



University  
of Glasgow

Wilbraham, L. , Mehr, S. H. M. and Cronin, L. (2021) Digitizing chemistry using the chemical processing unit: from synthesis to discovery. *Accounts of Chemical Research*, 54(2), pp. 253-262.

(doi: [10.1021/acs.accounts.0c00674](https://doi.org/10.1021/acs.accounts.0c00674))

This is the Author Accepted Manuscript.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<https://eprints.gla.ac.uk/232091/>

Deposited on: 29 January 2021

# Digitizing Chemistry Using the Chemical Processing Unit: From Synthesis to Discovery

Liam Wilbraham, S. Hessam M. Mehr, and Leroy Cronin

School of Chemistry, The University of Glasgow, University Avenue, Glasgow G12 8QQ, United Kingdom

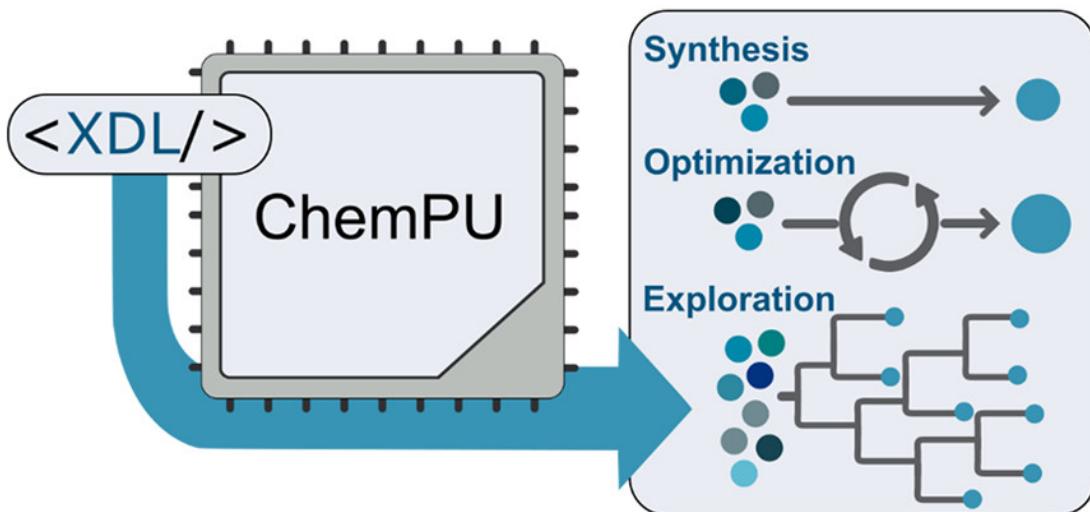
Corresponding Author: [Lee.Cronin@glasgow.ac.uk](mailto:Lee.Cronin@glasgow.ac.uk)

## Abstract

### Conspectus:

The digitization of chemistry is not simply about using machine learning or artificial intelligence systems to process chemical data, or about the development of ever more capable automation hardware; instead, it is the creation of a hard link between an abstracted process ontology of chemistry and bespoke hardware for performing reactions or exploring reactivity. Chemical digitization is therefore about the unambiguous development of an architecture, a chemical state machine, that uses this ontology to connect precise instruction sets to hardware that performs chemical transformations. This approach enables a universal standard for describing chemistry procedures via a chemical programming language and facilitates unambiguous dissemination of these procedures. We predict that this standard will revolutionize the ability of chemists to collaborate, increase reproducibility and safety, as well as optimize for cost and efficiency. Most importantly, the digitization of chemistry will dramatically reduce the labor needed to make new compounds and broaden accessible chemical space. In recent years, the developments of automation in chemistry have gone beyond flow chemistry alone, with many bespoke workflows being developed not only for automating chemical synthesis but also for materials, nanomaterials, and formulation production. Indeed, the leap from fixed-configuration synthesis machines like peptide, nucleic acid, or dedicated cross-coupling engines is important for developing a truly universal approach to “dial-a-molecule”. In this case, a key conceptual leap is the use of a batch system that can encode the chemical reagents, solvent, and products as packets which can be moved around the system, and a graph-based approach for the description of hardware modules that allows the compilation of chemical code that runs on, in principle, any hardware. Further, the integration of sensor systems for monitoring and controlling the state of the chemical synthesis machine, as well as high resolution spectroscopic tools, is vital if these systems are to facilitate closed-loop autonomous experiments. Systems that not only make molecules and materials, but also optimize their function, and use algorithms to assist with the development of new synthetic pathways and process optimization are also possible. Here, we discuss how the digitization of chemistry is happening, building on the plethora of technological developments in hardware and software. Importantly, digital-chemical robot systems need to integrate feedback from simple sensors, e.g., conductivity or temperature, as well as online analytics in order to navigate process space autonomously. This will open the door to accessing known molecules (synthesis), exploring whether known compounds/reactions are possible under new conditions (optimization), and searching chemical space for unknown and unexpected new molecules, reactions, and modes of reactivity (discovery). We will also discuss

the role of chemical knowledge and how this can be used to challenge bias, as well as define and expand synthetically accessible chemical space using programmable robotic chemical state machines.



## Key References

- Steiner, S.; Wolf, J.; Glatzel, S.; Andreou, A.; Granda, J. M.; Keenan, G.; Hinkley, T.; Aragon-Camarasa, G.; Kitson, P. J.; Angelone, D.; Cronin, L. Organic Synthesis in a Modular Robotic System Driven by a Chemical Programming Language. *Science* 2019, 363, eaav2211.(1) The *Chempoter*—our first demonstration of the Chemical Processing Unit (ChemPU) within a general-purpose automated platform for laboratory synthesis.
- 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.(2) Expressing the abstraction of batch synthetic chemistry in a language connected to the literature on one hand and to the specification of hardware on the other.
- 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.(3) Exploration of chemical space made possible by embedding a machine learning algorithm into an automated platform capable of executing a precise process space.
- Kitson, P. J.; Marie, G.; Francoia, J.-P.; Zalessky, S. S.; Sigerson, R. C.; Mathieson, J. S.; Cronin, L. Digitization of Multistep Organic Synthesis in Reactionware for On-Demand Pharmaceuticals. *Science* 2018, 359, 314–319.(4) Modularization of chemical transformations allows a cascade of chemical reactions to be carried out by a combination of 3D-printed hardware cartridges. These “reactionware” modules can be printed in bespoke combinations, enabling construction of chemical hardware from code.

