

## Self-Driving Laboratories for Chemistry and Materials Science

Gary Tom,\* Stefan P. Schmid, Sterling G. Baird, Yang Cao, Kourosh Darvish, Han Hao, Stanley Lo, Sergio Pablo-García, Ella M. Rajaonson, Marta Skreta, Naruki Yoshikawa, Samantha Corapi, Gun Deniz Akkoc, Felix Strieth-Kalthoff,\* Martin Seifrid,\* and Alán Aspuru-Guzik\*



Cite This: *Chem. Rev.* 2024, 124, 9633–9732



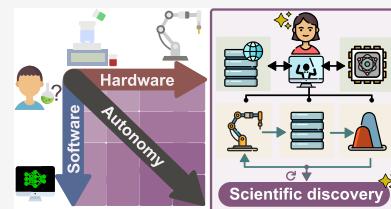
Read Online

ACCESS |

Metrics & More

Article Recommendations

**ABSTRACT:** Self-driving laboratories (SDLs) promise an accelerated application of the scientific method. Through the automation of experimental workflows, along with autonomous experimental planning, SDLs hold the potential to greatly accelerate research in chemistry and materials discovery. This review provides an in-depth analysis of the state-of-the-art in SDL technology, its applications across various scientific disciplines, and the potential implications for research and industry. This review additionally provides an overview of the enabling technologies for SDLs, including their hardware, software, and integration with laboratory infrastructure. Most importantly, this review explores the diverse range of scientific domains where SDLs have made significant contributions, from drug discovery and materials science to genomics and chemistry. We provide a comprehensive review of existing real-world examples of SDLs, their different levels of automation, and the challenges and limitations associated with each domain.



### CONTENTS

1. Introduction	9634
1.1. Brief History	9635
2. Infrastructure	9636
2.1. Hardware	9636
2.1.1. Specialized Hardware	9636
2.1.2. General-Purpose Robot Applied for Chemistry	9637
2.1.3. Open Hardware for Lab Automation	9638
2.1.4. Perception and Computer Vision	9638
2.1.5. Manipulation Skill Learning and Digital Twin	9638
2.1.6. Robotics Safety in Laboratories	9639
2.2. Software	9639
2.2.1. Orchestration	9639
2.2.2. Communication and Protocol Management for SDLs	9640
2.2.3. Data Management	9641
2.2.4. Role of Artificial Intelligence in Chemical Discovery	9641
2.2.5. Experiment Planning	9642
2.2.6. Bayesian Optimization and Active Learning	9644
3. Analytical Process Optimization	9645
3.1. Composition and Detection Process Optimization	9646
3.1.1. Sample Preparation	9646
3.1.2. Separation and Chromatography	9647
3.2. Other Properties	9647
4. Reaction Optimization	9649

4.1. Specialized Hardware and Software	9650
4.1.1. Reaction Execution	9650
4.1.2. Reaction Analysis	9651
4.1.3. Early Examples of Autonomous Condition Optimization	9651
4.2. Single-Step Organic Reactions	9652
4.2.1. Self-Optimizing Flow Reactors and Analytical Advances	9652
4.2.2. Discrete and Categorical Optimization and Batch-Type Reactors	9663
4.2.3. Pareto Optimizations and Further Algorithmic Advances	9666
4.3. Multi-Step Organic Reactions	9668
4.3.1. Sequential Single-Step Optimizations	9669
4.3.2. Simultaneous Multi-Step Optimization	9670
4.4. Further Solution-Phase Reactions	9672
4.5. Catalyst and Reaction Discovery	9673
4.5.1. New Catalyst Materials	9673
4.5.2. New Reactions and Reaction Types	9675
4.6. Determination of Reaction Kinetics	9676
4.7. Solid State Materials Synthesis	9677
4.8. Outlook and Perspectives	9680
5. Drug Discovery and Biochemistry	9680

Published: August 13, 2024



5.1. Drug Discovery Pipeline	9681
5.1.1. Target Identification and Validation	9681
5.1.2. Hit Discovery	9682
5.1.3. Hit-to-Lead and Lead Optimization	9683
5.1.4. Formulation Optimization and Bioavailability	9685
5.2. Synthetic Biology	9687
5.2.1. Biosynthetic Pathways Optimization	9687
5.2.2. Engineering Protein Stability	9688
5.3. Outlook and Perspectives	9689
6. Structural Materials	9690
6.1. Alloy Design	9690
6.2. Concrete Formulations	9691
6.3. Non-Alloy Additive Manufacturing	9692
6.4. Adhesives	9693
6.5. Outlook and Perspectives	9693
7. Optoelectronics	9693
7.1. In Solution and Crystals	9693
7.1.1. Perovskites	9693
7.1.2. Nanoparticles	9694
7.1.3. Molecules in Solution	9698
7.2. Thin Films	9699
7.3. Devices	9701
7.4. Outlook and Perspectives	9702
8. Energy Storage Materials	9702
8.1. Materials Synthesis and Characterization	9703
8.1.1. Lithium-Ion Batteries	9703
8.1.2. Alternatives to LIBs	9705
8.1.3. Redox Flow Batteries	9705
8.1.4. Hydrogen and Other Fuels	9706
8.2. Device Design, Assembly and Optimization	9707
8.2.1. Cell Batteries	9707
8.2.2. Flow Reactors	9709
8.3. Outlook and Perspectives	9709
9. Conclusion and Outlook	9709
Author Information	9710
Corresponding Authors	9710
Authors	9710
Author Contributions	9711
Notes	9711
Biographies	9711
Acknowledgments	9712
References	9712

## 1. INTRODUCTION

In the face of pressing global challenges such as climate change, energy sustainability, and current or emerging healthcare crises, we must seek efficient solutions in the context of a growing global population and increasing resource demands. The accelerated development of materials, technology, and scientific understanding emerges as a potential avenue for tackling these challenges. Traditional research methods, often characterized by gradual progress with limited efficiency, may prove insufficient for the urgency these challenges demand. The integration of laboratory automation and data-driven decision-making can potentially facilitate a more rapid and efficient exploration of solutions, while offering multiple advantages over traditional scientific discovery.<sup>1,2</sup> Notably, automated experiments can attain higher throughput and precision, while data-driven search algorithms can quickly and efficiently explore experimental space based on feedback from available data (“closed-loop” experimentation). Additionally,

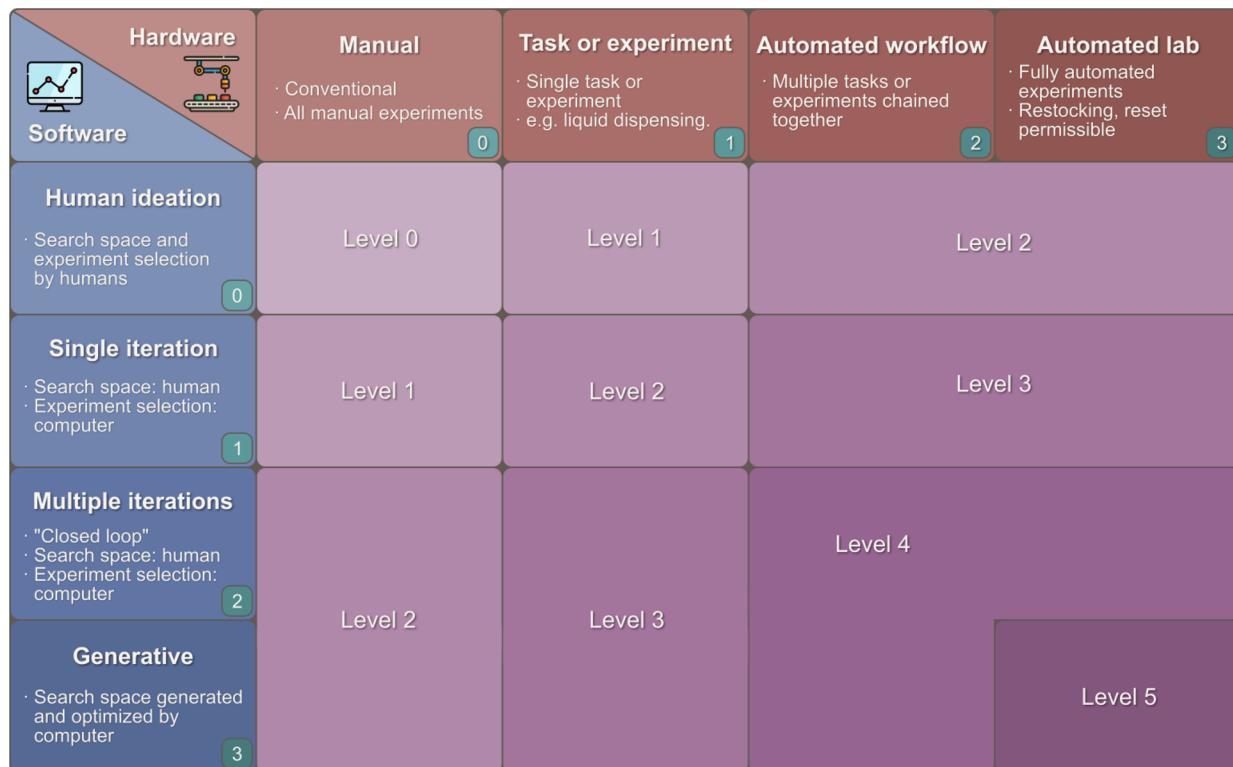
issues such as reproducibility challenges and the under-representation of null results in the scientific literature have been identified.<sup>3–6</sup> At the same time, automation encourages the further digitization of research.<sup>7</sup> The utilization of automated systems enables more precise documentation of experimental protocols, enhancing repeatability and reproducibility, while digitization facilitates data recording and sharing, with particular emphasis on the significance of null results, contributing to a more comprehensive and accurate portrayal of scientific endeavors. High quality large datasets made possible by autonomous experimentation would aid in the development of artificial intelligence (AI) for materials science and chemistry, creating better machine learning (ML) and deep learning (DL) models, and enhancing the decision-making capabilities of data-driven algorithms.

Against this background, the concept of a **self-driving laboratory** (SDL) describes a system in which automated experiments are integrated with data-driven decision-making, with the goal of accelerating the application of the scientific method. Early efforts in this direction date back to the 1970s, and have been referred to as “autonomous laboratories,” “closed-loop experimentation,” or “materials acceleration platforms” since. The term SDL is more commonly used in the 21st century; to the best of our knowledge, the term SDL was first introduced by Maruyama and co-workers.<sup>8</sup> In the context of chemistry and materials science, the main use cases of SDLs have encompassed the discovery of molecules or materials with optimized properties, or the optimization of reaction and process conditions. Before discussing these examples in detail in the later sections of the review, we want to start with a conceptual overview, introducing the main components of SDLs. The concept of SDLs has two critical defining dimensions: the automation of data-driven decision-making (software), and the automation of experimental workflows (hardware). Hence, we classify SDLs according to the degrees of autonomy in these two axes.

Concerning software autonomy, which pertains to experiment selection, SDLs can be classified into three categories: (1) single iteration of automated experimentation with a data-driven method to design the experiments, (2) multiple iterations on a closed-loop system, in which the experimental results feedback to guide another round of automated experiments, and (3) generative approaches, in which multiple iterations of closed-loop optimization are performed in a search space or chemical space that is generated by an algorithm.

On the other hand, hardware autonomy, which pertains to experiment execution, classifies SDLs into: (1) single-task setups primarily aimed at conducting a single type of experiments, (2) workflow configurations involving multiple tasks or experiments for design or discovery purposes, and (3) fully automated labs capable of executing a diverse range of experiments without human intervention. These dual categorizations offer a comprehensive framework for understanding the diverse landscape of SDLs and their applications in scientific exploration (Figure 1).

Similar to the hierarchical framework used to define levels of autonomy in self-driving cars, a similar classification system is applicable to SDLs.<sup>9</sup> Level 5 SDLs attain the highest level of autonomy, achieving category 3 in both software and hardware autonomy. Level 4 SDLs, while not reaching the pinnacle of autonomy, achieve category 3 in either software or hardware, or category 2 in both. Level 3 SDLs attain a baseline of



**Figure 1.** Schematic for the autonomy levels of SDLs based on the category of hardware and software autonomy achieved.

autonomy by achieving at least category 1 in either software or hardware and at least category 2 in the other. Level 2 SDLs are characterized by relying on human ideation (software category 0) in combination with at least an automated workflow (hardware category 2) or on manual experiments (hardware category 0) with at least multiple iterations of computer proposed experiments (software category 2). Further, experiments with both software and hardware category 1 are also classified as Level 2 SDLs. The categorization extends further to include experiments that reach category 1 or above in hardware autonomy but rely on human ideation (software category 0), which are termed automated experiments. Conversely, experiments that attain category 1 or higher in software autonomy but necessitate entirely manual experiments (hardware category 0) are termed ML-informed experiments. This hierarchical structure provides a nuanced understanding of the varying degrees of autonomy exhibited by SDLs, offering a comprehensive framework for characterizing their capabilities.

For the purposes of this review, we cover in detail the SDLs that attain a minimum of category 1 in both hardware and software autonomy. Level 2 and 3 SDLs make up the vast majority of SDL examples to date. For robotically “simple” tasks, a series of Level 4 SDLs have been demonstrated, however, a true Level 5 SDL remains an unattained goal in the field. Initially, the automation of laboratory workflows focused on more elementary tasks such as liquid handling, or data analysis. More advanced SDLs combining data-driven decision-making with automated experimentation have flourished, driven by advancements in AI and ML. More recent developments in robotics and computer vision have allowed for automation of more complicated and general-purpose chemistry laboratory tasks, typically performed by human chemists.<sup>10–14</sup> While still in the early stages of development,

such general-purpose SDLs have promising prospects for the future of scientific research.

In this review, we provide a comprehensive overview of the development of SDLs and their applications in the domain of chemistry and materials science. We first provide a historical perspective, and then a discussion on the required infrastructure for SDLs. Before discussing SDLs for materials discovery in detail, we provide a comprehensive overview of SDLs for optimizing chemical processes and, in particular, chemical synthesis. This directly lays the foundations for SDLs in the field of drug discovery and biochemistry. In fact, the pharmaceutical industry has been a key driver in the field of SDL technologies, due to the industrial and commercial importance of drug discovery, pioneering both experimental and computational high-throughput experimentation (HTE) and screening. Subsequently, we shift our attention to material design and discovery, covering structural materials, optoelectronic materials, and energy storage materials. Finally, we provide our perspective on the important challenges that need to be addressed by SDLs and the future outlook for their further development. A previous version of this review was published online as a preprint.<sup>15</sup>

### 1.1. Brief History

The concept of SDLs has its roots in the broader field of laboratory automation, which began in the mid-20th century. The initial concepts of the SDL were focused on the design of experiments (DoE),<sup>16,17</sup> a systematic approach of exploring parameters within a chemical process or experiment in order to optimize a particular outcome, for example the yield of a chemical reaction, and identify influential parameters within the process or experiment. The idea of using machines and robots to perform repetitive and time-consuming tasks in laboratories gained momentum with the advent of industrial

automation in manufacturing processes. The first steps towards autonomous laboratories involved the automation of elementary laboratory tasks, such as liquid handling, sample preparation, and data analysis.<sup>18–20</sup>

In the 1940s, industrial forces led by chemists Bosch and Mittasch devised highly parallelizable robots with continuous flow platform that could perform grid searches for ammonia fixation catalysts for the Haber-Bosch process, testing up to 4000 compounds.<sup>21</sup> Industrially, one of the first purpose-built robots appeared around 1960, developed by Unimation Inc., and was developed for automated die-casting processes.<sup>22</sup> In 1966, Merrifield *et al.*<sup>23</sup> demonstrated an automated stepwise synthesis platform for peptides, utilizing flow systems and batch reactors controlled by a “stepping-drum” programmer—a mechanical computer that activated microswitches through plugs on a rotating drum. Around this time, methods for optimizing black-box functions were being applied to analytical chemistry, for example, the closed-loop optimization of measurement parameters to maximize the response signal.<sup>24</sup> With more precise robotics and computer controls, smaller and more modular platforms were developed to automate specific chemistry tasks, allowing for automated workflows involving multiple instruments. By the 1980s, Zymark led the development and commercialization of laboratory robotics which automated parts of sample production and data analysis of immunoassays, and created one of the first robotic arms for chemical laboratories.<sup>20,25,26</sup>

With the rise of computers and computational power, experimental planning involved more sophisticated computational methods, such as simplex optimization,<sup>27,28</sup> regression techniques,<sup>29</sup> Bayesian optimization,<sup>30,31</sup> and evolutionary algorithms.<sup>32</sup> Using AI to plan robotic experiments was first discussed by T. L. Isenhour in 1985.<sup>33</sup> Dubbed the analytical director, the AI system would be capable of applying domain knowledge to reduce the search space in a DoE. Since then, the rise of AI and advancements in ML algorithms were instrumental in shaping the development of SDLs. AI technologies enabled these labs to process large amounts of data, make predictions, and even optimize experiments in real-time. Most importantly, AI methods can allow for the domain expert (i.e., the chemist) to be removed from the experimental planning process, as DL and ML techniques, with a sufficient amount of data, could be trained to recognize complex statistical patterns. Such models are capable of making data-driven decisions on future experiments, and learning from the experimental feedback. Such capabilities opened up new possibilities for exploring complex chemical and biological spaces in materials and drug discovery, and these SDLs will be the focus of this review.

While the earliest examples of autonomous closed-loop workflows have been demonstrated in the 1970s, the integration of laboratory robotics with more sophisticated ML algorithms started to gain traction in the 2000s. As the cost and capabilities of higher-throughput robotics systems and computers became accessible to individual laboratories, the development of higher levels of SDLs became more widespread. For example, continuous flow systems with digital controls and automated characterization became more commonly used, particularly in the discovery of pharmaceutical compounds.<sup>34</sup> In 2007, Krishnadasan *et al.*<sup>35</sup> demonstrated a closed-loop flow-based microfluidics SDL capable of synthesizing size-controlled CdSe nanoparticles by optimizing a custom utility function based on the data from an on-line

spectrometer.<sup>36</sup> And in 2009, King *et al.*<sup>37</sup> created *Adam*, an SDL capable of generating genomics hypotheses from bioinformatics models, designing experiments, and performing biological assays. Subsequently, *Eve* was developed by Williams *et al.*<sup>38</sup> exploring a large library of drug molecules for hit identification, performing assays and feeding back the results into a quantitative structure-activity relationship (QSAR) cheminformatics model.

While still relatively early in their development, SDLs hold immense promise for the future of scientific research and are poised to revolutionize how we approach complex scientific challenges in the coming years. As the technology continues to mature and gain wider acceptance, SDLs have expanded to various scientific domains beyond materials and drug discovery. The continuous evolution and integration of AI, robotics, and laboratory automation will likely unlock even greater potential for these autonomous laboratories in the decades to come.

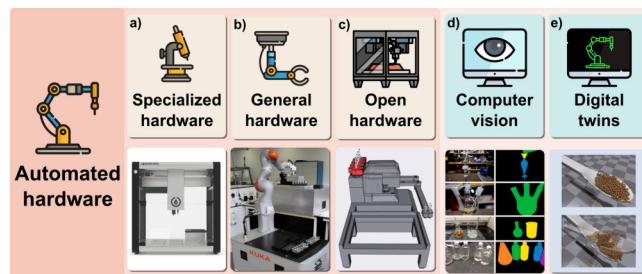
## 2. INFRASTRUCTURE

SDLs encompass three fundamental components: (i) automated laboratory devices proficient in executing complex chemical operations, (ii) software packages designed to seamlessly handle laboratory operations and the resulting data, and (iii) an experimental planner capable of processing acquired data and guiding subsequent laboratory procedures. In this section, we provide an overview of the essential constituents of SDLs and discuss ongoing efforts to harmonize their integration and control.

### 2.1. Hardware

Chemical experiments require different types of operations including chemical handling, reaction execution, post-reaction processing/purification, formulation, device fabrication, and chemical property measurements. For certain steps, task-specific automated hardware systems have existed for decades, and many have been commercialized as standard laboratory instruments. Most of these systems have not been designed for fully automated workflows, but rather for interfacing with a human researcher. Prominent examples stem from the field of analytical chemistry, where automated solutions, often covering multi-step analytical workflows (e.g., chromatography-mass spectrometry from an autosampler), are routinely available in chemistry laboratories. The integration of such platforms with further automated solutions to enable SDLs presents a major challenge, and can be approached through different strategies. Fixed purpose-built automated systems couple multiple platforms in a static fashion, whereas partially automated workflows, requiring a human-in-the-loop, allow for platforms that can be adapted and repurposed for different experiments.<sup>39</sup> Development in general-purpose robotic systems that can perform basic chemistry tasks and interface with the modules have allowed for completely automated yet modular SDLs.<sup>40</sup> Moreover, open hardware for lab automation has been proposed to lower the financial barrier to building a SDL.<sup>41</sup> In this subsection, we will review the various automated hardware modules, and provide a brief discussion on the development of general-purpose chemistry robotics platforms. (Figure 2) A summary of the distinctions between the hardware types are shown in Table 1.

**2.1.1. Specialized Hardware.** The automation of laboratory operations has come a long way since the high-throughput catalyst screening campaigns performed by Bosch



**Figure 2.** Examples of types of automated hardware. (a)-(c) Categories of automated hardware for SDLs and (d)-(e) supporting software for automation. Bottom half show examples of each category: (a) OT-2 platform manufactured by OpenTrons;<sup>42</sup> (b) robotic arm for chemical operations in Cooper's autonomous laboratory, adapted with permission from reference.<sup>11</sup> Copyright 2020, Springer Nature; (c) 3D schematics of Sidekick, a low cost liquid dispensing platform developed by Keesey *et al.*, adapted with permission from reference.<sup>43</sup> Copyright 2022, Elsevier; (d) computer vision framework for laboratory glassware developed by Eppel *et al.*;<sup>12</sup> and (e) solid weighting simulation software developed by Kadokawa *et al.*, adapted with permission from reference.<sup>44</sup> Copyright 2023, Institute of Electrical and Electronics Engineers (IEEE).

and Mittasch using continuous-flow platforms. As of today, chemistry and materials discovery laboratories host a range of automated solutions for routine tasks. At the same time, there are still many operations that are routinely performed by human researchers in traditional laboratories, due to the need for operational flexibility and adaptive decision-making.<sup>45</sup> Finding automated solutions for these workflows requires interdisciplinary efforts within chemistry and engineering, prompting a re-conceptualization of central laboratory processes.

At the core of most laboratory routines lies a set of fundamental operations that are essential across various types of experiments. These include, most importantly, the handling and transfer of materials (most often as liquids or solids), or the precise control of vessel or reactor conditions like temperature, atmosphere, and pressure. Whereas the latter have largely benefited from technological advances outside of chemistry, the challenge of automated reagent handling remains specific to the chemical laboratory.

The most straightforward and widely applied solution to automated liquid dispensing is the use of syringe pumps or peristaltic pumps. Commercial solutions to these technologies are widespread, and have facilitated the transfer and dispensing liquids in numerous SDLs. The automation of positive-displacement pipettes (PDPs) has also emerged as an alternative for robotic liquid dispensing, particularly in the context of biological experimentation. Gantry-based systems using PDPs (e.g. SPT LabTech Mosquito, or the OpenTrons OT systems), allow for substantial throughput increases of parallelized experiments in multi-well plates. Remarkably, the capabilities of PDPs to dispense microliter quantities have enabled the miniaturization of many experiments, particularly

for biological assays. The downsides of PDP usage include a limited measurement range, along with challenges in handling e.g., highly viscous liquids and slurries. In contrast to liquid dispensing, dispensing powders or other types of solids presents a more significant challenge for laboratory automation. While automated liquid dispensing benefits from precise robotic volumetric displacement, automated solid dispensing requires real-time measurements of the dispensed quantities, making it more rare and costly. As a consequence, automated laboratories often resort to working with stock solutions of solid reagents when possible.

In any SDL, these basic reagent handling operations are coupled to more problem-specific modules, including reaction execution (in environment-controlled reactors), separation and purification, or device fabrication. Given the large diversity between these modules, they will be discussed in the respective sections of this review. Eventually, the necessary characterization feedback to “close the loop” is provided by the diverse library of analytical instrumentation, which are already used in a (semi-)automated fashion in traditional laboratories, but require dedicated integration into SDL workflows. One popular solution is the static combination of individual modules into a continuous flow sequential workflow connected through tubing. As a particularly prominent example, this strategy has laid the foundation for the field of flow chemistry and microfluidics (as discussed in more detail in section **Reaction Optimization**). Owing to the simplicity of this hardware setup, it has also found applications in a series of higher level SDLs, as discussed throughout the course of this review.

An alternative strategy to statically coupling individual modules is the idea of flexible automation.<sup>40</sup> This approach emphasizes dynamic connections between modules using robotic systems for transfer between different workstations. This approach, imitating a human researcher operating the different modules, is particularly evident in the use of robotic arms, and, for example, has been used in foundational SDLs in drug discovery (Adam<sup>37</sup> and Eve,<sup>38</sup> see section **Drug Discovery and Biochemistry**), and thin-film material synthesis (Ada, see section **Optoelectronics**). The Chempoter by Cronin and co-workers connects different modules, including vessels, reactors, pumps and further specialized units, using selection valves, enabling a diverse range of automated synthetic chemistry workflows.<sup>46</sup> Recently, Cooper and co-workers have extended the concept of flexible automation to the use of mobile robots, operating multiple workstations which are distributed across the laboratory, mimicking a human researcher.<sup>11</sup> The idea of flexible automation has recently spurred commercial solutions, particularly from companies such as Chemspeed Technologies and Unchained Labs, based on gantry systems reminiscent of the HTE platforms discussed above. Despite higher costs, these solutions have garnered considerable interest in both industrial and academic SDLs.

### 2.1.2. General-Purpose Robot Applied for Chemistry.

While specialized chemistry hardware excels in conducting

**Table 1. Comparison of Special-Purpose, General-Purpose, and Open Hardware for SDLs.**

	Special-purpose hardware	General-purpose hardware	Open hardware
Provider	Commercial company	Commercial company	Community
Price	High (~\$1M)	Medium (~\$10K)	Low (\$100 - 2K)
Production-readiness	High (typically shipped in working condition)	Medium (additional cost of software development)	Low (additional cost of software development and hardware building)

predefined experiments, their limited modularity can prove inconvenient for specific SDL configurations. Therefore, the application of general-purpose robotic arms for chemistry has been investigated due to their flexibility and multi-purpose nature. A well-known example of demonstration of general-purpose hardware is the mobile robotic chemist by Burger *et al.*<sup>11</sup> (Figure 2b) In the study, they used a mobile robot arm, capable of moving around a traditional laboratory and operating various instruments, to search for optimal photocatalyst mixtures. They also demonstrated the reconfigurability of the setup, repurposing the system to perform solubility screening and crystallization.<sup>47</sup> General-purpose robots have advantages over purpose-built flow platforms in that they can perform experiments that require physical interaction with tools and objects in the laboratory, thereby minimizing the reconfiguration and/or adaption of proprietary equipment or instruments designed for humans. However, major challenges in perception and decision-making limit the robust deployment of general-purpose robotic systems for flexible lab automation. For this reason, many works in the literature address lab automation for specific tasks—for example, mechanical tasks such as retrieving samples of crystals by scraping the wall of a vial<sup>48</sup> and grinding powder with a soft jig.<sup>49</sup> Pouring liquid using visual feedback<sup>50</sup> and weight feedback<sup>51</sup> have been studied as an alternative method of transferring liquid. Custom hardware built to assist robots in handling liquids have also been proposed, for example, Lim *et al.* used a custom syringe pump operated by a robot arm to conduct a molecule synthesis experiment.<sup>52</sup> Knobbe *et al.* developed a robotic finger for operating electronic pipettes,<sup>53</sup> and Zhang *et al.* used a designed end-effector for operating manual pipettes. Solid dispensing has also been demonstrated using a dual-arm robotic manipulator.<sup>54</sup> Yoshikawa *et al.* demonstrated the use of a robotic arm for the more specific task of polishing electrodes used in electrochemical experiments.<sup>55</sup> Nevertheless, besides the advantages of generality, multi-purpose robotic arm systems are lower in efficiency and hard to parallelize compared to specialized systems.

**2.1.3. Open Hardware for Lab Automation.** The cost of hardware automation is a limiting factor for SDLs. As a means of lowering the hardware cost and crowd-sourcing development and testing, various open hardware for lab equipment has been proposed.<sup>56</sup> Users typically print the published design files with their own 3D printer and build the equipment. In addition to labware for human use, lab automation devices such as liquid handlers have been developed as open hardware. FINDUS<sup>57</sup> is an open-source liquid handling workstation that costs less than US\$400. OTTO<sup>58</sup> demonstrated qPCR with a 3D printed liquid handler. Both systems benefit from readily accessible parts and sensors for error checking, though space efficiency and generalizability remain as challenges. PHIL<sup>59</sup> is a personal pipetting robot that is compatible with microscopes, making it ideal for live cell studies but implementation in chemistry is limited. EvoBot<sup>60</sup> is a reconfigurable liquid handling robot that improved modularity by introducing layers and modules. Building upon a well-established 3D printer technology, it is easy to implement, but the fixed tool design makes it challenging for complicated tasks. Jubilee<sup>61</sup> is an open-source multi-tool gantry-style motion platform also based on 3D printing technology, which has been used to demonstrate liquid-handling tasks for synthesis of nanoparticles (NPs).<sup>62</sup> It can mount/dismount tools automatically to perform multiple tasks, while community contributions are

needed to develop more tools for chemistry applications. Sidekick<sup>43</sup> is a liquid dispenser that features an armature-based motion system with a fully 3D-printed chassis and home-built syringe-pumps to realize lower costs, and can handle only a limited number of liquid identities at a time. 3D printers have been utilized for producing microfluidic devices<sup>63</sup> or building a pipette for a two-finger robot hand to enable accurate liquid handling.<sup>64</sup> Open hardware is beneficial to lowering the cost of building SDLs and their customizability is helpful in meeting individual requirements in different experimental settings which are not met by existing commercial hardware.<sup>65</sup> However, the technical difficulty of setting up open hardware and the wide variety of similar hardware proposals hinder widespread adoption in laboratories other than the developers of the hardware. Further support by user communities is needed in facilitating the adaptation, and efforts in growing user communities have been made using online communication platforms.<sup>66</sup>

**2.1.4. Perception and Computer Vision.** Execution of chemistry experiments autonomously requires several layers of feedback. Mimicking the visual feedback of a chemist's eyes, a perception system should track the progress of the chemistry experiment and provide information to the robot such that it can achieve the high-level goal or direction of a given experiment. Computer vision can play a key role in this aspect. For example, HeinSight can provide perceptual information about the chemistry experiments.<sup>14,67,68</sup> Connected to an experiment planning algorithm, that information could be used to guide the robot throughout the experiment. More recently, Sun *et al.* presented a vision-guided liquid-liquid extraction platform, using image processing and computer vision to identify phase boundaries.<sup>69</sup> In another example, authors used visual feedback to train a 3D-CNN model for viscosity estimation of fluids.<sup>70</sup> At a lower level, the robot also requires visual and kinesthetic perceptual feedback in order to perform manipulation tasks successfully and robustly.

Robots need to be equipped with accurate perception skills to work in unconstrained open workspaces. One of the characteristics in a chemistry laboratory is the use of transparent objects, such as glass containers. Transparent objects have different optical properties from opaque objects that make object detection challenging. Transparent glassware detection algorithms using depth completion<sup>13</sup> and multiple images<sup>71</sup> have been proposed. Public datasets, such as Vector-LabPics dataset,<sup>12</sup> have been published in order to accelerate the development of ML models for laboratory related computer vision.

**2.1.5. Manipulation Skill Learning and Digital Twin.** In SDLs, general-purpose robots can interact with tools, objects, and materials within the workspace and require a repertoire of many laboratory skills. Those tools and objects can be in different forms, for example rigid objects like glassware, articulated objects with joints like cabinet doors, or soft objects like rubber tubes or powders and liquids. Some skills can be completed with existing heterogeneous instruments and sensors in chemistry laboratories, such as scales, stir plates, and heating instruments. Other skills are currently done either manually by humans in the lab or with expensive specialized instruments. In an SDL, robots should acquire those skills by effectively using different sensory inputs to compute appropriate robot commands. To effectively endow the robots with many skills in a scalable fashion, one approach

would be to allow robots to “learn” those skills in a digital twin, a simulated laboratory environment in which the robotic system can interact with, using AI techniques.<sup>72</sup> Digital twins can also be used for testing the workflows, algorithms, and scale-up developments.<sup>73</sup> For example, ChemGymRL<sup>74</sup> was proposed as an interactive framework for reinforcement learning in chemistry.

Some examples of physics-based simulators include Gazebo,<sup>75</sup> MuJoCo,<sup>76</sup> and NVIDIA Isaac Sim.<sup>77,78</sup> Recently, NVIDIA Isaac Lab,<sup>77</sup> a modular framework on top of Isaac Sim, was introduced to simplify common workflows for robot learning that is pivotal for robot foundation model training. Some examples of robotic simulation environments and benchmarks are iGibson,<sup>79</sup> MetaWorld,<sup>80</sup> and BEHAVIOR-1K.<sup>81</sup> Closer to tasks related to laboratory automation, RB2 proposed a robotics simulation benchmark with pouring, scooping, and insertion tasks.<sup>82</sup> In another work, a differentiable environment FluidLab<sup>83</sup> was proposed for simulating complex fluid manipulation tasks. An example of using digital twin for the SDL is provided in Vescovi *et al.*, where simulated environments have been used to visualize and compare tools, verifying the laboratory operations.<sup>74</sup> These simulators leverage different physics engines, such as Bullet,<sup>84</sup> FleX<sup>85</sup> and PhysX<sup>86</sup> and rendering happens *via* OpenGL<sup>87</sup> or Unity.<sup>88</sup> Although robot actions can be trained in a simulation environment at low cost, there are gaps between simulation and real-world settings. Multiple efforts have been made to close this sim-to-real gap,<sup>89</sup> including for chemistry laboratory robotics. For example, Kadokawa *et al.* have trained a powder weighing action in a simulator and realized precise weighing in the real world.<sup>44</sup> Nevertheless, high-fidelity and performant simulation of deformable objects and particle systems (such as fluids and powders), as well as the simulation of chemical phenomena in the context of SDLs, remain areas for future exploration.

**2.1.6. Robotics Safety in Laboratories.** In chemistry labs, several types of safety risks put humans and the environment in danger, including mechanical, electrical, and chemical hazards. Therefore, multiple levels of regulations and guidelines are implemented in chemistry labs for safety and accidents.<sup>90,91</sup>

The presence of robotic systems in chemistry labs can affect the risk of accidents in several ways, necessitating a diligent focus on safety and risk management. Generally, automated experiments with robotic systems inherently create a safer workplace for humans, as the users are less exposed to hazardous materials. Even when autonomous experimentation is not possible, chemists can tele-operate robotic systems to perform experiments with hazardous materials—which has been pioneered, for example, in the handling of radioactive materials,<sup>92</sup> or explosive compounds.<sup>93</sup>

However, particular attention should be given when humans are in proximity to robots or in the same lab space, especially when employing mobile robots for tasks such as sample transfer. The choice of robotic systems, whether collaborative or industrial, can affect the safety protocols. For example, when using industrial robots, safety fences or laser scanners are commonly used in the robot workspace. Ensuring human safety in shared spaces requires a comprehensive approach that encompasses both physical and psychological aspects. For physical safety, the literature advocates employing control and motion planning techniques to facilitate safe physical interactions and address pre- and post-collision scenarios. In

the realm of psychological safety, considerations such as robot motion, speed, adaptability, and appearance play pivotal roles in reducing stress and fostering a sense of safety in human-robot interactions. More information about robotics safety standards can be found in ISO 10218<sup>82</sup>, ISO/TS 15066<sup>83</sup>, and survey papers by Lasota *et al.*<sup>80</sup> and Zacharaki *et al.*<sup>81</sup>

Additional safety issues may arise from the manipulation and perception capabilities of robotic systems, as these remain open problems in the community. When deploying such robotic systems in chemistry labs, if manipulation policies and the robot's decision-making abilities are not robust enough, it may lead to failures, increasing the risk of accidents. An approach to rectify this shortcoming is to consider constraints on the robot policies. For example, Yoshikawa *et al.*<sup>94</sup> used constrained motion planning when transferring liquids with a robotic arm to reduce the risk of spillage. Moreover, the ability of robots to detect accidents, take immediate actions, and notify humans are other important considerations. Overall, safety in laboratories with robotic systems is a multifaceted challenge that requires additional research at several levels, from generating potentially hazardous chemicals to experimental planning and automated experiments where robots are used.

## 2.2. Software

The software component of an SDL is composed of three distinct parts, which are executed by some orchestration software (Figure 3): (1) the control and communication

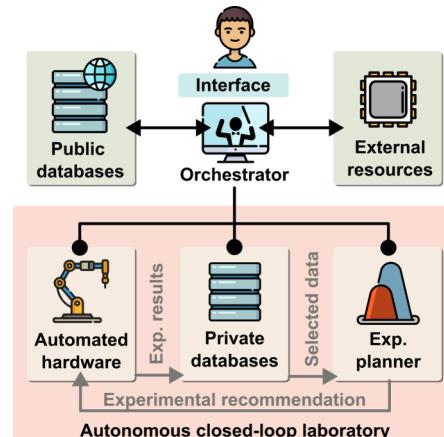


Figure 3. Summary of the software components of SDLs, and how they interface with the hardware, and with each other.

system of the automated hardware of the laboratory, (2) the data extraction, management, and analysis of experimental output, and (3) the decision-making experimental planner. In recent years, the fields of chemistry and materials sciences have undergone a paradigm shift with the rise of AI. ML algorithms, particularly DL models, have proven to be indispensable tools in deciphering complex patterns, predicting chemical properties, and accelerating the design of novel materials with tailored properties.

**2.2.1. Orchestration.** The true potential of individual automated devices in chemistry is most evident when they are interconnected in order to orchestrate some comprehensive chemical tasks. Consider, for instance, the typical course of a chemical analysis, which involves a sequence of actions, including compound synthesis, thorough characterization, and meticulous processing of resulting raw data. The intricate

interplay between these sequential steps underscores the indispensable need for standardized protocols to ensure the coherence and reproducibility of chemical experiments. Traditionally, these protocols were conveyed through research articles and manually executed by chemists. However, contemporary practices enable the translation of these protocols into orchestrated workflows executed by computational software, signifying a pivotal departure from historical methods.

The integration of automated workflows has flourished within the field of computational chemistry, where the automation of repetitive and error-prone tasks is straightforward due to the programmatic nature of the field. This is evident in the emergence of tools such as AiiDA,<sup>95</sup> Fireworks<sup>96</sup> and Snakemake<sup>97</sup> among others,<sup>98–107</sup> which excel in constructing and managing software workflows for *ab initio* simulations. On the experimental side, autonomous chemical laboratories constitute an emerging field where the adoption of such orchestration techniques has been hampered by the physical challenges inherent to wet laboratories, the lack of an orchestration standard, and the scarcity of resources for automated instrumentation.<sup>108</sup> Thus, it is imperative to acknowledge that the development of orchestration software for SDLs faces numerous challenges rooted in the methods used by the majority of researchers. These challenges include:

- The absence of standardized application programming interfaces (APIs) provided by instrument manufacturers, often necessitating the development and utilization of workarounds that place a substantial burden on researchers;
- The inherent software complexity of managing and orchestrating the transporting of items between chemicals processing stations;<sup>41</sup>
- Limited exposure to programming in current chemistry and materials science curricula.

Addressing these challenges requires collective efforts from various stakeholders involved in current research, implementation, and deployment of SDLs. The widespread adoption of SDLs, particularly in industry, will compel manufacturers to provide user-friendly SDL solutions, efficient transfer systems between devices, and graphical user interfaces (GUIs) that facilitate the seamless integration of these platforms into the workflows of chemists.

Numerous tools have surfaced in recent years to bridge this gap, with initiatives like the SiLA2 standard,<sup>109</sup> a communication protocol aiming to replicate a robot operating system (ROS) and adapt it for chemical devices. Within this context, various in-house orchestrators have emerged in different laboratories across diverse chemical fields, with notable examples including ChemOS,<sup>110,111</sup> Helao,<sup>112</sup> and AresOS,<sup>113</sup> among others.<sup>47,74,114–121</sup> These experimental orchestration platforms have achieved significant advancements in key orchestration features that are standard in computationally-oriented platforms, such as queue management, logging capabilities, data handling, and, more recently, the implementation of asynchronous execution of laboratory tasks and their integration with computational frameworks.<sup>122</sup> However, a lack of consensus between these platforms still prevails, and they often remain tailored to specific laboratories, lacking the required level of generalizability to cater to the diverse spectrum of SDLs.

## 2.2.2. Communication and Protocol Management for SDLs.

In the context of SDLs, where human researchers are the intended users, effective communication between researchers and the orchestration manager is paramount. This communication enables researchers to issue complex commands to the orchestrator that will be transformed into chemical operations, while receiving feedback and real-time updates on the status of laboratory processes in a readable format. In this regard, programming languages serve as the essential communication bridge, allowing users to convey instructions and request information from the orchestrator. Although general programming languages are frequently used to program chemistry hardware,<sup>123</sup> for example the MOCCA<sup>124</sup> open-source Python package which directly analyzes HPLC raw data and extracts relevant information, or Chemspyd<sup>125</sup> open-source Python software for communication with proprietary Chemspeed software, specialized programming languages are proposed to efficiently describe chemistry experiments. Chemical Description Language ( $\chi$ DL)<sup>126</sup> is an XML-based language used to describe chemistry experimental procedures, which was demonstrated by translating chemistry literature into  $\chi$ DL, and then synthesizing the described molecules. Chemical Markdown Language<sup>127</sup> is another chemistry domain specific language to describe or assist in experimental documentation. While such languages are more tailored for communicating chemistry specific tasks, they will require a low learning barrier to ensure adoption in other SDLs.

In recent years, there have been multiple examples of asynchronous workflows, in which SDLs operating in separate regions, with different research teams and equipment, work on the same discovery or optimization task. This requires extensive software infrastructure for the communication and coordination of results between the SDLs and the respective research teams. Multiple studies have demonstrated the use of internet cloud servers to manage and control distributed laboratory equipment.<sup>116,128</sup> Decentralized databases can then allow for communication of experimental protocols, experimental results, and coordinated experiment planning over multiple laboratories, which have been demonstrated in some SDL orchestration softwares.<sup>110,111,129</sup> Dynamic knowledge graphs have been proposed<sup>130</sup> and demonstrated<sup>131,132</sup> as an effective way of coordinating distributed SDLs. Ontologies are developed to capture various aspects of chemical research, including reactions, design of experiments, and hardware setups. Software agents are deployed at each lab site and act as executable knowledge components that can query, update, and restructure the knowledge graph autonomously, as the campaign progresses.

Given the recent rapid developments in their capabilities, large language models (LLMs) have been investigated to accelerate chemistry research.<sup>133–136</sup> In terms of communication, LLMs are able to interface with human users through text and conversation, translating between natural and machine language. For example, Boiko *et al.* demonstrated that LLMs can design and perform chemical experiments with a liquid handler based on natural language input from a user.<sup>137</sup> The ability of LLMs can be expanded for specialized use cases by collaborating with external programs. ChemCrow<sup>138</sup> is an LLM specially designed for chemical tasks, being able to observe, plan, and execute actions with integrated chemistry tools. CLAIRify<sup>139</sup> introduced an iterative prompting strategy using automated verifiers to generate  $\chi$ DL, and demonstrated

chemical experiments with a general-purpose robot. Likewise, ORGANA<sup>140</sup> is an experimental planner that uses a LLM to communicate with chemists, and then plans and interfaces with a robotic arm to perform parallel tasks in an SDL experiment. The role of LLMs is discussed further in the subsequent sections.

**2.2.3. Data Management.** Automation accelerates data generation, and the large datasets must be managed efficiently in order to process and disseminate the generated data, particularly for the downstream use in data-driven techniques such as ML. Data management can be categorized into private databases, adept at housing all laboratory-generated data, and public databases designed to share curated and processed data for widespread use.

Individual research laboratories often use private databases to facilitate record-keeping of chemical processes within the laboratory, and track chemical inventory and equipment availability. Traditionally, researchers have relied on laboratory notebooks and inventory software for these purposes, manually recording and annotating changes in experimental procedures. Annotated data would then be transferred for curation and processing, although inconsistent information tracking, and missing or biased data due to human error remain as issues. However, improvements in information technologies have changed data collection practices, with electronic laboratory notebooks emerging as modern alternatives to traditional notebooks.<sup>141,142</sup> Efforts have been made in integrating private databases with SDL orchestration frameworks to keep track of the status of the laboratory<sup>47,110–112</sup> or the status of simulations.<sup>95</sup> However, it's worth noting that the adoption of these tools is not standardized across the chemistry community.

Conversely, public databases play an indispensable role in the open science paradigm, adhering to the FAIR (findable, accessible, interoperable, and reusable) data principles<sup>143</sup> by providing transparent access to experimental data for other scientists, thereby enhancing reproducibility. Computational chemists hold a long tradition of publishing standalone computational databases hosted on cloud platforms like Zenodo.<sup>144</sup> These encompass a broad spectrum of materials, including MOFs, organic molecules, and heterogeneous catalysis. More advanced platforms such as the Harvard Clean Energy Project,<sup>145</sup> IoChem-BD,<sup>146</sup> Materials Project,<sup>147</sup> NOMAD,<sup>148</sup> The Protein Databank,<sup>149</sup> Materials Cloud,<sup>150,151</sup> Open Quantum Materials Database (OQMD)<sup>152</sup> and Catalysis-Hub<sup>153</sup> serve as noteworthy examples of public materials databases. These are typically built on general-purpose database frameworks; for example, Materials Project<sup>147</sup> uses MongoDB and OQMD<sup>152</sup> uses SQL. The variety of public materials databases for computational data typically provide supplementary tools for data parsing, querying, and publishing. However, in the realm of experimental chemistry, there is a lack of tradition in publishing chemical results in structured and open databases; reaction and characterization data are commonly published as standalone datasets or in commercial databases. Notable examples include the Spectral Database for Organic Compounds,<sup>154</sup> Reaxys,<sup>155</sup> SpectraBase,<sup>156</sup> SciFinder,<sup>157</sup> and the chemical reaction patents from the United States Patent and Trademark Office.<sup>158</sup> However, substantial efforts have been made to establish dedicated databases for storing experimental reactions and characterization data, with platforms such as Pubchem,<sup>159</sup> Open Reaction Database,<sup>160</sup> GNPS,<sup>161</sup> Mass Bank of North

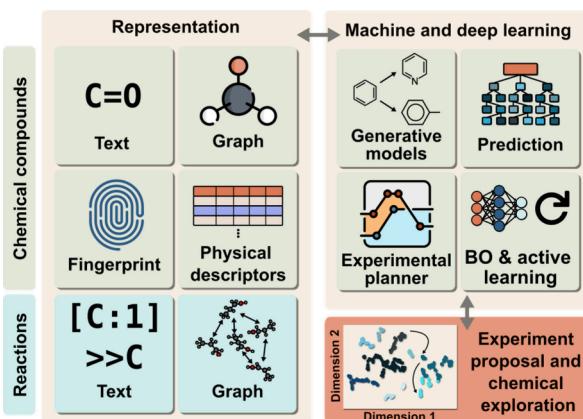
America,<sup>162</sup> Crystallography Open Database,<sup>163</sup> MNRShiftDB<sup>164</sup> and Molar.<sup>165</sup> Due to the automation capabilities of SDLs, these databases are poised to play a critical role in the expansion of SDLs as they serve as a common interface bridging diverse research laboratories, facilitating seamless collaboration and data sharing among geographically dispersed research teams.

Before concluding this subsection, it is worth mentioning the recent emergence of HuggingFace Hub,<sup>166</sup> introducing an open database focused on collecting datasets and ML models. While the effort is currently focused on DL research, the future standardization of SDLs will likely require the adoption of similar solutions. This will enable laboratories to share components of research workflows more effectively. For instance, a synthesis, characterization, and *ab initio* simulation workflow could be assembled for an SDL setup by downloading independent parts from the repository, connecting them, and subsequently customizing them to align with the specific laboratory needs. This collaborative approach facilitates the sharing and improvement of research components among different laboratories, fostering innovation and efficiency within the scientific community, as has been demonstrated for the AI community.

**2.2.4. Role of Artificial Intelligence in Chemical Discovery.** As large datasets of experimental and computational chemical data became accessible, data-driven statistical methods became more relevant to chemical discovery.<sup>167</sup> Cheminformatics have been developed since the 1960s, particularly driven by advances in computing technology, and the development of *ab initio* techniques such as density functional theory (DFT).<sup>168</sup> Early work focused on prediction of chemical properties, identifying quantitative structure-activity relationships (QSAR), for virtual screening of large libraries of pharmaceutical compounds.<sup>169,170</sup> Statistical analysis of feature importance, such as through Shapley additive explanation (SHAP) values,<sup>171</sup> have been used to provide intuition into the effect of certain chemical structures, properties, or experimental parameters in the model performance.<sup>172</sup>

ML methods such as (1) tree-based methods: random forests (RF)<sup>173</sup> or gradient-boosted trees;<sup>174</sup> (2) kernel-based methods:<sup>175</sup> Gaussian process (GP)<sup>176</sup> or support vector machine (SVM);<sup>177,178</sup> and (3) clustering algorithms: k-nearest neighbors (kNN)<sup>179</sup> or k-means clustering,<sup>180</sup> were used to capture complex QSAR in chemical and material space. Chemical compounds can be described by machine-readable chemical descriptors (Figure 4), represented by vectors of physicochemical descriptors,<sup>181,182</sup> unique fingerprints<sup>183,184</sup> (e.g. extended-connectivity, path-based fingerprints),<sup>185,186</sup> graph representations,<sup>187</sup> and structured strings (e.g. SMILES,<sup>188</sup> SELFIES,<sup>189</sup> and group SELFIES<sup>190</sup>).<sup>191,192</sup> More complex forms of chemical representations include 3D information, such as through Z-matrix or cartesian XYZ coordinates.<sup>149</sup> Additionally, chemical transformations can be represented as SMIRKS<sup>193</sup> and SMARTS,<sup>194</sup> which extend beyond SMILES to facilitate the textual representation of chemical reactions, while graph encoding has emerged as a powerful approach for capturing the complexity of reaction networks.<sup>195–198</sup>

More recently, DL methods using neural networks have had successes in chemical applications, with the downside of sacrificing interpretability and requiring large amounts of training data.<sup>199,200</sup> Neural networks are highly expressive non-



**Figure 4.** Depiction of the different encoding techniques (left) and ML/DL models (right) used in SDLs. They are commonly combined to obtain a deeper knowledge of the studied system and acceleration of chemical exploration through experimental proposals.

linear models that can be fit to complex data through backpropagation, capturing complex relationships in high-dimensional input data. DL methods are now state-of-the-art for many chemical prediction and classification tasks, for example, graph neural networks (GNN) on molecular chemical data.<sup>201–204</sup> Additionally, successes in natural language processing have led to LLMs which are able to extract meaning and context from natural language, and generate coherent responses.<sup>205</sup> Molecular language of string representations have been incorporated with language models for property prediction.<sup>206</sup> Various applications of language models to SDLs include allowing for orchestrator-to-human interactions through language, or translating natural language to robotic commands.<sup>139</sup> Language models have also been used to gather data from the scientific literature, generating datasets in an automated way.<sup>207,208</sup>

Additionally, DL allows for data-driven generative modeling and ideation, reaching category 3 in software automation (Figure 1). Generative models incorporating neural networks have been used to generate novel chemical compounds and materials without human intervention, through the use of architectures like variational autoencoders (VAEs),<sup>209,210</sup> generative-adversarial networks (GANs),<sup>211,212</sup> gradient flow (i.e., diffusion) models,<sup>213,214</sup> deep genetic algorithms,<sup>215,216</sup> language models for chemical strings,<sup>217,218</sup> and deep reinforcement learning (RL).<sup>219–221</sup> By directly learning the chemical space of a dataset, the model can interpolate and extrapolate new compounds, and even directly optimize within the latent space through inverse design.<sup>222</sup> Various *in silico* campaigns in generative inverse design and benchmarking have already been demonstrated,<sup>217,218,223–225</sup> but issues with synthesizability and chemical stability of generative compounds remain a barrier to automated empirical validation.

DL methods are capable of transfer learning, a technique commonly used in low-data settings, in which the model is pre-trained with more readily available data that provides the model with implicit information about the main task.<sup>221,226,227</sup> By leveraging the libraries of computational results, and historical empirical results, models can be preconditioned with physicochemical information for SDL campaigns that typically start in the low-data regime. Such models can even be used to encode chemical compounds as task-specific descriptors, compressing the chemical information into expressive abstract

representations.<sup>187,228</sup> Both traditional ML and DL techniques are now commonly used as part of optimization algorithms and experimental planners, which will be discussed in the next section.

**2.2.5. Experiment Planning.** The availability of data, coupled with the robust ML and DL models mentioned earlier, has created a demand for tools adept at processing the datasets. For practical examples, we direct the reader to various reviews that illustrate the utility of these techniques.<sup>199,229–235</sup>

Traditionally, brute force methods like combinatorial grid-search and random sampling<sup>236,237</sup> have been combined with high-throughput techniques to sample systems of interest. For instance, the Haber-Bosch process,<sup>238</sup> which involved testing up to 4000 catalysts in 6500 experiments<sup>239</sup> to identify suitable catalysts and experimental conditions for ammonia synthesis, required significant human effort and resources to complete. While such methods may work well and be preferable when dealing with a small number of parameters and low experimental costs, they quickly become unfeasible as the number of variables increases. In such cases, a methodical approach in the experimental space becomes necessary, particularly when computational or experimentation costs are a concern. Similarly, to make use of the improved precision of modern chemical apparatuses and sample preparation devices, exploring continuous variables in finer increments necessitates an increased number of experiments.

Conventionally, scientists and engineers have used DoE strategies to systematically scan the experimental space, in an effort to reach the optimum, and identify the important parameters.<sup>240–243</sup> A naive approach may be the one-factor-at-a-time (OFAT) design, which involves manipulating a single parameter, assuming there are no correlated effects between the factors. In response surface methods such as Box-Wilson central composite design (CCD) and Box-Behnken design, the experiment list is populated by equally spaced points in design space, followed by polynomial fitting of the parameters to the response variable(s) for creating a response surface. This response surface can be then used for finding the optimum while the fitting parameters can be evaluated through statistical tests such as t-test,<sup>244</sup> or analysis of variance (ANOVA)<sup>245,246</sup> to assess the relative importance of variables as well as to confirm validity of the strategy.

While amenable to the computation power available at the time, a disadvantage of DoE methods is the rigidity of the list of experiments, which remains the same as the results are collected. Furthermore, the equal spacing of selected points gives a diverse yet coarse-grained sample of design space, sacrificing precision in identifying the optimum. As the dimensionality of design space increases, DoE strategies become impractical, and likely insufficient. Especially when targeting software autonomy levels 2 and 3, advanced experiment planning algorithms must be capable of realizing closed-loop workflows. Such iterative global optimization algorithms must fulfill the following requirements: (1) the algorithm must take into account experimental observations from previous iterations, and use this knowledge to make more informed experimental recommendations; (2) as experiments are generally expensive and time consuming, optimization should proceed with the minimum number of required experiments; (3) the algorithm must treat the underlying response surface as a "black-box"—the functional form of the optimization surface, or any gradient information, is usually not available from experiment.

**Table 2. Comparative Overview of Some Common Optimization Software Packages for SDLs**

Name	Published	Type	Natively supported variable types			Natively supported optimization features				Highlights
			Continuous	Discrete	Categorical	Multi-fidelity	Multi-objective	Constrained	Batch	
Nelder-Mead Simplex <sup>253</sup>	1965	Heuristic	Yes	No	No	No	No	No	No	Simple to operate with quick convergence to local minima via updating simplexes.
SNOBFIT <sup>335</sup>	2008	Heuristic	Yes	No	No	Yes	No	No	Yes	Branching for global exploration with local quadratic fit.
HyperOpt <sup>300</sup>	2014	Heuristic	Yes	Yes	Yes	No	No	No	Yes	An easy-to-use optimizer with scikit-learn compatibility.
GPyOpt <sup>301</sup>	2016	BO	Yes	Yes	Yes	No	No	Yes	Yes	Off-the-shelf BO that allows custom surrogate and acquisition functions.
Deep Reaction Optimizer <sup>336</sup>	2017	RL	Yes	No	No	No	No	Yes	Yes	Unique with deep RL offering faster optimization than heuristic approaches.
Phoenics <sup>296</sup>	2018	BO	Yes	Periodic Only	No	No	Yes	No	Yes	Increased sample-efficiency by BNN based kernel density estimation.
TS-EMO <sup>318</sup>	2018	BO	Yes	No	No	No	Yes	Yes	Yes	Bayesian framework with smart sampling to efficiently deal with high-dimensionality less prior knowledge.
Ax/Bo-Torch <sup>305,306</sup>	2019	BO	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Mature library with MC acquisition function, a good base for custom BO libraries.
Dragonfly <sup>313</sup>	2020	BO	Yes	Yes	Yes	Yes	No	Yes	Yes	Emphasis on robustness via adaptive self-hyperparameter tuning along with scalability.
pymoo <sup>337</sup>	2020	Multiple	Yes	Yes	Yes	No	Yes	Yes	Yes	Large framework with multiple algorithms, from EA and BO to simplex optimizers.
CMA-ES <sup>252</sup>	2021	EA	Yes	No	No	No	No	No	Yes	Robust noise handling and self-tuning of internal parameters.
GPax <sup>320</sup>	2021	BO	Yes	No	No	Yes	Yes	No	Yes	Unique with physics-informed guidance of GPs for more robust optimization.
Gryffin <sup>297</sup>	2021	BO	Yes	Yes	Yes	No	No	No	Yes	Improved handling and interpretation of categorical variables.
EDBO+ <sup>295</sup>	2022	BO	No	Yes	Yes	No	Yes	No	Yes	Provides user-friendly web GUI with high performance batch suggestions.
SMAC3 <sup>314</sup>	2022	BO	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Convenient library that performs well on small design spaces.
Atlas <sup>304</sup>	2024	BO	Yes	Yes	Yes	Yes	Yes	Yes	Yes	A comprehensive BO toolbox including real and simulated chemical test data via Olympus.

Some early approaches to the black-box optimization challenges in chemical and material domains is mimicking the successful strategies observed in biology, such as the evolutionary algorithms or genetic algorithms (GA) inspired by natural selection.<sup>247,248</sup> In the context of experiment planning, each experimental setting refers to an individual species in a population. The fitness of each individual, which is associated with the quantities the algorithm is maximizing, is then used to assign a chance of producing offspring via *crossover* operations with other high-fitness species. With each generation, the population evolves to a greater fitness while random variations can be introduced through *mutations* that help prevent the GA from getting stuck in local minima. Similarly, particle swarm optimization (PSO) is an optimization tool inspired by flying flocks of birds where each particle represents a point in experimental design space and the velocities of the particles are analogous to update rates of the experiment parameters.<sup>249–251</sup> The covariance matrix adaptation-evolution strategy (CMA-ES) is another evolutionary strategy, in which the species of a generation are selected by a probability distribution based on high-fitness individuals.<sup>252</sup>

Many of today's pressing challenges such as catalyst discovery for sustainable energy applications, drug discovery, and synthesis optimization are analogous to finding a needle in the haystack, where only some narrow areas of the experimental space are highly promising and require detailed

exploration. While GAs and PSO have proven suitable for complex problems, such methods tend to require a large number of samples as the design space grows. Simplex optimization, notably the modified simplex,<sup>253,254</sup> has been another common heuristic optimization strategy that is also relatively simple and straightforward. Without any mathematical assumptions, simplex optimization gradually and systematically alters a virtual simplex constructed by vertices of previous experiments within the experimental design space, shrinking and expanding towards optimal settings with each subsequent evaluation. Despite multiple successful applications in chemistry<sup>255–263</sup> since its first use in analytical chemistry<sup>27</sup>, simplex optimization's final performance may be hindered by local optimality traps, noisy data and large number of variables. Another common systematic approach for experiment planning is Stable Noisy Optimization by Branch and Fit (SNOBFIT)<sup>35,264–266</sup> which uses branching for exploration and bounding for elimination of irrelevant parts of experiment design space. SNOBFIT further aims to improve optimization efficiency by including local searches and deals with noisy data via robust sampling combined with statistical modeling of the noise. Finally, gradient-based numerical methods for optimization have also been used in chemical process optimization, such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm.<sup>267,268</sup> These algorithms require more computational power, but typically converge faster than non-gradient

methods, like SNOBFIT or simplex optimization.<sup>269</sup> We provide a comparison of some common experimental planning algorithms in Table 2.

There are a number of aspects to consider when designing or choosing experimental planning algorithms in the context of SDLs. For example, SDLs in chemistry often involve the simultaneous optimization of continuous, discrete (i.e., ordinal discrete parameters), and categorical parameters. For example, the optimization of a reaction may involve parameters which include continuous variation of temperature, discrete increments of reaction times, and categorical selection of reagents. In scenarios where there are small amounts of data, or when experiments are too expensive to perform, multi-fidelity optimization may be used to accelerate optimization. Multi-fidelity learning leverages information from both low-fidelity (cheaper, faster, but less accurate), such as *ab initio* calculations, and high-fidelity (expensive, slower, but more accurate) data sources to guide the search for optimal conditions. This can be done, for example, through transfer learning with DL methods,<sup>270–273</sup> or delta-learning optimization.<sup>274–276</sup> And in many automated experimental platforms, batch experimentation is typically used to parallelize synthesis or characterization, such as through the use of well plates. Experimental planners need to be able to optimally suggest batches of solutions while balancing exploration and exploitation of the search space. This is also known as the multi-armed bandit problem.<sup>277–279</sup>

Multi-objective optimization is particularly important in the optimization of chemical processes, and the design of new materials or molecules—for example, maximizing the activity of drug molecules, while targeting a specific solubility. Scalarizing functions are one of the simplest ways to incorporate multiple objectives into a single objective, with the simplest function being a weighted average of the various targets. Other more sophisticated scalarizers have been proposed, such as Chimera,<sup>280</sup> which allows for user-specified hierarchical optimization of each objective, or the Pareto front hypervolume. The Pareto front represents the set of solutions where no single objective can be improved without degrading at least one other objective. Maximizing the hypervolume of the Pareto front yields a set of solutions that dominates a larger portion of the objective space, indicating better overall performance across all objectives and providing decision-makers with a diverse range of trade-off solutions. Some such methods include: parallel efficient global optimization (ParEGO),<sup>281</sup> which uses the Chebyshev scalarization; non-dominant sorting genetic algorithm (NSGA)-II,<sup>282</sup> and NSGA-III,<sup>283</sup> which are GA approaches that optimizes by considering crowding distances of solutions within the Pareto front; symmetric evolutionary multi-objective optimization algorithm (SMS-EMOA),<sup>284</sup> which directly optimizes the hypervolume indicator. For more detailed discussion, we refer readers to Sharma and Kumar,<sup>285</sup> Rangaiah *et al.*,<sup>286</sup> Vel *et al.*,<sup>287</sup> and Angelo *et al.*<sup>288</sup>

### 2.2.6. Bayesian Optimization and Active Learning.

One of the most common strategies today for SDL experimental planning is Bayesian optimization (BO), which aims to maximize or minimize some black-box function, such as a measurable chemical property, as a function of controllable experimental parameters.<sup>289,290</sup> To do this, a surrogate model with some prior distribution is fit, or trained, on the available data, and the predicted posterior distribution is used to create an acquisition function. The acquisition function contains

information about the prediction and the uncertainty of the prediction based on the posterior distribution, and can be used to control the exploitative or explorative nature of the subsequent experiments. Some common acquisition functions include the upper confidence bound (UCB), a linear combination of the mean and standard deviation of the posterior, and the expected improvement (EI), the expectation value of the next point that most improves upon the best value. In multi-objective optimization tasks, some commonly used acquisition functions include the expected hypervolume improvement (EHVI),<sup>291</sup> the noisy EHVI (NEHVI),<sup>292</sup> or ParEGO.

BO algorithms have since incorporated more complex non-linear models using ML and DL models as surrogates, and have found success in more difficult chemical optimization tasks. Similarly, active learning is a sequential optimization algorithm, however the goal is to improve the performance of the surrogate model with each additional data point. The goal of both algorithms is to optimize to the respective goals in as few evaluations as possible, minimizing the number of expensive black-box function calls. This entire loop of experimentation, model training, and decision-making is repeated until a given experimental budget is reached or until the given target is achieved.

While any model can serve as a surrogate model, GPs and RFs are typically chosen, as they train quite well with small amounts of data, which is typically the case in the early iterations of an SDL campaign. While GP models learn distributions of functions and intrinsically provide uncertainty estimates, RFs provide an uncertainty based on the ensemble of decision trees. DL surrogate models like multi-layer perceptrons (MLPs), convolutional neural networks (CNN), recurrent neural networks (RNN), or GNNs are less common in SDL applications, and usually require pre-training from larger datasets, (e.g., computational datasets). However, probabilistic DL models such as Bayesian neural networks (BNNs)<sup>293</sup> have been shown to work relatively well with lower amounts of data, due to the regularization effect of the neural network weight distributions.

Given the versatility and recent success of the BO methods for experiment planning, there has been an intense effort to provide software libraries targeting multi-objective optimization, compatibility with specialized material/chemical optimization tasks, better handling of categorical variables, increased accessibility as well as benchmarking. Shields *et al.* introduced an open-source Python package Experimental Design via Bayesian Optimization (EDBO)<sup>294</sup> and provided multiple benchmarks. The authors showed BO with GP performs statistically better when compared to common DoE techniques suitable for both continuous and discrete variables in maximizing Suzuki and Buchwald-Hartwig reaction yields. While EDBO was capable of recommending experiments in batches with seemingly no performance loss, Tores *et al.* published EDBO+<sup>295</sup> that extended the platform to tackle multi-objective tasks. They further augmented the platform with a cloud powered web-interface to provide accessibility to non-coder scientists as well.

Häse *et al.* developed the Bayesian optimizer Phoenics<sup>296</sup> that addressed the issue of large numbers of samples required for chemical global optimization tasks, particularly where evaluation of a point in chemical design space is costly. By utilizing an autoencoder-like BNN for kernel density estimation from observations, a surrogate function can be

constructed with higher efficiency. Phoenics was benchmarked on a reduced Oregonator model for the Belousov–Zhabotinsky reaction, and, when compared to RF, GP, PSO, and CMA-ES, Phoenics outperformed the other methods after only 25 evaluations. Later on, Häse *et al.* introduced Gryffin,<sup>297</sup> an extension of Phoenics with a particular focus on handling categorical variables. Additionally, Gryffin considers the correlation among the variables through the use of descriptors; for example, the physicochemical descriptors of selectable solvents. Among multiple optimization strategies and packages such as PyEvolve,<sup>298</sup> SMAC,<sup>299</sup> HyperOpt,<sup>300</sup> and GPyOpt,<sup>301</sup> the authors reported the best performance with the Gryffin optimizer. Other chemistry specific BO algorithms include Gemini,<sup>274</sup> which extends Gryffin to multi-fidelity optimization, Golem,<sup>302</sup> which identifies optima that are robust to input or measurement uncertainties, and Anubis,<sup>303</sup> which incorporates unknown experimental constraints into the acquisition function. Recently, Hickman *et al.* introduced Atlas,<sup>304</sup> an open-source software library incorporating the functionalities of the aforementioned BO softwares, including mixed parameter BO with *a priori* known and unknown constraints, as well as multi-objective, multi-fidelity and robust optimization capabilities. With an emphasis on integration with SDLs, the authors showcased the Atlas library embedded in the ChemOS 2.0<sup>111</sup> SDL orchestrator for oxidation potential optimization of metal complexes using electrochemical measurements.

It is also worth noting that there are many off-the-shelf and general-purpose BO packages available. BoTorch<sup>306</sup> is a modern BO library built on top of PyTorch<sup>307</sup>, offering modular components, Monte Carlo (MC) acquisition functions, and a variety of advanced optimization features such as high-dimensional, multi-fidelity, multi-objective, mixed-variable, and constrained optimization. Notably, the Ax (Adaptive eXperimentation) platform is a high-level wrapper to BoTorch managed by the same developers which significantly reduces the required learning curve and has seen recent usage in materials informatics.<sup>308–310</sup> GAUCHE, by Griffiths *et al.*,<sup>311</sup> is an open-source GP framework built atop GPyTorch,<sup>312</sup> BoTorch<sup>306</sup> and RDKit,<sup>184</sup> with a suite of custom kernel functions for GPs, and built-in performance and BO benchmarks for molecular and reaction discovery. Dragonfly<sup>313</sup> is an optimization library for handling both multi-fidelity and high-dimensional optimizations, emphasizing adaptability to various domains. SMAC3 (Sequential Model-based Algorithm Configuration)<sup>314,315</sup> combines BO with RFs, and is particularly suited for algorithm configuration tasks. GPyOpt<sup>301</sup> utilizes GPs as surrogate models for BO, offering a variety of GP-based acquisition functions. HEBO<sup>316</sup> (Hierarchical Evolutionary Bayesian Optimization) integrates evolutionary strategies with BO, presenting a hierarchical approach to enhance search efficiency. HyperOpt<sup>300</sup> is a Python library for serial and parallel optimization over challenging search spaces, utilizing techniques like tree-structured Parzen Estimator (TPE)<sup>317</sup> rather than traditional BO. Thompson sampling efficient multi-objective optimization (TS-EMO)<sup>318</sup> is a general-purpose BO with a GP surrogate allowing both multi-objective optimization and batch suggestions.

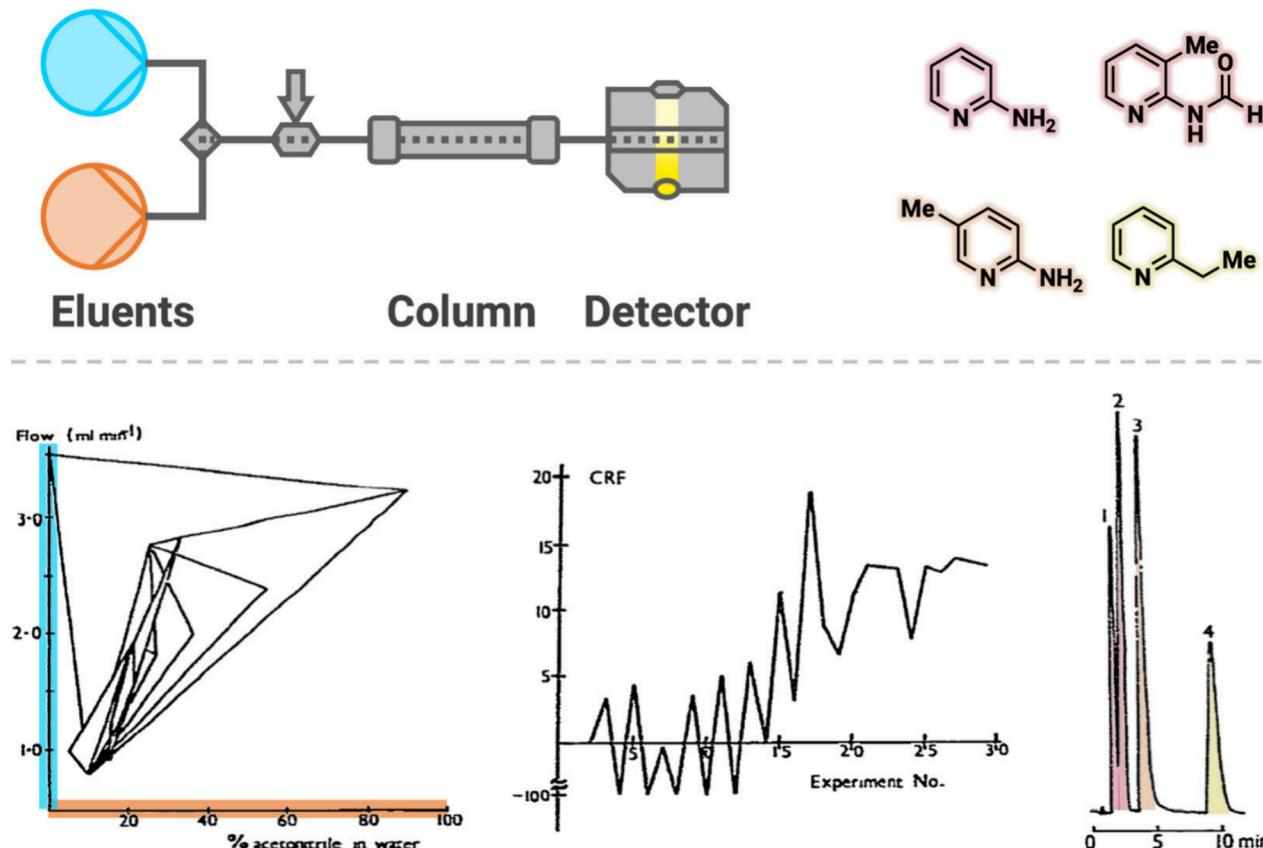
While off-the-shelf models are convenient, studies have shown that careful selection of surrogate models can greatly influence the performance BO or active learning campaign. For example, for the GP surrogate, Noack and Reyes stress the importance of understanding the physical system in selecting

hyperparameters, kernel and mean functions.<sup>319</sup> Many studies commonly use standard kernels like the radial basis function, or Matérn kernels, without considering the underlying physics of the system, anisotropy in the input features, or stationarity of the data points. Ziatdinov *et al.* proposed GPax,<sup>320</sup> a novel approach that augments GPs with structured probabilistic models to incorporate prior physical knowledge into BO and active learning tasks. Unlike standard GP-based BO, GPax balances the flexibility of non-parametric GPs with the rigidity of parametric models encoding known physical behaviors. The authors demonstrate GPax's capabilities on synthetic test functions, as well as physical lattice models like the 1D and 2D Ising models, where it outperforms classical GPs in discovering optimal regions and reconstructing phase boundaries with fewer observations. Further studies have demonstrated the potential for GPax in improving optimization in high-throughput experimental studies,<sup>321–323</sup> increasing explainability of the surrogate model in hypothesis learning.<sup>324–327</sup>

There can be significant performance variance between experiment planning approaches for different chemical and material optimization/engineering tasks, even for slightly different tasks within the same domain.<sup>328</sup> Therefore, the benchmarks have been developed to evaluate the various methods, which can be useful when initially choosing an experiment planning algorithm and the associated hyperparameters. Olympus<sup>329,330</sup> and Summit<sup>331</sup> are examples of BO benchmarking platforms with realistic chemical tests and experiments. For finding efficient black-box approaches, Tom *et al.* studied the effect of different chemical featurizations and surrogate models on the predictive performance and uncertainty calibration on different small chemical datasets, and the optimization performance in the context of BO.<sup>332</sup> Liang *et al.* benchmarked different BO flavors specifically on the materials science datasets covering a wide domain ranging from electrical conductivity of drop-casted composite blends to shape scores of 3D printed materials.<sup>333</sup> The authors further defined useful metrics for evaluating the acceleration and overall performances of optimization. Rohr *et al.* compared performances of linear ensembles (LEs), RFs and GPs for active learning based minimization of multi-metal oxide catalyst compositions' overpotential toward oxygen evolution reaction.<sup>334</sup> We summarize some commonly used BO tools for experimental planning in Table 2.

### 3. ANALYTICAL PROCESS OPTIMIZATION

The earliest examples of chemistry SDLs that automatically perform a sequence of experimental tasks, planned by a data-driven algorithm in a closed loop, largely stem from the field of analytical chemistry. As early as in the 1960s, the optimization of measurement parameters to maximize e.g., the response of a single instrument, has been addressed in an iterative closed-loop fashion. Whereas these approaches do not fall under the scope of this review, and are routinely implemented into modern (analytical) instruments, it is remarkable that this iterative optimization, taking into account data from previous iterations (in contrast to e.g., PID controllers) has already been achieved more than 50 years ago. As an example, Ernst *et al.* demonstrated the autonomous optimization of magnetic field homogeneity in a nuclear magnetic resonance (NMR) spectrometer by adjusting currents along the spinning axis, controlled by a gradient-based and a simplex-based algorithm.<sup>24</sup> To the best of our knowledge, this work represents the



**Figure 5.** Autonomous optimization of chromatographic separation by varying the eluent composition. Top: Schematic visualization of the experimental setup used by Berridge, and structures of four analytes in a sample mixture. Bottom: Visualization of the simplex optimization performance upon variation of eluent composition and flow rate (left). Development of the chromatographic response function (CRF), the objective, throughout the course of the optimization campaign (middle); the optimized chromatogram (right). Figure adapted with permission from reference.<sup>258</sup> Copyright 1982, Elsevier.

first published example of a simple Level 2 SDL in the field of chemistry.

Level 3 SDLs are realized when coupling an analytical technique to an upstream operation such as automated sample preparation or separation. In these setups, the goal definition of the SDL is the identification of those process conditions that optimize the detectability or quantifiability of specific materials. Whilst these tools can be regarded as components of larger SDLs in materials discovery, the development of robust analytical methods has been an active field of research for multiple decades, and has been addressed using SDL approaches early on. In this section, we will review the autonomous optimization of such analytical processes, and further related experimental procedures. Note that SDLs which address the condition optimization of chemical reactions are not included, but will be discussed in the section on **Reaction Optimization**. SDLs that target the conditions for the formulation of drugs are summarized in section **Drug Discovery and Biochemistry**.

### 3.1. Composition and Detection Process Optimization

The identification of ideal measurement parameters for analytical processes is of enormous importance across all chemical industries, and has attained considerable attraction from the standpoint of autonomous optimization. Beyond simple PID-type controllers, however, due to the limitations of computational power and automated laboratories before the 2000s, experimental planning was done primarily by simplex

optimization, and automation typically done through simple step motors and flow systems. Detailed reviews on these systems are provided by Deming and Parker,<sup>28</sup> Rozycki,<sup>338</sup> and Bezerra *et al.*<sup>254</sup> In the following sections, we will provide a brief overview of SDL progress for analytical processes, with a focus on the more recent studies.

**3.1.1. Sample Preparation.** The earliest examples of what would be considered a Level 3 SDL as per the definition of this review, have been reported in the 1970s, addressing the optimization of analytical procedures for spectroscopic detection. Typically, in these works, the optimization would target the automated sample preparation for a subsequent spectroscopic detection, in order to maximize the spectroscopic response for the material of interest.

One of the first examples of automated optimization of chemical detection was from King and Deming in 1974.<sup>339</sup> The SDL used a rotating motor with attached pumps to dispense various reagents into a flow system. The chemical system studied was the acid-dependent chromate–dichromate equilibrium. Characteristic peaks in the measured UV-Vis absorption spectra correspond to an equimolar solution. Using simplex optimization, the intensity of the characteristic absorbance peak was maximized as a function of the chromate pump speed over the course of 26 automated experiments.<sup>340</sup> Following this initial work, a series of similar publications appeared that used automated continuous flow analyzers, available commercially as AutoAnalyzer from the Technicon Corporation, to

perform Simplex optimizations over multiple variables for the detection of glucose.<sup>341,342</sup>

Mieling *et al.* developed a more sophisticated flow system by introducing automated flow-stopping, controlled by a magnetic-tape minicomputer with an analog-to-digital converter.<sup>343</sup> The flow stopping method allowed for automated solution preparation. The authors demonstrated the capabilities of the platform by creating a series of solutions of varying concentrations. The platform was then used in the detection of titanium through its reaction with hydrogen peroxide in the presence of ethylenediaminetetraacetic acid (EDTA), monitored by absorption spectroscopy and optimized by simplex optimization. Further advances in analytical laboratory automation led to works involving the Zymark robotic arm, capable of sample preparation, solution addition, and sample transfer into a spectrophotometer. Lochmüller *et al.* utilized this SDL to optimize the concentration of MgIn<sup>-</sup>, monitored by UV-Vis absorption spectra, through a reaction that is dependent on Ca<sup>2+</sup> ion concentration and pH.<sup>344</sup> Lochmüller and Lung used a similar system with a robotic arm to detect phosphate through the molybdenum blue reaction.<sup>345</sup> Both studies used simplex optimization for experimental planning. More recent studies of simplex optimization have included various spectroscopic methods and more automated methods, such as sequential injection analysis,<sup>255,346</sup> stopped-flow analysis,<sup>347,348</sup> and flow injection analysis.<sup>349–352</sup> Further applications of simplex optimization to analytical chemistry questions have been reviewed elsewhere.

