researchers,

Received: April 11, 2022

Published: August 10, 2022



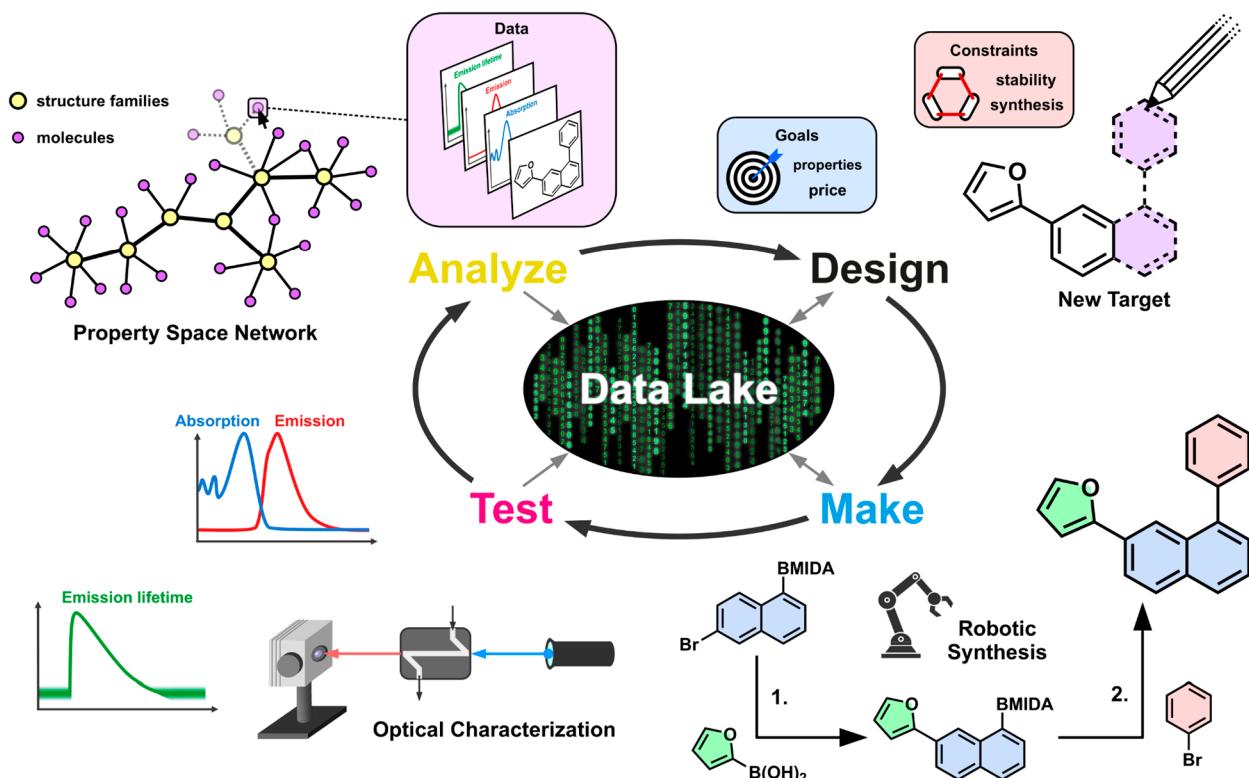


Figure 1. Diagram of the design–make–test–analyze cycle in our self-driving laboratory, showing the process for the development of new organic semiconductor laser materials.

automated instrumentation, and databases, is introduced and used to optimize color, robotic HPLC sampling and a cocktail recipe.

- Langner, S.; Häse, F.; Perea, J. D.; Stubhan, T.; Hauch, J.; Roch, L. M.; Heumueller, T.; Aspuru-Guzik, A.; Brabec, C. J. Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems. *Adv. Mater.* **2020**, *32* (14, 190780).² The “Phoenics” algorithm guides an automated thin-film fabrication and characterization platform to optimize the photostability of a multicomponent blend of organic photovoltaic materials.
- Christensen, M.; Yunker, L. P. E.; Adedeji, F.; Häse, F.; Roch, L. M.; Gensch, T.; dos Passos Gomes, G.; Zepel, T.; Sigman, M. S.; Aspuru-Guzik, A.; Hein, J. E. Data-Science Driven Autonomous Process Optimization. *Commun. Chem.* **2021**, *4* (1), 112.³ Bayesian optimization of discrete and continuous variables enables automated synthesis and characterization to find optimal ligands and conditions for a stereoselective Suzuki–Miyaura coupling.
- Seifrid, M.; Hickman, R. J.; Aguilar-Granda, A.; Lavigne, C.; Vestfrid, J.; Wu, T. C.; Gaudin, T.; Hopkins, E. J.; Aspuru-Guzik, A. Routescore: Punching the Ticket to More Efficient Materials Development. *ACS Cent. Sci.* **2022**, *8* (1), 122–131.⁴ Quantifying the costs of combined manual and automated synthetic routes enables the inverse design of materials that are cheaper to synthesize without sacrificing important properties.

1. INTRODUCTION

Recently, rapid progress in computer power and algorithmic efficiency has enabled the extensive computational exploration

of chemical space to design new materials. Yet, even the most rigorous simulations still cannot replace experimental data. Accordingly, synthesis is the bottleneck to progress in materials design, demanding a complete rethinking of conventional approaches. Hence, we believe that the implementation of self-driving laboratories in a closed-loop workflow is necessary to significantly speed up material design.

Self-driving laboratories unify artificial intelligence (AI) with automated robotic platforms to realize autonomous experimentation.^{5–9} The closed-loop workflow is generally subdivided into four steps: design, make, test, and analyze (DMTA; Figure 1). Autonomous experimentation is involved in both the make and test steps, but the interface to the design and analyze steps is also essential for autonomous experimentation. In this Account, we delineate our approach to implement a self-driving laboratory and discuss challenges and successes along the way.

2. WHY BUILD A SELF-DRIVING LAB?

Human and robotic strengths are roughly orthogonal. Actions that human researchers efficiently perform are difficult for robots and vice versa. For example, the actions of separating, dispensing solids, and extracting are relatively straightforward for humans but currently present significant obstacles for robotic systems. Notably, these processes all involve some combination of visual feedback, on-the-fly troubleshooting, and the possibility of results far outside the expected experimental parameters. In contrast, automated systems can efficiently handle well-behaved experiments within their design tolerances. These systems provide highly reproducible results and significantly higher throughput and collect large amounts of quantitative data. Mainly, they are well-suited to tasks where

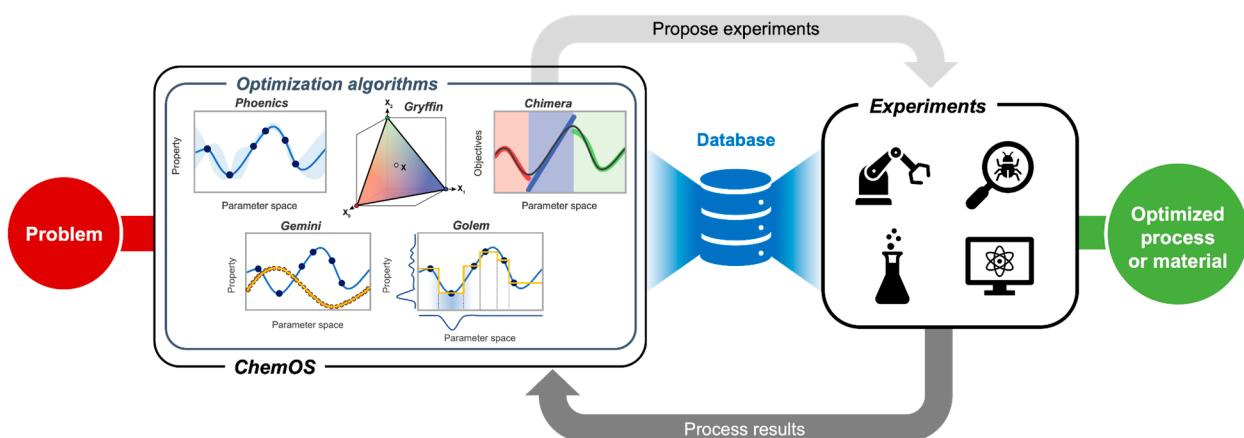


Figure 2. Integration of ChemOS and its most important algorithms into the process or material optimization workflow.

intensive physical properties characterize the entire system appropriately. While handling liquids tends to be easy to automate, the difficulty of dispensing powders depends on the total number of solid particles to be transferred. When that number is small, the local properties of each grain are important. However, when that number is large, the properties are more fluid-like, simplifying the process. When it comes to proposing new experiments, self-driving laboratories are better at learning from data due to their ability to handle high dimensionalities. However, human researchers can use their domain knowledge to transfer the results between distinct experiments and predict the outcomes of entirely new ones. Nevertheless, machine learning (ML) algorithms using transfer learning are steadily improving at imitating that ability.¹⁰

After the initial time and monetary investment, self-driving laboratories decrease the required human labor, freeing researchers up for higher-level scientific tasks such as formulating hypotheses, designing experimental campaigns, and interpreting data. Carefully designed autonomous experiments can also decrease material consumption and increase efficiency when paired with experiment planning algorithms by minimizing instrument downtime. Comprehensive online process control provides insight into the underlying mechanisms, enables more systematic optimizations, and allows the detection of potential hazards. Additionally, self-driving laboratories increase safety by handling hazardous materials with minimal human exposure.