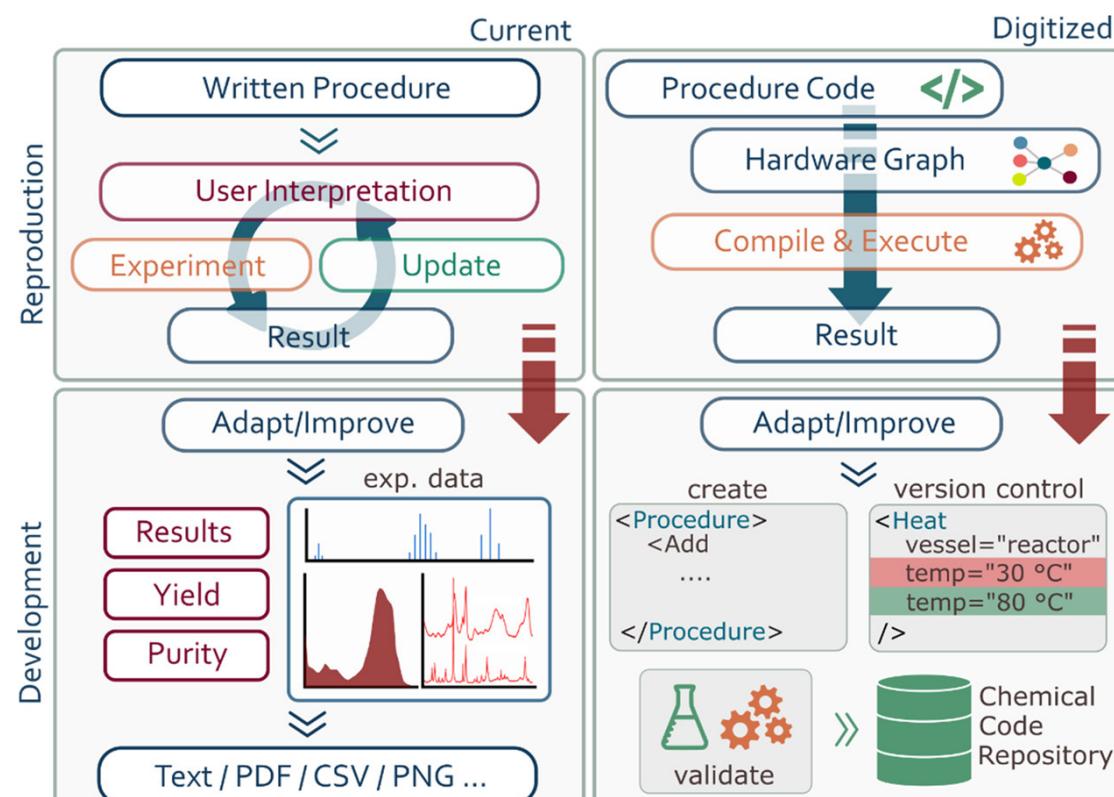
## Introduction

In recent years, the “digitization” of chemistry—that is, the integration of chemistry with techniques developed in computer science and robotics—has gathered pace, promising more reproducible methods and faster routes to discovery. While an emphasis on

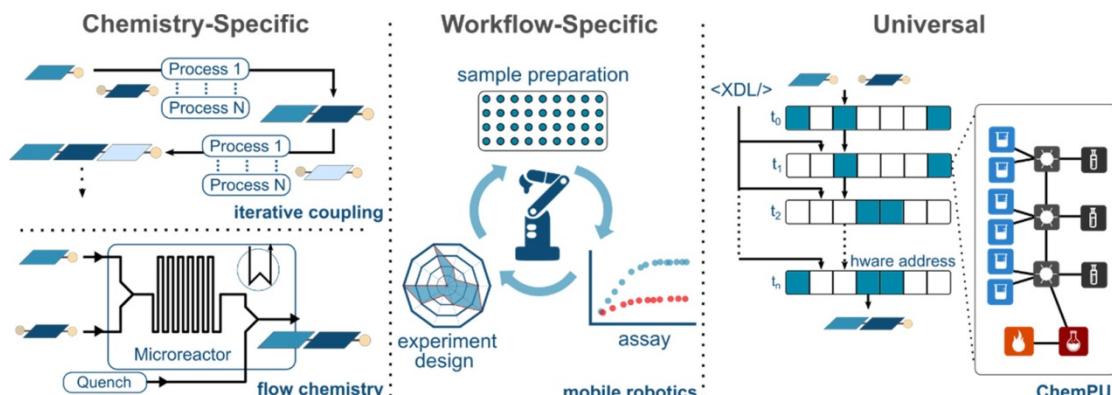
reducing labor-intensive lab work has led to advances in the robotic automation of, for instance, chemical synthesis, these approaches have seldom incorporated the broader needs of the chemistry field: namely, the reproducibility, storage, and dissemination of chemical data. In this Perspective, by making a distinction between the automation and complete digitization of chemistry, we outline what we perceive as the basic, necessary components of digital chemistry, the current state of the art, as well as the opportunities and challenges that these technologies afford.

## Automation of Lab Work

The automation of the many components of lab work, including synthesis, target assays, and formulation, represents a daunting set of milestones for digitization. In each case, automating these workflows involves translating tasks that would ordinarily be carried out by a human chemist into those that are executable by a robotic one, Figure 1. However, this translation is usually nontrivial. The often-unforgiving sensitivity of chemical reactions, external environmental effects, real-time analytical feedback, as well as the traditionally human-oriented nature of integral pieces of equipment represent a series of significant scientific challenges and goals. There have been numerous recent advances toward these that leverage distinct strategies. These strategies can be broadly cast into two groups: those which adapt or select specific chemistry to facilitate automation, and those which develop novel automation solutions that allow access to desired chemistry or processes, Figure 2.



**Figure 1.** Current paradigm (left) vs the digital paradigm (right). To reproduce and adapt procedures from the literature, a chemist currently relies on their interpretation of prose, which may contain ambiguities or lack crucial information. This leads to significant time being invested in refining this interpretation through trial and error. Any new results are disseminated in a similar fashion. In contrast, a fully digitized chemistry workflow allows the direct execution of procedure code with a verified outcome, and new or adapted procedures can easily be submitted to fully searchable, tracked repositories.



**Figure 2. Distinct solutions to chemical automation.** Chemistry-specific chemical automation: focusing on a productive subset of chemistry that lends itself to automation, e.g., iterative Suzuki–Miyaura cross coupling, solid-state peptide synthesis. At the expense of initial optimization of reaction conditions, continuous flow chemistry also provides a means to simplify automation and carries several other advantages. Workflow-specific chemical automation: robots can be designed to mimic the actions of the chemist and deployed within a specific workflow, e.g., the optimization of photocatalytic formulations for hydrogen evolution. Universal chemical automation: by abstracting chemical processes such that they may be described by a chemical programming language and using a hardware-agnostic description of physical resources, one can design a system that, in principle, can execute arbitrary sequences of chemical processes.