**3.1.2. Separation and Chromatography.** The most prominent and widespread class of coupled analytical techniques are chromatographic methods, in particular gas chromatography (GC) and liquid chromatography (LC). While, as of today, commercial instruments are sold as integrated solutions, the underlying process is composed of two major operations: first, separation of the analytes occurs on a stationary phase column, and the mobile phase stream is transferred to a downstream detector (most commonly, flame ionization detection (FID) in GC, absorption spectroscopy in LC) for measuring the analyte response as a function of time. In this context, identifying the right conditions that enable good and efficient separation of unknown compound mixtures, represents a major challenge, and has been an active research question in the field of chemometrics for over 50 years.

Foundational work towards the use of data-driven algorithms, particularly the simplex algorithm, for chromatographic separation optimization had been performed in the 1970s.<sup>27,261,353</sup> Shortly after, the first autonomous examples of separation optimization for high-performance liquid chromatography (HPLC) were reported by Berridge in 1982.<sup>258,354,355</sup> Using the simplex algorithm for experiment planning, the authors optimized the eluent composition to maximize the chromatographic resolution for mixtures of 4–5 organic compounds, detected on a single-wavelength UV spectrometer (Figure 5). Building on this foundational work, Berridge and co-workers reported a series of further advancements, including multi-parameter optimization,<sup>257</sup> constrained optimization,<sup>257</sup> multi-wavelength detection,<sup>262</sup> as well as case studies in drug manufacturing.<sup>356</sup> A similar study for ion chromatography was reported over 30 years later in 2015.<sup>357</sup> Key to success in all of these works was a “comprehensive, two-way communication between all units of the chromatograph and the computer controller,”<sup>358</sup> which allowed the authors to set-up a fully autonomous Level 3 SDL. As discussed in the

Software subsection, with increasing instrument complexity, and the economical driving force of proprietary software commercialization, such two-way communication interfaces have become rare, which poses a major barrier to the development of SDLs.

As a result, advances on SDLs for chromatographic condition optimization have stagnated in the early 1990s. Over 20 years after Berridge’s work, O’Hagan *et al.* reported an automated framework for optimizing separations in gas chromatography-mass spectrometry (GC-MS).<sup>359</sup> It is worth noting that the authors did not have access to a “comprehensive two-way communication” to control their system programmatically, but rather wrote a software to operate the supplier’s GUI by imitating mouse and keyboard inputs. The authors succeeded in optimizing multiple objectives including the signal-to-noise ratio, the number of peaks and the run time by automatically varying injection volume, flow rate and temperature gradient. All experiments were planned using a multi-objective genetic optimization algorithm. Using a similar framework, the authors expanded their work to two-dimensional GC-MS (GCxGC-MS)<sup>360</sup> and ultra-high performance LC (UPLC).<sup>361</sup> However, in all of these works, a custom GUI-controlling software was used. In order to circumvent this issue, and to develop a generalizable framework for self-driving chromatography-mass spectrometry (MS) systems, Jenkinson *et al.* developed MUSCLE as an optimization framework in which interfacing to the instrument is performed by user-defined *visual scripts* that standardize the imitation of mouse and keyboard commands. This software, however, has only been rarely used, for example in metabolomics separation development.<sup>362</sup>

As a notable exception from the trend of automating the use of GUIs in the early 21st century, I *et al.* reported a self-driving LC system in which direct programmatic control over the instrument is available.<sup>363</sup> Interestingly, the authors follow a fundamentally different software approach, relying on an expert system in which knowledge and heuristics are encoded into a decision tree algorithm. Roch *et al.* later presented ChemOS for the orchestration of SDLs.<sup>110</sup> In the work, the authors performed proof-of-concept automated experiments such as optimizing the color, pH, and density of mixture solutions. Additionally, ChemOS was able to coordinate a robotic arm system with an HPLC remotely, optimizing the parameters for maximal response from the chromatograph.

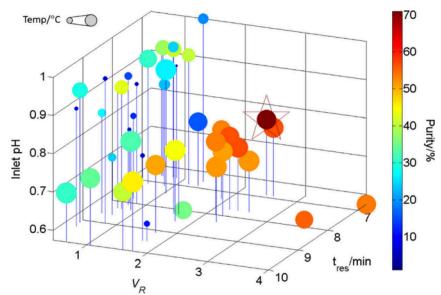
Very recently, Boelrijk *et al.* show the use of BO tools for a fully autonomous optimization of multi-step gradients in HPLC.<sup>364</sup> Using a multi-objective strategy for simultaneously optimizing resolution and elution time, the authors showcase the autonomous development of separation gradients for complex dye mixtures consisting of up to 50 different analytes while only performing ~30 experiments. Importantly, the use of a multi-objective BO approach is shown to be superior to a simple scalarization of multiple targets into a single objective, emphasizing the importance of advanced multi-objective algorithms for efficient navigation of chemical spaces.

### 3.2. Other Properties

Similar to the discussed works on automated sample preparation, other properties such as solution properties of pH or solute concentrations (for crystallization) have been addressed using automated liquid handling systems, instructed by data-driven algorithms in a closed-loop fashion. Solid state properties such as X-ray diffraction (XRD), and small-angle X-

ray scattering (SAXS) signals have also been optimized in a closed-loop manner.

As a prominent example, Clayton *et al.* used an automated flow system SDL to determine the solvent volume ratio and pH for liquid-liquid extraction.<sup>365</sup> The authors studied the separation of  $\alpha$ -methylbenzylamine and N-benzyl- $\alpha$ -methylbenzylamine dissolved in toluene, while the solvent and hydrochloric acid flow rates, temperature, and the residence times were varied. The outputs from the separator were analyzed by an on-line HPLC, with the goal of maximizing the amine purity for the two compounds. Rather than using simplex optimization, the SNOBFIT algorithm was used for this experiment. To extend into multi-step process development, the liquid-liquid extraction was performed in tandem with the reaction of  $\alpha$ -methylbenzylamine and benzyl bromide to form N-benzyl- $\alpha$ -methylbenzylamine. The purity of the product was the optimization objective, with the reaction mixture containing unreacted reactants and various amine-containing impurities. The optimal purity was at 71% which was identified in 53 experimental iterations (Figure 6).



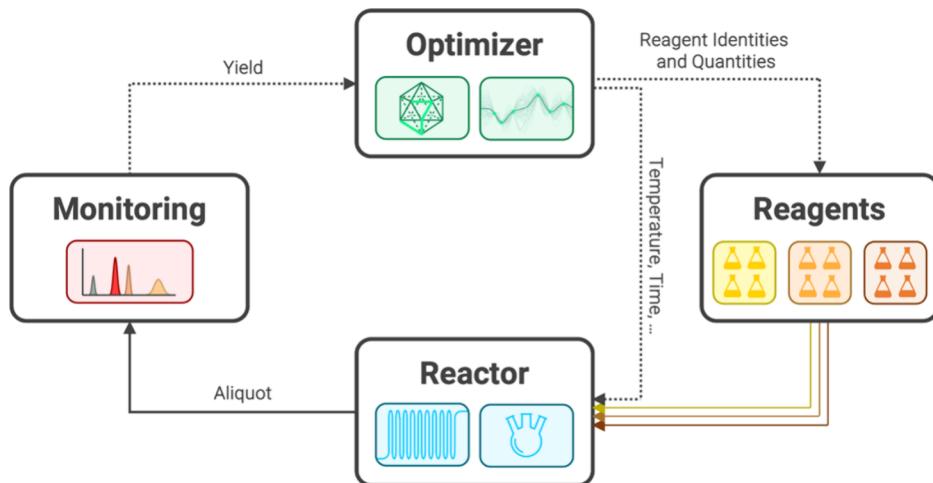
**Figure 6.** Purity optimization results in multi-step reaction-extraction process. Four parameters are varied: the temperature, solvent ratio ( $V_R$ ), the residence time, and the inlet pH. The size of the dots corresponds to the temperature, and the color represents the purity of N-benzyl- $\alpha$ -methylbenzylamine after the reaction. The star is the optimal purity obtained at experiment 53. Figure reproduced with permission from Clayton *et al.*<sup>365</sup> Copyright 2020, Springer Nature.

Pomberger *et al.* use BO to optimize pH of a solution using a robotics-integrated flow-platform.<sup>366</sup> While pH adjustment systems exist through control systems like proportional-integral-derivative (PID),<sup>367</sup> information from previous iterations are not incorporated in preparation of a solution. In this study, the authors used a BO framework to plan experiments, and compared the performance of linear regression, RF, GP, and an MLP predictor. The SDL starts with a buffer solution and some random amount of HCl and NaOH. Subsequent amounts of HCl and NaOH were decided by a greedy algorithm. Aiming for a target pH of 6, the SDL was able to converge in the fewest iterations using a GP surrogate model, while the linear regression model performed the worse, since the pH response is non-linear with the amount of acid/base. Chitre *et al.* further extended the above work through the use of a closed-loop robotic titration platform, capable of handling viscous solutions of unknown compositions.<sup>368</sup> The pH-bot used a translating stage with a probe head, capable of dispensing acid/base, measuring pH, and stirring the solution. For experimental planning, a GP surrogate was used, with the calculated effective strength of the acid/base incorporated into the feature space. The probability of improvement acquisition function was used to close the loop.

In addition to the optimization of solution-phase processes, recently, processes with surfaces, films, and solid crystalline materials have also been optimized in an SDL. Noack *et al.* presented SMART (Surrogate Model Autonomous expeRimenT)<sup>369</sup> for autonomous exploration of multi-dimensional parameter spaces in scientific experiments. Kriging,<sup>370</sup> a GP regression technique originating from geostatistics, is used to construct a surrogate model that fits the available experimental data and provides an associated error function. A GA iteratively maximizes this error function in an active learning campaign, suggesting the next sample location that maximally reduces model uncertainty. The method is validated on synthetic test functions, showing faster convergence to underlying test functions. The authors further applied SMART to autonomous SAXS experiments on block copolymer thin films and nanoparticle coatings. SAXS is a technique that measures the elastic scattering of X-rays by a sample at very low angles, providing information about the size, shape, and spatial arrangement of nanoscale and mesoscale structures. By integrating the SMART algorithm with beamline control, data acquisition, and analysis software, the authors demonstrate fully autonomous experiments without human intervention. The gradient of the surrogate model can be incorporated to emphasize high-gradient regions, enabling efficient reconstruction of sample features like edges. The SMART approach outperforms traditional grid-based and random sampling methods, rapidly converging to accurate approximations of the underlying experimental data while minimizing the number of measurements required.

The authors extended this method to identifying ordered regions of gold nanorod films,<sup>371</sup> studying the effects of fabrication parameters which varied spatially along the orthogonal axes of the film in the SAXS measurement. They demonstrate the importance of accounting for inhomogeneous measurement noise and anisotropy in GP surrogate decision-making algorithms for physical systems. The film was fabricated using a flow-coating method with gradients in coating velocity and substrate surface energy varying along the orthogonal axes of the film, enabling exploration of the effects of these parameters on the self-assembled nanoscale structure. The GP regressor, accounting for non-identically distributed noise levels and anisotropic correlation lengths in different parameter directions, guided the selection of subsequent measurement points to efficiently map the film's structure across the parameter space. The approach enabled identification of well-ordered nanorod regions within the first few hours, demonstrating the benefits of incorporating measurement noise heterogeneities and anisotropies for efficient autonomous experimentation.

Liu *et al.* presented an autonomously driven experimental workflow for scanning probe techniques, such as piezoresponse force microscopy (PFM) and spectroscopy (PFS).<sup>322</sup> In scanning probe techniques, an atomically sharp probe is raster scanned across a surface to measure interactions with the sample, providing a spatially-resolved mapping of the measurement. PFM was used to identify the ferroelectric domains in  $\text{PbTiO}_3$  thin films. In their study, the authors combine ML algorithms with physics insights to actively guide the microscope towards locations of interest for PFS, in which the sample-tip voltage is varied to generate a hysteresis loop. They use a deep kernel learning (DKL) approach, which uses a deep neural network to project high-dimensional data (e.g., domain structure images) into a low-dimensional latent space,



**Figure 7.** General components of a closed-loop system for reaction condition optimization, consisting of an automated reactor system for automated reaction execution, a monitoring system for quantifying the reaction outcome, and an optimization algorithm for decision-making.

where a GP establishes correlations between the domain structure and polarization-switching characteristics encoded in the hysteresis loops. The workflow iteratively acquires hysteresis loops, updates the DKL model, and selects the next measurement location based on the acquisition function and associated uncertainties. Liu *et al.* also utilized a similar autonomous workflow for the PFM study of ferroelectric domain writing on BaTiO<sub>3</sub> thin films, using Bayesian inference and GPs.<sup>325</sup> The physics informed surrogates were used for hypothesis testing in elucidating mechanisms of ferroelectric domain growth. The proposed methodology is generalizable to various scanning probe modalities, and electron microscopies, enabling autonomous experiments for investigating femtolitre-scale structure-property relationships in functional materials relevant for applications to solid state materials development.<sup>372</sup>

Szymanski *et al.* proposed ML guided XRD measurements in a closed loop.<sup>373</sup> The experiment starts with an automatic scan of 2θ between 10°–60°. With the user-defined elements that make up the chemical space, the algorithm, pre-trained on all Inorganic Crystal Structure Database (ICSD)<sup>374</sup> extracted phases, identifies probable phases within the chemical space and assigns each probable phase a confidence score, using a CNN. The confidence score enables the algorithm to identify the most informative regions, where the difference in confidence between different phases is highest. The algorithm guides the diffractometer to make additional scans, either with higher resolution in known regions of the spectrum or in new regions of the spectrum. The authors utilized their platform to demonstrate a higher detection rate for phases with low weight-percent compared to conventional XRD measurements in both the Li-La-Zr-O and Li-Ti-P-O chemical spaces, respectively. Further, the authors also demonstrated improved *in situ* XRD measurements, relevant for the synthesis of battery materials, where the adaptive algorithm strikes a balance between speed and accuracy, allowing to identify phases during the reaction that were not observed with either fast (high speed, low accuracy) or slow (high accuracy, low speed) conventional XRD approaches. The authors later incorporated this work into an SDL platform for synthesis of metal oxides and phosphate powders, discussed in [Solid state materials synthesis](#).

#### 4. REACTION OPTIMIZATION

For over a century, materials discovery has been governed and constrained by the ability to synthesize chemical compounds, and make materials. This applies to molecular discovery (drug discovery, agricultural chemistry, molecular optoelectronics) in particular, where synthesis often represents a tailored sequence of highly specific reaction steps, each of which comes with a set of variable parameters and process conditions. Both the discovery of optimal synthetic routes and the optimization of optimal reaction conditions for each step is therefore critical to all fields of materials discovery, and chemical industries. In fact, the industrial need for economic and ecological synthesis processes has led to the discipline of process and reaction chemistry, targeting route identification, reaction conditions, as well as design and engineering of reactors for synthesis on scale.

Due to the importance of reactions in chemistry and chemical applications, a range of closed-loop workflows for reaction optimization have been developed over the last decades, particularly targeting the identification of optimal reaction conditions. This section aims to summarize these literature-known approaches and examples of self-optimizing reactors. In contrast to the following chapters of this review, we will not focus on the optimization of materials properties, but rather on optimizing the ways to synthesize a specific material. Given that the vast majority of studies has focused on organic reactions in solution, this chapter will introduce the major concepts using this class of transformations. In the following subsections, approaches toward other solution-phase reactions, as well as non-solution-phase reactions, will be discussed. Afterwards, this section will provide a comprehensive overview of all works in which the automated integration of robotic reaction execution and data-driven optimization has been demonstrated in multiple iterations for enabling autonomous reaction or process optimization (i.e., Level 4 SDLs, [Figure 1](#)). For a more global discussion from the perspective of reaction optimization, we refer the reader to existing reviews and perspectives in the field.<sup>375–383</sup>

For solution-based reactions, the search space comprises a wide series of categorical (identity of reagents, catalysts, solvents, additives etc.) and continuous (relative stoichiometries, concentration, temperature(s), reaction time(s) etc.)

variables. With the goal of automating the optimization process, the choice of the appropriate automation platform for performing the reaction (see [Hardware](#)) facilitates or complicates the variation of specific parameters. Generally, the variation of continuous parameters, particularly reaction quantities, can be readily performed on a wide variety of experimental systems ([Figure 7](#)). This has led to a large number of studies for optimizing continuous reaction parameters in an automated fashion. On the other hand, optimizing over categorical parameters, such as specific choices of chemicals, poses additional challenges both from the hardware and the software side; the physical availability of reactants and reagents, as well as the storage capacities on the automated platforms, pose physical constraints to the number of available categorical options. As a consequence, most examples of closed-loop reaction condition optimization have operated on comparatively few options (usually  $< 10$ ) for categorical variables. Additionally, categorical parameters are not readily represented numerically, and lack an unambiguous order or measure of similarity (e.g., between solvents, or catalysts). This requires human decision in selecting an appropriate representation for chemicals for the optimization algorithm. Optimization algorithms that operate on molecular entities, as well as mixed continuous–categorical parameter spaces, are discussed in detail in [Software](#).

#### 4.1. Specialized Hardware and Software

The primary objective of optimizing a synthetic reaction is generally the reaction yield, i.e., the quantity of the desired reaction product that is formed. Reaction selectivity, defined as the ratio between the desired product and an undesired side product (e.g., a constitutional isomer or a stereoisomer) can be regarded as an auxiliary measure of product formation. In terms of prediction, this is a particularly challenging objective; reaction yields do not only depend on the rates of all steps in the desired reaction sequence, but also on the rates of a multitude of possibly unknown side reaction pathways, which renders the physics-inspired modeling of reaction yields highly difficult. Added complexity stems from the coupling of chemical reaction kinetics with physical transport phenomena (e.g., mass transfer/diffusion, and heat transfer). This dependence on unknown steps and mechanisms can lead to cross-dependencies between the assumedly independent variables, and unforeseen activity cliffs upon small variations. In these scenarios, traditional OFAT optimization approaches, which have been the method of choice for reaction optimization in most laboratories, face severe challenges. In this context, the optimization of chemical reaction yields is particularly well-suited for data-driven, system-agnostic optimization approaches (see [Software](#) for further details).

When it comes to larger-scale reactions and industrial process optimization campaigns, reaction yield or selectivity is no longer the sole optimization objective. Economic and ecological considerations present further constraints to the optimization problem, which can include: reagent and energy costs, atom economy, or environmental impact factors, such as measures of waste formation, or operational and environmental hazards.

**4.1.1. Reaction Execution.** On a laboratory scale, the execution of solution-phase chemical reactions can be classified into two complementary strategies: batch and continuous-flow operation ([Table 3](#)). Both strategies come with distinct chemical (dis)advantages for performing specific types of

**Table 3. Autonomous reaction optimization in batch and flow reactors**

	Batch Reactors	Flow Reactors
Achieving high throughput	Parallel experiments	Sequential experiments
Varying quantities (stoichiometry and concentration)	Easy	Easy
Varying other continuous variables (e.g., temperature, time)	Difficult (in parallel)	Easy
Varying categorical variables (e.g., reactant or reagent identities)	Easy	Difficult
Intermediate purification and multi-step reactions	Difficult	Easy
Increasing reaction scale	Difficult (requires increasing the reactor volume)	Easy (requires increasing the reaction time)

reactions, which have been thoroughly discussed in the literature, and are outside the scope of this review. Instead, we aim to provide a discussion of these strategies in the context of automation, and developing autonomous self-optimizing reactors.

For over a century, solution-phase reactions have been predominantly performed in batch reactors<sup>384</sup>—a fact well reflected in the practical chemistry education, where synthesis is mainly taught using beakers, flasks and vials as batch reaction vessels. In a teaching laboratory, and in many research laboratories, this approach is highly advantageous, as it allows the execution of a variety of different chemistries with minimal hardware requirements. In fact, most reactions can be executed by a human operator in standard, general-purpose glassware. However, batch reactions face severe challenges when it comes to isolation and purification, or the execution of multi-step reactions. The human-centric approaches to reaction workup and purification, such as extraction, crystallization or chromatography, are often based on specific operations that require large degrees of adaptive, intuition-guided decision-making. This has rendered their direct, programmatic translation into automated workflows challenging, and may be one of the reasons why batch reactors have not found widespread application in closed-loop discovery workflows yet.

Recent advances in biotechnology and the corresponding liquid handling systems (see section on [Hardware](#) for further details) have enabled the miniaturization and parallelization of batch reaction execution in multi-well plates. In such cases, analyzing directly on the crude reaction mixture can significantly increase the automated experimental throughput, particularly when it comes to varying categorical variables such as the identity of reactants, reagents, catalysts or solvents. Therefore, such HTE systems have been primarily used for large condition or substrate screening campaigns for important catalytic reactions.

As a complementary approach, the past decades have seen the development of flow reactors, in which reactions are performed in a continuously flowing stream of liquid. Importantly, with the requirement to continuously pump the reagent, reactant and reaction streams, this strategy to reaction execution is inherently automated. Still, flow reactions require highly specialized hardware and software, preventing the widespread adoption of this technology as a tool for chemical synthesis. Reactants and products must also be gaseous or liquid, and the liquid medium cannot be too viscous. Developments in microfabrication have led to miniaturization

of flow systems into microfluidic systems (sometimes called lab-on-a-chip).<sup>385</sup> The small footprint and low reagent consumption of microfluidic systems make them useful for high-throughput synthesis of compounds at small scales. The operation of chemical reactions in microfluidic systems provides a series of distinct advantages in terms of heat and mass transfer, interphase reactivity or safety. For more detailed discussions on flow and microfluidic reactors, we refer the reader to review articles in the literature.<sup>385–387</sup> As such, both academic researchers and industrial teams have focused on developing automated in-flow synthesis platforms, leading to a large variety of specific, custom-built setups, with few standardized solutions on the market.

From the perspective of autonomous optimization, the inherently automated nature of flow reactors made them ideal platforms for early explorations of autonomous operation modes, and a vast majority of examples of closed-loop reaction optimization (see below) have been performed on continuous-flow platforms. These platforms have allowed for optimization of continuous parameters such as stoichiometries, reaction times, and temperatures, which can be readily varied in sequential experiments, leading to high experimental throughput. The exploration of larger numbers of categorical entities such as reactants or reagents, however, comes with increased hardware requirements and experimental efforts. Recent work has also demonstrated autonomous flow systems in Schlenk lines, allowing for studies involving highly reactive or sensitive compounds.<sup>388</sup> Another major advantage of flow chemistry lies in the ability to telescope individual operations (including purification steps) into longer sequences, enabling the automated operation of multi-step sequences. This capacity, which has been reviewed comprehensively in the literature,<sup>389,390</sup> has enabled the closed-loop optimization of multi-step reactions in solution, which will be discussed in the [Multi-step organic reaction](#) section.

For screening and optimization purposes, segmented-flow approaches (often referred to as droplet reactors) provide an attractive approach for HTE in flow systems.<sup>386,391,392</sup> Rather than having a continuous flow of liquid in which the reaction occurs, the stream is divided into small separated segments by an inert gas (such as argon) or immiscible liquid (such as perfluorinated oil). Each of these segments can be regarded as an individual batch, and precise operational control can allow for screening distinct reaction conditions (e.g., reagent identities or quantities) in each of these batches. Moreover, the smaller reaction volumes and high surface-to-volume ratios have demonstrated significant acceleration of reactions, leading to micro-droplet approaches such as microfluidic flow droplet reactors, and even free-standing non-flow systems.<sup>392</sup>

**4.1.2. Reaction Analysis.** Irrespective of the reaction execution platform, the second major component of a self-optimizing reactor is a module to quantify the optimization objective, such as the reaction yield. The most widespread, general-purpose approach to this is the use of an automated chromatographic separation technique, usually LC, or GC, coupled to a quantitative detection technique. In this approach, an aliquot from the reaction mixture is taken and analyzed on the external instrument ([Figure 7](#)). Importantly, the required instruments have been commercialized for decades, and offer robust hardware solutions, which are available in most experimental laboratories.

As an alternative, *in situ* monitoring techniques can be used to directly analyze the crude reaction mixture and monitor

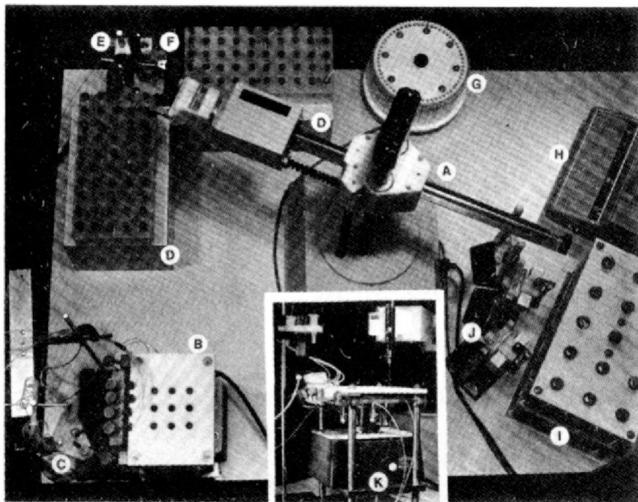
changes in the reaction composition to quantify possible optimization objectives. Spectroscopic tools such as NMR spectroscopy (implementable through benchtop spectrometers with flow cells) or infrared spectroscopy (flow cells or *in-situ* probes) can be used to identify and quantify compounds if unique, compound-specific signals exist. Further, *in situ* probes such as UV-Vis spectroscopy or conductivity cannot, in most cases, be used to quantify specific materials, but allow for the monitoring of global properties of the reaction mixture, which can serve as a valuable proxy for the actual optimization objective.

Analytical techniques that enable the quantification of reactants, intermediates or products throughout the course of a reaction can enable decision-making in real time, e.g., for adjusting reaction times, temperatures or reagent quantities. Such *adaptive optimization* of reaction conditions, however, does not follow the iterative closed-loop definition of SDLs, and therefore exceeds the scope of this review.

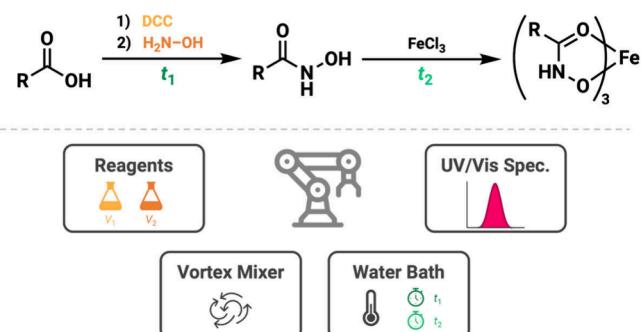
**4.1.3. Early Examples of Autonomous Condition Optimization.** Attempts to develop SDLs for autonomous reaction condition optimization date back to the 1970s, when Winicov *et al.*, from pharmaceutical company Smith, Kline & French, describe a fully automated batch reactor which shows remarkable similarities to modern open-source systems for automating batch reactions such as the *Chemputer*.<sup>393</sup> The authors describe automated modules for liquid addition (through pumps), stirring, heating and cooling, as well as reaction analysis by coupling to an HPLC-UV system. Remarkably, they even discuss the coupling of their platform with a simplex algorithm for automated experiment planning. However, no actual experiments have ever been publicly reported with this platform, neither in the initial publication from 1978, nor in any follow-up works; this is likely due to the proprietary nature of the research at Smith, Kline & French laboratories.

To the best of our knowledge, the first published example of a self-optimizing reactor stems from 1987. Matsuda *et al.* reported the autonomous optimization of the adduct formation reaction between phosphotungstic acid and basic drug molecules, namely, chlorpromazine hydrochloride and levo-mepromazine hydrochloride.<sup>260</sup> For this purpose, the authors had developed a robotic platform consisting of a Zymark robotic arm (see [Figure 8](#) for a related experimental setup by Frisbee *et al.*<sup>394</sup>) with two exchangeable tools for liquid transfer and vial transport, respectively. The reagents were added as solutions to a batch reactor vial, and the entire vial was first transported to a vortex mixer for stirring, and subsequently to a water bath for heating. Since the desired adducts are strongly colored, they can be detected quantitatively via steady-state absorption spectroscopy. After completion of the reaction, the vial is transferred to a UV-Vis spectrophotometer by the robotic arm, and the absorption at wavelength 538 nm was recorded as a proxy for the reaction yield. The simplex algorithm was used for iteratively planning the next experiment, varying the quantity of phosphotungstic acid and the reaction time. Remarkably, the authors show that the optimal reaction conditions can be found in less than ten iterations for both drug molecules.

Using the same experimental setup, the authors demonstrated the optimization of a significantly more complex reaction in 1988 (see [Figure 9](#)):<sup>259</sup> the conversion of a carboxylic acid to the corresponding hydroxamate using *N,N'*-dicyclohexylcarbodiimide (DCC) and hydroxylamine, followed



**Figure 8.** Example of an early automated robotic synthesizer centered around a Zymark robotic system. A: Robotic arm. B: Reactor station. C: Remote dispenser. D: Aliquot archive station. E: Workup station. F: Syringe and needle wash station. G: Turntable. H: Reporting integrator for analytical instrumentation. I: Reagent station. J: Robotic tool parking station. Figure reproduced with permission from Frisbee *et al.*<sup>394</sup> Copyright 1984, American Chemical Society.



**Figure 9.** Schematic of early example of autonomous reaction condition optimization in solution performed by Matsuda *et al.*<sup>259</sup> Coupling reaction of carboxylic acids with *N,N'*-dicyclohexylcarbodiimide (DCC) and hydroxylamine, and detection of the resulting hydroxamic acid as its iron(III) complex through UV-Vis absorption spectroscopy.

by the complexation of the hydroxamate with an iron(III) salt to give a colored, UV-Vis-detectable complex. The authors demonstrate the optimization of up to four continuous parameters (quantities of DCC and hydroxylamine, reaction times of both steps), showing that optimization can be performed in under 30 experiments. The authors benchmarked their optimizer against a grid search strategy, demonstrating a significant reduction (>75%) in the number of required experiments.

Inspired by these early results, and the sophisticated automation platforms developed in the 1980s (see e.g., Figure 8), Lindsey and co-workers made a series of contributions to early SDLs in solution-based synthesis. On the hardware side, the design of their “automated chemistry workstation”<sup>395</sup> is of note; in parallel with the developments of small-scale pipetting robots in biochemistry that spilled over to chemistry only a decade later, their system enables the miniaturization of chemical reactions (to  $\mu\text{L}$  scale), as well as advanced analytical

techniques, including, but not limited to, automated thin-layer chromatography.

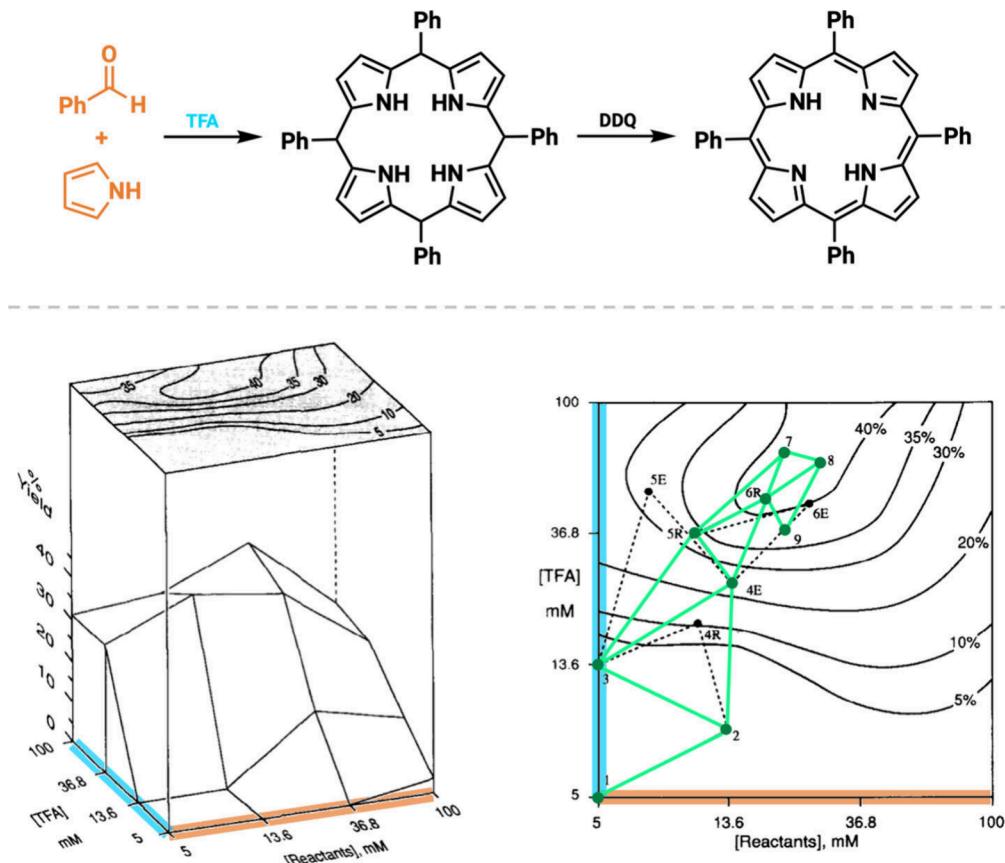
In addition to a series of automated data generation workflows, as a proof-of-concept, they demonstrated the closed-loop optimization of an “optical filter,” creating a specific absorption profile by mixing different dye solutions. Experiments are planned iteratively by the simplex algorithm, and are executed sequentially on the automated platform until the desired absorption profile is reached.<sup>396</sup> Showcasing an application in synthetic chemistry, the authors perform the closed-loop optimization of the synthesis of porphyrin dyes from aromatic aldehydes and pyrrole under acidic conditions.<sup>397</sup> The concentrations of both reactants (in 1:1 stoichiometry) and of the acid additive were used as independent variables. The yield of the product, obtained by quantitative DDQ oxidation, was determined by UV-Vis spectroscopy. The authors demonstrate accelerated experimentation by comparison with a full factorial design approach (Figure 10).

Beyond these works, Lindsey and co-workers made a series of important contributions to advance optimization and decision-making algorithms beyond the native simplex algorithm,<sup>375</sup> for example, using decision tree algorithms to handle screening and optimization of categorical variables. While such systems may have been utilized in industrial settings, given the research effort from both academic and industrial researchers,<sup>398</sup> publications from the late 1990s for SDL optimization of synthetic reactions are sparse. The cost and reproducibility of the robotic hardware, as well as the transferability of software may be influential factors in this regard.

#### 4.2. Single-Step Organic Reactions

The rise of flow chemistry as a versatile, automated tool for reaction execution, as well as the increased accessibility and distribution of software via the internet, have sparked new interest in the development of self-optimizing reactors in the late 2000s and early 2010s. Since then, a large number of examples focusing on the optimization of organic reactions in solution have been reported in the literature. This section will present the most important concepts and advances using selected examples. A complete, to the best of our knowledge, list of further examples from the literature, including the target reaction, independent optimization variables, hardware, software and optimization objectives is given in Table 4. Detailed discussions of single- and multi-step reactions optimization are provided after.

**4.2.1. Self-Optimizing Flow Reactors and Analytical Advances.** The first modern examples of self-optimizing reactors were reported by Jensen and co-workers in 2010.<sup>269,399</sup> In their first work, McMullen *et al.* show the optimization of reaction conditions for the Heck-coupling between 4-chlorobenzotrifluoride and 2,3-dihydrofuran in a flow microreactor.<sup>399</sup> The optimization was carried out to maximize the HPLC-determined yield of the mono-arylated reaction product, which is prone to undergo an undesired second coupling. Categorical parameters such as solvents, phosphine ligands and palladium sources were systematically screened to find conditions under which ammonium salts are soluble and the formation of palladium black is minimized, as this leads to clogging of the microreactor. Subsequently, a closed-loop optimization campaign over the continuous parameters residence time and alkene:aryl chloride ratio was



**Figure 10.** Closed-loop optimization of porphyrin synthesis conditions, as reported by Lindsey and co-workers. Optimization was performed under variation of the concentration of reactants, as well as the quantity of trifluoroacetic acid (TFA). The bottom row shows the response surface, as obtained by factorial design experiments (left), as well as the optimization trajectory of the simplex algorithm (right). Figure adapted with permission from reference.<sup>397</sup> Copyright 1992, Elsevier. DDQ: Dichlorodihydroquinone.

carried out using the Nelder-Mead simplex algorithm. The authors show that the optimal conditions can also be carried out in a meso-reactor while preserving the optimum yield, demonstrating the successful transfer from micro- to meso-scale systems. Notably, even though over 20 years had passed since the initial demonstration of self-optimizing reactors, the selected optimization algorithm is highly similar to the early works discussed above. Using a similar setup, McMullen *et al.* reported the evaluation of multiple “black-box” optimization algorithms,<sup>269</sup> namely the steepest descent algorithm, the Nelder-Mead Simplex algorithm<sup>253</sup> and SNOBFIT,<sup>335</sup> for the closed-loop optimization of the Knoevenagel condensation of *p*-anisaldehyde and malonitrile in a flow microreactor (see Figure 11).<sup>269</sup> All algorithms converged to essentially the same optimum conditions within 12 hours. The authors optimized a weighted objective function of product yield and flow rate, thereby showcasing the first multi-objective self-optimization.

Shortly after, in 2011, Poliakoff and co-workers reported a series of examples in which they employ their SDL for the optimization of reactions using  $\gamma$ -alumina as a heterogeneous catalyst in supercritical CO<sub>2</sub> as the solvent.<sup>400</sup> Parrott *et al.* optimized the yield of the dehydration of ethanol, the yield of the carboxymethylation of 1-pentanol with dimethyl carbonate (DMC), and the yield of the methylation of 1-pentanol with DMC. All of these optimization runs used a super-modified simplex algorithm to optimize temperature, pressure and CO<sub>2</sub> flow rate as variables (see Figure 12). The latter optimizations were each completed in approximately 1.5 days, whereas a

combinatorial search in the condition space would have taken more than 50 times longer, showcasing the efficiency of self-optimizations. Later, Bourne *et al.* re-evaluated the methylation of 1-pentanol using different methylating agents in a four-variable optimization, using the super-modified simplex algorithm.<sup>443</sup> In a following study, Jumbam *et al.* evaluated different objective functions for optimizing this transformation:<sup>401</sup> the yield, the space-time yield, the E-factor, E+ (the E-factor including all wastes) and the weighted space-time yield, calculated by the product of the space-time yield and the yield. The different criteria were shown to result in different optimal conditions. Surprisingly, a low E-factor led to a high value of E+. Overall, this shows the importance of designing an appropriate objective function when considering multiple inherently competitive optimization targets.

After these initial developments of self-optimizing reactors, a diversification of analytical techniques took place rapidly, allowing researchers to harness the different advantages of each technique (see discussion above). In 2012, Moore and Jensen reported an in-line flow IR cell to optimize the Paal-Knorr synthesis of pyrroles.<sup>402</sup> With this IR setup (Figure 13), steady-state conditions can be ensured before objective functions are evaluated. As a first objective function, the authors aimed to maximize the ratio between conversion and residence time. However, this led to poor conversions, and a quadratic loss function was applied to yields lower than 85%. The newly designed objective function resulted in an optimum of 81% conversion, demonstrating the difficulty of selecting combined

**Table 4. Comprehensive overview of “modern” examples of self-optimizing reaction platforms<sup>a</sup>**

Reaction	Optimization Parameters	Synthetic Hardware	Analytical Hardware	Optimization Algorithm	Optimization Objective
Heck coupling of 4-chlorobenzotrifluoride and 2,3-dihydrofuran <sup>399</sup>	Residence time, ratio of reactants	Microfluidic system	HPLC	Nelder-Mead Simplex	Yield of mono-coupled product
Knoevenagel condensation of p-anisaldehyde and malononitrile <sup>269</sup>	Temperature, residence time	Microfluidic system	HPLC	Steepest Descent; Nelder-Mead Simplex, SNOBFIT	Custom objective function consisting of flow rate and product yield
Oxidation of benzyl alcohol with CrO <sub>3</sub> <sup>269</sup>	Temperature, residence time, reactant concentration	Microfluidic system	HPLC	Simplex	Yield of benzaldehyde
Dehydration of ethanol in supercritical CO <sub>2</sub> <sup>400</sup>	Temperature, pressure, CO <sub>2</sub> flow rate	Flow	GLC	Super-Modified Simplex	Yield
Carboxymethylation of 1-pentanol with DMC in supercritical CO <sub>2</sub> <sup>400</sup>	Temperature, pressure, CO <sub>2</sub> flow rate	Flow	GLC	Super-Modified Simplex	Yield
Methylation of 1-pentanol with dimethyl carbonate in supercritical CO <sub>2</sub> <sup>400</sup>	Temperature, pressure, CO <sub>2</sub> flow rate	Flow	GLC	Super-Modified Simplex	Yield
Methylation of 1-pentanol with dimethyl carbonate in supercritical CO <sub>2</sub> <sup>400</sup>	Temperature, pressure, CO <sub>2</sub> flow rate, ratio of reactants	Flow	GLC	Super-Modified Simplex	Yield
Methylation of 1-pentanol with methanol in supercritical CO <sub>2</sub> <sup>400</sup>	Temperature, pressure, CO <sub>2</sub> flow rate, ratio of reactants	Flow	GLC	Super-Modified Simplex	Yield
Methylation of 1-pentanol with dimethyl carbonate in supercritical CO <sub>2</sub> <sup>401</sup>	Temperature, pressure, reagent flow rate	Flow	GLC	Super-Modified Simplex	Yield; space-time yield; E-factor; E-factor+ (E-factor including waste); product of space-time yield and yield
Paal-Knorr synthesis between 2,5-hexandione and ethanolamine <sup>402</sup>	Temperature, residence time	Microfluidic system	IR	steepest descent; conjugate gradient	Custom objective functions considering conversion and residence time
Methylation of 1-pentanol with dimethyl carbonate <sup>403</sup>	Temperature, residence time	Flow	IR, GLC	Super-Modified Simplex	Yield
Condensation of 4-fluorobenzaldehyde and aniline <sup>404</sup>	Residence time, ratio of reagents	Flow	NMR	modified Nelder-Mead Simplex	Custom objective function related to the space-time yield
Etherification of n-propanol in supercritical CO <sub>2</sub> <sup>28</sup>	Flow rates, temperature	Flow	GLC	Super-Modified Simplex	Yield
Methylation of 1-butanol with dimethyl carbonate in supercritical CO <sub>2</sub> <sup>128</sup>	Flow rates, temperature	Flow	GLC	Super-Modified Simplex	Yield
Mono-alkylation of trans-1,2-diaminocyclohexane with 4-methoxybenzyl chloride <sup>405</sup>	Temperature, residence time, reagent concentration	High-throughput microfluidic system	LC/MS	feedback DOE search algorithm	Yield
Hydrolysis of 3-Cyanopyridine <sup>117</sup>	Temperature, residence time, reactant concentration	Flow	MS	Modified Simplex	Ratio of product MS peak over reactant MS peak
Appel reaction of 1-phenylethanol <sup>117</sup>	Temperature, residence time, reagent equivalents, overall concentration	Flow	IR	Modified Simplex	Custom objective function considering terms for throughput, conversion, consumption
Amidation of methyl nicodimate with aqueous MeNH <sub>2</sub> <sup>406</sup>	Temperature, reactant flow rate, reactant equivalents	Flow	MS	SNOBFIT	Yield

Table 4. continued

Reaction	Optimization Parameters	Synthetic Hardware	Analytical Hardware	Optimization Algorithm	Optimization Objective
Reaction of 2,4-dimethoxyaniline to acrylamide derivative <sup>407</sup>	Temperature, reactant flow rate	Flow	HPLC	SNOBFIT	Yield
Synthesis of AZD9291 acrylamide <sup>407</sup>	Temperature, reactant flow rate	Flow	HPLC	SNOBFIT	Yield
Multiple Suzuki-Miyaura couplings <sup>408</sup>	Catalyst precursor, ligand, temperature, residence time, catalyst loading	Droplet microfluidic system	HPLC	Custom algorithm	TON with lower boundary for yield
Heck-Matsuda reaction for the arylation of cis-butene-1,4-diol <sup>409</sup>	Temperature, residence time, ratio of reagents, catalyst loading	Flow	GC/MS	modified Nelder-Mead Simplex	Yield, production cost; throughput
C–H activation of aliphatic secondary amine to generate azetidine <sup>410</sup>	Temperature, residence time, ratio of reagents, ratio of catalyst to reagents	Flow	GC	Active Learning	Cost; yield
Synthesis of o-xylenyl C <sub>60</sub> adducts <sup>266</sup>	Reagents flow rate, temperature	Flow	HPLC	SNOBFIT	Minimization of mole-fraction of third-order adduct; minimization of mole-fraction of third-order adduct, with additional constraints on total mole fraction of first- and second-order adducts; minimization of mole-fraction ratio of first-order adduct to second-order adduct; minimization of mole-fraction of third-order adduct, with additional constraints on total mole fraction of first- and second-order adducts and mole fraction ratio of second-order adduct to first-order adduct
Pomeranz-Fritsch synthesis of isoquinoline <sup>336</sup>	Flow rate, voltage, and pressure applied on the spray source	Microdroplet flow platform	MS	Deep RL	Yield
Friedländer synthesis of a substituted quinoline <sup>336</sup>	Flow rate, voltage, and pressure applied on the spray source	Microdroplet flow platform	MS	Deep RL	Yield
Synthesis of ribose phosphate <sup>336</sup>	Flow rate, voltage, and pressure applied on the spray source	Microdroplet flow platform	MS	Deep RL	Yield
Reaction between 2,6-dichlorophenolindophenol (DCIP) and ascorbic acid <sup>336</sup>	Flow rate, voltage, and pressure applied on the spray source	Microdroplet flow platform	MS	Deep RL	Yield
Photoredox Ir–Ni dual-catalyzed decarboxylative arylation with several substrates <sup>411</sup>	Temperature, residence time, base; Temperature, residence time, Ni precatalyst; Temperature, residence time	Segmented oscillatory flow ('microslug') reactor with custom photochemistry module	HPLC	Custom algorithm	Yield, productivity
Suzuki-Miyaura coupling with 3-chloropyridine and pyridine boronic ester <sup>412</sup>	Precatalyst scaffold, ligand, catalyst loading, temperature, residence time	Flow	LC	mixed-integer nonlinear program	TON with lower boundary for yield
[2+2] Paterno-Büchi reaction between furans and benzophenones <sup>413</sup>	Reagent flow rates	Flow system with photoreactor	IR	Modified Simplex	Conversion
Claisen-Schmidt condensation of benzaldehyde and acetone <sup>414</sup>	Temperature, reagent flow rates	Flow	HPLC	SNOBFIT	Yield
Semi-hydrogenation of 2-methyl-3-butyn-2-ol over Pd/SiO <sub>2</sub> <sup>415</sup>	Flow rate of substrate, flow rate of catalyst poison, flow rate of IPA solvent	Flow	GC	Super-Modified Simplex	Yield
Allylation of sesamol <sup>416</sup>	Temperature, residence time, stoichiometry	Flow	HPLC	Nelder-Mead Simplex and	Yield

Table 4. continued

Reaction	Optimization Parameters	Synthetic Hardware	Analytical Hardware	Optimization Algorithm	Optimization Objective
[3,3]-Claisen rearrangement of allyl sesamol <sup>416</sup>	Temperature, residence time	Flow	NMR	golden section search Nelder-Mead Simplex and golden section search	Product productivity
Isomerization to esmethoxy-carpacine <sup>416</sup>	Temperature, residence time, base loading	Flow	HPLC	Nelder-Mead Simplex and golden section search	Yield
Oxidative dimerization of desmethoxycarpacin to carpanone <sup>116</sup>	Reagent equivalents, temperature, residence time	Modular flow system	HPLC, MS, IR, Raman	Nelder-Mead Simplex and golden section search SNOBFIT	Yield
HWE Olefination of 4-phenylcyclohexanone <sup>265</sup>	Reagent equivalents, temperature, residence time	Modular flow system	HPLC, MS, IR, Raman	SNOBFIT	Yield
Reductive amination of o-Methoxybenzaldehyde with benzylamine <sup>265</sup>	Reagent equivalents, temperature, residence time	Modular flow system	HPLC, MS, IR, Raman	SNOBFIT	Yield
S <sub>N</sub> Ar of o-nitrochlorobenzene and tetra-hydro-quinoline <sup>265</sup>	Reagent equivalents, temperature, residence time	Modular flow system	HPLC, MS, IR, Raman	SNOBFIT	Yield
Multi-step photoredox-catalyzed oxidative $\alpha$ -functionalization of amines <sup>265</sup>	Reagent equivalents, temperature, residence time	Modular flow system	HPLC, MS, IR, Raman	SNOBFIT	Yield
Multi-step ketene generation and [2+2] cycloaddition <sup>265</sup>	Reagent equivalents, temperature, residence time	Modular flow system	HPLC, MS, IR, Raman	SNOBFIT	Custom objective function considering yield and selectivity
Grignard Addition in the second step in the synthesis of tramadol <sup>116</sup>	Temperature, residence time, equivalents of Grignard Reagent	Flow	IR	Modified Simplex	Custom objective function consisting of throughput, conversion and consumption
Amidation in the first step in the synthesis of lidocaine <sup>116</sup>	Temperature, residence time, reagent equivalents	Flow	IR	Modified Simplex	Custom objective function consisting of throughput, conversion and energy
Second step in the synthesis of lidocaine <sup>116</sup>	Temperature, residence time, reagent equivalents	Flow	IR	Modified Simplex	Custom objective function consisting of throughput and conversion
Alpha bromination of ketone in the first step in the synthesis of buproprion <sup>116</sup>	Temperature, residence time, stoichiometry	Flow	IR	Modified Simplex	Custom objective function consisting of throughput, conversion and consumption
Amine alkylation in the second step in the synthesis of buproprion <sup>116</sup>	Temperature, residence time, stoichiometry	Flow	IR	Modified Simplex	Custom objective function consisting of conversion
S <sub>N</sub> Ar reaction between 2,4-difluorotoluene and morpholine <sup>417</sup>	Residence time, reagent equivalents, reagent concentration, temperature	Flow	HPLC	TS-EMO	Pareto-front between space-time yield and E-factor
N-benzylation of $\alpha$ -methylbenzylamine with benzyl bromide <sup>417</sup>	Reagent flow rates, reagent equivalents, reagent concentration, temperature	Flow	HPLC	TS-EMO	Pareto-front between space-time yield and impurity formation

Table 4. continued

Reaction	Optimization Parameters	Synthetic Hardware	Analytical Hardware	Optimization Algorithm	Optimization Objective
Methanolysis of 2-cyanopyridine with MeONa <sup>418</sup>	Temperature, residence time, reagent equivalents	Flow	HPLC	modified Nelder-Mead Simplex	Yield
Acid-catalyzed condensation of imidate with $\alpha$ -amino alcohols <sup>418</sup>	Temperature, residence time, reagent equivalents	Flow	HPLC	modified Nelder-Mead Simplex	Yield
Hydrogenation of benzaldehyde over Pd/C <sup>419</sup>	Temperature, flow rate, H <sub>2</sub> pressure	H-Cube micro-reactor	IR	Simplex	Conversion
Hydrogenation of an alpha-ketoester over Pd/C <sup>419</sup>	Temperature, flow rate, H <sub>2</sub> pressure	H-Cube micro-reactor	IR	Simplex	Conversion
Hydrogenation of a quinoxaline over Ir NP on carbon nanotubes <sup>419</sup>	Temperature, flow rate, H <sub>2</sub> pressure	H-Cube micro-reactor	IR	Simplex	Conversion
Hydrogenation of a quinidine over Ir NP on carbon nanotubes <sup>419</sup>	Temperature, flow rate, H <sub>2</sub> pressure	H-Cube micro-reactor	IR	Simplex	Conversion
Cross-coupling between aniline and p-tolyl trifluoromethanesulfonate <sup>420</sup>	Temperature, residence time, amount of base, type of base	Microfluidic system	LC	Custom algorithm	Yield
Cross-coupling between benzamide and p-tolyl trifluoromethanesulfonate <sup>420</sup>	Temperature, residence time, amount of base, type of base	Microfluidic system	LC	Custom algorithm	Yield
Cross-coupling between 2-phenylethan-1-amine and p-tolyl trifluoromethanesulfonate <sup>420</sup>	Temperature, residence time, amount of base, type of base	Microfluidic system	LC	Custom algorithm	Yield
Cross-coupling between morpholine and p-tolyl trifluoromethanesulfonate <sup>420</sup>	Temperature, residence time, amount of base, type of base	Microfluidic system	LC	Custom algorithm	Yield
Songashira coupling of 3,5-dibromopyridine with 1-hexyne <sup>421</sup>	Residence time, reagent equivalents, temperature	Flow	HPLC	TS-EMO	Custom objective function considering the conversion and space-time-yield
Claisen-Schmidt condensation of benzaldehyde and acetone <sup>421</sup>	Flow rates of reagents, flow rates of aqueous and organic solvents, temperature	Flow in a miniaturized CSTR cascade	HPLC	TS-EMO	Custom objective function considering purity, space-time yield and reaction mass efficiency
Photocatalytic hydrogen evolution reaction <sup>11</sup>	Concentration of catalyst, concentration of hole scavenger, concentration of 8 other additives	Mobile chemist operating between workstations	GC	BO	H <sub>2</sub> Production
Palladium-catalyzed direct C-H arylation of indole-3-acetic acid derivatives with arene diazonium salts <sup>422</sup>	Reagent equivalents, temperature, residence time	Flow	HPLC	Simplex	Yield
Aldol-condensation of benzaldehyde and acetone <sup>423</sup>	Reagent equivalents, temperature, residence time	Flow	HPLC	TS-EMO	Pareto front between: yield & cost; space-time-yield & E-Factor
AuNP catalyzed reduction of 4-nitrophenoxy with NaBH <sub>4</sub> <sup>424</sup>	NP surface area, NaBH <sub>4</sub> conc., residence time	Flow	UV	SNOBFT	Conversion
Stereoselective Suzuki-Miyaura coupling <sup>425</sup>	Phosphine ligand, phosphine to Pd ratio, Pd loading,	Chemspeed SWING	HPLC	BO	Multi-objective with decreasing priorities: E-product yield (max), Z-product yield (min), Pd loading (min), Arylboronic acid equivalents (min)

Table 4. continued

Reaction	Optimization Parameters	Synthetic Hardware	Analytical Hardware	Optimization Algorithm	Optimization Objective
Thioquinazolinone with a teleposed lithium-halogen exchange and phenyl isocyanate addition <sup>426</sup>	aryboronic acid equivalents, temperature	Flow rate, reactor volume, temperature	Flow	FT-IR BO	Yield
Thioquinazolinone with a teleposed lithium-halogen exchange and phenyl isocyanate addition <sup>426</sup>	Flow rate, reactor volume, temperature, lithiating reagent	Flow	Flow	FT-IR BO	Yield
Oxidation of methyl phenyl sulfide to sulfoxide <sup>427</sup>	Temperature, residence time, H <sub>2</sub> O <sub>2</sub> concentration, catalyst concentration, temperature, residence time, catalyst	CSTR cascade	GC/MS HPLC	TS-EMO, EIM-EGO BO	Multi-objective considering conversion, selectivity, space-time-yield
Suzuki-Miyaura cross coupling <sup>428</sup>	LED Brightness, residence time, temperature, base	CSTR cascade	HPLC	BO	Yield
Metallaphotoredox-catalyzed sp <sup>3</sup> -sp <sup>2</sup> cross-coupling of carboxylic acids with alkyl halides <sup>428</sup>	Reagent ratio, temperature, residence time, photocatalyst	CSTR cascade	HPLC	BO	Yield, diastereoselectivity
Metallaphotoredox-catalyzed cross-coupling reaction between trans-4-hydroxy proline and 4-bromoaacetophenone <sup>428</sup>	Reagent equivalents, loading time	Flow	HPLC-IR	BO	Yield
Photocatalyzed C-C bond forming reaction between an aryl iodide and tert-butyl vinyl carbamate <sup>429</sup>	Reagent equivalents, loading of co-catalyst, residence time	Flow	HPLC-IR	BO	Yield
Photocatalyzed cyclization reaction <sup>429</sup>	Equivalents of oxidant, photocatalyst concentration	Flow	HPLC-IR	BO	Yield
S <sub>N</sub> Ar between dimethylmorpholine and a nitrohalopyridine <sup>430</sup>	Residence time, temperature, equivalent of reagents, equivalent of base, halide leaving group	Modular flow system	LC/MS & IR	BO	Custom objective function considering yield, productivity and cost
Nitro reduction and amide coupling for the synthesis of sonidegib <sup>430</sup>	Amide coupling activation agent, activation residence time, equivalents of reagents, temperature in amide coupling, reactor size in amide coupling	Modular flow system	LC/MS & IR	BO	Yield and productivity of product
Large set of Suzuki-Miyaura cross couplings <sup>431</sup>	Solvent base, catalyst and ligand, and temperature	Robotic system	LC/MS	BO	Generality (average yield over multiple reactions)
Suzuki coupling towards 4-(2,3-dimethoxyphenyl)-N- <sup>432</sup> pyrrol[2,3-b]pyridine	Catalyst, ligand, base, solvent	Batch	LC/MS	Hybrid dynamic optimization (GNN & BO)	Conversion, yield
Buchwald coupling towards N-(4-methoxyphenyl)-N-phenylpyrimidin-5-amine <sup>432</sup>	Catalyst, ligand, base, solvent	Batch	LS/MS	Hybrid dynamic optimization (GNN & BO)	Conversion, yield
Buchwald coupling towards N,N-diphenylquinoxalin-2-amine <sup>432</sup>	Catalyst, ligand, base, solvent	Batch	LC/MS	Hybrid dynamic optimization	Conversion, yield

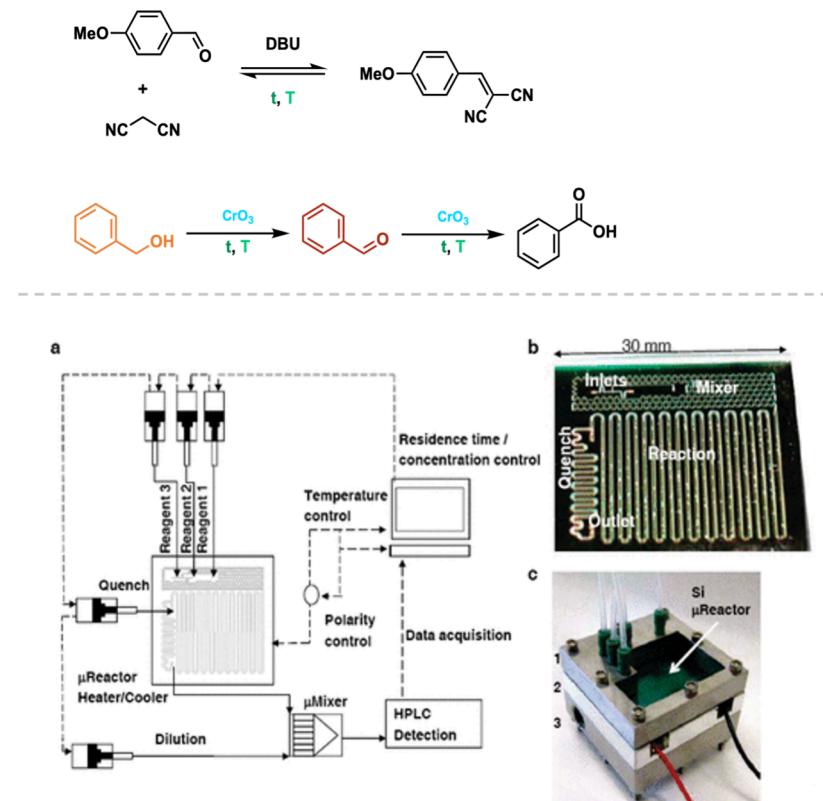
Table 4. continued

Reaction	Optimization Parameters	Synthetic Hardware	Analytical Hardware	Optimization Algorithm	Optimization Objective
Telescoped reaction from Boc-protected $\text{S}$ -Bromo-1-methyl-tetrahydroisoquinoline to Boc-protected $\text{S}$ -Acetyl-1-methyl-tetrahydroisoquinoline <sup>433</sup>	Residence time in first reactor, equivalents of reagent, temperature in first reactor, flow rates	Flow	HPLC (multi-point sampling)	BO	Yield
$\text{S}_{\text{N}}$ Ar reaction between 2,4-difluoronitrobenzene and morpholine <sup>434</sup>	Solvent, residence time, concentration, equivalent, temperature	Flow	HPLC	mixed variable multi-objective optimization (MVMOO), using GPs	Pareto-front between yield of ortho-product and yield of para-product
Sonogashira coupling of 2-Bromo-4-(trifluoromethyl)benzonitrile and 3,3-Dimethyl-1-butyne <sup>434</sup>	Phosphine ligand, residence time, reagent equivalents, temperature	Flow	HPLC	mixed variable multi-objective optimization (MVMOO), using GPs	Pareto-front between space-time-yield and reaction mass efficiency
Sulfide oxidation to sulfoxides <sup>435</sup>	Equivalents catalyst, equivalents $\text{H}_2\text{O}_2$ , temperature, residence time	CSTR cells	GC/MS	BO	Yield
Several palladium-catalyzed C–H activation reactions yielding oxindoles from their corresponding chloroacetanilides <sup>436</sup>	Residence time, temperature, catalyst concentration, solvent, ligand	Flow	HPLC	Multi-Task BO	Yield
Schotten-Baumann reaction for acetylation of benzylamine <sup>437</sup>	Reagent equivalents, flow rates, electrophile, solvent	Flow	HPLC	BO, TS-EMO	Space-time yield, F-factor
Lithium-halogen exchange <sup>438</sup>	Residence time, temperature, reagent equivalents	Flow	UPLC-MS	TS-EMO	Pareto front between: yield & impurity
C–H alkylation via photocatalytic HAT <sup>439</sup>	Substrate concentration, THF loading, catalyst loading, residence time, light intensity	Flow	Inline NMR	BO	Yield
R–H trifluoromethylthiolation via photocatalytic HAT (multiple substrates) <sup>439</sup>	Reactant concentration, H-Donor loading, catalyst loading, residence time, light intensity	Flow	Inline NMR	BO	Pareto front between: yield, throughput
Oxytrifluoromethylation via photocatalytic SET <sup>439</sup> (multiple substrates)	Reactant concentration, $\text{CF}_3$ source loading, $\text{CF}_3$ source, catalyst loading, residence time, light intensity	Flow	Inline NMR	BO	Pareto front between: yield, throughput
Aryl trifluoromethylation via photocatalytic SET <sup>439</sup> (multiple substrates)	Reactant concentration, Catalyst, Reagent loadings, residence time, light intensity	Flow	Inline NMR	BO	Pareto front between: yield, throughput
$\text{C}(\text{sp}^2)\text{-C}(\text{sp}^3)$ cross-electrophile coupling (multiple substrates) <sup>439</sup>	Reactant loading, reactant concentration, ligand type, photocatalyst type, photocatalyst loading, residence time, light intensity	Flow	Inline NMR	BO	Pareto front between: yield, throughput

**Table 4.** continued

Reaction	Optimization Parameters	Synthetic Hardware	Analytical Hardware	Optimization Algorithm	Optimization Objective
Alkyne iodination (multiple substrates, joint optimization) <sup>440</sup>	Alkyne group, iodinating reagent, equivalents, catalyst, catalyst equivalence, temperature	Robotic system	HPLC	BO-based proprietary algorithms	Conversion, yield
Hydroformylation of 1-octene (multiple optimization campaigns with different phosphine ligands) <sup>441</sup>	Reagent flow rates, total reaction pressure, temperature, dilution, ligand to metal ratio, olefin to metal ratio	Flow	GC	BO	Yield, selectivity
Four-component Ugi reaction <sup>442</sup>	Reactant volumes, solvent volume, time, temperature	Flow	<sup>19</sup> F NMR	BO	Yield
Van Leusen oxazole synthesis <sup>442</sup>	Reactant volumes, time, temperature	Flow	HPLC	SNOBFIT	Yield
Manganese-catalyzed styrene epoxidation <sup>442</sup>	Catalyst volume, reactant volume, addition speed, time	Flow	Online Raman	Phoenix BO	Product / reactant peak area ratio
Trifluoromethylation reactions after chemical space exploration <sup>442</sup>	Reactant volumes, temperature, time	Flow	<sup>19</sup> F NMR	BO	Yield
Cyclization reaction between toluenesulfonylmethyl isocyanide and benzylidene malononitrile <sup>442</sup>	Concentration, solvent volume, temperature, time	Flow	HPLC	BO	Yield
Cyclization between phloroglucinol, benzylidene malononitrile and 1,8-bis(dimethylamino)naphthalene <sup>442</sup>	Concentration, solvent volume, temperature, time	Flow	HPLC	TS-EMO	Pareto front between cost and yield
Aldol reaction between benzaldehyde and acetone <sup>131</sup>	Acetone equivalents, NaOH equivalents, time, temperature	Flow	HPLC	TS-EMO	Pareto front between cost and yield

<sup>a</sup>A description of the performed reaction, the parameters considered in the optimization, the used hardware and optimization algorithms, and the objective to be optimized are provided.



**Figure 11.** Micro-reactor self-optimizing platform for a Knoevenagel condensation and the oxidation of benzyl alcohol towards benzaldehyde. (a) Overview of the system consisting of a control center to adjust residence time, temperature and concentrations to control the flow from the syringe pumps in the micro-reactor. Products are detected via HPLC, which returns the data to the control unit. (b) Image of the used micro-reactor. (c) Packaging scheme for the microreactor including fluidic connections in the top plate (1), a recessed plate (2) to house the microreactor and TE device, and baffled heat exchanger (3) for sufficient heat removal and additional temperature control. Figure adapted with permission from McMullen *et al.*<sup>269</sup> Copyright 2010, American Chemical Society.

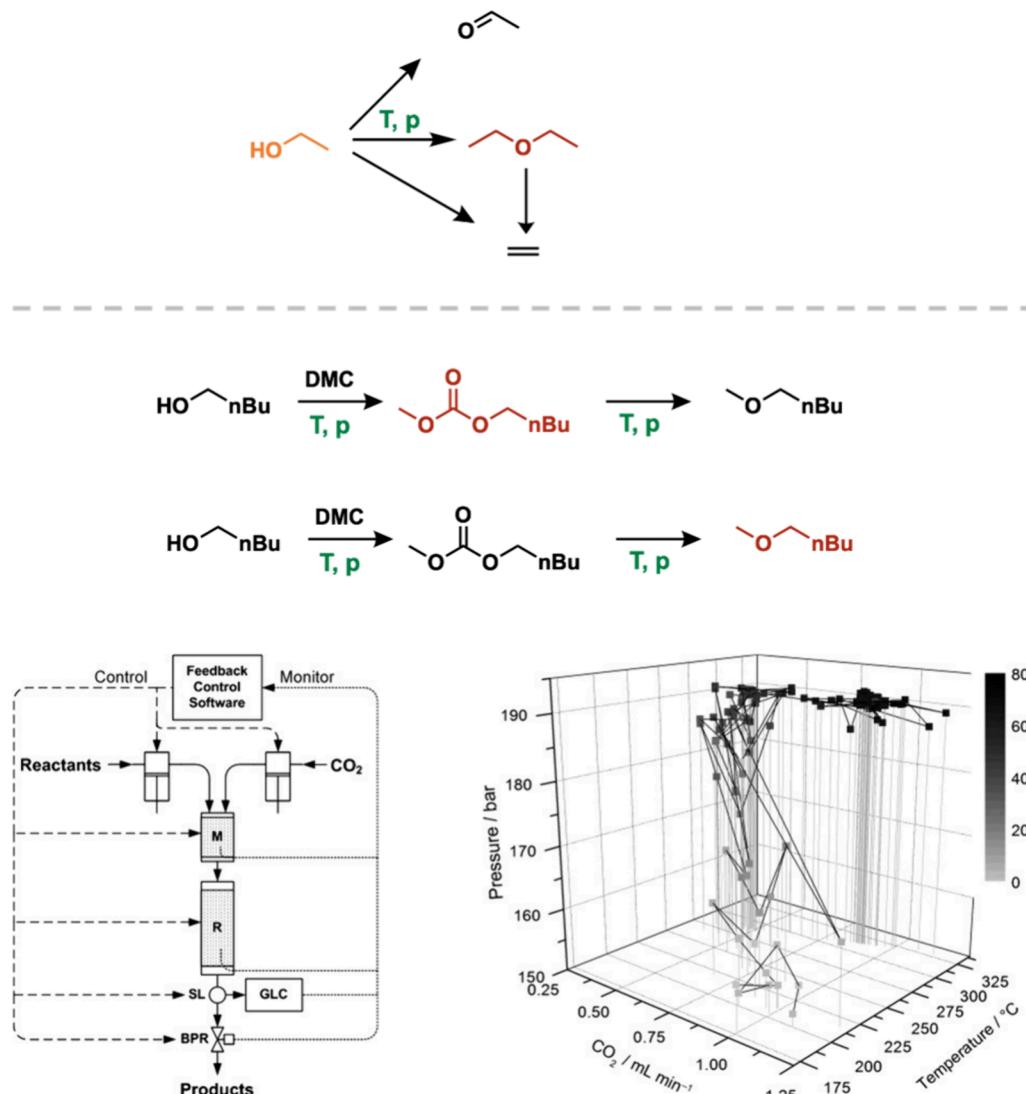
objectives in multi-objective optimization, and highlights the importance of multi-objective optimization algorithms. This initial development has inspired the adoption of in-line IR techniques in a variety of self-optimizing platforms, such as the optimization of an Appel reaction,<sup>117</sup> a [2+2] cycloaddition<sup>413</sup> and others,<sup>116,265,403,419</sup> pointing out the efficiency improvements, owing to the ability to circumvent time-intensive chromatographic methods.

In 2015, Sans *et al.* reported the use of in-line NMR spectroscopy for the self-optimization of the acid-catalyzed imine formation between 4-fluorobenzaldehyde and aniline.<sup>404</sup> Recorded <sup>1</sup>H NMR spectra were automatically phased and baseline-corrected, and the peak integrals were automatically evaluated to optimize an objective function related to the space-time-yield (Figure 14). Even though benchtop NMR spectrometers are commercially available at a reasonable price, their low sensitivity, as well as the difficulty of identifying and resolving characteristic signals, have prevented a more widespread usage in self-optimization platforms. A notable exception is the synthesis of the natural product Carpanone by Felpin and co-workers.<sup>416</sup>

In the same year, Holmes *et al.* demonstrated the usage of online quantitative MS for self-optimizing the synthesis of N'-methyl nicotinamide from methyl nicotinate and aqueous methylamine, varying the flow rate of methyl nicotinate, the quantity of methyl amine and the temperature as continuous independent variables (Figure 15).<sup>406</sup> Before the self-optimization experiments, HPLC was used to calibrate a

benchtop MS, in order to use the latter for product quantitation without prior purification in the SDL campaign. The authors compared a self-optimization using the SNOBFIT algorithm<sup>335</sup> with a classical DoE statistical design approach, finding that both methods found high-yielding conditions. However, while the SNOBFIT algorithm took 12 hours to find the optimum, the DoE approach only took 5.5 hours, due to the human intuition provided in the DoE: heating and cooling of the reactor is time-consuming, so avoid large jumps of temperature in the selected experimental parameters. Despite being commonly used for product identification in conjunction with HPLC, online MS was not widely adopted for quantification in self-optimizing platforms. Other examples are the hydrolysis of 3-Cyanopyridine by Ley and co-workers,<sup>117</sup> and the optimization of multiple reactions with RL by Zare and co-workers (vide infra).<sup>336</sup>

In addition to the integration of more analytical techniques into self-optimizing systems, the used reaction platforms have also seen a significant diversification. On the level of equipment, Fitzpatrick *et al.* have demonstrated the LeyLab, whose components were designed to communicate via the Internet and are thus accessible through every browser with Internet Access.<sup>117</sup> The LeyLab consists of four parts, a graphical user interface (GUI), a database for information storage, an equipment communication module and an equipment command module. Among others, they used this platform to optimize the Appel reaction of 1-phenylethanol, or the hydrolysis of 3-Cyanopyridine over a heterogeneous MnO<sub>2</sub>



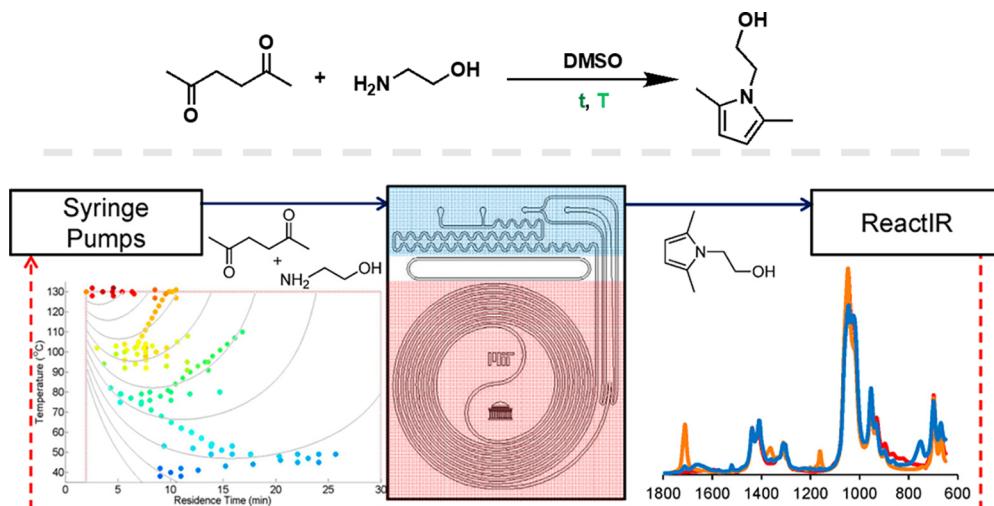
**Figure 12.** Self-optimization campaigns for the dehydration of ethanol and methylation of 1-pentanol in supercritical  $\text{CO}_2$  as reported by Poliakoff and co-workers. The optimizations varied the  $\text{CO}_2$  flow rate, temperature and pressure to optimize the yield of the products colored in brown. Bottom figures show the layout of the used reaction hardware (left) and the optimization campaign of the final reaction (right). Figure adapted with permission from Parrot *et al.*<sup>400</sup> Copyright 2011, John Wiley and Sons.

catalyst using a flow reactor setup. The same group further used their internet-based lab in an across-the-world optimization of the syntheses of multiple active pharmaceutical ingredients.<sup>116</sup> The optimization was initiated by a researcher residing in Los Angeles (California, USA), directed by remote servers in Japan and carried out in Cambridge (UK). Similarly, Skilton *et al.* demonstrated remote controlled self-optimizing reactors,<sup>128</sup> where collaborators from China, Ethiopia and Brazil directed the optimization of self-etherification of n-propanol and methylation of n-butanol and n-propanol through the cloud. In their commentary, the authors note that “watching an optimization in progress can be quite addictive, rather like watching the bids rising during an eBay auction” and further comment on the safety issues, intellectual property and financing of such cloud-based laboratories.

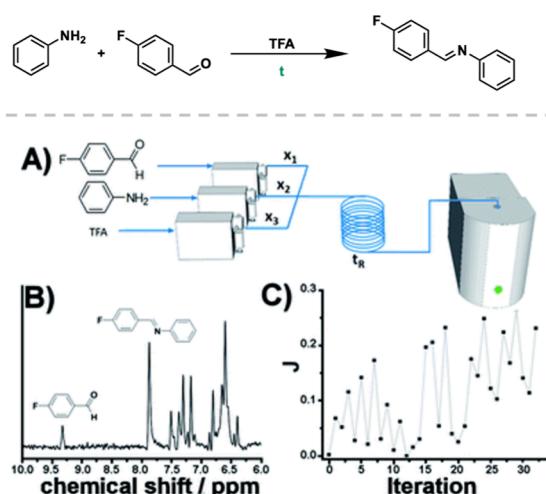
A modular flow system was introduced in 2018 by Bédard *et al.* for autonomous reaction optimization.<sup>265</sup> The system consisted of several bays that each could fit a modular unit, e.g., a photo-reactor, a heated reactor, a cooled reactor, a packed bed reactor, a liquid-liquid separator or a bypass (Figure 16).

Moreover, the system was connected to in-line analytics such as HPLC, MS, IR and Raman. The authors showcased the modularity of the platform by optimizing the conditions for maximal yield of a multitude of reactions, namely a Buchwald-Hartwig Cross coupling, a HWE Olefination, a reductive amination, a Suzuki-Miyaura cross coupling, an  $\text{S}_{\text{N}}\text{Ar}$  reaction, a photoredox reaction and a multi-step ketene generation followed by a 2+2 cycloaddition. For each reaction, the authors manually designed the appropriate flow system, which was then used to autonomously optimize reagent equivalents, residence times and bay temperatures as independent variables. After each successful optimization campaign, the optimal conditions were examined for different substrates. In one case, where the conditions did not lead to a satisfactory yield for a specific substrate, a re-optimization was conducted with a subset of variables within 6 hours, improving the yield from 67% to 97%, demonstrating the flexibility and efficiency of the flow platform.

Despite the prevalence of flow reactors, other reactor types have also been applied recently in self-optimization campaigns.



**Figure 13.** Demonstration of a ReactIR system in the self-optimization of the Paal-Knorr synthesis of a pyrrole. In the bottom, a scheme of the utilized flow system is demonstrated along with results of the two-dimensional optimization campaign (left) and a sample IR spectrum (right). Figure adapted with permission from Moore and Jensen.<sup>402</sup> Copyright 2012, American Chemical Society.



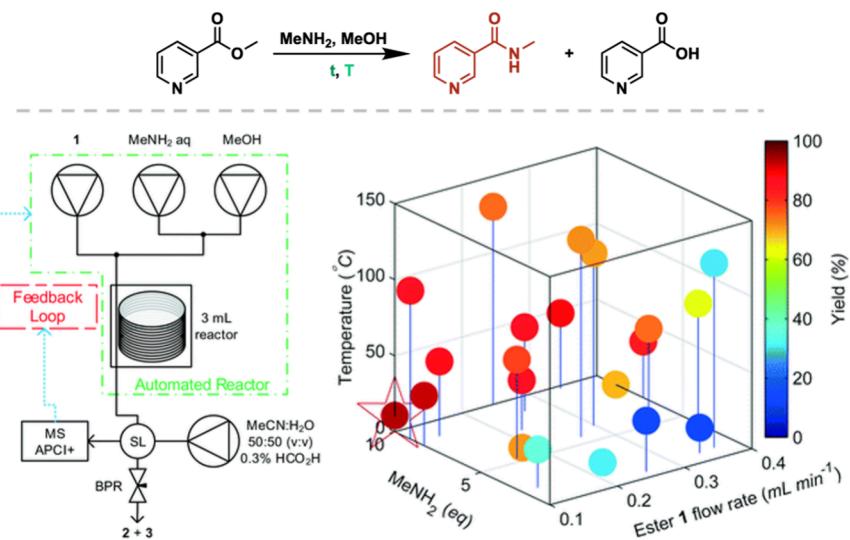
**Figure 14.** Demonstration of an NMR system in the self-optimization of the condensation between aniline and p-Fluorobenzaldehyde. (A) Scheme of the utilized flow system is demonstrated showing the input streams of the reagents and trifluoroacetic acid (TFA) as catalyst, leading to the NMR spectrometer for analytical measurement. (B) Sample NMR spectrum from the optimization campaign, indicating characteristic product and reactant peaks. (C) Results of the optimization campaign to optimize objective function  $J$  over multiple iterations. Figure adapted with permission from Sans *et al.*<sup>404</sup> Copyright 2014, Royal Society of Chemistry.