Finally, self-driving laboratories promise both increased throughput and precision in the production of high-quality experimental data needed for data-driven approaches.¹¹ They can increase reproducibility as they eliminate human error and maintain better records of “failed” experiments. Particularly, the publication of full data sets from self-driving laboratories, as well as standardized data formats, have the potential to promote the publication of negative results in easily accessible databases. This is important because the current endemic publication bias is detrimental to ML approaches as “negative” results are as important as “positive” ones for training models.^{12,13} Standardized digital procedures promise more straightforward transferability between laboratories, enabling delocalized and reproducible science.^{14–16}

3. OUR APPROACH

3.1. Experiment Planning and Digital Infrastructure

Experiment planning is an important part of the design node in the DMTA cycle (Figure 1), which integrates experimental design, resource scheduling, and hardware control. Previous approaches to self-driving laboratories involved tailor-made software applicable only to one specific application.¹⁷ However, since many processes are common across laboratories, our group designed ChemOS as a versatile software package agnostic to the experimental environment, supporting both fully autonomous workflows and the active involvement of researchers.^{1,18} ChemOS is flexible because it is agnostic to the specific hardware being controlled. It performs the higher-level tasks of orchestrating experiment scheduling and selecting future experiments by ML on the basis of feedback from previous results. Recent efforts by other groups have taken the abstraction of orchestration and hardware control even further by controlling a self-driving lab through a hierarchical framework of web servers.¹⁹ ChemOS also allows one to connect to multiple database management systems to facilitate flexible data transfer and storage. Accordingly, data from past campaigns can be reused to guide future research. Furthermore, ChemOS allows for the remote control of hardware, facilitating the decentralization of equipment to generate geographically distributed meta laboratories. In these meta laboratories, parallelization of research campaigns can be carried out more efficiently and cross-disciplinary collaborations are facilitated, accelerating innovative research.

Another crucial ingredient of self-driving laboratories is the creation and maintenance of a standardized data framework that can inform algorithmic choices throughout the loop, which is the central hub for every part of the DMTA cycle. To this end, we created a NewSQL database, Molar,²⁰ which is already in operation in our lab and interfaced with ChemOS. To ensure that no data is lost, it implements event sourcing, which allows one to roll the database back to any point in time. To facilitate information sharing between groups, we developed a Python client with tight integration with the *pandas* library.^{21,22}

Our group has developed various experiment planning algorithms, some of which are already implemented in ChemOS (Figure 2). Phoenics is a Bayesian global optimization algorithm based on kernel density estimation that proposes new experimental conditions on the basis of

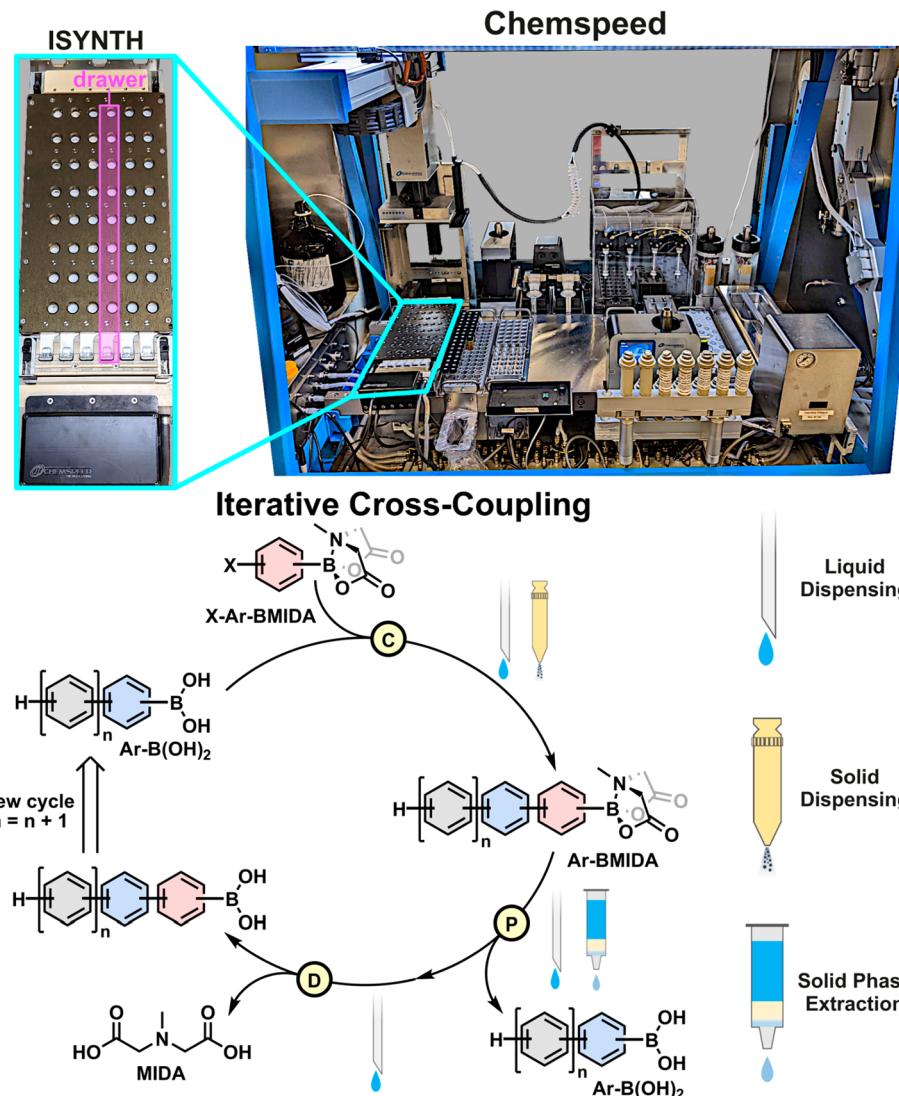


Figure 3. Top: Photo of the Chemspeed deck. The inset shows the top of the iSYNTH with one of the drawers (vertical row of wells) highlighted. Bottom: (Right) Icons for liquid dispensing, solid dispensing, and solid-phase extraction actions. (Left) Diagram of the iSMcc process along with icons indicating where different capabilities are used. Cross-coupling (C): X-Ar-BMIDA (1 equiv), Ar-B(OH)₂, 3 equiv, Pd-XPhos G2 (5 mol %), K₃PO₄ (2 equiv), THF, 16 h, 65 °C. Purification (P): precipitation from hexanes/THF 3:1. Deprotection (D): aqueous NaOH (1 M), 20 min, room temperature.

prior results, minimizing redundant evaluations.²³ Additionally, it can adopt strategies for both exploration and exploitation of the search space. Chimera is a flexible achievement scalarizing function for multiobjective optimization.²⁴ Multiobjective optimizations usually rely on handcrafted figures of merit, and it is challenging to design them without prior knowledge. Chimera relies on a user-specified hierarchy of the objectives together with their tolerances, allowing less important objectives to improve only if more important ones are not substantially degraded. Gryffin is a general-purpose Bayesian optimization framework for categorical variables.²⁵ Like Phoenics, it relies on kernel density estimation and uses smooth approximations to categorical distributions. Furthermore, Gryffin can use domain knowledge in the form of descriptors to approximate categorical variables in a continuous space. Gemini is a multifidelity ML algorithm that performs dynamic bias correction.²⁶ In the context of self-driving laboratories, it can be used to correct systematic biases of proxy experiments by learning from more expensive

measurements. Finally, Golem is an algorithm applicable to any experimental planning strategy that accounts for input uncertainties using probability distributions and locates optima that are robust to input variations arising from uncertainties in experimental conditions or instrument imprecision.²⁷

Our group has applied ChemOS and Phoenics to maximize the hole mobility of organic hole transport materials, which is critical for perovskite solar cell performance.²⁸ Their sensitivity to processing conditions and the large multidimensional search space make them difficult to optimize. Together with the Hein and Berlinguette groups,²⁸ we developed an autonomous setup capable of depositing, processing, and characterizing organic thin films by measuring pseudomobility as a surrogate for hole mobility. The setup varied the concentration (0–100%) of an FK 102 Co(III) TFSI salt (tris(2-(1*H*-pyrazol-1-yl)pyridine)-cobalt(III) tri[bis(trifluoromethane)sulfonimide]) and the annealing time (0–240 s) of the film in a forced convection annealing furnace. In less than 30 h each, two independent optimization campaigns converged on the same global

maximum with low dopant concentration and annealing times, demonstrating the reproducibility of our approach.

Additionally, our group along with the Brabec group² has used ChemOS and Phoenics to explore the formulation of quaternary blends to improve organic photovoltaic (OPV) stability. As the search space is inherently high-dimensional, we set up a parallel robotic workflow for the production of OPV films via drop-casting. Compared to a grid search, ChemOS located the most stable composition after only evaluating 7% of the space, highlighting the efficiency of Bayesian optimization.

Recently, ChemOS was employed to optimize the reaction conditions of a stereoselective Suzuki–Miyaura cross-coupling of vinyl (pseudo)halides.³ Undesirable E to Z inversion of the double bond depends on the phosphine ligand. Together with colleagues at Merck and in the Hein and Sigman groups,³ we used ChemOS to plan experiments on a Chemspeed SWING automated synthesis platform (ASP) coupled to an HPLC-UV system at Merck. Active learning was carried out with the experiments performed in batches of 8, where the conditions for each subsequent batch were determined by Phoenics and Griffin on the basis of the results of HPLC analysis. This setup facilitated a parallelized 192-reaction campaign carried out in 4 days, a 4-fold increase in throughput compared to a typical sequential approach. Using systematic data science tools and both quantum mechanical and geometrical ligand descriptors to facilitate ligand optimization, selectivity for the E product was greatly improved.