In the former case, it has been shown that by selecting a productive subset of chemistry one can gain access to a significant chemical space that can be traversed using a single automated procedure.(5–7) Analogously—that is, in another example of selecting chemistry to facilitate automation—enabling the use of continuous flow processes can simplify automation at the expense of the initial optimization of reaction conditions.(8,9) In the latter case, modern robotics have been applied to explore formulation spaces of photocatalysts(10) and hole transport materials,(11) and we have recently demonstrated the foundations for a universal synthesis platform, with the principal aim of having no restriction on the type of chemistry—and therefore chemical space—that can be accessed.(1,2)

### Iterative Synthesis Using a Single Automated Process

As outlined above, one attractive approach to the automation of chemical synthesis is to select a productive subset of chemistry which lends itself to automation and gives access to a chemical space of particular interest. This concept is exemplified by automated solid-phase peptide synthesis (SPPS),(12) where the versatility and inherent high affinities of peptides for their respective targets have led to tremendous progress for therapeutic applications in the last years.(13) Unconstrained small molecule synthesis—unlike SPPS, where one iterates between identically processed deprotection and coupling steps—involves process steps and conditions that are highly specific for each target and it is challenging, if not impossible, to establish a single, broadly applicable automated process that captures this space.

Burke and co-workers have taken a similar approach to SPPS, establishing an iterative cross-coupling method that takes advantage of a single automated process—in this case, the Suzuki–Miyaura reaction—and reported the synthesis of 14 distinct classes of small molecules.(5–7,14) The ability to apply this method iteratively and facilitate automation using a single set of processes is unlocked by a “catch-and-release” chromatographic purification protocol, which takes advantage of the MIDA group that is ubiquitous in all intermediates in this scheme. It follows that, with many building blocks that are suitable

for this platform already available, it yields greater potential coverage of the chemical space compared to—for instance—iterative peptide synthesis, while still relying only on a single automated process.

## Flow Chemistry

In a similar vein, adapting chemistry to operate in flow can simplify automation often at the expense of requiring initial adaptation/optimization of reaction conditions to work in flow.(15,16) Coley and colleagues demonstrated an integrated system of computer-aided synthesis planning and subsequent execution on a robotic flow system.(17) In their approach, a robotic arm is used to configure the synthesis apparatus by assembling the required unit operations and reagent lines as needed into a process stack, which is subsequently connected to form a continuous flow path. Further, the same robotic arm is used to connect this stack—through a robotically manipulated “switchboard”—to reagent feeds. Using this flow chemistry implementation, they demonstrated the computer-assisted planning and synthesis of 15 medicinally relevant small molecules.

Chatterjee and colleagues also demonstrated an automated approach described as “radial synthesis”,(18) which aims to eliminate the need for synthesis-specific platform reconfiguration. This system comprises individually accessible compartments for performing reactions. These compartments are arranged around a central hub that serves to coordinate reagent delivery, product sampling, and chemical analysis, as well as the temporary storage of intermediate compounds. This platform was used to demonstrate exploration of several strategies for the multistep synthesis of the anticonvulsant drug rufinamide and the synthesis of 18 compounds of two derivative libraries, all without reconfiguration of the platform.

## Mobile Robotics

In contrast to the selection of chemistry to facilitate automation, novel automation solutions have also been developed to allow access to specific processes or workflows of interest. Recently, Burger and colleagues demonstrated a mobile robotic chemist which was used to search for improved photocatalysts for the production of hydrogen from water.(10) Driven by a Bayesian search algorithm, the mobile robot (operating continuously for over 8 days) was able to navigate a process space of diverse instrument and measurement types, performing 688 experiments within a ten-variable formulation space. A similar approach using an (immobile) robotic arm has been applied to the discovery of thin-film materials.(11) These approaches are distinct in that they aim to automate the researcher rather than the instruments, which lends this system to the potential incorporation of further instrument types and experimental setups.

## Printed Reactionware

The development of 3D printed reactionware(4,19) marks a departure from each of the automated chemistry categories outlined above—that is, the selection of chemistry to facilitate automation or the development of specialized automation solutions to facilitate particular experimental workflows. Here, the aim was to establish a basic set of hardware modules that, when implemented in sequence, have the potential to access any chemical synthesis that can be performed in batch. This involved the translation of traditional bench-scale synthesis into (i) a decomposed set of printable modules and (ii) code for their manufacture. The code is then used to 3D print a bespoke monolithic device capable

of carrying out the entire synthetic route. Up to now, this approach has been deployed in the synthesis of several pharmaceutically relevant molecules, such as baclofen, lamotrigine, and zolimidine.

The reactionware approach is notable in its relationship with the spectrum of automation techniques discussed before. While the finished monolith is an extreme example of specialization, the ability to 3D print different configurations on-demand by combining modular components elevates it to a category of its own, where it is able to make a direct connection from bespoke chemistry all the way down to individual modules. Indeed, as alluded to above, this modular architecture led us to consider chemistry in terms of modular hardware components and the set of facts and relations—that is, the ontology—surrounding these hardware modules, the operation they can perform (alone or in combination), and the processes encountered in chemistry.

Figure 2 provides a summary and comparison of the solutions described in this section. Regardless of the approach, the benefits of these solutions are apparent, the most obvious of which is that chemists can spend less time doing repetitive and time-consuming tasks, freeing them up to consider new avenues of creative research, mentor others, or develop greater understanding of their problem. The development process of new molecules becomes streamlined and more cost-effective, removing emphasis from laborious synthetic tasks and placing it on to testing for efficacy (e.g., as a drug, in a device), as well as providing a route to on-demand synthesis of prohibitively expensive or unstable reagents. There are also potential safety benefits, as a robotic chemist is naturally able to work straightforwardly with substances that are toxic to humans. Lastly, particularly in the context of formulation and optimization, the inherent “tenacity” of robot chemists can lead to more optimal solutions than would be found manually, where the cost of the chemist’s time is a significant factor.

## Codified, Machine Readable Chemistry

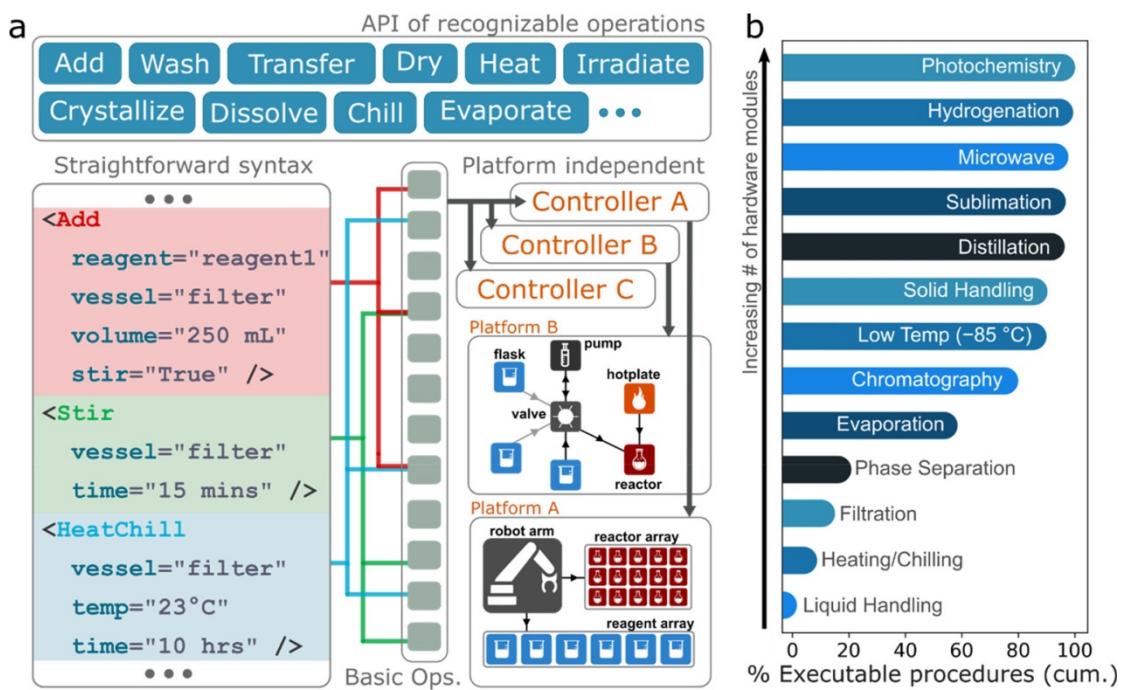
A lot of emphasis has been placed on the development of digital representations of molecules and reactions, for example, through molecular identifiers such as SMILES,(20) InChI strings,(21) and the more recent SELFIES.(22) These representations provide a convenient, searchable, and machine-readable way to describe molecules and have been pivotal in the development of cheminformatics workflows and machine learning in chemistry.