In particular, Clayton *et al.* demonstrated multiple cascaded CSTR reactors which can provide conditions similar to a flow system while decoupling mixing performance from flow rate, thereby facilitating multiphasic reactions.<sup>421</sup> Further, they also allow experiments with reactions involving solids and slurries, for which clogging is often a problem in conventional flow systems. The latter was utilized by Nandiwale *et al.* in 2022 for the successful self-optimization of a Pd-catalyzed Suzuki-Miyaura coupling, as well as two metallaphotoredox-catalyzed  $C_{sp^3}-C_{sp^3}$  and  $C_{sp^3}-C_{sp^2}$  couplings of alkyl carboxylic acids and halides, respectively.<sup>428</sup> Each of the investigated reactions involved at least one solid reactant, catalyst, additive or product, which could be transferred as a slurry in the CSTR.

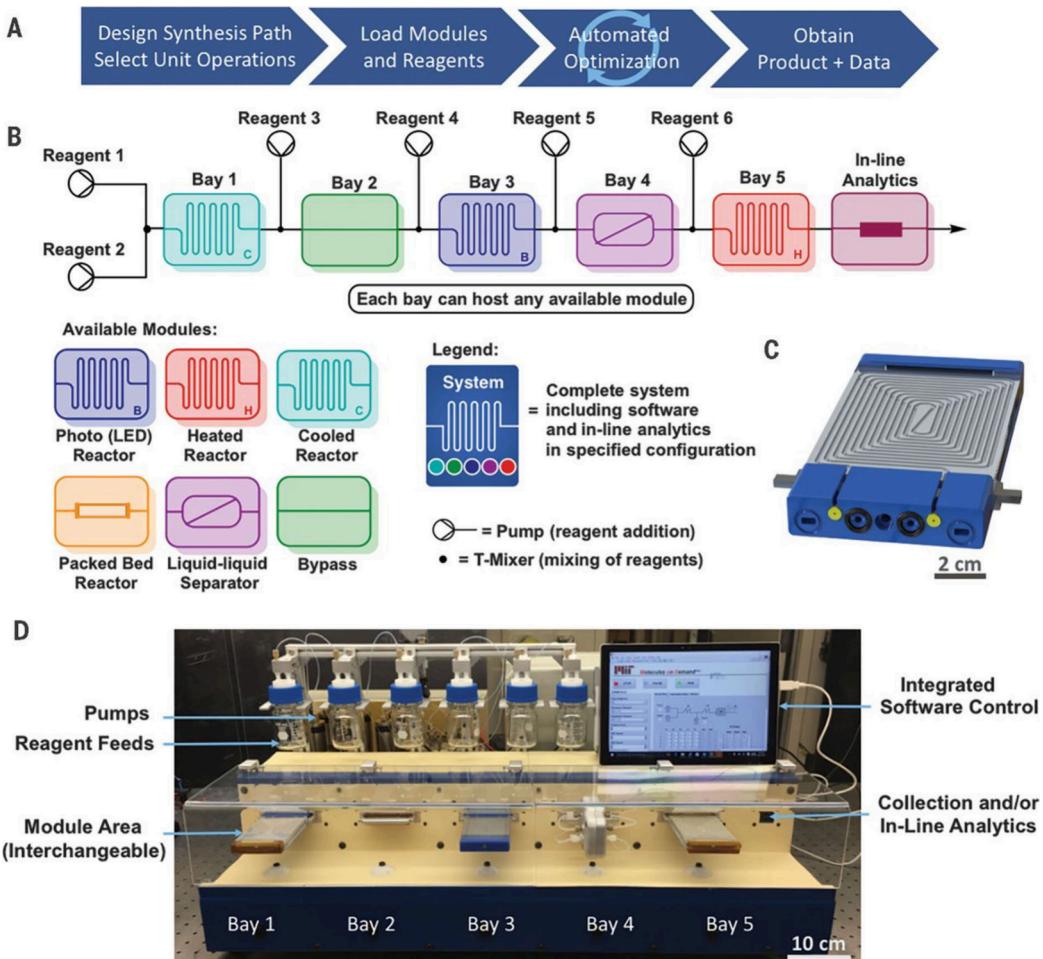
Leonov *et al.* report the development of an integrated self-optimizing programmable chemical synthesis and reaction engine.<sup>442</sup> They incorporated various sensors, including those for monitoring color, temperature, conductivity, pH, and liquid transfers, into their previously discussed Chemputer robotic platform. Additionally, they integrated analytical instruments like HPLC, NMR, and Raman spectroscopy, enabling closed-loop reaction optimization via feedback control. Adaptive execution of chemical procedures on the Chemputer was made possible by the dynamic  $\chi$ DL programming language. The authors demonstrated the platform's capabilities through temperature-controlled reagent additions, optical endpoint detection, and hardware failure detection. The authors optimized several organic reactions, including the Ugi four-component reaction, Van Leusen oxazole synthesis, manganese-catalyzed epoxidation, and trifluoromethylation reactions, utilizing various optimization algorithms like BO with GP surrogate, Phoenics BO, SNOBFIT, and genetic algorithms. The optimization led to improved product yields of up to 50%. Furthermore, the authors showcased an experimental pipeline for exploring unknown reaction spaces, combining digital discovery and optimization, exemplified by the discovery and optimization of two previously unreported reactions.

**4.2.2. Discrete and Categorical Optimization and Batch-Type Reactors.** The previously discussed studies have primarily focused on the optimization of continuous variables such as reagent stoichiometry, temperature or reaction time. Chemical reactions, however, are highly governed not only by the continuous parameters defining the process details, but also, most importantly, by the involved reactants and reagents, which are inherently categorical parameters. As discussed above, such optimization over categorical variables requires specific adaptations both in terms of software and hardware, and are often better suited for parallel batch reactor setups.

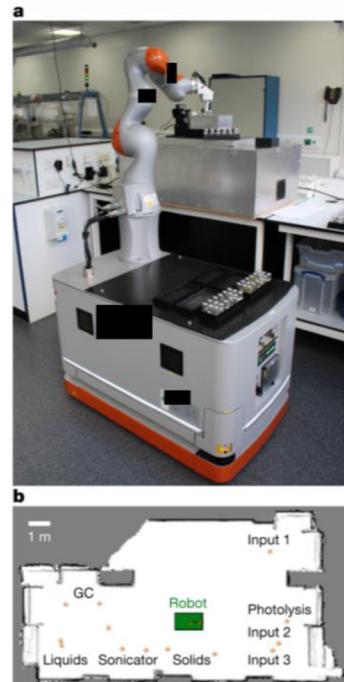
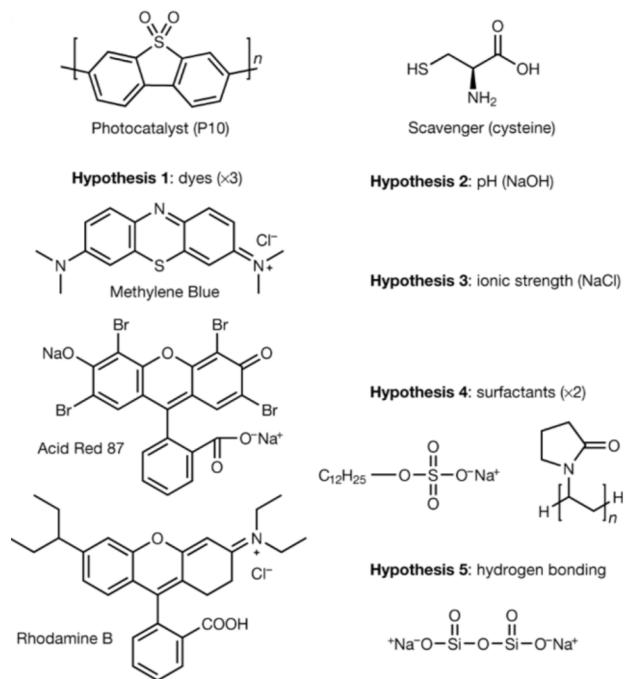
To the best of our knowledge, the first example of autonomous, closed-loop reaction optimization in batch reactors (since the early examples from the 1980s) was reported by Burger *et al.* in 2020,<sup>11</sup> tackling the homogeneous photocatalytic water splitting reaction. Notably, their work stands out owing to the highly advanced robotic setup used for performing and analyzing reactions (Figure 17). In this work,



**Figure 15.** Demonstration of an MS spectrometer in the self-optimization for the amidation of methyl nicotinate with aqueous MeNH<sub>2</sub>. The optimization considered the temperature, the reactant flow rate and MeNH<sub>2</sub> equivalents to optimize the yield of the product drawn in brown. In the bottom row, a schematic overview over the used hardware is given (left) and the experimentally tested conditions in the optimization campaign (right). Figure adapted with permission from Holmes *et al.*<sup>406</sup> Copyright 2016, Royal Society of Chemistry.



**Figure 16.** Modular reaction platform as proposed by Bédard *et al.* (A) The general four-step workflow consisting of the design of the synthesis path, loading the module, the self-optimization and obtaining the final results. (B) Construction of the modular system consisting of several bays that can contain multiple different modules and reagent inlets, leading to an in-line analytics system. (C) CAD representation of the LED reactor. (D) Schematic picture of the modular reactor platform. Figure adapted with permission from Bédard *et al.*<sup>265</sup> Copyright 2018, American Association for the Advancement of Science.

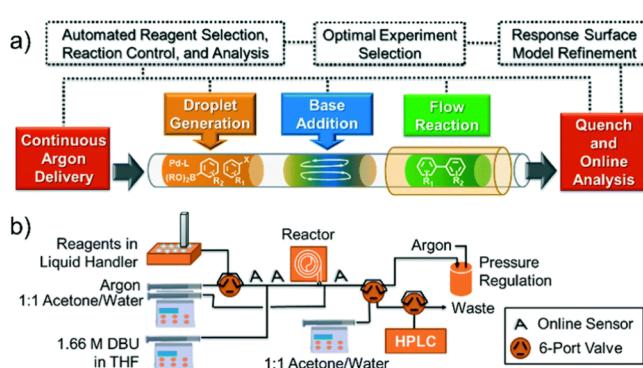


**Figure 17.** Mobile robotic chemist as demonstrated by Burger *et al.* Schematic (left) shows the ten chemicals for which the concentrations were varied to optimize H<sub>2</sub> production. Picture (right) portrays the (a) mobile robotic chemist with several workstations as well as (b) the general layout of the laboratory, with the mobile robot transferring the samples between the stations. Figure adapted with permission from Burger *et al.*<sup>11</sup> Copyright 2020, Springer Nature.

the authors introduce their “mobile robotic chemist” (for a more detailed discussion, see section on [Hardware](#)), a KUKA mobile robot that is designed to operate human-centric workstations. The robot transfers reactors between workstations for solid dispensing, liquid dispensing, inertization, capping and GC analysis. They utilized their robot to investigate the water splitting reaction catalyzed by the photoactive polymer P10. However, to circumvent the need for categorical optimization, the authors treated the quantities of each additive as a continuous variable, enabling the use of an off-the-shelf GP surrogate with an upper confidence bound acquisition function for BO. With their highly advanced experimental setup, the authors demonstrate 43 fully autonomous batches of experiments in approximately 8 days, resulting in an almost 10-fold increase in hydrogen evolution.

Similarly, Ha *et al.* recently reported SynBot,<sup>432</sup> a platform for autonomous organic synthesis in batch reactors, which was demonstrated for carbon-coupling reactions. SynBot consists of an AI layer, an AI–Robot layer and a Robot layer. As the AI layer, the authors trained a retrosynthesis model as well as a GNN that proposes suitable reaction conditions in combination with BO on a search space that consists of commonly used catalysts, bases and solvents for multiple reactions: Suzuki coupling, Buchwald amination, and Ullmann reaction. The SDL features an integrated robotic system capable of executing various tasks, including chemical dispensing, reaction handling, sampling, and analysis. The system aims to iteratively refine and optimize synthetic routes and reaction conditions in order to maximize the reaction yields.

As an alternative to classical batch reactors, Jensen and co-workers reported a series of SDLs using a segmented-flow system in which each droplet—i.e., each “batch”—contains a specific reaction with a unique set of conditions ([Figure 18](#)).<sup>405</sup> In their first work from 2015, Reizman *et al.* use this platform



**Figure 18.** Demonstration of an example of a micro-droplet reactor for a Suzuki-Miyaura cross-coupling. (a) In the beginning of the reactor, droplets are generated with reagents selected by the optimization algorithm. Further down in the reactor, a base is added and the reaction mixture reacts further down in the reactor within the droplet, creating “batch in flow” conditions. After the reaction, the droplet is quenched and directed to an online analysis station. (b) Diagram shows the integration of the micro-droplet reactor into a self-optimization platform. Figure adapted with permission from reference Reizman *et al.*<sup>408</sup> Copyright 2016, Royal Society of Chemistry.

for screening potential solvents and optimizing continuous reaction conditions for the mono-alkylation of *trans*-1,2-diaminocyclohexane. To address this mixed categorical–continuous optimization, the authors performed an initial fractional factorial DoE for every solvent, followed by another fractional factorial design at experimental conditions close to the predicted optimum, and a feedback DoE search to minimize the uncertainty on the maximum predicted yield for each solvent separately. Subsequently, insufficiently performing solvents were disregarded and an automated

gradient-based search around the predicted optimum was carried out for the remaining solvents to optimize the yield. Similar principles for the incorporation of categorical parameters in self-optimizations were subsequently used to select optimal precatalyst scaffolds and ligands for a Suzuki–Miyaura coupling,<sup>408,412</sup> organic base and Ni precatalyst for a photoredox Ir–Ni dual-catalyzed decarboxylative arylation,<sup>411</sup> organic base for several Pd-catalyzed C–N coupling reactions,<sup>420</sup> catalyst for a Suzuki–Miyaura coupling, base for a metallaphotoredox-catalyzed  $sp^3$ – $sp^3$  cross-coupling of carboxylic acids with alkyl halides or photocatalyst for a metallaphotoredox-catalyzed decarboxylative cross-coupling reaction.<sup>428</sup>

Slattery *et al.* made use of readily available internet-of-things phase sensors to detect the relevant reaction slugs in their flow system, dubbed RoboChem.<sup>439</sup> The authors integrated off-the-shelf hardware and custom software to build a modular platform, which contained a GUI to enable operation by non-expert chemists. The platform contained a light source with tunable intensity, enabling the autonomous optimization and scale-up of a multitude of photo-catalyzed reactions. The RoboChem platform employs multi-objective BO, as implemented in Dragonfly, to autonomously plan and execute experiments. This autonomous experimentation capability allows the system to explore complex chemical spaces efficiently, identifying optimal reaction conditions tailored to each substrate. The authors demonstrated the platform's versatility by optimizing a diverse set of 19 photocatalytic transformations, including hydrogen atom transfer photocatalysis, photoredox catalysis, and metallaphotocatalysis, which are relevant to pharmaceutical and agrochemical synthesis.

An alternate approach integrates the selection of categorical reaction variables directly through a suitable encoding of chemicals into appropriate optimization variables, rather than creating a separate response surface for each categorical reaction variable and comparing the response surfaces. This was done by framing the choice of each categorical variable through one-hot encoding all categorical possibilities or calculating descriptors for each categorical variable. The former method was used to find an optimal base for a regioselective S<sub>N</sub>Ar reaction, a suitable phosphine ligand for a Sonogashira coupling,<sup>434,444</sup> optimal solvents and phosphine ligands for multiple C–H activation reactions,<sup>436</sup> optimal electrophiles and solvents for a Schotten–Baumann reaction.<sup>445</sup>

The selection of expert-crafted, physically meaningful descriptors is an important strategy to introduce additional knowledge into the optimization campaign.<sup>366</sup> In 2021, Christensen *et al.* reported on this concept for the optimization of a stereoselective Suzuki–Miyaura coupling.<sup>425</sup> Notably, the authors used a batch system for reaction execution, namely a Chemspeed SWING platform coupled to a HPLC-UV system, to run parallel reactions in 96-well plates. The use of this batch reactor system enabled the authors to optimize a wide, representative set of 23 phosphine ligands selected in a fully data-driven fashion. Furthermore, the authors considered several other continuous parameters such as reaction temperature, palladium loading, boronic acid equivalents and phosphine to palladium ratio to optimize the yield of the *E*-diastereomer, while minimizing the *Z*-diastereomer yield and the quantities of used reagents. Mixed continuous-categorical optimization was performed using the Gryffin package relying on a BNN surrogate.<sup>446</sup> While finding a similar optimum, the

descriptor-based optimization campaign converged slower than a reference campaign based on one-hot encoding only, which was attributed to the introduction of unproductive bias through the selected descriptors.

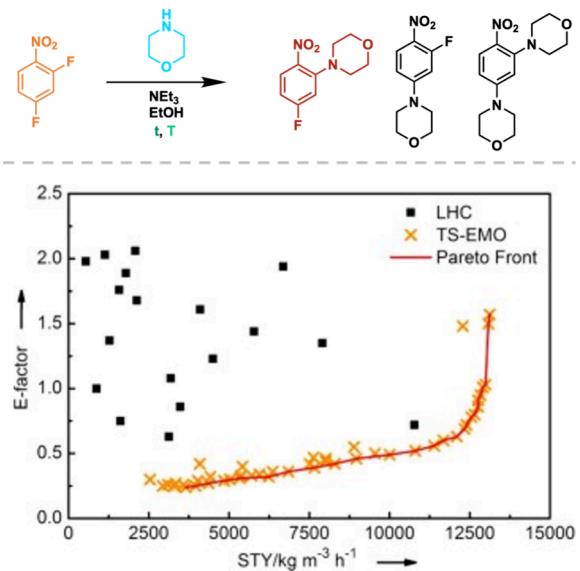
A remarkable example of categorical optimization was reported by Angello *et al.* in 2022, who—rather than optimizing reaction conditions for a single substrate combination—targeted the discovery of general reaction conditions.<sup>431</sup> The authors defined the most general conditions of a reaction type as those conditions that provide the highest average yield across the widest range of substrate space. The authors showcase this concept at the example of heteroaryl Suzuki–Miyaura couplings using protected boronic acids. To identify the “widest range of substrate space”, data-driven clustering techniques were employed to identify a set of 11 representative reactions for which general conditions should be identified. Optimization was performed over the identity of solvent, base, catalyst and ligand, and the reaction temperature as independent variables. Experiments are performed on a custom-built automated reaction platform that is capable of performing 36 parallel batch reactions with 20 distinct reagents under inert gas conditions.<sup>447</sup> The authors developed a custom BO workflow for maximizing generality, the yield over multiple reactions. Notably, the fully explorative acquisition strategy is designed in a way that it does not require evaluating each of the 11 representative reactions in every iteration. Using this approach, the authors managed to efficiently cover a wide space of conditions and substrates, ultimately identifying conditions that double the average yield compared to benchmark general conditions. Related work in optimization of generality of reaction conditions have also been done, although no automation is involved.<sup>448</sup>

While not optimizing for general reaction conditions, Schilter *et al.* recently performed a simultaneous optimization over multiple substrates.<sup>440</sup> Using a robotic batch system containing six reactors, the authors performed a single optimization campaign in which the yield and conversion for an alkyne iodination was jointly optimized for multiple substrates. Notably, in the optimization campaign, the substrate was a parameter that could be chosen by the optimization algorithm to optimize the reaction conditions. In order to find the optimal conditions for all of the substrates, a substrate could not be selected by the algorithm after a satisfactory performance (conversion > 80%) was obtained. Remarkably, the optimization campaigns showed high transferability as the optimization run was primarily conducted on one of the substrates, and after a satisfactory performance was obtained for this substrate, the same was achieved for the other substrates of interest, requiring a total of only 23 experiments to find suitable conditions for all substrates.

**4.2.3. Pareto Optimizations and Further Algorithmic Advances.** So far, all discussed SDL reaction optimization campaigns were conducted as a single-objective optimization, where the objective is reaction yield in most cases. Optimizing for multiple objectives allows researchers to consider multiple metrics of a reaction, such as yield, conversion, productivity or ecological factors. The most straightforward approach is to scalarize multiple objectives into a single objective value, which, however, requires pre-defining an often unknown trade-off between different objective values. In an earlier example of this section, we have shown that this can lead to undesirable outcomes of the optimization campaign, particularly if the two objective functions show opposing trends. In such a scenario,

an algorithm which uncovers the Pareto front of non-dominant optima would be highly desirable.

In 2018, Schweidtmann *et al.* used the TS-EMO algorithm<sup>318</sup> to identify the Pareto front between the space-time yield and E-factor for a regioselective S<sub>N</sub>Ar reaction (Figure 19), as well as between the space-time yield and



**Figure 19.** Exemplary results for a Pareto front optimization of the E-factor and the space-time yield for an S<sub>N</sub>Ar reaction. Identified Pareto optimal points are marked in orange, the interpolated Pareto front is drawn in red. In the left part of the Pareto front, a strong increase in space-time yield can be achieved by a minimal increase in E-factor, whereas the right side of the plot shows a strong increase in E-factor with only a small increase in space-time yield. Figure adapted with permission from reference Schweidtmann *et al.*<sup>417</sup> Copyright 2018, Elsevier.

impurity yield for a benzylation reaction of a primary amine.<sup>417</sup> The elucidation of the entire Pareto front allowed the researchers to identify suitable trade-offs, which was particularly useful in the S<sub>N</sub>Ar reaction, where the space-time yield could be significantly improved while almost not impacting the E-factor. Such a relation would not have been uncovered if only one optimal point was identified.

Building on this work, Jeraal *et al.* evaluated the Pareto front between yield and cost of the mono-alcohol-condensation of acetone and benzaldehyde using the TS-EMO algorithm.<sup>423</sup> In order to benchmark the algorithm's performance, the authors ran the campaign twice, once with 20 experiments and once with two low-yield (3% and 5%) experiments as a starting set. Both campaigns converged to the same Pareto front, even though the latter run needed roughly twice as many experiments. In addition, the authors demonstrated the general applicability of their approach by further uncovering the Pareto front between the space-time yield and the E-factor. Similarly, Karan *et al.* also employed the TS-EMO algorithm for the Pareto optimization of the yield and impurity for an ultra-fast lithium-halogen exchange reaction.<sup>438</sup> The authors performed three optimization campaigns with either different initial experiments or different reactant mixing equipment, showing that the algorithm efficiently converges to similar Pareto fronts.

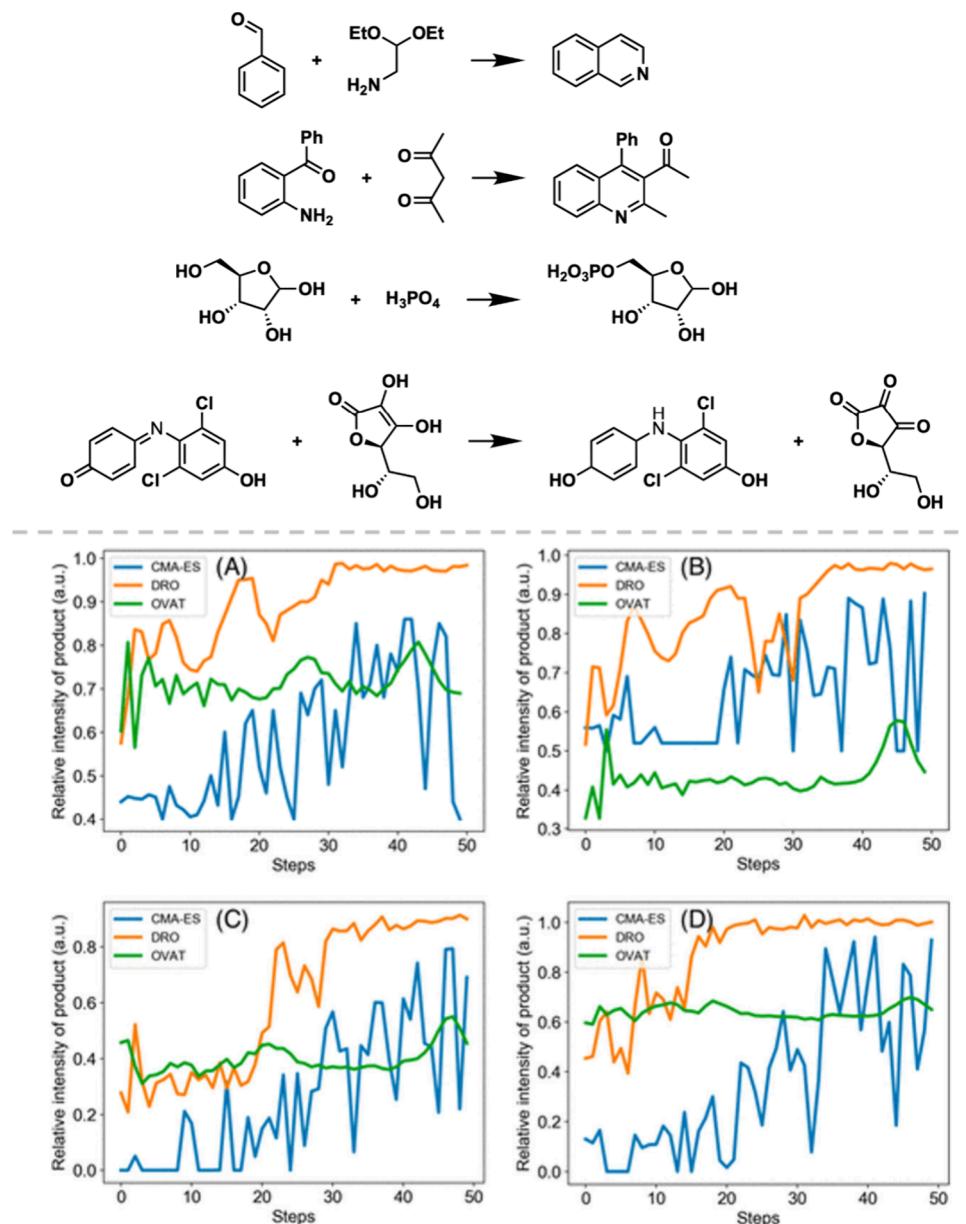
Since the TS-EMO algorithm is computationally expensive for categorical parameters, a GP-based BO with the qNEHVI acquisition function<sup>291</sup> to find the Pareto front for an

optimization with continuous and categorical parameters.<sup>445</sup> After demonstrating improved efficiency over TS-EMO *in silico*, experimental optimization revealed the Pareto front between the space-time yield and E-factor for a Schotten-Baumann reaction. Similarly, using their GP-based mixed-variable multi-objective optimization (MVMOO) algorithm,<sup>444</sup> Kershaw *et al.* identified the Pareto front for the yield of ortho- and para-products of a S<sub>N</sub>Ar reaction, where the optimization variables included continuous and categorical (solvent) variables.<sup>434</sup> Interestingly, their method showed that different solvents are responsible for different regions of the Pareto front, enabling researchers to select the right solvents for the desired product. In a further experiment, the authors uncovered the entire Pareto front between the reaction mass efficiency and space-time yield for a Sonogashira cross-coupling, finding that the Pareto front is almost exclusively dominated by one phosphine ligand.

In 2017, Zhou *et al.* demonstrated the applicability of RL for optimizing chemical reactivity.<sup>336</sup> RL was used to learn a policy that determines the next experiment to conduct, where RNNs were used to fit the policy function. Owing to the high cost of experimental data, the algorithm was pre-trained on cheap simulated reaction data, obtained from non-convex mixture Gaussian density functions with multiple local minima. The performance of the RL algorithm was benchmarked against the Nelder-Mead simplex method, the SNOBFIT algorithm and the CMA-ES<sup>449</sup> on the simulated data, and was found to outperform all of the established methods on average. However, no benchmarking against standard BO algorithms was performed. The pre-trained policy was then integrated into an SDL using micro-flow reactors with MS quantification, and was used to optimize the conditions of four different reactions: the Pomeranz-Fritsch synthesis of isoquinoline, the Friedländer Synthesis of a Substituted Quinoline, the synthesis of Ribose phosphate and the reaction between 2,6-Dichlorophenolindophenol and ascorbic acid. Again, the RL algorithm was compared with CMA-ES and a OFAT optimization, showcasing that RL consistently outperforms the other two methods. The optimization campaigns were carried out successively, with the policy improving after each completed optimization campaign, demonstrating the function and generalizability of the pre-trained DL agent model (Figure 20).

Recently, Bennett *et al.* developed Fast-Cat,<sup>441</sup> a gas-liquid segmented flow platform suitable for high temperatures and pressures. The platform enables the rapid identification of Pareto fronts for transition-metal catalyzed reactions through BO with the qNEHVI acquisition function. The authors utilized Fast-Cat to identify the Pareto front between the yield and the linear/branched selectivity of the hydroformylation of 1-octene for six different phosphine ligands. Each ligand encompasses different trade-offs between yield and selectivity, demonstrating the importance of efficient automation to uncover optimal conditions. The modular system integrates advanced process automation, in-line reaction characterization using GC, and closed-loop feedback algorithms to dynamically update its belief model and autonomously select new experimental conditions. By leveraging AI approaches, Fast-Cat accelerates reaction space exploration, rapidly identifies optimized conditions, and generates high-quality in-house experimental data to construct digital twins of the catalytic reactions under study.

Bai *et al.* demonstrated a closed-loop distributed SDL within The World Avatar project, aimed at creating a comprehensive

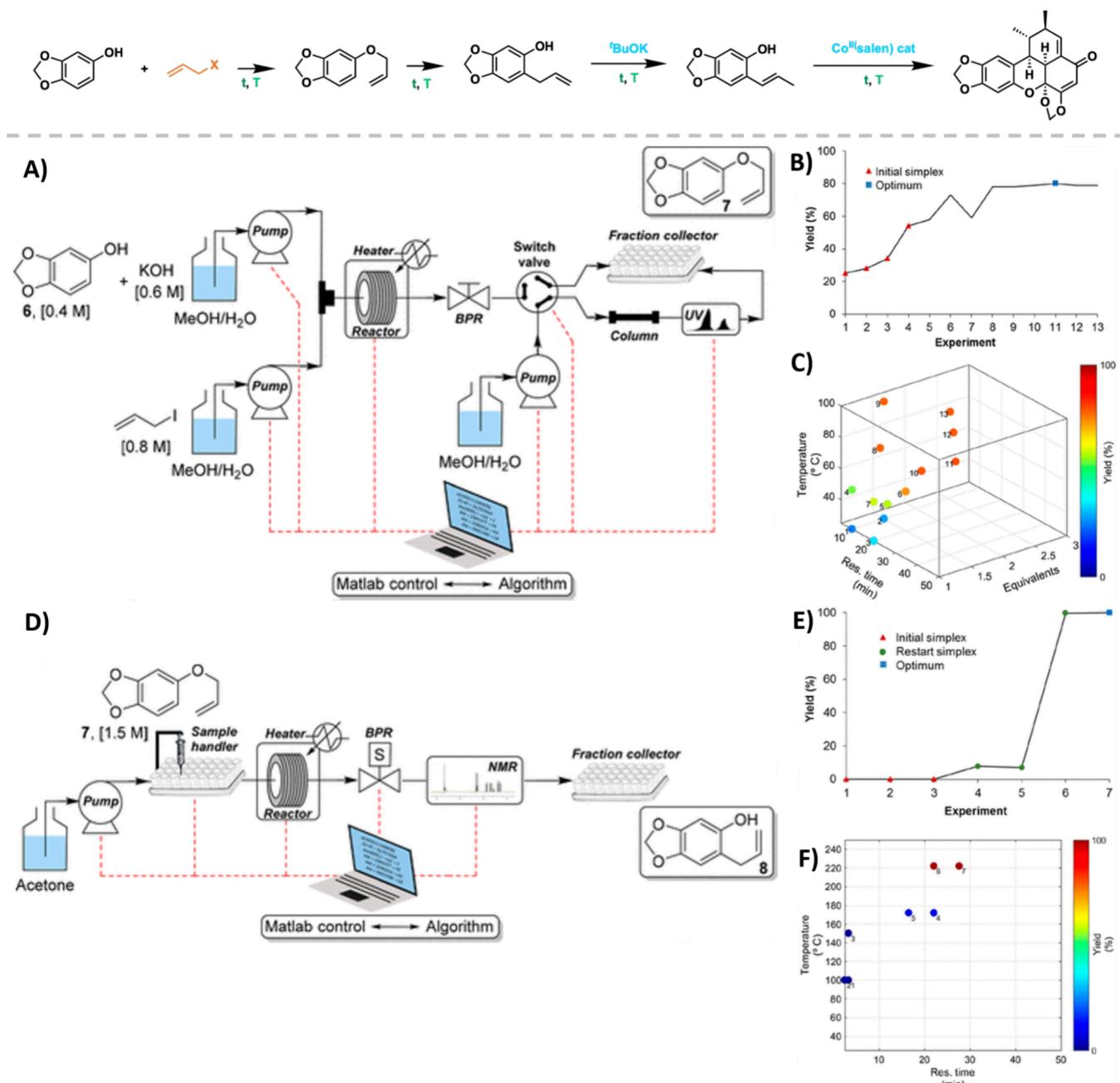


**Figure 20.** Reactions considered for self-optimization using deep RL. The bottom shows the different optimization runs for each of the reactions, respectively; comparing the algorithmic performance of RL (orange), one variable at a time (green) and the CMA-ES (blue). In all cases, the RL methods performed the optimization most efficiently. Figure adapted with permission from reference Zhou *et al.*<sup>336</sup> Copyright 2017, American Chemical Society.

digital twin based on a dynamic knowledge graph.<sup>131</sup> This architecture utilizes ontologies to capture data and material flows in the design-make-test-analyze cycle, and employs autonomous agents to execute the experimental workflows. The authors demonstrated the framework's application by linking two robotic systems in Cambridge and Singapore for a collaborative optimization of a pharmaceutically relevant aldol condensation reaction, mapping out the Pareto front for cost-yield optimization within three days. The optimization was done with the TS-EMO algorithm. This setup involved flow chemistry platforms with automated liquid handling and reagent sourcing, showcasing the integration of dynamic ontological knowledge graphs to streamline and coordinate separate SDLs.

#### 4.3. Multi-Step Organic Reactions

The synthesis of most organic molecules can hardly be achieved in a single step, and can easily require tens of steps for complex natural products. From an SDL standpoint, multi-step reactions can be approached in two distinct ways: (1) each reaction step considered and optimized separately, and the reaction product is purified and isolated before being subjected to the subsequent step; while purification, particularly in batch systems, poses significant hardware challenges, condition optimization can be performed following the approaches discussed in the previous section; (2) alternatively, all steps are run sequentially in the same batch reactor or sequential flow reactors, which is referred to as “one-pot synthesis” or “telescoped synthesis,” respectively. In the latter approach, optimization does not only become a higher-dimensional problem, but the presence of impurities and by-products can

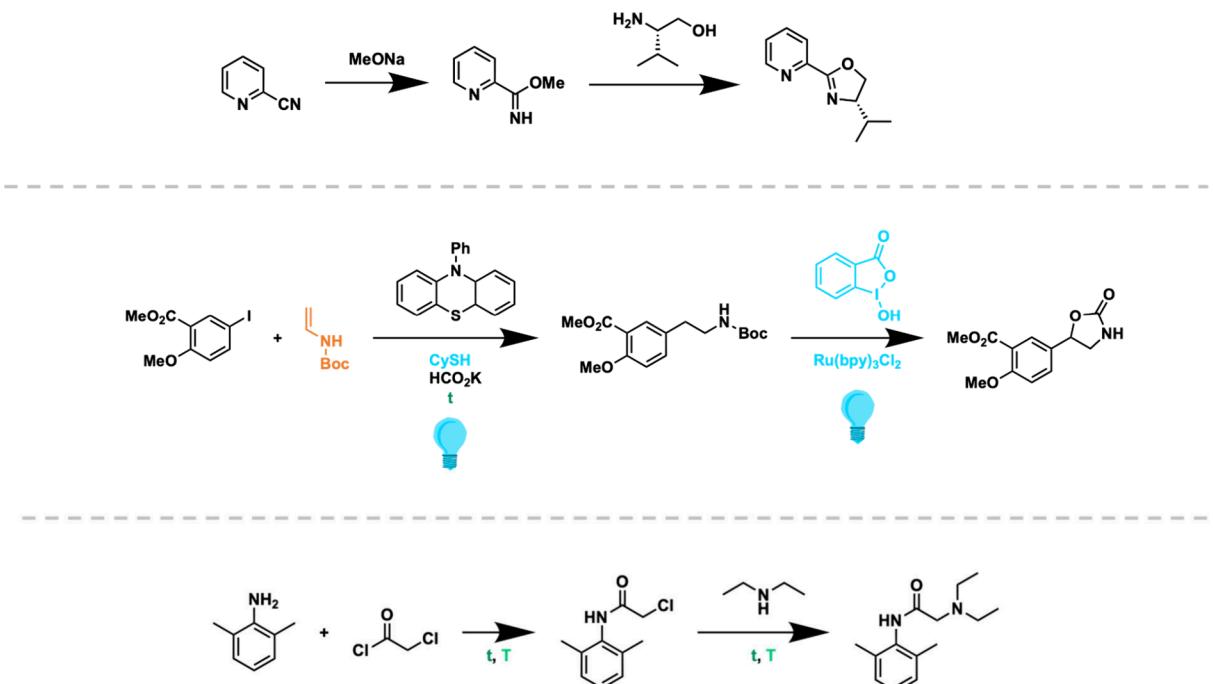


**Figure 21.** Multi-Step Synthesis of the natural product Carpanone, where each reaction was considered as a separate self-optimization experiment. The synthesis was performed via a four-step route, consisting of an allylation of a phenol, a [3,3]-Claisen rearrangement, an isomerization and ultimately an oxidative dimerization. (A) shows the schematic overview of the reactor system for the first step with an HPLC/UV unit as an analytical unit, as well as the results of the optimization campaign (B) per experiment, and (C) as a function of the varied parameters. (D) Schematic experimental overview over the second step is provided, with an in-line NMR as an analytical unit. No further purification was necessary, since the product was obtained in 100% NMR yield. (E-F) The results of the optimization of the second reaction are shown. Figure adapted with permission from reference Cortés-Borda *et al.*<sup>416</sup> Copyright 2018, American Chemical Society.

complicate the optimization of down-stream steps. For example, Coley *et al.* demonstrated a system with a robotic arm that can assemble the required unit operations (reactors, separators) into a continuous flow path according to the recipe, connect reagent lines, and carry out the telescoped reactions.<sup>450</sup> Furthermore, in telescoped systems, flow rates and reaction times cannot be modified independently, posing an additional optimization constraint. This section will first summarize examples that fall under approach (1) and optimize

each step individually, before discussing SDLs that feature self-optimizing telescoped reactors (approach (2)).

**4.3.1. Sequential Single-Step Optimizations.** To our knowledge, the first published example of autonomous optimization of a multi-step reaction was presented by Cortés-Borda *et al.* in 2018, where the authors described the synthesis of the natural product Carpanone.<sup>416</sup> For the four-step synthesis, the authors performed four different self-optimization campaigns, involving allylation, [3,3]-Claisen rearrangement, base-catalyzed isomerization and oxidative



**Figure 22.** Examples of multi-step reactions by combining multiple single step self-optimization platforms. Top: Exemplary synthesis of PyOX ligands by Felpin and co-workers.<sup>418</sup> Middle: Photo-catalytic two-step synthesis of a 2-oxazolidinone derivative performed by Jensen and co-workers.<sup>429</sup> Bottom: Two-step synthesis of lidocaine by Ley and co-workers.<sup>116</sup>

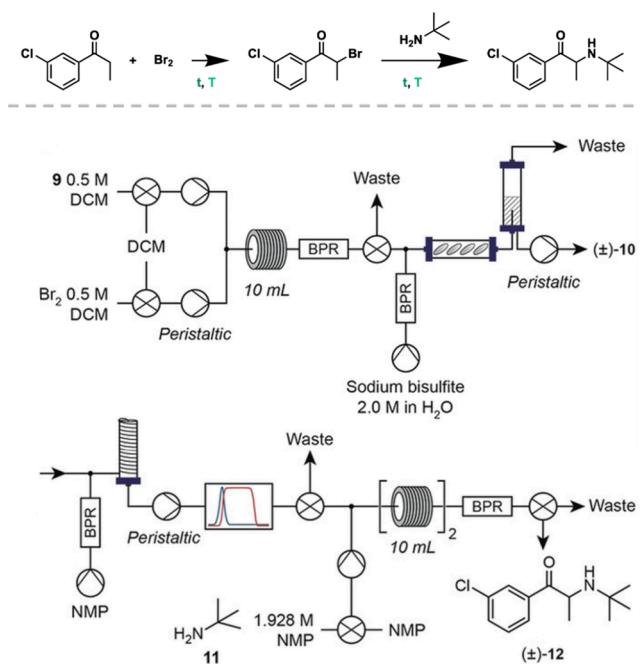
dimerization (Figure 21). For each campaign, up to three continuous variables, corresponding to temperature, residence time and stoichiometry/loading of one reactant species were optimized using a modified simplex algorithm. Depending on the reaction, either the HPLC or an in-line benchtop NMR spectrometer was used. Overall, the authors managed to optimize the synthesis to yield 67% of the natural product Carpanone with a total of only 66 experiments. The fact that it was manageable to conduct multiple different reactions resulting in a highly complex product on the same self-optimizing platform demonstrates the adaptability and efficiency of such systems.

A similar example of multi-step synthesis was reported by the same group in 2019, targeting the two-step synthesis of pyridine-oxazoline (PyOX) ligands (Figure 22, top panel).<sup>418</sup> Performing two sequential optimization campaigns (three and four continuous variables, respectively) using a custom modification of the Nelder-Mead Simplex algorithm, Wimmer *et al.* managed to obtain a yield of 75% with only 34 experiments. Notably, the use of the flow system allowed for a significant divergence from the conditions originally reported in batch reactors: Whereas the first step of the original batch route took place at room temperature overnight, due to the thermal instability of the reaction product, the high heat transfer efficiency of flow systems allowed for shorter reaction times under thermal activation, as revealed by the sequential optimization. Transferability of the conditions was further demonstrated through the synthesis of six similar ligands, with yields ranging from 66%–92%. Related examples of multi-step synthesis SDLs were reported by Jensen and co-workers (Figure 22, middle panel), as well as Ley and co-workers (Figure 22, bottom panel).

From a practical standpoint, maximizing the yield alone is not a sufficient criterion for successful synthesis—the product needs to be isolated from the crude reaction mixture in high

purity, which is usually achieved through phase transfers and phase separations (e.g., extraction, filtration, chromatography). Ley and co-workers, using the LeyLab, reported the autonomous optimization of two two-step syntheses of lidocaine and bupropion, respectively, where each step was optimized separately.<sup>116</sup> In the case of bupropion synthesis, after successful optimization of the reaction conditions for both steps, the authors demonstrated the telescoping of both steps into a single, continuous synthesis process (Figure 23). For this, the authors joined the crude product stream of the first step (bromination) with an aqueous sodium bisulfite stream to quench excess bromine. After mixing and subsequent phase separation, the organic phase was joined with the solvent stream for the second reaction (amination) before being transferred to a thin-film evaporation column, in which the dichloromethane from the first reaction step was selectively evaporated. The outflow of this evaporation column, ideally containing the purified product, was then transferred to the reactor in which the amination occurs. This discussion illustrates the hardware considerations required for successfully telescoping individually optimized reactions into a single production workflow—and showcases the existing constraints to a simultaneous optimization of telescoped reaction sequences.

**4.3.2. Simultaneous Multi-Step Optimization.** Owing to the hardware challenges regarding purification, the first examples of telescoped reactor SDLs did not involve any purification steps, but performed the second step directly using the crude reaction mixture from the previous step. Whilst this enables the use of simpler hardware setups, it not only requires that both reaction steps are compatible with the same solvent, but also necessitates some chemical “cross-compatibility.” In other words, the first reaction step either needs to proceed in a clean fashion without producing major by-products, or the



**Figure 23.** Two-step synthesis of bupropion performed by Ley and co-workers. Initially, the two steps are optimized separately. The two optimized reactions are then combined with a work-up consisting of the addition of an aqueous sodium bisulfite solution and subsequent phase separation to yield one optimized reaction platform. Figure adapted with permission from reference Fitzpatrick *et al.*<sup>116</sup> Copyright 2018, John Wiley and Sons.

second reaction must be robust and selective enough that side products do not interfere with the desired reaction step.

While telescoped syntheses had been reported in the flow chemistry literature for some time, the first examples of autonomous optimization have been described by Bédard *et al.* in their report on the modular flow platform, as described above. In this work, the authors show the sequential combination of multiple reactor bays to a telescoped reactor system, with the addition of further reagent streams between two reactors. Using this setup, the automation of two two-step sequences is shown: a photoredox-catalyzed oxidative  $\alpha$ -functionalization of amines, and a Lewis-acid-catalyzed [2+2]-cycloaddition of phenylacetic acid chlorides with alkenes. In both cases, the first reaction step consists of the generation of a reactive intermediate (an iminium ion or a ketene, respectively), which is subsequently reacted with an appropriate reaction partner.

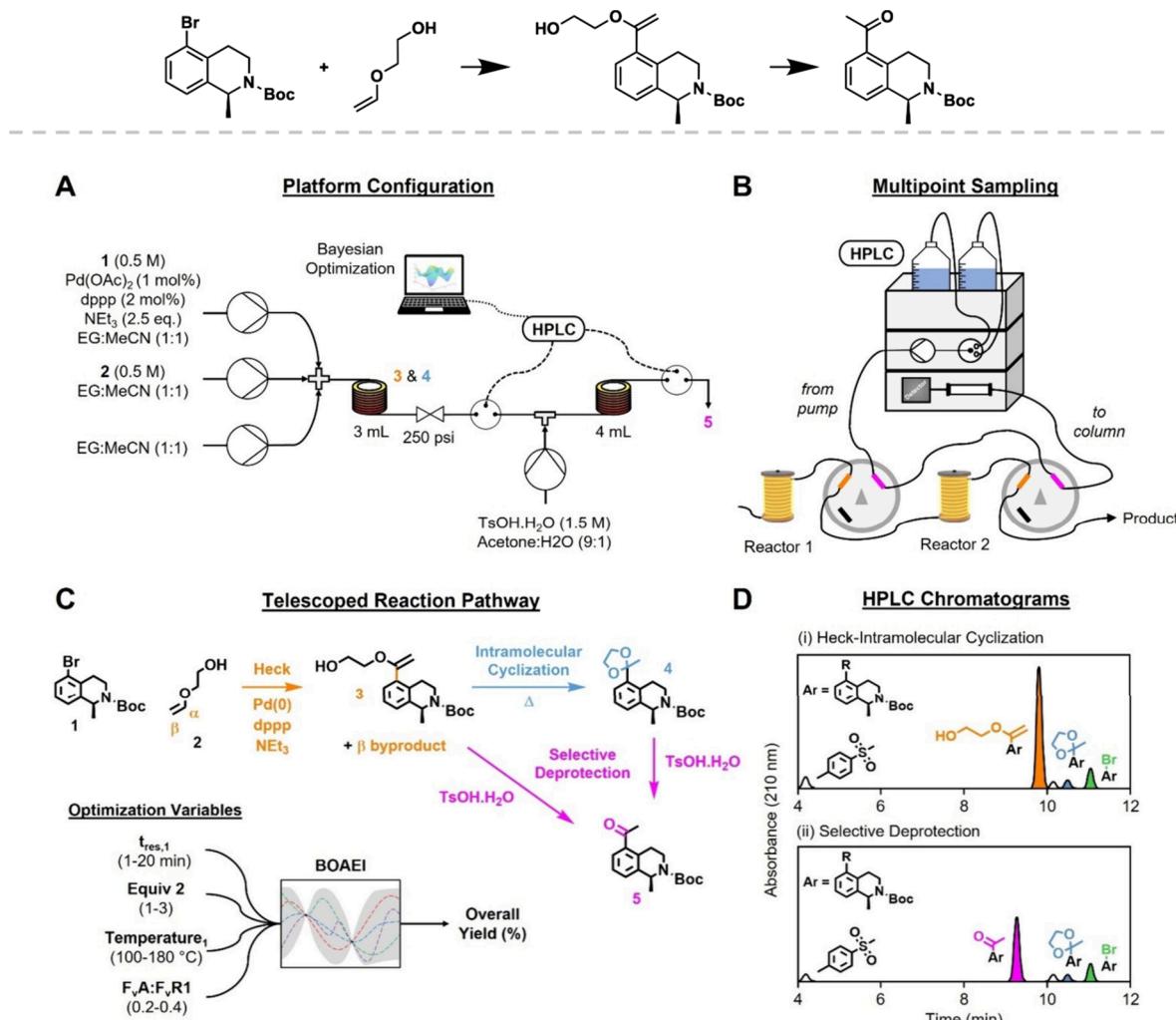
A further example of a telescoped reaction was shown by Ahn *et al.*,<sup>426</sup> where they conducted an ultrafast lithium-halogen exchange reaction directly followed by an addition-cyclization reaction of phenyl isocyanate. The authors designed an automated microreactor platform, which integrates a microreactor system with syringe pumps, solenoid valves, a thermostat and an in-line FT-IR spectrometer for real-time reaction monitoring. The authors use this platform to optimize the synthesis of a biologically active thioquinazolinone compound. The authors performed optimization campaigns over both only continuous (temperature, flow rate, reactor volume) as well as continuous and categorical (lithiating reagent) parameters. The BO algorithm employed by the authors achieved the same yields within 10 experiments that the authors previously found within 80 experiments of manual

planning. Lastly, the authors also optimized the conditions to synthesize a library of S-benzylidene thioquinazolinone derivatives.

A telescoped Heck coupling of a vinyl ether, followed by selective O-deprotection, was reported by Clayton *et al.* in 2022 (Figure 24).<sup>433</sup> The authors utilized a flow system combined with HPLC for quantifying the reaction yield, and a GP-based BO algorithm for iterative experiment planning. Notably, in order to obtain insights into their reaction, HPLC multi-point sampling, inspired by daisy-chaining from electrical engineering, allowed the sampling and investigating reactor outputs from both reactors separately. With this, the authors were able to uncover an alternative (but preferred) reaction mechanism for the deprotection step, which deviated from the initial working hypothesis, and turned out to be crucial for the identified deprotection conditions. This was only possible since the reaction was optimized as a telescoped process; if all of the three originally assumed steps had been optimized individually, a suboptimal process would have been found.

A highly complex example of multi-step synthesis was reported by Nambiar *et al.* in 2022 for the synthesis of Sonidegib.<sup>430</sup> The authors started out with the planning of the synthesis by a Computer-assisted Synthesis Planning (CASP) algorithm, which proposed a two-step route, consisting of an  $S_NAr$  reaction and an amide coupling reaction. Due to unfavorable electronics in the  $S_NAr$  step, the authors opted to synthesize the product via a three-step route, consisting of an  $S_NAr$  reaction, a nitro reduction and an amide coupling (Figure 25). The reactions were carried out in a robotically reconfigurable continuous-flow synthesis platform that allowed for the exchange of different modules by a robotic arm. As analytical modules, FT-IR and LC-MS were integrated to allow for monitoring reactor outputs. In their optimization, the authors considered a series of continuous parameters, including reaction times and stoichiometries, as well as multiple categorical parameters, such as the leaving group for the  $S_NAr$  reaction, the identity of the amide coupling reagent, or the reactor size for the last reaction step. Their modular platform allowed the robot to exchange the reactor, which in turn enabled the researcher to alleviate constraints on the interdependencies on residence times due to the flow rate of earlier steps.

The computer-proposed and human-refined synthesis pathway was subsequently attempted to be optimized in one telescoped reaction. In preliminary experiments, the LC-MS module after the first reaction showed that the  $S_NAr$  reaction proceeded with > 80% yield, however the FT-IR module after the nitro reduction revealed catalyst deactivation. Further experiments showed that this deactivation was caused by a by-product of the  $S_NAr$  reaction, rendering a fully telescoped process without thorough intermediate purification impossible. Thus, the authors decided to run the first reaction separately, and subsequently perform a telescoped reaction for the last two stages. As a consequence, the  $S_NAr$  reaction was run as a multi-objective optimization campaign, optimizing the yield, productivity and cost with respect to temperature, residence time, stoichiometry of reagent and base, as well as the leaving group as a categorical parameter. Optimal conditions were found in thirty experiments over 10 hours, with the algorithm providing multiple Pareto optimal points. The offline-purified product was subsequently used as a starting material for the telescoped reaction towards Sonidegib, optimizing yield and productivity simultaneously. Optimal conditions were found



**Figure 24.** Demonstration of the telescoped synthesis of a Heck-coupling followed by a selective O-deprotection, as reported by Bourne and co-workers. (A) A schematic overview of the used reaction platform. In the initial reactor, the Heck coupling is performed. TsOH is added to the output of the first reactor to perform the selective O-deprotection. The output of both reactors is analyzed via one HPLC device connected via multipoint sampling (B). (C) The reaction pathway of the telescoped reaction. Analysis of the optimization data shows a different dominant pathway ( $3 \rightarrow 5$ ) compared to the *pre hoc* assumed one ( $3 \rightarrow 4 \rightarrow 5$ ), underlining the utility of the optimization of a telescoped reaction. (D) Selected demonstrations of the HPLC chromatograms of the output of the first (top) and second (bottom) reactor. Figure adapted with permission from reference Clayton *et al.*<sup>433</sup> Copyright 2022, John Wiley and Sons.

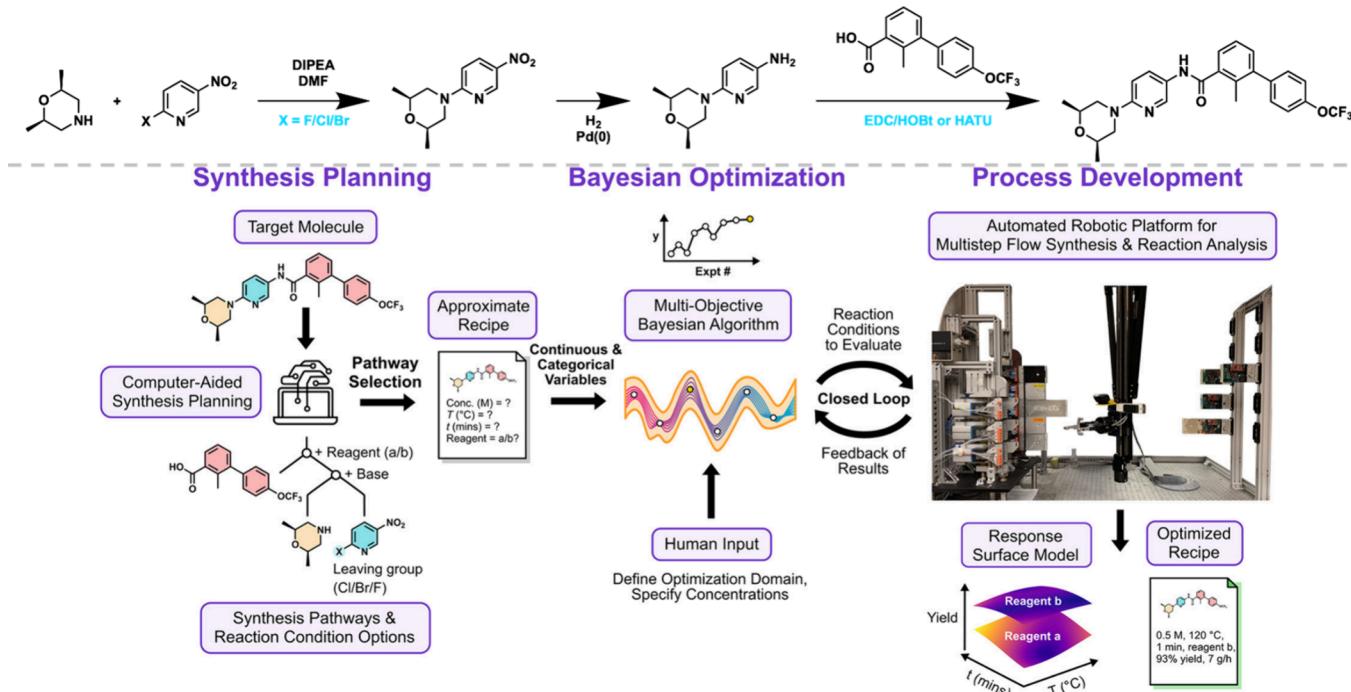
after fifteen experiments and 13 hours with a total yield of > 90%.

The above-mentioned SDL example reflects the challenges in fully autonomous, self-driving systems for organic synthesis particularly well. On the hardware side, telescoping multiple reaction steps offers a highly attractive solution to operating complex multi-step synthesis in a continuous fashion. However, generalizability of this strategy requires the development of advanced purification modules to minimize undesired cross-influences between individual reaction steps, e.g., to remove side products, or to enable solvent exchange. From an analytical standpoint, the introduction of automated reaction monitoring systems at multiple stages of the process provides access to important data that, in turn, can enable invaluable insights into the reaction progress and potential failure modes.<sup>451</sup> At the same time, the fully automated interpretation of this data, as well as downstream open-ended decision-making, usually require large degrees of expert knowledge, laboratory experience and adaptive decision-making (often referred to as “chemical intuition”). This applies to the

integration of automated algorithms for synthesis planning in particular, where, at the current stage, human decision-making is required for ranking routes or identifying reasonable condition search spaces. Integration of these advanced, and often open-ended, decision-making capabilities into AI systems represents an active challenge for the field and leaves room for future developments towards true, reliable SDLs for small-molecule synthesis.

#### 4.4. Further Solution-Phase Reactions

The concepts discussed above can readily be translated to synthetic chemistry domains beyond traditional small-molecule synthesis. Importantly, many polymers—with numerous applications in plastics, fibers, electronics, or drug delivery (materials-focused SDLs are discussed in later sections)—are synthesized in solution-phase processes, which makes these amenable to self-optimization. The major distinction to the previous discussions of small-molecule synthesis is the analytical methodology. Whereas for small molecules, a single, well-defined molecular entity needs to be determined in a



**Figure 25.** Multi-step self-optimization of the synthesis of Sonidegib as proposed by Jensen and co-workers. The synthesis was performed over three steps: an S<sub>N</sub>Ar reaction with a morpholine derivative as a nucleophile followed by a hydrogenation and an amide coupling. The latter two steps were performed as one telescoped synthesis. From a computer proposed and human-refined synthetic route, an approximate recipe was generated and continuous and categorical parameters were optimized in an automated robotic platform using a multi-objective BO algorithm (middle panel). Figure adapted with permission from reference Nambiar *et al.*<sup>430</sup> Copyright 2022, American Chemical Society.

quantitative fashion, the quantification of a “polymer yield” is less straightforward; in addition to the amount of formed polymer, the targeted size distribution, degree of (co-)polymerization, or other physical properties need to be controlled, which leads to a greater variability in the analytical methods and the resulting optimization objective.

One of the earliest examples of polymer SDL was performed in 2002. Vieira *et al.* demonstrated the closed-loop optimization of molecular weight and composition for copolymer latex.<sup>452</sup> Using a series of pumps and agitators, the authors automated emulsion polymerization, in which the monomers are dispersed as tiny droplets in aqueous phase, with emulsifiers and stabilizers to initiate and terminate polymerization, respectively. The copolymers were characterized by a near infrared spectroscopy (NIRS) probe of the solution, detecting the monomer concentration, polymer holdup, and the mean polymer size through the use of the partial least squares (PLS) model.<sup>453</sup> The goal was then to minimize the fitness, a weighted sum of differences between the desired and current molecular weights, over the feed rates of precursors, which was done using the iterative dynamic programming (IDP) method.<sup>454,455</sup> IDP considers discrete time intervals of previous iterations, and adjusts the flow rates to drive the system toward the desired synthesized polymers.

Houben *et al.* performed similar experiments with the use of multi-objective ML techniques to optimize the recipes of emulsion copolymerization reactions.<sup>456</sup> The authors used a setup similar to the one described before, however the analysis of particle sizes and conversion rates were done off-line, using dynamic light scattering and chromatography, respectively. Rather than only considering the flow rates, the 12 other experimental parameters were also varied. After each iteration of experimentation, the results were fed into the multi-

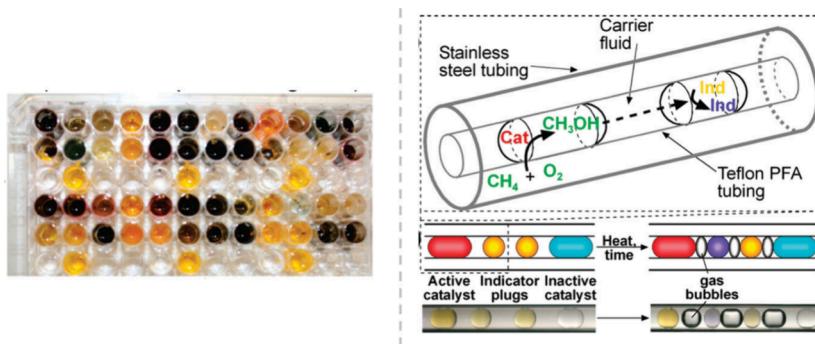
objective active learner (MOAL) algorithm, with suggestions produced by a GA, and predictions generated from a GP model.<sup>457</sup> Starting with 5 random initial experiments, and 15 additional experiments guided by MOAL, the authors found the conditions needed to produce high conversion polymers with particle sizes of 10 nm.

Rubens *et al.* used continuous flow microreactors, rather than batch reactors, to develop an SDL capable of high-throughput synthesis of reversible addition fragmentation chain transfer (RAFT)<sup>458</sup> polymers with precise molecular weights.<sup>459</sup> The polymer from the flow reactors were then characterized *in situ* by size exclusion chromatography (SEC), measuring the molecular weight, and dispersity of the polymers. The results were then fitted using a linear regression model with the results at each iteration. The flow rates with the best predicted results were then used for the next iterations.

Most recently, Knox *et al.* studied the same RAFT polymers with an SDL guided by BO, with the temperature and residence time as continuous optimization parameters.<sup>460</sup> Furthermore, the automated characterization techniques included both an in-line chromatography and an online NMR spectrometer. Using TS-EMO with a GP regressor surrogate, the authors were able to map out the Pareto-front for the polymer conversion and molar mass dispersion with higher resolution when compared to DoE. The BO iterations would suggest the next experimental parameters: the temperature and the residence time of the reactor.

#### 4.5. Catalyst and Reaction Discovery

**4.5.1. New Catalyst Materials.** Beyond the optimization of reaction conditions for a specific synthesis process, the discovery of novel highly active catalysts can allow for novel and more efficient synthetic processes, and can open up new



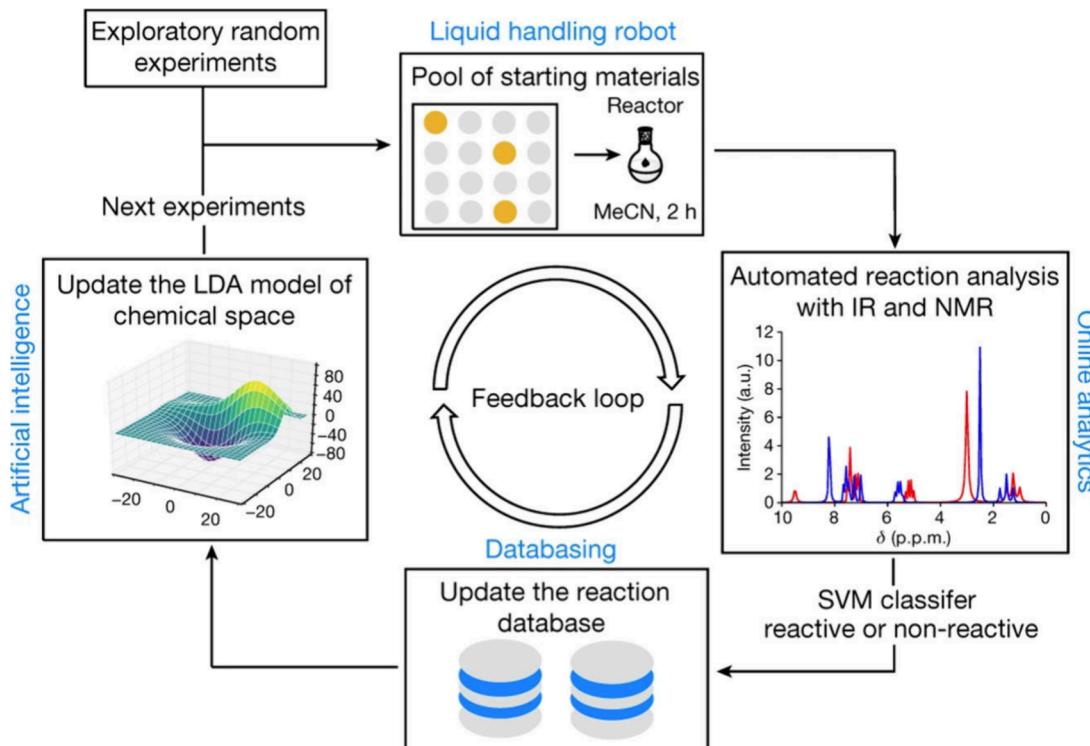
**Figure 26.** Experimental setup of the microfluidic platform for methane oxidation catalyst discovery. Left: Photograph of the 96-well plate for pre-mixing the catalyst components. Right: Schematic depiction of the microdroplet reactor containing a reaction droplet (left), and an indicator droplet (right). Figure adapted with permission from Kreutz *et al.*<sup>465</sup> Copyright 2010, American Chemical Society.

production avenues. While a catalyst is formally defined as a species that accelerates a given reaction, in reality, catalysis *enables* reactions that would otherwise only occur under impossible conditions. As such, catalysis has an enormous economic value, and it is assumed that >80% of all synthetic consumer products have gone through at least one catalytic process in their production. At the same time, discovering new catalysts is a considerable challenge, since their design requires the knowledge of a series of reaction pathways and modes of action, which also makes it extremely difficult to simulate catalytic efficiency from first principles. As a result, the last century has mainly seen empirically or heuristically driven campaigns for catalyst discovery. One of the most prominent examples is Mittasch's large-scale screening for heterogeneous catalysts for the Haber-Bosch process,<sup>21</sup> where they empirically test over 4000 possible catalysts—yielding an optimal catalyst that is still used as of today in almost unaltered form. More recently, Lai *et al.* demonstrated a LLM capable of suggesting catalyst synthesis conditions, drawing from the decades of results in the scientific literature.<sup>461</sup>

Major challenges in automating such a discovery process, and implementing it into an SDL, stem from the requirement to first synthesize and purify the catalyst candidate, which can involve a series of intricate experimental steps, and subsequently evaluate its activity in the catalytic reaction of interest. This challenge is illustrated in a pioneering example from Corma *et al.*, who tackled the challenge of identifying heterogeneous titanium silicate catalysts for olefin epoxidation.<sup>462</sup> Here, the catalyst synthesis alone involves gel formation from all involved reagents, followed by hydrothermal crystallization and post-synthesis treatment. The authors use a sophisticated robotic setup to automate these steps. The efficacy of the newly synthesized catalyst in the epoxidation of cyclohexene with a peroxide oxidant is then evaluated in a parallel batch reactor, which is coupled to ultrafast GC for on-line analysis. Even though the authors demonstrate an advanced level of automation (especially given that the work was published in 2005), transfer of samples between the workstations required a human experimentalist. Experiment planning is performed through a GA, enhanced by a neural network for applying a selection pressure on the newly proposed candidate generation, to optimize the quantities of four catalyst ingredients. The authors demonstrate three generations of 21 experiments each, and show the discovery of two new families of catalysts with improved activity, which are structurally characterized in detail.

In 2010, Kreutz *et al.* reported an SDL for homogeneous catalyst discovery for the partial oxidation of methane with molecular oxygen.<sup>465</sup> The catalytic system, which can be prepared by mixing all ingredients in an aqueous solution, is composed of three components: the active metal, a co-catalyst, and a ligand. These three categorical variables are optimized through a GA. To perform the required experiments, the authors have developed a sophisticated experimental setup based on droplet-flow reactors (Figure 26). Solutions containing the different catalyst compositions are prepared in 96-well plates, and are then injected into a microfluidic reactor. Both methane and oxygen are added by diffusion through the teflon walls of the flow reactor. The formed methanol was quantified by diffusion into neighboring microdroplets that contained a methanol-selective indicator, thereby allowing for semiquantitative analysis using UV-Vis spectroscopy. Per generation, the authors performed 48 experiments (in quadruplicate), and demonstrated that over 8 generations, a significant improvement in methanol formation (up to 3-fold increased catalytic activity) can be obtained.

Zhu *et al.* reported an autonomous system for discovering catalysts for the electrochemical oxygen evolution reaction (OER) from Martian meteorites, simulating the development of an oxygen-generating system on Mars.<sup>466</sup> For this purpose, the authors demonstrate a complex synergistic workflow consisting of multiple experimental and computational components. By analyzing the available Martian ores through automated atomic emission spectroscopy, the available elements—and therefore, the accessible materials search space—are defined autonomously by the platform. Within this search space (>3 million combinations of 6 metals in discrete quantity steps), a diverse set of ~30,000 possible catalyst compositions are first screened computationally, by molecular dynamics and DFT calculations. This data is used to train a surrogate neural network model for the computed catalytic properties as a function of the elemental composition. These computed properties, together with the elemental composition, are then used as inputs to a second neural network for predicting the experimental catalytic activity. The latter network was trained on a small seed dataset of < 300 experiments, which were conducted in a fully automated fashion using a robotic arm operating multiple workstations for dissolving the raw ores, creating reagent stock solutions, precipitating, drying and formulating the catalysts, and determining their catalytic activity in an electrochemical measurement. The authors then performed virtual BO within the entire search space, using the predictions of the trained



**Figure 27.** Schematic workflow of the reaction discovery SDL developed by Cronin and co-workers. Figure reproduced with permission from Granda *et al.*<sup>474</sup> Copyright 2020, Springer Nature.

neural network as their objective, and validated that the identified best candidate indeed outperforms all previously obtained catalysts. Even though the authors do not demonstrate multiple iterations of closed-loop of experiments and data-driven decision-making, the computational definition of the search space, as well as the advanced automation workflows are remarkable—making this work a Level 3 SDL, as by the definition of **Figure 1**.

In 2024, Ramirez *et al.* demonstrated the optimization of a heterogeneous catalyst for the reduction of CO<sub>2</sub> using BO.<sup>467</sup> As a catalyst, the authors explored systems containing up to three metals among iron, cobalt, copper, zinc, iridium and cerium with a maximum loading of 5 wt%. Additionally, the algorithm could choose between the presence or absence of potassium as a promoter, the amount of water as solvent as well as having silica, alumina, titania or zirconia as support. The authors synthesized 144 catalysts over six generations. Over the performed experiments, the BO algorithm was able to identify a catalytic system that maximizes the CO<sub>2</sub> conversion and MeOH selectivity while minimizing the CH<sub>4</sub> selectivity and the cost, where the latter was only considered throughout five generations to demonstrate the adaptability of the algorithm. Even though the algorithm is capable of finding a performant catalyst, the authors point out that this is performed within a well-studied and expert-restricted chemical space, demonstrating the hurdles for autonomous novel catalyst discovery.

The discussed works showcase examples of how catalyst discovery could be addressed in a closed-loop fashion—provided that the search space is sufficiently narrow, and the experiments can be automated in a useful manner. Particularly catalyst synthesis poses a major challenge in this regard; the diversity of catalyst space, and the fine nuances that can influence catalytic activity, however, render the development of

generalizable automation schemes difficult. In homogeneous catalysis, *making* new catalysts requires synthesizing new molecular species, which usually require multi-step reaction and purification sequences, which, in turn, we had previously identified as a major challenge for automation. On a purely computational level, this bottleneck can be circumvented, which has led to impressive and experimentally validated examples of closed-loop catalyst design, for example in organocatalysis.<sup>468</sup> In heterogeneous catalysis, on the other hand, synthesis requires intricate thermal treatment and annealing steps, which possess inherent automation constraints, and can often lead to structurally ill-defined materials, adding further complexity to the data-driven prediction problem. As a consequence, the last decade has produced rare examples of true SDLs for catalyst discovery, which remains a grand challenge for autonomous discovery, both from the software and the hardware standpoint.

**4.5.2. New Reactions and Reaction Types.** While all previous discussions have focused on specific reactions—the product (and reactants) are given, and the goal is to find the catalysts, reactants, reagents or reaction conditions that maximize the product quantity—SDLs can also be used to search for new reactions or products.<sup>469</sup> In fact, this problem of discovering new reactivity or catalytic activity has been an active field of research in organic chemistry for more than a century. While the predominant search strategy in this field has been rational design, the importance of “serendipitous” discoveries has been emphasized numerous times.<sup>470</sup> As an example from an SDL, Amara *et al.* reported the detection of an unexpected side product when attempting the self-optimization of a γ-Al<sub>2</sub>O<sub>3</sub>-catalyzed methylation in supercritical CO<sub>2</sub> (for a more detailed discussion of these reactions and the self-optimization algorithms used, see the section on **Self-optimizing flow reactors**).<sup>471</sup> Careful characterization of the

side product by a human researcher allowed for its unambiguous identification, and a second closed-loop campaign towards the yield of this side product was carried out, which eventually resulted in the discovery of optimized conditions for a new reaction type. This example demonstrates the possibility of discovering new reactions through SDLs. At the same time, especially from the standpoint of automation and experiment planning, it poses the open-ended analytical challenge of detecting and identifying newly formed, unknown reaction products from a crude reaction mixture, which is often addressed by analyzing changes in the bulk properties of the reaction mixture (UV-Vis spectra, IR spectra, NMR spectra), or through coupled separation–detection techniques (GC- or HPLC-MS).

Early examples in the field of “untargeted” reaction discovery have focused on non-iterative screening campaigns using combinatorial chemistry and HTE, which have been reviewed elsewhere.<sup>472,473</sup> The first example of a truly closed-loop campaign for discovering new reactivity was reported by Cronin and co-workers in 2018, who developed an SDL for finding new two- or three-component reactions in a pool of reactants (Figure 27).<sup>474</sup> In a proof-of-concept work, Granda *et al.* selected a set of 18 reactants with diverse functional groups, which can be reacted under fixed reaction conditions in a fully automated fashion. Crude reaction mixtures were analyzed by automated IR and <sup>1</sup>H-NMR spectroscopy, and the spectra, along with the spectra of the starting materials, were processed by a pre-trained SVM classifier to label the reaction as “reactive” or “non-reactive”. Based on this data, a linear discriminant analysis (LDA) model was trained to predict reactivity across the entire search space, and new experiments were selected in a fully exploitative fashion. With this search strategy, the authors demonstrate a significantly improved hit rate compared to trivial random search algorithms, and report a series of nontrivial reactions which had not been published before.

Later work from the same group, Caramelli *et al.* used a similar platform to discover new unreported reactions in an automated fashion:<sup>475</sup> the photochemical reaction of phenyl hydrazine and bromoacetonitrile, and the reaction of p-toluenesulfonylmethyl isocyanide (TosMIC) and diethyl bromomalonate. For decision-making, *Reactify* is a CNN that is trained on the NMR spectral data of 440 reactions with reactivity classified by a chemist. A neural network, using the junction-tree VAE embedding of the molecules as features, is trained to then suggest new reactants for the SDL platform. Both the *Reactify* and the surrogate neural networks were retrained at each iteration. The reaction mechanisms of the novel reactions were further studied by the authors. In related work, Mehr *et al.* demonstrated a probabilistic approach to reaction discovery, both *in silico* and as part of an SDL.<sup>476</sup> Reactants were assigned prior distributions which were then combined to form a joint probability prediction of the reactivity between them. Following Bayes’ theorem, the distributions were updated based on the feedback results of an automated HTE platform. The experiments were carried out in a flow-based system, with on-line NMR, HPLC, and MS characterization. The authors were able to rediscover known reactions such as the Buchwald-Hartwig amination, and the Wittig-Horner reactions.

The question of whether an identified reaction can be considered “novel” has been subject of an ongoing debate. While the previously unknown formation of a reaction

product—the definition used by Granda *et al.* and Caramelli *et al.*—clearly constitutes a *new* reaction, the term *novelty* lacks an unambiguous definition. In both studies, the authors were maximizing the *reactivity*, or rather, maximizing the number of reactions classified as reactive. Any SDL targeting the discovery of *novel* reactions therefore requires a series of assumptions and simplifications for defining the optimization objective. Porwol *et al.* later applied a similar discovery strategy for finding new polyoxometalate clusters composed of metal ions and bridging ligands.<sup>477</sup> Following *in situ* assembly of the ligands in a three-component coupling, a metal precursor is added to form potential polyoxometalates. By using a series of characterization techniques including UV-Vis spectroscopy, MS, and pH measurements, *novelty* is measured as the cumulative difference between the data of starting materials and products, respectively. As independent variables, the authors selected the ligand precursor identities, metal ion, reagent volumes, reaction temperature and reaction time. To maximize the novelty, the authors used a custom surrogate-free search algorithm, which samples each experiment in a given distance from the previous experiment, depending on the novelty of the previous experiment. Following this strategy, the authors discovered a range of new polyoxometalate clusters. This is discussed further in the section on state materials synthesis optimization.

#### 4.6. Determination of Reaction Kinetics

Especially on the process chemistry level, knowledge about the reaction conditions that lead to optimized reaction yields is not sufficient for safe and reliable reactor operation. In these contexts, detailed information about the kinetics of a reaction is required in order to predict and adjust the behavior of a reactor system. At the same time, kinetic knowledge can enable important insights into the mechanism of a reaction—which is of high relevance for informed decision-making, both at the discovery and at the process stage. SDLs can, and have been, used to iteratively acquire kinetic data, refine kinetic hypotheses, and eventually obtain reliable kinetic models. This has, in a simple proof-of-concept study, already been demonstrated in the late 1970s in the context of derivatization reactions for analytical chemistry.<sup>478</sup>

Decades later, in 2011, McMullen and Jensen utilized a microfluidic system to optimize the parameters of a kinetics model for the Diels-Alder reaction of isoprene with maleic anhydride.<sup>479</sup> Following the Box and Hill method, the probability of a particular rate model describing an experiment can be formulated in a Bayesian context by a posterior probability function based on the experimental conditions and the outcome concentration.<sup>480</sup> Using an in-line HPLC, the microfluidic system returns the output concentration of isoprene, which is used to update the distribution until a predefined probability threshold is met. After deciding the rate law, the microreactor was then used to optimize the parameters of the rate constant through plug-flow reactor kinetics. Finally, for validation, the authors performed 4 additional experiments and found good agreement with predictions from the optimized rate law.

In a related study, Reizman and Jensen presented a continuous-flow SDL for studying multi-step reaction kinetics.<sup>481</sup> Using high-throughput synthesis methods enabled by flow reactors, the authors studied the conversion of 2,4-dichloropyrimidine to 4,4'-(2,4-pyrimidinediy)bis-morpholine. There are two reaction pathways, each with two reactions,

which are all modeled as second-order bi-molecular reactions. The product concentrations were measured after the reaction by online HPLC, and the kinetic model parameters were least-squares fit to the results. The sensitivity coefficients, a measure of how sensitive the predicted concentrations are to the synthesis parameters, were then calculated for the optimal parameters. By minimizing the sensitivity coefficient, the next experimental conditions were generated, and the reaction kinetic models' parameters were iteratively optimized.

Most recently, Sheng *et al.* applied a closed-loop SDL to study the electrochemical reaction of cobalt tetraphenylporphyrin (CoTPP) with organohalides.<sup>482</sup> The electrochemical platform uses a flow system to control the flow of reactants into a 3-electrode cell, which is monitored by a potentiostat for cyclic voltammetry (CV). The platform first identified reactions which can be modeled by the EC mechanism, which consists of an electron transfer step followed by a solution reaction. This was done by analyzing the CV data with a ResNet CNN previously trained to extract relevant electrochemical quantities.<sup>483</sup> In the second stage, the EC mechanism is probed by optimizing the rate constant ( $k_0$ ) of the solution reaction step as a function of the voltammetric scan rate, and the organohalide concentration. Both stages were guided by a Bayesian optimizer from Dragonfly.