3.2. Synthesis

In the make node of the DMTA cycle (Figure 1), we are mainly interested in automating manual batch syntheses, i.e., the majority of synthetic procedures reported in the literature. This is important as the translation to batch protocols to flow and vice versa can be difficult.^{29,30} Although batch reactors allow straightforward handling of heterogeneous mixtures, significant progress in that regard has also been achieved using oscillatory flow reactors.³¹ Additionally, batch reactions are easily performed in parallel, which is important for the increase of throughput and the acceleration of discovery. We focus primarily on the development of general batch methods on the micromolar scale, allowing both the discovery of novel molecules and the optimization for specific targets with minimal procedural changes.

We are implementing a mix of turn-key and home-built systems in our ASP. Turn-key systems promise to be easier to set up but are often much more expensive, which restricts their use to laboratories with sufficient financial resources.³² Home-built systems, which require much more initial time investment, can be tuned to the specific needs of each lab. Our ASP consists of a glovebox with two linked chambers containing a Chemspeed ISYNTH CATSCREEN Robotic Workflow Platform (Figure 3), which we will refer to simply as “the Chemspeed”. It can dispense solids and liquids and cap and uncap vials and provides various ways to control temperature and pressure. The ISYNTH reactor accommodates up to 48, 8 mL reaction vials and has the following capabilities: vortex stirring, temperature control, reflux, vacuum, and inertization. Notably, the glovebox allows us to perform air- and moisture-sensitive reactions. We designed this ASP for high versatility. Thus, it offers a wide range of methods, providing access to a large chemical space. However, this type of general system potentially suffers from lower performance on tasks for which it is not optimized. In addition, not all components can be used

simultaneously. In contrast, dedicated systems can be highly optimized for one specific task, leading to better performance, reproducibility, and robustness. The Chemspeed and our high-performance liquid chromatography-mass spectrometer (HPLC-MS) are primarily turn-key systems for which we have written custom control software. However, the optical characterization setup, which is described in detail below, is a home-built system that has been designed for the specific tasks described here.

Our first goal was to implement a well-established reaction with a large substrate scope. We selected the iterative Suzuki–Miyaura cross-coupling (iSMcc) strategy on the basis of *N*-methyliminodiacetic acid (MIDA)-protected boronates developed by Gillis and Burke (Figure 3).³³ Previously, the Burke group demonstrated the automated synthesis of macro- and polycyclic materials rich in sp³-carbons as well as pharmaceuticals.³⁴ This is achieved by iteratively assembling bifunctional MIDA-protected building blocks using iSMcc reactions to form carbon–carbon bonds. Many organic electronic materials can be synthesized in a similar manner: by assembling (hetero)aromatic building blocks from prefunctionalized starting materials via cross-coupling chemistry.³⁵ Accordingly, we decided to target the autonomous synthesis of organic laser molecules via iSMcc.³⁶ We recently reported the results of an initial screening campaign³⁷ and are actively pursuing closed-loop optimization of these materials.

The iSMcc method comprises three automated steps: deprotection of the MIDA boronate (BMIDA) to yield a reactive boronic acid (BA), cross-coupling with the halide of a building block containing both BMIDA and halide functional groups, and rapid purification of the product. The resulting BMIDA-containing product can be used as a substrate for a subsequent coupling cycle. A wide variety of halogenated and BA-containing starting materials are commercially available with bifunctional halo-BMIDA starting materials becoming more accessible. For instance, employing strict selection rules regarding price, aromatic (hetero)cycle identity, and functional groups, we picked 116 purchasable substrates as building blocks (47 MIDA boronates, 22 difunctionalized substrates, and 47 halides) for organic laser molecules. Using only four coupling cycles, we will be able to access approximately one million distinct organic laser molecules in a fully automated fashion.

We have adapted each step of the iterative coupling sequence to be performed in our ASP. The gravimetric dispensing unit dispenses solids from dedicated cartridges into the ISYNTH, and the four-needle head tool transfers solvents and solutions to the ISYNTH from dedicated solvent bottles or from vials on the Chemspeed deck. The solutions are sealed by closing the ISYNTH drawers, then heated, and stirred to start the reactions. The atmosphere within the vials is regulated by a valve that exposes them to nitrogen or vacuum and by the position of the drawers, which can be adjusted to simultaneously control the state of the eight vials in each column, allowing for them to be open to the ambient atmosphere, to be fully isolated, or have a vacuum or inert atmosphere applied. The temperature within the ISYNTH is controlled by a circulation thermostat, which pumps oil through the ISYNTH reactor block. After the reactions are completed, the solvents are evaporated. In the purification step, the product is isolated using the “catch-and-release” solid phase extraction (SPE) method.³⁴ The products containing a BMIDA group (Figure 3) are precipitated from a mixture of

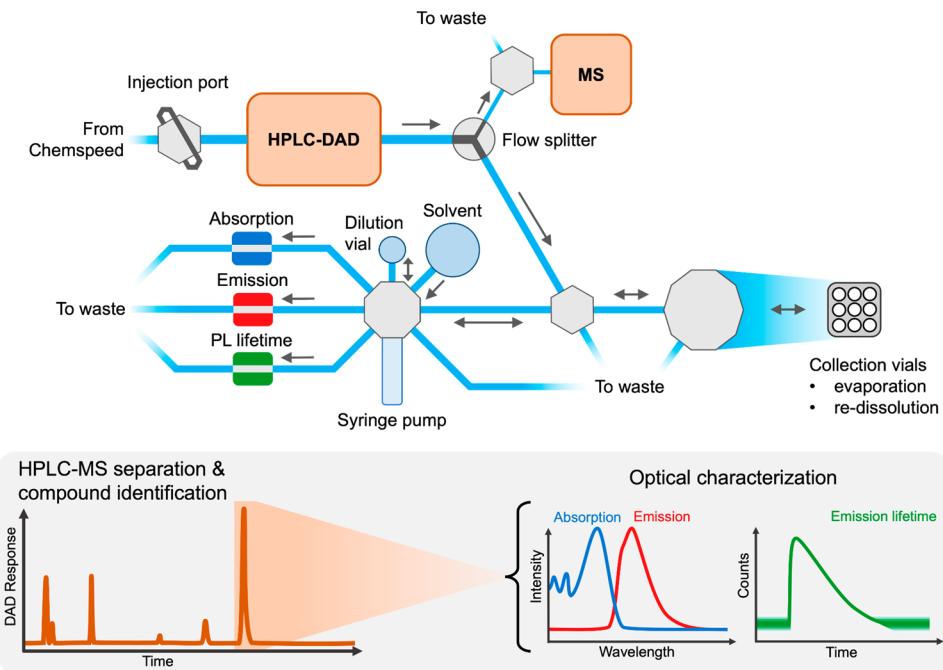


Figure 4. Schematic diagram of our analysis, purification, and optical characterization setup. The gray box is a schematic diagram of how a specific HPLC fraction is selected for further evaluation and how its properties are measured. Absorption measurements are carried out in the “absorption” flow cell. Photoluminescence (PL), PL quantum yield (PLQY), and photodegradation rate measurements are carried out in the “emission” flow cell. PL lifetime is measured in the “PL lifetime” flow cell. Gray polygons represent valves with the number of ports corresponding to the number of sides, and arrows represent the directions of sample transport.

solvents on top of a short silica column. The desired product is released by elution via THF and collected. Subsequently, the solvent is evaporated in the ISYNTH. Finally, in the deprotection step, the newly synthesized BMIDA-containing product is redissolved in THF and treated with a strong aqueous base (1 M NaOH), yielding the unprotected BA for the next cycle.

3.3. Analysis and Purification

After a batch of reactions is completed, samples from each reaction mixture are diluted with solutions containing an internal standard and injected into an in-line HPLC-MS system equipped with an additional diode array detector (DAD), as depicted in Figure 4. The substrates and product peaks are located by monitoring the selected ion chromatogram at the m/z value of the expected ionized compounds. Isotope pattern matching is used to confirm their chemical identity.³⁸ The components of the mixture are quantified using the 3D chromatogram of retention time, wavelength, and absorbance intensity. This 3D chromatogram is collapsed into 2D (i.e., retention time and absorbance intensity) by integration over the spectral range of interest. The resulting chromatogram is used to determine the retention time of the molecules and approximate their concentration.

In “discovery mode”, it is impractical to determine calibration curves for every possible product in order to estimate its concentration. To address this challenge, we are developing a method to automatically quantify peaks by their maximum absorbance wavelength. The absorbance spectrum of particular peaks would be extracted to determine the corresponding maximum absorbance wavelength. Next, the maximum absorbance wavelength would be used to estimate peak area, height, and other peak parameters. This approach avoids the pitfall of quantification using a single wavelength

that can readily result in misestimation of the concentration of various components based on differences between their respective absorption coefficients at the chosen wavelength. Nevertheless, it is challenging to determine the true concentration of the components, which have a wide range of possible absorption coefficients at their respective maximal absorbance wavelengths, since we cannot make reference samples in “discovery” mode. However, this method has the potential to improve our accuracy and be more robust since our quantification will be based on the compound’s maximal absorption coefficient, which may minimize errors in the response compared to performing the same analysis at an arbitrary wavelength. Additionally, we are exploring this method because we use the maximal absorption wavelength to adjust concentrations for subsequent optical characterization, as described below. Because the response is directly related to the signal, the adjustment of the concentration based on absorbance at the maximal absorbance wavelength makes the most sense.