The same level of development, however, has not been reached for procedures, which are still almost entirely written in plain text format. As a result, these often contain ambiguities, cannot be automatically verified, and are not readily searchable without the development of additional technologies such as chemical text mining software.(23–25) To illustrate the problem, consider the software repository GitHub ([github.com](https://github.com)). Here, procedures are described by sets of instructions in the form of computer code, written in a programming language, and these instructions can be carried out by any computer with the necessary hardware and software, with the same results every time. Moreover, these procedures are “versionable”, meaning that they can be updated or superseded by superior workflows, while tracking the changes that were made by each update. In contrast, in the chemistry literature, the plain text must be manually interpreted, can often contain ambiguities or lack crucial detail, and there is no explicit link between versions of the same methodology. This, together with a lack of standardized or well-described “hardware”, makes chemistry procedures more difficult to reproduce, threatens the ability of the chemistry field to leverage its own literature toward the development of data-driven solutions, and carries an associated cost related to time spent rectifying errors or covering old ground.

Naturally, this calls for a “chemical programming language” to replace—or exist alongside—text-based procedures. That is, an abstraction of the chemistry process space that leads to a series of basic commands that can be sequenced to create procedures. These sequences of commands should be exportable in a consistent, universally recognized format and be executable on any robotic platform with access to hardware with equivalent capability. It follows that “chemical code” can in principle benefit from many of the developments that have been made for computer code, namely, the mechanisms for sharing, testing, versioning, and collaborating. Additionally, the publishing of “negative” or suboptimal results becomes trivial, with the ability to share negative results that include detailed process variables, addressing this imbalance currently present in the literature and allowing negative and positive results to be placed—at least digitally—on equal footing. Indeed, significant effort to standardize and balance the chemical reaction literature is already underway.(26)

Such a chemical data repository also trivially unlocks advanced search capability across the literature. For example, current databases such as Reaxys and SciFinder, while they already allow for specific queries, are limited and are susceptible to noisy data sets. With the introduction of procedures described by unambiguous chemical code, it would be possible to perform queries of almost arbitrary granularity (e.g., “show me all procedures that contain this molecule, are performed within this temperature range, and were not stirred”).

Considering the needs of software and hardware together, we can outline what we perceive as the basic requirements for fully digital chemistry. On the software front (Figure 3a), it is essential to be able to write code that deals in process language that is instantly recognizable to chemists (e.g., “Heat”, “Evaporate”, “Crystallize”). To facilitate adoption, the syntax for this language should be as simple as possible, and its interpreter should ideally be constructed to permit facile transferability to any robotic platform with the appropriate hardware. This way, the same syntax can be adopted by specialist and general solutions alike, paving the way for a universally recognized, machine readable description of experimental procedures in chemistry. On the hardware front, in order to maximize the amount of chemical space we can explore, we do not want to be restricted by the types of chemistry we can perform (whether these are currently known or not). Therefore, the foundations of any effort to address this should be designed with expansion in mind. The benefit of this is demonstrated by Figure 3b. For a number of procedures selected from the literature, we show the relationship between the scope of a robotic platform and the fraction of procedures that can be performed, while at the same time outlining an implicit roadmap for module development. This analysis suggests, in agreement with the conclusions of others,(27) that a modular approach to hardware development—i.e., the development of bespoke pieces of hardware optimized to perform particular tasks—is attractive, as new chemistries can be sequentially accommodated by expanding a repertoire of modules.



**Figure 3.** Requirements placed on software and hardware for the full digitization of chemistry. (a) Basic requirements and scope of a programming language for chemistry. The application programming interface (API) should contain operations that are easily recognized by chemists, and syntax should be as straightforward as possible. Ideally, the language is not tied to a particular robotic platform. Any platform that possesses the minimum hardware required for a given procedure should be able to execute a given procedure script. (b) Hardware requirements of literature procedures. The data set contains every procedure from *Organic Syntheses* volumes 77–95 (published between 2000 in 2018) and required hardware was extracted using *SynthReader*. The cumulative number of executable procedures with the successive addition of each hardware module is shown.(2)

## Impact on the Discovery Process

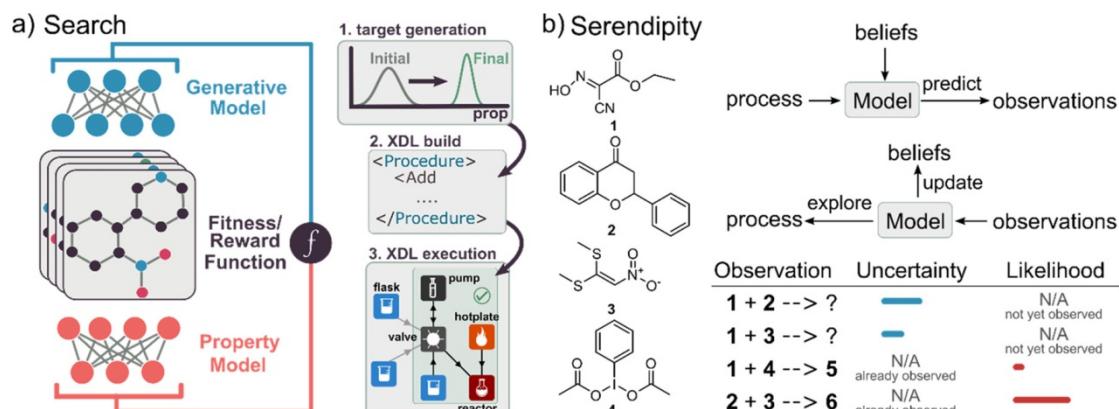
The combined tools of chemical code and automated chemistry can have a profound impact on the discovery process. At this point, it is helpful to distinguish two distinct concepts for which “discovery” is used as an umbrella term. (i) **Search** (Figure 4a) usually refers to the optimization or exploration of chemical space according to a tailored “fitness” function within the domain of a defined parameter space, leading to the discovery of high-performing molecules or materials. Examples of search-based discovery efforts include optimizing processing conditions to maximize material performance,(28) searching defined chemical spaces for molecules with high binding affinity to protein targets(29) or fine-tuned optoelectronic properties,(30) and the exploration of the behavior space of multicomponent droplets.(31) (ii) **Serendipity** (Figure 4b), on the other hand, refers to observations that defy current expectations and understanding, often leading to breakthroughs and revisions in existing models. Given these distinct facets of discovery, the significance of digitization is threefold:

- Precise navigation of chemical and process space:** We can leverage knowledge gained in a robotic context to guide responsive search algorithms in real time. Full digitization—that is, a programming language and a “chemical processor” with which this code can be executed—means we are able to adjust these inputs with arbitrary granularity, allowing degrees of freedom which are otherwise inaccessible by other discovery systems. The manipulation of these degrees of freedom can be trivially orchestrated by some optimization strategy, or algorithm, which seeks to discover new

optima in chemical or process space with the minimum experimental budget. Indeed, there has been increasing movement in this area in recent years, with attention directed toward, among others, Bayesian(32) or noisy optimization methods.(33) Paired with executable chemical code, such orchestration software holds the potential to completely alter how common tasks—for instance, the optimization of reaction conditions—are routinely performed.

**2. Possibility to define more intelligent discovery strategies:** This is especially relevant to the search for genuine novelty. Rather than using a brute-force or grid-based search of process space, exploration can use a curiosity-based approach, challenging and evolving the current state of understanding (encoded in a sufficiently realistic model). Unexpected positive or negative outcomes can then be attributed either to experimental error, which can be verified by repeating the same experiment or attempting related experiments, or to genuine discovery.

**3. Provide a quantitative framework for defining discoveries:** A precise definition of true discovery versus search and optimization has been elusive and, thus far, delegated to chemists' personal interpretation and experience. With machines capable of producing an overwhelming number of potentially novel observations, providing a quantitative definition is especially relevant today. It is important to understand the difference between natural extensions of what is known, i.e., interpolation, versus a truly novel observation that can serve as a starting point for future discoveries.



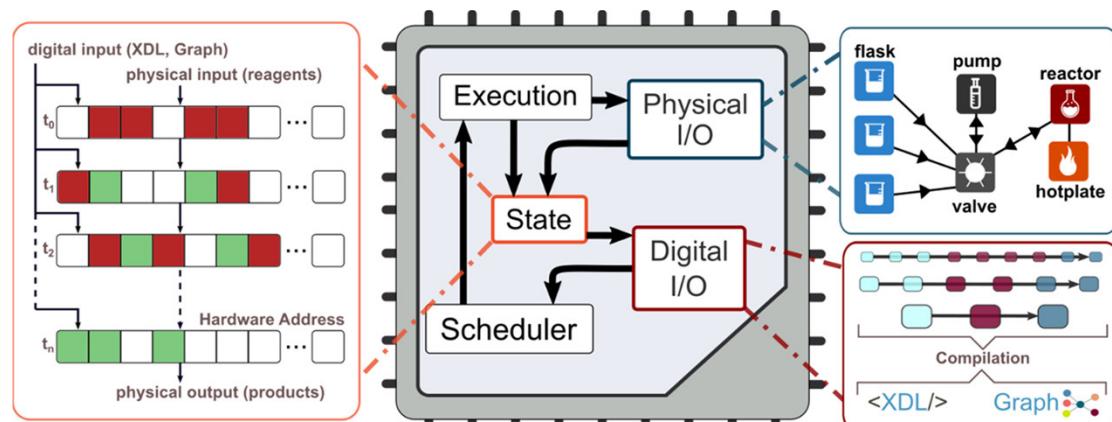
**Figure 4.** Two distinct concepts under the umbrella of discovery. (a) Search: Navigation of chemical space according to a fitness function through a molecular generation-evaluation cycle. When optimal points in chemical space (i.e., molecules) have been identified, these can be exported for automatic synthesis on a robotic platform. (b) Serendipity: Discovery is defined in the context of current chemical knowledge, that is, a probabilistic model connecting experiments and their outcomes. Experiments are a form of making observations in process space, and discoveries are those observations that defy chemical knowledge at the time.

Digitization clarifies and makes explicit the trail of experiments that led to discovery. Therefore, with all procedures described in a machine-readable way, it is much easier to make discoveries reproducible. When experiments are captured as points in process space, the principal challenge is to quantify the sensitivity of the outcome to variations in process parameters, so it can be generalized to adjacent points. In practice, no two experimental setups are exactly identical: there will always be differences in device manufacturers/models, reagent purity, ambient conditions, and scale. In a specific implementation, one can imagine that careful observation of points in process space and the resulting outcomes, e.g., yield, can be interpreted within a presupposed Bayesian model to estimate the unknown sensitivities of the chemical process to conditions. Being able to automatically capture the outcome in these different scenarios unlocks the ability

to quantify reproducibility, discover optimal conditions for a transformation, or relax unnecessary constraints.