#### 4.7. Solid State Materials Synthesis

Solid state materials, such as molecular crystals, zeolites, metal-organic frameworks (MOFs), covalent organic frameworks (COFs), polyoxometalates and alloys, have a variety of applications, particularly in catalysis of reactions. Porous materials like molecular crystals, MOFs, COFs, and zeolites are characterized by voids in the crystalline structure, typically on the nanometer to micrometer scale, and high surface areas, giving the material the ability to adsorb molecules for storage or catalysis. This has applications in gas storage and separation (i.e., methane, hydrogen gas), filtration, and drug delivery.<sup>484,485</sup> While there are many SDLs focused on optimizing the function of solid state materials, the SDLs discussed in this section are focused on finding optimal synthesis conditions for the structure and crystallinity of the material. For SDLs related to energy storage and optoelectronic applications, we refer the reader to the respective sections.

The primary advantage of solid state materials is their tunability beyond the chemical component; by varying both the composition and synthesis parameters, the material properties and structure can be tuned for specific applications. Considering the space of possible materials and structures is intractably large, traditional approaches based on manual synthesis are insufficient to explore these materials efficiently. A number of high-throughput methods, both computational,<sup>373,486–489</sup> and experimental,<sup>490–492</sup> have been developed and successfully applied to combinatorially exploring the material space. These have resulted in large datasets of possible structures and materials, as well as their measured or predicted properties, paving the way for data-driven strategies.<sup>493,494</sup>

Conventional data-driven applications of these high-throughput methods are through compound screening: a predefined space of compounds and structures are filtered down based on predictions from statistical models trained on the datasets, or theoretical calculations.<sup>495,496</sup> There are extensive works in the literature completely within the computational domain: developing descriptors and models that can predict the properties of interest from the data-

sets,<sup>486,497–499</sup> and extending these models for computational SDLs, performing active learning campaign based on *in silico* model predictions from models trained on high-throughput experimental or computational results.<sup>500,501</sup>

In the synthesis of zeolites, Moliner *et al.* utilized a high-throughput robotic arm platform capable of liquid/solid handling, stirring, and crystallization to generate a combinatorial DoE study of 144 triethylamine:SiO<sub>2</sub>:Na<sub>2</sub>O:Al<sub>2</sub>O<sub>3</sub>:H<sub>2</sub>O zeolites.<sup>491</sup> Using the MLP model, the authors were able to attain better predictions of the crystallinity of the zeolites from the experimental dataset than the typical multivariable quadratic models. The crystallinity is measured via XRD: the spectral peaks are fitted with Gaussian functions, and the average full-width half-maximum (FWHM) of the peaks are used as a measure of crystallinity. In a related study, Corma *et al.* performed a similar study for SiO<sub>2</sub>:GeO<sub>2</sub>:Al<sub>2</sub>O<sub>3</sub>:F<sup>-</sup>:H<sub>2</sub>O:4-(2-methane sulfonylphenyl)-1,2,3,6-tetrahydropyridine hydrochloride zeolites, which have demonstrated successful crystallization into ITQ-21 and ITQ-31 zeolites.<sup>502</sup> The authors improved the crystallinity predictions by including structural descriptors derived from the XRD spectra, along with the synthesis descriptors, in the MLP neural network input.

Nikolaev *et al.* demonstrated an Autonomous Research System (ARES)<sup>503</sup> capable of autonomously conducting iterative materials experiments to study carbon nanotube (CNT) synthesis—a pioneering example of an autonomous SDL for materials research. Experiments were conducted by heating catalyst-coated silicon pillars, which each serve as CNT microreactors, with a laser while varying growth parameters like temperature, pressure, and gas composition. Raman spectroscopy measured the CNT growth rate in real-time. Using linear regression models, the authors were able to map out the effect of experimental conditions on the resulting growth of single-wall or multi-wall CNTs.<sup>504</sup> In a later study, the same system was providing feedback to a RF model and genetic algorithm to propose new experimental conditions. Over hundreds of closed-loop iterations with minimal human intervention, ARES successfully learned to grow CNTs at targeted growth rates by optimizing the multi-dimensional parameter space. More recent work from the group modified ARES to use BO with GP surrogates for the maximization of CNT growth rate.<sup>505</sup> These demonstrations showcase ARES's ability to autonomously navigate complex experimental domains and obtain insights into growth kinetics, which is valuable for controlled nanotube synthesis. As one of the first implementations of SDL for materials science, this work highlights the potential of autonomous research systems to accelerate the scientific understanding and development of complex functional materials.

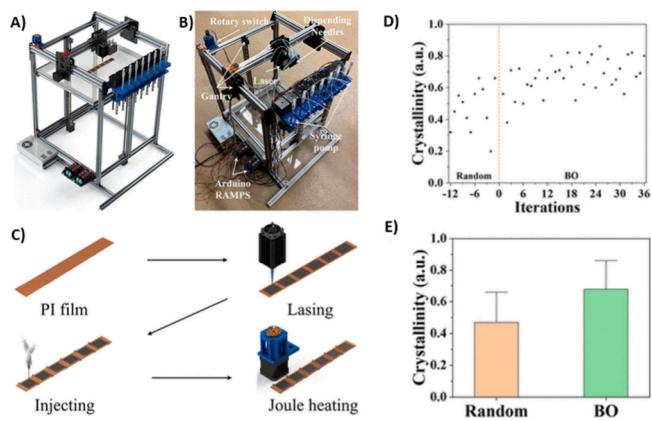
Similar ML directed discovery have been demonstrated in the synthesis of MOFs, for example Raccuglia *et al.* further incorporated reactant and reaction descriptors in the prediction of successful synthesis and crystallization of organic templated vanadium selenite materials.<sup>506</sup> Training a SVM on experimental results from both failed and successful reactions, and comparing the recommended reactions from a human chemist, the model was shown to have a higher success rate and provide more diverse reactions. More recently, Xie *et al.* utilized XGBoost, a gradient boosting tree-based model, to determine the reaction parameters for crystallization of metal-organic nano-crystals.<sup>507</sup> To test their model, validation experiments not found in the training set were conducted to demonstrate the use of the XGBoost model for extrapolating to

new MOF nano-crystals. Luo *et al.* later developed the MOF Synthesis Prediction tool, using natural language processing DL models to extract synthesis conditions of MOFs from the literature and create a dataset and prediction tool for synthesizing new MOFs.<sup>508</sup> The experiments in these works were not conducted in an automated fashion. The earliest examples of closed-loop SDLs for solid state materials were by Corma *et al.* in 2005, previously discussed in greater detail in the section on catalyst discovery. The authors were interested in optimizing the catalysis of olefin using a Ti-based zeolite catalyst.<sup>462</sup> Various concentrations of hydroxide, titanium and surfactants were combined in the hydrothermal synthesis of the zeolite using a robotics system. The batches of zeolites were then tested for catalytic activity using ultrafast GC.

In the development of MOFs, Moosavi *et al.* developed an SDL that optimizes the crystallinity of the HKUST-1, first synthesized by Chiu *et al.*<sup>509</sup> at the Hong Kong University of Science and Technology.<sup>510</sup> The synthesis was performed using a high-throughput robotic platform, capable of handling and stirring reactants, transferring the samples into a microwave reactor cavity for synthesis and to a powder X-ray diffractometer for crystallinity measurement. The exploration of the parameter space was done using a GA dubbed the SyCoFinder, over the course of three generations, with 30 synthesis conditions tested in each. Similar to previous work,<sup>506</sup> results from successful and failed experiments were collected in order to train a RF model to identify the synthesis parameters of importance. By weighting the 9 dimensional parameter space by the identified importance, the parameter space becomes smaller and more confined, allowing for more efficient exploration guided by chemical intuition. Further optimizations were not performed.

Xie *et al.* performed a similar analysis with the zeolite imidazolate framework (ZIF), ZIF-67.<sup>511</sup> They developed a new ZIF synthesis protocol based on a custom low-cost gantry-style robot SDL platform that injects precursors onto laser-induced graphene microreactors fabricated on a thin film (Figure 28).<sup>512</sup> The microreactors were then Joule-heated to create ZIF-67 in a high-throughput manner. The synthesized samples were transferred for XRD characterization. Rather than using a GA in the experiment planning, the authors used BO with a RF surrogate model. For the synthesis, the molar ratio of metal ions to organic molecules, the volume of precursors, the applied DC voltage, and the heating duration was varied. After an initial 12 random samples, three additional generations with 12 samples each were suggested by the BO algorithm using the expected improvement acquisition function. Figure 28 shows the improvement in the crystallinity as a function of BO iterations.

Extending into thin films of MOFs, Pilz *et al.* developed an SDL optimizing surface anchored MOFs that are formed layer-by-layer.<sup>513</sup> Like previous work,<sup>510</sup> the authors used the SyCoFinder GA for synthesis planning, with the goal of optimizing multiple objectives: the crystallinity, the [111]-orientation of the crystal, and the phase purity, all of which are measured from the XRD spectra. The objectives are combined with a summation and then normalized to a fitness between 0 and 1. The parameter space included the metal and linker concentrations, the amount of water, and cleaning time via sonication and spray cleaning. The samples were transferred across the various modules via a 6-axis robotic arm. The SDL started with a diverse random set, and two more generations

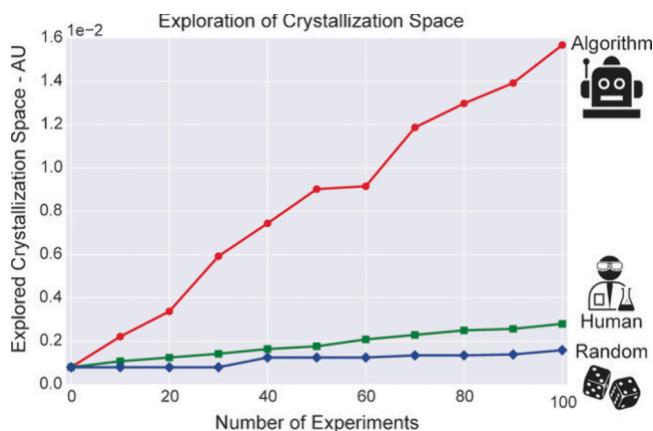


**Figure 28.** HTE platform used by Lin and co-workers, and the results of the closed-loop optimization. (A) Schematic and (B) picture of gantry-style SDL with multiple heads (C) to perform laser fabrication of microreactors, injection of precursors, and Joule heating synthesis of ZIFs. (D) Using a BO algorithm seeded with initial random samples of experimental parameters, the crystallinity is optimized with each iteration. (E) When compared to random sampling, BO achieves a higher crystallinity, measured by XRD. Figure adapted with permission from reference Xie *et al.*<sup>511</sup> Copyright 2021, American Chemical Society.

were carried out, with increasing fitness found with subsequent generations.

Harris *et al.* demonstrate an autonomous synthesis platform for pulsed laser deposition (PLD) of thin films by combining real-time diagnostics, automated synthesis and characterization, and ML algorithms. The platform utilizes GP regression and BO to autonomously explore a 4D parameter space of background pressure, substrate temperature, and laser fluences on two targets, tungsten and selenium, aiming to optimize the crystallinity of WSe<sub>2</sub> thin films based on *in situ* Raman spectroscopy feedback—sharper peaks indicate higher crystallinity. Having only sampled 0.25% of the parameter space, the autonomous workflow discovered two distinct growth windows and mapped the process-property relationships governing film quality. Notably, the automation achieved at least a 10-fold increase in throughput compared to traditional manual PLD workflows. The combination *in situ* Raman spectroscopy monitoring, and ML driven decision-making can be used for PLD fabrication of other solid state thin film systems.

Duros *et al.* studied the crystallization of a new polyoxometalate structure with an SDL driven by active learning, and also provided a comparison with random and human-guided experimental planning.<sup>514</sup> A series of syringe pumps fed aqueous precursor solutions into a reactor, and the products were visually inspected for crystallization. The platform was capable of performing batches of 10 crystallization experiments per day, and an initial dataset of 89 points was acquired to start as a training set. A SVM classifier was trained on this dataset to classify successful crystallization experiments. The subsequent experiments were then conducted using an active learning loop, with the goal of maximizing the number of polyoxometalate structures and the explored synthesis parameter space. When compared to human and random exploration of the space, the SDL explored more of crystallization space, while still finding a similar number of crystallization points as human decision (Figure 29).



**Figure 29.** Explored space of possible polyoxometalate crystals explored with each iteration. Duros *et al.* performed human-guided and random searches of the crystallization space as comparison for the algorithmic approach. The space is defined by the experimental parameters. Figure reproduced with permission from Duros *et al.*<sup>514</sup> Copyright 2017, John Wiley and Sons.

Beyond MOFs and COFs, van der Waals superlattices—i.e., stacks of graphene-like atomic monolayers bound through dispersion interactions—have emerged as an attractive class of 2D crystals with multiple applications in e.g., semi- and superconductors, or topological insulation. The layer-by-layer assembly of these materials could allow precise control over materials properties, but requires delicate physical handling. As an important step towards SDLs for van der Waals superlattices, Masubuchi *et al.* developed a multi-step robotic workflow: in the first step, pre-synthesized 2D crystals deposited on Si chips are automatically detected and characterized using optical microscopy and computer vision.<sup>515</sup> Subsequently, the detected crystals are robotically transferred to a stamping apparatus, aligned and assembled to the desired superlattice. While this work does employ iterative data-driven decision-making, the advanced automation and computer vision approaches can justify the classification as a Level 3 SDL, laying the foundation for autonomous materials discovery for van der Waals superlattices.

Kusne *et al.* developed CAMEO for the self-driven discovery of phase-change memory (PCM) materials.<sup>516</sup> These are inorganic materials capable of switching between amorphous and crystalline states, altering the optical and electrical properties of the material. CAMEO uses a physics-guided ML model for BO of Ge-Sb-Te ternary PCM. Synthesis was not part of the design process; rather, a combinatorial library of Ge-Sb-Te material was loaded onto the system, along with data from DFT simulations. Because the target property was dependent on the phase of the material, the first iterations maximize the phase map of the material. After some defined threshold for phase map exploration, the BO algorithm, based the predictions GP models with an UCB acquisition function modified with an additional term based on the distance from the phase boundary, selected the next material for automatic synchrotron XRD characterization and human-in-the-loop evaluation of the optical gap. While not fully automated, the authors were able to discover a new photonic PCM with an optical gap difference between crystalline and amorphous phases of  $0.76 \pm 0.03$  eV, over three times larger than the conventional GST225 material.

In the quest to understand the phases of specific solid state inorganic materials, Ament *et al.* demonstrated a self-driven high-throughput platform for determining the phase boundaries of Bi<sub>2</sub>O<sub>3</sub> system.<sup>517</sup> Bi was sputtered in an atmosphere of Ar and O<sub>2</sub> onto Si wafers to create thin-films of Bi<sub>2</sub>O<sub>3</sub>, which were annealed in stripes using a laser. By varying the annealing temperature and time, different phases of Bi<sub>2</sub>O<sub>3</sub> can be observed. The samples were characterized by optical microscopy and reflectance spectroscopy to determine the phase boundaries, and the next conditions are suggested by GP models with custom kernels based on the physics of the experiment; the algorithm was dubbed Scientific Autonomous Reasoning Agent (SARA). The authors were able to map the phase boundaries of the system two orders of magnitude faster than random or exhaustive search methods.

A major obstacle to the development of a fully automated SDL for solid state materials is the need for powder handling and XRD characterization. Lunt *et al.* developed the Powder-Bot, an autonomous robot capable system capable of automated Powder XRD.<sup>10</sup> Powder-Bot successfully synthesized molecular crystals using a Chemspeed liquid-handling platform. A single-arm mobile robotic manipulator transfers the crystalline material to a grinding station where a dual-arm stationary robot produces the powder, and then takes the powder XRD samples to a diffractometer for analysis, totaling thirteen distinct steps. The manipulator operates the diffractometer as a human chemist would, and the XRD spectra is recorded. While the work is not a true closed-loop SDL due to the lack of intelligent experimental design, the authors demonstrated a landmark single iteration of automated synthesis and powder XRD characterization using conventional processing and characterization equipment. In another notable advancement, Chen *et al.* present ASTRAL, a robotic platform that seamlessly integrates powder-precursor synthesis including powder dispensing, ball milling and oven-firing into XRD characterization of reaction products.<sup>518</sup>

Most recently, Szymanski *et al.* presented A-Lab,<sup>519</sup> an SDL for solid state synthesis of metal oxides and phosphate powders, with fully automated sample preparation, heating, and XRD characterization capabilities. Solid state synthesis pathways were selected using the ML-based precursor selecting algorithm ARROWS<sup>3</sup>, which incorporates decomposition energies from both *ab initio* calculations and previous experimental outcomes to find the best reaction pathways.<sup>520</sup> Air-stable synthesis targets were identified based on *ab initio* calculations from the Materials Project, and a dataset from Google DeepMind.<sup>521</sup> Recipes obtained from text-mining sources in the literature were used to train ML models to generate recipes for compounds not found in the training dataset. We note that this is an unguided systematic search of the proposed synthesis routes; however, if these recipes fail to produce high enough yields (> 50%), A-Lab defaults to the ARROWS<sup>3</sup> algorithm, which utilizes information from prior experimental results. The autonomous platform then carries out the recipe, performing dosing, syntheses, and analysis on three different stations, with a robotic arm transporting the sample between stations. The collected XRD spectra were analyzed using a probabilistic ML model trained on the ICSD, as discussed previously in *Analytical Process Optimization*. The resulting weight fractions of the synthesis products were fed back into the orchestrator of A-Lab to inform further experimentation. Over the course of 17 days of continuous experimentation and 355 experiments, A-Lab successfully

synthesized 41 out of 58 target compounds, of which 9 of the targets were optimized by the data-driven ARROWS<sup>3</sup> algorithm for improved yields. The authors further claim the discovery of multiple new compounds and structures, although this has been called into question due to the non-standard analysis of XRD results, and the under-characterization of the compounds.<sup>522,523</sup> Still, the A-Lab has demonstrated advancements in the development of inorganic solid state SDLs.

These examples represent significant steps towards accelerating the discovery of feasible solid-state materials in a design space that contains a large fraction of unstable and metastable materials, and closing the automation design loop for arguably the most difficult-to-automate piece. We expect these endeavors will set precedents for application-driven, inorganic solid-state SDLs to come.

#### 4.8. Outlook and Perspectives

Within this chapter, we have provided a comprehensive overview of SDLs for chemical reaction optimization, which has arguably been the most widespread application of SDLs as per definition of this review. While first, foundational examples of autonomous reaction optimization have been laid in the 1980s, the field has seen an enormous boost in the 21<sup>st</sup> century, owing to advances in digitization, computational resources and software distribution. The largest body of work has focused on the autonomous optimization of single-step reactions in solution. Notable examples include: heterogeneous catalysis, photochemical reactions and photocatalysis, nanoparticle catalysis, the use of supercritical fluids as reaction solvents, and many others. These works have also led to notable automation and advances in related disciplines of modern synthesis, including catalytic technologies like electrocatalysis<sup>524,525</sup> and organocatalysis,<sup>526</sup> or economically important applications like biomass or waste valorization.<sup>527</sup> We expect to see pioneering examples of SDLs in these fields in the years to come, leading to a further diversification of SDLs for chemical reaction optimization. Importantly, optimization campaigns have not been limited to maximize the yield of a chemical reaction, but have been extended to economic considerations (e.g., time, cost, and produced waste), kinetic information, or the information content of the obtained reaction data.<sup>528</sup>

It is important to note that all of these works have relied on two main pillars. First, the availability of open-source solutions for both automated reaction hardware and optimization software has enabled the implementation of autonomous systems across a variety of labs, and has proven to be a (figurative) catalyst for the spread of SDLs. We highly advocate for such open-source initiatives—accessible solutions (such as EDBO+ platform from the Doyle group<sup>295</sup>) have shown to serve as inspiration for further groups to adopt important SDL technologies.<sup>529</sup> Secondly, domain expertise and laboratory experience has been instrumental to set up the required hardware and, more importantly, define and constrain the experimental search problem.

The use of AI for those open-ended decision-making tasks represents an important open challenge to the community, in addition to adaptive decision-making in synthetic laboratory scenarios. These software requirements go hand in hand with the development of flexible, reconfigurable hardware systems that enable such adaptive operations. Addressing these challenges, as discussed in detail throughout this chapter of the review, can build the foundation for the next generation of

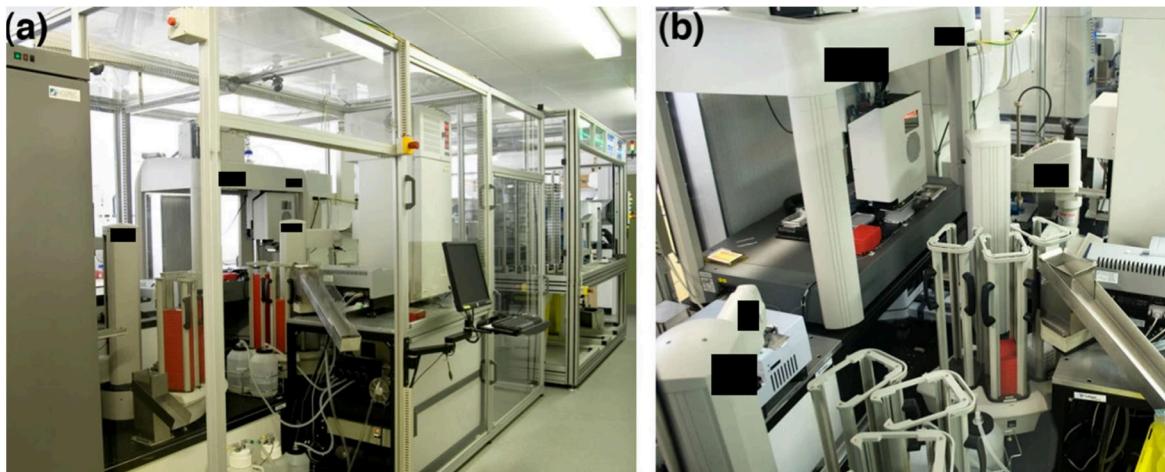
SDLs for chemical synthesis, and eventually bring us one step closer to the dream of autonomously synthesizing any molecule (or material) on-demand. As such, autonomous synthesis can be an integral component of any autonomous materials discovery initiative, including the efforts detailed in the following sections.

### 5. DRUG DISCOVERY AND BIOCHEMISTRY

Drug discovery plays a pivotal role in modern society and in the chemical industry, not only as a major consumer of chemical compounds, but also as a driving force behind chemical innovations: indeed, the pharmaceutical industry invests billions in research and development (R&D) every year,<sup>530</sup> and some of the first examples of automated and high-throughput experiments were first developed by pharmaceutical companies. The reason behind this huge investment is the high cost associated with drug development: it usually takes US\$2.6 billion and 10 years to put a single drug on the market.<sup>531</sup> This long and costly pipeline can be roughly split into five main stages: early-stage discovery, preclinical studies, clinical trials, FDA review and approval and finally, post-market monitoring. Early-stage discovery includes disease-related proteins target identification, compound screening against selected target, assay development and compound property optimization. Preclinical studies focus on drug profiling, delivery and dose range finding.

However, while the R&D budget increases over the years, the composite average approval rate of drugs keeps falling down.<sup>530</sup> Analyses of clinical trial data from 2010 to 2017 show four possible reasons attributed to 90% of the clinical failures of drug development: (i) lack of clinical efficacy (40%–50%), (ii) unmanageable toxicity (30%), (iii) poor drug-like properties (10%–15%), and (iv) lack of commercial needs and poor strategic planning (10%).<sup>532</sup> Given those statistics, it is apparent that success in early-stage discovery and preclinical studies stages is key to overcoming the high attrition rate. In those stages, researchers are confronted with multi-objective optimization problems that span the chemical and biological space. Not only are those vast, but the understanding of them is also incomplete. For efficient exploration, the pharmaceutical industry has thought to employ automation relatively early compared to other industries:<sup>533</sup> Automation in drug discovery dates back to the 1980s with the advent of high-throughput screening platforms, which leverage robotics to manage the handling of thousands of bioassays.<sup>534</sup> Spurred by large investments by pharmaceutical companies, robotic drug discovery platforms have evolved towards a higher level of automation and complexity. A notable example is Eli Lilly's state-of-the-art automated synthesis laboratory, among others.<sup>535</sup>

Along with hardware automation, the field has benefited significantly from advances in computational molecular design and synthesis planning, which have proven to be powerful tools for accelerating drug discovery.<sup>536,537</sup> Notably, while the idea of applying ML methods to drug discovery dates back to the 1990s, the recent achievements of DL methods sparked a high interest in the field for AI-driven early-stage drug discovery.<sup>538,539</sup> Indeed, exploiting the capacity of DL to leverage vast amounts of data to create efficient biochemical representations could transform how early-stage research is conducted. For example, Stokes *et al.* used a DL GNN to identify new antibiotic compounds, and were able to successfully demonstrate the repurposing of halicin, originally



**Figure 30.** (a) Photo of the external of Adam, with Eve on the far right. (b) Photo of Adam's robotic platform inside the system. Figures adapted with permission from Sparks *et al.*<sup>549</sup> Copyright 2010, Springer Nature.

used in the treatment of diabetes, as a lead compound for inhibiting *E. coli* bacterial growth.<sup>540</sup> Additionally, the release of AlphaFold<sup>541</sup> has revolutionized the approach to computational protein structure prediction, holding great implications on structure-based high-throughput virtual screening, a routinely used method in early-stage drug discovery.<sup>542,543</sup> Generative DL approaches have recently been used to design new small molecules and proteins,<sup>224,544,545</sup> with multiple drug discovery companies now progressing AI-driven designed molecules into clinical trials.<sup>538</sup> Notably, in 2019, Zhavoronkov *et al.* showed one of the first examples of generative DL accelerated drug discovery, with the 6 possible lead compounds for DDR1 kinase inhibitors verified by manual biological assays.<sup>224</sup> Ren *et al.* later demonstrated that the same workflow was effective in finding lead compounds for dark proteins—those with no experimentally known structure—using AlphaFold to find the protein structure and binding pocket.<sup>546</sup>

While automation is now routinely used in the pharmaceutical industry, and AI has made its debut into the pipeline, these components have mostly remained disconnected from each other. Therefore, extensive human input, interface between different steps, and external control is still needed. By combining both into a closed-loop manner, SDLs could help reduce the current bottlenecks and also eliminate human biases in hypothesis generation.<sup>547</sup> However, there are two important challenges that drug development does not share with any of the other topics discussed in this review: (i) drug development spans vast length- and time-scales unlike any other SDL system and (ii) biological experiments provide very noisy responses, especially as the complexity of the organism increases.

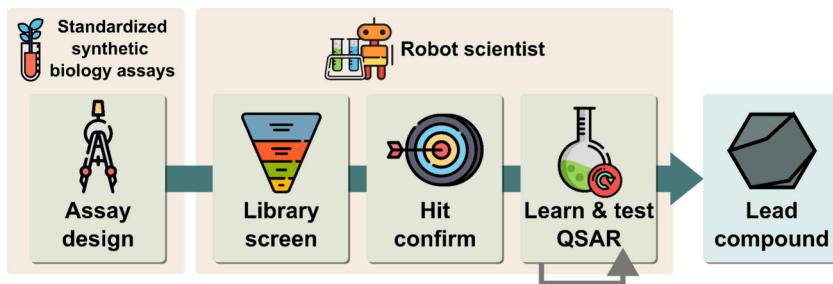
Since the stages of drug discovery typically occur sequentially with target identification, hit discovery, hit-to-lead, and lead optimization being distinct stages, it is not surprising that SDLs for drug discovery typically focus on optimizing one stage of the pipeline at a time. Therefore, we will assess the progress in the adoption of SDLs in the pharmaceutical industry by looking at their implementation at different stages of the small-molecule discovery pipeline, mainly focusing on early stage research and preclinical studies. We also dedicate a section to discuss the broader application of SDLs to protein engineering and synthetic biology. We limit our discussion to SDLs applied to biochemistry, such as the

development of small molecule drugs, molecules and polymers for biologics, nanomedicines, and production of chemical matter through biological systems.

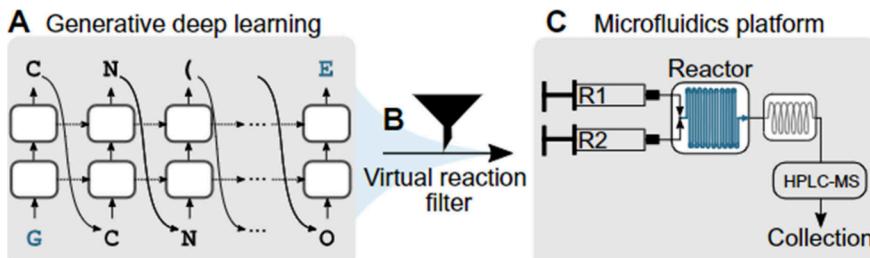
### 5.1. Drug Discovery Pipeline

**5.1.1. Target Identification and Validation.** In modern early-stage drug discovery, identifying a target, a gene or protein that is involved in a disease, is a critical initial step.<sup>548</sup> A great demonstration of the benefit of automation in target identification is the robot Adam that was developed by King *et al.* to perform high-throughput automated microbial batch growth experiments which are individually designed.<sup>37</sup> Adam was used to identify which genes encoded locally orphan enzymes in *Saccharomyces cerevisiae* (i.e. enzymes with unknown encoding genes).<sup>549</sup> The stages of Adam's workflow included generating hypotheses; generating, designing, and performing experiments, collecting optical density (OD) data, forming growth curves from the OD data; recording and analyzing data; relating the data back to the hypotheses. The hypotheses suggested potential encoding genes for locally orphan enzymes. They were generated using bioinformatics software and databases. For the experiments, several modules including a robotic arm, plate slides, plate centrifuges, and plate washers were embedded in the high-level automation workflow, shown in Figure 30. Notably, the hardware did not require human intervention other than replacing materials, and could hypothetically run for a few days without human supervision. However, it was still at risk of encountering problems where a human would be needed to solve them. In addition, its hypotheses were indirect and required additional experiments and literature searches by the authors to verify Adam's hypotheses.

Along with the hardware-enabled acceleration of target discovery, AI has emerged as a powerful engine in finding targets. Recent developments in AI for target identification and validation were reviewed by Pun *et al.*<sup>548</sup> While the authors suggest that combining AI with automated target validation and screening can potentially increase the efficiency of these stages of early drug discovery, the integration of AI approaches into SDLs has remained elusive. The lack of robotic automation in target identification studies could be due to the fact that biological experiments have inherent challenges including the extrapolation of results from small-scale experi-



**Figure 31.** Diagram of the early stage drug discovery pipeline. Robot scientist Eve<sup>38</sup> is designed to automate the library screening, hit confirmation, and QSAR steps of the pipeline.



**Figure 32.** Automated pipeline for liver X receptor (LXR) agonist discovery. (A) The DL generative model was used to design candidate molecules. (B) A virtual reaction filter screened for synthesizability of the candidates. (C) Finally, select candidates were synthesized using a microfluidic platform. Figure reproduced with permission from Grisoni *et al.*<sup>553</sup> Copyright 2021, American Association for the Advancement of Science.

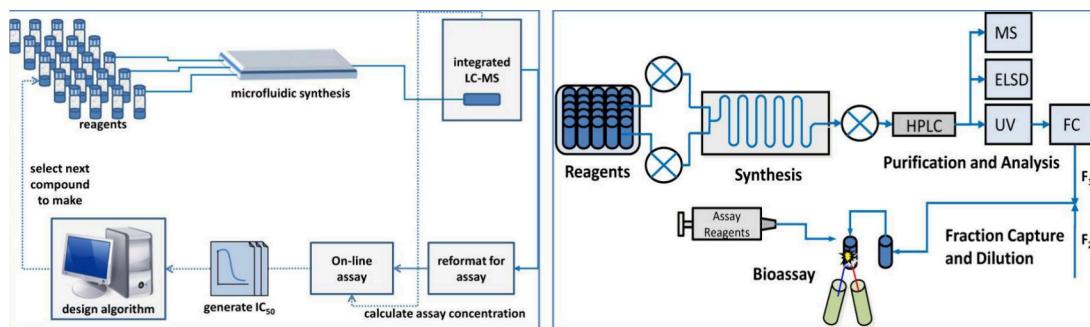
ments to emergent behaviors in biological systems, and predicting the phenotype of systems with altered DNA.<sup>550</sup>

**5.1.2. Hit Discovery.** Once a target is identified, the traditional drug discovery pipeline enters the compound screening phase, where compounds are screened to find “hits,” compounds that display interaction with the target or desired activity during screening.<sup>551,552</sup> This includes assay development and high-throughput screening to conduct pharmacological, chemical, and genetic tests. In recent years, developments have been made in the automation of various aspects of hit discovery, such as virtual and experimental screening, and assay optimization, which represent important steps towards closed-loop drug discovery.<sup>38,543,553–555</sup>

In 2015, Williams *et al.* reported the development of the robot scientist, Eve (Figure 30). Eve was developed to perform high-throughput screening of more than 10000 compounds per day for drug discovery.<sup>38</sup> Eve operates in three modes: a library-screening mode which involves grid search testing of a randomly chosen set of compounds from its library, a hit confirmation mode in which Eve re-assays hits, and an “intelligent screening” mode where Eve autonomously hypothesizes and tests QSARs. Figure 31 shows how these three modes fit into the greater early stage drug discovery pipeline. Eve generates QSARs using a GP with a linear kernel.<sup>176</sup> In addition, active learning using a greedy strategy is implemented to select batches of 64 compounds to test Eve’s hypotheses. The authors made a semantic data model of the screening assay results. The flexibility of Eve’s design allows for the easy definition and modification of assays, including, e.g., general, standardized assays (such as computational assays), targeted assays (such as biochemical assays), and biologically realistic assays and screens for toxicity (such as a cell-based assay). All three modes are integrated with software that communicates with the robotics within Eve’s framework. For the robotics, Eve uses off-the-shelf automation equipment for

laboratories. Examples include robotic arms and linear actuators for plate transfer, liquid handling systems for sample transfer, and shaking incubators for screening reactions. For analysis, Eve can measure fluorescence, absorbance, cell morphology (using microplate readers), and bright-field and fluorescence images, with an automated microscope. Once the assay is created and the QSAR problem is defined, Eve can run with minimal human intervention. Remarkably, Eve can further be used to discover new targets for existing drugs; Eve uncovered a second target for an anti-cancer drug which makes it a potential candidate for treating malaria. Eve was also used to compare its intelligent screening with grid search screening, with the authors concluding that intelligent screening is less expensive than grid search screening for pharmaceutical screening which uses large libraries and expensive compounds. While Eve shows great strides towards an SDL since it can optimize the activity of drug molecules for a particular target in a closed-loop fashion, and is proven to be useful for repositioning drugs, one drawback of the platform is that it is not connected to an automated synthesis platform, and therefore it is limited to only testing compounds in its library. Integrating the automated synthesis of new compounds into the pipeline would greatly expand the capabilities of Eve.

More recently, Grisoni *et al.* developed an automated pipeline for hit discovery of liver X receptor (LXR) agonists.<sup>553</sup> They combine a DL generative model and automated synthesis in one platform. This modular system, shown in Figure 32, consists of a design module that uses a RNN based generative model with long-short term memory cells to design new molecules as SMILES strings, a verification module that virtually confirms the synthesizability of the designed molecules, and an automated bench-top microfluidics platform that runs the synthesis. The microfluidics platform retrieves reagents, optimizes reaction conditions, and performs one-step reactions to synthesize compounds. The reactions are



**Figure 33.** (left) Schematic of the integrated design, synthesis, and screening platform illustrating the fully automated processes implemented for closed-loop drug discovery. Following initiation of the process the system completes multiple iterations of design, synthesis, and screening without manual intervention. (right) Schematic showing the continuous fluidic path taken by reagents and products on the platform. Figure reprinted with permission from Desai *et al.*<sup>558</sup> Copyright 2013, American Chemical Society.

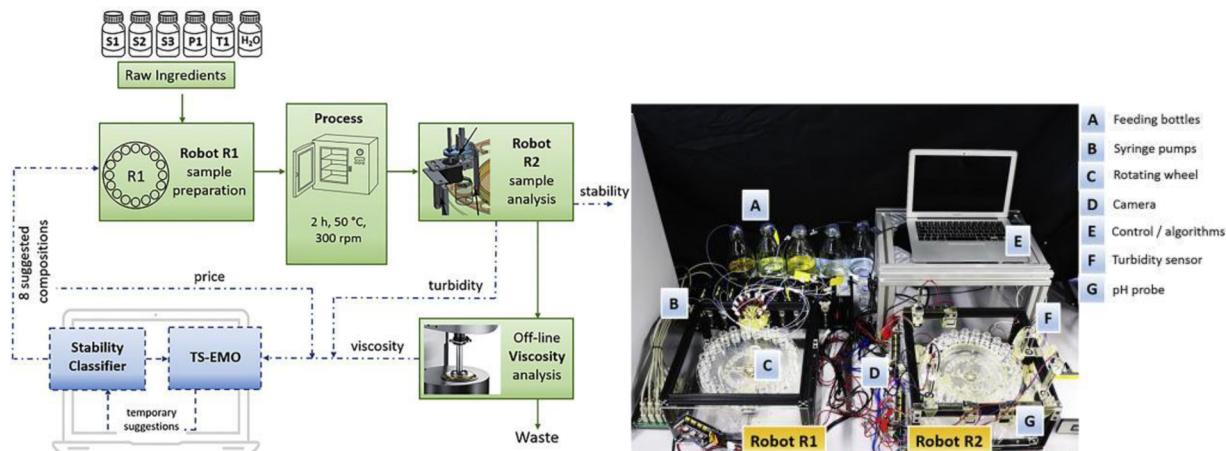
monitored using HPLC-MS, and the crude reaction mixtures are collected automatically. The only human intervention needed to operate the entire platform is selecting the compounds for pretraining and fine-tuning the model. The authors demonstrate one “iteration” of their pipeline, and do not feed results back from the reactions to the design module. The platform synthesized 61% of the computationally designed molecules in this study. In addition to the automated experiments, the authors performed batch synthesis and further screening of select compounds to confirm activity. Through this study, 12 novel, active LXR agonists were found. Although this platform is not closed-loop, it is a successful example of automated drug design and synthesis, and shows potential to be incorporated in a closed-loop platform.

An enzyme assay is an experimental method which qualitatively or quantitatively assesses the activity of an enzyme.<sup>556</sup> With assays being an important part of high-throughput screening, optimizing assays is an area of research in itself, and therefore, automating the assay optimization process is pertinent to creating an SDL for hit discovery. One demonstration of automated assay optimization comes from Elder *et al.*<sup>554</sup> They used a cloud-based BO based algorithm, along with automated experiments to optimize a cell-free papain biochemical enzymatic assay for papain inhibitors. The optimization involves minimizing final enzyme concentration, final substrate concentration, and incubation time, while maximizing the value of  $K'$ , which is a statistical parameter that uses control data to assess the quality of assays.<sup>557</sup> The automated platform included liquid dispensers, microplate reader for fluorescence measurements, and automated microplate washing. Their platform tested, on average, 21 assay conditions in order to find the best conditions, therefore being more efficient and less expensive than other methods such as grid search which requires testing all 294 conditions. This demonstrates the advantage of a closed-loop experimental platform, where the experimental results are fed into the optimizer to suggest future experiments. In addition, the automated platform allows the optimization process to be controlled remotely. Other assays could be optimized on this platform and the technology can be applied to other areas of drug discovery, such as reaction screening and hit selection.

Finally, Kanda *et al.* reported BO combined with automated experiments studied in another context: optimizing a cell culture to produce induced pluripotent stem cell-derived retinal pigment epithelial (iPSC-RPE) cells.<sup>555</sup> In this study, the target protocol (differentiation of iPS cells to RPE cells),

seven parameters (one parameter for reagent concentration, four parameters for the duration of certain steps, and two pipetting parameters), and validation function are defined by users. The robot booth included a microscope, dry bath, plate and tube racks, an aspirator, a dust bin, a tip sensor, pipette tips, micropipettes, a CO<sub>2</sub> incubator, and a dual arm robot. While the seeding, preconditioning, passage, RPE differentiation, and RPE maintenance steps of the experiments were performed by the robot, there was still a considerable amount of human labour involved in the process: initiating and preparing cell suspensions, preparing various reagents, importing plates into and out of the robot booth, taking images of the samples and analyzing them, further processing and testing the cells and media collected from the experiments. In addition, the conditions used for this study, including the robotic equipment, parameters, and scores, are not necessarily directly transferable to different protocols, and must be re-evaluated when designing a new study. This platform was able to improve iPSC-RPE production by 88% in 111 days through testing 143 different cell culture conditions. The authors also found that the robot generated cells which satisfy the criteria for research applications in regenerative medicine. While the work focused on the study of regenerative medicine, the authors’ method is not unlike the other examples shown above for hit discovery, and may be applicable to hit discovery platforms as well. This platform has the advantage of being closed-loop, with three rounds of BO performed with a GP surrogate, however, there is room for improvement. Making the platform more flexible to accommodate different types of experiments, and increasing the amount of automation could reduce the amount of human labour required to run and design the experiments, bringing this platform closer to an ideal SDL.

**5.1.3. Hit-to-Lead and Lead Optimization.** The goal of the hit-to-lead stage is to evaluate and perform optimization on the “hit” compounds from the previous substage to identify which ones are most susceptible to turn into “lead” compounds. Once a lead is found, it usually undergoes multiple rounds of optimization to improve potency and reduce side effects. The integration of SDLs at this substage would answer one of the core demands of the pharmaceutical industry. In fact, while it is relatively straightforward to identify numerous hit compounds virtually or via HTS, prioritizing those for further stages requires medicinal chemistry intuition and testing those hypotheses more thoroughly. Since this process is iterative, it is well amenable to the DMTA paradigm, and therefore to SDL integration.



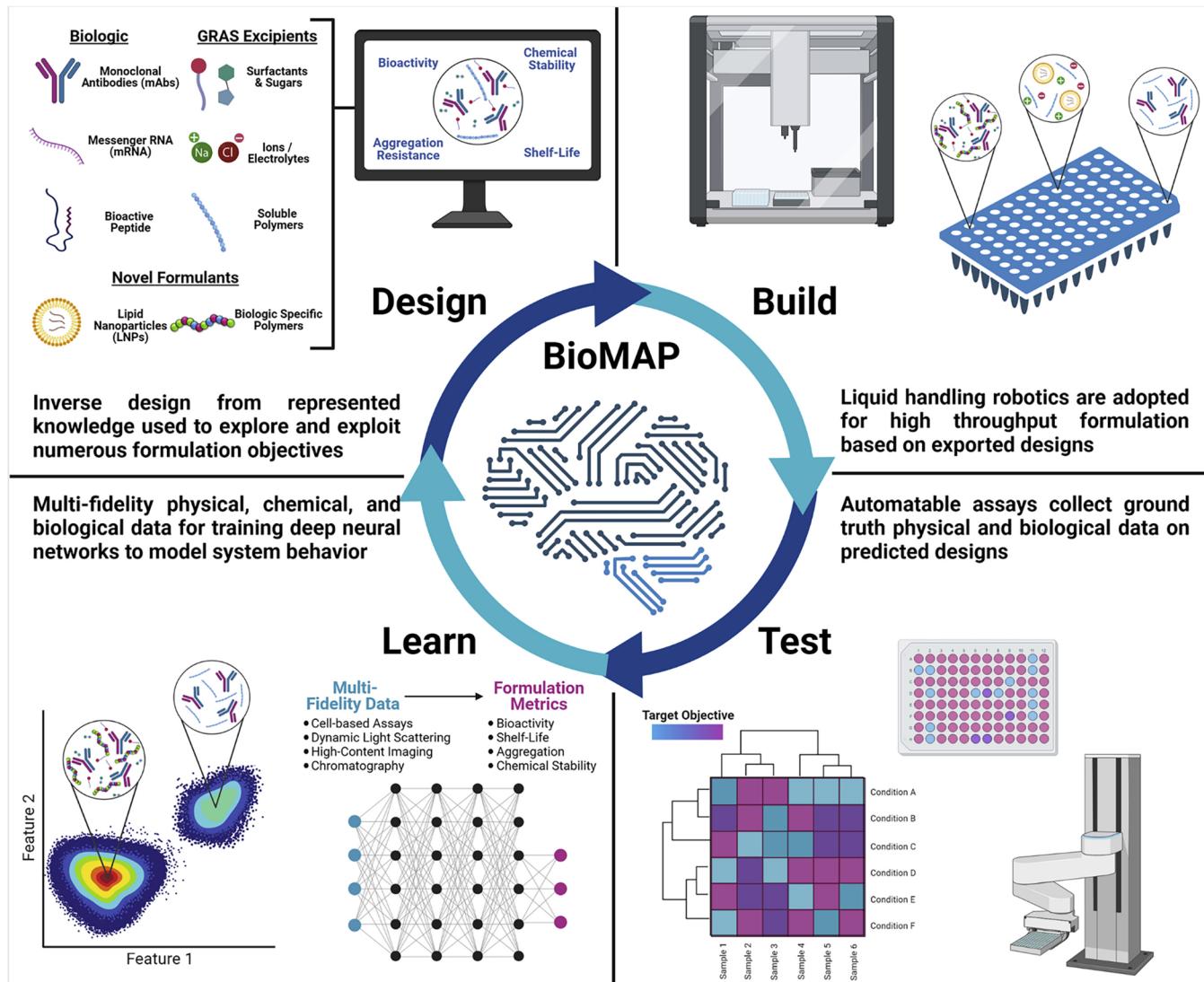
**Figure 34.** (left) Scheme of the adopted closed-loop optimization workflow. Material flow (continuous lines) and information flow (dashed lines) are reported. Ingredients are mixed following the suggested recipes in robot R1, processed, and analyzed with a combination of in-line automated operations and offline manual analyses. The results are then collected and processed by the algorithm to suggest a new set of experiments to run for the next iteration. (right) Image of the experimental setup based on the two formulation robots, R1 and R2. The picture shows the actual experimental setup as used for the experiments. Automated syringe pumps (B) are connected to feeding bottles (A) to dispense ingredients to different vials located on the rotating wheel (C) of robot R1. Samples are then moved to the offline incubator for processing and placed in robot R2, where image collection (D), turbidity (F), and pH (G) analyses can be run. The platforms are controlled by the PC (E), where data is stored and fed to the algorithm for the generation of the next iteration. Figure adapted with permission from Cao *et al.*<sup>563</sup> Copyright 2021, Elsevier.

In 2013, Desai *et al.* designed a fully integrated flow-based autonomous platform assisted by an algorithm design (CyclOps) to perform hit-to-lead optimization, showcasing its use in the case of AbI Kinase inhibitors (Figure 33).<sup>558</sup> Starting from ponatinib as a hit compound, the authors defined a chemical space of 270 molecules that could be synthesized in the automated workflow by structural analysis of potanib-bound AbI Kinase. The design algorithm would then select compounds from this space to be synthesized on the platform using Sonogashira reactions in flow, purified by in-line preparative HPLC, and analyzed for kinase activity in real-time. The authors used a RF model for activity prediction that used drug-like molecular descriptors involving the Lipinski rules and molecular fingerprints, initially trained on 36 literature compounds. Three design strategies were set up : (i) “chase potency,” an exploitative strategy selecting top-scoring compounds based on predicted activity, (ii) “most active under sampled one,” an explorative strategy accounting for the number of times certain reactants have previously been employed and (iii) a hybrid strategy combining (i) and (ii). Overall, the flow chemistry, purification, and bioassay proceeded with a success rate of 71%. In all, 11 key compounds were identified as potent inhibitors of Abl1/Abl2, with IC<sub>50</sub> values in the low nanomolar range. Those were retested with conventional bioassay methods, and the data generally showed a high level of correlation with data generated via the microfluidic platform. In a subsequent paper, Czechizky *et al.* demonstrated the reproducibility and consistency of their platform by applying it to replicate xanthine-based dipeptidyl peptidase 4 (DPP4) inhibitors.<sup>559</sup> This time, the compounds were synthesized via a two-step synthetic protocol using a Vapourtec R4 flow chemistry system. Overall, 29 compounds were prepared in high purity and tested in only three days with a chemistry success rate of 93%. Close correlation between the microfluidics platform data and data generated within traditional approaches was observed once again.

Recently, the CyclOps platform was used to develop hepsin inhibitors selective against urokinase-type plasminogen activator (uPA).<sup>560</sup> Over the course of 9 days, 142 novel compounds were generated and assayed with hepsin and uPA. The algorithm explored a virtual chemical space of 5472 molecules, spanning three types of commercially available reagents—a sulfonylating/acylating agent, an amino acid and an amino amidine. Each closed-loop cycle took approximately 90 min on the platform. The authors alternated between exploitative and explorative strategies, but also conducted several grid-search rounds focused on the variation of a specific reagent. The progression from the initial hit to the lead compound was accompanied by an improvement in inhibitory activity against hepsin from ~1 μM to 22 nM. The selectivity over uPA was improved from 30-fold to >6000-fold. The lead compound found was also further ADMET-profiled (i.e., absorption, distribution, metabolism, excretion, and toxicity) and tested in oncogenic functional assays. When assayed against a panel of 10 serine proteases, it displayed promising selectivity.

The CyclOps platform is a great example of concrete application of SDLs to drug discovery development. Leveraging microfluidics for compound synthesis in a combinatorial fashion and coupling it to the RF algorithm allowed saved experimentation time and chemical resources, while leading to the discovery of compounds with enhanced properties. A weakness of such demonstration was that the RF algorithm was only optimizing the compound activity, so it could not be part of decision-making in the event of any synthesis- or process-related issues, e. g. poor reactivity or solubility. One can find more discussion on SDLs integrating Reaction optimization.

Recent work from Novartis Medical Research addresses synthesis optimization within hit-to-lead optimization in their microscale SDL. Brocklehurst *et al.* developed the Micro-Cycle<sup>561</sup> platform, an integrated workflow that connects the infrastructure of Novartis with software tools and a robotics system to create a closed-loop cycle. Candidates are designed through an RF model trained on in-house data for QSAR of



**Figure 35.** Overview of the BioMAP platform for biologic formulation. Biologic formulation is performed entirely through autonomous workflows. Multisource data from physical and biological experiments are exploited by deep neural networks to map complex structure-function landscapes and inform downstream design campaigns. Figure reproduced with permission from Tamasi *et al.*<sup>569</sup> Copyright 2022, Elsevier.

physicochemical and biochemical properties. The RF model is then incorporated in a BO campaign, in some cases along with protein docking results, for selecting molecules in the synthesis step. From acquired building blocks, the robotics platform, equipped with automated solid dispensing, liquid handling, and a robotic arm, autonomously performs optimization of reaction conditions and high-throughput microscale synthesis. For the test stage, the MicroCycle platform includes an integrated plating process to prepare microscale assay-ready plates and can perform many types of assays automatically, including physicochemical assays, ADME (i.e., absorption, distribution, metabolism, and excretion) *in vitro* assays, and target-specific biochemical assays. Starting from a hit compound with moderate activity, the authors used MicroCycle to generate 13 libraries of compounds from 8 reaction types, showcasing the use of their predictive models, automated synthesis, and purification. Over 440 molecules were made and an average success rate of about 50% was achieved, meaning that about half of the syntheses were sufficient for running an assay. With additional analysis and contributions from medicinal chemists, molecules with improved activities and potency were

identified, while maintaining good solubilities, and appropriate molecular weights.

#### 5.1.4. Formulation Optimization and Bioavailability.

Drug formulation is an essential stage in the discovery and development of new medicines, allowing to improve bioavailability and targeted delivery. Traditionally, designing drug formulation relies on iterative trial-and-error, requiring a large number of resource-intensive and time-consuming *in vitro* and *in vivo* experiments. However, the field has recently experienced a growing interest in integrating ML and automation approaches into the design process, as described in the review of Bao *et al.*<sup>562</sup> As optimizing drug formulations implies varying multiple parameters related to the drug, excipients, and manufacturing conditions, SDLs could help navigate this highly dimensional space.

One example of formulation optimization using an SDL is the work conducted by Cao *et al.*,<sup>563</sup> although this example is not directly tailored to a pharmaceutical application. In this work, a commercial formulation consisting of a mixture of three different surfactants, a polymer and a thickener was optimized in a closed-loop fashion according to the following

multi-objective (Figure 34): (i) stability and low turbidity, (ii) high viscosity and (iii) low ingredients costs. The TS-EMO algorithm was chosen to suggest formulation parameters<sup>318</sup> and coupled to an SVM classifier (trained on initial experimental runs) that classified its temporary suggestions based on their stability. The algorithm was run until the classifier identified eight stable formulations amongst the suggested ones, which were then synthesized automatically using a first robot, and transferred to a second one that performed pH, turbidity, and stability tests. Unfortunately, the samples had to be taken offline to measure viscosity. In 15 working days and without providing any explicit physical intuition to the system, the authors were able to obtain satisfactory formulations.

Another example in the literature of SDL for formulation is the work of Grizou *et al.*<sup>564</sup> in which a new high-throughput droplet dispensing robot was coupled to a Curiosity Algorithm (CA) to study the behavior of dynamic oil-in-water droplets, which serve as promising protocells models—a synthetic cell-like entity that contains non-biologically relevant components. The authors defined their parameter space by choosing mixtures of four oils and set a budget of 1000 experiments to observe how varying the oil mixture impacted the speed of the droplets and their division. This observation space was chosen for its simple life-forms-like behavior, which can move and replicate. To do so, small oil droplets are placed at the surface of an aqueous medium, the droplet movements are then video recorded and analyzed using traditional image processing techniques to deduce the speed and division of the droplets. To select the next oil mixture to be tested, the CA first feeds previous observations to a locally weighted linear regressor that approximates the mapping between input parameters and observations. A random target observation is then selected and fed to the numerical inverse of the regressor to infer the most probable experimental parameters that will lead to the target observation. The fully closed-loop platform can conduct more than 30 experiments per hour by leveraging parallelization, a six-time throughput increase from previously reported ones. By leveraging the CA, the speed observation space was more efficiently explored, with only 128 experiments needed to cover the portion of the observation space that random parameter search covered in 1000 experiments. The number of droplets deemed active (with speed >3 mm/s) was also improved 14-fold, without it being an explicit objective.

Moreover, two whitepapers on SDL concepts have recently been reported in the formulation development literature, demonstrating the high interest of this field for automation. Hickman *et al.* have proposed an SDL named NanoMAP, which focuses on the development of nanomedicines for pharmaceutical formulations.<sup>565</sup> Nanomedicines commonly consist of a combination of polymer and/or lipid-based materials or excipients that encapsulate small molecules or biologic-based active agents.<sup>566</sup> The authors propose to automate the preparation of nanomedicines for screening using nanoprecipitation via liquid-handling robots, while coupling it with active learning strategies. Importantly, this experimental protocol has previously been successfully implemented and can be scaled up by leveraging a microfluidics platform.<sup>567,568</sup> On the characterization side, the drug loading capacity (DLC) and encapsulation efficiency (EE) would be automated with appropriate extraction methods and analysis via HPLC. The authors also plan on automating high-throughput *in vitro* stability and release assays in biorelevant

media using 96-well dialysis plates, as well as particle size measurement using dynamic light scattering (DLS) plate readers.

Tamasi *et al.* proposed the development of BioMAP for biologic formulation design (Figure 35).<sup>569</sup> Indeed, while therapeutic proteins and vaccines—commonly called biologics—have proven their therapeutic efficacy, they remain extremely fragile under standard pharmaceutical storage and handling conditions, c.a. -78°C. Therefore, extensive formulation efforts are routinely required to avoid their denaturation, using additives such as small-molecule stabilizers, polymer excipients, or surfactants. This is also observed for monoclonal antibodies (mAbs). The authors' plan on building on their previous experience of coupling automation and ML (see below section on [Engineering the stability of proteins](#) for a detailed discussion) to create a fully autonomous platform for optimization of tailored polymer additives. They also aim at increasing their materials library to generally recognized as safe (GRAS) excipients as well as expanding the platform to liquid NP formulation. This ambitious project necessitates careful design of the automated instrumentation, as it must remain flexible for each biologic type. The authors plan on providing extra supportive modules such as a multimode reagent dispenser, a plate heater/shaker, a plate sealer and a vacuum filtration system on top of a multi-purpose liquid-handling robotic system for cell culture and reagent mixing. For liquid NP production, microfluidics and continuous-flow fluidics would be leveraged. Testing the stability of formulations over a range of storage and handling conditions could be carried out using an automated microplate incubator with humidity and CO<sub>2</sub> regulation, which could also be used to support cell-based assays. For characterization UV-visible (UV-vis) and DLS plate readers, size-exclusion chromatography (SEC), and a high content imager (HCI) would be employed.

A proof-of-principle synthesis and formulation platform by Adamo *et al.* showcases technological advances in continuous-flow synthesis and formulation of pharmaceuticals that could be incorporated into SDLs for formulation optimization.<sup>570</sup> The platform's capabilities included multistep synthesis, purification, crystallization, real-time process monitoring, and formulation. In addition, it was reconfigurable to produce pharmaceuticals with diverse chemical structures and synthesis routes on-demand. With the entire system being approximately the size of a refrigerator, and total cycle time for synthesis and formulation being up to 48 hours, the authors demonstrated a competitive alternative to the usual batch synthesis of pharmaceuticals which can take up to 12 months and involves multiple synthesis steps and formulations occurring in different locations. The authors showed the successful synthesis and formulation of four common drugs in liquid formulations, and were able to achieve a capacity of up to 4500 doses per day of diphenhydramine hydrochloride. Automated components of the platform include pumps, heating reactors, multichannel valves, and gravity-based separators, as well as precipitation, filtration, and crystallization tanks. Only one user is required to operate the entire system. While the purpose of their platform was to demonstrate the production of liquid formulations of common drugs, the design and technology integrated into the SDL could be advantageous not only for formulation optimization, but also for time and cost effective synthesis of molecules in hit discovery or hit-to-lead workflows. However, one drawback of this system that would make it difficult to use in a high-throughput setting is the turnaround time for

reconfiguring and cleaning the system, which could take as long as two hours, or potentially longer depending on the complexity of the synthesis. In addition, it is not equipped for solid formulations.

Ortiz-Perez *et al.*<sup>571</sup> proposed an integrated and semi-automated iterative workflow that combines microfluidic-assisted nanoparticle formulation, automated fluorescence imaging and analysis with BO to design poly(lactic-co-glycolic acid)-polyethylene glycol (PLGA-PEG) NPs with high uptake in human breast cancer cells. To maximize the uptake, one process variable—the flow rate ratio between solvent and antisolvent—and 4 polymer components that directly influence physicochemical properties (size, PEGylation and charge) were varied. The polymer mixture can be automatically prepared using a syringe pump, and injected into a microfluidic chip at a constant flow rate while the antisolvent rate is adjusted. This automatically produces NPs with controllable size and composition, which can be labeled *in situ* by incorporating a fluorescent dye during formulation. To measure uptake in cells, NPs are added to cells in 96 well plates and fluorescence microscopy is used to acquire and process widefield fluorescence images in an automated way. This measured response per NP is used to train a BNN to predict nanoparticle uptake from nanoparticle formulation. The uptake, polydispersity index and size of a virtual library of 100,000 NPs spanning the entire design space homogeneously are then predicted using the BNN and tree-based models respectively. The next formulations are then selected from this pool, based on the models' predictions. With two 5-day experimental cycles, the authors were able to triple the measured NP uptake.

While concrete examples of SDLs for drug formulation optimization are still sparse, the excitement of the field for ML techniques and automation foresee a bright future for SDLs. Nevertheless, several questions are left to be answered; notably, how the diversity of therapeutic compounds should be handled. Indeed, contrary to most of the SDLs mentioned in other sections, one would expect the platform to be adaptable to a wide range of drugs, as proposed in the work of Tamasi *et al.* This is likely to complicate the coordination of experiments both on the software and hardware side. Finally, while this can also be relevant to other stages of the pipeline—and applies to other fields—Lammers *et al.* highlighted the lack of standardization in studies conducted to characterize formulations such as nanoparticles.<sup>572</sup> Better guidelines and reporting is thus needed to provide the best conditions for ML application to thrive in the field.

## 5.2. Synthetic Biology

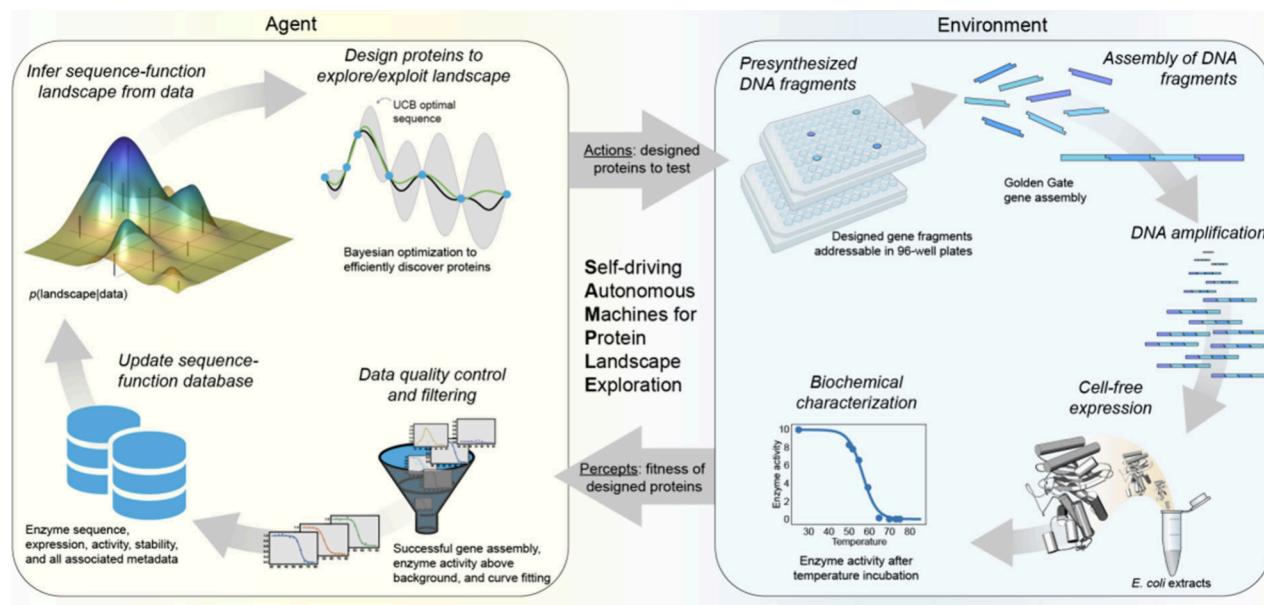
While so far mostly the discussion has been focused on reviewing SDLs for small molecule therapeutics development, it is important to highlight that the deployment of SDLs has also sparked interest in the fields of protein engineering and synthetic biology,<sup>550,573</sup> even inspiring space biologists.<sup>574</sup> This naturally extends the potential applications of SDLs to a broader range of therapeutics, but also to the design of new biomaterials or biofuels for instance. The challenge resides in the inherent complexity and non-linearity of biological phenotypes, the high-dimensionality of genomic search spaces, and the error-prone and difficult-to-automate nature of biological experiments.

**5.2.1. Biosynthetic Pathways Optimization.** The microbial synthesis of chemicals offers a viable alternative to

widely employed chemical manufacturing methods. Biosynthetic production is appealing due to its ability to leverage a diverse range of organic feedstocks, operate under benign physiological conditions, and circumvent the generation of environmentally harmful byproducts. However, natural cells are rarely fine-tuned to efficiently generate a specific molecule. To attain economically feasible production, significant alterations to the metabolism of host cells are frequently necessary to enhance metabolite titer, production rate, and overall yield.<sup>575</sup> Unfortunately, the complexity of biological systems and their multiple components and many unknown interactions among them lead to having to perform many DMTA cycles. Employing automation combined with computational approaches can help expedite this process.

In their paper, Carbonell *et al.* improved the production of flavonoid (2S)-pinocembrin in *E.Coli*, a natural product, by optimizing a 4-gene pathway (2592 possible configurations) by 500-folds, going through two DMTA cycles.<sup>576</sup> First, rule-based pathway and enzyme selection tools were employed to define a synthetic route for (2S)-pinocembrin. A combinatorial approach was then utilized to design 2592 synthetic plasmids, each corresponding to a pathway that varied in terms of gene expression and combination. Assembly recipes and robotics worklists were generated to automate the plasmid assembly—commercial DNA synthesis, part preparation via PCR followed by ligase cycling reaction. After this, growth of microbial production cultures was conducted in a high-throughput manner, products were automatically extracted, and screened via fast-LC and MS. Finally, the data was analyzed using standard linear regression to identify important experimental factors to aid in the design of the next iteration. The pipeline was designed in a modular fashion which would allow other laboratories to replace individual pieces of equipment or protocols to adopt their own methods. The authors tried to demonstrate the adaptability of their pipeline to other compounds production than (2S)-pinocembrin, choosing to optimize the expression of the (S)-reticuline—an alkaloid—pathway in *E.Coli*. While some of the obtained constructs matched literature titers, experimental difficulties were encountered, suggesting that the proposed arrangement was either unstable or negatively selected against in the cloning host. Further conversion of (S)-reticuline to (S)-scoulerine yielded modest results, although this alkaloid had not been produced in *E. Coli* previously. Each step of their pipeline leveraged automation, but the entire workflow was not fully integrated to enable autonomous operation. Indeed, PCR clean-up and host-cell transformation were carried out off deck, and plates needed to be manually transferred between certain platforms. Another point that weakens the SDL character of this work is that the “analyzing” component was conducted using a standard least squares linear regression, which could identify trends across the experimental parameters, but did not actively make design suggestions for the next cycle.

Hamedirad *et al.* focused the production of lycopene in *E.Coli*—a food additive and colorant recently proposed as anticarcinogenic—and demonstrated that BO could help optimize gene expression of a 3-gene pathway, outperforming random screening by 77% while evaluating less than 1% of possible variants (over a total of 13,824 possibilities) through three DMTA cycles.<sup>577</sup> The experiments were conducted using the iBioFAB biofoundry, a fully automated and versatile robotic platform consisting of a 6-degree-freedom articulated



**Figure 36.** Schematic of the SAMPLE workflow. An intelligent agent (left) learns sequence-function relationships and designs proteins to test hypotheses. The agent sends designed proteins to a laboratory environment (right) that performs fully automated gene assembly, protein expression, and biochemical characterization, and sends the resulting data back to the agent, which refines its understanding of the system and repeats the process. Figure reproduced with permission from Rapp *et al.*<sup>582</sup> Copyright 2024, Springer Nature.

robotic arm that travels along a 5-meter-long track to transfer microplates among more than 20 instruments installed on the platform and a 3-degree-of-freedom arm moves labware inside a liquid handling station.<sup>578</sup> Each instrument is in charge of a unit operation, such as pipetting and incubation. They are linked by the two robotic arms into various process modules, such as DNA assembly and transformation, and then further organized into workflows such as pathway construction and genome engineering. An overall scheduler orchestrates the unit operations and allows hierarchical programming of the workflows. In this work, the iBioFAB platform was used to automate the lycopene pathway DNA assembly with different expression levels of genes using the Golden Gate method,<sup>579</sup> as well as the transformation, cell cultivation and lycopene extraction. The BO algorithm proposed DNA assembly designs, which were then converted into robotic commands for iBioFAB to conduct the complex pipetting work using a Tecan liquid handler. The best mutant found produced 1.77-fold higher lycopene titer than the best mutant found using random sampling and the number of evaluations was at least eight times less than the regression-based optimization scheme.

This work greatly demonstrates how a flexible platform like iBioFAB and BO can benefit from each other to efficiently explore the gene expression landscape. It is important to mention that the lycopene pathway was chosen as a proof of concept for its straightforward methods of extraction and quantification, which facilitated high-throughput execution using iBioFAB. Indeed, the biofoundry faced limitations linked to compound extraction methods and analytical/quantification methods requiring equipment more complex than a plate reader (e.g., GC-MS or LC-MS instruments). Those challenges could be overcome in the future by the development of larger-scale and more sophisticated biofoundries. Another possible improvement highlighted by the authors was that no initial assumptions about the landscape were made prior to using BO. Using the trained model for one system as the starting point

for a similar system could potentially result in reducing the number of evaluations to find the optimum.

**5.2.2. Engineering Protein Stability.** Proteins are sparking increasing interest in the context of biomedicine and pharmaceutical sciences. However, the moderate stability of proteins, and specially enzymes, is the major drawback hindering the generalized application of these bioactive molecules at the industrial scale. Indeed, the current process conditions may include extreme temperatures, pH values, or presence of organic solvents that are outside the operating stability window of the biomolecule, but that are often necessary to solubilize poorly-water-soluble substrates in high concentration values. Therefore, ensuring protein stability is of tremendous importance to unleash their potential and is a very active field of research.<sup>580</sup> Enabling protein stability optimization via SDLs could help to efficiently explore the combinatorial search space that relates protein sequences to their function, and ensure reproducibility of such results.

Tamasi *et al.* investigated the use of active learning for protein-polymer hybrids (PPHs) using copolymers. PPHs are a promising way to address challenges such as solubility and low physical stability of proteins by conjugating them with synthetic polymers.<sup>581</sup> They leveraged a GP regression model with BO to identify candidate copolymers for three chemically distinct enzymes, namely HRP, GOx, and Lip, to maximize the retained enzyme activity (REA). While the synthesis of the proposed copolymers was automated, the PPHs formation and REA characterization were undertaken manually. Their discovery process invoked five DMTA cycles and resulted in enhanced thermostability for the three distinct enzymes (46.2%, 31.5%, and 87.6% improvement in comparison to the initial seed batch for HRP, Gox, and Lip, respectively). While this work is not an SDL in itself, the same group has then later proposed the conceptual outline of BioMAP (*vide supra*), demonstrating the incremental nature of SDL improvements.

Recently, Rapp *et al.* introduced the Self-driving Autonomous Machines for Protein Landscape Exploration (SAMPLE)<sup>582</sup> platform for fully autonomous protein engineering and were able to engineer glycoside hydrolase family 1 (GH1) enzymes with an enhanced thermal tolerance (Figure 36). The protein engineering task was framed as a BO problem that was tackled using a multi-output GP, combining GP regression on thermostability and GP classification on protein activity. First, the authors designed a GH1 combinatorial sequence space composed of sequence elements from natural GH1 family members, elements designed using Rosetta,<sup>583</sup> and elements designed using evolutionary information. This yielded a space containing 1352 unique GH1 sequences. To pick sequence candidates for the experiments, the authors designed a custom sampling strategy that constrained selection to the subset of sequences predicted as active by GPC. Within this subset, candidates for improved thermostability were then selected. To access the protein sequence space experimentally, SAMPLE relies on combining pre-synthesized DNA fragments using the Golden Gate method<sup>579</sup> to produce a gene, which can be amplified using PCR then expressed into the desired protein using T7-based cell-free protein expression reagents. Finally, the expressed protein is characterized using colorimetric/fluorescent assays to evaluate its biochemical activity and properties. The procedure took approximately one hour for gene assembly, one hour for PCR, three hours for protein expression, three hours to measure thermostability, and overall, nine hours to go from a requested protein design to a physical protein sample to a corresponding data point. The experimental pipeline was fully automated and implemented on the Strateos Cloud Lab.<sup>584</sup> To ensure reproducibility, four diverse GH1 enzymes from Streptomyces species were optimized for thermostability, each trial was composed of 20 DMTA cycles. Each resulting enzyme was at least 12 °C more stable than the starting protein sequences while exploring less than 2% of the defined protein sequence-function landscape.

This work demonstrates the use of SDLs for improved protein thermostability. Importantly, it confirms the reproducibility of the results across several GH1 enzymes, and reports, to our knowledge, the first example of BO coupled to a fully-automated Cloud Lab platform. A cloud lab is a fully-automated decentralized laboratory in which scientists can run multiple experiments simultaneously and remotely, all through a single digital interface. This type of facility allows researchers to have full control over their experiment without having to be physically present in the lab. Moreover, research can be conducted without purchasing costly lab instruments or leasing physical laboratory space.<sup>584</sup> Another strength of this work is the exception handling and data quality control mechanism implemented to further increase the reliability of the SAMPLE platform, allowing to flag experiments as inconclusive and add the associated sequence back to the potential experiment queue.

### 5.3. Outlook and Perspectives

The use of animal models has played a crucial role in the advancement of modern biomedical research, allowing the exploration of basic pathophysiological mechanisms, but also the development of new medicines. Indeed, because of their role in evaluating new therapeutic approaches, animal models bear the weight of the “go” or “no-go” decision to carry new drug candidates forward into clinical trials. However, the discordances between animal and human studies are frequent,

thus drug candidates may be eliminated for lack of efficacy in animals, or discovery of hazards or toxicity in animals that might not be relevant to humans.<sup>585</sup> The impressive expansion of the organ on a chip field bears the promise to address this issue by leveraging the latest advances in microfabrication engineering, microfluidics, genome editing and cell culture capabilities.<sup>586</sup> Recently, the FDA Modernization Act 2.0 was approved, allowing alternatives to animal testing for drug and biological product applications.<sup>587</sup> This change in legislature could potentially improve the adoption of SDLs in the preclinical phase of the drug discovery pipeline.

Given the great challenges of automating biotechnology, it is fundamental that laboratories collaborate and collectively develop guidelines and protocols by establishing global alliances<sup>588</sup> and consortia.<sup>589</sup> For instance, the adoption of the FAIR guidelines<sup>143</sup> for data-sharing is crucial for efficient use of DL and ML, especially in a field where many processes are yet to be explained and reproducibility can be an issue. The use of “cloud labs” could also allow researchers to access standardized equipment from anywhere at any time.<sup>584</sup> Furthermore, the stellar increase in interest for data-driven approaches applied to drug discovery calls for a reflection upon open science adoption in this industry. Indeed, a lot of success stories in DL stem from data availability and open-source libraries.<sup>166</sup> The AI for science excitement could therefore bring us to a long-awaited moment.<sup>590,591</sup> While we begin to see success stories where academia and industry join forces in an open science setting,<sup>592</sup> it is important to carefully consider and control the risks of misuse associated with sharing biological data and models.<sup>593</sup>

Throughout this section, we have reviewed the adoption of SDLs across the different stages of the drug discovery pipeline and in the field of protein engineering and synthetic biology. The works we highlighted showcase the great potential of SDLs to transform the current drug discovery process, while shining light upon the challenges ahead. On the algorithmic side, the recent DL achievements have highly increased the community’s excitement and its expectations. Nevertheless, several challenges are already foreseeable.<sup>594–596</sup> For example, the ligand-protein binding data from AI predicted protein structures is not as accurate as that which uses experimentally determined protein structures, and it is often not experimentally validated. In addition, more transparency from companies is needed in order to move the field forward as it helps to build trust in their results and allows larger sets of data to become available for ML. Federated learning is a way to overcome this challenge because it can keep company data confidential while using data from multiple companies in one ML model. On the automation side, we expect future advancements in microfluidics technology to greatly improve some of the systems covered above. Moreover, BioFoundries represent sophisticated hardware that is likely to have a great impact in the adoption of SDLs in this context, and could inspire other fields too. Furthermore, while proof-of-concept SDLs exist at various stages, the lack of seamless integration of feedback across these stages poses a potential limitation. Addressing this challenge and fostering collaboration between different phases could enhance the overall efficiency of SDLs. Overall, recent legislative changes, the growing significance of DL and the current advancement in hardware components suggest a favorable landscape for the adoption of SDLs in pharmaceutical science, paving the way for transformative advancements in drug discovery and biotechnology.

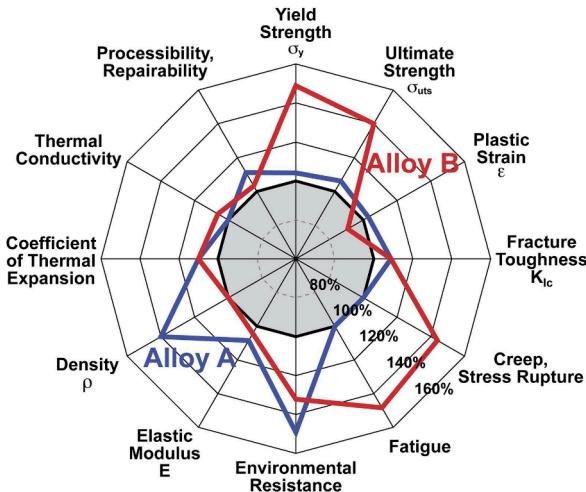
## 6. STRUCTURAL MATERIALS

This section covers materials geared towards structural applications with a focus on mechanical performance of materials. Naturally, we focus on materials and techniques that involve automated and data-driven methods including robotics, optimization algorithms, software orchestration, experiments, and simulations. We do so for alloy design, concrete formulations, non-alloy additive manufacturing, and adhesives. Many of these materials are inorganic solid state materials. All synthesis, processing, characterization, and testing methods of mechanical properties of solid samples require mechanical motion of some kind—whether through the transfer of solid samples or the movement of experimental apparatus around or in contact with solid samples. To date and to the authors' best knowledge, there are no published demonstrations of fully autonomous SDLs that use data-driven methods to iteratively explore inorganic solid state materials design spaces for large-scale structural applications with mechanical performance measurements. However, there has been a large amount of progress towards automated, and in some cases autonomous, synthesis and characterization methods for inorganic solid-state materials such as with Powder-Bot, A-Lab, and ASTRAL, as described in section on **Solid State Material Synthesis** optimization. Within each section and where appropriate, we point to the data-driven approaches that accelerate the discovery of these materials and which provide the foundation for structural-focused SDLs to come.

### 6.1. Alloy Design

From extreme-temperature Inconel alloys in jet turbines, lightweight aluminum alloys for engine blocks, biocompatible titanium alloys for hip replacement joints, to the classic corrosion-resistant 304 stainless steel found in the kitchen sink, the enabling effects of alloys are immense and often invisible. Events like the fatigue-induced catastrophic failure of “de Havilland Comet” commercial jet airliners in the 1950s, the steel corrosion-induced oil spill of the Alaskan Oil Pipeline in 2006, and the tens of thousands of metallic poisoning cases from cobalt-based hip implants up to 2010, is a reminder of how much the world relies on this class of materials and how the performance can enable (or the lack of performance can hinder) advancements in the transportation, energy, and medical fields.