In principle, it is possible to selectively isolate and transfer the desired compounds to the in-line characterization devices in real time. However, for the time being, we use separate injections because in “discovery mode” we are examining numerous reactions for new compounds with varying reaction yields. In addition, the properties of the target compounds such as the absorption coefficient are unknown in advance and vary widely. Hence, it is better to adjust the quantity of sample transferred to the characterization devices on the basis of the absorption data obtained from the first injection.

3.4. Characterization

Once the synthesized compounds have been isolated, they are transferred to several measurement flow cells for characterization (Figure 4). This configuration allows for greater

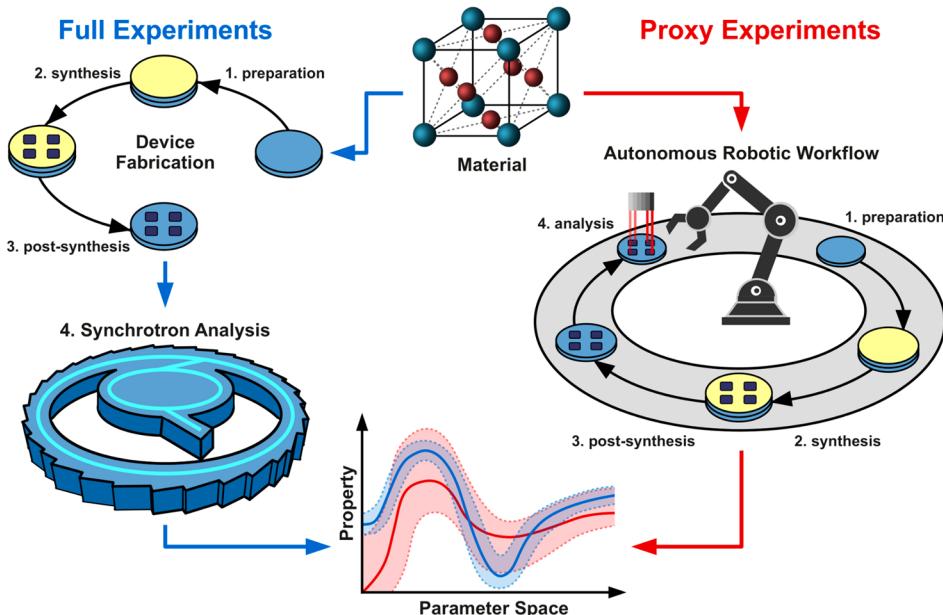


Figure 5. Autonomous robotic workflows can accelerate the discovery of solid-state inorganic materials using proxy experiments. These can then be used in conjunction with more accurate full experiments to perform multifidelity optimization of the inorganic materials.

flexibility since the sample can be optimized for each measurement by adjusting the concentration, mixing with other reagents, or even changing the solvent via evaporation and redissolution.

Automated experiments where the instruments are linked by stepwise robotic sample transport^{28,39} require less process optimization for each measurement as the intermediates can simply be purified, and samples can be prepared between successive experiments. In contrast, in-line processes must be optimized so that the initial sample is suitable for all subsequent experiments. Nevertheless, properly optimized in-line processes have the potential to lead to substantial time savings. The advantage of in-line “discovery mode” synthesis is that only very small amounts of material are required for full characterization, which makes it possible to run reactions on a much smaller scale and results in more efficient utilization of both money and materials. However, an important drawback of the in-line approach is that it is limited mainly to characterizing samples dissolved in a single solvent. To address this challenge, we have built a “quasi-in-line” setup, where samples can be dried and redissolved with a different solvent in the collection vials. This can be done after the samples are collected from the HPLC or between multiple rounds of optical characterization in order to measure a sample’s properties in multiple solvents.

We have built custom in-line instruments to measure optical properties such as UV–vis absorbance and photoluminescence (PL) spectra, relative PL quantum yield (PLQY),⁴⁰ photo-degradation rate, and PL lifetime (Figure 4). For the absorption and PL instruments, a charge-coupled device (CCD) spectrometer is coupled to the cell holders for each measurement via a Y-junction optical fiber. In order to estimate the relative PLQY by measuring the absorbed power, the PL instrument has also been equipped with a power meter. In principle, both absorbance and PL spectra can be obtained from the same flow cell. However, we have separated them to maximize the signal-to-noise response for minimal analyte amounts by optimizing the respective cell designs. To measure the PL lifetime of a molecule, the time-resolved PL is recorded

using time-correlated single-photon counting (TCSPC)⁴¹ with a picosecond pulse laser. The instruments described above are controlled by modules written in Python. All of the measurement parameters such as exposure time for the PL spectrum and excitation intensity and frequency for the PL lifetime measurements can be automatically adapted to the sample response.

3.5. Solid-State Materials

The experimental methods described in the previous sections have focused on synthesis, analysis, and characterization in solution. However, the structure–property relationships of solid-state materials are central to the performance of virtually all materials.⁴² To address the gap in both data and knowledge of the inorganic structure–property realm,⁴³ we are building an autonomous robotic workflow for inorganic thin-film materials. The system is designed to be highly modular, accounting for a wide range of materials synthesis and characterization. However, given the vast realm of inorganic chemistry, we have chosen to focus on thin-film catalysts for the oxygen evolution reaction (OER)⁴⁴ with planned expansions to CO₂ reduction and hydrogen evolution.

The various characterization tools are a combination of custom-built and turn-key instrumentation orchestrated by ChemOS. Central to this system is a multichannel potentiostat with custom-built electrochemical three-electrode cells. This platform is facilitated by a robotic arm that can manipulate and access all of the stations within the workflow (Figure 5). The arm may conduct an on-the-fly transfer of samples between various characterization systems in a custom order. For example, as electrochemical characterization is a batch process, a cleaning station is incorporated to clean the corresponding cells. While a cell is being cleaned, the working electrode can be transferred to a separate cell for further electrochemical characterization or to a different instrument. In conjunction with the development of ML algorithms for constrained optimization, this has the potential to minimize instrument downtime by assigning an optimal order of the experiments.

These algorithms, which are currently in development in our group, account for time constraints as well as the destructive or nondestructive nature of the respective characterization technique.

One of the key challenges to automated experimentation with solid-state materials is the use of proxy experiments.²⁸ There are many experiments currently performed by human experimenters that are impractical to incorporate into a standalone self-driving lab due to their cost, size, or complexity. For example, alternatives must be found for experiments that rely on synchrotron radiation or device fabrication. It can be difficult to design meaningful proxy experiments, or the proxy experiments, if possible, may not adequately reflect real performance. As a result, researchers must perform a cost-benefit analysis of the integration of proxy experiments into their self-driving lab. ML algorithms such as Gemini²⁶ that learn from both proxy experiments and full experiments coupled with smart autonomous workflows can maximize the accuracy and efficiency of material characterization workflows to save time and cost.

4. CHALLENGES AND LESSONS LEARNED

4.1. Replacing Cognitive Processes

The challenges of building a self-driving lab can generally be broken down into two broad categories: cognitive processes and motor function. Namely, the challenge of replacing human cognitive processes comes from bringing ML algorithms into the “real world” of chemistry, encountering unexpected or difficult to predict results, and trying to automate instruments that are designed to be used by humans. A salient example of this challenge can be taken from the area of molecular design by AI-guided synthetic route planning for which a central consideration is what design spaces are accessible. While a small set of couplings under a limited range of conditions can be useful for specific applications, it is a much greater challenge to build a self-driving lab that can make molecules of arbitrary structure. However, this outstanding challenge is too complex to be addressed here and continues to see significant ongoing developments that lead to expanded capabilities.^{4,9,15,45–48} In summary, to run smoothly, a self-driving lab must be controlled by robust algorithms and code. Human cognitive capabilities can usually handle such tasks with ease; however, their automation can be very difficult.

Most optimization algorithms assume the absence of holes and that all constraints are known.⁴⁹ However, real data has unforeseen holes as specific points can be challenging or even impossible to explore. Common reasons include reaction scope limitations, measurement errors, simulation failures, or even the inability to purchase or synthesize chemicals. Additionally, known inequality constraints may be present, either due to physical constraints, e.g., mixing liquids without exceeding the container volume, or scientific considerations, e.g., testing molecule mixtures while keeping the total concentration constant. Unfortunately, generally effective solutions for arbitrary constraints are still unavailable. We have recently extended our Phoenics and Gryffin algorithms to be optimized on noncompact domains that result from interdependent, nonlinear constraints.⁵⁰ However, Bayesian optimization with unknown constraints remains a challenge.