## Our Paradigm: The Chemical Processing Unit

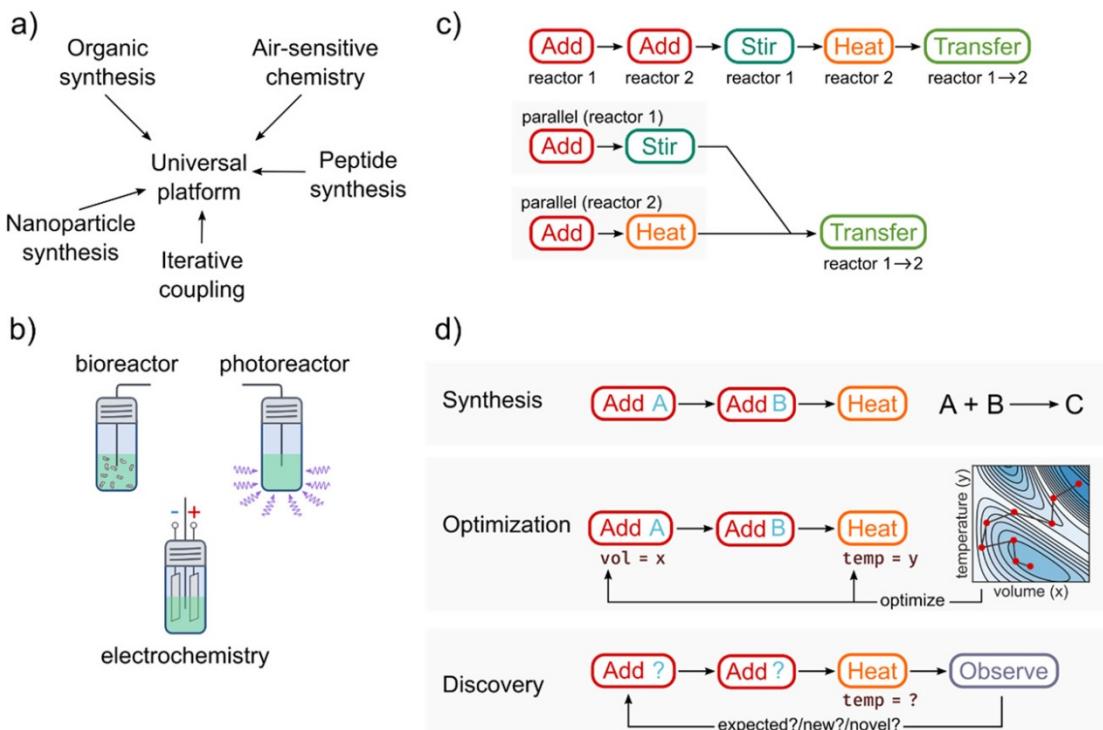
At the heart of our solution to what we have outlined up to now is the implementation of a chemical analogue to the architecture of a computer, namely, a “chemical processing unit” (ChemPU, Figure 5), a formal architecture within which chemistry can be executed or simulated deterministically. This is analogous to the combination of hardware logic and firmware in digital computers where programmed instructions are also processed in a deterministic manner. To realize such a chemical processor, an integrated approach to the development of software and hardware for chemistry is essential. In doing this, we have begun to address the broad requirements of various digital chemistry workflows. On the hardware front, from the outset our platform has been designed to map the (batch) processes typically carried out by the bench chemist to modular, reconfigurable hardware that can be conceptually described by a hardware “graph”. This (directed) graph is simply a machine-readable description of the available hardware, and how it is connected. It is composed of nodes (pieces of hardware with a known, exposed functionality), and edges (electronic, fluid, or logical connections between pieces of hardware). As such, simple expansion of the types of graph “nodes” that are available allows us to trivially accommodate new chemistries, several of which are under active development. For example, one aim for the ChemPU is to demonstrate the “convergence” of several conceptually and operationally different chemistries within our platform, and this work is ongoing and will be published soon.(34)



**Figure 5. Chemical processing unit (ChemPU) concept.** The transformation of a chemical “state” from inputs (reagents) to outputs (products) is dictated by a series of hardware manipulations at different time steps. These manipulations are mediated by a digital input (in this case, XDL), where low-level instructions are obtained via a “compilation” stage, which considers the high-level instructions and available hardware, represented by a hardware graph.

The reconfigurable nature of this graph allows the platform to be geared toward significantly different and unrelated tasks without further hardware development, essentially providing further “task-based convergence”, which goes beyond the chemical convergence outlined above. Specifically, a given “basis set” of hardware nodes can be configured to perform, for instance, synthesis (reproduction of literature procedures), optimization (searching for optimal reaction conditions), and discovery tasks (searching for new reactivity). To date, we have successfully demonstrated the synthesis(1,2) and discovery(3,35) configurations and are actively working on optimization. As we have alluded to above, just as a computer’s CPU is instructed by computer code, a ChemPU

must be given instructions that correspond to a chemical procedure. Therefore, we have developed such a programming language for chemistry (XDL) and an associated interpreter which can carry out these instructions on our—and, in principle, any—robotic platform.(2) Importantly, this interpreter is designed to be (i) “platform independent”, meaning that it is largely trivial to accommodate not only new hardware but entirely different robotic setups and philosophies and (ii) “task independent”, meaning that the same application programming interface can be repurposed for synthesis, optimization, and discovery tasks. With the design of the ChemPU architecture, we seek to simultaneously address several requirements and desirable features of a universal digital chemistry platform (Figure 6), which are discussed below in more detail.



**Figure 6. Goals and target features of a universal digital chemistry platform.** (a) The platform must be able to accommodate different types of chemistry. (b) New hardware modules (graph nodes) can be added to expand the range of accessible chemistry within the existing platform. (c) Uses scheduling to enable dynamic allocation of hardware resources. (d) Unifies synthesis, optimization, and discovery within the same platform.

## Perform almost All Types of Chemical Reactions

The first programmable digital machines to enter the chemistry lab were analytical instruments, followed by programmable labware intended for synthetic chemistry, such as rotary evaporators and automated column chromatography systems. These single-task robots have since evolved to automate entire application-specific workflows, such as solid-phase peptide synthesis. In spite of their convenience and efficiency, these systems have significant gaps in their ontology. Because of their fixed function, they are not designed to execute a universal chemical code or even a subset thereof. Rather than a proliferation of automation islands with no interoperability, we propose a common standard architecture for all automated platforms, controlled using a universal high-level chemical code. Within this architecture, each system might still only be capable of executing a small subset of this code (determined by available hardware modules and their relationships) and using a bespoke implementation. Still, all the benefits above hold. To validate our implementation of this goal, we have used XDL to program our

Chemputer system for such varied tasks as general organic synthesis, air-sensitive transition metal chemistry, solid-phase peptide synthesis, iterative coupling of MIDA boronates, and the synthesis of nanoparticles.

## Extensibility to New Hardware and Chemistries

Early implementations of digital chemistry platforms often rely on a core set of basic operations, such as addition of reagents, stirring/heating, phase separation, and evaporation. These processes are important as they underpin the bulk of the existing literature procedures, and their robust implementation will allow existing chemistry to be replicated in automated platforms in a reliable manner. In order to serve as a viable research platform, however, digital chemistry must be able to integrate emerging or less common process types and the requisite hardware without requiring fundamental changes to the paradigm. New process types of interest include organic electrosynthesis, photochemistry, as well as flow and bioreactors, Figure 6b.

To this end, we have devised a flexible abstraction of a hardware device module, that is, a physical device with an associated set of inputs and outputs (digital and chemical), a low-level instruction set, and a specification that allows compilation of XDL to these elementary operations. Continuous addition of new hardware modules has provided us the opportunity to test how effective this approach is. For instance, with the recent introduction of a bespoke digital pneumatic manifold—the automated counterpart to the Schlenk line—a formal specification of the manifold’s capabilities allowed the platform to take advantage of it for evacuation, evaporation, and purging operations in existing procedures.

## Intelligent Resource Allocation

Synthetic chemists in research are limited in the number of reactions they can carry out simultaneously. The difficulty of planning and monitoring parallel chemical processes means that only a small fraction of a lab’s capacity is in use at one time and any naive digitization of chemistry using a linear workflow would bear the same limitation. With the introduction of chemical code, however, the inherent barriers to concurrent execution of chemical code are conceptually removed. All that is needed is a scheduling module which, by considering the resources (graph) available, is able to distribute tasks efficiently across graph nodes. All this can be augmented by a robust array of sensors to ensure safe operation of the system. In this sense, our platform is well suited to perform chemistry procedures concurrently. With available hardware expressed as a graph, the system is at all times aware of the current availability of resources (i.e., graph nodes that are locked/unlocked) as well as the tasks it has yet to perform (expressed through chemical code). Therefore, synthesis tasks can be scheduled and updated dynamically at run time so as to achieve maximum efficiency according to some target (e.g., total execution time, minimum reagent idle time), Figure 6c. This is another example of how the “digitization” of chemistry allows us to leverage well-known concepts from computer science.