While many alloys are known and in usage, the alloy discovery space is a largely unexplored and high-dimensional search space. In the case of multi-principal element alloys (MPEAs)—i.e., alloys with many constituent components—Miracle *et al.* estimates that there are nearly 200 million potential MPEAs systems with three-to-six constituent elements. Note that these are individual systems, meaning the tunable parameters of stoichiometry and processing conditions are ignored.<sup>597</sup> Over the course of twelve years from their discovery and documentation (2004–2017), the authors estimated that only 122 MPEA systems had been identified. Forecasting the current rate to the year 2117, and using only traditional methods, the risk of missing the best possible MPEAs system for a given application is over 99.999%, pointing to the need for physics-based, data-driven, automated, and high-throughput approaches. The alloy discovery space is not only high-dimensional, it is also multi-objective. See, for example, a spider plot with twelve performance properties for structural applications (Figure 37) such as yield strength, fracture toughness, thermal

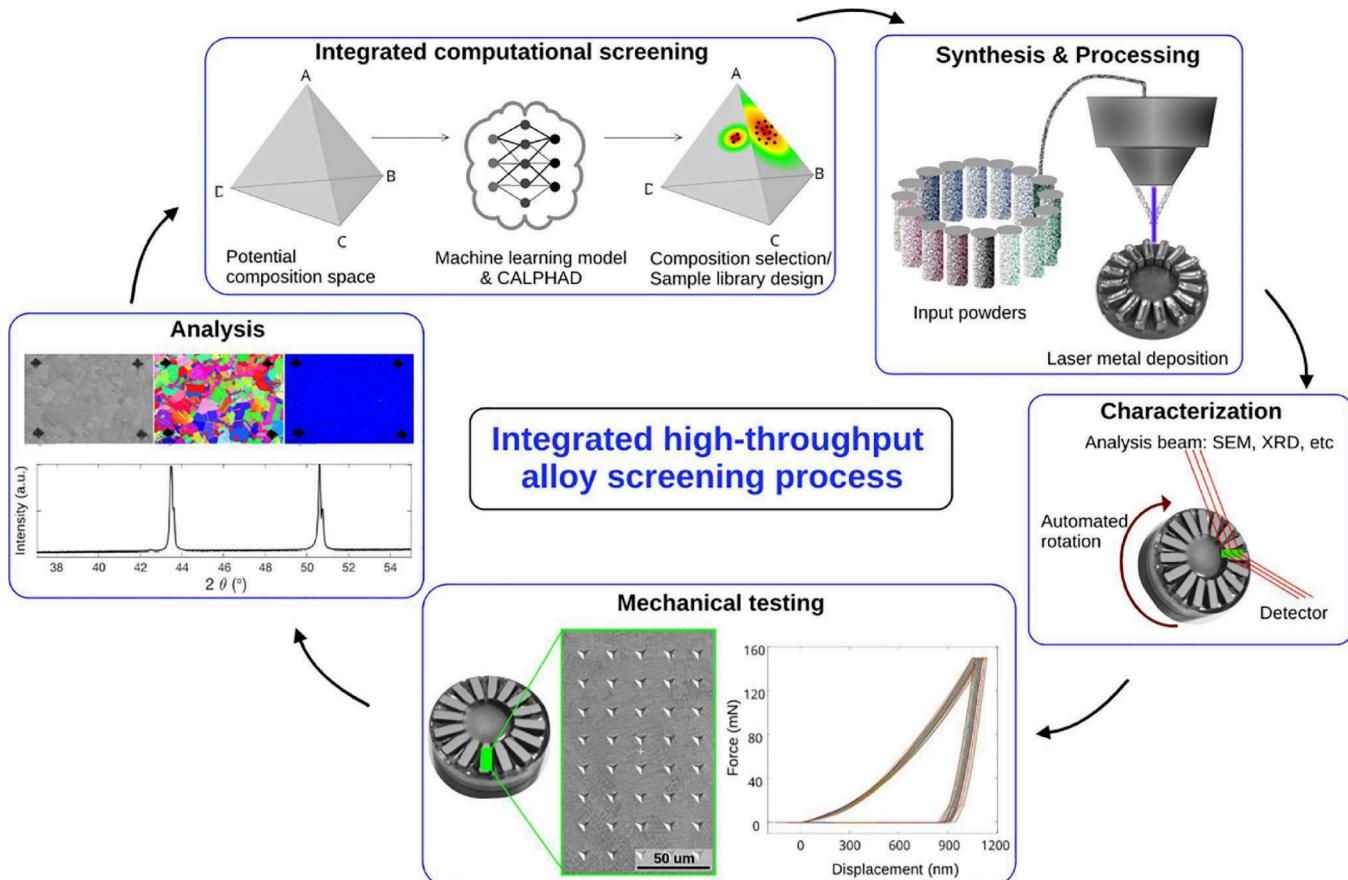


**Figure 37.** A spider plot illustrating twelve performance properties for structural applications. Two example alloys are evaluated along the twelve axes. Figure reproduced with permission from Miracle *et al.*<sup>597</sup> Copyright 2017, Elsevier.

expansion, and fatigue. While not every structural application incorporates all of these objectives, depending on the application, many of the objectives must be met simultaneously for commercial viability.

While fully autonomous setups that iteratively suggest new experiments with automated synthesis, processing, characterization, testing, and sample transfer are rare in materials discovery for structural applications, there has been a large amount of progress towards automating complex and difficult tasks with inorganic solid-state materials. For example, Vecchio *et al.* demonstrates the use of the FormAlloy tool to automatically mix powder precursors and additively manufacture alloys.<sup>598</sup> While characterization and property testing requires manual sample transfer and intervention, the authors developed a unique platform for high-throughput characterization using a turntable-style sample holder with multiple sample positions (Figure 38). The unique benefit of this design is that they integrated it with a variety of characterization tools, which both drastically reduces the amount of sample prep time and simultaneously makes it much easier to correlate multimodal data between instruments with sample batches. For example, scanning electron microscopy (SEM) and electron backscattered diffraction (EBSD) data are spatially correlated with nanoindentation hardness measurements. While no iterative optimization took place, Vecchio *et al.* used both thermodynamics-based CALPHAD simulations and ML methods for property prediction, and set the stage for an autonomous and robust SDL for alloy discovery.

Aside from macroscale structural applications, Kusne *et al.* demonstrated, in a data-driven and automated synchrotron characterization setup, a large reduction in the number of required experiments to identify optimal epitaxial nanocomposite phase-change memory materials, important for non-volatile data storage applications.<sup>516</sup> These materials leverage heat-induced and reversible transitions between amorphous and crystalline phases to mimic the “0” and “1” binary states of conventional transistors but with higher permanence. We note that the materials search was restricted to a single ternary alloy system (Ge-Sb-Te) and the processing parameters were fixed. Additionally, the sequential characterization experiments were carried out on a pool of several



**Figure 38.** HT-READ incorporates physics- and thermodynamics-based models with automated synthesis and high-throughput characterization methods for virtual alloy screening. Figure reproduced with permission from Vecchio *et al.*<sup>598</sup> Copyright 2021, Elsevier.

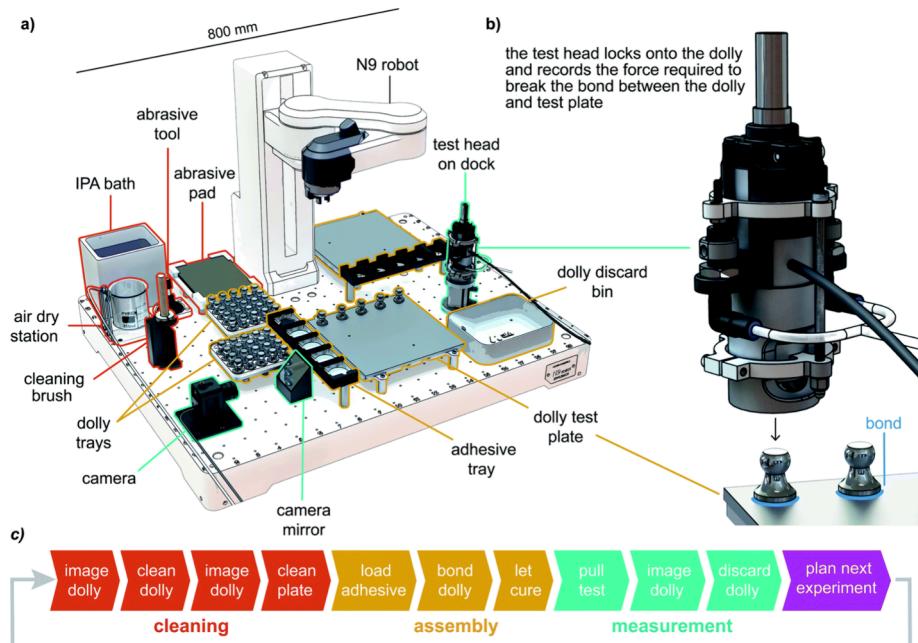
hundred pre-synthesized samples on a single silicon wafer via combinatorial sputtering; meaning that only one iteration of synthesis was performed. Their benchmarking results demonstrated that the incorporation of physics-based phase mapping information led to more efficient discovery relative to both random search and BO without phase mapping awareness.

In addition to mechanical and phase-change properties, automated methods have been used to search for corrosion-resistant alloys. DeCost *et al.* built an autonomous scanning droplet cell to accelerate the discovery of a novel Al-Ni-Ti alloy composition for corrosion resistance.<sup>599</sup> This SDL has automated serial electrodeposition with adjustable solution compositions and online processing characterization (i.e. optical camera and laser reflectance for assessing continuity, coloration, uniformity, and qualitative roughness of electrodeposits; a potentiostat for measured potential, and current; and a pH probe and thermometer for monitoring pH, and temperature). To close the loop in combining all the various process conditions and measured objectives, the authors adopt the active learning strategy using GP to predict and optimize for corrosion resistance (i.e., passivation current, passivation potential, and the slope of the passivation plateau). After several iterations, they successfully found Al-Ni-Ti alloy compositions that were near the Pareto frontier.

## 6.2. Concrete Formulations

If all of the artificial materials in the world were to be categorized and placed on a scale, concrete would be the heaviest. Concrete is the most prevalent human-made material and the second most consumed commodity after water

(approximately 30 billion tonnes per year<sup>600</sup>). Concrete is not a trending scientific topic, like one might expect for topics like superconductors, long-range EV batteries, or quantum computers; however, it is one of the most practical materials topics related to environmental sustainability. The energy-intensive nature of manufacturing concrete leads to a weight-share of approximately 8% of human-derived CO<sub>2</sub> emissions.<sup>600</sup> Unfortunately, there are only a few public examples of iterative and accelerated science methods applied to concrete formulations and processing (at least ones that move past property prediction based on classical ML models). While the reason for this may be complex, in addition to the “hype” factor mentioned previously, there are other practical factors that constrain against deviations from existing concrete technologies, such as the safety concerns of new formulations, or the lack of long-term test data. On a related note, perhaps much of the ML and robotics focus in the field has been concentrated on detection rather than discovery—extending the life of existing concrete rather than seeking to replace it, as may be indicated by the large fraction of detection-focused ML manuscripts in the literature.<sup>601–605</sup> Despite these considerations, one promising alternative to traditional concrete formulations involves swapping the energy-intensive “Portland cement” with eco-friendly cements based on alkali-activated binders or “geopolymers.” In an effort to validate and identify optimal data-driven routes for optimization of such concrete, Völker *et al.* used a set of 131 experimental data from the literature to conduct computational benchmarks, exploring the effect of algorithm choice and parallelization on the efficiency



**Figure 39.** A robotic platform with custom tooling for processing and testing adhesive specimens where batches of next experiments are guided via BO. (a) The automated platform with N9 arm capable of picking up and using various attachments. (b) The testing head used to characterize the force required to break the adhesive material. (c) The workflow of the SDL, note that adhesives were created and loaded manually. Figure reproduced with permission from Rooney *et al.*<sup>308</sup> Copyright 2022, Royal Society of Chemistry.

with which high compression strength materials can be identified.<sup>606</sup> Notably, this is one of the only studies that shows modern adaptive experimentation being used in the context of concrete optimization.

Concrete formulation and processing optimization also has a unique challenge relative to many other materials discovery tasks: it is highly affected by locally available materials. Concrete is dense and is used in large quantities for buildings and other structural applications at a relatively cheap cost-per-weight, so transporting it large distances is infeasible. What this means is that an optimal concrete in one part of the world does not translate directly to another part of the world. While the considerations for applying accelerated science tools to the field of concrete discovery are complex, the potential positive impacts are large. We anticipate a fully autonomous concrete formulation and processing optimization tool in the near future, which will require awareness, incorporation of AI and robotics into civil engineering repertoires, and a strong understanding of the limitations and opportunities within the field.

### 6.3. Non-Alloy Additive Manufacturing

The use of AI and robotics in additive manufacturing settings holds great promise: in terms of synthesis and optical characterization, there is a relatively low barrier to automation; the equipment typically has an API and native programming abilities; the equipment is adaptable and available, and the processing parameters are straightforward.

For example, Erps *et al.* used multi-objective BO to simultaneously maximize toughness, compression modulus, and maximum compression strength as a function of six primary formulations to form a composite formulation for photocurable resins. All processing parameters were held fixed.<sup>607</sup> We note that this is a semi-automated platform which requires human intervention for transfer of materials between each step in the sample fabrication pipeline while all of the

individual steps of dispensing, mixing, 3D printing, post-processing, and testing are completed individually without human intervention. Such formulation systems can be adapted to other domains such as surfactants, cosmetics, foods, and paints.

Brown and co-workers have also demonstrated SDL studies of mechanical properties of 3D printed structural materials. While the works predominantly focus on the effects of macroscopic structural designs rather than the intrinsic material property or formulation, the studies demonstrate the potential of autonomous BO for searching various design spaces in additive manufacturing. Gongora *et al.* introduced Bayesian Experimental Autonomous Researcher (BEAR), combining simulations and experimental observations to optimize the twist angle and struts of a 3D printed structure for optimal toughness.<sup>608</sup> The authors use a GP regression model to explore the search space, while a novel automated platform provides high-throughput printing and testing of manufactured parts. Simulations through finite element analysis were shown to strengthen the Bayesian prior of BEAR, increasing the speed of optimization in the experiments.<sup>609</sup> In their most recent work, Snapp *et al.*<sup>610</sup> introduce new mechanical designs (i.e., a generalized cylindrical shell with 8 variable parameters) and filament combinations to improve the material's mechanical properties. Although their approach employs the SDL framework to accelerate toward an optimal mechanical design, there is an absence of chemical synthesis or materials design.

While getting the formulation right and automating the exploration of a wide variety of combinations of starting materials is difficult, optimizing processing parameters within a certain material family is much more attainable. One example of this processing parameter optimization is a low-cost example that uses BO with a modified 3D printer to optimize the print characteristics of a silicone material.<sup>113</sup>

#### 6.4. Adhesives

Similar to many applications in this review, adhesives are a classic formulation optimization problem. The relative fractional share of constituents and the combinations of precursors can have a large effect on the bond strength exhibited by the adhesive. In this vein, Rooney *et al.* developed a semi-automated SDL platform for adhesive synthesis and characterization using a SCARA-type N9 (North Robotics) workstation, and substrates ("dollies") to be coated with the adhesive and used with a custom shear-stress "pull-off" tester (Figure 39).<sup>308</sup> Adhesive coating, preparation, and testing were all performed automatically via the robotic platform, while mixing of the adhesive formulations was carried out manually. This system used BO with a GP surrogate to maximize the bond strength as a function of resin to hardener ratio in a two-part epoxy system over four iterations of 5-sample batches (20 experiments). Notably from an algorithm standpoint, the BO algorithm from the Ax package uses a sophisticated batch-aware acquisition function.

#### 6.5. Outlook and Perspectives

Researchers in structural materials are quickly adopting ML techniques to accelerate the discovery of novel materials, but many have yet to adopt robotics technology to automate the synthesis and characterization of such materials. Specifically in structural materials, hardware automation remains a challenge because of the inherent difficulties in solid-dispensing, extreme temperatures, and complex testing instrumentation. By reimagining our approach to synthesis and characterization, shifting from conventional human-oriented instrumentation to a hardware-centric perspective, we can harness the power of robotics to automate intricate tasks in diverse ways.

### 7. OPTOELECTRONICS

Optoelectronic materials play a pivotal role in modern technological advancements by enabling the manipulation and control of light-matter interactions. These materials are integral to a wide array of applications spanning from telecommunications and displays to solar cells and medical devices. The significance of optoelectronic materials lies in their ability to absorb, emit, and modulate light, thereby facilitating the conversion of electrical signals into optical signals and vice versa. This functionality underpins the development of high-speed data communication systems, energy-efficient lighting, and sensitive imaging devices, among other innovations.

Designing new optoelectronic materials and making them technologically useful requires a comprehensive understanding of the complex relationship between their composition, structure, processing, and physical properties such as electronic structure and optical characteristics. Experimentally characterizing these attributes often requires sophisticated techniques. The synthesis and fabrication of these materials with the desired properties can be difficult and resource-intensive; optoelectronic materials are often part of devices in crystallized form or thin films, with blends and mixes of other optoelectronic materials, which can have dramatic effects on the performance of the device. In the context of SDLs, this introduces additional parameters that require further optimization, and will be dependent on the material-specific properties such as stability, scalability, and cost.

Due to the complexity of the design-make-test-analyze cycle for optoelectronics, efforts have been made to combine ML

and DoE without an automated laboratory, aiming only to more efficiently search the design space of possible compounds and devices. Cao *et al.* used SVMs to optimize only device processing parameters for an organic photovoltaic device active layer composed of a PCDTBT donor and PC<sub>71</sub>BM acceptor.<sup>611</sup> Subsequently, Kirkey *et al.* extended this method to study multiple acceptor compounds.<sup>612</sup> Conversely, Wu *et al.* leveraged high-throughput synthesis and characterization to speed up discovery of organic semiconductor laser materials without ML-guided experiment selection by exploring molecules similar to the prototypical BSBCz.<sup>613–615</sup>

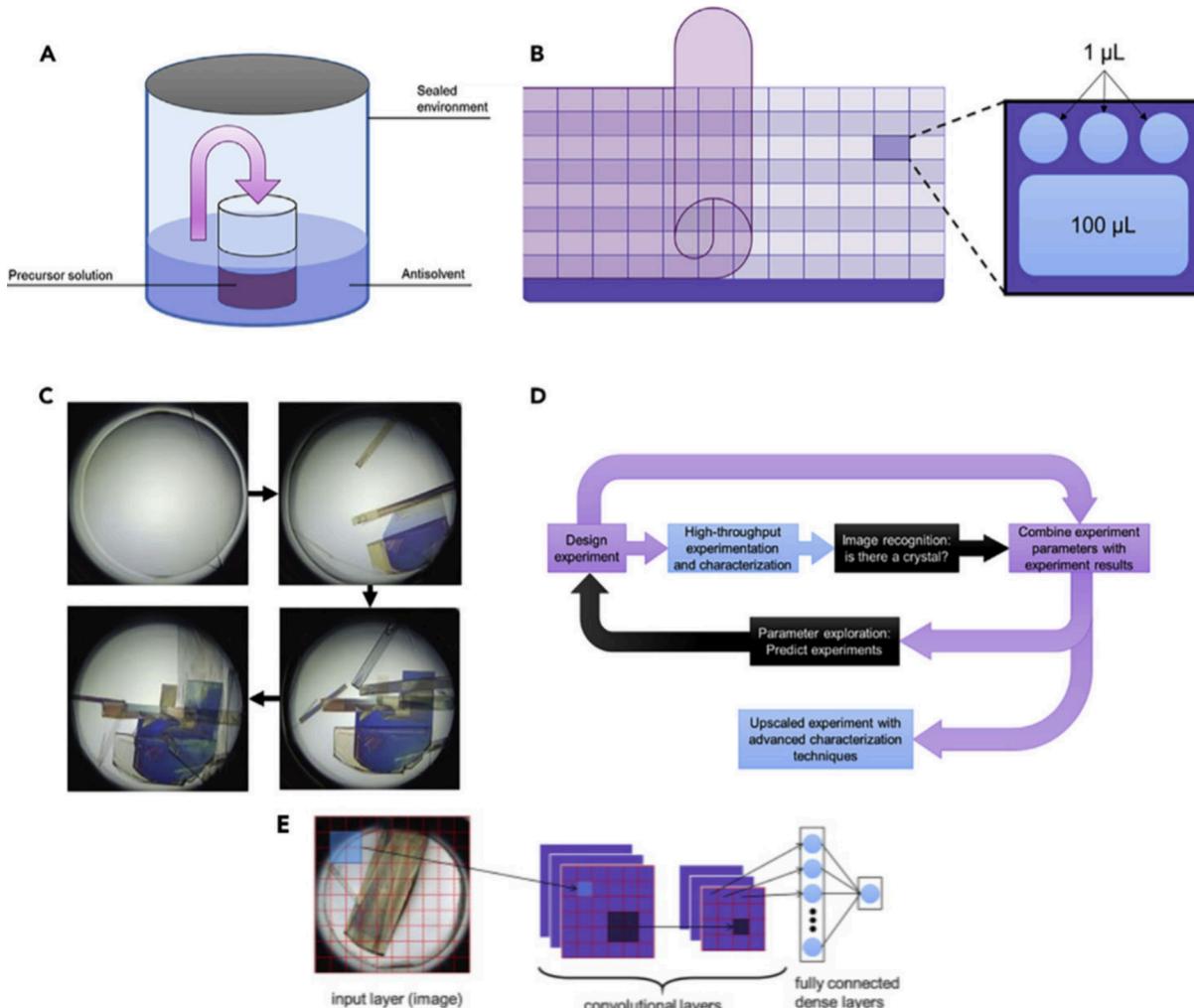
Although these efforts did not involve SDLs, they served as initial steps toward the development of SDLs for optoelectronic materials and devices. In the following sections, we discuss SDLs for optoelectronics based on the sophistication of the proxy measurements: from solution-state and single crystals to full device fabrication and testing.

#### 7.1. In Solution and Crystals

Proxy measurements of materials in solution and as single crystals offer two primary advantages. Both proxies are simple enough that they can be used to gain fundamental insights into the materials at the atomic level, especially in conjunction with quantum chemical calculations. Additionally, they are relatively amenable to high-throughput experiments.<sup>616</sup> In the case of solution-based testing there are many options for highly parallel and high-throughput experiments to study the relationship between composition and important optoelectronic properties using simple analytical techniques such as optical spectroscopy.

**7.1.1. Perovskites.** An example of a commonly studied crystal optoelectronic material is the perovskites. In an SDL developed by Higgins *et al.*, the authors utilized a well plate to automatically and parallelly synthesize 96 multi-component perovskites and analyze the photoluminescence of the compounds in a high-throughput manner.<sup>617</sup> Four perovskites systems were studied with varying compositions with the goal of maximizing the photoluminescence of the dissolved crystals. The authors also added a temporal axis to the data, looking at the spectra as a function of time, in order to capture the stability of the compounds. The results of the photoluminescence were decomposed using non-negative matrix factorization (NMF) into spectral information dependent only on the material composition, which was then fed into a GP regression model. The GP model interpolated between the low amounts of data to give a predictive map of the best compositions for the relevant perovskite systems. While the authors did not perform a second iteration based on the predictions of the GP model, the model prediction and uncertainty could have been used in a BO scheme to suggest the next round of experiments.

There have also been SDL studies of only the crystallization process of perovskites. Crystallization is not only dependent on the composition of the perovskite, but also the process conditions, which will affect the structure and performance of the optoelectronic device. Li *et al.* studied the perovskite crystallization using a robotics-accelerated micropipetting system.<sup>618</sup> The solution was gradually cooled in an inverse temperature crystallization (ITC) reaction,<sup>619</sup> and the resulting crystals were visually categorized into 4 levels of crystal quality. The ITC reaction parameters were the same for all reactions, however the concentrations of the inorganic, organic, and formic acid precursors were varied. All reactions contained the



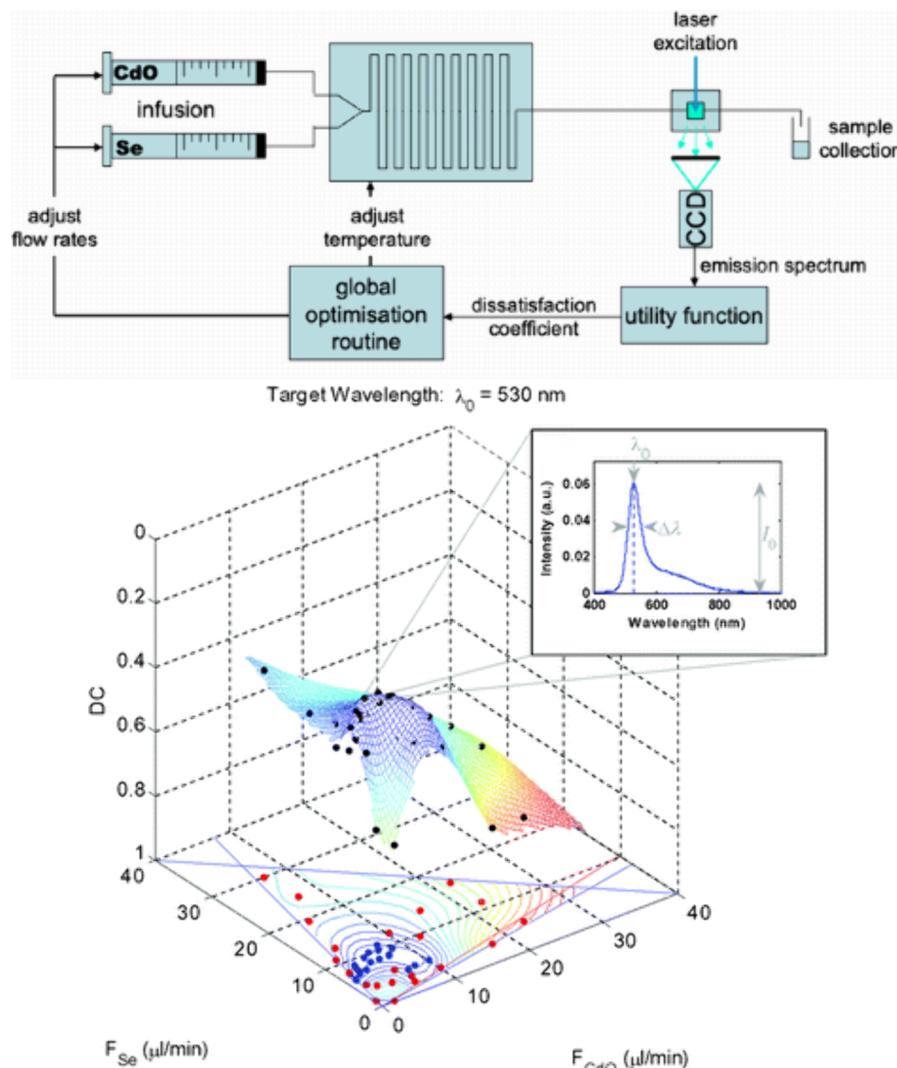
**Figure 40.** Schematic of SDL from Sargent and co-workers, studying the crystallization of perovskites. (A) A diagram of experimental setup used for antisolvent vapor-assisted crystallization. (B) A tray with 96 containers, each containing 3 drops of precursor solution, and a well of antisolvent, which were then sealed for the crystallization. (C) Microscopy images over the course of 3 hours of the crystallization. (D) The SDL workflow. (E) A schematic ML method used. The convolution layers perform operations on spatially connected pixels of the images, finally being flattened and put through a fully connected layer to output a classification. Figure adapted with permission from Kirman *et al.*<sup>620</sup> Copyright 2020, Elsevier.

lead (II) iodide, and one of 45 structurally diverse organoammonium cations. In total, 8172 reactions were performed, and the most novel structures were further studied in manual experiments, however, no ML was utilized during the crystallization experiment. Instead, the authors performed a retrospective study, and determined that a SVM model with a Pearson VII universal function kernel was most accurate in crystal quality prediction, using the reaction conditions and the organic and inorganic precursor chemical descriptors, however suffers when trying to generalize to different precursors or perform with small amounts of data. This study demonstrates the potential for a second round of high-throughput experiments based on the predictions of the model.

In another work studying the crystallization of perovskites, Kirman *et al.* incorporated a CNN and a kNN regressor model in their SDL for automated metal halide perovskites.<sup>620</sup> Repurposing a protein crystallization robot, the authors were able to study 96 experiments with different concentrations of precursors in parallel (Figure 40). The prepared solutions are then sealed in a chamber with an antisolvent to induce the crystallization.<sup>621</sup> Initially, studies focused on phenethylammonium lead bromide,  $(C_8H_{12}N)_2PbBr_4$  (PEAPbBr) crystalliza-

tion, with 7000 images classified from an initial run as either no/bad crystals, or good crystals. This dataset was used to train a CNN, achieving 95% accuracy in crystal detection. Kirman *et al.* then applied the workflow to a new perovskite system, with 3-picolylammonium (3-PLA) as the ligand, and different lead halides. Additionally, a kNN model was trained to map the experimental conditions, as well as DFT-based descriptors of the precursors, to the success of the crystallization experiment. Another round of experiments was performed based on the predicted reaction parameters most likely to yield crystallization, and the authors were able to increase their rate of crystallization success from ~1% to ~10%. While the SDL still required human intervention to perform the crystallization, Kirman *et al.* demonstrated a fully closed-loop platform for perovskite crystallization, using AI methods to learn from an experiment to suggest the subsequent experiments, and improving upon their original results.

**7.1.2. Nanoparticles.** Nanoparticles (NPs) are another type of commonly studied optoelectronic material. By controlling their size, shape, structure, and composition, the absorption/emission intensity and wavelength, and their self-assembly behavior, can be controlled. A variety of methods

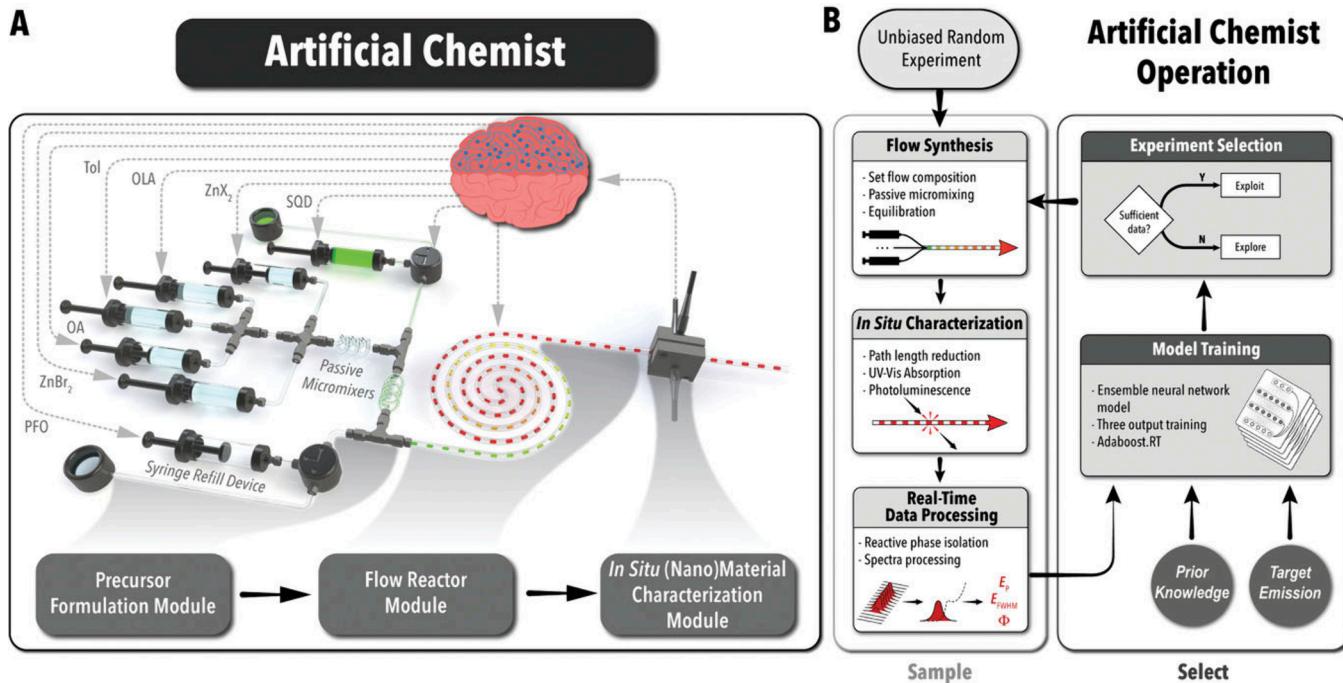


**Figure 41.** Schematic of the microfluidics platform, and the results of the SDL optimization from deMello and co-workers. (top) A schematic of the microfluidic SDL, with the channels etched onto a glass substrate. Precursors flowed through the chip using syringes, and the resulting NPs were characterized by emission spectra. (bottom) The dissatisfaction coefficient was calculated from the characterization of the NPs, and was minimized using SNOBFIT. The minimum in the dissatisfaction coefficient is shown as a function of experimental parameters (flow rate of precursors). Figure adapted with permission from Krishnadasan *et al.*<sup>35</sup> Copyright 2001, Royal Society of Chemistry.

have been developed to control NP synthesis, many of which are suitable for automated synthesis platforms. One of the first examples of SDL for NP optoelectronics is from Krishnadasan *et al.* in 2007.<sup>35</sup> The authors used a microfluidics platform to control the injection of CdO and Se solutions into a reactor. The goal of the SDL was to optimize the emission of CdSe NP to some target wavelength, and maximize the intensity, combined together in a dissatisfaction utility function. The goal would be to minimize the dissatisfaction as a function of the controllable parameters using the SNOBFIT algorithm,<sup>622</sup> which proceeds as follows: (1) the search space is discretized around sampled data points, (2) quadratic models are fit around each point, (3) the next point to sample is the model minima, (4) the model is refit using new experimental data. Krishnadasan *et al.* demonstrated a closed-loop SDL capable of minimizing the dissatisfaction function in a 3D search space, varying the flow rates of CdO, Se, and the reaction temperature. The optimization trace is seen in Figure 41, with the best experimental conditions identified at evaluation 71. In a related study from the same group, Li *et al.* studied the

ligand-mediated synthesis of nanocrystals in cesium lead bromide NPs.<sup>623</sup> Linear ligands of base:acid precursors are used to define the size and shape of the colloidal NPs. Rather than utilizing an automated experimental design, the authors systematically varied the reaction temperature and the ratio of base:acid ligands, and the effect on the photoluminescence spectra.

Salley *et al.* developed a liquid-handling robotics platform for controlled seed-mediated synthesis of Au nanoparticles.<sup>624</sup> Three different experiments were performed, manipulating not only the size, but also the shape of the synthesized nanoparticles. The fitness function maximized is dependent on the absorbance at particular wavelengths in the UV-Vis spectra. In order to close the loop, a GA was used to select for the next experimental conditions. In the first experiment, the authors maximized the absorbance of spherical AuNPs at 553 nm, corresponding to a NP diameter of  $\sim 80 \text{ nm}$ , by manipulating the volume of dispensed precursors. Then they moved onto Au nanorods; starting from a randomly selected set of parameters, the platform synthesized 10 generations of



**Figure 42.** Schematic of the artificial chemist developed by Abolhasani and co-workers. Variations of this platform were used in future work from the group. (A) Continuous flow systems powered by a series of syringe pumps. A custom *in situ* spectrometer system was used to monitor the absorption and emission spectra. (B) A schematic of the workflow of the artificial chemist. Starting from random experiments, the flow synthesis and characterization were followed by a boosted ensemble of neural networks. Figure reproduced with permission from Epps *et al.*<sup>630</sup> Copyright 2020, John Wiley and Sons.

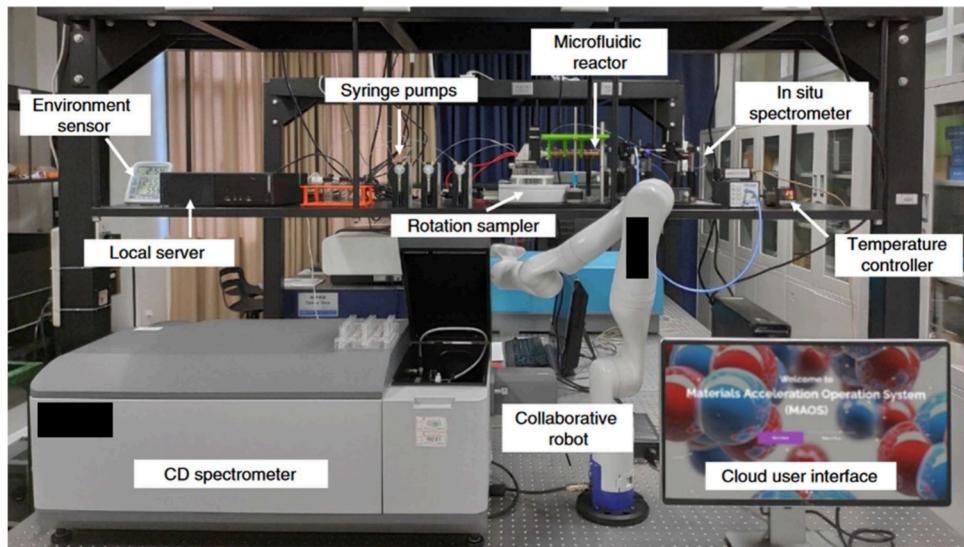
nanorods, with 15 experiments each, ultimately maximizing the fitness function, and converging close to the original parameters proposed in the literature.<sup>625</sup> In the final experiment, the nanorods from the second experiment were used as the crystallization seed, generating NPs of octahedral shape, a shape with unknown optimal synthesis outcomes. The authors later extended their study to include automated exploration and optimization of multiple levels of seed-mediated AuNP synthesis.<sup>626</sup> Jiang *et al.* extended the previously described automated platform with in-line optical spectroscopy in UV-Vis-IR, allowing for high-throughput characterization. The various hierarchical levels of seed-mediated synthesis were also automated and controlled via directed graphs. The results are fed into a custom GA based exploration algorithm based on the Multi-dimensional Archive of Phenotypic Elites (MAP-Elites) algorithm, as well as the sparsity of the synthesis conditions.<sup>627</sup>

Tao *et al.* further extended the use of microfluidics in the SDL development of metal nanoparticles by incorporating a machine learned surrogate model.<sup>628</sup> Multiple rounds of closed-loop experiments were performed based on suggestions from the surrogate models trained on the results of previous experiments. The authors studied the formation of gold nanoparticles (AuNPs) under various concentrations of aqueous precursor compounds, and different reaction times. Solution concentrations were manually loaded into the platform based on the suggestions of the algorithm. Uniquely, the SDL optimized for multiple objectives calculated from the UV-Vis absorption spectra, such as the position, full-width half-maximum, and intensity of the peak, through the use of a novel BO algorithm which allows for hierarchical multi-objective optimization.<sup>280,446</sup> The authors successfully demonstrate optimization of NP spectral properties for both large and

small AuNP, using kernel density regression, with the kernels estimated via a BNNs.

Most recently, Low *et al.* presented an SDL that optimized the synthesis of silver NPs (AgNPs) using a multi-objective optimization algorithm dubbed Evolution-Guided Bayesian Optimization (EGBO).<sup>629</sup> They develop a fully automated SDL for seed-mediated AgNP synthesis, integrating microfluidics, inline hyperspectral imaging, and closed-loop optimization. The optimization goals were to target a desired spectral signature for specific optoelectronic applications, maximize the reaction rate for high throughput, and minimize costly seed particle usage. The various objectives were modeled by a GP surrogate. The EGBO algorithm then combines a batched BO with qNEHVI acquisition function with an evolutionary algorithm, leveraging selection pressure to balance exploration and exploitation toward the Pareto front. Applying EGBO to the nanoparticle synthesis and various synthetic test problems, the authors demonstrate improved performance over state-of-the-art methods in terms of hypervolume convergence, uniform coverage of the Pareto front, and constraint handling. They also investigate pre-repair and post-repair strategies for handling input and output constraints, underscoring the importance of careful constraint treatment in self-driving laboratories.

Colloidal synthesis of NPs has also been applied to inorganic lead halide perovskite NPs. Epps *et al.* developed a high-throughput microfluidic reactor platform with an *in situ* characterization module.<sup>630</sup> Starting with  $\text{CsPbBr}_3$  quantum dots, the bandgap of the NPs was tuned via halide exchange reactions, via introductions of zinc halide precursors.<sup>631</sup> The various precursors were varied to optimize for a joint fitness value comprised of the PLQY, emission linewidth, and the emission energy, which is related to the bandgap. The colloidal



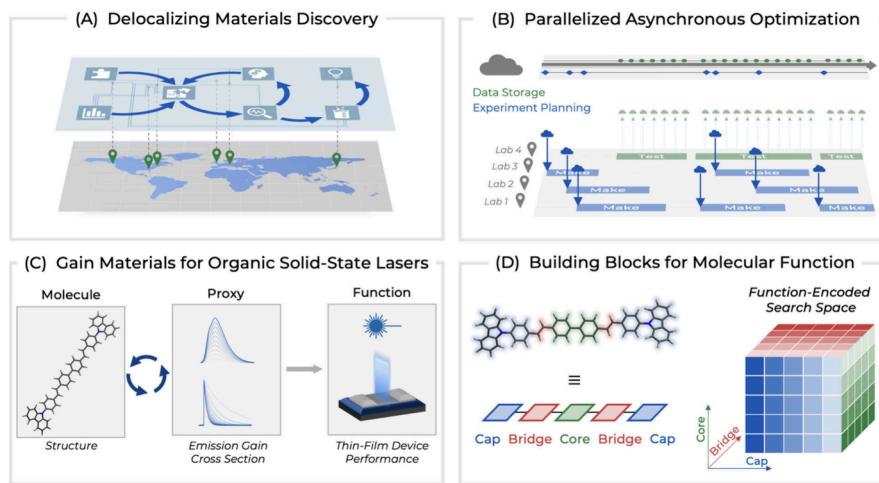
**Figure 43.** MAOSIC platform with automated synthesis and characterization connected via a cloud platform. Synthesis was performed with a microfluidic reactor, followed by an *in situ* spectrometer. The samples were then transferred by a robotic arm to the CD spectrometer. All operation, data management, and optimization algorithms were done remotely and collaboratively in the cloud. Figure reproduced with permission from Li *et al.*<sup>119</sup> Copyright 2020, Springer Nature.

lead halide perovskite NPs were flowed through a custom in-line module capable of absorption and photoluminescence UV-Vis spectrometry (Figure 42). The results were fed into a boosted ensemble of neural networks and the next synthesis conditions were selected via BO. The authors compared the optimization with other commonly used methods, e.g., SNOBFIT, and CMA-ES, and found superior performance with the neural networks. Additional performance gain was observed after pre-training the networks with supplemental experimental data. In related work from the research group, Abdel-Latif *et al.* modified the aforementioned platform to include multi-phase reactions (i.e., gas-liquid), allowing for in-series synthesis of  $\text{CsPbBr}_3$  quantum dots, and expanding the synthesis parameter space for lead halide perovskite NPs.<sup>632</sup> To improvement the ensemble of neural networks, an initial round of 200 experiments were performed to pretrain the networks to predict the FWHM and energy of the photoluminescence spectra. Epps *et al.* further studied optimization of AI guided experimental design agent used in their SDL through a simulated experimentation platform.<sup>633</sup> Using 1000 experimental data points on metal halide perovskite NPs, a surrogate model comprised of a series of GP models served as the HTE platform, and the model, fitness functions, and acquisition functions were tested and compared.

Li *et al.* developed the MAOSIC (materials acceleration operating system in cloud) in order to look at chiral perovskite nanocrystals.<sup>119</sup> This class of optoelectronics have shown strong optical activity, and have possible applications in spintronics, sensing, or optical communications.<sup>634</sup> However, the controlling the chirality of such semiconductors is non-trivial. Li *et al.* utilized a microfluidics SDL with a cloud server for data storage and communication. A robotic arm is used for automated transfer of the synthesized NPs into a spectrometer, returning data on the absorbance and circular dichroism (CD) spectra (Figure 43). The SNOBFIT algorithm was used for experimental design, varying the temperature and the precursor concentrations. Synthesized NPs with strong CD intensities were extracted for further analysis via XRD and transmission electron microscopy (TEM).

Vikram *et al.* applied the automated microfluidics approach to optimization of indium phosphide nanocrystals.<sup>635</sup> In order to understand the kinetics of the growth and nucleation of the InP NPs, the SDL had a growth stage that spatially separates the various stages of NP synthesis for sampling and characterization. The experimental design of the SDL uses an ensemble of 25 neural networks for uncertainty estimation, predicting the polydispersity, and bandgap of the NPs from the synthesis conditions. And while the kinetics were not involved in the SDL optimization, the additional data on the stages of InP growth were analyzed afterwards.

Additional work on lead halide perovskite NPs was conducted by Bateni *et al.*,<sup>636</sup> doping the nanocrystals with cations in a flow reactor similar to those discussed prior (Figure 42).<sup>630,632</sup> Similar to the halide exchange mechanism, cation doping reactions were performed in the microfluidics platform through the introduction of manganese acetate, dissolved in 1-octadecene and activated with oleic acid. Spectroscopy data from the synthesized NPs were then fed into an ensemble of 100 neural networks, and the next experiments were suggested based on a greedy BO strategy. The closed-loop optimization campaigns optimized for the peak energy, and the Mn:exciton emission peak area from the photoluminescence spectra, producing on-demand bandgaps and doping levels of the lead halide perovskite NPs. In a recent study from the same authors, Bateni *et al.* demonstrated Smart Dope, an SDL for multi-cation doping of the lead halide perovskite NP system.<sup>637</sup>  $\text{CsPbCl}_3$  quantum dots were doped with both Mn and Yb cations; the successful doping was confirmed through off-line characterization. Varying the reaction temperature and the precursor flow rates, the optical features in the absorbance spectra, measured using *in situ* spectrometry, were optimized as proxies for the reaction yield, and Mn and Yb emission. Experimental design was done using BO with a similar ensemble of neural networks, first pretrained on an unbiased dataset of 150 NP synthesis experiments. The optimized Mn-Yb doped NP resulted in an impressive PLQY of 158%.



**Figure 44.** Schematic describing the closed-loop optimization campaign for discovery of organic solid-state lasers. (A) Multiple laboratories from across the globe ran asynchronous experiments, with experimental planning and results coordinated through an online server. (B) The parallel asynchronous optimization visualized as a timeline, demonstrating how 4 different labs can coordinate experiments. (C) The organic molecules were created and the lasing performances were approximated using the emission gain factor. After closed-loop automated discovery, the best molecules were used to test thin-film device performance. (D) The chemical space was defined through the use of cap, bridge, and core fragments, linked together via Suzuki-Miyaura coupling reactions. Figure reproduced with permission from Strieth-Kalthoff *et al.*<sup>641</sup> Copyright 2024, American Association for the Advancement of Science.

Zhao *et al.* studied colloidal perovskite NPs in a high-throughput platform consisting of a robotic arm, and a series of modules for pipetting, and UV-Vis spectroscopic analysis.<sup>638</sup> While the authors still use liquid precursors, the robot is capable of selecting solutions based on the suggestions of a ML algorithm without human intervention, and has the potential to perform more complex chemical tasks. The authors considered two different systems, AuNPs, and lead-free double-perovskite NPs ( $\text{Cs}_2\text{AgIn}_{1-x}\text{Bi}_x\text{Cl}_6$ ). To start, a literature search was performed to determine the best starting concentrations for AuNP synthesis, and the best surfactants and solvents to use for perovskite NP synthesis. Based on these results, a series of NPs were synthesized while systematically varying the experimental parameters, generating a database of absorption and photoluminescence data for AuNPs and double-perovskite NPs, respectively, along with data on the aspect ratio of the NPs from TEM and SEM images. The resulting datasets were then used to train a sure independence screening and sparsifying operator (SISSO) model, which identifies correlations between the target and compressed input descriptors.<sup>639</sup> Based on the prediction of the models, the authors ran an additional iteration experiment, varying the concentrations and volumes of precursors to verify the predictions of the model. For AuNPs, the aspect ratios were measured, and for the double-perovskite NPs, the sizes of the crystals were measured; the created NPs matched the predictions provided by the SISSO model. However, no additional model training with the new results were performed, and no additional iterations were done.

With further development of ML algorithms, more sophisticated methods of optimization were studied in the context of optimizing reaction conditions. Deep RL utilizes a neural network agent to decide the next experiments based on some policy. This policy is refined with each experiment, as the agent receives feedback from the environment, in this case the experimental result, in the form of rewards or punishments. Zhou *et al.* applied this optimization algorithm to finding the optimal conditions for organic reactions in a microdroplet

reactor, as discussed in a previous section.<sup>336</sup> The agent is a modified long short-term memory network (LSTM) capable of recursively learning from time-series data, such as data acquired over each iteration of experimentation. To overcome overfitting in the low-data regime, the authors pretrained the network on simulated data. Not only was the pretrained neural network based optimizer capable of optimizing the yield of the reactions, the model was able to successfully optimize the SDL synthesis of silver NPs for maximal absorbance at a particular wavelength.

In a more recent study, Volk *et al.* presented AlphaFlow, an RL-driven SDL capable of optimizing CdSe/CdS core-shell NPs with a modular microfluidics platform, optimizing the optoelectronic material over 40 experimental parameters.<sup>640</sup> Experimental planning was done using an LSTM agent over 20 steps, with a belief model comprised of an ensemble neural network regressor and a decision tree classifier. The regressor maps the action-state pair to the corresponding reward (based on spectral data), and the classifier determines if the action-state pair is viable; both models are retrained over each iteration. The authors demonstrate AlphaFlow's capability to optimize the sequence of injected precursors, and the volume and reaction time at each iteration.

**7.1.3. Molecules in Solution.** In addition to nanoparticles and crystalline materials, optoelectronic molecules are often the precursor to forming thin films and devices. Drawing from a long history of organic chemistry, small organic molecules can be formed from a myriad of organic reactions with careful control of initial organic fragments, much like the precursor solutions in synthesis of nanoparticles. While molecules in solution do not behave exactly the same as when in thin film form or in devices, they are more easily characterized and serve as a proxy to more complex morphologies of optoelectronics.

In 2023, Koscher *et al.*<sup>642</sup> presented an SDL for designing dye molecules integrated with computer-aided synthesis planning, first exploring unknown regions through synthesizing diverse examples to ground the property models, then exploiting the trained models to realize top-performing

candidates. The platform leverages automated molecular generation using a graph-completion model trained on existing data to propose new candidate molecules. Viable synthetic routes for these candidates are identified through automated reaction pathway planning with ASKCOS (Autonomous Synthesis Knowledge Cloud Organized System).<sup>450,643</sup> Ensembles of message-passing GNNs are employed for automated property prediction, evaluating candidates for specific optoelectronic properties like absorption, lipophilicity, and photostability. Robotic arms, batch reactors, and an automated liquid handler are integrated for automated synthesis to execute the recommended multi-step reaction pathways and isolate products. Crucially, the property prediction models are continually retrained with new experimental data in a closed automation loop, improving their accuracy iteratively. This platform demonstrated both the exploration of unknown parts of chemical space, and the exploitation of important optoelectronic properties in dye-like molecules.

Strieth-Kalthoff *et al.* demonstrated the closed-loop discovery of organic laser molecules across three different SDL platforms, asynchronously.<sup>641</sup> The chemical space is defined through the combination of organic fragment building blocks into organic pentamer molecules (Figure 44), similar to previous work done by Wu *et al.*<sup>644</sup> in which the fragments are joined together through iterative Suzuki-Miyaura couplings. The synthesis was performed through a generalizable two-step one-pot protocol, handled by automated experimental platforms. Absorption and emission spectra were recorded for the in-solution molecules, from which the lasing performance is estimated using the spectral gain factor.<sup>645</sup> Results were then uploaded to a database for coordination with the other laboratories. For decision-making, the authors used a GP model, with the molecules represented as embeddings extracted from a GNN. To overcome the issue of low amounts of experimental data, time-dependent density functional theory (TD-DFT) calculations were performed for the enumerated chemical space, and the descriptors generated from the calculations were used to train the GNN. In this transfer learning approach, the embeddings extracted from the GNN provided a stronger set of features for the GP regression task, which informed the subsequent experiments. Ultimately, this work discovered 21 novel gain materials with state-of-the-art lasing performance, of which the top three compounds were successfully tested in devices.

The work of Angello *et al.* demonstrated an SDL focused on discovering organic optoelectronics with good photostability, particularly for organic photovoltaic (OPV) applications.<sup>646</sup> Like in the previously discussed works on automated organic molecule synthesis,<sup>613,641</sup> the chemical space was predefined through the combination of molecular fragments through iterative Suzuki-Miyaura coupling reactions.<sup>431</sup> In this case, the fragments were acceptor and donor complexes connected by a bridge fragment; this is a common design for OPV applications, with light-induced charge separation encouraged by the difference in local electronic energy levels. The platform was capable of synthesis, purification, and structural characterization of the final compounds. The in-solution photostability was then approximated as a product of the spectral decay time, and the spectral overlap of the molecular absorbance and the solar irradiance spectra. In total, the closed-loop synthesis and characterization was repeated 5 times, with the experiments guided by Gryffin, a BNN-based BO algorithm capable of handling categorical parameters (such as the selected frag-

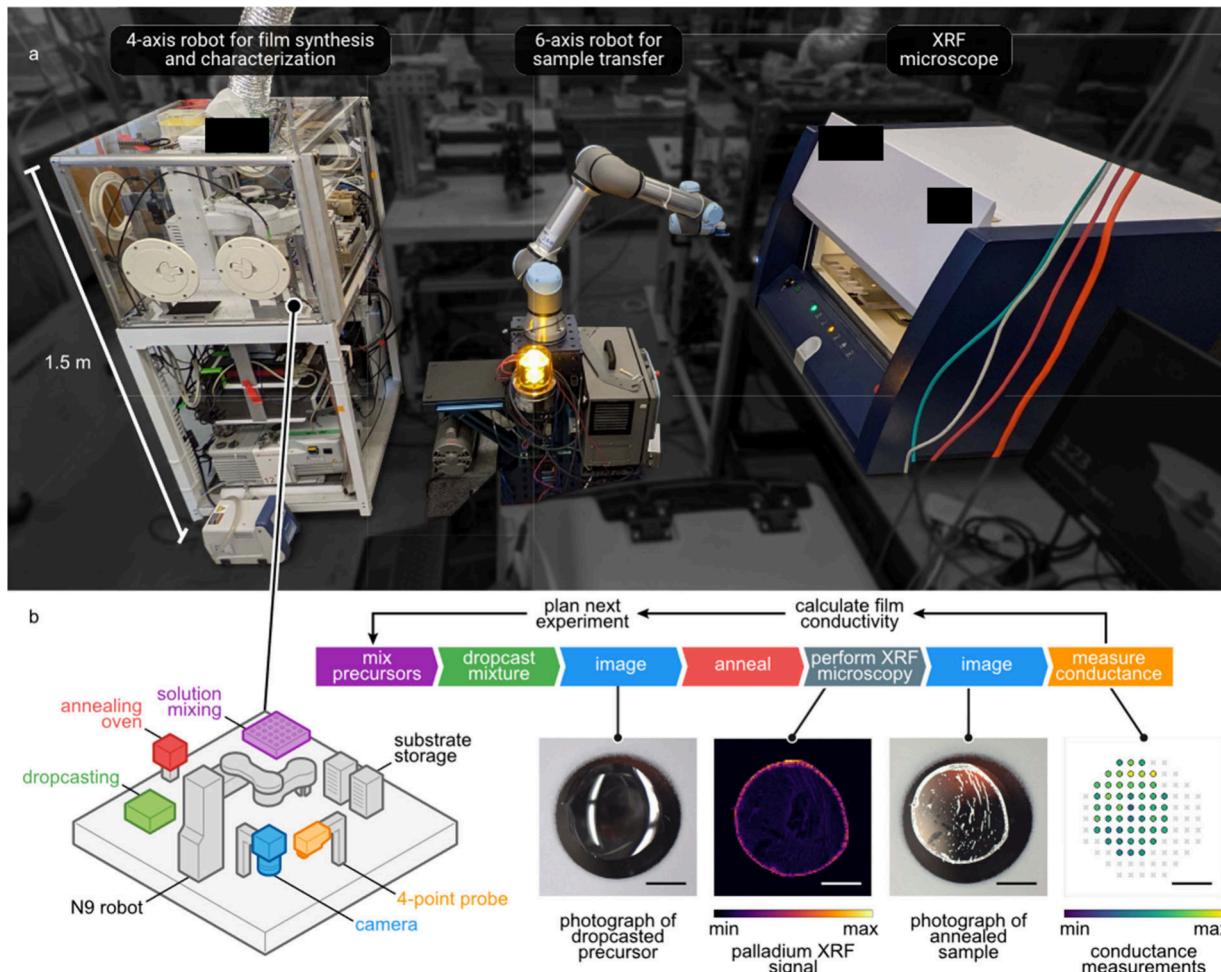
ments).<sup>446</sup> After the optimization, the authors further extended the work by using the experimental results from the SDL to perform physics-informed discovery. Whole molecule DFT calculations were performed on the entire space of possible molecules, and the extracted physicochemical descriptors were used to train SVMs. In this way, the experimental results of the SDL campaign were extended to the entire chemical space, and the predicted best and worst 7 molecules were synthesized to confirm the model predictions.

## 7.2. Thin Films

Thin films offer another useful proxy for optoelectronic devices without the need for fabricating an entire device. Optoelectronic devices (e.g., light-emitting diodes, LEDs, photo-detectors, and photovoltaics, PVs) are based on thin films (nm to  $\mu\text{m}$  in thickness) in order to simultaneously balance charge transport, and light absorption or emission. For example, in PV devices a thicker film maximizes the number of photons absorbed by the active layer. On the other hand, it is easier to extract charges from thinner films. Thin films are a useful proxy because they offer the ability to investigate larger length scales and the impact of processing and microstructure on important material properties such as photoluminescence, stability, or charge carrier mobility. Finally, thin films can be fabricated with relative ease through processes including spin-coating and thermal evaporation, which can be easily automated and integrated into an SDL. A detailed discussion of data-driven automated synthesis and characterization of thin film optoelectronics and electronic polymers is also provided in other perspective articles.<sup>647,648</sup>

A study on the SDL synthesis of colloidal and thin film chalcogenide quantum dots handily demonstrates the strong effect the thin film configuration has on measured performance. Chalcogenides are a class of compounds primarily composed of chalcogen elements, such as sulfur, or selenium, combined with various other elements, and are often used in semiconductor technology and materials science. Stroyuk *et al.* used a novel method of using aqueous precursor solutions of chalcogenide NPs to form multinary quantum dots with composition  $\text{Cu}_{1-x}\text{Ag}_x\text{InS}_y\text{Se}_{1-y}$  (CAISSe).<sup>649</sup> Previous work from the authors demonstrate that this method produces NPs of similar spectral properties as those directly formed from precursor metal complexes.<sup>650</sup> The use of aqueous forms of the precursors allows for automated synthesis using microfluidics platforms. By varying the precursor solutions, the produced NPs vary in Ag/Cu metallic composition ( $x$ ), and S/Se chalcogen composition ( $y$ ). By depositing and evaporating the colloidal mixture, solid thin films were formed on glass plates analyzed alongside the colloidal form. Several photoluminescent properties were measured in the experiments, such as photoluminescence lifetime, energy, and rate constant. The experiments were repeated for the quantum dots with a shell composed of ZnS. Due to the relatively small parameter space of the synthesis, the authors simply interpolated between the data points, creating a 2D map of the best CAISSe compositions. In particular, we can see the quenching effects due to the different forms of the NPs, with the photoluminescence lifetime significantly suppressed for the thin film. A second iteration was not performed, however the authors described possible future work involving a ML approach for more complex experimental parameter spaces.

MacLeod *et al.* demonstrate an SDL, named Ada, capable of optimizing thin film fabrication parameters.<sup>651</sup> With a robotic

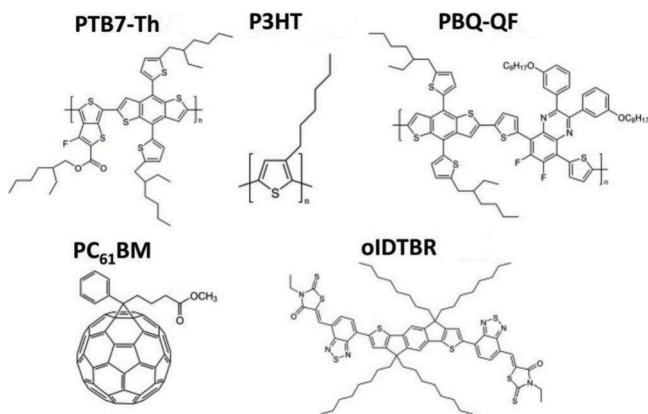


**Figure 45.** Configuration of Ada robot used by Berlinguette and co-workers to study the conductivity and annealing temperature of Pd films. A similar platform was used for the optimization of thin film fabrication. The modular nature of Ada allows for different instruments to be connected to the platform. (a) A robotic arm transferred the sample, after thin film deposition, from the 4-axis robot platform to the XRF microscope, and back to the 4-axis robot for characterization. (b) An expanded diagram of the 4-axis robot platform, as well as the various stages of operation of the automated synthesis and characterization. Figure reproduced with permission from MacLeod *et al.*<sup>653</sup> Copyright 2022, Springer Nature.

arm, Ada is able to move samples between various stations that are responsible for the stages of thin film fabrication. The entire process starts with measuring out appropriate amounts of precursor solution, spin-coating the glass substrate with the material, and then annealing for a specific amount of time. Characterization involves measuring the reflection and transmission spectra in UV-Vis-NIR, and measuring the conductance. The material studied were thin films of spiro-OMeTAD, an organic hole transport material used in perovskite solar cells, doped with cobalt (III). By varying the dopant concentration, and the annealing time, Ada maximized the electron hole mobility in the material, approximated as a ratio of the conductance and the absorbance. The results were fed into the BNN-based Phoenics BO algorithm.<sup>652</sup> Ada performs subsequent experiments based on suggestions from Phoenics, with the best parameters for global maximum hole mobility identified within 35 experiments. Exploiting the capabilities of Ada, the same group later demonstrated the autonomous optimization of synthesis parameters for the combustion synthesis of Pd thin films. Notably, MacLeod *et al.* extended Ada with an X-ray fluorescence (XRF) microscope for localizing the Pd in the annealed film before performing the conductance measurements (Figure 45).<sup>653</sup> Through the

variation of annealing temperature and combustion fuel composition, the authors were able to optimize the Pareto front between annealing temperature and conductivity.

Advances in thin film devices often include multinary films, blends of multiple optoelectronic materials that can affect the stability and performance of the devices. Langner *et al.* developed an SDL capable of fabricating up to 6048 organic polymer films a day, with the experimental planning done by the Phoenics algorithm.<sup>654,655</sup> Two quaternary systems with different compositions were explored (Figure 46). The first was composed of P3HT, PBQ-QF, PCBM, and oIDTBr, while the second replaced PBQ-QF with the more common PTB7-Th (i.e., P3HT, PTB7-Th, PCBM, and oIDTBr). The liquid handling robotics platform drop-casts the organic semiconductors onto glass substrates, with variation in the four components that make up the thin films. The films were then exposed to metal halide lamps; absorbance spectra taken before and after the exposure were used to determine the photostability of the quaternary thin film blends. The authors performed a grid-search method in addition to the BO experiments. They found that the full SDL with ML based experiment planning was able to find the blends that were as stable as the grid-search in 27 samples, on average, showing the



**Figure 46.** Compounds used in quaternary OPV film systems. Two mixes were studied by Langner *et al.*:<sup>654</sup> (1) P3HT, PBQ-QF, PCBM, and oIDTBr, and (2) P3HT, PTB7-Th, PCBM, and oIDTBr.

efficiency of combining HTE with data-driven experimental design.

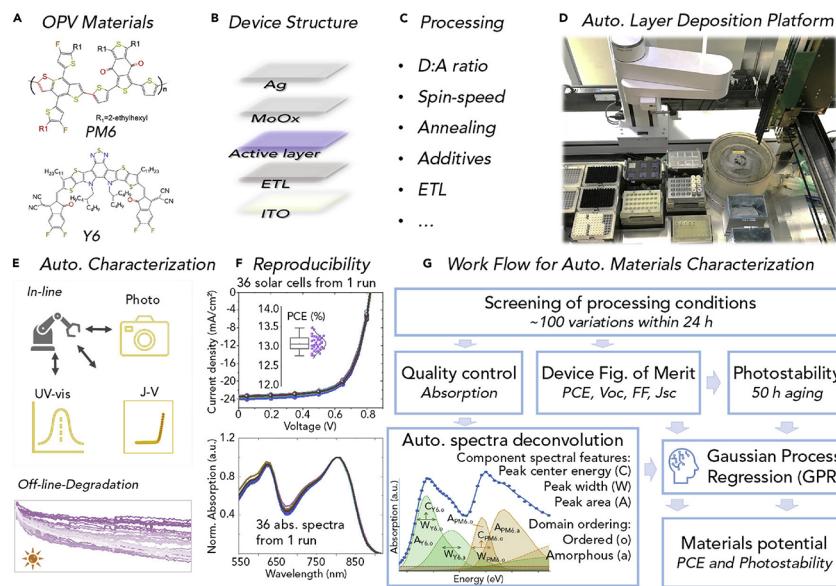
Sanchez *et al.*<sup>323</sup> proposes a workflow that combines structured GP models with custom physics-motivated mean functions and automated synthesis methods for the optimization of hybrid perovskite thin films with tunable bandgaps. The approach aims to accelerate optimization of properties like bandgap, photoluminescence, and absorption spectra by guiding experiments and reducing the required number of thin film preparations. By incorporating domain knowledge through custom mean functions, the structure GP converges more rapidly to the underlying ground truth compared to classical GP. The article demonstrates the application of this approach to study the bandgap evolution, photoluminescence peak shifts, and absorption spectra changes of  $\text{MA}_{1-x}\text{GA}_x\text{Pb}(\text{I}_{1-x}\text{Br}_x)_3$ , a mixed-halide perovskite system relevant for tandem solar cells and tunable light emission. Experimental character-

ization included measuring bandgaps from absorption onsets, tracking photoluminescence peak positions and intensities, and monitoring absorption spectral features over a range of compositions. The workflow's adaptability to automated synthesis platforms is highlighted, enabling the exploration of higher-dimensional compositional spaces. The authors suggest that this approach could facilitate the discovery and optimization of advanced materials for optoelectronic applications in self-driving laboratories.

### 7.3. Devices

SDLs that can optimize whole devices are incredibly complex because they require integrating and automating multiple workflows with many highly complex experimental systems. However, this also makes it possible to directly test the performance of a device and control aspects from composition all the way to device architecture, which are rarely optimized simultaneously. As a result of the high degree of complexity of SDLs that optimize optoelectronic devices, there are only a few groups in the world with the resources to conduct such research. Despite this limitation, significant progress has been attained in recent years.

Du *et al.* in 2021 developed AMANDA Line One, a robotic platform capable of automated multi-layer device fabrication and characterization.<sup>655</sup> Rather than exploring chemical space, the device parameters were varied, similar to the group's previous work in quaternary systems, described above. The platform used a robotic arm to move the sample between stations on AMANDA Line One for deposition of layers, thermal treatment, and optoelectronic measurements. The active compounds were PM6 and Y6, acting as donor (D) and acceptor (A) organic semiconductors, respectively. Various layers were deposited via spin-coating, with the PM6:Y6 active layer sandwiched between electron and hole conducting layers to form the device, shown in Figure 47. In total, 10 different processing parameters were varied in the device fabrication,



**Figure 47.** Summary of AMANDA Line One from the SDL work of Brabec and co-workers. (A) the OPV materials of interest, with PM6 as the donor, and Y6 as the acceptor. (B) The device structure, with the active layer containing the D:A compounds. (C) Some of the dimensions varied in the device fabrication. (D) Picture of the automated layer deposition platform used to fabricate the devices. (E) Schematic of the automated characterization methods, with additional off-line studies of the degradation. (F) Plots of the current density and absorption spectra measurements of multiple solar cells, demonstrating the reproducibility. (G) The workflow of AMANDA Line One. Figure reproduced with permission from Du *et al.*<sup>656</sup> Copyright 2021, Elsevier.

optimizing for four figures of merit: open current voltage ( $V_{oc}$ ), short circuit current ( $J_{sc}$ ), fill factor (FF), and the PCE. Due to the parallelized high-throughput nature of the platform, ~100 process conditions were systematically explored; without an experimental planning algorithm, the best fabrication parameters were identified within these experiments, producing a device with a PCE of ~14% in ambient conditions, aligning with the results from the literature.<sup>657</sup> The authors utilized the data from the automated platform to train a GP regression model, correlating spectral features obtained from the absorption spectra to the figures of merit, which gave some physical insight for the differences in performance. However, the platform is not yet integrated with an automated experiment planner, nor was a second iteration performed based on the feedback from the GP model predictions.

The following year, Liu *et al.* published their work on BO of perovskite solar cell devices fabrication parameters.<sup>658</sup> Motivated by the rapid spray plasma processing (RSPP)<sup>659</sup> technique for high-throughput fabrication of open-air perovskite cells, the authors aimed to find the best process parameters, varying the substrate temperature (°C), speed of the spray and plasma nozzles (cm/s), flow rate of precursor liquid (mL/min), gas flow rate into plasma nozzle (L/min), height of plasma nozzle (cm), and plasma duty cycle (ratio of time plasma receives DC power). A Gaussian process was trained on batches of experiments, with the next iterations informed by the upper-confidence bound (UCB) acquisition function, in a BO setting. In all, 5 rounds with 20 devices each were performed, with significant manual work involved in the manufacturing and testing of the devices. The authors were able to more quickly identify parameters for higher PCE devices using their ML-guided experiment planning than previous experiments led by human decisions.

While both works demonstrate significant advancements in the automation of the hardware and software for device optimization, there is still a lack of a fully closed-loop SDL for process optimization or material discovery for optoelectronic devices. However, the pieces to the puzzle show great potential, and an optoelectronic device SDL may only be a few years away.

#### 7.4. Outlook and Perspectives

SDLs for characterizing optoelectronic materials in solution and in crystals are the most developed because of the ease of carrying out the requisite experiments. In-solution experiments, in particular, are the simplest to set up and execute, relying only on pumps and well-established spectroscopic equipment. SDLs for thin film characterization or device fabrication, on the other hand, require substantially more complex systems such as robotic arms for transporting samples and vacuum chambers for depositing interlayers and contacts. At the same time, these highly complex SDLs possess the greatest potential for accelerating design and discovery in optoelectronics.

There are a number of challenges that stand in the way of fully realized SDLs for optoelectronic devices. Optimizing optoelectronic materials and devices requires taking into account numerous processes that occur at length scales from Å to μm. While this is challenging, larger quantities of data and improved ML or DL models can potentially overcome it. Next, the sheer number of possible variables—for example, selecting a material with appropriate properties, depositing a thin film and fabricating a device—can easily reach a design space that

becomes challenging for BO. At the same time, the cost of the experiments is simply untenable for optimization algorithms such as RL or evolutionary algorithms. While a simple solution might be to optimize devices based on a handful of accessible materials, simultaneously optimizing material synthesis and processing would be transformative. Synthesizing a small amount of a new material in order to tweak how it responds to processing conditions has the potential to significantly accelerate the development of optoelectronics. However, this remains a long term vision due to the many challenges enumerated in the Reaction Optimization section on top of those discussed here.

## 8. ENERGY STORAGE MATERIALS

Efficient energy storage systems are imperative to exploit the full potential of renewable energy sources, such as solar and wind, to reduce reliance on fossil fuels. The substantial amount of solar energy accessible on Earth could theoretically satisfy all human energy demands by powering photovoltaics and solar thermal systems. However, the intermittent nature of sunlight significantly limits the growth of solar power. For example, in California, peak solar power production during the day drives down the price of electricity, sometimes to negative territories, reducing the incentives for more installations.<sup>660</sup> Efficient and powerful energy storage technologies can ensure a stable power supply by capturing the excess energy during the day and releasing it at night, mitigating reliance on fossil fuels. While optimization of battery designs and devices are possible, here we focus on the SDL development of new materials for improved energy storage.

Electrochemical energy storage can be roughly divided into two broad categories based on the length of intended storage and speed of power delivery. Short duration energy storage technologies include capacitors and supercapacitors which charge and discharge within seconds to deliver high power. There are industrial efforts to automate the manufacturing of supercapacitors,<sup>661</sup> however, SDL-driven discovery of new chemistry and materials is limited.

Long-duration electrochemical energy storage is possible in batteries and redox flow batteries, as well as by converting the energy to liquid fuels such as alcohols or ammonia.<sup>662</sup> Batteries store energy in the form of reversible chemical reactions in static and enclosed cells. They are relatively compact and inexpensive, suitable for mobile applications such as consumer products and electric vehicles. To scale up, cells can be assembled into battery systems, which require a dedicated battery management system. Redox flow batteries (RFBs) also use reversible chemical reactions. However, the redox-active materials (RAMs) are solutions that can be stored in tanks and circulated through electrochemical reactors to generate power. RFBs can scale energy and power independently with higher tank volume and larger flow reactors, respectively, making them more scalable and cost-effective than batteries for grid applications, such as offsetting energy production and demand peaks.<sup>663</sup> H<sub>2</sub> gas and liquid fuels can store electrochemical energy off the grid, typically generated through non-reversible chemistry in flow reactors such as fuel cells and electrolyzers. Research in this area mainly focuses on cost-effective electrocatalysts that interconvert chemical energy and electricity.

The major challenge in developing the aforementioned energy storage technologies lies in the need to develop specific materials and systems for different use cases. Compromises

often have to be made to strike a balance between requirements. For example, lithium iron phosphate (LFP) batteries, a type of LIB, are widely used in electrical vehicles due to low cost, high thermal stability and long cycling life. But they are not suitable for cell phones because of lower energy density than most other LIBs. Even with the same choice of electrode materials, there is still a large and high-dimensional parameter space to explore to optimize the performance of batteries, including electrolyte formulation, cell configurations and assembly methods. SDLs can effectively address these complex problems, and offer large time and resource savings compared to traditional manual or high-throughput experiments.

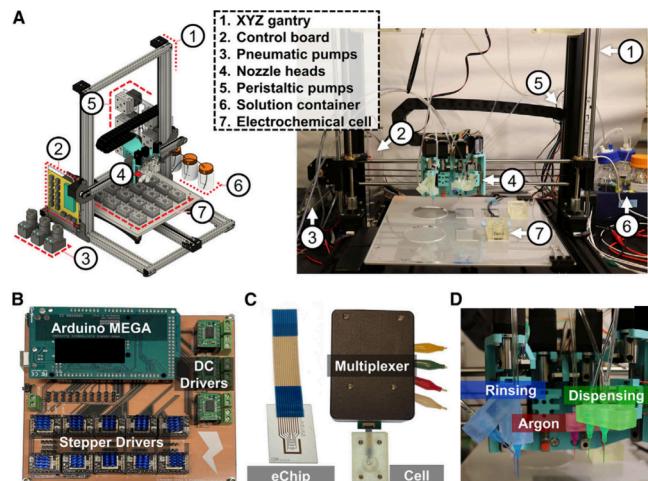
An ideal SDL for energy storage should be able to automatically design, make, assemble, and test energy storage technologies at different scales. An end-to-end platform for battery or flow battery development without human intervention is a major challenge. The industrial production of batteries has undergone significant automation to achieve high-throughput and capital efficiency. These processes aim to carry out precise actions to maximize consistency and productivity in large-scale manufacturing processes, at the cost of flexibility for research and development. SDLs for battery research should focus on the ability to test new materials and optimize electrochemical processes in batteries of standardized form factors. In comparison, flow batteries operate on a large variety of redox chemistries, many of which have not been scaled up to industrial relevant levels. Therefore, SDL for flow batteries should focus on the screening and scaling of molecules and materials, and the optimization of these materials in realistic flow reactors.

### 8.1. Materials Synthesis and Characterization

A major focus of energy storage technologies is to develop materials that can improve device performance and longevity. The synthesis of materials for energy storage requires low cost and high scalability because they are aimed to be produced at massive scales. Therefore, low-cost feedstocks such as metals, metal oxides, and products and wastes of petrochemical processes are greatly preferred. There is also a drive to simplify preparation steps, and minimizing the need for solvents and purifications. In comparison, there is a stronger motivation to characterize materials as detailed as possible, whether as synthesized, *in situ* or even *operando*, to fully understand underlying processes. SDLs for energy storage will likely have a relatively small synthesis component, but relatively complex characterization capabilities.

The common characterization methods include XRD, SEM, and solid-state NMR for solid state materials; thermal analysis for polymers; and most importantly, various electrochemical methods. The ability to conduct electrochemical analysis in a fully automated fashion is the prerequisite of any SDL that studies batteries, flow batteries, fuel cells or electrolyzers. However, this is often a challenge due to the fact that most commercial potentiostats, the instrument that performs electrochemical tests, are expensive and only operate with closed-source softwares, making them difficult to integrate into automated workflows.

An example of such effort is the Electrolab, a modular electrochemistry platform by Oh *et al.* able to automatically formulate redox electrolytes and characterize them using cyclic voltammetry (CV) across various conditions without human intervention (Figure 48).<sup>664</sup> Their platform integrates a liquid-



**Figure 48.** Electrolab is an automated electrochemistry platform that can dispense solutions onto electrochemical cells and run cyclic voltammograms, all without human intervention. (A) A schematic and picture of the Electrolab gantry-style SDL. (B) The control system of the Electrolab. (C) The vital electrochemical modules, with a microfabricated “eChip” electrode array for CV scans. (D) Fluidic nozzle system controlled by the gantry that can dispense and dispose fluids, rinse and flush with solvent, and dry and sparge with Ar gas. Figure reproduced with permission from Oh *et al.*<sup>664</sup> Copyright 2023, Elsevier.

handling robot to mix solutions and dispense them onto a series of cells containing an electrode array connected to a potentiostat (Figure 48C). After measurements are done, the robot can also de-gas and clean the cells. Electrolab was able to run a series of 200 CV scans in 2 hours (along with cleaning) on a known redox species under a variety of concentrations and scan rates. They then demonstrated a grid search to find the optimal conditions for a supporting electrolyte when scanning a candidate redox polymer for a nonaqueous RFB. Given the modularity of their setup, it seems possible to extend to smarter data-driven search algorithms to find interesting molecular candidates or remove the need for exhaustive grid search in the future.

In recent works initially introduced by Hickman *et al.*,<sup>304</sup> they demonstrated a low-cost SDL platform for electrochemistry discovery. The platform combines a synthesis platform, MEDUSA, and open source potentiostat for end-to-end automated complexation and electrochemical characterization. Adapted to the ChemOS 2.0 orchestration framework,<sup>111</sup> a closed-loop optimization of the redox potential of metal complexes for flow battery application was demonstrated using the Atlas optimization library.<sup>304</sup> The low-cost and open-source features of such a platform make it accessible to a broad scope of researchers, therefore allowing for the democratization of SDL for community-based research.

**8.1.1. Lithium-Ion Batteries.** Lithium-ion batteries (LIB) are among the most influential technologies today, enabling the establishment of two hundred-billion markets: portable electronics and the electric vehicles.<sup>665</sup> Since their commercialization in 1991, the energy density of LIBs has been improved by over 200%, but as the market grows and demand diversifies, there is a need to optimize their performance, stability, safety, cost, and environmental footprint. Much of the conventional workflow is centered on trial-and-error approaches to find better materials. Given the enormous space

of possible electrolytes, it is difficult and unreliable to screen active materials by manual experimentation. This motivates the need for SDL systems to improve LIBs. Data-driven ML methods for battery design have already been demonstrated on limited experimental dataset, such as for electrolyte formulation<sup>666,667</sup> and battery lifetimes.<sup>668</sup> In general, a fully automated workflow is difficult and expensive due to sophisticated engineering requirements, whereas semi-automated platforms that investigate some aspects of the material are more feasible for researchers.<sup>669</sup>

Electrolyte formulation is one of the key research areas of LIB research. Most LIBs nowadays require a liquid electrolyte to conduct electricity with minimal undesirable chemical reactions. Dave *et al.* developed a pipeline to automatically measure the electrochemical properties of 10 different solutions in different compositions (251 total) for use in LIBs using a Bayesian optimizer they developed called Dragonfly.<sup>670</sup> They later extended their pipeline to non-aqueous LIB electrolytes, which is a larger search space than aqueous electrolytes on account of co-solvents (although not necessarily a harder search space, as finding electrolytes that work in water is difficult).<sup>671</sup> Their system could automatically create and characterize different solution compositions, although some human assistance was required when transferring electrolytes into pouch cells for characterization. Their search space consisted of over 1000 points over three axes: solvent mass fraction, co-solvent ratios, and salt molality. In both cases, Dragonfly found electrolyte compositions that were novel or non-intuitive and better than benchmark electrolytes. They also noted that their experiment planner resulted in a six-fold speed increase in finding the optimum compared to random searching by their robotics platform and postulated the same process would take far longer through manual search. Svensson *et al.* developed an automated screening platform for different electrolyte formulations, consisting of two platforms: a system to formulate and characterize electrolyte solutions and a system for coin cell assembly/disassembly and electrochemistry characterization, which are linked together using a robotic arm.<sup>672</sup> While they only screened one electrolyte formulation as a proof-of-concept, their robotic platform was able to obtain similar measurements for assembled batteries compared with batteries assembled by hand, showing that this part of the battery development pipeline can be automated as well.