Difficulties with unexpected or difficult to predict results persist at the stage of analysis and purification as well: automated identification of unknown compounds is challeng-

ing.^{51,52} Forward reaction prediction combined with searching for the expected products facilitates identification. However, it is also common for unexpected or unknown side products to form. The identification of them and determination of their structure automatically is incredibly challenging. Although efforts have been made to automate the prediction of side products,^{53–57} this is an area where autonomous laboratories, and indeed humans, still struggle. Furthermore, the susceptibility to ionization and manner in which molecules ionize is not always predictable *a priori*. The recognition of molecular ions in MS is mostly straightforward to automate; however, it can be challenging to find general analysis conditions. State-of-the-art MS instruments can rapidly switch between positive and negative ionization, expanding the scope of detectable ions. While tandem MS provides structural information, fragmentation is difficult to predict,^{58–60} and few databases exist for fragmentation patterns.^{61–63} In principle, benchtop flow nuclear magnetic resonance (NMR) can provide additional structural information as well as purity and yield estimates.^{64–67} However, its limited resolution and sensitivity currently hamper general applicability.

Finally, few manufacturers develop their software to consider self-driving laboratories. Therefore, sometimes, there is no commercial equipment with an API for comprehensive external control, and programming support is hardly offered. Consequently, the adaptation of an instrument for use in self-driving laboratories often requires a significant time investment to write custom code.³² “Hacking” instruments without a sufficient API introduces potential points of failure. Nevertheless, we believe this will change in the future as more laboratories move to automation, but for the moment, vigilance is required when choosing equipment.

4.2. Replacing or Replicating Motor Function

The other major type of challenge is replicating or replacing certain actions that are easy for humans because of their fine motor skills and hand-eye coordination. For example, troubleshooting the electrochemical experiments described above is more suited to human actions because it requires dexterity and coordination between visual inputs and motor response. Electrolyte salts may wick up the working electrode during bulk electrolysis, causing interference in the current response. A human researcher can wipe the electrode gently without interrupting the experiment, but current automated systems cannot. Rather than adapt automated procedures, the three-electrode electrochemical cell must be redesigned to suit automated functionality, as it can increase precision and reproducibility.

This is especially true because ASPs struggle to replicate some of these tasks, particularly accurately dispensing small amounts of solids. Dispensing solids automatically is a well-known challenge, especially for solids with very different properties and for amounts less than approximately 20 mg.^{68,69} The achievement of both accuracy and precision requires substantial calibration and testing. For “discovery mode” synthesis, such an investment is not always feasible as hundreds of solids can require customized settings. Advances are ongoing; however, there is considerable room for improvement.⁷⁰ Accordingly, many self-driving laboratories circumvent that by relying on stock solutions and well-established liquid-handling technologies, allowing the precise down-scaling of procedures. Additionally, the preparation of stock solutions can also be automated with the help of computer vision.⁷¹

Another major challenge is handling heterogeneous mixtures, especially for purification. Heterogeneous mixtures present a significant challenge to systems designed around liquid transfer via pumps due to their risk of being damaged or malfunctioning. The avoidance of critical failure requires careful experimental design and continuous process monitoring. Furthermore, heterogeneous mixtures can complicate solution-state characterization as homogeneity is typically assumed. For purification in our ASP, we use disposable cartridges to perform solid-phase extraction (SPE). However, SPE is prone to blockages and leaks, potentially requiring human intervention. Additionally, liquid–liquid extraction is particularly cumbersome without phase boundary detection. To minimize product loss, we perform additional extraction cycles, increasing solvent consumption and the duration of the experiments. Furthermore, the subsequent precipitation of solid can cause clogging, requiring manual intervention. Finally, when synthesizing many compounds, the development of robust purification protocols suitable for all products is challenging, and it is likely necessary to have multiple alternative protocols.

More generally, the automation of every reaction may be unfeasible, or even impossible, due to instrument limitations. Hence, we have developed an approach to quantify the costs of combined manual and automated synthesis routes.⁴ Human chemists would carry out steps that are impossible for the ASP while taking advantage of its higher throughput and lower cost to make molecules that would otherwise be unattainable to the ASP alone. On the path to self-driving laboratories, this combined approach is analogous to level two or three self-driving cars, which can perform some complex tasks but still require human intervention.⁷²

5. FUTURE PERSPECTIVES

In this Account, we have described our progress toward the realization of a self-driving lab in our group. Starting at the digital infrastructure, we have developed ChemOS as a versatile and robust software for comprehensive campaign orchestration and experiment planning, making use of state-of-the-art Bayesian optimization algorithms for sample-efficient operation. At the heart of our robotic workflow, we have incorporated the Chemspeed as a robotic synthesis engine performing iterative Suzuki–Miyaura cross-couplings to access a wide range of organic electronic materials. This platform is directly connected to purification and characterization via HPLC-MS and our custom-built optical characterization setup. For inorganic solid-state materials, we have developed an automated multichannel potentiostat for electrochemical characterization facilitated by a robotic arm and are currently incorporating it into an autonomous robotic workflow for thin-film materials.

While we have successfully implemented self-driving components, we have not arrived at a fully autonomous system yet. Accordingly, we are pursuing multiple avenues to extend our self-driving laboratory, some of which will be outlined in this section. The chemical space we can access is limited by the implemented synthetic protocols. Currently, we rely solely on iSMcc reactions. To increase our scope, we will implement other Pd-catalyzed cross-couplings, such as Buchwald–Hartwig aminations,⁷³ Sonogashira couplings,⁷⁴ and the Cu-catalyzed azide–alkyne Huisgen cycloaddition,⁷⁵ and potentially other reactions amenable to iterative synthetic schemes.⁷⁶ To broaden the repertoire of chemical reactions,

the enhancement of the capabilities of our system and the development of new procedures will be necessary. Some steps still require human intervention such as loading new vials or chemicals into the Chemspeed. To automate them, we are implementing robotic arms directly into our workflow that imitate human actions. We are also adding an online process control via computer vision, which will make the system more robust and versatile.⁷⁷ The idea is to monitor every operation and detect experimental failure early on. Moreover, we are implementing the Chemical Description Language, XDL, as a standardized synthesis specification to make our experimental procedures readily transferable to other laboratories and simplify the adoption of external protocols.^{15,78,79} To increase throughput, we want to perform product identification and collection with a single HPLC injection, which requires communication between multiple devices in real time. Lastly, we want to connect the solution and solid-state capabilities that we have described above. At the moment, there is no direct link between solution phase synthesis and solid-state characterization, and much work remains to be done to connect these two areas of focus in self-driving laboratories. In that regard, robotic arms seem to be the most promising solution to connect these workflows.

Finally, automation is expensive with the short-term implication that it will be limited to groups and countries with sufficient budgets. One solution is to build user facilities targeting a large number of researchers, providing both training and research opportunities. Prospective users could submit research proposals similar to how scientists can apply for beamtime at synchrotrons or propose experiments to be conducted at CERN. Additionally, these facilities could be decentralized but connected to form networks of self-driving laboratories, each with different capabilities, forming the meta laboratories of the future. To achieve that, academic–industry partnerships will play an important role in making self-driving laboratories more accessible. Recently, the Acceleration Consortium has been established at the University of Toronto.⁸⁰ Its purpose is to drive innovation in self-driving laboratories and offer training and experience to researchers in academia, industry, and government. We believe that the Acceleration Consortium and similar initiatives around the world will drive the transformation of scientific laboratories into the self-driving laboratories of the future.

■ AUTHOR INFORMATION

Corresponding Author

Alán Aspuru-Guzik – Department of Chemistry and Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3H6, Canada; Department of Chemical Engineering & Applied Chemistry, University of Toronto, Toronto, Ontario M5S 3E5, Canada; Department of Materials Science, University of Toronto, Toronto, Ontario M5S 3E4, Canada; Vector Institute for Artificial Intelligence, Toronto, Ontario M5S 1M1, Canada; Lebovic Fellow, Canadian Institute for Advanced Research, Toronto, Ontario M5S 1M1, Canada;  orcid.org/0000-0002-8277-4434; Email: alan@aspuru.com

Authors

Martin Seifrid – Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada;  orcid.org/0000-0001-5238-0058

Robert Pollice – Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada;
ORCID: [0000-0001-8836-6266](https://orcid.org/0000-0001-8836-6266)

Andrés Aguilar-Granda – Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada; Present Address: Facultad de Química, Universidad Nacional Autónoma de México, 04510 Ciudad de México, Mexico

Zamyla Morgan Chan – Department of Chemistry and Acceleration Consortium, University of Toronto, Toronto, Ontario M5S 3H6, Canada

Kazuhiko Hotta – Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada; Science & Innovation Center, Mitsubishi Chemical Corporation, Yokohama, Kanagawa 227-8502, Japan

Cher Tian Ser – Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada

Jenya Vestfrid – Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada

Tony C. Wu – Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.accounts.2c00220>

Author Contributions

[†]A. Aguilar-Granda, Z.M.C., K.H., C.T.S., J.V., and T.C.W. contributed equally to this work. M.S. conceived the general outline and structure of this manuscript, and all authors contributed toward refining the structure. The manuscript was written through contributions of all authors. All authors have approved the final version of the manuscript.

Notes

The authors declare the following competing financial interest(s): A. Aspuru-Guzik is the co-founder and Chief Visionary Officer of Kebotix Inc.

Biographies

Martin Seifrid received a Ph.D. in Chemistry from the University of California, Santa Barbara, where he studied the relationship between molecular structure, processing, and solid-state structure in organic semiconducting materials under the supervision of Professor Guillermo Bazan. He is currently a postdoctoral fellow in the group of Professor Alán Aspuru-Guzik at the University of Toronto, where he is building a self-driving lab for the design of organic semiconductor laser materials.