The prohibitive cost of setting up a chemistry lab is a factor hampering innovation in the field. The ability to test new ideas is exclusive to large companies or investigators in well-funded academic or industrial research laboratories,(36) and many high-risk/high-reward ideas are never attempted. With this concept in mind, we can consider the possibility that digital infrastructure opens the door to “synthesis servers”, analogous to now-ubiquitous cloud-based computing services, whose large chemical robotic infrastructure can be utilized by researchers and industry on-demand. Within these

services, hardware resources can be optimally utilized, with experiments only charged on the basis of the resources used, such as machine time, vessels, reagents, or energy.

In our approach to enabling dynamic scheduling in an automated platform, unit operations must be annotated with invariants. These invariants express requirements on the state of the system at each point in time within the confines of which operations can be reordered or executed concurrently to improve efficiency. Examples of such invariants are

- The devices involved being available, i.e., not busy.
- The necessary chemical having reached the correct vessel.
- Prerequisite operation or reactions having concluded, e.g., the reaction mixture having reached its target temperature before the addition of a reagent.

Some of these invariants can be inferred from the procedure, e.g., which reactions depend on which reagents or upstream reactions. Others are related to the physical limitations of the system and inferred from the hardware ontology. For instance, a distribution valve can only be in one position at a given time or reactors need to be empty and clean before they can be used to carry out reactions.

With the original chemical code and the above list of requirements in hand, multiple avenues for scheduling are possible. The list of precise requirements can be reformulated as a Boolean satisfiability problem, for which efficient solvers exist.(37) Alternatively, the original steps can be substituted with more efficient equivalents in a stepwise fashion using a knowledge base of rewrite rules expressing the above constraints.(38)

## Synthesis, Optimization, and Discovery in One Platform

XDL's abstraction of unit operations are not limited to synthesis and extend to optimization and discovery with the addition of a feedback loop, Figure 6d. In this context, feedback can be implemented as special operations within the chemical code that influence the course of execution of other "static" steps. This feedback construct was initially devised to express measurement-dependent actions that would otherwise be executed in a "static" fashion, such as phase separation, where the boundary between the two layers of liquid needs to be detected. Expanding the use of these dynamic steps to monitor device operations, reaction progress, and platform safety naturally led to us to the realization that hardware intended for synthesis could be used for discovery without any hardware modifications.

The range of possible feedback sources includes, but is not limited to, those traditionally available in benchtop settings, such as a pH indicator or thin-layer chromatography. The advantages of this expanded range are twofold. First, entire analytical workflows, such as HPLC/MS and flow NMR can be integrated, providing live measurements far richer than previously possible. Second, these richer sources of data and higher rates of data collection are a boon to reproducibility and better diagnostics, preventing situations where only poor or inconclusive measurements are available.

## Conclusions and Outlook

By distinguishing contributions from robotic automation and digital descriptions of chemistry, we have outlined what we believe to be the basic and necessary requirements for the full digitization of chemistry. In response to these requirements, we have developed an integrated hardware and software framework—the ChemPU—which attempts to address these directly, as well as additional desirable features of a digital chemistry platform. By the simultaneous development of a modular hardware philosophy and a programming language for chemistry, we provide a means to encode, execute, and disseminate chemistry procedures. The combination of modular hardware and chemical

code possesses several attractive features, among which are the convergence of diverse chemistries in a single platform, the convergence of tasks (i.e., discovery, optimization, synthesis) within a single platform, the possibility for concurrent task execution, and trivial module expansion. Crucially, because this architecture does not rely on any specific properties of a given chemistry domain, it is in principle able to access any point in (accessible) chemical space through arbitrary known synthetic transformations. We believe that these concepts have far reaching consequences for the field of chemistry, and we are working to help others both use and build on the standard for digitizing chemistry realized by the ChemPU. To do this we are establishing an open standard, developed transparently and collaboratively, for use both in academia and industry, to help us work toward and reap the benefits offered by the universal digitization of chemistry; see [www.chemify.org](http://www.chemify.org).

We gratefully acknowledge financial support from the EPSRC (Grant Nos EP/L023652/1, EP/R020914/1, EP/S030603/1, EP/R01308X/1, EP/S017046/1, EP/S019472/1), the ERC (Project 670467 SMART-POM), the EC (project 766975 MADONNA), The John Templeton Foundation (Projects 60625 and 61184), and DARPA (projects W911NF-18-2-0036, W911NF-17-1-0316, HR001119S0003).

The authors declare the following competing financial interest(s): LC is the founder of DeepMatter and Chemify; both companies are aiming to digitize chemistry, including chemical synthesis and discovery.

## References

- (1) Steiner, S.; Wolf, J.; Glatzel, S.; Andreou, A.; Granda, J. M.; Keenan, G.; Hinkley, T.; Aragon-Camarasa, G.; Kitson, P. J.; Angelone, D.; Cronin, L. Organic Synthesis in a Modular Robotic System Driven by a Chemical Programming Language. *Science* 2019, 363, eaav2211.
- (2) 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.
- (3) 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.
- (4) Kitson, P. J.; Marie, G.; Francoia, J.-P.; Zalesskiy, S. S.; Sigerson, R. C.; Mathieson, J. S.; Cronin, L. Digitization of Multistep Organic Synthesis in Reactionware for On-Demand Pharmaceuticals. *Science* 2018, 359, 314–319.
- (5) 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.
- (6) Trobe, M.; Burke, M. D. The Molecular Industrial Revolution: Automated Synthesis of Small Molecules. *Angew. Chem., Int. Ed.* 2018, 57, 4192–4214.
- (7) Gillis, E. P.; Burke, M. D. Iterative Cross-Coupling with MIDA Boronates: Towards a General Platform for Small Molecule Synthesis. *Aldrichimica Acta* 2009, 42, 17–27.
- (8) Jebrail, M. J.; Bartsch, M. S.; Patel, K. D. Digital Microfluidics: A Versatile Tool for Applications in Chemistry, Biology and Medicine. *Lab Chip* 2012, 12, 2452–2463.