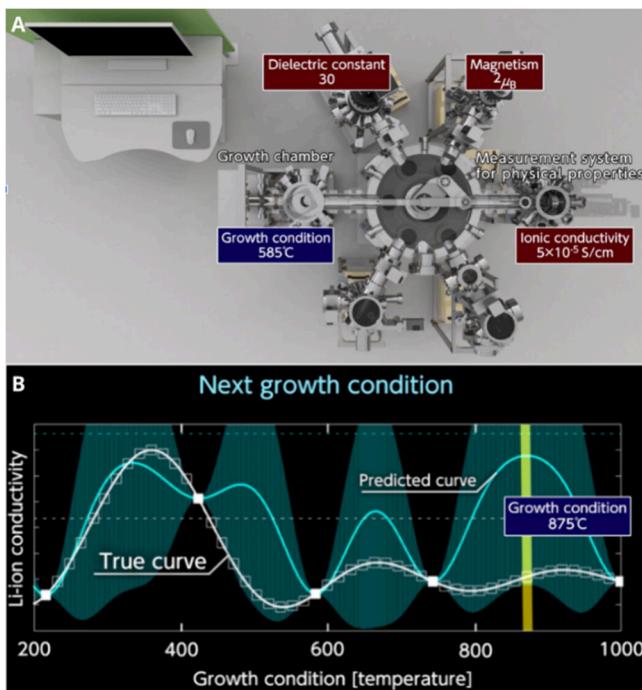
Vogler *et al.*<sup>132</sup> demonstrate a brokering approach to orchestrate modular and asynchronous research workflows, enabling the integration of multiple laboratories for LIB electrolyte development. They implement a passive brokering server called FINALES to mediate communication between various tenants, which can be physical modules like experimental equipment, or digital modules such as machine learning agents or simulations software. The SDL comprises an experimental setup for automated synthesis and analysis of LIB electrolytes through a pump and valve system with stock solutions, and connected densimeter and viscometer. As another tenant, a simulation orchestrator using Pipeline Pilot for molecular dynamics simulations is used to calculate ionic transport coefficients, radial distribution functions, and other properties critical for electrolyte performance. An AiiDA interface provides ML models to predict low-fidelity conductivity values,<sup>673</sup> and also for BO surrogate modeling. As a proof of concept, the authors aimed to maximize viscosity while minimizing density, using a GP optimizer combined with

Chimera for multi-objective optimization. The demonstration successfully orchestrated these tenants across five countries, optimizing electrolyte formulations based on lithium hexafluorophosphate salt in carbonate solvents like ethylene carbonate and ethyl methyl carbonate. This SDL approach enables efficient screening and optimization of new electrolyte compositions to improve critical performance metrics like ionic conductivity, essential for developing next-generation LIBs.

Another key research direction of solid-state electrolytes (SSE) for LIBs, such as metal oxide and polymer ion conductors, is to avoid the fire hazard and degradation issues caused by organic solvents. SSEs are also more compact than liquid electrolyte giving rise to higher energy density in batteries.<sup>674</sup> However, finding solid-state materials with high ionic conductivity, low electrical conductivity, and high electrochemical stability is a major challenge.<sup>675</sup> Currently there is no single best solid state electrolyte material for LIBs, partially due to limited understanding of lithium ion transport in bulk solids and at interfaces of different materials.<sup>676</sup> Computationally, a number of works have developed frameworks to featurize solid state conductors and train ML models to either predict properties or recommend new conductors.<sup>677</sup> He *et al.* developed a high-throughput screen platform which integrates a large database with modules that calculate ion-transport-related properties and a hierarchical search algorithm to propose promising candidates.<sup>678</sup> Laskowski *et al.* reported ML-guided synthesis of Si-doped Li<sub>3</sub>BS<sub>3</sub> using solid-state reactions, reaching superionic conductivity above 10<sup>-3</sup> S cm<sup>-1</sup>, surpassing most reported SSEs.<sup>679</sup> However, their discovery approach is neither automatic nor closed-loop.

To our knowledge, fully automated close-loop SDL that discovers/optimizes solid-state electrolyte materials is much needed but very rare. One of the closest examples of an SDL was developed by Shimizu *et al.*<sup>680</sup> The Connected, Autonomous, Shared and High-throughput (CASH) laboratory integrated several components of an automated SDL, with some human-in-the-loop steps in initializing the synthesis step. The authors minimize the electrical resistance of Nb-doped TiO<sub>2</sub> thin films by varying the oxygen partial pressure during the deposition of the film. Human intervention was required to load substrates and prepare the system for sputter deposition, after which the thin-film deposition and resistance characterization were carried out automatically. A robotic arm transferred the sample between the dedicated chambers (Figure 49). For experimental planning, a BO approach was utilized, with a GP regressor as the surrogate. The CASH SDL identified the global minimal resistance within 18 experimental samples for two different sputtering targets. The authors also outlined future plans to expand the characterization platform for multiple physical properties.

Active electrode materials store ions over many charging-discharging cycles and determine the battery's cell voltage. The chemical space of possible active electrode materials is large, ranging from graphite and Si-metal alloys to mixed-metal oxides and phosphate salts. By 2010, there were over 25,000 real and hypothetical negative electrode materials investigated, but experimental verification remains a bottleneck.<sup>681</sup> SDL for the discovery of new electrode materials does not yet exist, to our best knowledge. However, the development of high-throughput methods, such as through physical vapor deposition or solution-based methods, have allowed for combinatorial exploration of electrode materials, for example, negative electrode Si-metal alloys,<sup>682,683</sup> and positive electrode



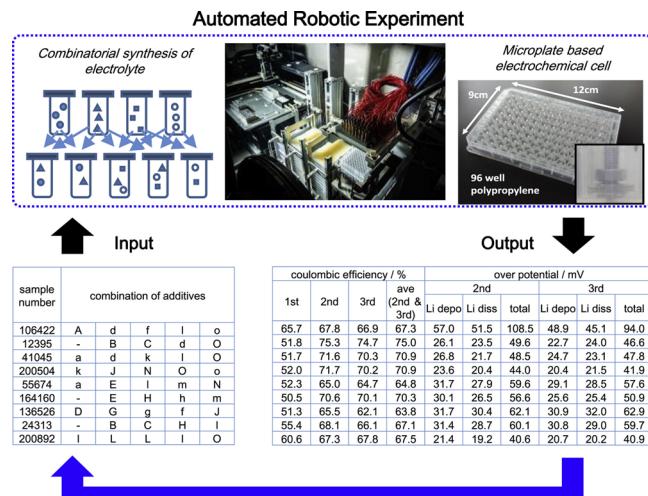
**Figure 49.** Illustration of CASH (Connected, Autonomous, Shared, and High-throughput) (A) proposed by Shimizu *et al.* (B) The system synthesizes thin films using deposition conditions commanded by BO, after which the film's resistance is evaluated. Adapted with permission from Shimizu *et al.*<sup>680</sup> Copyright 2020, American Institute of Physics Publishing.

Li-Ni-Mn-Co-O or Li-M-PO<sub>4</sub>.<sup>684</sup> In addition to the high-throughput synthesis of these materials, there are various high-throughput characterization methods for both structural and electrochemical characterization of electrode materials.<sup>685,686</sup> By performing characterization in large batches, combinatorial searches of the material space can generate large datasets for future data-driven applications.

McCalla outlined efforts and engineering bottlenecks using automated workflows to accelerate the design of battery materials, including solid-state electrolytes and electrode materials.<sup>669</sup> Currently, semi-automated systems might be more feasible for most academic researchers because they balance speed, cost, and adaptability. Nonetheless, a review by Szymanski *et al.* discussed in detail the challenges and opportunities in close-loop optimization of inorganic materials for batteries, highlighting the importance of future SDLs for not only liberating human researchers from low-level manual tasks but also possessing the ability to explore new materials without being limited by the development of theories.<sup>686</sup>

**8.1.2. Alternatives to LIBs.** There has been extensive work on alternatives to LIBs, such as Li-O<sub>2</sub><sup>687</sup> and Li-S<sup>688</sup> batteries, looking to achieve many folds higher energy density; and sodium-ion batteries, aiming to reduce cost and the reliance on metal resources such as lithium, cobalt, and nickel.<sup>689</sup> SDL can help develop crucial materials for these technologies which are still in early phases of research. Matsuda *et al.* demonstrated an SDL for the automated electrolyte synthesis for Li-O<sub>2</sub> batteries.<sup>689</sup> These batteries suffer from poor cycle performance due to low reaction efficiencies for the oxygen-generating (positive) and lithium-forming (negative) electrodes. As a result of the high-throughput screening guided by ML, they found multi-

component electrolyte additives for Li-O<sub>2</sub> batteries that gave rise to a stable solid electrolyte interface. Their automated experiments covered a total of 14,460 samples, with 4,320 samples allocated to random search, another 4,320 for hill climbing involving the top 10 samples, an additional 4,320 for hill climbing with the top 50 samples, and finally, 1,500 samples for BO (Figure 50). Combination of the hill-climbing method with BO resulted in significantly improved Coulombic efficiency with all top 10 samples exceeding 91%.

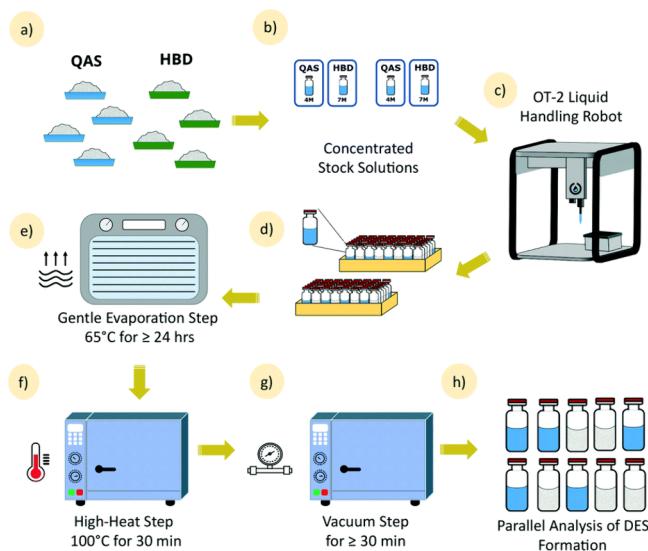


**Figure 50.** (left) Diagram illustrating the steps of the data-driven high-throughput automated robotic experiments designed to assess the Coulombic efficiencies of multi-component electrolyte additives. (right) Graph illustrating the average Coulombic efficiency (CE) achieved by the leading 5 samples over the course of the experiments. The x-axis represents the experiment iteration, distinguishing between random search (black), hill climbing with the top 10 samples (red), hill climbing with the top 50 samples (green), and BO (blue). Figure adapted with permission from Matsuda *et al.*<sup>689</sup> Copyright 2022, Elsevier.

**8.1.3. Redox Flow Batteries.** Since the debut of RFBs in the 1970s, researchers developed a variety of redox chemistries and technologies for RFBs, yet none of them have reached the scale of LIBs. Prior to 2015, RFBs primarily rely on inorganic salt RAMs, such as vanadium flow batteries (VFBs) and zinc bromide batteries. However, these batteries rely on scarce metal resources or highly corrosive operating conditions, which leads to a high cost of energy storage and maintenance. To achieve wide deployment, significant cost-reductions are needed to reach a target installation cost of \$100/kWh and a leveled storage cost of \$0.05/kWh.<sup>690</sup> This is achievable by optimization of the electrolyte solution, and discovery of more cost-effective chemistries.

Similar to LIBs, formulation of the electrolyte solution can result in enhanced performance. Gao *et al.* presented the Solubility of Organic Molecules in Aqueous Solution (SOMAS) dataset for advancing ML algorithms in the exploration of aqueous organic RFBs.<sup>691</sup> In the case of non-aqueous flow batteries which use organic solvents instead of water, mixed-solvent and mixed-electrolyte systems can be explored. Deep eutectic solvents (DESs) are an attractive choice of solvent with low toxicity, broad commercial availability, and low costs.<sup>692</sup> The properties of DESs can be fine-tuned with their composition. A recent study by Rodriguez *et al.* demonstrated a high throughput and data-driven search

for solvent formulation using open hardware (Figure 51).<sup>693</sup> They first outlined 3477 hydrogen bond donor (HBD) and



**Figure 51.** Schematic illustration of the high-throughput synthesis protocol for DESs used by Pozzo and co-workers. (a) The chemical space is defined by the QASs and HBDs used. (b) Solutions were prepared for the liquid handling robot (c). (d) The DES samples in 48-well plates were transferred to a dehydrator (e). (f) The samples were heated, and then (g) placed in a vacuum. (h) Final parallel analysis of the formed DESs. Reproduced with permission from Rodriguez *et al.*<sup>693</sup> Copyright 2016, Institution of Chemical Engineers (IChemE) and the Royal Society of Chemistry.

185 quaternary ammonium salt (QAS) molecules identified as good candidate components for DES and synthesized DESs using liquid handling robots to combine these components. They tested several physical properties, including melting point, electrochemical potential window and ionic conductivity. It is worth noting this work is based on Jubilee,<sup>61</sup> an open-source hardware machine based on 3D printing hardware with automatic tool-changing and interchangeable bed plates.

Another major research direction is to find low-cost electrolytes made from earth-abundant and widely available resources, and operate in mild aqueous environments. Over the past decade, researchers have explored various small organic molecules,<sup>694</sup> polymers,<sup>695</sup> and metal coordination compounds<sup>696</sup> to address the limitations of inorganic salts. The structural flexibility of organic molecules has facilitated the exploration of a broad spectrum of chemical and physical properties. They also resulted in a massive chemical space that is difficult to navigate with traditional computational and experimental methods.

Currently, there is no SDL capable of completing the DMAT cycles of new redox materials. The design step can be achieved based on the computational screening of different molecular classes, such as generating new analogues by combining core structures and making peripheral substitutions,<sup>697,698</sup> or using generative models and principles of inverse design.<sup>699</sup> For example, Jinich *et al.* computationally assessed 315,000 metabolic-inspired redox reactions<sup>700</sup> while S. V. *et al.* showcased the de novo design of radical species as both catholytes and anolytes.<sup>701,702</sup>

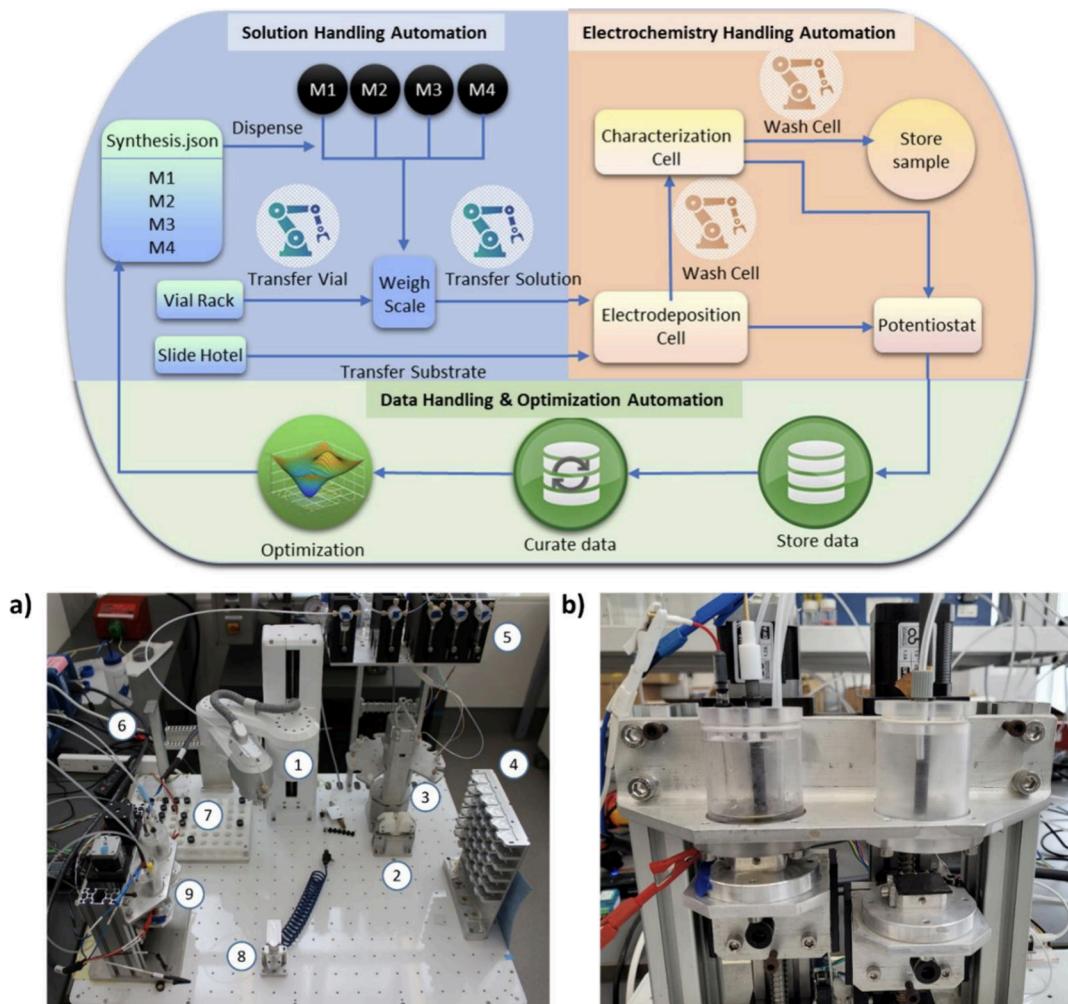
The above workflow narrows down the number of candidates that can be practically synthesized. However, the

“make” capabilities in SDLs are restricted to producing molecules within the same class that can be synthesized under similar conditions. Additionally, precise electrochemical assessment of RAMs demands samples of high purity. Conducting tests in real batteries necessitates a considerable quantity of samples, prompting the need for scaling up synthesis (see the Reaction Optimization section). Recently, a new class of radical-based organic RAMs showed promise for higher-throughput exploration due to a simple  $S_N2$  substitution reaction scheme.<sup>703</sup> A low-hanging fruit is the relatively straightforward synthesis of metal-ligand coordination compounds. Porwol *et al.* reported an autonomous chemical robot that can explore this process and discover the rules of coordination chemistry.<sup>477</sup> This is an important step towards generalizable synthesis of different redox-active metal-ligand complexes, which can be used to generate suitable RAMs on demand.

The test of new electrolytes focuses on the evaluation of key performance metrics, such as redox potentials, chemical stability and solubility. Liang *et al.* reported an important work on the high-throughput and automated solubility determination.<sup>704</sup> Solubility is important because it determines the highest possible energy density of the electrolyte solution. There is a significant computational challenge in quantitative prediction of the solubility of organic electrolytes. The authors assembled a robotic system in an argon glovebox to experimentally measure solubility of electrolytes. They showcased the effects of additives on solubility in aqueous flow batteries and the development of solubility databases for non-aqueous systems.

Most recently, Noh *et al.*<sup>705</sup> of the same research group presented an integrated high-throughput robotic platform and BO approach for accelerated discovery of optimal electrolyte formulations for non-aqueous RFBs, specifically the RAM 2,1,3-benzothiadiazole (BTZ). The goal of the study was to improve solubility of BTZ in organic solvents. The SDL carried out automated sample preparation through powder and liquid dispensing with a robotic arm. The solubility of BTZ in the solvent was measured via quantitative NMR spectroscopy. The BO component employs a surrogate GP model, operating on molecular features derived from physicochemical properties and DFT calculations of the solvent and solute. The authors identified multiple binary solvent systems with remarkable solubility thresholds exceeding 6.20 M from a comprehensive library of over 2000 potential solvents. Notably, their integrated strategy necessitated solubility assessments for fewer than 10% of these candidates, underscoring the efficiency of their approach.

**8.1.4. Hydrogen and Other Fuels.** Other than enclosed LIBs and flow batteries, electrical energy can also be stored in the chemical bonds of fuels.  $H_2$  has the highest *gravimetric energy density*, or *specific energy*, of any known chemical, although specialized materials and conditions are needed for its safe storage and transportation.<sup>706,707</sup> Liquid hydrocarbons offer a high *volumetric energy density*,<sup>708</sup> making them easy to store and transport indefinitely. Currently, only hydrogen and methanol can be directly converted back to electricity in fuel cells.<sup>709</sup> Ammonia is considered a good carbon-free energy carrier for the future,<sup>662</sup> although the electrosynthesis of  $NH_3$  from  $N_2$  is still challenging. Hydrocarbons are challenging for direct fuel cell consumption due to CO poisoning and carbon deposition on catalytic surfaces. They have to be reformed to generate  $H_2$  for hydrogen fuel cells.



**Figure 52.** Diagram illustrating the SDL setup designed for the electrochemical deposition and evaluation of the performance of OER catalysts made of amorphous mixed-metal oxide. (a) The EMAP (Electrocatalysis material acceleration platform) is comprised of (1) a robotic arm with an integrated pipette holder, (2) weigh scale and gripper, (3) liquid dispensing carousel, (4) slide hotel, (5) syringe pumps, (6) pipette tip holder, (7) vial rack, (8) slide gripper and (9) automated electrochemical cell. (b) Zoomed-in view of automated electrochemical cells. Adapted with permission from Fatehi *et al.*<sup>710</sup>

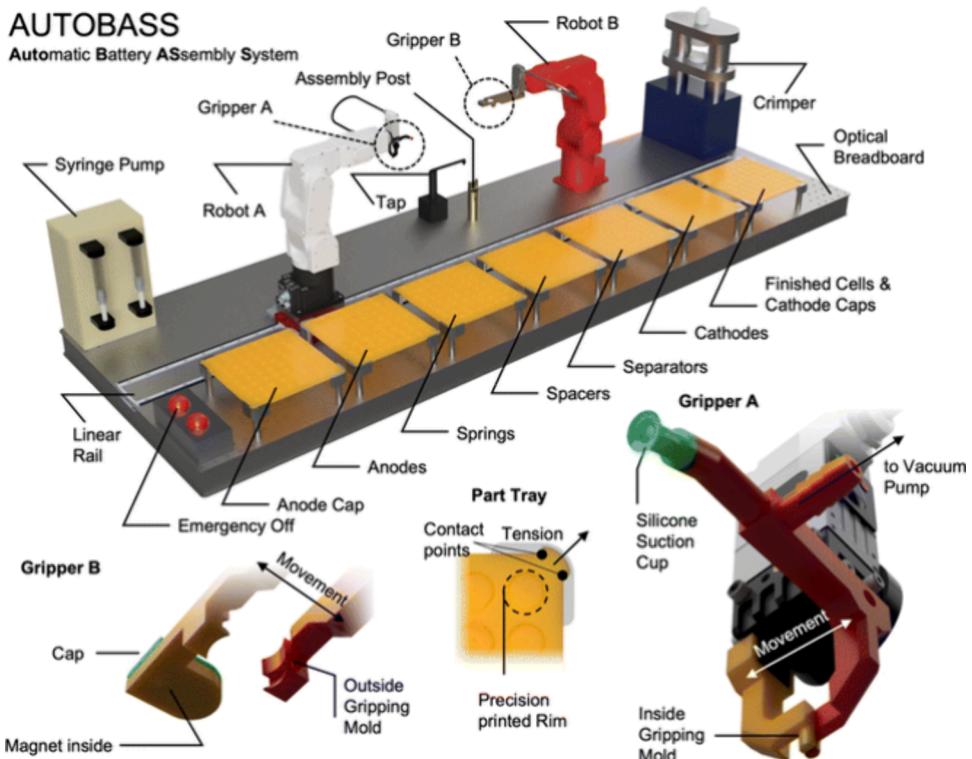
The development of electrocatalysts is central to the development of both fuel cells and electrolyzers. One important topic is the sourcing of hydrogen from water via electrolysis using earth-abundant catalysts. Fatehi *et al.* outlined the design of an SDL that is designed to find such catalysts to address the sluggish oxygen evolution reaction in water electrolysis.<sup>710</sup> They developed a framework for electrocatalyst SDLs consisting of three automated components: liquid handling, electrochemistry, and software that handles data and optimize experiment parameters (Figure 52). They use GP-based BO to find ideal material composition in a closed loop fashion. Their ultimate goal is to discover earth-abundant mixed-metal oxide catalysts for OERs in an acidic medium. They demonstrated the optimization of CoFeMnP-bO<sub>x</sub> electrodeposited catalyst materials through multiple campaigns. Within hours, they were able to identify successful formulae for catalyst synthesis and operational conditions that are corroborated with results in scientific literature. One interesting aspect of this work involved developing proxy measurements for target properties since the ideal characterization machinery was difficult to incorporate into the SDL. Fatehi *et al.* created a proxy measurement for stability by

holding the material at an overpotential for an extended period of time, and found that it was helpful for optimizing within a space of materials.

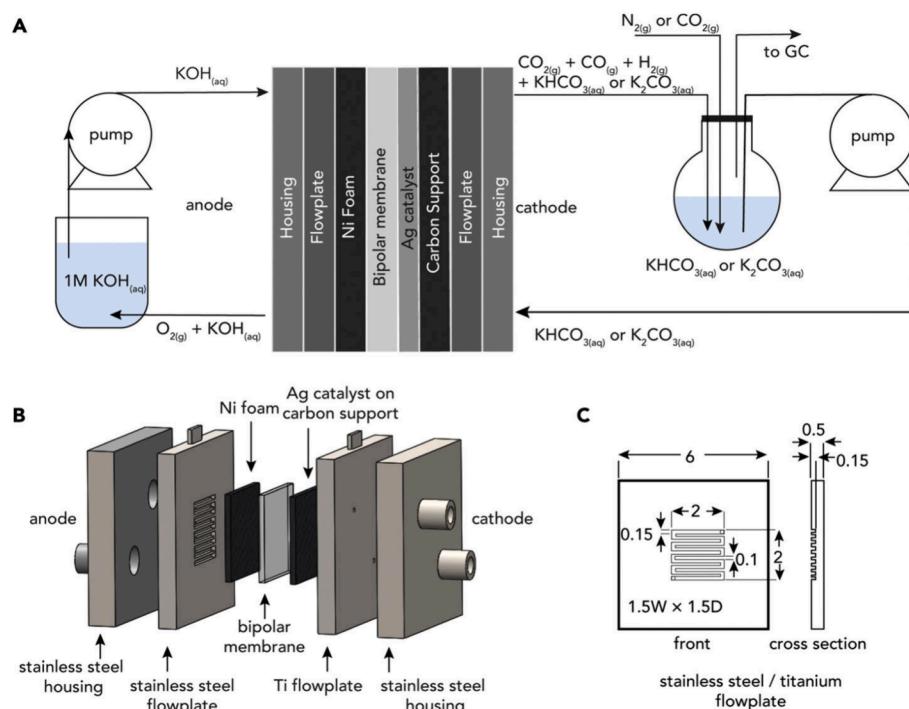
Another important work described the accelerated discovery of solid-state material in fuel cells which was discussed in a previous section on **Solid state materials synthesis**. The goal of the work by Ament *et al.*<sup>517</sup> was to autonomously design bismuth oxide thin films. The automated fabrication of Bi<sub>2</sub>O<sub>3</sub> films of different phases have possible applications to thin film solid oxide fuel cells.<sup>711</sup>

## 8.2. Device Design, Assembly and Optimization

**8.2.1. Cell Batteries.** Another important area of automation in battery development is cell assembly. This is typically done manually in a research setting. Coin cells are relatively easier to prepare than other cell types and have low material costs, which are beneficial for quick battery prototyping.<sup>713</sup> Zhang *et al.* developed AutoBASS, which automatically assembles 64 coin cell batteries per batch (Figure 53).<sup>712</sup> They found that their system produces consistent and reproducible results across batches and parameters for a single electrolyte, which is promising both for improving quality control and increasing the speed of lead discovery. Yik *et al.*



**Figure 53.** A visual representation of the automated battery assembly system (AutoBASS) is depicted, showcasing part trays designed for the assembly of CR2023 cells. The components are selected from these trays and positioned onto the assembly post using a gentle silicone suction cup attached to robot A. The cells are constructed with a downward-facing anode cap, and gripper B is responsible for flipping them. Gripper B also moves the filled and assembled cells to the crimper. The extraction from the crimper is facilitated by a magnetic pickup mechanism. Figure reproduced with permission from Zhang *et al.*<sup>712</sup> Copyright 2022, Royal Society of Chemistry.



**Figure 54.** A flow reactor setup for bicarbonate reduction converts captured  $\text{CO}_2$  into  $\text{CO}$ , which can be converted into other fuels. (A) A schematic demonstrating the electrochemical flow cell experiment, with corresponding reactions. (B) An expanded view of the flow cell within the stainless steel housing. (C) The flow plates used for the cathode and anode. Figure reproduced with permission from Li *et al.*<sup>713</sup> Copyright 2019, Elsevier.

created ODACell, a 4-robot system which combines electrolyte formulation with automated coin cell assembly.<sup>714</sup> While their batch throughput is smaller than that of AutoBASS (16 vs 64), their system can formulate different electrolyte compositions using a liquid handling robot, potentially allowing for easier integration with optimization algorithms in the future to search for ideal compositions. ODACell was used to test the impact of contaminants (specifically water) in non-aqueous batteries. The degradation of batteries was measured after being contaminated with different water concentrations, and it was found that the variance of the experiments increased at higher water concentrations when replicated multiple times. This illustrates how automation can effectively identify instances of failure or conditions characterized by elevated performance uncertainty.

**8.2.2. Flow Reactors.** The flow reactors in flow cells, flow batteries, and fuel cells share some common design elements, which in themselves have an enormous parameter space to explore. An electrochemical flow reactor is intricate, typically composed of a complex stack of multiple layers, including separators (commonly ion-exchange membranes), electrode materials, current collectors, gaskets, flow plates that regulate flow fields, inlets of liquids and gasses, etc. In most cases, they are assembled manually, as is the system demonstrated by Li *et al.* The electrochemical flow reactor assembly is difficult to automate, but tests can be performed (1) in parallel reactors and (2) sequentially using the sample reactor by cleaning out the reactor before the measurement. The research goal on the device level is often the optimization of performance by searching the parameter space of reactor design (flow field and materials) operating conditions (electrical and flow system management), and exhaustive monitoring of device degradation or failure over time.

As shown in Figure 54, the complex configuration of the electrochemical flow reactors often leads to reproducibility issues since slight misalignments and different applied pressures likely lead to varied device performance. Currently, only fuel cells based on proton exchange membranes are known to be assembled automatically in industry, using highly expensive commercial setups.<sup>716,717</sup> For instance, Thyssenkrupp Automation Engineering GmbH demonstrated a commercial plant that produces one electrolyzer per second, or at least 50,000 fuel cell stacks yearly. Such manufacturing maturity and scalability has not been realized in the production of flow reactors for flow batteries and electrolyzers.

Other than the exploration of the chemical space of RAMs, a secondary discovery process is conducted to explore the electrochemical parameter space of the RAMs, with a primary focus on optimizing battery performance. This exploration encompasses various factors, such as formulating electrolyte solutions, pairing posolyte and negolytes, determining properties of the membrane and electrode materials, designing the flow field, specifying flow rates, and other related considerations. The electrochemical parameter space is extensive and can become complex as more realistic factors are taken into account. While this exploration has been extensively conducted in a few inorganic systems, particularly in strongly acidic vanadium batteries, the realm of emerging organic RAMs operates under distinct conditions. This necessitates the use of new materials and battery designs for both aqueous and non-aqueous systems with different testing conventions.<sup>718</sup> Recently, the Aziz group reported an important step toward

high-throughput flow battery testing by miniaturizing the flow batteries using a modular design.<sup>719</sup>

### 8.3. Outlook and Perspectives

Energy storage technologies play a crucial role in achieving sustainability objectives. Generally, there is a strong industry effort on automation which guarantees robustness in production and product reliability. There is also strong academic research on ML for sustainable energy that expands beyond electrified energy storage to more general electrochemical sciences.<sup>720,721</sup> It is necessary to combine efforts in both communities, i.e., automation and ML expertise, to supercharge the discovery of advanced energy storage materials. For instance, LIB manufacturing has been highly automated in the industry. However, there is still a large design space for battery material discovery and optimization and battery design improvements.

There exist a number of opportunities to advance SDLs in the field of electrochemical energy storage. Integrating *operando* spectroscopic and electrochemical analysis of materials and devices, especially in flow reactors, will provide additional training data and a better understanding of failure mechanisms. As with optoelectronic materials and devices, simultaneously optimizing (co-designing) the different component materials in a device is an important potential contribution of SDLs. In most cases, RAMs for flow batteries are developed separately from the membrane or electrode materials. This could result in a mismatch of new RAMs and existing reactors, and thus subpar battery performance. Finally, SDLs capable of carrying out large scale or long term experiments on top candidates generated by another SDL could provide most realistic data about materials and devices for large, long-term energy storage systems.

We also predict the explosive growth of SDLs due to the rapid development of ML, automation, and significant investments from both private and government-led initiatives. The latter includes campaign to automate the discovery of new batteries and flow battery materials, including the European Battery2030 initiative,<sup>722</sup> the Department of Energy's efforts at Argonne National Laboratory<sup>723</sup> and Lawrence Berkeley National Laboratory,<sup>724</sup> and Canada's CA\$200 million investment in the Acceleration Consortium.<sup>589</sup> The U.S. government also recently announced a US\$7 billion investment from the Department of Energy to build seven hubs of H<sub>2</sub> infrastructure in the USA, which can significantly accelerate the implementation of SDLs for fuel-based energy storage.

## 9. CONCLUSION AND OUTLOOK

The evolution of SDLs within chemical and materials science promises to usher in a transformative era of scientific exploration. In this review, we have provided a comprehensive overview of SDL systems, both past and present, for a variety of applications. Early autonomous systems have been enhanced by rapid development of better automated chemistry platforms, improved AI-based experimental planning algorithms, and the availability of large, high-quality datasets fueled by advancements in information technology and computational power. Many Level 3 and 4 SDLs have already demonstrated impact in accelerating reaction process optimization, functional property refinement, and novel chemical and materials discovery. Further development of both custom and general automation systems has also significantly reduced the barrier to entry. Progress towards next-generation SDLs signals a

paradigm shift in which we believe SDLs will transition from systems designed and operated by specialists to everyday tools, similar to those brought about by the development of other now ubiquitous tools such as NMR, HPLC, MS, XRD, TEM, and SEM.

However, there are also important potential challenges in the future of SDLs. Most immediately, many contemporary SDL systems are very complex and expensive. Realizing the full potential of SDLs requires a concerted effort from the scientific community to embrace open-source software and hardware, democratizing access to these technologies and fostering collaboration. Numerous challenges must be addressed, including the development of robust and user-friendly interfaces, the establishment of standardized protocols and data formats, and adherence to the FAIR principles for data sharing. Effective implementation of FAIR data practices is crucial in enabling researchers worldwide to leverage the wealth of information generated by SDLs, and promoting transparency and reproducibility in chemistry and materials science. Initiatives, such as the creation of low-cost SDL prototypes and educational programs, play a vital role in empowering future scientists to navigate and contribute to this evolving multidisciplinary field. The growing importance of ML, automation, and SDLs is also forcing us to rethink education and workforce development in the physical sciences, where curricula often remain unchanged from the late 20th century. We also envision independent non-profit organizations capable of developing the talent within the community, building the ecosystem for collaboration, and supporting a platform to include private industry in collaborations that respect proprietary concerns.

As we continue development of SDLs, a key consideration lies in the role of human researchers in the scientific discovery process. As SDL technologies become more mature and widespread, the role of the researcher may shift toward translating the results from autonomous experimentation into scientific understanding,<sup>725</sup> which may be coupled with advances in explainable AI.<sup>726–729</sup> We foresee that human ingenuity will remain important in the discovery of new chemical and physical phenomena, novel classes of materials, and advanced laboratory techniques and technologies. Furthermore, while the Level 5 SDL is the pinnacle of autonomous experimentation, the human-in-the-loop Level 4 SDL will continue to be valuable, and perhaps even preferred, due to the adaptable and innovative nature of human problem-solving.<sup>730</sup> For such diverse and multidisciplinary fields as chemistry and materials science, the flexibility and modularity provided by semi-autonomous systems will be vital to SDL development. Moreover, as the barrier to chemistry and materials discovery becomes lower, and the process becomes faster, the potential misuse of SDLs for malicious purposes underscores the societal responsibility of researchers, the need for ethical guidelines, and the promotion of responsible implementation in industry. Striking a balance of economic considerations, ethical standards, and societal welfare is imperative to ensure the constructive and beneficial deployment of SDLs.

While the challenges are formidable, the potential benefits of a fully realized SDL ecosystem are substantial, and such an ecosystem will be the future of chemical discovery. By fostering a collaborative environment, promoting transparency, and aligning efforts towards a shared objective, the chemistry and materials science communities can accelerate the pace of

scientific discovery, explore new frontiers of knowledge, and drive innovation in ways that were previously unattainable.

## AUTHOR INFORMATION

### Corresponding Authors

**Gary Tom** — Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada; Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada; Vector Institute for Artificial Intelligence, Toronto, Ontario MSG 1M1, Canada; [orcid.org/0000-0002-8470-6515](https://orcid.org/0000-0002-8470-6515); Email: [gtom@cs.utoronto.ca](mailto:gtom@cs.utoronto.ca)

**Felix Strieth-Kalthoff** — Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada; Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada; School of Mathematics and Natural Sciences, University of Wuppertal, 42119 Wuppertal, Germany; Email: [strieth-kalthoff@uni-wuppertal.de](mailto:strieth-kalthoff@uni-wuppertal.de)

**Martin Seifrid** — Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada; Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada; Department of Materials Science and Engineering, North Carolina State University, Raleigh, North Carolina 27695, United States of America; [orcid.org/0000-0001-5238-0058](https://orcid.org/0000-0001-5238-0058); Email: [m\\_seifrid@ncsu.edu](mailto:m_seifrid@ncsu.edu)

**Alán Aspuru-Guzik** — Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada; Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada; Vector Institute for Artificial Intelligence, Toronto, Ontario MSG 1M1, Canada; Acceleration Consortium, Toronto, Ontario M5S 3H6, Canada; Department of Chemical Engineering & Applied Chemistry, University of Toronto, Toronto, Ontario M5S 3E5, Canada; Department of Materials Science & Engineering, University of Toronto, Toronto, Ontario M5S 3E4, Canada; Lebovic Fellow, Canadian Institute for Advanced Research (CIFAR), Toronto, Ontario MSG 1M1, Canada; [orcid.org/0000-0002-8277-4434](https://orcid.org/0000-0002-8277-4434); Email: [alan@aspuru.com](mailto:alan@aspuru.com)

### Authors

**Stefan P. Schmid** — Department of Chemistry and Applied Biosciences, ETH Zurich, CH-8093 Zurich, Switzerland; [orcid.org/0000-0002-0965-0208](https://orcid.org/0000-0002-0965-0208)

**Sterling G. Baird** — Acceleration Consortium, Toronto, Ontario M5S 3H6, Canada

**Yang Cao** — Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada; Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada; Acceleration Consortium, Toronto, Ontario M5S 3H6, Canada; [orcid.org/0000-0001-8636-7919](https://orcid.org/0000-0001-8636-7919)

**Kourosh Darvish** — Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada; Vector Institute for Artificial Intelligence, Toronto, Ontario MSG 1M1, Canada; Acceleration Consortium, Toronto, Ontario M5S 3H6, Canada

**Han Hao** — Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada; Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada; Acceleration Consortium, Toronto, Ontario M5S 3H6, Canada

**Stanley Lo** — Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada;  [orcid.org/0000-0003-0278-5318](https://orcid.org/0000-0003-0278-5318)

**Sergio Pablo-García** — Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada; Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada

**Ella M. Rajaonson** — Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada; Vector Institute for Artificial Intelligence, Toronto, Ontario M5G 1M1, Canada

**Marta Skreta** — Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada; Vector Institute for Artificial Intelligence, Toronto, Ontario M5G 1M1, Canada

**Naruki Yoshikawa** — Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada; Vector Institute for Artificial Intelligence, Toronto, Ontario M5G 1M1, Canada;  [orcid.org/0000-0003-1546-8709](https://orcid.org/0000-0003-1546-8709)

**Samantha Corapi** — Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada

**Gun Deniz Akkoc** — Forschungszentrum Jülich GmbH, Helmholtz Institute for Renewable Energy Erlangen-Nürnberg, 91058 Erlangen, Germany; Department of Chemical and Biological Engineering, Friedrich-Alexander Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acs.chemrev.4c00055>

## Author Contributions

CRediT: **Gary Tom** conceptualization, data curation, investigation, project administration, supervision, writing-original draft, writing-review & editing; **Stefan P. Schmid** conceptualization, data curation, investigation, writing-original draft, writing-review & editing; **Sterling G. Baird** investigation, writing-original draft, writing-review & editing; **Yang Cao** conceptualization, data curation, investigation, writing-original draft, writing-review & editing; **Kourosh Darvish** conceptualization, data curation, investigation, writing-original draft, writing-review & editing; **Han Hao** conceptualization, data curation, writing-original draft, writing-review & editing; **Stanley Lo** conceptualization, data curation, investigation, writing-original draft, writing-review & editing; **Sergio Pablo-García** conceptualization, data curation, investigation, writing-original draft, writing-review & editing; **Ella M. Rajaonson** conceptualization, data curation, investigation, writing-original draft, writing-review & editing; **Marta Skreta** conceptualization, data curation, investigation, writing-original draft, writing-review & editing; **Naruki Yoshikawa** conceptualization, data curation, investigation, writing-original draft, writing-review & editing; **Samantha Corapi** writing-original draft, writing-review & editing; **Gun Deniz Akkoc** writing-original draft, writing-review & editing; **Felix Strieth-Kalthoff** conceptualization, data curation, investigation, supervision, writing-original draft, writing-review & editing; **Martin Seifrid** conceptualization, data curation, investigation, supervision, writing-original draft, writing-review & editing; **Alán Aspuru-Guzik** conceptualization, funding acquisition, project administration, supervision, writing-review & editing.

## Notes

The authors declare the following competing financial interest(s): A.A.-G. is a founder of Kebotix, Inc., a company

specializing in closed-loop molecular discovery, and Intrepid Labs, Inc. a company using self-driving laboratories for pharmaceuticals.

## Biographies

**Gary Tom** is a PhD student in chemistry at the University of Toronto and the Vector Institute, working on machine learning applied to data-driven discovery in chemistry and materials science. He received his BSc from McGill University in physics, and MSc in experimental condensed matter from University of British Columbia. He is interested in generative models for materials inverse design, probabilistic modeling, and Bayesian optimization in chemical space exploration.

**Stefan P. Schmid** is a doctoral researcher in the Digital Chemistry Laboratory in the Department of Chemistry and Applied Biosciences at ETH Zurich. He obtained his MSc from ETH Zurich working on developing chemistry-informed representations for reactivity prediction. In his doctoral studies, he is working on developing experimental design algorithms to facilitate experimental catalyst discovery.

**Sterling G. Baird** is the Director of Training and Programs at the Acceleration Consortium within the University of Toronto. He received his BSc in Applied Physics and his MSc in Mechanical Engineering at Brigham Young University. He completed his PhD in Materials Science & Engineering at the University of Utah. Sterling is accelerating materials discovery through advanced Bayesian optimization, self-driving laboratories, and educational platforms.

**Yang Cao** is a staff scientist at the Acceleration Consortium of the University of Toronto. He obtained his BSc from Peking University and a PhD in chemistry at the University of British Columbia. He received training in organic synthesis, main-group chemistry and materials science, and now is working on building self-driving laboratories to accelerate the discovery of materials for sustainability.

**Kourosh Darvish** is currently a staff scientist at the Acceleration Consortium at the University of Toronto. He received his BSc and MSc degrees in aerospace engineering from K.N. Toosi University of Technology and Sharif University of Technology in Tehran, Iran, in 2012 and 2014, respectively, and his PhD in bioengineering and robotics from the University of Genoa in 2019. He was previously a post-doctoral researcher with the Italian Institute of Technology until May 2022. Later, he was a post-doctoral researcher with the Computer Science and Robotics Institute at the University of Toronto and was a member of the Vector Institute. His research focuses on robotics, shared autonomy, teleoperation, human-robot collaboration, and humanoid robotics.

**Han Hao** is a staff scientist at the Acceleration Consortium and a post-doctoral researcher at the University of Toronto. He obtained his PhD of chemistry at the University of British Columbia, working on early-transition metal catalyst development under the supervision of Prof. Laurel L. Schafer. After getting his degree, Han moved to the University of Toronto and joined the group of Prof. Alán Aspuru-Guzik to conduct research on development of the next generation of self-driving laboratories for the accelerated discovery of novel materials.

**Stanley Lo** is a PhD student at the University of Toronto, working on photocatalysis and polymer chemistry using self-driving laboratories. Stanley obtained his B.Sc. from the University of Toronto, where he worked on ML for organic photovoltaics and polymers.

**Sergio Pablo-García** is a post-doctoral researcher at University of Toronto. He obtained his BSc in Chemistry and his MSc in computer

modeling for physics and chemistry at University of Barcelona in Spain. He did his PhD research under the supervision of Prof. Núria López at the Institut Català d'Investigació Química, focusing his research in cheminformatics, automation and machine learning techniques applied to theoretical heterogeneous catalysis. His current research mainly focused in chemistry automation and AI for theoretical chemistry.

Ella M. Rajaonson is a PhD student in the Department of Chemistry at the University of Toronto and the Vector Institute, working on the development of high-throughput virtual screening and deep learning techniques for drug discovery.

Marta Skreta is a PhD student in the Department of Computer Science at the University of Toronto and the Vector Institute, where she works on ML for the design of self-driving laboratories and learning representations for molecular discovery.

Naruki Yoshikawa is a PhD student at the Department of Computer Science of the University of Toronto. He is working on automation of chemistry experiments under the supervision of Alán Aspuru-Guzik. He received his master's degree in 2020 at the University of Tokyo and his bachelor's degree in 2018 at the same institution.

Samantha Corapi is a MSc student in Chemistry at the University of Toronto. She completed her BSc in Chemistry at the University of Toronto where her research focused on photocatalysis, computational chemistry, and ML applications for chemistry. Her current research interests include ML applications for materials discovery.

Gun Deniz Akkoc received both his BSc and MSc degrees in chemistry from Izmir Institute of Technology during which he also carried out multiple projects ranging from building high-throughput synthesis platforms for battery materials to automated on-line analysis of hydrocarbons through spectroscopy and ML. He now continues his research as a PhD student at The Helmholtz Institute Erlangen-Nürnberg for Renewable Energy working on high-throughput electrocatalysis.

Felix Strieth-Kalthoff is an assistant professor of Digital Chemistry at the University of Wuppertal, Germany. Initially trained as an organic chemist, Felix obtained his PhD from the University of Münster, working on systematic and computational strategies for reaction discovery in homogeneous catalysis. As a Schmidt Futures AI in Science postdoctoral fellow at the University of Toronto, he extended his focus to the integration of machine learning with lab automation. Felix's research interests center on data-rich experimentation and decision-making strategies for accelerated discovery in organic synthesis.

Martin Seifrid is an assistant professor in the Department of Materials Science and Engineering at North Carolina State (NC State) University. He received his PhD from the University of California, Santa Barbara, where he studied relationships between molecular structure, processing, solid-state structure, and material and device properties of organic semiconductors. As a postdoctoral fellow at the University of Toronto, Martin integrated automated experiments and ML to enable accelerated discovery of organic materials by SDLs. At NC State, his group is interested in developing SDLs for soft matter, and accelerating the development of organic mixed-ionic-electronic conductors.

Alán Aspuru-Guzik is a professor of Chemistry, Computer Science, Chemical Engineering & Applied Chemistry, and Materials Science & Engineering at the University of Toronto. He is also the Canada 150 Research Chair in Theoretical Chemistry and a Canada CIFAR AI Chair at the Vector Institute. He is a CIFAR Lebovic Fellow in the Biologically Inspired Solar Energy program. Alán is the director of the

Acceleration Consortium, a University of Toronto-based strategic initiative that aims to gather researchers from industry, government and academia around pre-competitive research topics related to the self-driving laboratory of the future.

## ACKNOWLEDGMENTS

We thank Riley Hickman for his contributions and advice in writing this review. We also thank the anonymous peer reviewers, and the members of the SDL community for their valuable feedback and suggestions. S.P.-G was responsible for the custom figures. This work was supported by 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 U.S. government. G.T. acknowledges the support of the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the Vector Institute. K.D. is supported by the Acceleration Consortium. University of Toronto's Acceleration Consortium receives funding from the Canada First Research Excellence Fund. S.L. acknowledges the support from the Government of Ontario through the Ontario Graduate Scholarship. S.P.-G. acknowledges that 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. E.R. acknowledges support from the Vector Institute. M.S is supported by the NSERC-Google Industrial Research Chair award, and the NSERC Canadian Graduate Scholarship-Doctoral award. N.Y. is supported by the NSERC-Google Industrial Research Chair award. G.D.A gratefully acknowledges the German Ministry of Education and Research for financial support within the project 03HY108A. Furthermore, part of this work by G.D.A was supported by Mitacs through the Mitacs Globalink program. F.S.-K.'s contributions were supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship Program, a program by Schmidt Futures. A.A.-G. acknowledges the generous support of Anders G. Frøseth, the CIFAR, and the Canada 150 Research Chair program.

## REFERENCES

- (1) Abolhasani, M.; Kumacheva, E. The Rise of Self-Driving Labs in Chemical and Materials Sciences. *Nat. Synth.* **2023**, *2*, 483–492.
- (2) Zhu, Q.; Zhang, F.; Huang, Y.; Xiao, H.; Zhao, L.; Zhang, X.; Song, T.; Tang, X.; Li, X.; He, G.; et al. An All-Round AI-Chemist with a Scientific Mind. *Natl. Sci. Rev.* **2022**, *9*, nwac190.
- (3) Bergman, R. G.; Danheiser, R. L. Reproducibility in Chemical Research. *Angew. Chem. Int. Ed.* **2016**, *55*, 12548–12549.
- (4) Baker, M. 1,500 Scientists Lift the Lid on Reproducibility. *Nature*. **2016**, *533*, 452–454.
- (5) Beker, W.; Roszak, R.; Wołos, 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.
- (6) 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*, No. e202204647.

- (7) Greenaway, R. L.; Jelfs, K. E.; Spivey, A. C.; Yaliraki, S. N. From Alchemist to AI Chemist. *Nat. Rev. Chem.* **2023**, *7*, 527–528.
- (8) Maruyama, B. Air Force Research Laboratory. Personal Communication with Benji Maruyama, 2018.
- (9) J3016\_202104: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles - SAE International. [https://www.sae.org/standards/content/j3016\\_202104/](https://www.sae.org/standards/content/j3016_202104/) (accessed 2023-11-13).
- (10) Lunt, A. M.; Fakhruldeen, H.; Pizzuto, G.; Longley, L.; White, A.; Rankin, N.; Clowes, R.; Alston, B. M.; Cooper, A. I.; Chong, S. Y. Powder-Bot: A Modular Autonomous Multi-Robot Workflow for Powder X-Ray Diffraction. *arXiv* 2023. DOI: [10.48550/arXiv.2309.00544](https://doi.org/10.48550/arXiv.2309.00544) (accessed October 31, 2023).
- (11) Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; et al. A Mobile Robotic Chemist. *Nature*. **2020**, *583*, 237–241.
- (12) 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.
- (13) Xu, H.; Wang, Y. R.; Eppel, S.; Aspuru-Guzik, A.; Shkurti, F.; Garg, A. Seeing Glass: Joint Point Cloud and Depth Completion for Transparent Objects. *arXiv* 2021. DOI: [10.48550/arXiv.2110.00087](https://doi.org/10.48550/arXiv.2110.00087) (accessed March 15, 2023).
- (14) El-khawaldeh, R.; Guy, M. A.; Bork, F.; Taherimakhousi, N.; Jones, K. N.; Hawkins, J.; Han, L.; Pritchard, R. P.; Cole, B.; Monfette, S.; et al. Keeping an “Eye” on the Experiment: Computer Vision for Real-Time Monitoring and Control. *Chem. Sci.* **2024**, *15*, 1271.
- (15) Tom, G.; Schmid, S. P.; Baird, S. G.; Cao, Y.; Darvish, K.; Hao, H.; Lo, S.; Pablo-García, S.; Rajaonson, E. M.; Skreta, M. et al. Self-Driving Laboratories for Chemistry and Materials Science. *ChemRxiv* 2024. DOI: [10.26434/chemrxiv-2024-rj946-v2](https://doi.org/10.26434/chemrxiv-2024-rj946-v2) (accessed June 25, 2024).
- (16) Fisher, R. A. *The Design of Experiments*; Oliver and Boyd: Edinburgh, London, 1937.
- (17) Smallwood, H. M. Design of Experiments in Industrial Research. *Anal. Chem.* **1947**, *19*, 950–952.
- (18) Spinrad, R. J. Automation in the Laboratory. *Science*. **1967**, *158*, 55–60.
- (19) Seinfeld, J. H.; McBride, W. L. Optimization with Multiple Performance Criteria. Application to Minimization of Parameter Sensitivities in a Refinery Model. *Ind. Eng. Chem. Process Des. Dev.* **1970**, *9*, 53–57.
- (20) Owens, G. D.; Eckstein, R. J.; Franz, T. P. Laboratory Robotics - Past, Present, and Future. *Microchim. Acta*. **1986**, *89*, 15–30.
- (21) Topham, S. A. The History of the Catalytic Synthesis of Ammonia. In *Catalysis: Science and Technology*; Anderson, J. R.; Boudart, M., Eds.; Springer: Berlin, Heidelberg, 1985; pp 1–50. DOI: [10.1007/978-3-642-93281-6\\_1](https://doi.org/10.1007/978-3-642-93281-6_1).
- (22) Boyd, J. Robotic Laboratory Automation. *Science* **2002**, *295*, 517–518.
- (23) Merrifield, R. B.; Stewart, J. M.; Jernberg, N. Instrument for Automated Synthesis of Peptides. *Anal. Chem.* **1966**, *38*, 1905–1914.
- (24) Ernst, R. R. Measurement and Control of Magnetic Field Homogeneity. *Rev. Sci. Instrum.* **1968**, *39*, 998–1012.
- (25) Owens, G. D.; Eckstein, R. J. Robotic Sample Preparation Station. *Anal. Chem.* **1982**, *54*, 2347–2351.
- (26) Little, J. N. Advances in Laboratory Robotics for Automated Sample Preparation. *Chemom. Intell. Lab. Syst.* **1993**, *21*, 199–205.
- (27) Morgan, S. L.; Deming, S. N. Simplex Optimization of Analytical Chemical Methods. *Anal. Chem.* **1974**, *46*, 1170–1181.
- (28) Deming, S. N.; Parker, L. R.; Bonner Denton, M. A Review of Simplex Optimization in Analytical Chemistry. *Crit. Rev. Anal. Chem.* **1978**, *7*, 187–202.
- (29) Lucia, A.; Xu, J. Chemical Process Optimization Using Newton-like Methods. *Comput. Chem. Eng.* **1990**, *14*, 119–138.
- (30) Mockus, J. On Bayesian Methods for Seeking the Extremum. In *Proceedings of the IFIP Technical Conference*; Springer-Verlag: Berlin, Heidelberg, 1974; pp 400–404. DOI: [10.1007/3-540-07165-2\\_55](https://doi.org/10.1007/3-540-07165-2_55).
- (31) Wolpert, D. H.; Macready, W. G. No Free Lunch Theorems for Optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82.
- (32) Androulakis, I. P.; Venkatasubramanian, V. A Genetic Algorithmic Framework for Process Design and Optimization. *Comput. Chem. Eng.* **1991**, *15*, 217–228.
- (33) Isenhour, T. L. Robotics In The Laboratory. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 292–295.
- (34) Bourne, S. L.; Amann, F.; Ley, S. V. The Evolution of Flow Chemistry: An Opinion on Factors Driving Innovation. *CHIMIA* **2023**, *77*, 288–288.
- (35) Krishnadasan, S.; Brown, R. J. C.; deMello, A. J.; deMello, J. C. Intelligent Routes to the Controlled Synthesis of Nanoparticles. *Lab. Chip.* **2007**, *7*, 1434–1441.
- (36) Krishnadasan, S.; Tovilla, J.; Vilar, R. J.; deMello, A. C.; deMello, J. On-Line Analysis of CdSe Nanoparticle Formation in a Continuous Flow Chip-Based Microreactor. *J. Mater. Chem.* **2004**, *14*, 2655–2660.
- (37) King, R. D.; Rowland, J.; Oliver, S. G.; Young, M.; Aubrey, W.; Byrne, E.; Liakata, M.; Markham, M.; Pir, P.; Soldatova, L. N.; et al. The Automation of Science. *Science* **2009**, *324*, 85–89.
- (38) Williams, K.; Bilsland, E.; Sparkes, A.; Aubrey, W.; Young, M.; Soldatova, L. N.; De Grave, K.; Ramon, J.; de Clare, M.; Sirawaraporn, W.; et al. Cheaper Faster Drug Development Validated by the Repositioning of Drugs against Neglected Tropical Diseases. *J. R. Soc. Interface*. **2015**, *12*, 20141289.
- (39) El-khawaldeh, R.; Hein, J. E. Balancing Act: When to Flex and When to Stay Fixed. *Trends Chem.* **2024**, *6*, 1.
- (40) MacLeod, B. P.; Parlame, F. G. L.; Brown, A. K.; Hein, J. E.; Berlinguette, C. P. Flexible Automation Accelerates Materials Discovery. *Nat. Mater.* **2022**, *21*, 722–726.
- (41) Lo, S.; Baird, S. G.; Schrier, J.; Blaiszik, B.; Carson, N.; Foster, I.; Aguilar-Granda, A.; Kalinin, S. V.; Maruyama, B.; Politi, M.; et al. Review of Low-Cost Self-Driving Laboratories in Chemistry and Materials Science: The “Frugal Twin” Concept. *Digit. Discov.* **2024**, *3*, 842–868.
- (42) OT-2 Robot - Opentrons. <https://shop.opentrons.com/ot-2-robot/> (accessed 2024-01-17).
- (43) Keesey, R.; LeSuer, R.; Schrier, J. Sidekick: A Low-Cost Open-Source 3D-Printed Liquid Dispensing Robot. *HardwareX* **2022**, *12*, No. e00319.
- (44) Kadokawa, Y.; Hamaya, Masashi; Tanaka, K. Learning Robotic Powder Weighing from Simulation for Laboratory Automation. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: Detroit, MI, USA, 2023; pp 2932–2939. DOI: [10.1109/IROS5555.2023.10342463](https://doi.org/10.1109/IROS5555.2023.10342463).
- (45) 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*, 15473–15490.
- (46) Steiner, S.; Wolf, J.; Glatzel, S.; Andreou, A.; Granda, J. M.; Keenan, G.; Hinkley, T.; Aragon-Camarasa, G.; Kitson, P. J.; Angelone, D.; et al. Organic Synthesis in a Modular Robotic System Driven by a Chemical Programming Language. *Science*. **2019**, *363*, No. eaav2211.
- (47) Fakhruldeen, H.; Pizzuto, G.; Glowacki, J.; Cooper, A. I. ARChemist: Autonomous Robotic Chemistry System Architecture. In *2022 International Conference on Robotics and Automation (ICRA)*; IEEE: Philadelphia, PA, USA, 2022; pp 6013–6019. DOI: [10.1109/ICRA46639.2022.9811996](https://doi.org/10.1109/ICRA46639.2022.9811996).
- (48) Pizzuto, G.; Wang, H.; Fakhruldeen, H.; Peng, B.; Luck, K. S.; Cooper, A. I. Accelerating Laboratory Automation Through Robot Skill Learning For Sample Scraping. 2022. [http://arxiv.org/abs/2209.14875](https://arxiv.org/abs/2209.14875) (accessed 2023-06-04). DOI: [10.48550/arXiv.2209.14875](https://doi.org/10.48550/arXiv.2209.14875)
- (49) Nakajima, Y.; Hamaya, M.; Suzuki, Y.; Hawai, T.; Drigalski, F. V.; Tanaka, K.; Ushiku, Y.; Ono, K. Robotic Powder Grinding with a Soft Jig for Laboratory Automation in Material Science. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems*

- (IROS); IEEE: Kyoto, Japan, 2022; pp 2320–2326. DOI: 10.1109/IROS47612.2022.9981081.
- (50) Kennedy, M.; Schmeckpeper, K.; Thakur, D.; Jiang, C.; Kumar, V.; Daniilidis, K. Autonomous Precision Pouring From Unknown Containers. *IEEE Robot. Autom. Lett.* **2019**, *4*, 2317–2324.
- (51) Huang, Y.; Wilches, J.; Sun, Y. Robot Gaining Accurate Pouring Skills through Self-Supervised Learning and Generalization. *Robot. Auton. Syst.* **2021**, *136*, 103692.
- (52) Lim, J. X.-Y.; Leow, D.; Pham, Q.-C.; Tan, C.-H. Development of a Robotic System for Automatic Organic Chemistry Synthesis. *IEEE Trans. Autom. Sci. Eng.* **2021**, *18*, 2185–2190.
- (53) Knobbe, D.; Zwirnmann, H.; Eckhoff, M.; Haddadin, S. Core Processes in Intelligent Robotic Lab Assistants: Flexible Liquid Handling. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: Kyoto, Japan, 2022; pp 2335–2342. DOI: 10.1109/IROS47612.2022.9981636.
- (54) Jiang, Y.; Fakhruldeen, H.; Pizzuto, G.; Longley, L.; He, A.; Dai, T.; Clowes, R.; Rankin, N.; Cooper, A. I. Autonomous Biomimetic Solid Dispensing Using a Dual-Arm Robotic Manipulator. *Digit. Discov.* **2023**, *2*, 1733–1744.
- (55) Yoshikawa, N.; Akkoc, G. D.; Pablo-García, S.; Cao, Y.; Hao, H.; Aspuru-Guzik, A. Does One Need to Polish Electrodes in an Eight Pattern? Automation Provides the Answer. *ChemRxiv*. February 13, 2024. DOI: 10.26434/chemrxiv-2024-ttxnr (accessed June 10, 2024).
- (56) Baden, T.; Chagas, A. M.; Gage, G.; Marzullo, T.; Prieto-Godino, L. L.; Euler, T. Open Labware: 3-D Printing Your Own Lab Equipment. *PLOS Biol.* **2015**, *13*, No. e1002086.
- (57) Barthels, F.; Barthels, U.; Schwickert, M.; Schirmeister, T. FINDUS: An Open-Source 3D Printable Liquid-Handling Workstation for Laboratory Automation in Life Sciences. *SLAS Technol.* **2020**, *25*, 190–199.
- (58) Florian, D. C.; Odziomek, M.; Ock, C. L.; Chen, H.; Guelcher, S. A. Principles of Computer-Controlled Linear Motion Applied to an Open-Source Affordable Liquid Handler for Automated Micro-pipetting. *Sci. Rep.* **2020**, *10*, 13663.
- (59) Dettinger, P.; Kull, T.; Arekatla, G.; Ahmed, N.; Zhang, Y.; Schneiter, F.; Wehling, A.; Schirmacher, D.; Kawamura, S.; Loeffler, D.; et al. Open-Source Personal Pipetting Robots with Live-Cell Incubation and Microscopy Compatibility. *Nat. Commun.* **2022**, *13*, 2999.
- (60) Faiña, A.; Nejati, B.; Stoy, K. EvoBot: An Open-Source, Modular, Liquid Handling Robot for Scientific Experiments. *Appl. Sci.* **2020**, *10*, 814.
- (61) Vasquez, J.; Twigg-Smith, H.; O'Leary, J.; Peek, N. Jubilee: An Extensible Machine for Multi-Tool Fabrication; Association for Computing Machinery, 2020; p 13. DOI: 10.1145/3313831.3376425.
- (62) Politi, M.; Baum, F.; Vaddi, K.; Antonio, E.; Vasquez, J. P.; Bishop, B.; Peek, N. C.; Holmberg, V. D.; Pozzo, L. A High-Throughput Workflow for the Synthesis of CdSe Nanocrystals Using a Sonochemical Materials Acceleration Platform. *Digit. Discov.* **2023**, *2*, 1042–1057.
- (63) Nielsen, A. V.; Beauchamp, M. J.; Nordin, G. P.; Woolley, A. T. 3D Printed Microfluidics. *Annu. Rev. Anal. Chem.* **2020**, *13*, 45–65.
- (64) Yoshikawa, N.; Darvish, K.; Vakili, M. G.; Garg, A.; Aspuru-Guzik, A. Digital Pipette: Open Hardware for Liquid Transfer in Self-Driving Laboratories. *Digit. Discov.* **2023**, *2*, 1745–1751.
- (65) Pablo-García, S.; García, Á.; Akkoc, G. D.; Sim, M.; Cao, Y.; Somers, M.; Hattrick, C.; Yoshikawa, N.; Dworschak, D.; Hao, H. et al. An Affordable Platform for Automated Synthesis and Electrochemical Characterization. *ChemRxiv*. February 9, 2024. DOI: 10.26434/chemrxiv-2024-cwnwc (accessed June 10, 2024).
- (66) Accelerated Discovery - AI and automation to accelerate materials discovery. Accelerated Discovery. <https://accelerated-discovery.org/> (accessed 2024-06-07).
- (67) Shir, P.; Lai, V.; Zepel, T.; Griffin, D.; Reifman, J.; Clark, S.; Grunert, S.; Yunker, L. P. E.; Steiner, S.; Situ, H.; et al. Automated Solubility Screening Platform Using Computer Vision. *iScience*. **2021**, *24*, 102176.
- (68) El-khawaldeh, R.; Mandal, A.; Yoshikawa, N.; Zhang, W.; Corkery, R.; Prieto, P.; Aspuru-Guzik, A.; Darvish, K.; Hein, J. E. From Eyes to Cameras: Computer Vision for High-Throughput Liquid-Liquid Separation. *Device*. **2024**, 100404.
- (69) Sun, A. C.; Jurica, J. A.; Rose, H. B.; Brito, G.; Deprez, N. R.; Grosser, S. T.; Hyde, A. M.; Kwan, E. E.; Moor, S. Vision-Guided Automation Platform for Liquid-Liquid Extraction and Workup Development. *Org. Process Res. Dev.* **2023**, *27*, 1954.
- (70) Walker, M.; Pizzuto, G.; Fakhruldeen, H.; Cooper, A. I. Go with the Flow: Deep Learning Methods for Autonomous Viscosity Estimations. *Digit. Discov.* **2023**, *2*, 1540–1547.
- (71) Wang, Y. R.; Zhao, Y.; Xu, H.; Eppel, S.; Aspuru-Guzik, A.; Shkurti, F.; Garg, A. MVTrans: Multi-View Perception of Transparent Objects. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*; IEEE: London, United Kingdom, 2023; pp 3771–3778. DOI: 10.1109/ICRA48891.2023.10161089.
- (72) Klami, A.; Damoulas, T.; Engkvist, O.; Rinke, P.; Kaski, S. Virtual Laboratories: Transforming Research with AI. *TechRxiv2022*. DOI: 10.36227/techrxiv.20412540.v1 (accessed October 25, 2023).
- (73) Vescovi, R.; Ginsburg, T.; Hippe, K.; Ozgulbas, D.; Stone, C.; Stroka, A.; Butler, R.; Blaiszik, B.; Brettin, T.; Chard, K. et al. Towards a Modular Architecture for Science Factories. *arXiv2023*. DOI: 10.48550/arXiv.2308.09793 (accessed September 22, 2023)
- (74) Beeler, C.; Subramanian, S. G.; Sprague, K.; Chatti, N.; Bellinger, C.; Shahen, M.; Paquin, N.; Baula, M.; Dawit, A.; Yang, Z. et al. ChemGymRL: An Interactive Framework for Reinforcement Learning for Digital Chemistry. *arXiv2023*. <http://arxiv.org/abs/2305.14177> (accessed 2023-10-24). DOI: 10.48550/arXiv.2305.14177
- (75) Koenig, N.; Howard, A. Design and Use Paradigms for Gazebo, an Open-Source Multi-Robot Simulator. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE Cat. No.04CH37566); 2004; Vol. 3, pp 2149–2154. DOI: 10.1109/IROS.2004.1389727.
- (76) Todorov, E.; Erez, T.; Tassa, Y. MuJoCo: A Physics Engine for Model-Based Control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*; IEEE, 2012; pp 5026–5033. DOI: 10.1109/IROS.2012.6386109.
- (77) Isaac Sim - Robotics Simulation and Synthetic Data | NVIDIA Developer. <https://developer.nvidia.com/isaac/sim> (accessed 2024-06-07).
- (78) Mittal, M.; Yu, C.; Yu, Q.; Liu, J.; Rudin, N.; Hoeller, D.; Yuan, J. L.; Singh, R.; Guo, Y.; Mazhar, H.; et al. Orbit: A Unified Simulation Framework for Interactive Robot Learning Environments. *IEEE Robot. Autom. Lett.* **2023**, *8*, 3740–3747.
- (79) Li, C.; Xia, F.; Martín-Martín, R.; Lingelbach, M.; Srivastava, S.; Shen, B.; Vainio, K.; Gokmen, C.; Dharan, G.; Jain, T.; et al. iGibson 2.0: Object-Centric Simulation for Robot Learning of Everyday Household Tasks. *arXiv2021*. DOI: 10.48550/arXiv.2108.03272 (accessed October 12, 2023).
- (80) Yu, T.; Quillen, D.; He, Z.; Julian, R.; Narayan, A.; Shively, H.; Bellathur, A.; Hausman, K.; Finn, C.; Levine, S. Meta-World: A Benchmark and Evaluation for Multi-Task and Meta Reinforcement Learning. *arXiv2021*. DOI: 10.48550/arXiv.1910.10897 (accessed October 12, 2023).
- (81) Li, C.; Zhang, R.; Wong, J.; Gokmen, C.; Srivastava, S.; Martín-Martín, R.; Wang, C.; Levine, G.; Lingelbach, M.; Sun, J. et al. BEHAVIOR-1K: A Benchmark for Embodied AI with 1,000 Everyday Activities and Realistic Simulation; *arXiv2022*. DOI: 10.48550/arXiv.2403.09227
- (82) Dasari, S.; Wang, J.; Hong, J.; Bahl, S.; Lin, Y.; Wang, A.; Thankaraj, A.; Chahal, K.; Calli, B.; Gupta, S.; et al. RB2: Robotic Manipulation Benchmarking with a Twist. *arXiv2022*. DOI: 10.48550/arXiv.2203.08098 (accessed October 25, 2023).
- (83) Xian, Z.; Zhu, B.; Xu, Z.; Tung, H.-Y.; Torralba, A.; Fragiadaki, K.; Gan, C. FluidLab: A Differentiable Environment for Benchmarking Complex Fluid Manipulation; *arXiv2022*. DOI: 10.48550/arXiv.2303.02346

- (84) Coumans, E.; Bai, Y. PyBullet, a Python Module for Physics Simulation for Games, Robotics and Machine Learning, 2016. <http://pybullet.org> (accessed 2023-10-12).
- (85) NVIDIA Flex - 1.2.0, 2023. <https://github.com/NVIDIAAGameWorks/FleX> (accessed 2023-10-25).
- (86) NVIDIA PhysX, 2023. <https://github.com/NVIDIA-Omniverse/PhysX> (accessed 2023-10-25).
- (87) OpenGL - The Industry Standard for High Performance Graphics. <https://opengl.org/> (accessed 2024-06-07).
- (88) Unity Real-Time Development Platform | 3D, 2D, VR & AR Engine. Unity. <https://unity.com/> (accessed 2024-06-07).
- (89) Zhao, W.; Queraltà, J. P.; Westerlund, T. Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: A Survey. In 2020 IEEE Symposium Series on Computational Intelligence (SSCI); IEEE, 2020; pp 737–744. DOI: [10.1109/SSCI47803.2020.9308468](https://doi.org/10.1109/SSCI47803.2020.9308468).
- (90) Laboratory Safety Guidance; OSHA, 2011.
- (91) WHMIS.org | Canada's National WHMIS Portal. <https://whmis.org/> (accessed 2024-06-07).
- (92) Harden, T. A.; Lloyd, J. A.; Turner, C. J. *Robotics for Nuclear Material Handling at LANL: Capabilities and Needs*. ASME IDETC/CIE Conference ; August 30, 2009; San Diego, CA; Los Alamos National Laboratory, 2009; <https://digital.library.unt.edu/ark:/67531/metadc934538/> (accessed 2024-06-07).
- (93) Salley, D.; Manzano, J. S.; Kitson, P. J.; Cronin, L. Robotic Modules for the Programmable Computation of Molecules and Materials. *ACS Cent. Sci.* **2023**, 9 (8), 1525–1537.
- (94) Yoshikawa, N.; Li, A. Z.; Darvish, K.; Zhao, Y.; Xu, H.; Kuramshin, A.; Aspuru-Guzik, A.; Garg, A.; Shkurti, F. Chemistry Lab Automation via Constrained Task and Motion Planning. *arXiv* 2023. [http://arxiv.org/abs/2212.09672](https://arxiv.org/abs/2212.09672) (accessed 2023-06-04). DOI: [10.48550/arXiv.2212.09672](https://doi.org/10.48550/arXiv.2212.09672)
- (95) Uhrin, M.; Huber, S. P.; Yu, J.; Marzari, N.; Pizzi, G. Workflows in AiiDA: Engineering a High-Throughput, Event-Based Engine for Robust and Modular Computational Workflows. *Comput. Mater. Sci.* **2021**, 187, 110086.
- (96) Jain, A.; Ong, S. P.; Chen, W.; Medasani, B.; Qu, X.; Kocher, M.; Brafman, M.; Petretto, G.; Rignanese, G.; Hautier, G.; et al. FireWorks: A Dynamic Workflow System Designed for High-throughput Applications. *Concurr. Comput. Pract. Exp.* **2015**, 27, S037–S059.
- (97) 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. Sustainable Data Analysis With Snakemake. *F1000Research* **2021**, 10, 33.
- (98) Alegre-Requena, J. V.; Sowndarya, S. V. S.; Pérez-Soto, R.; Alturaifi, T. M.; Paton, R. S. AQME: Automated Quantum Mechanical Environments for Researchers and Educators. *WIREs Comput. Mol. Sci.* **2023**, 13, No. e1663.
- (99) Guan, Y.; Ingman, V. M.; Rooks, B. J.; Wheeler, S. E. AARON: An Automated Reaction Optimizer for New Catalysts. *J. Chem. Theory Comput.* **2018**, 14, 5249–5261.
- (100) Żurański, A. M.; Wang, J. Y.; Shields, B. J.; Doyle, A. G. Auto-QChem: An Automated Workflow for the Generation and Storage of DFT Calculations for Organic Molecules. *React. Chem. Eng.* **2022**, 7, 1276–1284.
- (101) Rosales, A. R.; Wahlers, J.; Limé, E.; Meadows, R. E.; Leslie, K. W.; Savin, R.; Bell, F.; Hansen, E.; Helquist, P.; Munday, R. H.; et al. Rapid Virtual Screening of Enantioselective Catalysts Using CatVS. *Nat. Catal.* **2019**, 2, 41–45.
- (102) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell—a Modular Software Package for QM/MM Simulations. *WIREs Comput. Mol. Sci.* **2014**, 4, 101–110.
- (103) Ioannidis, E. I.; Gani, T. Z. H.; Kulik, H. J. molSimplify: A Toolkit for Automating Discovery in Inorganic Chemistry. *J. Comput. Chem.* **2016**, 37, 2106–2117.
- (104) Jacob, C. R.; Beyhan, S. M.; Bulo, R. E.; Gomes, A. S. P.; Götz, A. W.; Kiewisch, K.; Sikkema, J.; Visscher, L. PyADF - A Scripting Framework for Multiscale Quantum Chemistry. *J. Comput. Chem.* **2011**, 32, 2328–2338.
- (105) Zapata, F.; Ridder, L.; Hidding, J.; Jacob, C. R.; Infante, I.; Visscher, L. QMflows: A Tool Kit for Interoperable Parallel Workflows in Quantum Chemistry. *J. Chem. Inf. Model.* **2019**, 59, 3191–3197.
- (106) Lavigne, C.; Aspuru-Guzik, A. FunSies: A Minimalist, Distributed and Dynamic Workflow Engine. *J. Open Source Softw.* **2021**, 6, 3274.
- (107) Corbeil, C. R.; Thielges, S.; Schwartzenruber, J. A.; Moitessier, N. Toward a Computational Tool Predicting the Stereochemical Outcome of Asymmetric Reactions: Development and Application of a Rapid and Accurate Program Based on Organic Principles. *Angew. Chem. Int. Ed.* **2008**, 47, 2635–2638.
- (108) Seifrid, M.; Pollice, R.; Aguilar-Granda, A.; Morgan Chan, Z.; Hotta, K.; Ser, C. T.; Vestfrid, J.; Wu, T. C.; Aspuru-Guzik, A. Autonomous Chemical Experiments: Challenges and Perspectives on Establishing a Self-Driving Lab. *Acc. Chem. Res.* **2022**, 55, 2454–2466.
- (109) Bromig, L.; Leiter, D.; Mardale, A.-V.; von den Eichen, N.; Bieringer, E.; Weuster-Botz, D. The SiLA 2 Manager for Rapid Device Integration and Workflow Automation. *SoftwareX* **2022**, 17, 100991.
- (110) 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, No. e0229862.
- (111) Sim, M.; Vakili, M. G.; Strieth-Kalthoff, F.; Hao, H.; Hickman, R. J.; Miret, S.; Pablo-García, S.; Aspuru-Guzik, A. ChemOS 2.0: An Orchestration Architecture for Chemical Self-Driving Laboratories. *Matter* **2024**. DOI: [10.1016/j.matt.2024.04.022](https://doi.org/10.1016/j.matt.2024.04.022).
- (112) Rahmanian, 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.
- (113) Deneault, J. R.; Chang, J.; Myung, J.; Hooper, D.; Armstrong, A.; Pitt, M.; Maruyama, B. Toward Autonomous Additive Manufacturing: Bayesian Optimization on a 3D Printer. *MRS Bull.* **2021**, 46, 566–575.
- (114) Tamura, R.; Tsuda, K.; Matsuda, S. NIMS-OS: An Automation Software to Implement a Closed Loop between Artificial Intelligence and Robotic Experiments in Materials Science. *Sci. Technol. Adv. Mater. Methods* **2023**, 3, 2232297.
- (115) Kusne, A. G.; McDannald, A. Scalable Multi-Agent Lab Framework for Lab Optimization. *Matter* **2023**, 6, 1880–1893.
- (116) Fitzpatrick, D. E.; Maujean, T.; Evans, A. C.; Ley, S. V. Across-the-World Automated Optimization and Continuous-Flow Synthesis of Pharmaceutical Agents Operating Through a Cloud-Based Server. *Angew. Chem. Int. Ed.* **2018**, 57, 15128–15132.
- (117) Fitzpatrick, D. E.; Battilocchio, C.; Ley, S. V. A Novel Internet-Based Reaction Monitoring, Control and Autonomous Self-Optimization Platform for Chemical Synthesis. *Org. Process Res. Dev.* **2016**, 20, 386–394.
- (118) Leong, C. J.; Low, K. Y. A.; Recatala-Gomez, J.; Quijano Velasco, P.; Vissol-Gaudin, E.; Tan, J. D.; Ramalingam, B. I.; I Made, R.; Pethe, S. D.; Sebastian, S.; et al. An Object-Oriented Framework to Enable Workflow Evolution across Materials Acceleration Platforms. *Matter* **2022**, 5, 3124–3134.
- (119) Li, J.; Li, J.; Liu, R.; Tu, Y.; Li, Y.; Cheng, J.; He, T.; Zhu, X. Autonomous Discovery of Optically Active Chiral Inorganic Perovskite Nanocrystals through an Intelligent Cloud Lab. *Nat. Commun.* **2020**, 11, 2046.
- (120) Wierenga, R. P.; Golas, S. M.; Ho, W.; Coley, C. W.; Esveld, K. M. PyLabRobot: An Open-Source, Hardware-Agnostic Interface for Liquid-Handling Robots and Accessories. *Device* **2023**, 1, 100111.
- (121) Fei, Y.; Rendy, B.; Kumar, R.; Darts, O.; Sahasrabuddhe, H. P.; McDermott, M. J.; Wang, Z.; Szymanski, N. J.; Walters, L. N.; Milsted, D. et al. AlabOS: A Python-Based Reconfigurable Workflow Management Framework for Autonomous Laboratories. *arXiv* 2024. [http://arxiv.org/abs/2405.13930](https://arxiv.org/abs/2405.13930) (accessed 2024-06-03). DOI: [10.48550/arXiv.2405.13930](https://doi.org/10.48550/arXiv.2405.13930)

- (122) Guevarra, D.; Kan, K.; Lai, Y.; Jones, R. J. R.; Zhou, L.; Donnelly, P.; Richter, M.; Stein, H. S.; Gregoire, J. M. Orchestrating Nimble Experiments across Interconnected Labs. *Digit. Discov.* **2023**, *2*, 1806.
- (123) Prabhu, G. R. D.; Urban, P. L. Elevating Chemistry Research with a Modern Electronics Toolkit. *Chem. Rev.* **2020**, *120*, 9482–9553.
- (124) Haas, C. P.; Lübbesmeyer, M.; Jin, E. H.; McDonald, M. A.; Koscher, B. A.; Guimond, N.; Di Rocco, L.; Kayser, H.; Leweke, S.; Niedenführ, S.; et al. Open-Source Chromatographic Data Analysis for Reaction Optimization and Screening. *ACS Cent. Sci.* **2023**, *9*, 307–317.
- (125) Seifrid, M.; Strieth-Kalthoff, F.; Haddadnia, M.; Wu, T. C.; Alca, E.; Bodo, L.; Arellano-Rubach, S.; Yoshikawa, N.; Skreta, M.; Keunen, R.; et al. Chemspyd: An Open-Source Python Interface for Chemspeed Robotic Chemistry and Materials Platforms. *Digit. Discov.* **2024**, DOI: 10.1039/D4DD00046C.
- (126) 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.
- (127) Park, N. H.; Manica, M.; Born, J.; Hedrick, J. L.; Erdmann, T.; Zubarev, D. Y.; Adell-Mill, N.; Arrechea, P. L. Artificial Intelligence Driven Design of Catalysts and Materials for Ring Opening Polymerization Using a Domain-Specific Language. *Nat. Commun.* **2023**, *14*, 3686.
- (128) Skilton, R. A.; Bourne, R. A.; Amara, Z.; Horvath, R.; Jin, J.; Scully, M. J.; Streng, E.; Tang, S. L. Y.; Summers, P. A.; Wang, J.; et al. Remote-Controlled Experiments with Cloud Chemistry. *Nat. Chem.* **2015**, *7*, 1–5.
- (129) Caramelli, D.; Salley, D.; Henson, A.; Camarasa, G. A.; Sharabi, S.; Keenan, G.; Cronin, L. Networking Chemical Robots for Reaction Multitasking. *Nat. Commun.* **2018**, *9*, 3406.
- (130) Bai, J.; Cao, L.; Mosbach, S.; Akroyd, J.; Lapkin, A. A.; Kraft, M. From Platform to Knowledge Graph: Evolution of Laboratory Automation. *JACS Au.* **2022**, *2*, 292–309.
- (131) Bai, J.; Mosbach, S.; Taylor, C. J.; Karan, D.; Lee, K. F.; Rihm, S. D.; Akroyd, J.; Lapkin, A. A.; Kraft, M. A Dynamic Knowledge Graph Approach to Distributed Self-Driving Laboratories. *Nat. Commun.* **2024**, *15*, 462.
- (132) Vogler, M.; Busk, J.; Hajiyani, H.; Jørgensen, P. B.; Safaei, N.; Castelli, I. E.; Ramirez, F. F.; Carlsson, J.; Pizzi, G.; Clark, S.; et al. Brokering between Tenants for an International Materials Acceleration Platform. *Matter.* **2023**, *6*, 2647–2665.
- (133) Taylor, R.; Kardas, M.; Cucurull, G.; Scialom, T.; Hartshorn, A.; Saravia, E.; Poulton, A.; Kerkez, V.; Stojnic, R. Galactica: A Large Language Model for Science. *arXiv2022*. DOI: 10.48550/arXiv.2211.09085 (accessed October 10, 2023).
- (134) Jablonka, K. M.; Ai, Q.; Al-Feghali, A.; Badhwar, S.; Bocarsly, J. D.; Bran, A. M.; Bringquier, S.; Brinson, L. C.; Choudhary, K.; Circi, D.; et al. 14 Examples of How LLMs Can Transform Materials Science and Chemistry: A Reflection on a Large Language Model Hackathon. *Digit. Discov.* **2023**, *2*, 1233–1250.
- (135) Guo, T.; Guo, K.; Nan, B.; Liang, Z.; Guo, Z.; Chawla, N. V.; Wiest, O.; Zhang, X. What Can Large Language Models Do in Chemistry? A Comprehensive Benchmark on Eight Tasks. *arXiv2023*. DOI: 10.48550/arXiv.2305.18365 (accessed October 10, 2023).
- (136) Hatakeyama-Sato, K.; Yamane, N.; Igarashi, Y.; Nabae, Y.; Hayakawa, T. Prompt Engineering of GPT-4 for Chemical Research: What Can/Cannot Be Done? *Sci. Technol. Adv. Mater. Methods.* **2023**, *3*, 2260300.
- (137) Boiko, D. A.; MacKnight, R.; Kline, B.; Gomes, G. Autonomous Chemical Research with Large Language Models. *Nature* **2023**, *624*, 570–578.
- (138) Bran, A.; Cox, S.; Schilter, O.; Baldassari, C.; White, A. D.; Schwaller, P. Augmenting Large Language Models with Chemistry Tools. *Nat. Mach. Intell.* **2024**, *6*, 525–535.
- (139) Yoshikawa, N.; Skreta, M.; Darvish, K.; Arellano-Rubach, S.; Ji, Z.; Bjørn Kristensen, L.; Li, A. Z.; Zhao, Y.; Xu, H.; Kuramshin, A.; et al. Large Language Models for Chemistry Robotics. *Auton. Robots.* **2023**, *47*, 1057.
- (140) Darvish, K.; Skreta, M.; Zhao, Y.; Yoshikawa, N.; Som, S.; Bogdanovic, M.; Cao, Y.; Hao, H.; Xu, H.; Aspuru-Guzik, A. et al. ORGANA: A Robotic Assistant for Automated Chemistry Experimentation and Characterization. *arXiv2024*. <http://arxiv.org/abs/2401.06949> (accessed 2024-01-17). DOI: 10.48550/arXiv.2401.06949
- (141) CARPi, N.; Minges, A.; Piel, M. eLabFTW: An Open Source Laboratory Notebook for Research Labs. *JOSS* **2017**, *2*, 146.
- (142) Barillari, C.; Ottoz, D. S. M.; Fuentes-Serna, J. M.; Ramakrishnan, C.; Rinn, B.; Rudolf, F. openBIS ELN-LIMS: An Open-Source Database for Academic Laboratories. *Bioinformatics* **2016**, *32*, 638–640.
- (143) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, Ij. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data.* **2016**, *3*, 160018.
- (144) Zenodo - Research. Shared. <https://zenodo.org/> (accessed 2023-09-22).
- (145) Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrénk, S.; Amador-Bedolla, C.; Sánchez-Carrera, R. S.; Gold-Parker, A.; Vogt, L.; Brockway, A. M.; Aspuru-Guzik, A. The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid. *J. Phys. Chem. Lett.* **2011**, *2*, 2241–2251.
- (146) Álvarez-Moreno, M.; de Graaf, C.; López, N.; Maseras, F.; Poblet, J. M.; Bo, C. Managing the Computational Chemistry Big Data Problem: The ioChem-BD Platform. *J. Chem. Inf. Model.* **2015**, *55*, 95–103.
- (147) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1*, 011002.
- (148) Scheffler, M.; Aeschlimann, M.; Albrecht, M.; Bereau, T.; Bungartz, H.-J.; Felser, C.; Greiner, M.; Groß, A.; Koch, C. T.; Kremer, K.; et al. FAIR Data Enabling New Horizons for Materials Research. *Nature* **2022**, *604*, 635–642.
- (149) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (150) The Materials Cloud Team. *Materials Cloud*. <https://www.materialscloud.org/> (accessed 2023-10-10).
- (151) Bobbitt, N. S.; Shi, K.; Bucior, B. J.; Chen, H.; Tracy-Amoroso, N.; Li, Z.; Sun, Y.; Merlin, J. H.; Siepmann, J. I.; Siderius, D. W.; et al. MOFX-DB: An Online Database of Computational Adsorption Data for Nanoporous Materials. *J. Chem. Eng. Data.* **2023**, *68*, 483–498.
- (152) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501–1509.
- (153) Winther, K. T.; Hoffmann, M. J.; Boes, J. R.; Mamun, O.; Bajdich, M.; Bligaard, T. Catalysis-Hub.Org, an Open Electronic Structure Database for Surface Reactions. *Sci. Data.* **2019**, *6*, 75.
- (154) AIST. *AIST:Spectral Database for Organic Compounds, SDBS*. [https://sdbs.db.aist.go.jp/sdbs/cgi-bin/direct\\_frame\\_top.cgi](https://sdbs.db.aist.go.jp/sdbs/cgi-bin/direct_frame_top.cgi) (accessed 2023-10-10).
- (155) Reaxys. <https://www.reaxys.com/> (accessed 2023-10-10).
- (156) John Wiley & Sons, Inc. *SpectraBase*. <https://spectrabase.com/> (accessed 2023-10-10).
- (157) CAS SciFinder - Chemical Compound Database. <https://www.cas.org/solutions/cas-scifinder-discovery-platform/cas-scifinder> (accessed 2024-07-19).
- (158) Lowe, D. Chemical Reactions from US Patents 2017, 1494665893 Bytes. DOI: 10.6084/M9.FIGSHARE.5104873.V1.