Robert Pollice received his Ph.D. in Chemistry in 2019 under the supervision of Professor Peter Chen at ETH Zurich, where he investigated London dispersion in molecular systems. Subsequently, he joined the group of Professor Alán Aspuru-Guzik as an SNSF postdoctoral fellow at the University of Toronto to work on the inverse design of organic electronic materials and molecular catalysts.

Andrés Aguilar-Granda studied Industrial Chemistry at the University of Veracruz and received his Ph.D. from the Institute of Chemistry, Universidad Nacional Autónoma de México (UNAM). Afterward, he completed a postdoctoral fellowship at the University of Toronto (in the group of Prof. Alán Aspuru-Guzik). In April 2021, he began his independent career as an associate professor in the Department of Organic Chemistry at the School of Chemistry at UNAM. His research interests are in the areas of automated organic synthesis for functional materials and the digitization of organic chemistry.

Zamyla Morgan Chan is the Associate Director of the Acceleration Consortium at the University of Toronto. She received a Ph.D. in Chemistry from Harvard University and joined the Vector Institute for Artificial Intelligence as a Postdoctoral Research Fellow. Her research seeks to facilitate accelerated discovery of scalable and stable materials by leveraging robotics, machine learning, and fundamental science for inverse design.

Kazuhiko Hotta received a Ph.D. in Chemistry from Tohoku University, where he studied optical biosensors based on nanoporous materials as a sensing platform. He then joined Mitsubishi Chemical Corporation where he is currently a senior scientist working on laboratory automation and the development of a self-driving laboratory.

Cher Tian Ser is a Ph.D. student at the University of Toronto supervised by Prof. Alán Aspuru-Guzik. His research interests involve the application of machine learning methods for the discovery of catalytic and energy materials.

Jenya Vestfrid received her Ph.D. in Chemistry from Technion – Israel Institute of Technology. Afterward, she completed postdoctoral fellowships in the departments of Chemical Engineering & Applied Chemistry and Chemistry at the University of Toronto. Currently, she is an experienced researcher at StoreDot, a motor vehicle parts manufacturing company, developing extreme fast charging batteries for electric vehicles.

Tony C. Wu is a postdoctoral fellow at the University of Toronto and the Vector Institute. With a cross-disciplinary background, his current research passion is in the development of autonomous chemistry and machine learning models for accelerated materials development. He received his B.Sc. in both Electrical Engineering and Physics from the National Taiwan University in 2011 and his M.A. and Ph.D. in Electrical Engineering from the Massachusetts Institute of Technology in 2018. During his Ph.D., his research focused on optoelectronics and excitonics engineering with applications in OLEDs and organic solar cells.

Alán Aspuru-Guzik is a Professor of Chemistry and Computer Science at the University of Toronto, a Canada 150 Research Chair in Theoretical Chemistry, a Canada CIFAR AI Chair at the Vector Institute, a CIFAR Lebovic Fellow in the Biologically Inspired Solar Energy program, and a Google Industrial Research Chair in Quantum Computing. He received a Ph.D. in Chemistry from the University of California, Berkeley, where he was also a postdoctoral fellow. He began his independent career at Harvard University and became a Full Professor before moving to the University of Toronto. His research interests span chemistry, automation, machine learning, and quantum information.

ACKNOWLEDGMENTS

We thank all our co-workers and collaborators who contributed to the projects highlighted in this Account. R.P. acknowledges funding through a Postdoc.Mobility fellowship by the Swiss National Science Foundation (SNSF; Project No. 191127). T.C.W. acknowledges funding through University of Toronto Arts & Science Postdoctoral Fellowship. We acknowledge the Defense Advanced Research Projects Agency (DARPA) under the Accelerated Molecular Discovery Program under Cooperative Agreement No. HR00111920027 dated August 1, 2019. The content of the information presented in this work does not necessarily reflect the position or the policy of the Government. We also acknowledge funding from the National Research Council Canada (MCF-106) as part of the Materials for Clean Fuels

Challenge program. A. Aspuru-Guzik thanks Anders G. Frøseth for his generous support. A. Aspuru-Guzik also acknowledges the generous support of Natural Resources Canada and the Canada 150 Research Chairs program. We also acknowledge the Department of Navy awards (N00014-19-1-2134 and N00014-21-1-2137) issued by the Office of Naval Research. The United States Government has a royalty-free license throughout the world in all copyrightable material contained herein. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Office of Naval Research.

■ REFERENCES

- (1) Roch, L. M.; Häse, F.; Kreisbeck, C.; Tamayo-Mendoza, T.; Yunker, L. P. E.; Hein, J. E.; Aspuru-Guzik, A. ChemOS: An Orchestration Software to Democratize Autonomous Discovery. *PLoS One* **2020**, *15*, e0229862.
- (2) Langner, S.; Häse, F.; Perea, J. D.; Stubhan, T.; Hauch, J.; Roch, L. M.; Heumueller, T.; Aspuru-Guzik, A.; Brabec, C. J. Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems. *Adv. Mater.* **2020**, *32*, 1907801.
- (3) Christensen, M.; Yunker, L. P. E.; Adedeji, F.; Häse, F.; Roch, L. M.; Gensch, T.; dos Passos Gomes, G.; Zepel, T.; Sigman, M. S.; Aspuru-Guzik, A.; Hein, J. E. Data-Science Driven Autonomous Process Optimization. *Commun. Chem.* **2021**, *4*, 112.
- (4) Seifrid, M.; Hickman, R. J.; Aguilar-Granda, A.; Lavigne, C.; Vestfrid, J.; Wu, T. C.; Gaudin, T.; Hopkins, E. J.; Aspuru-Guzik, A. Routescore: Punching the Ticket to More Efficient Materials Development. *ACS Cent. Sci.* **2022**, *8*, 122–131.
- (5) Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Next-Generation Experimentation with Self-Driving Laboratories. *Trends Chem.* **2019**, *1*, 282–291.
- (6) MacLeod, B. P.; Parlane, F. G. L.; Rupnow, C. C.; Dettelbach, K. E.; Elliott, M. S.; Morrissey, T. D.; Haley, T. H.; Proskurin, O.; Rooney, M. B.; Taherimakhsousi, N.; Dvorak, D. J.; Chiu, H. N.; Waizenegger, C. E. B.; Ocean, K.; Mokhtari, M.; Berlinguette, C. P. A Self-Driving Laboratory Advances the Pareto Front for Material Properties. *Nat. Commun.* **2022**, *13*, 995.
- (7) Rooney, M. B.; MacLeod, B. P.; Oldford, R.; Thompson, Z. J.; White, K. L.; Tungjuyatham, J.; Stankiewicz, B. J.; Berlinguette, C. P. A Self-Driving Laboratory Designed to Accelerate the Discovery of Adhesive Materials. *Digit. Discovery* **2022**, DOI: [10.1039/D2DD00029F](https://doi.org/10.1039/D2DD00029F).
- (8) Tao, H.; Wu, T.; Kheiri, S.; Aldeghi, M.; Aspuru-Guzik, A.; Kumacheva, E. Self-Driving Platform for Metal Nanoparticle Synthesis: Combining Microfluidics and Machine Learning. *Adv. Funct. Mater.* **2021**, *31*, 2106725.
- (9) Gao, W.; Raghavan, P.; Coley, C. W. Autonomous Platforms for Data-Driven Organic Synthesis. *Nat. Commun.* **2022**, *13*, 1075.
- (10) Gupta, A.; Ong, Y.; Feng, L. Insights on Transfer Optimization: Because Experience Is the Best Teacher. *IEEE Trans. Emerg. Top. Comput. Intell.* **2018**, *2*, 51–64.
- (11) Shi, Y.; Prieto, P. L.; Zepel, T.; Grunert, S.; Hein, J. E. Automated Experimentation Powers Data Science in Chemistry. *Acc. Chem. Res.* **2021**, *54*, 546–555.
- (12) Strieth-Kalthoff, F.; Sandfort, F.; Kühnemund, M.; Schäfer, F. R.; Kuchen, H.; Glorius, F. Machine Learning for Chemical Reactivity: The Importance of Failed Experiments. *Angew. Chem., Int. Ed.* **2022**, *61*, e202204647.
- (13) Beker, W.; Roszak, R.; Wolos, A.; Angello, N. H.; Rathore, V.; Burke, M. D.; Grzybowski, B. A. Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki–Miyaura Coupling. *J. Am. Chem. Soc.* **2022**, *144*, 4819–4827.
- (14) Wilbraham, L.; Mehr, S. H. M.; Cronin, L. Digitizing Chemistry Using the Chemical Processing Unit: From Synthesis to Discovery. *Acc. Chem. Res.* **2021**, *54*, 253–262.
- (15) Rohrbach, S.; Šiaučiulis, M.; Chisholm, G.; Pirvan, P.-A.; Saleeb, M.; Mehr, S. H. M.; Trushina, E.; Leonov, A. I.; Keenan, G.; Khan, A.; Hammer, A.; Cronin, L. Digitization and Validation of a Chemical Synthesis Literature Database in the ChemPU. *Science* **2022**, *377*, 172–180.
- (16) Bubliauskas, A.; Blair, D. J.; Powell-Davies, H.; Kitson, P. J.; Burke, M. D.; Cronin, L. Digitizing Chemical Synthesis in 3D Printed Reactionware. *Angew. Chem., Int. Ed.* **2022**, *61*, e202116108.
- (17) Nikolaev, P.; Hooper, D.; Webber, F.; Rao, R.; Decker, K.; Krein, M.; Poleski, J.; Barto, R.; Maruyama, B. Autonomy in Materials Research: A Case Study in Carbon Nanotube Growth. *Npj Comput. Mater.* **2016**, *2*, 16031.
- (18) Roch, L. M.; Häse, F.; Kreisbeck, C.; Tamayo-Mendoza, T.; Yunker, L. P. E.; Hein, J. E.; Aspuru-Guzik, A. ChemOS: Orchestrating Autonomous Experimentation. *Sci. Robot.* **2018**, *3* (19), 1.
- (19) Rahamanian, F.; Flowers, J.; Guevarra, D.; Richter, M.; Fichtner, M.; Donnelly, P.; Gregoire, J. M.; Stein, H. S. Enabling Modular Autonomous Feedback-Loops in Materials Science through Hierarchical Experimental Laboratory Automation and Orchestration. *Adv. Mater. Interfaces* **2022**, *9*, 2101987.
- (20) Gaudin, T.; Benlolo, I.; Cui, Z. Y.; Hickmann, R.; Tamblyn, I.; Aspuru-Guzik, A. Molar. In Zenodo; 2022; <https://zenodo.org/record/6809290>.
- (21) McKinney, W. Data Structures for Statistical Computing in Python. In Proc. of the 9th Python in Science Conference, Austin, Texas, 2010; pp 56–61.
- (22) Reback, J.; McKinney, W.; Jbrockmendel; Van den Bossche, J.; Roeschke, M.; Augspurger, T.; Hawkins, S.; Cloud, P.; Gfyoung; Sinhrks; Hoefer, P.; Klein, A.; Petersen, T.; Tratner, J.; She, C.; Ayd, W.; Naveh, S.; Darbyshire, J. H. M.; Shadrach, R.; Garcia, M.; Schendel, J.; Hayden, A.; Saxton, D.; Gorelli, M. E.; Li, F.; Wörtwein, T.; Zeitlin, M.; Jancauskas, V.; McMaster, A.; Li, T. Pandas-Dev/Pandas: Pandas 1.4.3. In Zenodo, 2022; <https://zenodo.org/record/3509134>.
- (23) Häse, F.; Roch, L. M.; Kreisbeck, C.; Aspuru-Guzik, A. Phoenics: A Bayesian Optimizer for Chemistry. *ACS Cent. Sci.* **2018**, *4*, 1134–1145.
- (24) Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Chimera: Enabling Hierarchy Based Multi-Objective Optimization for Self-Driving Laboratories. *Chem. Sci.* **2018**, *9*, 7642–7655.
- (25) Häse, F.; Aldeghi, M.; Hickman, R. J.; Roch, L. M.; Aspuru-Guzik, A. Gryffin: An Algorithm for Bayesian Optimization of Categorical Variables Informed by Expert Knowledge. *Appl. Phys. Rev.* **2021**, *8*, 031406.
- (26) Hickman, R. J.; Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Gemini: Dynamic Bias Correction for Autonomous Experimentation and Molecular Simulation. *arXiv*, **2021**, 2103.03391.
- (27) Aldeghi, M.; Häse, F.; Hickman, R. J.; Tamblyn, I.; Aspuru-Guzik, A. Golem: An Algorithm for Robust Experiment and Process Optimization. *arXiv*, **2021**, 2103.03716.
- (28) 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.; Lai, V.; Ng, G. J.; Situ, H.; Zhang, R. H.; Elliott, M. S.; Haley, T. H.; Dvorak, D. J.; Aspuru-Guzik, A.; Hein, J. E.; Berlinguette, C. P. Self-Driving Laboratory for Accelerated Discovery of Thin-Film Materials. *Sci. Adv.* **2020**, *6*, eaaz8867.
- (29) Glasnov, T. N.; Kappe, C. O. The Microwave-to-Flow Paradigm: Translating High-Temperature Batch Microwave Chemistry to Scalable Continuous-Flow Processes. *Chem. – Eur. J.* **2011**, *17*, 11956–11968.
- (30) Plutschack, M. B.; Pieber, B.; Gilmore, K.; Seeberger, P. H. The Hitchhiker's Guide to Flow Chemistry. *Chem. Rev.* **2017**, *117*, 11796–11893.