- (9) Köhler, J. M.; Henkel, T.; Grodrian, A.; Kirner, T.; Roth, M.; Martin, K.; Metze, J. Digital Reaction Technology by Micro Segmented Flow Components, Concepts and Applications. *Chem. Eng. J.* 2004, 101, 201–216.
- (10) Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I. A Mobile Robotic Chemist. *Nature* 2020, 583, 237–241.
- (11) MacLeod, B. P.; Parlane, F. G. L.; Morrissey, T. D.; Hase, F.; Roch, L. M.; Dettelbach, K. E.; Moreira, R.; Yunker, L. P. E.; Rooney, M. B.; Deeth, J. R.; Lai, V.; Ng, G. J.; Situ, H.; Zhang, R. H.; Elliott, M. S.; Haley, T. H.; Dvorak, D. J.; Aspuru-Guzik, A.; Hein, J. E.; Berlinguette, C. P. Self-Driving Laboratory for Accelerated Discovery of Thin-Film Materials. *Science Advances* 2020, 6, eaaz8867.
- (12) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* 1963, 85, 2149–2154.
- (13) Vargas-Casanova, Y.; Poveda, J. C. V.; Rivera-Monroy, Z. J.; Garzon, A. C.; Fierro-Medina, R.; Pape, P. L.; García-Castañeda, J. E.; Giraldo, C. M. P. Palindromic Peptide LfcinB (21–25)Pal Exhibited Antifungal Activity against Multidrug-Resistant Candida. *ChemistrySelect* 2020, 5, 7236–7242.
- (14) Woerly, E. M.; Roy, J.; Burke, M. D. Synthesis of Most Polyene Natural Product Motifs Using Just 12 Building Blocks and One Coupling Reaction. *Nat. Chem.* 2014, 6, 484–491.
- (15) Pastre, J. C.; Browne, D. L.; Ley, S. V. Flow Chemistry Syntheses of Natural Products. *Chem. Soc. Rev.* 2013, 42, 8849–8869.
- (16) Plutschack, M. B.; Pieber, B.; Gilmore, K.; Seeberger, P. H. The Hitchhiker’s Guide to Flow Chemistry. *Chem. Rev.* 2017, 117, 11796–11893.
- (17) 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.; Hicklin, R. W.; Plehiers, P. P.; Byington, J.; Piotti, J. S.; Green, W. H.; Hart, A. J.; Jamison, T. F.; Jensen, K. F. A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning. *Science* 2019, 365, eaax1566.
- (18) Chatterjee, S.; Guidi, M.; Seeberger, P. H.; Gilmore, K. Automated Radial Synthesis of Organic Molecules. *Nature* 2020, 579, 379–384.
- (19) Symes, M. D.; Kitson, P. J.; Yan, J.; Richmond, C. J.; Cooper, G. J. T.; Bowman, R. W.; Vilbrandt, T.; Cronin, L. Integrated 3DPrinted Reactionware for Chemical Synthesis and Analysis. *Nat. Chem.* 2012, 4, 349–354.
- (20) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* 1988, 28, 31–36.
- (21) Heller, S. R.; McNaught, A.; Pletnev, I.; Stein, S.; Tchekhovskoi, D. InChI, the IUPAC International Chemical Identifier. *J. Cheminf.* 2015, 7, 23.
- (22) Krenn, M.; Hase, F.; Nigam, A.; Friederich, P.; Aspuru-Guzik, A. Self-Referencing Embedded Strings (SELFIES): A 100% Robust Molecular String Representation. *Machine Learning: Science and Technology* 2020, 1, No. 045024.
- (23) Hawizy, L.; Jessop, D. M.; Adams, N.; Murray-Rust, P. ChemicalTagger: A Tool for Semantic Text-Mining in Chemistry. *J. Cheminf.* 2011, 3, 17.
- (24) Lowe, D. M.; Sayle, R. A. LeadMine: A Grammar and Dictionary Driven Approach to Entity Recognition. *J. Cheminf.* 2015, 7, S5.
- (25) Vaucher, A. C.; Zipoli, F.; Geluykens, J.; Nair, V. H.; Schwaller, P.; Laino, T. Automated Extraction of Chemical Synthesis Actions from Experimental Procedures. *Nat. Commun.* 2020, 11, 3601.

- (26) Open Reaction Database; <https://open-reaction-database.org/> (accessed Nov 27, 2020).
- (27) Flores-Leonar, M. M.; Mejía-Mendoza, L. M.; Aguilar-Granda, A.; Sanchez-Lengeling, B.; Tribukait, H.; Amador-Bedolla, C.; Aspuru-Guzik, A. Materials Acceleration Platforms: On the Way to Autonomous Experimentation. *Current Opinion in Green and Sustainable Chemistry* 2020, 25, 100370.
- (28) Collins, S. P.; Daff, T. D.; Piotrkowski, S. S.; Woo, T. K. Materials Design by Evolutionary Optimization of Functional Groups in Metal-Organic Frameworks. *Science Advances* 2016, 2, No. e1600954.
- (29) Ma, D.-L.; Chan, D. S.-H.; Fu, W.-C.; He, H.-Z.; Yang, H.; Yan, S.-C.; Leung, C.-H. Discovery of a Natural Product-Like c-Myc GQuadruplex DNA Groove-Binder by Molecular Docking. *PLoS One* 2012, 7, e43278.
- (30) 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.; Schrier, J.; Norquist, A. J.; Chan, E. M. Robot-Accelerated Perovskite Investigation and Discovery. *Chem. Mater.* 2020, 32, 5650–5663.
- (31) Grizou, J.; Points, L. J.; Sharma, A.; Cronin, L. A Curious Formulation Robot Enables the Discovery of a Novel Protocell Behavior. *Science Advances* 2020, 6, eaay4237.
- (32) Griffiths, R.-R.; Hernandez-Lobato, J. M. Constrained Bayesian Optimization for Automatic Chemical Design Using Variational Autoencoders. *Chem. Sci.* 2020, 11, 577–586.
- (33) Huyer, W.; Neumaier, A. SNOBFIT – Stable Noisy Optimization by Branch and Fit. *ACM Trans. Math. Softw.* 2008, 35, 1–25.
- (34) Angelone, D.; Hammer, A. J. S.; Rohrbach, S.; Krambeck, S.; Granda, J. M.; Wolf, J.; Zalesskiy, S.; Chisholm, G.; Cronin, L. Convergence of Multiple Synthetic Paradigms in a Universally Programable Chemical Synthesis Machine. *Nat. Chem.* 2021, 13, 63–69.
- (35) Caramelli, D.; Granda, J.; Cambie, D.; Mehr, H.; Henson, A.; Cronin, L. An Artificial Intelligence That Discovers Unpredictable Chemical Reactions. *ChemRxiv* 2020, DOI: 10.26434/chemrxiv.12924968.v1.
- (36) Webb, S. Business Sense: Starting an Academic Lab. *Science* 2009, 1 DOI: 10.1126/science.caredit.a0900088.
- (37) Bogdanova, V. G.; Gorsky, S. A. Scalable Parallel Solver of Boolean Satisfiability Problems. In 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO); 2018; pp 0222–0227.
- (38) Jones, S. P.; Tolmach, A.; Hoare, T. Playing by the Rules: Rewriting as a Practical Optimisation Technique in GHC. In Haskell workshop; 2001; Vol. 1, pp 203–233.