- (159) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem 2023 Update. *Nucleic Acids Res.* **2023**, *51*, D1373–D1380.
- (160) Kearnes, S. M.; Maser, M. R.; Wleklinski, M.; Kast, A.; Doyle, A. G.; Dreher, S. D.; Hawkins, J. M.; Jensen, K. F.; Coley, C. W. The Open Reaction Database. *J. Am. Chem. Soc.* **2021**, *143*, 18820–18826.
- (161) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; et al. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837.
- (162) *MassBank of North America*. <https://mona.fiehnlab.ucdavis.edu/> (accessed 2023-10-10).
- (163) Gražulis, S.; Daškevič, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quirós, M.; Serebryanaya, N. R.; Moeck, P.; Downs, R. T.; Le Bail, A. Crystallography Open Database (COD): An Open-Access Collection of Crystal Structures and Platform for World-Wide Collaboration. *Nucleic Acids Res.* **2012**, *40*, D420–D427.
- (164) Kuhn, S.; Schlörer, N. E. Facilitating Quality Control for Spectra Assignments of Small Organic Molecules: Nmrshiftdb2 - a Free in-House NMR Database with Integrated LIMS for Academic Service Laboratories. *Magn. Reson. Chem.* **2015**, *53*, 582–589.
- (165) Gaudin, T.; Benlolo, I.; Cui, Z. Y.; Hickmann, R.; Tamblyn, I.; Aspuru-Guzik, A. Molar. *Zenodo*2022. DOI: [10.5281/zenodo.6809291](https://doi.org/10.5281/zenodo.6809291).
- (166) Hugging Face - The AI community building the future. <https://huggingface.co/> (accessed 2023-09-28).
- (167) Walters, W. P.; Barzilay, R. Applications of Deep Learning in Molecule Generation and Molecular Property Prediction. *Acc. Chem. Res.* **2021**, *54*, 263–270.
- (168) Varnek, A.; Baskin, I. I. Chemoinformatics as a Theoretical Chemistry Discipline. *Mol. Inform.* **2011**, *30*, 20–32.
- (169) Capecchi, A.; Probst, D.; Reymond, J.-L. One Molecular Fingerprint to Rule Them All: Drugs, Biomolecules, and the Metabolome. *J. Cheminformatics.* **2020**, *12*, 43.
- (170) Zagidullin, B.; Wang, Z.; Guan, Y.; Pitkänen, E.; Tang, J. Comparative Analysis of Molecular Fingerprints in Prediction of Drug Combination Effects. *Brief. Bioinform.* **2021**, *22*, bbab291.
- (171) Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv*2017. <http://arxiv.org/abs/1705.07874> (accessed 2023-11-20). DOI: [10.48550/arXiv.1705.07874](https://doi.org/10.48550/arXiv.1705.07874)
- (172) Rodríguez-Pérez, R.; Bajorath, J. Feature Importance Correlation from Machine Learning Indicates Functional Relationships between Proteins and Similar Compound Binding Characteristics. *Sci. Rep.* **2021**, *11*, 14245.
- (173) Ho, T. K. Random Decision Forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*; 1995; Vol. 1, pp 278–282 vol.1. DOI: [10.1109/ICDAR.1995.598994](https://doi.org/10.1109/ICDAR.1995.598994).
- (174) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232.
- (175) Cho, Y.; Saul, L. Kernel Methods for Deep Learning. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc., 2009; Vol. 22.
- (176) Williams, C. K.; Rasmussen, C. E. *Gaussian Processes for Machine Learning*; MIT press Cambridge, MA, 2006; Vol. 2.
- (177) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297.
- (178) Platt, J. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Advances in Large-Margin Classifiers*; MIT Press, 2000.
- (179) Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory.* **1967**, *13*, 21–27.
- (180) MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*; University of California Press, 1967; Vol. 5.1, pp 281–298.
- (181) Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R. DRAGON Software: An Easy Approach to Molecular Descriptor Calculations. *MATCH Commun. Math. Comput. Chem.* **2006**, *56*, 237–248.
- (182) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *J. Cheminformatics.* **2018**, *10*, 4.
- (183) Himanen, L.; Jäger, M. O. J.; Morooka, E. V.; Federici Canova, F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. DSCRIBE: Library of Descriptors for Machine Learning in Materials Science. *Comput. Phys. Commun.* **2020**, *247*, 106949.
- (184) Landrum, G.; Tosco, P.; Kelley, B.; Ric; Cosgrove, D.; Sriniker; Gedeck; Vianello, R.; Schneider, N.; Kawashima, E. et al. Rdkit/Rdkit: 2023\_09\_1 (Q3 2023) Release Beta, *Zenodo*2023. DOI: [10.5281/ZENODO.591637](https://doi.org/10.5281/ZENODO.591637) (accessed October 6, 2023).
- (185) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113.
- (186) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
- (187) Duvenaud, D.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Gómez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. *arXiv*2015. DOI: [10.48550/arXiv.1509.09292](https://doi.org/10.48550/arXiv.1509.09292) (accessed October 23, 2023).
- (188) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
- (189) Krenn, M.; Häse, F.; Nigam, A.; Friederich, P.; Aspuru-Guzik, A. Self-Referencing Embedded Strings (SELFIES): A 100% Robust Molecular String Representation. *Mach. Learn. Sci. Technol.* **2020**, *1*, 045024.
- (190) Cheng, A.; Cai, A.; Miret, S.; Malkomes, G.; Philipp, M.; Aspuru-Guzik, A. Group SELFIES: A Robust Fragment-Based Molecular String Representation. *Digit. Discov.* **2023**, *2*, 748–758.
- (191) David, L.; Thakkar, A.; Mercado, R.; Engkvist, O. Molecular Representations in AI-Driven Drug Discovery: A Review and Practical Guide. *J. Cheminformatics.* **2020**, *12*, S6.
- (192) Krenn, M.; Ai, Q.; Barthel, S.; Carson, N.; Frei, A.; Frey, N. C.; Friederich, P.; Gaudin, T.; Gayle, A. A.; Jablonka, K. M.; et al. SELFIES and the Future of Molecular String Representations. *Patterns.* **2022**, *3*, 100588.
- (193) Daylight Theory: SMIRKS - A Reaction Transform Language. <https://www.daylight.com/dayhtml/doc/theory/theory.smirks.html> (accessed 2023-10-23).
- (194) Daylight Theory: SMARTS - A Language for Describing Molecular Patterns. <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html> (accessed 2023-10-23).
- (195) Kim, Y.; Kim, J. W.; Kim, Z.; Kim, W. Y. Efficient Prediction of Reaction Paths through Molecular Graph and Reaction Network Analysis. *Chem. Sci.* **2018**, *9*, 825–835.
- (196) Garay-Ruiz, D.; Bo, C. Chemical Reaction Network Knowledge Graphs: The OntoRXN Ontology. *J. Cheminformatics.* **2022**, *14*, 29.
- (197) Hashemi, A.; Bougueroua, S.; Gaigeot, M.-P.; Pidko, E. A. ReNeGate: A Reaction Network Graph-Theoretical Tool for Automated Mechanistic Studies in Computational Homogeneous Catalysis. *J. Chem. Theory Comput.* **2022**, *18*, 7470–7482.
- (198) McDermott, M. J.; Dwaraknath, S. S.; Persson, K. A. A Graph-Based Network for Predicting Chemical Reaction Pathways in Solid-State Materials Synthesis. *Nat. Commun.* **2021**, *12*, 3097.
- (199) Keith, J. A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K.-R.; Tkatchenko, A. Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. *Chem. Rev.* **2021**, *121*, 9816–9872.
- (200) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature.* **2018**, *559*, 547–555.
- (201) Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.; et al. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* **2019**, *59*, 3370–3388.

- (202) Sanchez-Lengeling, B.; Reif, E.; Pearce, A.; Wiltschko, A. B. A Gentle Introduction to Graph Neural Networks. *Distill.* **2021**, *6*, No. e33.
- (203) Satorras, V. G.; Hoogeboom, E.; Welling, M. E(n) Equivariant Graph Neural Networks. *arXiv*, 2022. <http://arxiv.org/abs/2102.09844> (accessed 2023-11-20). DOI: [10.48550/arXiv.2102.09844](https://doi.org/10.48550/arXiv.2102.09844)
- (204) Rampášek, L.; Galkin, M.; Dwivedi, V. P.; Luu, A. T.; Wolf, G.; Beaini, D. Recipe for a General, Powerful, Scalable Graph Transformer. *arXiv* 2023. <http://arxiv.org/abs/2205.12454> (accessed 2023-11-20). DOI: [10.48550/arXiv.2205.12454](https://doi.org/10.48550/arXiv.2205.12454)
- (205) Zheng, Z.; Zhang, O.; Nguyen, H. L.; Rampal, N.; Alawadhi, A. H.; Rong, Z.; Head-Gordon, T.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. ChatGPT Research Group for Optimizing the Crystallinity of MOFs and COFs. *ACS Cent. Sci.* **2023**, *9*, 2161.
- (206) Ross, J.; Belgodere, B.; Chenthamarakshan, V.; Padhi, I.; Mroueh, Y.; Das, P. Large-Scale Chemical Language Representations Capture Molecular Structure and Properties. *Nat. Mach. Intell.* **2022**, *4*, 1256–1264.
- (207) Dong, Q.; Cole, J. M. Snowball 2.0: Generic Material Data Parser for ChemDataExtractor. *J. Chem. Inf. Model.* **2023**, *63*, 7045.
- (208) Kononova, O.; Huo, H.; He, T.; Rong, Z.; Botari, T.; Sun, W.; Tshitoyan, V.; Ceder, G. Text-Mined Dataset of Inorganic Materials Synthesis Recipes. *Sci. Data.* **2019**, *6*, 203.
- (209) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4*, 268–276.
- (210) Jin, W.; Barzilay, R.; Jaakkola, T. Junction Tree Variational Autoencoder for Molecular Graph Generation. In *Proceedings of the 35th International Conference on Machine Learning*; PMLR, 2018; pp 2323–2332.
- (211) De Cao, N.; Kipf, T. MolGAN: An Implicit Generative Model for Small Molecular Graphs. *arXiv* 2022. DOI: [10.48550/arXiv.1805.11973](https://doi.org/10.48550/arXiv.1805.11973) (accessed November 20, 2023).
- (212) Sanchez-Lengeling, B.; Outeiral, C.; Guimaraes, G. L. Optimizing Distributions over Molecular Space. An Objective-Reinforced Generative Adversarial Network for Inverse-Design Chemistry (ORGANIC). *ChemRxiv*, 2017. DOI: [10.26434/chemrxiv.5309668.v2](https://doi.org/10.26434/chemrxiv.5309668.v2)
- (213) Bengio, E.; Jain, M.; Korablyov, M.; Precup, D.; Bengio, Y. Flow Network Based Generative Models for Non-Iterative Diverse Candidate Generation; *arXiv* 2021. DOI: [10.48550/arXiv.2106.04399](https://doi.org/10.48550/arXiv.2106.04399)
- (214) Bengio, Y.; Lahou, S.; Deleu, T.; Hu, E. J.; Tiwari, M.; Bengio, E. GFlowNet Foundations. *J. Mach. Learn. Res.* **2023**, *24*, 1–55.
- (215) Nigam, A.; Pollice, R.; Aspuru-Guzik, A. Parallel Tempered Genetic Algorithm Guided by Deep Neural Networks for Inverse Molecular Design. *Digit. Discov.* **2022**, *1*, 390–404.
- (216) Nigam, A.; Friederich, P.; Krenn, M.; Aspuru-Guzik, A. Augmenting Genetic Algorithms with Deep Neural Networks for Exploring the Chemical Space. 2020. <http://arxiv.org/abs/1909.11655> (accessed 2023-11-20).
- (217) Gao, W.; Fu, T.; Sun, J.; Coley, C. W. Sample Efficiency Matters: A Benchmark for Practical Molecular Optimization. 2022. <http://arxiv.org/abs/2206.12411> (accessed 2023-11-20).
- (218) Brown, N.; Fiscato, M.; Segler, M. H. S.; Vaucher, A. C. GuacaMol: Benchmarking Models for de Novo Molecular Design. *J. Chem. Inf. Model.* **2019**, *59*, 1096–1108.
- (219) Blaschke, T.; Arús-Pous, J.; Chen, H.; Margreitter, C.; Tyrchan, C.; Engkvist, O.; Papadopoulos, K.; Patronov, A. REINVENT 2.0: An AI Tool for De Novo Drug Design. *J. Chem. Inf. Model.* **2020**, *60*, 5918–5922.
- (220) Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular De Novo Design through Deep Reinforcement Learning. *arXiv*, 2017. DOI: [10.1186/s13321-017-0235-x](https://doi.org/10.1186/s13321-017-0235-x) (accessed November 20, 2023).
- (221) Zhou, Z.; Kearnes, S.; Li, L.; Zare, R. N.; Riley, P. Optimization of Molecules via Deep Reinforcement Learning. *Sci. Rep.* **2019**, *9*, 10752.
- (222) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018**, *361*, 360–365.
- (223) Nigam, A.; Pollice, R.; Tom, G.; Jorner, K.; Willes, J.; Thiede, L.; Kundaje, A.; Aspuru-Guzik, A. Tartarus: A Benchmarking Platform for Realistic And Practical Inverse Molecular Design. *arXiv* 2023. DOI: [10.48550/arXiv.2209.12487](https://doi.org/10.48550/arXiv.2209.12487)
- (224) Zhavoronkov, A.; Ivanenkov, Y. A.; Aliper, A.; Veselov, M. S.; Aladinskiy, V. A.; Aladinskaya, A. V.; Terentiev, V. A.; Polykovskiy, D. A.; Kuznetsov, M. D.; Asadulaev, A.; et al. Deep Learning Enables Rapid Identification of Potent DDR1 Kinase Inhibitors. *Nat. Biotechnol.* **2019**, *37*, 1038–1040.
- (225) Polykovskiy, D.; Zhebrak, A.; Sanchez-Lengeling, B.; Golovanov, S.; Tatanov, O.; Belyaev, S.; Kurbanov, R.; Artamonov, A.; Aladinskiy, V.; Veselov, M. et al. Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models. *Front. Pharmacol.* **2020**, *11*. DOI: [10.3389/fphar.2020.565644](https://doi.org/10.3389/fphar.2020.565644).
- (226) King-Smith, E. Transfer Learning for a Foundational Chemistry Model. *ChemRxiv* 2023. DOI: [10.26434/chemrxiv-2023-gnzpf](https://doi.org/10.26434/chemrxiv-2023-gnzpf) (accessed June 17, 2024)
- (227) Loeffler, H.; He, J.; Tibo, A.; Janet, J. P.; Voronov, A.; Mervin, L.; Engkvist, O. REINVENT4: Modern AI-Driven Generative Molecule Design. *ChemRxiv* 2023. DOI: [10.26434/chemrxiv-2023-xt65x](https://doi.org/10.26434/chemrxiv-2023-xt65x) (accessed June 17, 2024)
- (228) Sanchez-Lengeling, B.; Wei, J. N.; Lee, B. K.; Gerkin, R. C.; Aspuru-Guzik, A.; Wiltschko, A. B. Machine Learning for Scent: Learning Generalizable Perceptual Representations of Small Molecules. *arXiv* 2019. DOI: [10.48550/arXiv.1910.10685](https://doi.org/10.48550/arXiv.1910.10685) (accessed November 20, 2023).
- (229) Eyke, N. S.; Koscher, B. A.; Jensen, K. F. Toward Machine Learning-Enhanced High-Throughput Experimentation. *Trends Chem.* **2021**, *3*, 120–132.
- (230) Bender, A.; Schneider, N.; Segler, M.; Patrick Walters, W.; Engkvist, O.; Rodrigues, T. Evaluation Guidelines for Machine Learning Tools in the Chemical Sciences. *Nat. Rev. Chem.* **2022**, *6*, 428–442.
- (231) Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. Interpretable Machine Learning for Knowledge Generation in Heterogeneous Catalysis. *Nat. Catal.* **2022**, *5*, 175–184.
- (232) Mou, T.; Pillai, H. S.; Wang, S.; Wan, M.; Han, X.; Schweitzer, N. M.; Che, F.; Xin, H. Bridging the Complexity Gap in Computational Heterogeneous Catalysis with Machine Learning. *Nat. Catal.* **2023**, *6*, 122–136.
- (233) Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K. Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catal.* **2020**, *10*, 2260–2297.
- (234) Oliveira, J. C. A.; Frey, J.; Zhang, S.-Q.; Xu, L.-C.; Li, X.; Li, S.-W.; Hong, X.; Ackermann, L. When Machine Learning Meets Molecular Synthesis. *Trends Chem.* **2022**, *4*, 863–885.
- (235) Dara, S.; Dhamercherla, S.; Jadav, S. S.; Babu, C. M.; Ahsan, M. J. Machine Learning in Drug Discovery: A Review. *Artif. Intell. Rev.* **2022**, *55*, 1947–1999.
- (236) Bergstra, J.; Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305 <http://jmlr.org/papers/v13/bergstra12a.html>.
- (237) Bergstra, J.; Bardenet, R.; Bengio, Y.; Kégl, B. Algorithms for Hyper-Parameter Optimization. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc., 2011; Vol. 24. [https://papers.nips.cc/paper\\_files/paper/2011/hash/86e8f7ab32cfcd12577bc2619bc635690-Abstract.html](https://papers.nips.cc/paper_files/paper/2011/hash/86e8f7ab32cfcd12577bc2619bc635690-Abstract.html).
- (238) Haber, F. The Synthesis of Ammonia from Its Elements Nobel Lecture, June 2, 1920. *Resonance* **2002**, *7*, 86–94.
- (239) Chen, S.; Perathoner, S.; Ampelli, C.; Centi, G. Electrochemical Dinitrogen Activation: To Find a Sustainable Way to Produce Ammonia. In *Studies in Surface Science and Catalysis*; Elsevier, 2019; Vol. 178, pp 31–46. DOI: [10.1016/B978-0-444-64127-4.00002-1](https://doi.org/10.1016/B978-0-444-64127-4.00002-1).
- (240) Anderson, M. J.; Whitcomb, P. J. *DOE Simplified: Practical Tools for Effective Experimentation*, 3rd ed.; CRC Press, 2017.

- (241) Fisher, R. A. *The Design of Experiments*, 7th ed.; Oliver & Boyd, 1960.
- (242) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*; Wiley series in probability and mathematical statistics; Wiley: New York, 1978.
- (243) Brereton, R. G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; John Wiley & Sons, 2003.
- (244) Student. The Probable Error of a Mean. *Biometrika*. **1908**, *6*, 1–25.
- (245) Girden, E. R. *ANOVA: Repeated Measures*; Sage, 1992.
- (246) Sthle, L.; Wold, S. Analysis of Variance (ANOVA). *Chemom. Intell. Lab. Syst.* **1989**, *6*, 259–272.
- (247) Holland, J. H. Genetic Algorithms. *Sci. Am.* **1992**, *267*, 66–73.
- (248) Booker, L. B.; Goldberg, D. E.; Holland, J. H. Classifier Systems and Genetic Algorithms. *Artif. Intell.* **1989**, *40*, 235–282.
- (249) Kennedy, J.; Eberhart, R. Particle Swarm Optimization. In *Proceedings of ICNN'95-international conference on neural networks*; IEEE, 1995; Vol. 4, pp 1942–1948. DOI: [10.1109/ICNN.1995.488968](https://doi.org/10.1109/ICNN.1995.488968)
- (250) Shi, Y.; Eberhart, R. A Modified Particle Swarm Optimizer. In *1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360)*; IEEE, 1998; pp 69–73. DOI: [10.1109/ICEC.1998.699146](https://doi.org/10.1109/ICEC.1998.699146)
- (251) Clerc, M.; Kennedy, J. The Particle Swarm-Explosion, Stability, and Convergence in a Multidimensional Complex Space. *IEEE Trans. Evol. Comput.* **2002**, *6*, 58–73.
- (252) Jastrebski, G. A.; Arnold, D. V. Improving Evolution Strategies through Active Covariance Matrix Adaptation. In *2006 IEEE International Conference on Evolutionary Computation*; IEEE, 2006; pp 2814–2821. DOI: [10.1109/CEC.2006.1688662](https://doi.org/10.1109/CEC.2006.1688662).
- (253) Nelder, J. A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313.
- (254) Bezerra, M. A.; dos Santos, Q. O.; Santos, A. G.; Novaes, C. G.; Ferreira, S. L. C.; de Souza, V. S. Simplex Optimization: A Tutorial Approach and Recent Applications in Analytical Chemistry. *Microchem. J.* **2016**, *124*, 45–54.
- (255) Pasamontes, A.; Callao, P. Fractional Factorial Design and Simplex Algorithm for Optimizing Sequential Injection Analysis (SIA) and Second Order Calibration. *Chemom. Intell. Lab. Syst.* **2006**, *83*, 127–132.
- (256) Xiong, Q.; Jutan, A. Continuous Optimization Using a Dynamic Simplex Method. *Chem. Eng. Sci.* **2003**, *58*, 3817–3828.
- (257) Berridge, J. C. Automated Multiparameter Optimisation of High-Performance Liquid Chromatographic Separations Using the Sequential Simplex Procedure. *Analyst*. **1984**, *109*, 291–293.
- (258) Berridge, J. C. Unattended Optimisation of Reversed-Phase High-Performance Liquid Chromatographic Separations Using the Modified Simplex Algorithm. *J. Chromatogr. A*. **1982**, *244*, 1–14.
- (259) Matsuda, R.; Ishibashi, M.; Takeda, Y. Simplex Optimization of Reaction Conditions with an Automated System. *Chem. Pharm. Bull. (Tokyo)*. **1988**, *36*, 3512–3518.
- (260) Matsuda, R.; Ishibashi, M.; Uchiyama, M. Simplex Optimization of Reaction Condition Using Laboratory Robotic System. *Yakugaku Zasshi*. **1987**, *107*, 683–689.
- (261) Watson, M. W.; Carr, P. W. Simplex Algorithm for the Optimization of Gradient Elution High-Performance Liquid Chromatography. *Anal. Chem.* **1979**, *51*, 1835–1842.
- (262) Wright, A. G.; Fell, A. F.; Berridge, J. C. Sequential Simplex Optimization and Multichannel Detection in HPLC: Application to Method Development. *Chromatographia*. **1987**, *24*, 533–540.
- (263) MacConnell, A. B.; Price, A. K.; Paegel, B. M. An Integrated Microfluidic Processor for DNA-Encoded Combinatorial Library Functional Screening. *ACS Comb. Sci.* **2017**, *19*, 181–192.
- (264) Huyer, W.; Neumaier, A. SNOBFIT - Stable Noisy Optimization by Branch and Fit. *ACM Trans. Math. Softw.* **2008**, *35*, 1–25.
- (265) Bédard, A.-C.; Adamo, A.; Aroh, K. C.; Russell, M. G.; Bedermann, A. A.; Torosian, J.; Yue, B.; Jensen, K. F.; Jamison, T. F. Reconfigurable System for Automated Optimization of Diverse Chemical Reactions. *Science*. **2018**, *361*, 1220–1225.
- (266) Walker, B. E.; Bannock, J. H.; Nightingale, A. M.; deMello, J. C. Tuning Reaction Products by Constrained Optimisation. *React. Chem. Eng.* **2017**, *2*, 785–798.
- (267) Hestenes, M. R.; Stiefel, E. Methods of Conjugate Gradients for Solving Linear Systems. *J. Res. Natl. Bur. Stand.* **1952**, *49*, 409–436.
- (268) Lucia, A.; Xu, J. Chemical Process Optimization Using Newton-like Methods. *Comput. Chem. Eng.* **1990**, *14*, 119–138.
- (269) McMullen, J. P.; Jensen, K. F. An Automated Microfluidic System for Online Optimization in Chemical Synthesis. *Org. Process Res. Dev.* **2010**, *14*, 1169–1176.
- (270) Pan, S. J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359.
- (271) Tighineanu, P.; Skubch, K.; Baireuther, P.; Reiss, A.; Berkenkamp, F.; Vinogradsk, J. Transfer Learning with Gaussian Processes for Bayesian Optimization. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*; PMLR, 2022; pp 6152–6181.
- (272) Fare, C.; Fenner, P.; Benatan, M.; Varsi, A.; Pyzer-Knapp, E. O. A Multi-Fidelity Machine Learning Approach to High Throughput Materials Screening. *Npj Comput. Mater.* **2022**, *8*, 1–9.
- (273) Swersky, K.; Snoek, J.; Adams, R. P. Multi-Task Bayesian Optimization. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc., 2013; Vol. 26.
- (274) Hickman, R. J.; Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Gemini: Dynamic Bias Correction for Autonomous Experimentation and Molecular Simulation. *arXiv2021*. DOI: [10.48550/arXiv.2103.03391](https://doi.org/10.48550/arXiv.2103.03391) (accessed June 11, 2024).
- (275) Tran, A.; Wildey, T.; McCann, S. sMF-BO-2CoGP: A Sequential Multi-Fidelity Constrained Bayesian Optimization Framework for Design Applications. *J. Comput. Inf. Sci. Eng.* **2020**, *20*. DOI: [10.1115/1.4046697](https://doi.org/10.1115/1.4046697).
- (276) Tran, A.; Tranchida, J.; Wildey, T.; Thompson, A. P. Multi-Fidelity Machine-Learning with Uncertainty Quantification and Bayesian Optimization for Materials Design: Application to Ternary Random Alloys. *J. Chem. Phys.* **2020**, *153*, 074705.
- (277) Robbins, H. Some Aspects of Sequential Design of Experiments. *Bull. Am. Math. Soc.* **1952**, *58*, 527–535.
- (278) Kuleshov, V.; Precup, D. Algorithms for the Multi-Armed Bandit Problem. *arXiv2014*. DOI: [10.48550/arXiv.1402.6028](https://doi.org/10.48550/arXiv.1402.6028) (accessed June 13, 2024).
- (279) Kikkawa, N.; Ohno, H. Materials Discovery Using Max K-Armed Bandit. *J. Mach. Learn. Res.* **2024**, *25*, 140.
- (280) 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.
- (281) Knowles, J. ParEGO: A Hybrid Algorithm with on-Line Landscape Approximation for Expensive Multiobjective Optimization Problems. *IEEE Trans. Evol. Comput.* **2006**, *10*, 50–66.
- (282) Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197.
- (283) Deb, K.; Jain, H. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints. *IEEE Trans. Evol. Comput.* **2014**, *18*, 577–601.
- (284) Beume, N.; Naujoks, B.; Emmerich, M. SMS-EMOA: Multiobjective Selection Based on Dominated Hypervolume. *Eur. J. Oper. Res.* **2007**, *181*, 1653–1669.
- (285) Sharma, S.; Kumar, V. A Comprehensive Review on Multi-Objective Optimization Techniques: Past, Present and Future. *Arch. Comput. Methods Eng.* **2022**, *29*, 5605–5633.
- (286) Rangaiah, G. P.; Feng, Z.; Hoadley, A. F. Multi-Objective Optimization Applications in Chemical Process Engineering: Tutorial and Review. *Processes*. **2020**, *8*, 508.

- (287) Senthil Vel, A.; Cortes-Borda, D.; Felpin, F.-X. A Chemist's Guide to Multi-Objective Optimization Solvers for Reaction Optimization. *React. Chem. Eng.* **2024**, DOI: 10.1039/D4RE00175C.
- (288) Angelo, J. S.; Guedes, I. A.; Barbosa, H. J. C.; Dardenne, L. E. Multi-and Many-Objective Optimization: Present and Future in de Novo Drug Design. *Front. Chem.* **2023**, 11. DOI: 10.3389/fchem.2023.1288626.
- (289) Mockus, J. *Bayesian Approach to Global Optimization: Theory and Applications*; Springer Netherlands: Dordrecht, 1989.
- (290) Di Fiore, F.; Nardelli, M.; Mainini, L. Active Learning and Bayesian Optimization: A Unified Perspective to Learn with a Goal. *arXiv*2023. <http://arxiv.org/abs/2303.01560> (accessed 2023-11-11). DOI: 10.48550/arXiv.2303.01560
- (291) Daulton, S.; Balandat, M.; Bakshy, E. Differentiable Expected Hypervolume Improvement for Parallel Multi-Objective Bayesian Optimization. *arXiv*2020. DOI: 10.48550/arXiv.2006.05078 (accessed November 25, 2023).
- (292) Daulton, S.; Balandat, M.; Bakshy, E. Parallel Bayesian Optimization of Multiple Noisy Objectives with Expected Hypervolume Improvement. *arXiv*2021. DOI: 10.48550/arXiv.2105.08195 (accessed June 7, 2024).
- (293) Neal, R. M. *Bayesian Learning for Neural Networks*; Springer Science & Business Media, 2012; Vol. 118.
- (294) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian Reaction Optimization as a Tool for Chemical Synthesis. *Nature*. **2021**, 590, 89–96.
- (295) Torres, J. A. G.; Lau, S. H.; Anchuri, P.; Stevens, J. M.; Tabora, J. E.; Li, J.; Borovika, A.; Adams, R. P.; Doyle, A. G. A Multi-Objective Active Learning Platform and Web App for Reaction Optimization. *J. Am. Chem. Soc.* **2022**, 144, 19999–20007.
- (296) Häse, F.; Roch, L. M.; Kreisbeck, C.; Aspuru-Guzik, A. Phoenics: A Bayesian Optimizer for Chemistry. *ACS Cent. Sci.* **2018**, 4, 1134–1145.
- (297) 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.
- (298) Perone, C. S. Pyevolve: A Python Open-Source Framework for Genetic Algorithms. *ACM SIGEVolution*. **2009**, 4, 12–20.
- (299) Hutter, F.; Hoos, H. H.; Leyton-Brown, K. Sequential Model-Based Optimization for General Algorithm Configuration. In *Learning and Intelligent Optimization*; Coello, C. A. C., Ed.; Lecture Notes in Computer Science; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; Vol. 6683, pp 507–523. DOI: 10.1007/978-3-642-25566-3\_40.
- (300) Bergstra, J.; Yamins, D.; Cox, D. D. Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms. In *Proceedings of the 12th Python in science conference*; SciPy, 2013; Vol. 13, p 20. DOI: 10.25080/Majora-8b375195-003
- (301) The GPyOpt Authors. GPyOpt: A Bayesian Optimization Framework in Python, 2016. <http://sheffieldml.github.io/GPyOpt/> (accessed 2023-11-29).
- (302) Aldeghi, M.; Häse, F.; Hickman, R. J.; Tamblyn, I.; Aspuru-Guzik, A. Golem: An Algorithm for Robust Experiment and Process Optimization. *Chem. Sci.* **2021**, 12, 14792–14807.
- (303) Hickman, R.; Aldeghi, M.; Aspuru-Guzik, A. Anubis: Bayesian Optimization with Unknown Feasibility Constraints for Scientific Experimentation. *ChemRxiv*2023. DOI: 10.26434/chemrxiv-2023-s5qnw (accessed November 4, 2023)
- (304) Hickman, R.; Sim, M.; Pablo-García, S.; Woolhouse, I.; Hao, H.; Bao, Z.; Bannigan, P.; Allen, C.; Aldeghi, M.; Aspuru-Guzik, A. Atlas: A Brain for Self-Driving Laboratories. *ChemRxiv*2023. DOI: 10.26434/chemrxiv-2023-8nrxx (accessed June 17, 2024)
- (305) Bakshy, E.; Dworkin, L.; Karrer, B.; Kashin, K.; Letham, B.; Murthy, A.; Singh, S. AE: A Domain-Agnostic Platform for Adaptive Experimentation. In *Conference on neural information processing systems*; 2018; pp 1–8.
- (306) Balandat, M.; Karrer, B.; Jiang, D. R.; Daulton, S.; Letham, B.; Wilson, A. G.; Bakshy, E. BOTORCH: A Framework for Efficient Monte-Carlo Bayesian Optimization.
- (307) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems* 32; Curran Associates, Inc., 2019; pp 8024–8035.
- (308) Rooney, M. B.; MacLeod, B. P.; Oldford, R.; Thompson, Z. J.; White, K. L.; Tungunyatham, J.; Stankiewicz, B. J.; Berlinguette, C. P. A Self-Driving Laboratory Designed to Accelerate the Discovery of Adhesive Materials. *Digit. Discov.* **2022**, 1, 382–389.
- (309) Baird, S. G.; Hall, J. R.; Sparks, T. D. Compactness Matters: Improving Bayesian Optimization Efficiency of Materials Formulations through Invariant Search Spaces. *Comput. Mater. Sci.* **2023**, 224, 112134.
- (310) Baird, S. G.; Liu, M.; Sparks, T. D. High-Dimensional Bayesian Optimization of 23 Hyperparameters over 100 Iterations for an Attention-Based Network to Predict Materials Property: A Case Study on CrabNet Using Ax Platform and SAASBO. *Comput. Mater. Sci.* **2022**, 211, 111505.
- (311) Griffiths, R.-R.; Klärner, L.; Moss, H.; Ravuri, A.; Truong, S. T.; Du, Y.; Stanton, S. D.; Tom, G.; Ranković, B.; Jamasb, A. R.; et al. GAUCHE: A Library for Gaussian Processes in Chemistry; *arXiv*2023. DOI: 10.48550/arXiv.2212.04450
- (312) Gardner, J.; Pleiss, G.; Weinberger, K. Q.; Bindel, D.; Wilson, A. G. GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration. *Adv. Neural Inf. Process. Syst.* **2018**, 31.
- (313) Kandasamy, K.; Vysyaraju, K. R.; Neiswanger, W.; Paria, B.; Collins, C. R.; Schneider, J.; Poczos, B.; Xing, E. P. Tuning Hyperparameters without Grad Students: Scalable and Robust Bayesian Optimisation with Dragonfly. *arXiv*2019 DOI: 10.48550/arXiv.1903.06694
- (314) Lindauer, M.; Eggensperger, K.; Feurer, M.; Biedenkapp, A.; Deng, D. SMAC3: A Versatile Bayesian Optimization Package for Hyperparameter Optimization. *J. Mach. Learn. Res.* **2022**, 23, 1–9 <http://jmlr.org/papers/v23/21-0888.html>.
- (315) Lindauer, M.; Eggensperger, K.; Feurer, M.; Biedenkapp, A.; Deng, D.; Benjamins, C.; Ruhkopf, T.; Sass, R.; Hutter, F. SMAC3: A Versatile Bayesian Optimization Package for Hyperparameter Optimization. *Journal of Machine Learning Research* **2022**, 23, 1–9 <https://www.jmlr.org/papers/volume23/21-0888/21-0888.pdf> (accessed 2023-11-29).
- (316) Cowen-Rivers, A. I.; Lyu, W.; Tutunov, R.; Wang, Z.; Grosnit, A.; Griffiths, R. R.; Maravel, A. M.; Jianye, H.; Wang, J.; Peters, J.; et al. HEBO: Pushing The Limits of Sample-Efficient Hyperparameter Optimisation. *arXiv*2020 DOI: 10.48550/arXiv.2012.03826
- (317) Duin, On the Choice of Smoothing Parameters for Parzen Estimators of Probability Density Functions. *IEEE Trans. Comput.* **1976**, 100, 1175–1179.
- (318) Bradford, E.; Schweidtmann, A. M.; Lapkin, A. Efficient Multiobjective Optimization Employing Gaussian Processes, Spectral Sampling and a Genetic Algorithm. *J. Glob. Optim.* **2018**, 71, 407–438.
- (319) Noack, M. M.; Reyes, K. G. Mathematical Nuances of Gaussian Process-Driven Autonomous Experimentation. *MRS Bull.* **2023**, 48, 153–163.
- (320) Ziatdinov, M. A.; Ghosh, A.; Kalinin, S. V. Physics Makes the Difference: Bayesian Optimization and Active Learning via Augmented Gaussian Process. *Mach. Learn. Sci. Technol.* **2022**, 3, 015003.
- (321) Ahmadi, M.; Ziatdinov, M.; Zhou, Y.; Lass, E. A.; Kalinin, S. V. Machine Learning for High-Throughput Experimental Exploration of Metal Halide Perovskites. *Joule*. **2021**, 5, 2797–2822.
- (322) Liu, Y.; Kelley, K. P.; Vasudevan, R. K.; Funakubo, H.; Ziatdinov, M. A.; Kalinin, S. V. Experimental Discovery of Structure-Property Relationships in Ferroelectric Materials via Active Learning. *Nat. Mach. Intell.* **2022**, 4, 341–350.

- (323) Sanchez, S. L.; Foadian, E.; Ziatdinov, M.; Yang, J.; Kalinin, S. V.; Liu, Y.; Ahmadi, M. Physics-Driven Discovery and Bandgap Engineering of Hybrid Perovskites. *arXiv* 2023. DOI: 10.48550/arXiv.2310.06583 (accessed June 7, 2024).
- (324) Slautin, B. N.; Pratiush, U.; Ivanov, I. N.; Liu, Y.; Pant, R.; Zhang, X.; Takeuchi, I.; Ziatdinov, M. A.; Kalinin, S. V. Co-Orchestration of Multiple Instruments to Uncover Structure-Property Relationships in Combinatorial Libraries. *arXiv* 2024. DOI: 10.48550/arXiv.2402.02198 (accessed June 7, 2024).
- (325) Liu, Y.; Morozovska, A. N.; Eliseev, E. A.; Kelley, K. P.; Vasudevan, R.; Ziatdinov, M.; Kalinin, S. V. Autonomous Scanning Probe Microscopy with Hypothesis Learning: Exploring the Physics of Domain Switching in Ferroelectric Materials. *Patterns*. 2023, 4, 100704.
- (326) Ziatdinov, M. A.; Liu, Y.; Morozovska, A. N.; Eliseev, E. A.; Zhang, X.; Takeuchi, I.; Kalinin, S. V. Hypothesis Learning in Automated Experiment: Application to Combinatorial Materials Libraries. *Adv. Mater.* 2022, 34, 2201345.
- (327) Ziatdinov, M.; Liu, Y.; Kelley, K.; Vasudevan, R.; Kalinin, S. V. Bayesian Active Learning for Scanning Probe Microscopy: From Gaussian Processes to Hypothesis Learning. *ACS Nano*. 2022, 16, 13492–13512.
- (328) Wolpert, D. H.; Macready, W. G. No Free Lunch Theorems for Optimization. *IEEE Trans. Evol. Comput.* 1997, 1, 67–82.
- (329) Häse, F.; Aldeghi, M.; Hickman, R. J.; Roch, L. M.; Christensen, M.; Liles, E.; Hein, J. E.; Aspuru-Guzik, A. Olympus: A Benchmarking Framework for Noisy Optimization and Experiment Planning. *Mach. Learn. Sci. Technol.* 2021, 2, 035021.
- (330) Hickman, R.; Parakh, P.; Cheng, A.; Ai, Q.; Schrier, J.; Aldeghi, M.; Aspuru-Guzik, A. Olympus, Enhanced: Benchmarking Mixed-Parameter and Multi-Objective Optimization in Chemistry and Materials Science. *ChemRxiv* 2023. DOI: 10.26434/chemrxiv-2023-74w8d (accessed June 21, 2023).
- (331) Felton, K. C.; Rittig, J. G.; Lapkin, A. A. Summit: Benchmarking Machine Learning Methods for Reaction Optimisation. *Chemistry-Methods*. 2021, 1, 116–122.
- (332) Tom, G.; Hickman, R. J.; Zinzuwadia, A.; Mohajeri, A.; Sanchez-Lengeling, B.; Aspuru-Guzik, A. Calibration and Generalizability of Probabilistic Models on Low-Data Chemical Datasets with DIONYSUS. *Digit. Discov.* 2023, 2, 759–774.
- (333) Liang, Q.; Gongora, A. E.; Ren, Z.; Tiihonen, A.; Liu, Z.; Sun, S.; Deneault, J. R.; Bash, D.; Mekki-Berrada, F.; Khan, S. A.; et al. Benchmarking the Performance of Bayesian Optimization across Multiple Experimental Materials Science Domains. *Npj Comput. Mater.* 2021, 7, 188.
- (334) Rohr, B.; Stein, H. S.; Guevarra, D.; Wang, Y.; Haber, J. A.; Aykol, M.; Suram, S. K.; Gregoire, J. M. Benchmarking the Acceleration of Materials Discovery by Sequential Learning. *Chem. Sci.* 2020, 11, 2696–2706.
- (335) Huyer, W.; Neumaier, A. SNOBFIT - Stable Noisy Optimization by Branch and Fit. *ACM Trans. Math. Softw.* 2008, 35, 1–25.
- (336) Zhou, Z.; Li, X.; Zare, R. N. Optimizing Chemical Reactions with Deep Reinforcement Learning. *ACS Cent. Sci.* 2017, 3, 1337–1344.
- (337) Blank, J.; Deb, K. Pymoo: Multi-Objective Optimization in Python. *IEEE Access*. 2020, 8, 89497–89509.
- (338) Rozycski, C. Application of the Simplex Method for Optimization of the Analytical Methods. *Chem. Anal.* 1993, 38, 681–698.
- (339) King, P. G.; Deming, S. N. UNIPLEX. Single-Factor Optimization of Response in the Presence of Error. *Anal. Chem.* 1974, 46, 1476–1481.
- (340) Vosburgh, W. C.; Cooper, G. R. Complex Ions. I. The Identification of Complex Ions in Solution by Spectrophotometric Measurements. *J. Am. Chem. Soc.* 1941, 63, 437–442.
- (341) Krause, R. D.; Lott, J. A. Use of the Simplex Method to Optimize Analytical Conditions in Clinical Chemistry. *Clin. Chem.* 1974, 20, 775–782.
- (342) Lott, J. A.; Turner, K. Evaluation of Trinder's Glucose Oxidase Method for Measuring Glucose in Serum and Urine. *Clin. Chem.* 1975, 21, 1754–1760.
- (343) Mieling, G. E.; Taylor, R. W.; Hargis, L. G.; English, J.; Pardue, H. L. Fully Automated Stopped-Flow Studies with a Hierarchical Computer Controlled System. *Anal. Chem.* 1976, 48, 1686–1693.
- (344) Lochmüller, C. H.; Lung, K. R.; Cousins, K. R. Applications of Optimization Strategies in the Design of Intelligent Laboratory Robotic Procedures. *Anal. Lett.* 1985, 18, 439–448.
- (345) Lochmüller, C. H.; Lung, K. R. Applications of Laboratory Robotics in Spectrophotometric Sample Preparation and Experimental Optimization. *Anal. Chim. Acta* 1986, 183, 257–262.
- (346) Horstotte, B.; Tovar Sánchez, A.; Duarte, C. M.; Cerdà, V. Sequential Injection Analysis for Automation of the Winkler Methodology, with Real-Time SIMPLEX Optimization and Shipboard Application. *Anal. Chim. Acta* 2010, 658, 147–155.
- (347) Pulgarín, J. A. M.; Molina, A. A.; Pardo, M. T. A. Simplex Optimization and Kinetic Determination of Nabumetone in Pharmaceutical Preparations by Micellar—Stabilized Room Temperature Phosphorescence. *Anal. Chim. Acta* 2005, 528, 77–82.
- (348) Santos, Q. O. d.; Novaes, C. G.; Bezerra, M. A.; Lemos, V. A.; Moreno, I.; Silva, D. G. d.; Santos, L. d. Application of Simplex Optimization in the Development of an Automated Online Preconcentration System for Manganese Determination. *J. Braz. Chem. Soc.* 2010, 21, 2340–2346.
- (349) Ensafi, A. A.; Chamjngali, M. A. Flow-Injection Spectrophotometric Determination of Periodate and Iodate by Their Reaction with Pyrogallol Red in Acidic Media. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* 2002, 58, 2835–2839.
- (350) Silva, H. A. D. F. O.; Álvares-Ribeiro, L. M. B. C. Optimization of a Flow Injection Analysis System for Tartaric Acid Determination in Wines. *Talanta*. 2002, 58, 1311–1318.
- (351) Motyka, K.; Onjia, A.; Mikuška, P.; Večera, Z. Flow-Injection Chemiluminescence Determination of Formaldehyde in Water. *Talanta*. 2007, 71, 900–905.
- (352) Shakerian, F.; Dadfarnia, S.; Shabani, A. M. H.; Rohani, M. MultiSimplex Optimization of On-Line Sorbent Proconcentration and Determination of Iron by FI-AAS and Microcolumn of Immobilized Ferron. *Talanta*. 2008, 77, 551–555.
- (353) Holderith, J.; Tóth, T.; Váradi, A. Minimizing the Time for Gas Chromatographic Analysis. *J. Chromatogr. A*. 1976, 119, 215–222.
- (354) Berridge, J. C. Unattended Optimisation of Normal Phase High-Performance Liquid Chromatography Separations with a Microcomputer Controlled Chromatograph. *Chromatographia*. 1982, 16, 172–174.
- (355) James, E. B.; Simons, J. V.; Bateman, D. C.; Adams, M. J.; Black, I.; Berridge, J. C.; Braithwaite, A.; Ferrige, A. G.; Strutt, A. C. R.; Everett, A. J.; et al. Annual Chemical Congress. *Anal. Proc.* 1982, 19, 462–483.
- (356) Wright, A. G.; Fell, A. F.; Berridge, J. C.; Kelly, H. C.; Davies, B. E.; Harding, A.; Robertson, S. M.; Bagon, D. A.; Lynch, I. R.; Buxton, P. C.; et al. Short Papers in Pharmaceutical Analysis. *Anal. Proc.* 1988, 25, 300–308.
- (357) Horstotte, B.; Jarošová, P.; Chocholouš, P.; Sklenářová, H.; Solich, P. Sequential Injection Chromatography with Post-Column Reaction/Derivatization for the Determination of Transition Metal Cations in Natural Water Samples. *Talanta*. 2015, 136, 75–83.
- (358) Berridge, J. C. Techniques for the Automated Optimization of HPLC Separations. *TrAC Trends Anal. Chem.* 1984, 3, 5–10.
- (359) O'Hagan, S.; Dunn, W. B.; Brown, M.; Knowles, J. D.; Kell, D. B. Closed-Loop, Multiobjective Optimization of Analytical Instrumentation: Gas Chromatography/Time-of-Flight Mass Spectrometry of the Metabolomes of Human Serum and of Yeast Fermentations. *Anal. Chem.* 2005, 77, 290–303.
- (360) O'Hagan, S.; Dunn, W. B.; Knowles, J. D.; Broadhurst, D.; Williams, R.; Ashworth, J. J.; Cameron, M.; Kell, D. B. Closed-Loop, Multiobjective Optimization of Two-Dimensional Gas Chromatog-

- raphy/Mass Spectrometry for Serum Metabolomics. *Anal. Chem.* **2007**, *79*, 464–476.
- (361) Zelena, E.; Dunn, W. B.; Broadhurst, D.; Francis-McIntyre, S.; Carroll, K. M.; Begley, P.; O'Hagan, S.; Knowles, J. D.; Halsall, A.; HUSERMET Consortium; et al. Development of a Robust and Repeatable UPLC-MS Method for the Long-Term Metabolomic Study of Human Serum. *Anal. Chem.* **2009**, *81*, 1357–1364.
- (362) Jenkinson, C.; Bradbury, J.; Taylor, A. S.; Adams, J.; He, S. R.; Viant, M.; Hewison, M. Automated Development of an LC-MS/MS Method for Measuring Multiple Vitamin D Metabolites Using MUSCLE Software. *Anal. Methods.* **2017**, *9*, 2723–2731.
- (363) I, T.-P.; Smith, R.; Guhan, S.; Taksen, K.; Vavra, M.; Myers, D.; Hearn, M. T. W. Intelligent Automation of High-Performance Liquid Chromatography Method Development by Means of a Real-Time Knowledge-Based Approach. *J. Chromatogr. A.* **2002**, *972*, 27–43.
- (364) Boelrijk, J.; Ensing, B.; Forré, P.; Pirok, B. W. J. Closed-Loop Automatic Gradient Design for Liquid Chromatography Using Bayesian Optimization. *Anal. Chim. Acta.* **2023**, *1242*, 340789.
- (365) Clayton, A. D.; Power, L. A.; Reynolds, W. R.; Ainsworth, C.; Hose, D. R. J.; Jones, M. F.; Chamberlain, T. W.; Blacker, A. J.; Bourne, R. A. Self-Optimising Reactive Extractions: Towards the Efficient Development of Multi-Step Continuous Flow Processes. *J. Flow Chem.* **2020**, *10*, 199–206.
- (366) Pomberger, A.; Jose, N.; Walz, D.; Meissner, J.; Holze, C.; Kopczynski, M.; Müller-Bischof, P.; Lapkin, A. A. Automated pH Adjustment Driven by Robotic Workflows and Active Machine Learning. *Chem. Eng. J.* **2023**, *451*, 139099.
- (367) Zhu, Y.; Nishigori, S.; Shimura, N.; Nara, T.; Fujimori, E. Development of an Automatic pH Adjustment Instrument for the Preparation of Analytical Samples Prior to Solid Phase Extraction. *Anal. Sci.* **2020**, *36*, 621–625.
- (368) Chitre, A.; Cheng, J.; Ahmed, S.; Querimit, R.; Hippalgaonkar, K.; Lapkin, A. pHbot: Self-Driven Robot for pH Adjustment of Viscous Formulations via Physics-Informed-ML. *ChemRxiv* 2023. DOI: [10.26434/chemrxiv-2023-c46mv](https://doi.org/10.26434/chemrxiv-2023-c46mv) (accessed June 17, 2024)
- (369) Noack, M. M.; Yager, K. G.; Fukuto, M.; Doerk, G. S.; Li, R.; Sethian, J. A. A Kriging-Based Approach to Autonomous Experimentation with Applications to X-Ray Scattering. *Sci. Rep.* **2019**, *9*, 11809.
- (370) Cressie, N. The Origins of Kriging. *Math. Geol.* **1990**, *22*, 239–252.
- (371) Noack, M. M.; Doerk, G. S.; Li, R.; Streit, J. K.; Vaia, R. A.; Yager, K. G.; Fukuto, M. Autonomous Materials Discovery Driven by Gaussian Process Regression with Inhomogeneous Measurement Noise and Anisotropic Kernels. *Sci. Rep.* **2020**, *10*, 17663.
- (372) Brown, K. A. Scanning Probes as a Materials Automation Platform with Extremely Miniaturized Samples. *Matter.* **2022**, *5*, 3112–3123.
- (373) Szymanski, N. J.; Bartel, C. J.; Zeng, Y.; Diallo, M.; Kim, H.; Ceder, G. Adaptively Driven X-Ray Diffraction Guided by Machine Learning for Autonomous Phase Identification. *Npj Comput. Mater.* **2023**, *9*, 1–8.
- (374) Zagorac, D.; Müller, H.; Ruehl, S.; Zagorac, J.; Rehme, S. Recent Developments in the Inorganic Crystal Structure Database: Theoretical Crystal Structure Data and Related Features. *J. Appl. Crystallogr.* **2019**, *52*, 918–925.
- (375) Dixon, J. M.; Lindsey, J. S. Performance of Search Algorithms in the Examination of Chemical Reaction Spaces with an Automated Chemistry Workstation. *JALA J. Assoc. Lab. Autom.* **2004**, *9*, 364–374.
- (376) Fabry, D. C.; Sugiono, E.; Rueping, M. Self-Optimizing Reactor Systems: Algorithms, On-Line Analytics, Setups, and Strategies for Accelerating Continuous Flow Process Optimization. *Isr. J. Chem.* **2014**, *54*, 341–350.
- (377) Houben, C.; Lapkin, A. A. Automatic Discovery and Optimization of Chemical Processes. *Curr. Opin. Chem. Eng.* **2015**, *9*, 1–7.
- (378) Sans, V.; Cronin, L. Towards Dial-a-Molecule by Integrating Continuous Flow, Analytics and Self-Optimisation. *Chem. Soc. Rev.* **2016**, *45*, 2032–2043.
- (379) Reizman, B. J.; Jensen, K. F. Feedback in Flow for Accelerated Reaction Development. *Acc. Chem. Res.* **2016**, *49*, 1786–1796.
- (380) Clayton, A. D.; Manson, J. A.; Taylor, C. J.; Chamberlain, T. W.; Taylor, B. A.; Clemens, G.; Bourne, R. A. Algorithms for the Self-Optimisation of Chemical Reactions. *React. Chem. Eng.* **2019**, *4*, 1545–1554.
- (381) Mateos, C.; Nieves-Remacha, M. J.; Rincón, J. A. Automated Platforms for Reaction Self-Optimization in Flow. *React. Chem. Eng.* **2019**, *4*, 1536–1544.
- (382) Labes, R.; Bourne, R. A.; Chamberlain, T. W. Automated Reaction Optimisation in Continuous Flow. *Chem. Today.* **2020**, *38*.
- (383) Taylor, C. J.; Pomberger, A.; Felton, K. C.; Grainger, R.; Barecka, M.; Chamberlain, T. W.; Bourne, R. A.; Johnson, C. N.; Lapkin, A. A. A Brief Introduction to Chemical Reaction Optimization. *Chem. Rev.* **2023**, *123*, 3089.
- (384) Lindsey, J. S. A Retrospective on the Automation of Laboratory Synthetic Chemistry. *Chemom. Intell. Lab. Syst.* **1992**, *17*, 15–45.
- (385) Fredrickson, C. K.; Fan, Z. H. Macro-to-Micro Interfaces for Microfluidic Devices. *Lab. Chip.* **2004**, *4*, 526–533.
- (386) Plutschack, M. B.; Pieber, B.; Gilmore, K.; Seeberger, P. H. The Hitchhiker's Guide to Flow Chemistry. *Chem. Rev.* **2017**, *117*, 11796–11893.
- (387) Capaldo, L.; Wen, Z.; Noël, T. A Field Guide to Flow Chemistry for Synthetic Organic Chemists. *Chem. Sci.* **2023**, *14*, 4230–4247.
- (388) Bell, N. L.; Boser, F.; Publiauskas, A.; Willcox, D. R.; Luna, V. S.; Cronin, L. Autonomous Execution of Highly Reactive Chemical Transformations in the Schlenkputer. *Nat. Chem. Eng.* **2024**, *1*, 180–189.
- (389) Britton, J.; Raston, C. L. Multi-Step Continuous-Flow Synthesis. *Chem. Soc. Rev.* **2017**, *46*, 1250–1271.
- (390) Jiao, J.; Nie, W.; Yu, T.; Yang, F.; Zhang, Q.; Aihemaiti, F.; Yang, T.; Liu, X.; Wang, J.; Li, P. Multi-Step Continuous-Flow Organic Synthesis: Opportunities and Challenges. *Chem. - Eur. J.* **2021**, *27*, 4817–4838.
- (391) Gutmann, B.; Cantillo, D.; Kappe, C. O. Continuous-Flow Technology—A Tool for the Safe Manufacturing of Active Pharmaceutical Ingredients. *Angew. Chem. Int. Ed.* **2015**, *54*, 6688–6728.
- (392) Zhang, M.; Vokoun, A. E.; Chen, B.; Deng, W.; Dupont, R. L.; Xu, Y.; Wang, X. Advancements in Droplet Reactor Systems Represent New Opportunities in Chemical Reactor Engineering: A Perspective. *Can. J. Chem. Eng.* **2023**, *101*, 5189–5207.
- (393) Winicov, H.; Schainbaum, J.; Buckley, J.; Longino, G.; Hill, J.; Berkoff, C. E. Chemical Process Optimization by Computer — a Self-Directed Chemical Synthesis System. *Anal. Chim. Acta.* **1978**, *103*, 469–476.
- (394) Frisbee, A. R.; Nantz, M. H.; Kramer, G. W.; Fuchs, P. L. Robotic Orchestration of Organic Reactions: Yield Optimization via an Automated System with Operator-Specified Reaction Sequences. *J. Am. Chem. Soc.* **1984**, *106*, 7143–7145.
- (395) Lindsey, J. S.; Corkan, L. A.; Erb, D.; Powers, G. J. Robotic Work Station for Microscale Synthetic Chemistry: On-line Absorption Spectroscopy, Quantitative Automated Thin-layer Chromatography, and Multiple Reactions in Parallel. *Rev. Sci. Instrum.* **1988**, *59*, 940–950.
- (396) Plouvier, J.-C.; Andrew Corkan, L.; Lindsey, J. S. Experiment Planner for Strategic Experimentation with an Automated Chemistry Workstation. *Chemom. Intell. Lab. Syst.* **1992**, *17*, 75–94.
- (397) Andrew Corkan, L.; Plouvier, J.-C.; Lindsey, J. S. Application of an Automated Chemistry Workstation to Problems in Synthetic Chemistry. *Chemom. Intell. Lab. Syst.* **1992**, *17*, 95–105.
- (398) Harre, M.; Tilstam, U.; Weinmann, H. Breaking the New Bottleneck: Automated Synthesis in Chemical Process Research and Development. *Org. Process Res. Dev.* **1999**, *3*, 304–318.

- (399) McMullen, J. P.; Stone, M. T.; Buchwald, S. L.; Jensen, K. F. An Integrated Microreactor System for Self-Optimization of a Heck Reaction: From Micro- to Mesoscale Flow Systems. *Angew. Chem. Int. Ed.* **2010**, *49*, 7076–7080.
- (400) Parrott, A. J.; Bourne, R. A.; Akien, G. R.; Irvine, D. J.; Poliakoff, M. Self-Optimizing Continuous Reactions in Supercritical Carbon Dioxide. *Angew. Chem. Int. Ed.* **2011**, *50*, 3788–3792.
- (401) Jumbam, D. N.; Skilton, R. A.; Parrott, A. J.; Bourne, R. A.; Poliakoff, M. The Effect of Self-Optimization Targets on the Methylation of Alcohols Using Dimethyl Carbonate in Supercritical CO<sub>2</sub>. *J. Flow Chem.* **2012**, *2*, 24–27.
- (402) Moore, J. S.; Jensen, K. F. Automated Multitrajectory Method for Reaction Optimization in a Microfluidic System Using Online IR Analysis. *Org. Process Res. Dev.* **2012**, *16*, 1409–1415.
- (403) Skilton, R. A.; Parrott, A. J.; George, M. W.; Poliakoff, M.; Bourne, R. A. Real-Time Feedback Control Using Online Attenuated Total Reflection Fourier Transform Infrared (ATR FT-IR) Spectroscopy for Continuous Flow Optimization and Process Knowledge. *Appl. Spectrosc.* **2013**, *67*, 1127–1131.
- (404) 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.
- (405) Reizman, B. J.; Jensen, K. F. Simultaneous Solvent Screening and Reaction Optimization in Microliter Slugs. *Chem. Commun.* **2015**, *51*, 13290–13293.
- (406) Holmes, N.; Akien, G. R.; Savage, R. J. D.; Stanetty, C.; Baxendale, I. R.; Blacker, A. J.; Taylor, B. A.; Woodward, R. L.; Meadows, R. E.; Bourne, R. A. Online Quantitative Mass Spectrometry for the Rapid Adaptive Optimisation of Automated Flow Reactors. *React. Chem. Eng.* **2016**, *1*, 96–100.
- (407) Holmes, N.; Akien, G. R.; Blacker, A. J.; Woodward, R. L.; Meadows, R. E.; Bourne, R. A. Self-Optimisation of the Final Stage in the Synthesis of EGFR Kinase Inhibitor AZD9291 Using an Automated Flow Reactor. *React. Chem. Eng.* **2016**, *1*, 366–371.
- (408) Reizman, B. J.; Wang, Y.-M.; Buchwald, S. L.; Jensen, K. F. Suzuki-Miyaura Cross-Coupling Optimization Enabled by Automated Feedback. *React. Chem. Eng.* **2016**, *1*, 658–666.
- (409) Cortés-Borda, D.; Kutonova, K. V.; Jamet, C.; Trusova, M. E.; Zammattio, F.; Truchet, C.; Rodriguez-Zubiri, M.; Felpin, F.-X. Optimizing the Heck-Matsuda Reaction in Flow with a Constraint-Adapted Direct Search Algorithm. *Org. Process Res. Dev.* **2016**, *20*, 1979–1987.
- (410) Echtermeyer, A.; Amar, Y.; Zakrzewski, J.; Lapkin, A. Self-Optimisation and Model-Based Design of Experiments for Developing a C-H Activation Flow Process. *Beilstein J. Org. Chem.* **2017**, *13*, 150–163.
- (411) Hsieh, H.-W.; Coley, C. W.; Baumgartner, L. M.; Jensen, K. F.; Robinson, R. I. Photoredox Iridium-Nickel Dual-Catalyzed Decarboxylative Arylation Cross-Coupling: From Batch to Continuous Flow via Self-Optimizing Segmented Flow Reactor. *Org. Process Res. Dev.* **2018**, *22*, 542–550.
- (412) Baumgartner, L. M.; Coley, C. W.; Reizman, B. J.; Gao, K. W.; Jensen, K. F. Optimum Catalyst Selection over Continuous and Discrete Process Variables with a Single Droplet Microfluidic Reaction Platform. *React. Chem. Eng.* **2018**, *3*, 301–311.
- (413) Poscharny, K.; Fabry, D. C.; Heddrich, S.; Sugiono, E.; Liauw, M. A.; Rueping, M. Machine Assisted Reaction Optimization: A Self-Optimizing Reactor System for Continuous-Flow Photochemical Reactions. *Tetrahedron* **2018**, *74*, 3171–3175.
- (414) Jeraal, M. I.; Holmes, N.; Akien, G. R.; Bourne, R. A. Enhanced Process Development Using Automated Continuous Reactors by Self-Optimization Algorithms and Statistical Empirical Modelling. *Tetrahedron* **2018**, *74*, 3158–3164.
- (415) Cherkasov, N.; Bai, Y.; Expósito, A. J.; Rebrov, E. V. OpenFlowChem - a Platform for Quick, Robust and Flexible Automation and Self-Optimization of Flow Chemistry. *React. Chem. Eng.* **2018**, *3*, 769–780.
- (416) Cortés-Borda, D.; Wimmer, E.; Gouilleux, B.; Barré, E.; Oger, N.; Goulamaly, L.; Peault, L.; Charrier, B.; Truchet, C.; Giraudeau, P.; et al. An Autonomous Self-Optimizing Flow Reactor for the Synthesis of Natural Product Carpanone. *J. Org. Chem.* **2018**, *83*, 14286–14299.
- (417) Schweidtmann, A. M.; Clayton, A. D.; Holmes, N.; Bradford, E.; Bourne, R. A.; Lapkin, A. A. Machine Learning Meets Continuous Flow Chemistry: Automated Optimization towards the Pareto Front of Multiple Objectives. *Chem. Eng. J.* **2018**, *352*, 277–282.
- (418) Wimmer, E.; Cortés-Borda, D.; Brochart, S.; Barré, E.; Truchet, C.; Felpin, F.-X. An Autonomous Self-Optimizing Flow Machine for the Synthesis of Pyridine-Oxazoline (PyOX) Ligands. *React. Chem. Eng.* **2019**, *4*, 1608–1615.
- (419) Fabry, D. C.; Heddrich, S.; Sugiono, E.; Liauw, M. A.; Rueping, M. Adaptive and Automated System-Optimization for Heterogeneous Flow-Hydrogenation Reactions. *React. Chem. Eng.* **2019**, *4*, 1486–1491.
- (420) Baumgartner, L. M.; Dennis, J. M.; White, N. A.; Buchwald, S. L.; Jensen, K. F. Use of a Droplet Platform To Optimize Pd-Catalyzed C-N Coupling Reactions Promoted by Organic Bases. *Org. Process Res. Dev.* **2019**, *23*, 1594–1601.
- (421) Clayton, A. D.; Schweidtmann, A. M.; Clemens, G.; Manson, J. A.; Taylor, C. J.; Niño, C. G.; Chamberlain, T. W.; Kapur, N.; Blacker, A. J.; Lapkin, A. A.; et al. Automated Self-Optimisation of Multi-Step Reaction and Separation Processes Using Machine Learning. *Chem. Eng. J.* **2020**, *384*, 123340.
- (422) Vasudevan, N.; Wimmer, E.; Barré, E.; Cortés-Borda, D.; Rodriguez-Zubiri, M.; Felpin, F.-X. Direct C-H Arylation of Indole-3-Acetic Acid Derivatives Enabled by an Autonomous Self-Optimizing Flow Reactor. *Adv. Synth. Catal.* **2021**, *363*, 791–799.
- (423) Jeraal, M. I.; Sung, S.; Lapkin, A. A. A Machine Learning-Enabled Autonomous Flow Chemistry Platform for Process Optimization of Multiple Reaction Metrics. *Chemistry-Methods* **2021**, *1*, 71–77.
- (424) Hall, B. L.; Taylor, C. J.; Labes, R.; Massey, A. F.; Menzel, R.; Bourne, R. A.; Chamberlain, T. W. Autonomous Optimisation of a Nanoparticle Catalysed Reduction Reaction in Continuous Flow. *Chem. Commun.* **2021**, *57*, 4926–4929.
- (425) 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.; et al. Data-Science Driven Autonomous Process Optimization. *Commun. Chem.* **2021**, *4*, 1–12.
- (426) Ahn, G.-N.; Kang, J.-H.; Lee, H.-J.; Park, B. E.; Kwon, M.; Na, G.-S.; Kim, H.; Seo, D.-H.; Kim, D.-P. Exploring Ultrafast Flow Chemistry by Autonomous Self-Optimizing Platform. *Chem. Eng. J.* **2023**, *453*, 139707.
- (427) Müller, P.; Clayton, A. D.; Manson, J.; Riley, S.; May, O. S.; Govan, N.; Notman, S.; Ley, S. V.; Chamberlain, T. W.; Bourne, R. A. Automated Multi-Objective Reaction Optimisation: Which Algorithm Should I Use? *React. Chem. Eng.* **2022**, *7*, 987–993.
- (428) Nandiwal, K. Y.; Hart, T.; Zahrt, A. F.; Nambiar, A. M. K.; Mahesh, P. T.; Mo, Y.; Nieves-Remacha, M. J.; Johnson, M. D.; García-Losada, P.; Mateos, C.; et al. Continuous Stirred-Tank Reactor Cascade Platform for Self-Optimization of Reactions Involving Solids. *React. Chem. Eng.* **2022**, *7*, 1315–1327.
- (429) Gérardy, R.; Nambiar, A. M. K.; Hart, T.; Mahesh, P. T.; Jensen, K. F. Photochemical Synthesis of the Bioactive Fragment of Salbutamol and Derivatives in a Self-Optimizing Flow Chemistry Platform. *Chem. - Eur. J.* **2022**, *28*, No. e202201385.
- (430) Nambiar, A. M. K.; Breen, C. P.; Hart, T.; Kulesza, T.; Jamison, T. F.; Jensen, K. F. Bayesian Optimization of Computer-Proposed Multistep Synthetic Routes on an Automated Robotic Flow Platform. *ACS Cent. Sci.* **2022**, *8*, 825–836.
- (431) 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.; et al. Closed-Loop Optimization of General Reaction Conditions for Heteroaryl Suzuki-Miyaura Coupling. *Science* **2022**, *378*, 399–405.
- (432) Ha, T.; Lee, D.; Kwon, Y.; Park, M. S.; Lee, S.; Jang, J.; Choi, B.; Jeon, H.; Kim, J.; Choi, H.; et al. AI-Driven Robotic Chemist for Autonomous Synthesis of Organic Molecules. *Sci. Adv.* **2023**, *9*, No. eadj0461.