- (31) Bianchi, P.; Williams, J. D.; Kappe, C. O. Oscillatory Flow Reactors for Synthetic Chemistry Applications. *J. Flow Chem.* **2020**, *10*, 475–490.
- (32) Christensen, M.; Yunker, L. P. E.; Shiri, P.; Zepel, T.; Prieto, P. L.; Grunert, S.; Bork, F.; Hein, J. E. Automation Isn't Automatic. *Chem. Sci.* **2021**, *12*, 15473.
- (33) Gillis, E. P.; Burke, M. D. Multistep Synthesis of Complex Boronic Acids from Simple MIDA Boronates. *J. Am. Chem. Soc.* **2008**, *130*, 14084–14085.
- (34) Li, J.; Ballmer, S. G.; Gillis, E. P.; Fujii, S.; Schmidt, M. J.; Palazzolo, A. M. E.; Lehmann, J. W.; Morehouse, G. F.; Burke, M. D. Synthesis of Many Different Types of Organic Small Molecules Using One Automated Process. *Science* **2015**, *347*, 1221–1226.
- (35) Anthony, J. E.; Heeney, M.; Ong, B. S. Synthetic Aspects of Organic Semiconductors. *MRS Bull.* **2008**, *33*, 698–705.
- (36) Kuehne, A. J. C.; Gather, M. C. Organic Lasers: Recent Developments on Materials, Device Geometries, and Fabrication Techniques. *Chem. Rev.* **2016**, *116*, 12823–12864.
- (37) Wu, T. C.; Granda, A. A.; Hotta, K.; Yazdani, S. A.; Pollice, R.; Vestfrid, J.; Hao, H.; Lavigne, C.; Seifrid, M.; Angello, N.; Bencheikh, F.; Hein, J. E.; Burke, M.; Adachi, C.; Aspuru-Guzik, A. A Materials Acceleration Platform for Organic Laser Discovery. *ChemRxiv* **2022**; DOI: [10.26434/chemrxiv-2022-9zm65](https://doi.org/10.26434/chemrxiv-2022-9zm65).
- (38) Yunker, L. P. E.; Donnecke, S.; Ting, M.; Yeung, D.; McIndoe, J. S. PythoMS: A Python Framework To Simplify and Assist in the Processing and Interpretation of Mass Spectrometric Data. *J. Chem. Inf. Model.* **2019**, *59*, 1295–1300.
- (39) Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I. A Mobile Robotic Chemist. *Nature* **2020**, *583*, 237–241.
- (40) Crosby, G. A.; Demas, J. N. Measurement of Photoluminescence Quantum Yields. Review. *J. Phys. Chem.* **1971**, *75*, 991–1024.
- (41) O'Connor, D. V. O.; Phillips, D. *Time-Correlated Single Photon Counting*; Academic Press: London, 1984.
- (42) Cava, R. J.; DiSalvo, F. J.; Brus, L. E.; Dunbar, K. R.; Gorman, C. B.; Haile, S. M.; Interrante, L. V.; Musfeldt, J. L.; Navrotsky, A.; Nuzzo, R. G.; Pickett, W. E.; Wilkinson, A. P.; Ahn, C.; Allen, J. W.; Burns, P. C.; Ceder, G.; Chidsey, C. E. D.; Clegg, W.; Coronado, E.; Dai, H.; Deem, M. W.; Dunn, B. S.; Galli, G.; Jacobson, A. J.; Kanatzidis, M.; Lin, W.; Manthiram, A.; Mrksich, M.; Norris, D.; Nozik, A. J.; Peng, X.; Rawn, C.; Rolison, D.; Singh, D. J.; Toby, B. H.; Tolbert, S.; Wiesner, U. B.; Woodward, P. M.; Yang, P. Future Directions in Solid State Chemistry: Report of the NSF-Sponsored Workshop. *Prog. Solid State Chem.* **2002**, *30*, 1–101.
- (43) Aspuru-Guzik, A.; Persson, K. *Materials Acceleration Platform: Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods and Artificial Intelligence*; Mission Innovation: Innovation Challenge 6; Canadian Institute for Advanced Research, 2018.
- (44) Fabbri, E.; Schmidt, T. J. Oxygen Evolution Reaction—The Enigma in Water Electrolysis. *ACS Catal.* **2018**, *8*, 9765–9774.
- (45) Coley, C. W.; Eyke, N. S.; Jensen, K. F. Autonomous Discovery in the Chemical Sciences Part II: Outlook. *Angew. Chem., Int. Ed.* **2020**, *59* (52), 23414–23436.
- (46) Molga, K.; Szymkuć, S.; Grzybowski, B. A. Chemist Ex Machina: Advanced Synthesis Planning by Computers. *Acc. Chem. Res.* **2021**, *54*, 1094–1106.
- (47) Shim, E.; Kammeraad, J. A.; Xu, Z.; Tewari, A.; Cernak, T.; Zimmerman, P. M. Predicting Reaction Conditions from Limited Data through Active Transfer Learning. *Chem. Sci.* **2022**, *13*, 6655–6668.
- (48) Grzybowski, B. A.; Badowski, T.; Molga, K.; Szymkuć, S. Network Search Algorithms and Scoring Functions for Advanced-Level Computerized Synthesis Planning. *WIREs Comput. Mol. Sci.* **e1630**;
- (49) Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R. P.; de Freitas, N. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* **2016**, *104*, 148–175.
- (50) Hickman, R. J.; Aldeghi, M.; Häse, F.; Aspuru-Guzik, A. Bayesian Optimization with Known Experimental and Design Constraints for Chemistry Applications. *arXiv* **2022**, 2203.17241.
- (51) Blaženović, I.; Kind, T.; Torbašinović, H.; Obrenović, S.; Mehta, S. S.; Tsugawa, H.; Wermuth, T.; Schauer, N.; Jahn, M.; Biedendieck, R.; Jahn, D.; Fiehn, O. Comprehensive Comparison of in Silico MS/MS Fragmentation Tools of the CASMI Contest: Database Boosting Is Needed to Achieve 93% Accuracy. *J. Cheminformatics* **2017**, *9*, 32.
- (52) De Vijlder, T.; Valkenborg, D.; Lemière, F.; Romijn, E. P.; Laukens, K.; Cuyckens, F. A Tutorial in Small Molecule Identification via Electrospray Ionization-mass Spectrometry: The Practical Art of Structural Elucidation. *Mass Spectrom. Rev.* **2018**, *37*, 607–629.
- (53) Cook, A.; Johnson, A. P.; Law, J.; Mirzazadeh, M.; Ravitz, O.; Simon, A. Computer-Aided Synthesis Design: 40 Years On. *WIREs Comput. Mol. Sci.* **2012**, *2*, 79–107.
- (54) Rappoport, D.; Aspuru-Guzik, A. Predicting Feasible Organic Reaction Pathways Using Heuristically Aided Quantum Chemistry. *J. Chem. Theory Comput.* **2019**, *15*, 4099–4112.
- (55) Rappoport, D. Reaction Networks and the Metric Structure of Chemical Space(s). *J. Phys. Chem. A* **2019**, *123*, 2610–2620.
- (56) Wołos, A.; Roszak, R.; Żądro-Dobrowolska, A.; Beker, W.; Mikulak-Klucznik, B.; Spólnik, G.; Dygas, M.; Szymkuć, S.; Grzybowski, B. A. Synthetic Connectivity, Emergence, and Self-Regeneration in the Network of Prebiotic Chemistry. *Science* **2020**, *369*, eaaw1955.
- (57) Arya, A.; Ray, J.; Sharma, S.; Simbron, R. C.; Lozano, A.; Smith, H. B.; Andersen, J. L.; Chen, H.; Meringer, M.; Cleaves, H. J. An Open Source Computational Workflow for the Discovery of Autocatalytic Networks in Abiotic Reactions. *Chem. Sci.* **2022**, *13*, 4838–4853.
- (58) Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. CFM-ID: A Web Server for Annotation, Spectrum Prediction and Metabolite Identification from Tandem Mass Spectra. *Nucleic Acids Res.* **2014**, *42*, W94–W99.
- (59) Djoumbou-Feunang, Y.; Pon, A.; Karu, N.; Zheng, J.; Li, C.; Arndt, D.; Gautam, M.; Allen, F.; Wishart, D. S. CFM-ID 3.0: Significantly Improved ESI-MS/MS Prediction and Compound Identification. *Metabolites* **2019**, *9*, 72.
- (60) Ji, H.; Deng, H.; Lu, H.; Zhang, Z. Predicting a Molecular Fingerprint from an Electron Ionization Mass Spectrum with Deep Neural Networks. *Anal. Chem.* **2020**, *92*, 8649–8653.
- (61) Xue, J.; Guijas, C.; Benton, H. P.; Warth, B.; Siuzdak, G. METLIN MS 2 Molecular Standards Database: A Broad Chemical and Biological Resource. *Nat. Methods* **2020**, *17*, 953–954.
- (62) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. MassBank: A Public Repository for Sharing Mass Spectral Data for Life Sciences. *J. Mass Spectrom.* **2010**, *45*, 703–714.
- (63) NIST. *NIST 20 Tandem Mass Spectral Libraries*; <https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:msms> (accessed 2022-07-19).
- (64) Sans, V.; Porwol, L.; Dragone, V.; Cronin, L. A Self Optimizing Synthetic Organic Reactor System Using Real-Time in-Line NMR Spectroscopy. *Chem. Sci.* **2015**, *6*, 1258–1264.
- (65) Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. Controlling an Organic Synthesis Robot with Machine Learning to Search for New Reactivity. *Nature* **2018**, *559*, 377–381.
- (66) Maschmeyer, T.; Prieto, P. L.; Grunert, S.; Hein, J. E. Exploration of Continuous-Flow Benchtop NMR Acquisition Parameters and Considerations for Reaction Monitoring. *Magn. Reson. Chem.* **2020**, *58*, 1234–1248.

- (67) Chatterjee, S.; Guidi, M.; Seeberger, P. H.; Gilmore, K. Automated Radial Synthesis of Organic Molecules. *Nature* **2020**, *579*, 379–384.
- (68) Bahr, M. N.; Damon, D. B.; Yates, S. D.; Chin, A. S.; Christopher, J. D.; Cromer, S.; Perrotto, N.; Quiroz, J.; Rosso, V. Collaborative Evaluation of Commercially Available Automated Powder Dispensing Platforms for High-Throughput Experimentation in Pharmaceutical Applications. *Org. Process Res. Dev.* **2018**, *22*, 1500–1508.
- (69) Bahr, M. N.; Morris, M. A.; Tu, N. P.; Nandkeolyar, A. Recent Advances in High-Throughput Automated Powder Dispensing Platforms for Pharmaceutical Applications. *Org. Process Res. Dev.* **2020**, *24*, 2752.
- (70) Tu, N. P.; Dombrowski, A. W.; Goshu, G. M.; Vasudevan, A.; Djuric, S. W.; Wang, Y. High-Throughput Reaction Screening with Nanomoles of Solid Reagents Coated on Glass Beads. *Angew. Chem., Int. Ed.* **2019**, *58*, 7987–7991.
- (71) Shiri, P.; Lai, V.; Zepel, T.; Griffin, D.; Reifman, J.; Clark, S.; Grunert, S.; Yunker, L. P. E.; Steiner, S.; Situ, H.; Yang, F.; Prieto, P. L.; Hein, J. E. Automated Solubility Screening Platform Using Computer Vision. *iScience* **2021**, *24*, 102176.
- (72) J3016C: *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*; 2021; https://www.sae.org/standards/content/j3016_202104.
- (73) Ruiz-Castillo, P.; Buchwald, S. L. Applications of Palladium-Catalyzed C–N Cross-Coupling Reactions. *Chem. Rev.* **2016**, *116*, 12564–12649.
- (74) Chinchilla, R.; Nájera, C. The Sonogashira Reaction: A Booming Methodology in Synthetic Organic Chemistry. *Chem. Rev.* **2007**, *107*, 874–922.
- (75) Breugst, M.; Reissig, H.-U. The Huisgen Reaction: Milestones of the 1,3-Dipolar Cycloaddition. *Angew. Chem., Int. Ed.* **2020**, *59*, 12293–12307.
- (76) Molga, K.; Szymkuć, S.; Golębiowska, P.; Popik, O.; Dittwald, P.; Moskal, M.; Roszak, R.; Mlynarski, J.; Grzybowski, B. A. A Computer Algorithm to Discover Iterative Sequences of Organic Reactions. *Nat. Synth.* **2022**, *1*, 49–58.
- (77) Eppel, S.; Xu, H.; Bismuth, M.; Aspuru-Guzik, A. Computer Vision for Recognition of Materials and Vessels in Chemistry Lab Settings and the Vector-LabPics Data Set. *ACS Cent. Sci.* **2020**, *6*, 1743–1752.
- (78) Steiner, S.; Wolf, J.; Glatzel, S.; Andreou, A.; Granda, J. M.; Keenan, G.; Hinkley, T.; Aragon-Camarasa, G.; Kitson, P. J.; Angelone, D.; Cronin, L. Organic Synthesis in a Modular Robotic System Driven by a Chemical Programming Language. *Science* **2019**, *363*, 1.
- (79) Mehr, S. H. M.; Craven, M.; Leonov, A. I.; Keenan, G.; Cronin, L. A Universal System for Digitization and Automatic Execution of the Chemical Synthesis Literature. *Science* **2020**, *370*, 101–108.
- (80) Acceleration Consortium; <https://acceleration.utoronto.ca> (accessed 2021-06-10).

□ Recommended by ACS

Generative Models as an Emerging Paradigm in the Chemical Sciences

Dylan M. Anstine and Olexandr Isayev

APRIL 13, 2023

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

READ ↗

Molecular Machine Learning for Chemical Catalysis: Prospects and Challenges

Sukriti Singh and Raghavan B. Sunoj

JANUARY 30, 2023

ACCOUNTS OF CHEMICAL RESEARCH

READ ↗

Recognition-Encoded Synthetic Information Molecules

Giulia Iadevaia and Christopher A. Hunter

MARCH 09, 2023

ACCOUNTS OF CHEMICAL RESEARCH

READ ↗

Machine-Learning-Guided Discovery of Electrochemical Reactions

Andrew F. Zahrt, Klavs F. Jensen, et al.

DECEMBER 02, 2022

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

READ ↗

Get More Suggestions >