- (433) Clayton, A. D.; Pyzer-Knapp, E. O.; Purdie, M.; Jones, M. F.; Barthelme, A.; Pavey, J.; Kapur, N.; Chamberlain, T. W.; Blacker, A. J.; Bourne, R. A. Bayesian Self-Optimization for Telescopied Continuous Flow Synthesis. *Angew. Chem.* **2023**, *135*, No. e202214511.
- (434) Kershaw, O. J.; Clayton, A. D.; Manson, J. A.; Barthelme, A.; Pavey, J.; Peach, P.; Mustakis, J.; Howard, R. M.; Chamberlain, T. W.; Warren, N. J.; et al. Machine Learning Directed Multi-Objective Optimization of Mixed Variable Chemical Systems. *Chem. Eng. J.* **2023**, *451*, 138443.
- (435) Mueller, P.; Vriza, A.; Clayton, A. D.; May, O. S.; Govan, N.; Notman, S.; Ley, S. V.; Chamberlain, T. W.; Bourne, R. A. Exploring the Chemical Space of Phenyl Sulfide Oxidation by Automated Optimization. *React. Chem. Eng.* **2023**, *8*, 538–542.
- (436) Taylor, C. J.; Felton, K. C.; Wigh, D.; Jeraal, M. I.; Grainger, R.; Chessari, G.; Johnson, C. N.; Lapkin, A. A. Accelerated Chemical Reaction Optimization Using Multi-Task Learning. *ACS Cent. Sci.* **2023**, *9*, 957–968.
- (437) Zhang, J.; Sugisawa, N.; Felton, K.; Fuse, S.; Lapkin, A. Multi-Objective Bayesian Optimisation Using q-Noisy Expected Hypervolume Improvement (qNEHVI) for Schotten-Baumann Reaction. *ChemRxiv* 2023. DOI: [10.26434/chemrxiv-2023-dlkgl](https://doi.org/10.26434/chemrxiv-2023-dlkgl) (accessed November 13, 2023).
- (438) Karan, D.; Chen, G.; Jose, N.; Bai, J.; McDaid, P. A.; Lapkin, A. A Machine Learning-Enabled Process Optimization of Ultra-Fast Flow Chemistry with Multiple Reaction Metrics. *React. Chem. Eng.* **2024**, *9*, 619–629.
- (439) Slattery, A.; Wen, Z.; Tenblad, P.; Sanjosé-Orduna, J.; Pintossi, D.; den Hartog, T.; Noël, T. Automated Self-Optimization, Intensification, and Scale-up of Photocatalysis in Flow. *Science* **2024**, *383*, No. eadj1817.
- (440) Schilter, O.; Gutierrez, D. P.; Folkmann, L. M.; Castrogiovanni, A.; García-Durán, A.; Zipoli, F.; Roch, L. M.; Laino, T. Combining Bayesian Optimization and Automation to Simultaneously Optimize Reaction Conditions and Routes. *Chem. Sci.* **2024**, *15*, 7732–7741.
- (441) Bennett, J. A.; Orouji, N.; Khan, M.; Sadeghi, S.; Rodgers, J.; Abolhasani, M. Autonomous Reaction Pareto-Front Mapping with a Self-Driving Catalysis Laboratory. *Nat. Chem. Eng.* **2024**, *1*, 240–250.
- (442) Leonov, A. I.; Hammer, A. J. S.; Lach, S.; Mehr, S. H. M.; Caramelli, D.; Angelone, D.; Khan, A.; O'Sullivan, S.; Craven, M.; Wilbraham, L.; et al. An Integrated Self-Optimizing Programmable Chemical Synthesis and Reaction Engine. *Nat. Commun.* **2024**, *15*, 1240.
- (443) Bourne, R. A.; Skilton, R. A.; Parrott, A. J.; Irvine, D. J.; Poliakoff, M. Adaptive Process Optimization for Continuous Methylation of Alcohols in Supercritical Carbon Dioxide. *Org. Process Res. Dev.* **2011**, *15*, 932–938.
- (444) Manson, J. A.; Chamberlain, T. W.; Bourne, R. A. MVMOO: Mixed Variable Multi-Objective Optimisation. *J. Glob. Optim.* **2021**, *80*, 865–886.
- (445) Zhang, J.; Sugisawa, N.; Felton, K. C.; Fuse, S.; Lapkin, A. A. Multi-Objective Bayesian Optimisation Using q-Noisy Expected Hypervolume Improvement (qNEHVI) for the Schotten-Baumann Reaction. *React. Chem. Eng.* **2024**, *9*, 706.
- (446) 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.
- (447) 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.
- (448) Wang, J. Y.; Stevens, J. M.; Kariofillis, S. K.; Tom, M.-J.; Golden, D. L.; Li, J.; Tabora, J. E.; Parasram, M.; Shields, B. J.; Primer, D. N.; et al. Identifying General Reaction Conditions by Bandit Optimization. *Nature* **2024**, *626*, 1025–1033.
- (449) Hansen, N. The CMA Evolution Strategy: A Comparing Review. In *Towards a New Evolutionary Computation: Advances in the Estimation of Distribution Algorithms*; Lozano, J. A.; Larrañaga, P.; Inza, I.; Bengoechea, E., Eds.; Studies in Fuzziness and Soft Computing; Springer: Berlin, Heidelberg, 2006; pp 75–102 DOI: [10.1007/3-540-32494-1\\_4](https://doi.org/10.1007/3-540-32494-1_4).
- (450) Coley, C. W.; Thomas, D. A.; Lummiss, J. A. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; et al. A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning. *Science* **2019**, *365*, No. eaax1566.
- (451) Daponte, J. A.; Guo, Y.; Ruck, R. T.; Hein, J. E. Using an Automated Monitoring Platform for Investigations of Biphasic Reactions. *ACS Catal.* **2019**, *9*, 11484–11491.
- (452) Vieira, R. A. M.; Sayer, C.; Lima, E. L.; Pinto, J. C. Closed-Loop Composition and Molecular Weight Control of a Copolymer Latex Using Near-Infrared Spectroscopy. *Ind. Eng. Chem. Res.* **2002**, *41*, 2915–2930.
- (453) Beebe, K. R.; Kowalski, B. R. An Introduction to Multivariate Calibration and Analysis. *Anal. Chem.* **1987**, *59*, 1007A–1017A.
- (454) Bojkov, B.; Luus, R. Use of Random Admissible Values for Control in Iterative Dynamic Programming. *Ind. Eng. Chem. Res.* **1992**, *31*, 1308–1314.
- (455) Sayer, C.; Arzamendi, G.; Asua, J. M.; Lima, E. L.; Pinto, J. C. Dynamic Optimization of Semicontinuous Emulsion Copolymerization Reactions: Composition and Molecular Weight Distribution. *Comput. Chem. Eng.* **2001**, *25*, 839–849.
- (456) Houben, C.; Peremezhney, N.; Zubov, A.; Kosek, J.; Lapkin, A. A. Closed-Loop Multitarget Optimization for Discovery of New Emulsion Polymerization Recipes. *Org. Process Res. Dev.* **2015**, *19*, 1049–1053.
- (457) Peremezhney, N.; Hines, E.; Lapkin, A.; Connaughton, C. Combining Gaussian Processes, Mutual Information and a Genetic Algorithm for Multi-Target Optimization of Expensive-to-Evaluate Functions. *Eng. Optim.* **2014**, *46*, 1593–1607.
- (458) Moad, G.; Rizzardo, E.; Thang, S. H. Radical Addition-Fragmentation Chemistry in Polymer Synthesis. *Polymer* **2008**, *49*, 1079–1131.
- (459) Rubens, M.; Vrijen, J. H.; Laun, J.; Junkers, T. Precise Polymer Synthesis by Autonomous Self-Optimizing Flow Reactors. *Angew. Chem. Int. Ed.* **2019**, *58*, 3183–3187.
- (460) Knox, S. T.; Parkinson, S. J.; Wilding, C. Y. P.; Bourne, R. A.; Warren, N. J. Autonomous Polymer Synthesis Delivered by Multi-Objective Closed-Loop Optimisation. *Polym. Chem.* **2022**, *13*, 1576–1585.
- (461) Lai, N. S.; Tew, Y. S.; Zhong, X.; Yin, J.; Li, J.; Yan, B.; Wang, X. Artificial Intelligence (AI) Workflow for Catalyst Design and Optimization. *Ind. Eng. Chem. Res.* **2023**, *62*, 17835–17848.
- (462) Corma, A.; Serra, J. M.; Serna, P.; Valero, S.; Argente, E.; Botti, V. Optimisation of Olefin Epoxidation Catalysts with the Application of High-Throughput and Genetic Algorithms Assisted by Artificial Neural Networks (Softcomputing Techniques). *J. Catal.* **2005**, *229*, 513–524.
- (463) Valero, S.; Argente, E.; Botti, V.; Serra, J. M.; Corma, A. SoftComputing Techniques Applied to Catalytic Reactions. In *Current Topics in Artificial Intelligence*; Conejo, R.; Urretavizcaya, M.; Pérez-de-la-Cruz, J.-L., Eds.; Lecture Notes in Computer Science; Springer: Berlin, Heidelberg, 2004; pp 536–545. DOI: [10.1007/978-3-540-25945-9\\_53](https://doi.org/10.1007/978-3-540-25945-9_53).
- (464) Cho, S.-B. Fusion of Neural Networks with Fuzzy Logic and Genetic Algorithm. *Integr. Comput.-Aided Eng.* **2002**, *9*, 363–372.
- (465) Kreutz, J. E.; Shukhaev, A.; Du, W.; Druskin, S.; Daugulis, O.; Ismagilov, R. F. Evolution of Catalysts Directed by Genetic Algorithms in a Plug-Based Microfluidic Device Tested with Oxidation of Methane by Oxygen. *J. Am. Chem. Soc.* **2010**, *132*, 3128–3132.
- (466) Zhu, Q.; Huang, Y.; Zhou, D.; Zhao, L.; Guo, L.; Yang, R.; Sun, Z.; Luo, M.; Zhang, F.; Xiao, H.; et al. Automated Synthesis of Oxygen-Producing Catalysts from Martian Meteorites by a Robotic AI Chemist. *Nat. Synth.* **2024**, *3*, 1–10.

- (467) Ramirez, A.; Lam, E.; Gutierrez, D. P.; Hou, Y.; Tribukait, H.; Roch, L. M.; Copéret, C.; Laveille, P. Accelerated Exploration of Heterogeneous CO<sub>2</sub> Hydrogenation Catalysts by Bayesian-Optimized High-Throughput and Automated Experimentation. *Chem. Catal.* **2024**, *4*, 100888.
- (468) Seumer, J.; Kirschner Solberg Hansen, J.; Brøndsted Nielsen, M.; Jensen, J. H. Computational Evolution Of New Catalysts For The Morita-Baylis-Hillman Reaction\*\*. *Angew. Chem. Int. Ed.* **2023**, *62*, No. e202218565.
- (469) Braconi, E. Bayesian Optimization as a Valuable Tool for Sustainable Chemical Reaction Development. *Nat. Rev. Methods Primer.* **2023**, *3*, 1–2.
- (470) Roberts, R. M. *Serendipity: Accidental Discoveries in Science*; Wiley, 1989.
- (471) Amara, Z.; Poliakoff, M.; Duque, R.; Geier, D.; Franciò, G.; Gordon, C. M.; Meadows, R. E.; Woodward, R.; Leitner, W. Enabling the Scale-Up of a Key Asymmetric Hydrogenation Step in the Synthesis of an API Using Continuous Flow Solid-Supported Catalysis. *Org. Process Res. Dev.* **2016**, *20*, 1321–1327.
- (472) Collins, K. D.; Gensch, T.; Glorius, F. Contemporary Screening Approaches to Reaction Discovery and Development. *Nat. Chem.* **2014**, *6*, 859–871.
- (473) Gromski, P. S.; Henson, A. B.; Granda, J. M.; Cronin, L. How to Explore Chemical Space Using Algorithms and Automation. *Nat. Rev. Chem.* **2019**, *3*, 119–128.
- (474) 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.
- (475) Caramelli, D.; Granda, J. M.; Mehr, S. H. M.; Cambié, D.; Henson, A. B.; Cronin, L. Discovering New Chemistry with an Autonomous Robotic Platform Driven by a Reactivity-Seeking Neural Network. *ACS Cent. Sci.* **2021**, *7*, 1821–1830.
- (476) Mehr, S. H. M.; Caramelli, D.; Cronin, L. Digitizing Chemical Discovery with a Bayesian Explorer for Interpreting Reactivity Data. *Proc. Natl. Acad. Sci.* **2023**, *120*, No. e2220045120.
- (477) Porwol, L.; Kowalski, D.; Henson, A.; Long, D.-L.; Bell, N.; Cronin, L. An Autonomous Chemical Robot Discovers the Rules of Inorganic Chemistry without Prior Knowledge. *Angew. Chem.* **2020**, *132* (28), 11352–11357.
- (478) Olansky, A. S.; Deming, S. N. Automated Development of a Kinetic Method for the Continuous-Flow Determination of Creatinine. *Clin. Chem.* **1978**, *24*, 2115–2124.
- (479) McMullen, J. P.; Jensen, K. F. Rapid Determination of Reaction Kinetics with an Automated Microfluidic System. *Org. Process Res. Dev.* **2011**, *15*, 398–407.
- (480) Box, G. E. P.; Hill, W. J. Discrimination Among Mechanistic Models. *Technometrics* **1967**, *9*, 57–71.
- (481) Reizman, B. J.; Jensen, K. F. An Automated Continuous-Flow Platform for the Estimation of Multistep Reaction Kinetics. *Org. Process Res. Dev.* **2012**, *16*, 1770–1782.
- (482) Sheng, H.; Sun, J.; Rodriguez, O.; Hoar, B.; Zhang, W.; Xiang, D.; Tang, T.; Hazra, A.; Min, D.; Doyle, A.; et al. Autonomous Closed-Loop Mechanistic Investigation of Molecular Electrochemistry via Automation. *ChemRxiv*, 2023. DOI: 10.26434/chemrxiv-2023-psqjx (accessed June 17, 2024).
- (483) Hoar, B. B.; Zhang, W.; Xu, S.; Deeba, R.; Costentin, C.; Gu, Q.; Liu, C. Electrochemical Mechanistic Analysis from Cyclic Voltammograms Based on Deep Learning. *ACS Meas. Sci. Au.* **2022**, *2*, 595–604.
- (484) Li, Y.; Yu, J. Emerging Applications of Zeolites in Catalysis, Separation and Host-Guest Assembly. *Nat. Rev. Mater.* **2021**, *6*, 1156–1174.
- (485) Yusuf, V. F.; Malek, N. I.; Kailasa, S. K. Review on Metal-Organic Framework Classification, Synthetic Approaches, and Influencing Factors: Applications in Energy, Drug Delivery, and Wastewater Treatment. *ACS Omega* **2022**, *7*, 44507–44531.
- (486) Pardakhti, M.; Moharreri, E.; Wanik, D.; Suib, S. L.; Srivastava, R. Machine Learning Using Combined Structural and Chemical Descriptors for Prediction of Methane Adsorption Perform-
- ance of Metal Organic Frameworks (MOFs). *ACS Comb. Sci.* **2017**, *19*, 640–645.
- (487) Shi, Z.; Yang, W.; Deng, X.; Cai, C.; Yan, Y.; Liang, H.; Liu, Z.; Qiao, Z. Machine-Learning-Assisted High-Throughput Computational Screening of High Performance Metal-Organic Frameworks. *Mol. Syst. Des. Eng.* **2020**, *5*, 725–742.
- (488) Tran, K.; Ulissi, Z. W. Active Learning across Intermetallics to Guide Discovery of Electrocatalysts for CO<sub>2</sub> Reduction and H<sub>2</sub> Evolution. *Nat. Catal.* **2018**, *1*, 696–703.
- (489) Zhang, X.; Xu, Z.; Wang, Z.; Liu, H.; Zhao, Y.; Jiang, S. High-Throughput and Machine Learning Approaches for the Discovery of Metal Organic Frameworks. *APL Mater.* **2023**, *11*, 060901.
- (490) Akporiaye, D. E.; Dahl, I. M.; Karlsson, A.; Wendelbo, R. Combinatorial Approach to the Hydrothermal Synthesis of Zeolites. *Angew. Chem. Int. Ed.* **1998**, *37*, 609–611.
- (491) Moliner, M.; Serra, J. M.; Corma, A.; Argente, E.; Valero, S.; Botti, V. Application of Artificial Neural Networks to High-Throughput Synthesis of Zeolites. *Microporous Mesoporous Mater.* **2005**, *78*, 73–81.
- (492) Greenaway, R. L.; Santolini, V.; Bennison, M. J.; Alston, B. M.; Pugh, C. J.; Little, M. A.; Miklitz, M.; Eden-Rump, E. G. B.; Clowes, R.; Shakil, A.; et al. High-Throughput Discovery of Organic Cages and Cateneanes Using Computational Screening Fused with Robotic Synthesis. *Nat. Commun.* **2018**, *9*, 2849.
- (493) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. *Chem. Rev.* **2020**, *120*, 8066–8129.
- (494) Muraoka, K.; Sada, Y.; Miyazaki, D.; Chaikittisilp, W.; Okubo, T. Linking Synthesis and Structure Descriptors from a Large Collection of Synthetic Records of Zeolite Materials. *Nat. Commun.* **2019**, *10*, 4459.
- (495) Greenaway, R. L.; Santolini, V.; Pulido, A.; Little, M. A.; Alston, B. M.; Briggs, M. E.; Day, G. M.; Cooper, A. I.; Jelfs, K. E. From Concept to Crystals via Prediction: Multi-Component Organic Cage Pots by Social Self-Sorting. *Angew. Chem. Int. Ed.* **2019**, *58*, 16275–16281.
- (496) Cui, P.; McMahon, D. P.; Spackman, P. R.; Alston, B. M.; Little, M. A.; Day, G. M.; Cooper, A. I. Mining Predicted Crystal Structure Landscapes with High Throughput Crystallisation: Old Molecules, New Insights. *Chem. Sci.* **2019**, *10*, 9988–9997.
- (497) Borboudakis, G.; Stergiannakos, T.; Frysali, M.; Klontzas, E.; Tsamardinos, I.; Froudakis, G. E. Chemically Intuited, Large-Scale Screening of MOFs by Machine Learning Techniques. *Npj Comput. Mater.* **2017**, *3*, 40.
- (498) Bucior, B. J.; Bobbitt, N. S.; Islamoglu, T.; Goswami, S.; Gopalan, A.; Yildirim, T.; Farha, O. K.; Bagheri, N.; Snurr, R. Q. Energy-Based Descriptors to Rapidly Predict Hydrogen Storage in Metal-Organic Frameworks. *Mol. Syst. Des. Eng.* **2019**, *4*, 162–174.
- (499) Altintas, C.; Altundal, O. F.; Keskin, S.; Yildirim, R. Machine Learning Meets with Metal Organic Frameworks for Gas Storage and Separation. *J. Chem. Inf. Model.* **2021**, *61*, 2131–2146.
- (500) Moliner, M.; Román-Leshkov, Y.; Corma, A. Machine Learning Applied to Zeolite Synthesis: The Missing Link for Realizing High-Throughput Discovery. *Acc. Chem. Res.* **2019**, *52*, 2971–2980.
- (501) Greenaway, R. L.; Jelfs, K. E. Integrating Computational and Experimental Workflows for Accelerated Organic Materials Discovery. *Adv. Mater.* **2021**, *33*, 2004831.
- (502) Corma, A.; Moliner, M.; Serra, J. M.; Serna, P.; Díaz-Cabañas, M. J.; Baumes, L. A. A New Mapping/Exploration Approach for HT Synthesis of Zeolites. *Chem. Mater.* **2006**, *18*, 3287–3296.
- (503) 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*, 1–6.
- (504) Nikolaev, P.; Hooper, D.; Perea-López, N.; Terrones, M.; Maruyama, B. Discovery of Wall-Selective Carbon Nanotube Growth Conditions via Automated Experimentation. *ACS Nano.* **2014**, *8*, 10214–10222.

- (505) Chang, J.; Nikolaev, P.; Carpena-Núñez, J.; Rao, R.; Decker, K.; Islam, A. E.; Kim, J.; Pitt, M. A.; Myung, J. I.; Maruyama, B. Efficient Closed-Loop Maximization of Carbon Nanotube Growth Rate Using Bayesian Optimization. *Sci. Rep.* **2020**, *10*, 9040.
- (506) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature*. **2016**, *533*, 73–76.
- (507) Xie, Y.; Zhang, C.; Hu, X.; Zhang, C.; Kelley, S. P.; Atwood, J. L.; Lin, J. Machine Learning Assisted Synthesis of Metal-Organic Nanocapsules. *J. Am. Chem. Soc.* **2020**, *142*, 1475–1481.
- (508) Luo, Y.; Bag, S.; Zaremba, O.; Cierpka, A.; Andreo, J.; Wuttke, S.; Friederich, P.; Tsotsalas, M. MOF Synthesis Prediction Enabled by Automatic Data Mining and Machine Learning. *Angew. Chem. Int. Ed.* **2022**, *61*, No. e202200242.
- (509) Chui, S. S.-Y.; Lo, S. M.-F.; Charmant, J. P. H.; Orpen, A. G.; Williams, I. D. A Chemically Functionalizable Nanoporous Material [Cu<sub>3</sub>(TMA)<sub>2</sub>(H<sub>2</sub>O)<sub>3</sub>]<sub>n</sub>. *Science*. **1999**, *283*, 1148–1150.
- (510) Moosavi, S. M.; Chidambaram, A.; Talirz, L.; Haranczyk, M.; Stylianou, K. C.; Smit, B. Capturing Chemical Intuition in Synthesis of Metal-Organic Frameworks. *Nat. Commun.* **2019**, *10*, 539.
- (511) Xie, Y.; Zhang, C.; Deng, H.; Zheng, B.; Su, J.-W.; Shutt, K.; Lin, J. Accelerate Synthesis of Metal-Organic Frameworks by a Robotic Platform and Bayesian Optimization. *ACS Appl. Mater. Interfaces*. **2021**, *13*, 53485–53491.
- (512) Xie, Y.; Zhang, C.; Su, J.-W.; Deng, H.; Zhang, C.; Lin, J. Rapid Synthesis of Zeolitic Imidazole Frameworks in Laser-Induced Graphene Microreactors. *ChemSusChem*. **2019**, *12*, 473–479.
- (513) Pilz, L.; Natzeck, C.; Wohlgemuth, J.; Scheuermann, N.; Weidler, P. G.; Wagner, I.; Wöll, C.; Tsotsalas, M. Fully Automated Optimization of Robot-Based MOF Thin Film Growth via Machine Learning Approaches. *Adv. Mater. Interfaces*. **2023**, *10*, 2201771.
- (514) Duros, V.; Grizou, J.; Xuan, W.; Hosni, Z.; Long, D.-L.; Miras, H. N.; Cronin, L. Human versus Robots in the Discovery and Crystallization of Gigantic Polyoxometalates. *Angew. Chem. Int. Ed.* **2017**, *56*, 10815–10820.
- (515) Masubuchi, S.; Morimoto, M.; Morikawa, S.; Onodera, M.; Asakawa, Y.; Watanabe, K.; Taniguchi, T.; Machida, T. Autonomous Robotic Searching and Assembly of Two-Dimensional Crystals to Build van Der Waals Superlattices. *Nat. Commun.* **2018**, *9*, 1413.
- (516) Kusne, A. G.; Yu, H.; Wu, C.; Zhang, H.; Hattrick-Simpers, J.; DeCost, B.; Sarker, S.; Oses, C.; Toher, C.; Curtarolo, S.; et al. On-the-Fly Closed-Loop Materials Discovery via Bayesian Active Learning. *Nat. Commun.* **2020**, *11*, 5966.
- (517) Ament, S.; Amsler, M.; Sutherland, D. R.; Chang, M.-C.; Guevarra, D.; Connolly, A. B.; Gregoire, J. M.; Thompson, M. O.; Gomes, C. P.; van Dover, R. B. Autonomous Materials Synthesis via Hierarchical Active Learning of Nonequilibrium Phase Diagrams. *Sci. Adv.* **2021**, *7*, No. eabg4930.
- (518) Chen, J.; Cross, S. R.; Miara, L. J.; Cho, J.-J.; Wang, Y.; Sun, W. Navigating Phase Diagram Complexity to Guide Robotic Inorganic Materials Synthesis. *arXiv* **2023**. DOI: [10.48550/arXiv.2304.00743](https://doi.org/10.48550/arXiv.2304.00743) (accessed November 24, 2023).
- (519) Service, R. AI-Driven Robots Start Hunting for Novel Materials without Help from Humans. *Science* **2023**. DOI: [10.1126/science.adj3613](https://doi.org/10.1126/science.adj3613).
- (520) Szymanski, N. J.; Nevatia, P.; Bartel, C. J.; Zeng, Y.; Ceder, G. Autonomous and Dynamic Precursor Selection for Solid-State Materials Synthesis. *Nat. Commun.* **2023**, *14*, 6956.
- (521) Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling Deep Learning for Materials Discovery. *Nature* **2023**, *624*, 80.
- (522) Buonassisi, T. On Characterization of “Novel Materials” from High-Throughput & Self-Driving Labs | LinkedIn. <https://www.linkedin.com/pulse/characterization-novel-materials-from-high-throughput-buonassisi-uoyke/> (accessed 2023-12-05).
- (523) Leeman, J.; Liu, Y.; Stiles, J.; Lee, S.; Bhatt, P.; Schoop, L.; Palgrave, R. Challenges in High-Throughput Inorganic Materials Prediction and Autonomous Synthesis. *ChemRxiv* **2024**. 3 DOI: [10.1103/PRXEnergy.3.011002](https://doi.org/10.1103/PRXEnergy.3.011002) (accessed January 16, 2024)
- (524) Kondo, M.; Sugizaki, A.; Khalid, M. I.; Wathsala, H. D. P.; Ishikawa, K.; Hara, S.; Takaai, T.; Washio, T.; Takizawa, S.; Sasai, H. Energy-, Time-, and Labor-Saving Synthesis of  $\alpha$ -Ketiminophosphonates: Machine-Learning-Assisted Simultaneous Multiparameter Screening for Electrochemical Oxidation. *Green Chem.* **2021**, *23*, 5825–5831.
- (525) Naito, Y.; Kondo, M.; Nakamura, Y.; Shida, N.; Ishikawa, K.; Washio, T.; Takizawa, S.; Atobe, M. Bayesian Optimization with Constraint on Passed Charge for Multiparameter Screening of Electrochemical Reductive Carboxylation in a Flow Microreactor. *Chem. Commun.* **2022**, *58*, 3893–3896.
- (526) Kondo, M.; Wathsala, H. D. P.; Sako, M.; Hanatani, Y.; Ishikawa, K.; Hara, S.; Takaai, T.; Washio, T.; Takizawa, S.; Sasai, H. Exploration of Flow Reaction Conditions Using Machine-Learning for Enantioselective Organocatalyzed Rauhut-Currier and [3+2] Annulation Sequence. *Chem. Commun.* **2020**, *56*, 1259–1262.
- (527) Jorayev, P.; Russo, D.; Tibbetts, J. D.; Schweidtmann, A. M.; Deutsch, P.; Bull, S. D.; Lapkin, A. A. Multi-Objective Bayesian Optimisation of a Two-Step Synthesis of p-Cymene from Crude Sulphate Turpentine. *Chem. Eng. Sci.* **2022**, *247*, 116938.
- (528) Eyke, N. S.; Green, W. H.; Jensen, K. F. Iterative Experimental Design Based on Active Machine Learning Reduces the Experimental Burden Associated with Reaction Screening. *React. Chem. Eng.* **2020**, *5*, 1963–1972.
- (529) Dunlap, J. H.; Ethier, J. G.; Putnam-Neel, A. A.; Iyer, S.; Luo, S.-X. L.; Feng, H.; Garrido Torres, J. A.; Doyle, A. G.; Swager, T. M.; Vaia, R. A.; et al. Continuous Flow Synthesis of Pyridinium Salts Accelerated by Multi-Objective Bayesian Optimization with Active Learning. *Chem. Sci.* **2023**, *14*, 8061–8069.
- (530) Mullard, A. R&D Budgets Boom, but Success Rates Falter. *Nat. Rev. Drug Discov.* **2022**, *21*, 249–249.
- (531) DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. *J. Health Econ.* **2016**, *47*, 20–33.
- (532) Sun, D.; Gao, W.; Hu, H.; Zhou, S. Why 90% of Clinical Drug Development Fails and How to Improve It? *Acta Pharm. Sin. B*. **2022**, *12*, 3049–3062.
- (533) Chapman, T. Lab Automation and Robotics: Automation on the Move. *Nature*. **2003**, *421*, 661–663 665–666.
- (534) Pereira, D. A.; Williams, J. A. Origin and Evolution of High Throughput Screening. *Br. J. Pharmacol.* **2007**, *152*, 53–61.
- (535) Schneider, G. Automating Drug Discovery. *Nat. Rev. Drug Discov.* **2018**, *17*, 97–113.
- (536) Graff, D. E.; Shakhnovich, E. I.; Coley, C. W. Accelerating High-Throughput Virtual Screening through Molecular Pool-Based Active Learning. *Chem. Sci.* **2021**, *12*, 7866–7881.
- (537) Sadybekov, A. V.; Katritch, V. Computational Approaches Streamlining Drug Discovery. *Nature*. **2023**, *616*, 673–685.
- (538) Jayatunga, M. K. P.; Xie, W.; Ruder, L.; Schulze, U.; Meier, C. AI in Small-Molecule Drug Discovery: A Coming Wave? *Nat. Rev. Drug Discov.* **2022**, *21*, 175–176.
- (539) Nagra, N. S.; van der Veken, L.; Stanzl, E.; Champagne, D.; Devereson, A.; Macak, M. The Company Landscape for Artificial Intelligence in Large-Molecule Drug Discovery. *Nat. Rev. Drug Discov.* **2023**, *22*, 949.
- (540) Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackermann, Z.; et al. A Deep Learning Approach to Antibiotic Discovery. *Cell*. **2020**, *180*, 688–702 e13.
- (541) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature*. **2021**, *596*, 583–589.
- (542) Sadybekov, A. A.; Sadybekov, A. V.; Liu, Y.; Iliopoulos-Tsoutsouvas, C.; Huang, X.-P.; Pickett, J.; Houser, B.; Patel, N.; Tran, N. K.; Tong, F.; et al. Synthon-Based Ligand Discovery in Virtual Libraries of over 11 Billion Compounds. *Nature*. **2022**, *601*, 452–459.

- (543) Gorgulla, C.; Boeszoeremenyi, A.; Wang, Z.-F.; Fischer, P. D.; Coote, P. W.; Padmanabha Das, K. M.; Malets, Y. S.; Radchenko, D. S.; Moroz, Y. S.; Scott, D. A.; et al. An Open-Source Drug Discovery Platform Enables Ultra-Large Virtual Screens. *Nature*. **2020**, *580*, 663–668.
- (544) Ingraham, J. B.; Baranov, M.; Costello, Z.; Barber, K. W.; Wang, W.; Ismail, A.; Frappier, V.; Lord, D. M.; Ng-Thow-Hing, C.; Van Slack, E. R.; et al. Illuminating Protein Space with a Programmable Generative Model. *Nature*. **2023**, *623*, 1070–1078.
- (545) Watson, J. L.; Juergens, D.; Bennett, N. R.; Trippe, B. L.; Yim, J.; Eisenach, H. E.; Ahern, W.; Borst, A. J.; Ragotte, R. J.; Milles, L. F.; et al. De Novo Design of Protein Structure and Function with RFdiffusion. *Nature*. **2023**, *620*, 1089–1100.
- (546) Ren, F.; Ding, X.; Zheng, M.; Korzinkin, M.; Cai, X.; Zhu, W.; Mantysyzov, A.; Aliper, A.; Aladinskiy, V.; Cao, Z.; et al. AlphaFold Accelerates Artificial Intelligence Powered Drug Discovery: Efficient Discovery of a Novel CDK20 Small Molecule Inhibitor. *Chem. Sci.* **2023**, *14*, 1443–1452.
- (547) Saikin, S. K.; Kreisbeck, C.; Sheberla, D.; Becker, J. S.; Aspuru-Guzik, A. Closed-Loop Discovery Platform Integration Is Needed for Artificial Intelligence to Make an Impact in Drug Discovery. *Expert Opin. Drug Discov.* **2019**, *14*, 1–4.
- (548) Pun, F. W.; Ozerov, I. V.; Zhavoronkov, A. AI-Powered Therapeutic Target Discovery. *Trends Pharmacol. Sci.* **2023**, *44*, 561–572.
- (549) Sparkes, A.; Aubrey, W.; Byrne, E.; Clare, A.; Khan, M. N.; Liakata, M.; Markham, M.; Rowland, J.; Soldatova, L. N.; Whelan, K. E.; et al. Towards Robot Scientists for Autonomous Scientific Discovery. *Autom. Exp.* **2010**, *2*, 1.
- (550) Carbonell, P.; Radivojevic, T.; García Martín, H. Opportunities at the Intersection of Synthetic Biology, Machine Learning, and Automation. *ACS Synth. Biol.* **2019**, *8*, 1474–1477.
- (551) Whittingham, H.; Ashenden, S. K. Chapter 5 - Hit Discovery. In *The Era of Artificial Intelligence, Machine Learning, and Data Science in the Pharmaceutical Industry*; Ashenden, S. K., Ed.; Academic Press, 2021; pp 81–102. DOI: [10.1016/B978-0-12-820045-2.00006-4](https://doi.org/10.1016/B978-0-12-820045-2.00006-4).
- (552) Hughes, J.; Rees, S.; Kalindjian, S.; Philpott, K. Principles of Early Drug Discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239–1249.
- (553) Grisoni, F.; Huisman, B. J. H.; Button, A. L.; Moret, M.; Atz, K.; Merk, D.; Schneider, G. Combining Generative Artificial Intelligence and On-Chip Synthesis for de Novo Drug Design. *Sci. Adv.* **2021**, *7*, No. eabg3338.
- (554) Elder, S.; Klumpp-Thomas, C.; Yasgar, A.; Travers, J.; Frebert, S.; Wilson, K. M.; Zakharov, A. V.; Dahlin, J. L.; Kreisbeck, C.; Sheberla, D.; et al. Cross-Platform Bayesian Optimization System for Autonomous Biological Assay Development. *SLAS Technol. Transl. Life Sci. Innov.* **2021**, *26*, 579–590.
- (555) Kanda, G. N.; Tsuzuki, T.; Terada, M.; Sakai, N.; Motozawa, N.; Masuda, T.; Nishida, M.; Watanabe, C. T.; Higashi, T.; Horiguchi, S. A.; et al. Robotic Search for Optimal Cell Culture in Regenerative Medicine. *eLife*. **2022**, *11*, No. e77007.
- (556) Bisswanger, H. Enzyme Assays. *Perspect. Sci.* **2014**, *1*, 41–55.
- (557) Zhang, J. H.; Chung, T. D.; Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen.* **1999**, *4*, 67–73.
- (558) Desai, B.; Dixon, K.; Farrant, E.; Feng, Q.; Gibson, K. R.; van Hoorn, W. P.; Mills, J.; Morgan, T.; Parry, D. M.; Ramjee, M. K.; et al. Rapid Discovery of a Novel Series of Abl Kinase Inhibitors by Application of an Integrated Microfluidic Synthesis and Screening Platform. *J. Med. Chem.* **2013**, *56*, 3033–3047.
- (559) Czechitzky, W.; Dedio, J.; Desai, B.; Dixon, K.; Farrant, E.; Feng, Q.; Morgan, T.; Parry, D. M.; Ramjee, M. K.; Selway, C. N.; et al. Integrated Synthesis and Testing of Substituted Xanthine Based DPP4 Inhibitors: Application to Drug Discovery. *ACS Med. Chem. Lett.* **2013**, *4*, 768–772.
- (560) Pant, S. M.; Mukonoweshuro, A.; Desai, B.; Ramjee, M. K.; Selway, C. N.; Tarver, G. J.; Wright, A. G.; Birchall, K.; Chapman, T. M.; Tervonen, T. A.; et al. Design, Synthesis, and Testing of Potent, Selective Hepsin Inhibitors via Application of an Automated Closed-Loop Optimization Platform. *J. Med. Chem.* **2018**, *61*, 4335–4347.
- (561) Brocklehurst, C. E.; Altmann, E.; Bon, C.; Davis, H.; Dunstan, D.; Ertl, P.; Ginsburg-Moraff, C.; Grob, J.; Gosling, D. J.; Lapointe, G.; et al. MicroCycle: An Integrated and Automated Platform to Accelerate Drug Discovery. *J. Med. Chem.* **2024**, *67*, 2118–2128.
- (562) Bao, Z.; Bufton, J.; Hickman, R. J.; Aspuru-Guzik, A.; Bannigan, P.; Allen, C. Revolutionizing Drug Formulation Development: The Increasing Impact of Machine Learning. *Adv. Drug Deliv. Rev.* **2023**, *202*, 115108.
- (563) Cao, L.; Russo, D.; Felton, K.; Salley, D.; Sharma, A.; Keenan, G.; Mauer, W.; Gao, H.; Cronin, L.; Lapkin, A. A. Optimization of Formulations Using Robotic Experiments Driven by Machine Learning DoE. *Cell Rep. Phys. Sci.* **2021**, *2*, 100295.
- (564) Grizou, J.; Points, L. J.; Sharma, A.; Cronin, L. A Curious Formulation Robot Enables the Discovery of a Novel Protocell Behavior. *Sci. Adv.* **2020**, *6*, No. eaay4237.
- (565) Hickman, R. J.; Bannigan, P.; Bao, Z.; Aspuru-Guzik, A.; Allen, C. Self-Driving Laboratories: A Paradigm Shift in Nanomedicine Development. *Matter.* **2023**, *6*, 1071.
- (566) Mitchell, M. J.; Billingsley, M. M.; Haley, R. M.; Wechsler, M. E.; Peppas, N. A.; Langer, R. Engineering Precision Nanoparticles for Drug Delivery. *Nat. Rev. Drug Discov.* **2021**, *20*, 101–124.
- (567) Fan, Y.; Yen, C.-W.; Lin, H.-C.; Hou, W.; Estevez, A.; Sarode, A.; Goyon, A.; Bian, J.; Lin, J.; Koenig, S. G.; et al. Automated High-Throughput Preparation and Characterization of Oligonucleotide-Loaded Lipid Nanoparticles. *Int. J. Pharm.* **2021**, *599*, 120392.
- (568) Sarode, A.; Fan, Y.; Byrnes, A. E.; Hammel, M.; Hura, G. L.; Fu, Y.; Kou, P.; Hu, C.; Hinz, F. I.; Roberts, J.; et al. Predictive High-Throughput Screening of PEGylated Lipids in Oligonucleotide-Loaded Lipid Nanoparticles for Neuronal Gene Silencing. *Nanoscale Adv.* **2022**, *4*, 2107–2123.
- (569) Tamasi, M. J.; Gormley, A. J. Biologic Formulation in a Self-Driving Biomaterials Lab. *Cell Rep. Phys. Sci.* **2022**, *3*, 101041.
- (570) Adamo, A.; Beingessner, R. L.; Behnam, M.; Chen, J.; Jamison, T. F.; Jensen, K. F.; Monbaliu, J.-C. M.; Myerson, A. S.; Revalor, E. M.; Snead, D. R.; et al. On-Demand Continuous-Flow Production of Pharmaceuticals in a Compact, Reconfigurable System. *Science*. **2016**, *352*, 61–67.
- (571) Ortiz-Perez, A.; van Tilborg, D.; van der Meel, R.; Grisoni, F.; Albertazzi, L. Machine Learning-Guided High Throughput Nanoparticle Design. *Digit. Discov.* **2024**, DOI: [10.1039/D4DD00104D](https://doi.org/10.1039/D4DD00104D).
- (572) Lammers, T.; Storm, G. Setting Standards to Promote Progress in Bio-Nano Science. *Nat. Nanotechnol.* **2019**, *14*, 626–626.
- (573) Martin, H. G.; Radivojevic, T.; Zucker, J.; Bouchard, K.; Sustarich, J.; Peisert, S.; Arnold, D.; Hillson, N.; Babnigg, G.; Marti, J. M.; et al. Perspectives for Self-Driving Labs in Synthetic Biology. *Curr. Opin. Biotechnol.* **2023**, *79*, 102881.
- (574) Sanders, L. M.; Scott, R. T.; Yang, J. H.; Qutub, A. A.; Garcia Martin, H.; Berrios, D. C.; Hastings, J. J. A.; Rask, J.; Mackintosh, G.; Hoarfrost, A. L.; et al. Biological Research and Self-Driving Labs in Deep Space Supported by Artificial Intelligence. *Nat. Mach. Intell.* **2023**, *5*, 208–219.
- (575) Du, J.; Shao, Z.; Zhao, H. Engineering Microbial Factories for Synthesis of Value-Added Products. *J. Ind. Microbiol. Biotechnol.* **2011**, *38*, 873–890.
- (576) Carbonell, P.; Jervis, A. J.; Robinson, C. J.; Yan, C.; Dunstan, M.; Swainston, N.; Vinaixa, M.; Hollywood, K. A.; Currin, A.; Rattray, N. J. W.; et al. An Automated Design-Build-Test-Learn Pipeline for Enhanced Microbial Production of Fine Chemicals. *Commun. Biol.* **2018**, *1*, 1–10.
- (577) Hamedirad, M.; Chao, R.; Weisberg, S.; Lian, J.; Sinha, S.; Zhao, H. Towards a Fully Automated Algorithm Driven Platform for Biosystems Design. *Nat. Commun.* **2019**, *10*, 5150.
- (578) Chao, R.; Liang, J.; Tasan, I.; Si, T.; Ju, L.; Zhao, H. Fully Automated One-Step Synthesis of Single-Transcript TALEN Pairs Using a Biological Foundry. *ACS Synth. Biol.* **2017**, *6*, 678–685.

- (579) Engler, C.; Kandzia, R.; Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS ONE*. **2008**, *3*, No. e3647.
- (580) Balcão, V. M.; Vila, M. M. D. C. Structural and Functional Stabilization of Protein Entities: State-of-the-Art. *Adv. Drug Deliv. Rev.* **2015**, *93*, 25–41.
- (581) Tamasi, M. J.; Patel, R. A.; Borca, C. H.; Kosuri, S.; Mugnier, H.; Upadhyay, R.; Murthy, N. S.; Webb, M. A.; Gormley, A. J. Machine Learning on a Robotic Platform for the Design of Polymer-Protein Hybrids. *Adv. Mater.* **2022**, *34*, 2201809.
- (582) Rapp, J.; Bremer, B.; Romero, P. Self-Driving Laboratories to Autonomously Navigate the Protein Fitness Landscape. *bioRxiv* 2023. DOI: 10.1101/2023.05.20.541582 (accessed October 23, 2023)
- (583) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031–3048.
- (584) Arnold, C. Cloud Labs: Where Robots Do the Research. *Nature*. **2022**, *606*, 612–613.
- (585) Seok, J.; Warren, H. S.; Cuenca, A. G.; Mindrinos, M. N.; Baker, H. V.; Xu, W.; Richards, D. R.; McDonald-Smith, G. P.; Gao, H.; Hennessy, L.; et al. Genomic Responses in Mouse Models Poorly Mimic Human Inflammatory Diseases. *Proc. Natl. Acad. Sci.* **2013**, *110*, 3507–3512.
- (586) Low, L. A.; Mummery, C.; Berridge, B. R.; Austin, C. P.; Tagle, D. A. Organs-on-Chips: Into the next Decade. *Nat. Rev. Drug Discov.* **2021**, *20*, 345–361.
- (587) Han, J. J. FDA Modernization Act 2.0 Allows for Alternatives to Animal Testing. *Artif. Organs*. **2023**, *47*, 449–450.
- (588) Hillson, N.; Caddick, M.; Cai, Y.; Carrasco, J. A.; Chang, M. W.; Curach, N. C.; Bell, D. J.; Le Feuvre, R.; Friedman, D. C.; Fu, X.; et al. Building a Global Alliance of Biofoundries. *Nat. Commun.* **2019**, *10*, 2040.
- (589) Acceleration Consortium. <https://acceleration.utoronto.ca/> (accessed 2023-11-13).
- (590) Munos, B. Can Open-Source R&D Reinvigorate Drug Research? *Nat. Rev. Drug Discov.* **2006**, *5*, 723–729.
- (591) Wang, H.; Fu, T.; Du, Y.; Gao, W.; Huang, K.; Liu, Z.; Chandak, P.; Liu, S.; Van Katwyk, P.; Deac, A.; et al. Scientific Discovery in the Age of Artificial Intelligence. *Nature*. **2023**, *620*, 47–60.
- (592) Boby, M. L.; Fearon, D.; Ferla, M.; Filep, M.; Koekemoer, L.; Robinson, M. C.; Chodera, J. D.; Lee, A. A.; London, N.; et al. Open Science Discovery of Potent Noncovalent SARS-CoV-2 Main Protease Inhibitors. *Science*. **2023**, *382*, No. eabo7201.
- (593) Urbina, F.; Lentzos, F.; Invernizzi, C.; Ekins, S. Dual Use of Artificial-Intelligence-Powered Drug Discovery. *Nat. Mach. Intell.* **2022**, *4*, 189–191.
- (594) Mock, M.; Edavettal, S.; Langmead, C.; Russell, A. AI Can Help to Speed up Drug Discovery — but Only If We Give It the Right Data. *Nature*. **2023**, *621*, 467–470.
- (595) AI's Potential to Accelerate Drug Discovery Needs a Reality Check. *Nature* **2023**, *622*, 217–217.
- (596) Arnold, C. AlphaFold Touted as next Big Thing for Drug Discovery — but Is It? *Nature*. **2023**, *622*, 15–17.
- (597) Miracle, D.; Majumdar, B.; Wertz, K.; Gorsse, S. New Strategies and Tests to Accelerate Discovery and Development of Multi-Principal Element Structural Alloys. *Scr. Mater.* **2017**, *127*, 195–200.
- (598) Vecchio, K. S.; Dippo, O. F.; Kaufmann, K. R.; Liu, X. High-Throughput Rapid Experimental Alloy Development (HT-READ). *Acta Mater.* **2021**, *221*, 117352.
- (599) DeCost, B.; Joress, H.; Sarker, S.; Mehta, A.; Hattrick-Simpers, J. Towards Automated Design of Corrosion Resistant Alloy Coatings with an Autonomous Scanning Droplet Cell. *JOM*. **2022**, *74*, 2941–2950.
- (600) Monteiro, P. J. M.; Miller, S. A.; Horvath, A. Towards Sustainable Concrete. *Nat. Mater.* **2017**, *16*, 698–699.
- (601) Forcael, E.; Ferrari, I.; Opazo-Vega, A.; Pulido-Arcas, J. A. Construction 4.0: A Literature Review. *Sustainability*. **2020**, *12*, 9755.
- (602) Koch, C.; Brilakis, I. Pothole Detection in Asphalt Pavement Images. *Adv. Eng. Inform.* **2011**, *25*, 507–515.
- (603) Koch, C.; Georgieva, K.; Kasireddy, V.; Akinci, B.; Fieguth, P. A Review on Computer Vision Based Defect Detection and Condition Assessment of Concrete and Asphalt Civil Infrastructure. *Adv. Eng. Inform.* **2015**, *29*, 196–210.
- (604) Zhang, A.; Wang, K. C. P.; Li, B.; Yang, E.; Dai, X.; Peng, Y.; Fei, Y.; Liu, Y.; Li, J. Q.; Chen, C. Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces Using a Deep-Learning Network. *Comput.-Aided Civ. Infrastruct. Eng.* **2017**, *32*, 805–819.
- (605) Cha, Y.-J.; Choi, W.; Suh, G.; Mahmoudkhani, S.; Büyüköztürk, O. Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 731–747.
- (606) Völker, C.; Moreno Torres, B.; Rug, T.; Firdous, R.; Jan Zia, G. A.; Lüders, S.; Scaffino, H. L.; Höpler, M.; Böhmer, F.; Pfaff, M.; et al. Data Driven Design of Alkali-Activated Concrete Using Sequential Learning. *J. Clean. Prod.* **2023**, *418*, 138221.
- (607) Erps, T.; Foshey, M.; Luković, M. K.; Shou, W.; Goetzke, H. H.; Dietsch, H.; Stoll, K.; von Vacano, B.; Matusik, W. Accelerated Discovery of 3D Printing Materials Using Data-Driven Multiobjective Optimization. *Sci. Adv.* **2021**, *7*, No. eabf7435.
- (608) Gongora, A. E.; Xu, B.; Perry, W.; Okoye, C.; Riley, P.; Reyes, K. G.; Morgan, E. F.; Brown, K. A. A Bayesian Experimental Autonomous Researcher for Mechanical Design. *Sci. Adv.* **2020**, *6*, No. eaaz1708.
- (609) Gongora, A. E.; Snapp, K. L.; Whiting, E.; Riley, P.; Reyes, K. G.; Morgan, E. F.; Brown, K. A. Using Simulation to Accelerate Autonomous Experimentation: A Case Study Using Mechanics. *iScience*. **2021**, *24*, 102262.
- (610) Snapp, K. L.; Verdier, B.; Gongora, A. E.; Silverman, S.; Adesiji, A. D.; Morgan, E. F.; Lawton, T. J.; Whiting, E.; Brown, K. A. Superlative Mechanical Energy Absorbing Efficiency Discovered through Self-Driving Lab-Human Partnership. *Nat. Commun.* **2024**, *15*, 4290.
- (611) Cao, B.; Adutwum, L. A.; Oliynyk, A. O.; Luber, E. J.; Olsen, B. C.; Mar, A.; Buriak, J. M. How To Optimize Materials and Devices via Design of Experiments and Machine Learning: Demonstration Using Organic Photovoltaics. *ACS Nano*. **2018**, *12*, 7434–7444.
- (612) Kirkey, A.; Luber, E. J.; Cao, B.; Olsen, B. C.; Buriak, J. M. Optimization of the Bulk Heterojunction of All-Small-Molecule Organic Photovoltaics Using Design of Experiment and Machine Learning Approaches. *ACS Appl. Mater. Interfaces*. **2020**, *12*, 54596.
- (613) 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. A Materials Acceleration Platform for Organic Laser Discovery. *Adv. Mater.* **2023**, *35*, 2207070.
- (614) Sandanayaka, A. S. D.; Matsushima, T.; Bencheikh, F.; Yoshida, K.; Inoue, M.; Fujihara, T.; Goushi, K.; Ribierre, J.-C.; Adachi, C. Toward Continuous-Wave Operation of Organic Semiconductor Lasers. *Sci. Adv.* **2017**, *3*, No. e1602570.
- (615) Sandanayaka, A. S. D.; Matsushima, T.; Bencheikh, F.; Terakawa, S.; Potsavage, W. J.; Qin, C.; Fujihara, T.; Goushi, K.; Ribierre, J.-C.; Adachi, C. Indication of Current-Injection Lasing from an Organic Semiconductor. *Appl. Phys. Express*. **2019**, *12*, 061010.
- (616) Wang, T.; Li, R.; Ardekani, H.; Serrano-Luján, L.; Wang, J.; Ramezani, M.; Wilmington, R.; Chauhan, M.; Epps, R. W.; Darabi, K.; et al. Sustainable Materials Acceleration Platform Reveals Stable and Efficient Wide-Bandgap Metal Halide Perovskite Alloys. *Matter*. **2023**, *6*, 2963–2986.
- (617) Higgins, K.; Valletti, S. M.; Ziatdinov, M.; Kalinin, S. V.; Ahmadi, M. Chemical Robotics Enabled Exploration of Stability in Multicomponent Lead Halide Perovskites via Machine Learning. *ACS Energy Lett.* **2020**, *5*, 3426–3436.
- (618) Li, Z.; Najeeb, M. A.; Alves, L.; Sherman, A. Z.; Shekar, V.; Cruz Parrilla, P.; Pendleton, I. M.; Wang, W.; Nega, P. W.; Zeller, M.,

- et al. Robot-Accelerated Perovskite Investigation and Discovery. *Chem. Mater.* **2020**, *32*, 5650–5663.
- (619) Saidaminov, M. I.; Abdellahy, A. L.; Murali, B.; Alarousu, E.; Burlakov, V. M.; Peng, W.; Dursun, I.; Wang, L.; He, Y.; Maculan, G.; et al. High-Quality Bulk Hybrid Perovskite Single Crystals within Minutes by Inverse Temperature Crystallization. *Nat. Commun.* **2015**, *6*, 7586.
- (620) Kirman, J.; Johnston, A.; Kuntz, D. A.; Askerka, M.; Gao, Y.; Todorović, P.; Ma, D.; Privé, G. G.; Sargent, E. H. Machine-Learning-Accelerated Perovskite Crystallization. *Matter.* **2020**, *2*, 938–947.
- (621) Shi, D.; Adinolfi, V.; Comin, R.; Yuan, M.; Alarousu, E.; Buin, A.; Chen, Y.; Hoogland, S.; Rothenberger, A.; Katsiev, K.; et al. Low Trap-State Density and Long Carrier Diffusion in Organolead Trihalide Perovskite Single Crystals. *Science.* **2015**, *347*, 519–522.
- (622) Huyer, W.; Neumaier, A. SNOBFIT - Stable Noisy Optimization by Branch and Fit. *ACM Trans. Math. Softw.* **2008**, *35*, 1–25.
- (623) Li, S. W.; Baker, R.; Lignos, I.; Yang, Z.; Stavrakis, S. D.; Howes, P. J.; deMello, A. Automated Microfluidic Screening of Ligand Interactions during the Synthesis of Cesium Lead Bromide Nanocrystals. *Mol. Syst. Des. Eng.* **2020**, *5*, 1118–1130.
- (624) Salley, D.; Keenan, G.; Grizou, J.; Sharma, A.; Martín, S.; Cronin, L. A Nanomaterials Discovery Robot for the Darwinian Evolution of Shape Programmable Gold Nanoparticles. *Nat. Commun.* **2020**, *11*, 2771.
- (625) Nikoobakht, B.; El-Sayed, M. A. Preparation and Growth Mechanism of Gold Nanorods (NRs) Using Seed-Mediated Growth Method. *Chem. Mater.* **2003**, *15*, 1957–1962.
- (626) Jiang, Y.; Salley, D.; Sharma, A.; Keenan, G.; Mullin, M.; Cronin, L. An Artificial Intelligence Enabled Chemical Synthesis Robot for Exploration and Optimization of Nanomaterials. *Sci. Adv.* **2022**, *8*, No. eab02626.
- (627) Mouret, J.-B.; Clune, J. Illuminating Search Spaces by Mapping Elites. *arXiv*2015. DOI: [10.48550/arXiv.1504.04909](https://doi.org/10.48550/arXiv.1504.04909) (accessed October 30, 2023)
- (628) 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.
- (629) Low, A. K. Y.; Mekki-Berrada, F.; Gupta, A.; Ostudin, A.; Xie, J.; Vissol-Gaudin, E.; Lim, Y.-F.; Li, Q.; Ong, Y. S.; Khan, S. A.; et al. Evolution-Guided Bayesian Optimization for Constrained Multi-Objective Optimization in Self-Driving Labs. *Npj Comput. Mater.* **2024**, *10*, 1–11.
- (630) Epps, R. W.; Bowen, M. S.; Volk, A. A.; Abdel-Latif, K.; Han, S.; Reyes, K. G.; Amassian, A.; Abolhasani, M. Artificial Chemist: An Autonomous Quantum Dot Synthesis Bot. *Adv. Mater.* **2020**, *32*, 2001626.
- (631) Abdel-Latif, K.; Epps, R. W.; Kerr, C. B.; Papa, C. M.; Castellano, F. N.; Abolhasani, M. Facile Room-Temperature Anion Exchange Reactions of Inorganic Perovskite Quantum Dots Enabled by a Modular Microfluidic Platform. *Adv. Funct. Mater.* **2019**, *29*, 1900712.
- (632) Abdel-Latif, K.; Epps, R. W.; Bateni, F.; Han, S.; Reyes, K. G.; Abolhasani, M. Self-Driven Multistep Quantum Dot Synthesis Enabled by Autonomous Robotic Experimentation in Flow. *Adv. Intell. Syst.* **2021**, *3*, 2000245.
- (633) Epps, R. W.; Volk, A. A.; Reyes, K.; Abolhasani, M. Accelerated AI Development for Autonomous Materials Synthesis in Flow. *Chem. Sci.* **2021**, *12*, 6025.
- (634) Ahn, J.; Lee, E.; Tan, J.; Yang, W.; Kim, B.; Moon, J. A New Class of Chiral Semiconductors: Chiral-Organic-Molecule-Incorporating Organic-Inorganic Hybrid Perovskites. *Mater. Horiz.* **2017**, *4*, 851–856.
- (635) Vikram, A.; Brudnak, K.; Zahid, A.; Shim, M. A.; Kenis, P. J. Accelerated Screening of Colloidal Nanocrystals Using Artificial Neural Network-Assisted Autonomous Flow Reactor Technology. *Nanoscale.* **2021**, *13*, 17028–17039.
- (636) Bateni, F.; Epps, R. W.; Antami, K.; Dargis, R.; Bennett, J. A.; Reyes, K. G.; Abolhasani, M. Autonomous Nanocrystal Doping by Self-Driving Fluidic Micro-Processors. *Adv. Intell. Syst.* **2022**, *4*, 2200017.
- (637) Bateni, F.; Sadeghi, S.; Orooji, N.; Bennett, J. A.; Punati, V. S.; Stark, C.; Wang, J.; Rosko, M. C.; Chen, O.; Castellano, F. N.; et al. Smart Dope: A Self-Driving Fluidic Lab for Accelerated Development of Doped Perovskite Quantum Dots. *Adv. Energy Mater.* **2024**, *14*, 2302303.
- (638) Zhao, H.; Chen, W.; Huang, H.; Sun, Z.; Chen, Z.; Wu, L.; Zhang, B.; Lai, F.; Wang, Z.; Adam, M. L.; et al. A Robotic Platform for the Synthesis of Colloidal Nanocrystals. *Nat. Synth.* **2023**, *2*, 1–10.
- (639) Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO: A Compressed-Sensing Method for Identifying the Best Low-Dimensional Descriptor in an Immensity of Offered Candidates. *Phys. Rev. Mater.* **2018**, *2*, 083802.
- (640) Volk, A. A.; Epps, R. W.; Yonemoto, D. T.; Masters, B. S.; Castellano, F. N.; Reyes, K. G.; Abolhasani, M. AlphaFlow: Autonomous Discovery and Optimization of Multi-Step Chemistry Using a Self-Driven Fluidic Lab Guided by Reinforcement Learning. *Nat. Commun.* **2023**, *14*, 1403.
- (641) Strieth-Kalthoff, F.; Hao, H.; Rathore, V.; Derasp, J.; Gaudin, T.; Angello, N. H.; Seifrid, M.; Trushina, E.; Guy, M.; Liu, J.; et al. Delocalized, Asynchronous, Closed-Loop Discovery of Organic Laser Emitters. *Science.* **2024**, *384*, No. eadk9227.
- (642) Koscher, B. A.; Canty, R. B.; McDonald, M. A.; Greenman, K. P.; McGill, C. J.; Bilodeau, C. L.; Jin, W.; Wu, H.; Vermeire, F. H.; Jin, B.; et al. Autonomous, Multiproperty-Driven Molecular Discovery: From Predictions to Measurements and Back. *Science.* **2023**, *382*, No. eadi1407.
- (643) Coley, C. Connorcoley/ASKCOS: First Public Release of ASKCOS, Zenodo2019. DOI: [10.5281/zenodo.3261361](https://doi.org/10.5281/zenodo.3261361) (accessed June 6, 2024)
- (644) 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. A Materials Acceleration Platform for Organic Laser Discovery. *Adv. Mater.* **2023**, *35*, 2207070.
- (645) Deshpande, A. V.; Beidoun, A.; Penzkofer, A.; Wagenblast, G. Absorption and Emission Spectroscopic Investigation of Cyanovinyldiethylaniline Dye Vapors. *Chem. Phys.* **1990**, *142*, 123–131.
- (646) Angello, N.; Friday, D.; Hwang, C.; Yi, S.; Cheng, A.; Torres-Flores, T.; Jira, E.; Wang, W.; Aspuru-Guzik, A.; Burke, M. et al. Closed-Loop Transfer Enables AI to Yield Chemical Knowledge. *ChemRxiv*2023. DOI: [10.26434/chemrxiv-2023-jqbqt](https://doi.org/10.26434/chemrxiv-2023-jqbqt) (accessed September 22, 2023)
- (647) Ludwig, A. Discovery of New Materials Using Combinatorial Synthesis and High-Throughput Characterization of Thin-Film Materials Libraries Combined with Computational Methods. *Npj Comput. Mater.* **2019**, *5*, 70.
- (648) Vriza, A.; Chan, H.; Xu, J. Self-Driving Laboratory for Polymer Electronics. *Chem. Mater.* **2023**, *35*, 3046.
- (649) Stroyuk, O.; Raievskaya, O.; Langner, S.; Kupfer, C.; Barabash, A.; Solonenko, D.; Azhniuk, Y.; Hauch, J.; Osvet, A.; Batentschuk, M.; et al. High-Throughput Robotic Synthesis and Photoluminescence Characterization of Aqueous Multinary Copper-Silver Indium Chalcogenide Quantum Dots. Part. Part. Syst. Charact. **2021**, *38*, 2100169.
- (650) Stroyuk, O.; Raievskaya, O.; Solonenko, D.; Kupfer, C.; Osvet, A.; Batentschuk, M. J.; Brabec, C. T.; Zahn, D. R. Spontaneous Alloying of Ultrasmall Non-Stoichiometric Ag-In-S and Cu-In-S Quantum Dots in Aqueous Colloidal Solutions. *RSC Adv.* **2021**, *11*, 21145–21152.
- (651) MacLeod, B. P.; Parlante, 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. Self-Driving Laboratory for Accelerated Discovery of Thin-Film Materials. *Sci. Adv.* **2020**, *6*, No. eaaz8867.
- (652) Häse, F.; Roch, L. M.; Kreisbeck, C.; Aspuru-Guzik, A. Phoenics: A Bayesian Optimizer for Chemistry. *ACS Cent. Sci.* **2018**, *4*, 1134–1145.

- (653) MacLeod, B. P.; Parlante, F. G. L.; Rupnow, C. C.; Dettelbach, K. E.; Elliott, M. S.; Morrissey, T. D.; Haley, T. H.; Proskurin, O.; Rooney, M. B.; Taherimakhousi, N.; et al. A Self-Driving Laboratory Advances the Pareto Front for Material Properties. *Nat. Commun.* **2022**, *13*, 995.
- (654) 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.
- (655) Xie, C.; Tang, X.; Berlinghof, M.; Langner, S.; Chen, S.; Späth, A.; Li, N.; Fink, R. H.; Unruh, T.; Brabec, C. J. Robot-Based High-Throughput Engineering of Alcoholic Polymer: Fullerene Nanoparticle Inks for an Eco-Friendly Processing of Organic Solar Cells. *ACS Appl. Mater. Interfaces.* **2018**, *10*, 23225–23234.
- (656) Du, X.; Lüer, L.; Heumueller, T.; Wagner, J.; Berger, C.; Osterrieder, T.; Wortmann, J.; Langner, S.; Vongsaysy, U.; Bertrand, M.; et al. Elucidating the Full Potential of OPV Materials Utilizing a High-Throughput Robot-Based Platform and Machine Learning. *Joule.* **2021**, *5*, 495–506.
- (657) Yuan, J.; Zhang, Y.; Zhou, L.; Zhang, G.; Yip, H.-L.; Lau, T.-K.; Lu, X.; Zhu, C.; Peng, H.; Johnson, P. A.; et al. Single-Junction Organic Solar Cell with over 15% Efficiency Using Fused-Ring Acceptor with Electron-Deficient Core. *Joule.* **2019**, *3*, 1140–1151.
- (658) Liu, Z.; Rolston, N.; Flick, A. C.; Colburn, T. W.; Ren, Z.; Dauskardt, R. H.; Buonassisi, T. Machine Learning with Knowledge Constraints for Process Optimization of Open-Air Perovskite Solar Cell Manufacturing. *Joule.* **2022**, *6*, 834–849.
- (659) Rolston, N.; Scheideler, W. J.; Flick, A. C.; Chen, J. P.; Elmaraghi, H.; Sleagh, A.; Zhao, O.; Woodhouse, M.; Dauskardt, R. H. Rapid Open-Air Fabrication of Perovskite Solar Modules. *Joule.* **2020**, *4*, 2675–2692.
- (660) Temple, J. *The lurking threat to solar power's growth*. MIT Technology Review. <https://www.technologyreview.com/2021/07/14/1028461/solar-value-deflation-california-climate-change/> (accessed 2023-11-14).
- (661) *Skeleton launching a fully automated supercapacitor production line*. <https://www.skeletontech.com/skeleton-blog/skeleton-will-launch-the-first-ever-fully-automated-supercapacitor-production-line> (accessed 2023-11-27).
- (662) Service, R. Ammonia—a renewable fuel made from sun, air, and water—could power the globe without carbon. <https://www.science.org/content/article/ammonia-renewable-fuel-made-sun-air-and-water-could-power-globe-without-carbon> (accessed 2023-11-14).
- (663) Bai, H.; Song, Z. Lithium-Ion Battery, Sodium-Ion Battery, or Redox-Flow Battery: A Comprehensive Comparison in Renewable Energy Systems. *J. Power Sources.* **2023**, *580*, 233426.
- (664) Oh, I.; Pence, M. A.; Lukhanin, N. G.; Rodríguez, O.; Schroeder, C. M.; Rodríguez-López, J. The Electrolab: An Open-Source, Modular Platform for Automated Characterization of Redox-Active Electrolytes. *Device* **2023**, *1*, 100103.
- (665) Lombardo, T.; Duquesnoy, M.; El-Bouysidy, H.; Åren, F.; Gallo-Bueno, A.; Jørgensen, P. B.; Bhowmik, A.; Demortière, A.; Ayerbe, E.; Alcaide, F.; et al. Artificial Intelligence Applied to Battery Research: Hype or Reality? *Chem. Rev.* **2022**, *122*, 10899–10969.
- (666) Sharma, V.; Giannoni, M.; Zubarev, D.; Tek, A.; Nuguyen, K.; Sundberg, L.; Congiu, D.; La, Y.-H. Formulation Graphs for Mapping Structure-Composition of Battery Electrolytes to Device Performance. *J. Chem. Inf. Model.* **2023**, *63*, 6998–7010.
- (667) Kim, S. C.; Oyakhire, S. T.; Athanatis, C.; Wang, J.; Zhang, Z.; Zhang, W.; Boyle, D. T.; Kim, M. S.; Yu, Z.; Gao, X.; et al. Data-Driven Electrolyte Design for Lithium Metal Anodes. *Proc. Natl. Acad. Sci.* **2023**, *120*, No. e2214357120.
- (668) Severson, K. A.; Attia, P. M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M. H.; Aykol, M.; Herring, P. K.; Fragedakis, D.; et al. Data-Driven Prediction of Battery Cycle Life before Capacity Degradation. *Nat. Energy.* **2019**, *4*, 383–391.
- (669) McCalla, E. Semiautomated Experiments to Accelerate the Design of Advanced Battery Materials: Combining Speed, Low Cost, and Adaptability. *ACS Eng. Au.* **2023**, *3*, 391.
- (670) Dave, A.; Mitchell, J.; Kandasamy, K.; Wang, H.; Burke, S.; Paria, B.; Póczos, B.; Whitacre, J.; Viswanathan, V. Autonomous Discovery of Battery Electrolytes with Robotic Experimentation and Machine Learning. *Cell Rep. Phys. Sci.* **2020**, *1*, 100264.
- (671) Dave, A.; Mitchell, J.; Burke, S.; Lin, H.; Whitacre, J.; Viswanathan, V. Autonomous Optimization of Non-Aqueous Li-Ion Battery Electrolytes via Robotic Experimentation and Machine Learning Coupling. *Nat. Commun.* **2022**, *13*, 5454.
- (672) Svensson, P. H.; Yushmanov, P.; Tot, A.; Kloof, L.; Berg, E.; Edström, K. Robotised Screening and Characterisation for Accelerated Discovery of Novel Lithium-Ion Battery Electrolytes: Building a Platform and Proof of Principle Studies. *Chem. Eng. J.* **2023**, *455*, 140955.
- (673) Rahamanian, F.; Vogler, M.; Wölke, C.; Yan, P.; Winter, M.; Cekic-Laskovic, I.; Stein, H. S. One-Shot Active Learning for Globally Optimal Battery Electrolyte Conductivity\*\*. *Batter. Supercaps.* **2022**, *5*, No. e202200228.
- (674) Ling, C. A Review of the Recent Progress in Battery Informatics. *Npj Comput. Mater.* **2022**, *8*, 1–22.
- (675) Benayad, A.; Diddens, D.; Heuer, A.; Krishnamoorthy, A. N.; Maiti, M.; Cras, F. L.; Legallais, M.; Rahamanian, F.; Shin, Y.; Stein, H.; et al. High-Throughput Experimentation and Computational Freeway Lanes for Accelerated Battery Electrolyte and Interface Development Research. *Adv. Energy Mater.* **2022**, *12*, 2102678.
- (676) Zhao, W.; Yi, J.; He, P.; Zhou, H. Solid-State Electrolytes for Lithium-Ion Batteries: Fundamentals, Challenges and Perspectives. *Electrochem. Energy Rev.* **2019**, *2*, 574–605.
- (677) Zhang, Y.; He, X.; Chen, Z.; Bai, Q.; Nolan, A. M.; Roberts, C. A.; Banerjee, D.; Matsunaga, T.; Mo, Y.; Ling, C. Unsupervised Discovery of Solid-State Lithium Ion Conductors. *Nat. Commun.* **2019**, *10*, 5260.
- (678) He, B.; Chi, S.; Ye, A.; Mi, P.; Zhang, L.; Pu, B.; Zou, Z.; Ran, Y.; Zhao, Q.; Wang, D.; et al. High-Throughput Screening Platform for Solid Electrolytes Combining Hierarchical Ion-Transport Prediction Algorithms. *Sci. Data.* **2020**, *7*, 151.
- (679) Laskowski, F. A. L.; McHaffie, D. B.; See, K. A. Identification of Potential Solid-State Li-Ion Conductors with Semi-Supervised Learning. *ChemRxiv2022*. DOI: [10.26434/chemrxiv-2022-2m3qb](https://doi.org/10.26434/chemrxiv-2022-2m3qb) (accessed November 27, 2023)
- (680) Shimizu, R.; Kobayashi, S.; Watanabe, Y.; Ando, Y.; Hitosugi, T. Autonomous Materials Synthesis by Machine Learning and Robotics. *APL Mater.* **2020**, *8*, 111110.
- (681) Ceder, G. Opportunities and Challenges for First-Principles Materials Design and Applications to Li Battery Materials. *MRS Bull.* **2010**, *35*, 693–701.
- (682) Fleischauer, M. D.; Topple, J. M.; Dahn, J. R. Combinatorial Investigations of Si-M (M = Cr+Ni, Fe, Mn) Thin Film Negative Electrode Materials. *Electrochem. Solid-State Lett.* **2005**, *8*, A137.
- (683) Fleischauer, M. D.; Hatchard, T. D.; Bonakdarpour, A.; Dahn, J. R. Combinatorial Investigations of Advanced Li-Ion Rechargeable Battery Electrode Materials. *Meas. Sci. Technol.* **2005**, *16*, 212.
- (684) Green, M. L.; Takeuchi, I.; Hattrick-Simpers, J. R. Applications of High Throughput (Combinatorial) Methodologies to Electronic, Magnetic, Optical, and Energy-Related Materials. *J. Appl. Phys.* **2013**, *113*, 231101.
- (685) Lyu, Y.; Liu, Y.; Cheng, T.; Guo, B. High-Throughput Characterization Methods for Lithium Batteries. *J. Materomics.* **2017**, *3*, 221–229.
- (686) Szymanski, N. J.; Zeng, Y.; Huo, H.; Bartel, C. J.; Kim, H.; Ceder, G. Toward Autonomous Design and Synthesis of Novel Inorganic Materials. *Mater. Horiz.* **2021**, *8*, 2169–2198.
- (687) Kwak, W.-J.; Rosy; Sharon, D.; Xia, C.; Kim, H.; Johnson, L. R.; Bruce, P. G.; Nazar, L. F.; Sun, Y.-K.; Frimer, A. A.; et al. Lithium-Oxygen Batteries and Related Systems: Potential, Status, and Future. *Chem. Rev.* **2020**, *120*, 6626–6683.

- (688) Zhao, M.; Li, B.-Q.; Zhang, X.-Q.; Huang, J.-Q.; Zhang, Q. A Perspective toward Practical Lithium-Sulfur Batteries. *ACS Cent. Sci.* **2020**, *6*, 1095–1104.
- (689) Matsuda, S.; Lambert, G.; Sodeyama, K. Data-Driven Automated Robotic Experiments Accelerate Discovery of Multi-Component Electrolyte for Rechargeable Li-O<sub>2</sub> Batteries. *Cell Rep. Phys. Sci.* **2022**, *3*, 100832.
- (690) Office of Electricity. *Accelerating Pathways towards the Long-Duration Storage Shot*. Energy.gov. <https://www.energy.gov/oe/storage-innovations-2030> (accessed 2023-12-04).
- (691) Gao, P.; Andersen, A.; Sepulveda, J.; Panapitiya, G. U.; Hollas, A.; Saldanha, E. G.; Murugesan, V.; Wang, W. SOMAS: A Platform for Data-Driven Material Discovery in Redox Flow Battery Development. *Sci. Data.* **2022**, *9*, 740.
- (692) Dean, W.; Muñoz, M.; Noh, J.; Liang, Y.; Wang, W.; Gurkan, B. Tuning and High Throughput Experimental Screening of Eutectic Electrolytes with Co-Solvents for Redox Flow Batteries. *Electrochimica Acta* **2024**, *474*, 143517.
- (693) Rodriguez, J.; Politi, M.; Adler, S.; Beck, D.; Pozzo, L. High-Throughput and Data Driven Strategies for the Design of Deep-Eutectic Solvent Electrolytes. *Mol. Syst. Des. Eng.* **2022**, *7*, 933–949.
- (694) Huskinson, B.; Marshak, M. P.; Suh, C.; Er, S.; Gerhardt, M. R.; Galvin, C. J.; Chen, X.; Aspuru-Guzik, A.; Gordon, R. G.; Aziz, M. J. A Metal-Free Organic-Inorganic Aqueous Flow Battery. *Nature* **2014**, *505*, 195–198.
- (695) Janoschka, T.; Martin, N.; Martin, U.; Fribe, C.; Morgenstern, S.; Hiller, H.; Hager, M. D.; Schubert, U. S. An Aqueous, Polymer-Based Redox-Flow Battery Using Non-Corrosive, Safe, and Low-Cost Materials. *Nature* **2015**, *527*, 78–81.
- (696) Robb, B. H.; Farrell, J. M.; Marshak, M. P. Chelated Chromium Electrolyte Enabling High-Voltage Aqueous Flow Batteries. *Joule* **2019**, *3*, 2503–2512.
- (697) Er, S.; Suh, C.; Marshak, M. P.; Aspuru-Guzik, A. Computational Design of Molecules for an All-Quinone Redox Flow Battery. *Chem. Sci.* **2015**, *6*, 885–893.
- (698) Cheng, L.; Assary, R. S.; Qu, X.; Jain, A.; Ong, S. P.; Rajput, N. N.; Persson, K.; Curtiss, L. A. Accelerating Electrolyte Discovery for Energy Storage with High-Throughput Screening. *J. Phys. Chem. Lett.* **2015**, *6*, 283–291.
- (699) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018**, *361*, 360–365.
- (700) Jinich, A.; Sanchez-Lengeling, B.; Ren, H.; Harman, R.; Aspuru-Guzik, A. A Mixed Quantum Chemistry/Machine Learning Approach for the Fast and Accurate Prediction of Biochemical Redox Potentials and Its Large-Scale Application to 315 000 Redox Reactions. *ACS Cent. Sci.* **2019**, *5*, 1199–1210.
- (701) Cao, Y.; Ser, C. T.; Skreta, M.; Jorner, K.; Kusanda, N.; Aspuru-Guzik, A. Reinforcement Learning Supercharges Redox Flow Batteries. *Nat. Mach. Intell.* **2022**, *4*, 667–668.
- (702) Sowndarya, S.; Law, J. N.; Tripp, C. E.; Duplyakin, D.; Skordilis, E.; Biagioli, D.; Paton, R. S.; St John, P. C. Multi-Objective Goal-Directed Optimization of de Novo Stable Organic Radicals for Aqueous Redox Flow Batteries. *Nat. Mach. Intell.* **2022**, *4*, 720–730.
- (703) Lv, X.-L.; Sullivan, P. T.; Li, W.; Fu, H.-C.; Jacobs, R.; Chen, C.-J.; Morgan, D.; Jin, S.; Feng, D. Modular Dimerization of Organic Radicals for Stable and Dense Flow Battery Catholyte. *Nat. Energy* **2023**, *8*, 1109–1118.
- (704) Liang, Y.; Job, H.; Feng, R.; Parks, F.; Hollas, A.; Zhang, X.; Bowden, M.; Noh, J.; Murugesan, V.; Wang, W. High-Throughput Solubility Determination for Data-Driven Materials Design and Discovery in Redox Flow Battery Research. *Cell Rep. Phys. Sci.* **2023**, *4*, 101633.
- (705) Noh, J.; Doan, H. A.; Job, H.; Robertson, L. A.; Zhang, L.; Assary, R. S.; Mueller, K.; Murugesan, V.; Liang, Y. An Integrated High-Throughput Robotic Platform and Active Learning Approach for Accelerated Discovery of Optimal Electrolyte Formulations. *Nat. Commun.* **2024**, *15*, 2757.
- (706) Hydrogen Storage. Energy.gov. <https://www.energy.gov/eere/fuelcells/hydrogen-storage> (accessed 2023-11-27).
- (707) Yue, M.; Lambert, H.; Pahon, E.; Roche, R.; Jemei, S.; Hissel, D. Hydrogen Energy Systems: A Critical Review of Technologies, Applications, Trends and Challenges. *Renew. Sustain. Energy Rev.* **2021**, *146*, 111180.
- (708) Tountas, A. A.; Ozin, G. A.; Sain, M. M. Solar Methanol Energy Storage. *Nat. Catal.* **2021**, *4*, 934–942.
- (709) Wasmus, S.; Küver, A. Methanol Oxidation and Direct Methanol Fuel Cells: A Selective review1In Honour of Professor W. Vielstich on the Occasion of His 75th Birthday and in Appreciation of His Contributions to Electrochemistry as Well as Fuel Cell Development. *J. Electroanal. Chem.* **1999**, *461*, 14–31.
- (710) Fatehi, E.; Thadani, M.; Birsan, G.; Black, R. W. A Critical Evaluation of a Self-Driving Laboratory for the Optimization of Electrodeposited Earth-Abundant Mixed-Metal Oxide Catalysts for the Oxygen Evolution Reaction (OER). *arXiv* 2023 DOI: [10.48550/arXiv.2305.12541](https://doi.org/10.48550/arXiv.2305.12541)
- (711) Evans, A.; Bieberle-Hütter, A.; Rupp, J. L. M.; Gauckler, L. J. Review on Microfabricated Micro-Solid Oxide Fuel Cell Membranes. *J. Power Sources* **2009**, *194*, 119–129.
- (712) Zhang, B.; Merker, L.; Sanin, A.; Stein, H. S. Robotic Cell Assembly to Accelerate Battery Research. *Digit. Discov.* **2022**, *1*, 755–762.
- (713) Dai, F.; Cai, M. Best Practices in Lithium Battery Cell Preparation and Evaluation. *Commun. Mater.* **2022**, *3*, 1–6.
- (714) Yik, J. T.; Zhang, L.; Sjölund, J.; Hou, X.; Svensson, P. H.; Edström, K.; Berg, E. J. Automated Electrolyte Formulation and Coin Cell Assembly for High-Throughput Lithium-Ion Battery Research. *Digit. Discov.* **2023**, *2*, 799–808.
- (715) Li, T.; Lees, E. W.; Goldman, M.; Salvatore, D. A.; Weekes, D. M.; Berlinguet, C. P. Electrolytic Conversion of Bicarbonate into CO in a Flow Cell. *Joule* **2019**, *3*, 1487–1497.
- (716) Fuel cell assembly. System Engineering. <https://www.thyssenkrupp-automation-engineering.com/en/automotive-industry/electric-motor-assembly/fuel-cell> (accessed 2023-11-14).
- (717) Find products of ruhlamat GmbH | Supplier | HYFINDR. <https://hyfindr.com/en/store/ruhlamat-GmbH>.
- (718) Li, M.; Odom, S. A.; Pancoast, A. R.; Robertson, L. A.; Vaid, T. P.; Agarwal, G.; Doan, H. A.; Wang, Y.; Suduwella, T. M.; Bheemireddy, S. R.; et al. Experimental Protocols for Studying Organic Non-Aqueous Redox Flow Batteries. *ACS Energy Lett.* **2021**, *6*, 3932–3943.
- (719) Fell, E. M.; Aziz, M. J. High-Throughput Electrochemical Characterization of Aqueous Organic Redox Flow Battery Active Material. *J. Electrochem. Soc.* **2023**, *170*, 100507.
- (720) Yao, Z.; Lum, Y.; Johnston, A.; Mejia-Mendoza, L. M.; Zhou, X.; Wen, Y.; Aspuru-Guzik, A.; Sargent, E. H.; Seh, Z. W. Machine Learning for a Sustainable Energy Future. *Nat. Rev. Mater.* **2023**, *8*, 202–215.
- (721) Mistry, A.; Franco, A. A.; Cooper, S. J.; Roberts, S. A.; Viswanathan, V. How Machine Learning Will Revolutionize Electrochemical Sciences. *ACS Energy Lett.* **2021**, *6*, 1422–1431.
- (722) Bhowmik, A.; Berecibar, M.; Casas-Cabanas, M.; Csanyi, G.; Dominko, R.; Hermansson, K.; Palacin, M. R.; Stein, H. S.; Vegge, T. Implications of the BATTERY 2030+ AI-Assisted Toolkit on Future Low-TRL Battery Discoveries and Chemistries. *Adv. Energy Mater.* **2022**, *12*, 2102698.
- (723) Improving energy storage using autonomous experiments | Argonne National Laboratory. <https://www.anl.gov/autonomous-discovery/improving-energy-storage-using-autonomous-experiments> (accessed 2023-11-14).
- (724) Biron, L. Meet the Autonomous Lab of the Future. Berkeley Lab News Center. <https://newscenter.lbl.gov/2023/04/17/meet-the-autonomous-lab-of-the-future/> (accessed 2023-11-14).
- (725) Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Next-Generation Experimentation with Self-Driving Laboratories. *Trends Chem.* **2019**, *1*, 282–291.

- (726) Krenn, M.; Pollice, R.; Guo, S. Y.; Aldeghi, M.; Cervera-Lierta, A.; Friederich, P.; dos Passos Gomes, G.; Häse, F.; Jinich, A.; Nigam, A.; et al. On Scientific Understanding with Artificial Intelligence. *Nat. Rev. Phys.* **2022**, *4*, 761–769.
- (727) Messeri, L.; Crockett, M. J. Artificial Intelligence and Illusions of Understanding in Scientific Research. *Nature*. **2024**, *627*, 49–58.
- (728) Kitano, H. Nobel Turing Challenge: Creating the Engine for Scientific Discovery. *Npj Syst. Biol. Appl.* **2021**, *7*, 1–12.
- (729) Gu, X.; Krenn, M. Generation and Human-Expert Evaluation of Interesting Research Ideas Using Knowledge Graphs and Large Language Models. arXiv:2405.17044 [cs]. May 27, arXiv2024. <http://arxiv.org/abs/2405.17044> (accessed 2024-06-10). DOI: 10.48550/arXiv.2405.17044
- (730) Hysmith, H.; Foadian, E.; Padhy, S. P.; Kalinin, S. V.; Moore, R. G.; Ovchinnikova, O. S.; Ahmadi, M. The Future of Self-Driving Laboratories: From Human in the Loop Interactive AI to Gamification. *ChemRxiv*2024. DOI: 10.26434/chemrxiv-2024-3xq9z (accessed June 10, 2024).