

Chemputation and the Standardization of Chemical Informatics

Alexander J. S. Hammer, Artem I. Leonov, Nicola L. Bell, and Leroy Cronin*

Cite This: *JACS Au* 2021, 1, 1572–1587

Read Online

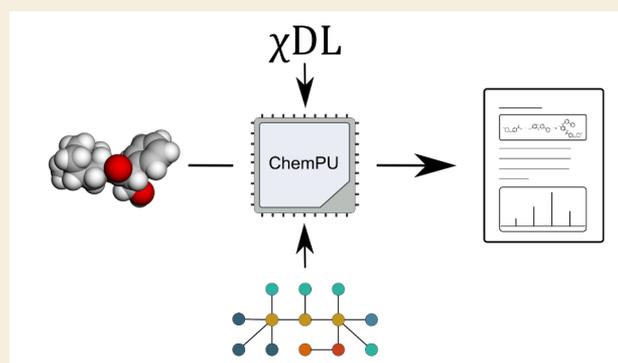
ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: The explosion in the use of machine learning for automated chemical reaction optimization is gathering pace. However, the lack of a standard architecture that connects the concept of chemical transformations universally to software and hardware provides a barrier to using the results of these optimizations and could cause the loss of relevant data and prevent reactions from being reproducible or unexpected findings verifiable or explainable. In this Perspective, we describe how the development of the field of digital chemistry or chemputation, that is the universal code-enabled control of chemical reactions using a standard language and ontology, will remove these barriers allowing users to focus on the chemistry and plug in algorithms according to the problem space to be explored or unit function to be optimized. We describe a standard hardware (the chemical processing programming architecture—the ChemPU) to encompass all chemical synthesis, an approach which unifies all chemistry automation strategies, from solid-phase peptide synthesis, to HTE flow chemistry platforms, while at the same time establishing a publication standard so that researchers can exchange chemical code (χ DL) to ensure reproducibility and interoperability. Not only can a vast range of different chemistries be plugged into the hardware, but the ever-expanding developments in software and algorithms can also be accommodated. These technologies, when combined will allow chemistry, or chemputation, to follow computation—that is the running of code across many different types of capable hardware to get the same result every time with a low error rate.

KEYWORDS: Automated Synthesis, Chemical Informatics, Digital Chemistry, Data Standards, Reaction Optimization



1. INTRODUCTION

The exploration and optimization of chemical reactions can be a laborious and time-consuming endeavor¹ as poor optimization strategies, combined with human intuition laced with biases, means that a large number of reactions must be undertaken to map a given chemical space.² Worse yet, published and previously optimized methods have well-known reproducibility issues.³ A host of different procedures must be carried out in any chemical synthesis leaving significant room for miscommunication, lack of detail, or omission of tacit knowledge.⁴ In recent years, attention has turned to addressing these problems through initiatives which aim to normalize data generation and improve data sharing standards as well as applying novel methods for analyzing available reaction data, including machine learning.^{5–11} These efforts can be seen as first steps toward the full digitization of chemistry (Figure 1).¹²

At first glance, the digitization of chemistry can be seen as the process of converting information, typically gained from physical experiments, into a digital format. However, on a deeper level the digitization of chemistry involves the full control of chemical processes by capturing all relevant input parameters, process operations and output data, and representing these in a machine-readable fashion to allow

consistent reproduction of processes and efficient dissemination of the knowledge obtained.^{13–16} We envision that future digital chemistry laboratories will run automated, multistep reactions on a variety of interoperable hardware. Processes will be monitored with an array of sensors and process analytical tools, enabling rapid sharing of experimental data and advanced algorithmic control, which will be key to fully exploit new understanding and develop new models.^{17,18} In addition, for full exploitation of the potential of digital chemistry, a unified set of data formats, data capture standards and sharing guidelines are required to prevent research groups and fields becoming siloed. Incorporation of digital tools into the workflow of chemistry has already begun to increase efficiency, discovery, and the pace of innovation and it is clear that the volume of scientific data will rapidly proliferate in the coming years.^{9,19–23} In chemical synthesis and reaction

Received: July 6, 2021

Published: August 31, 2021



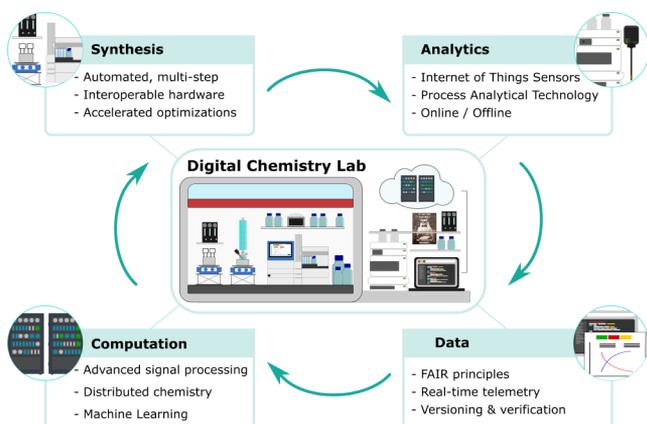


Figure 1. Digitization of a chemistry lab involves all aspects of lab work and data analysis. Automatic synthesis hardware should be able to undertake multiple unit operations or steps and use modular hardware that can communicate with different control systems and hardware systems. Analytics includes not only spectroscopy, spectrometry, and chromatography but also sensors for the continuous monitoring of process variables, such as a temperature, which are vital input/output parameters for optimizations. Data must be collected, stored, and shared using FAIR principles such that both suboptimal and optimal conditions are recorded.

analytics, it will be essential to capture all the key parameters of a preparation or experiment, including the context of the work. This approach will be essential for reproducibility, however, with only a few exceptions, this data has not been recorded reliably to date. Herein we describe how, in order to achieve this vision we need a standard architecture, comprised of a high-level machine and human readable code, to describe and record chemical process steps. It will be vital to connect this via a standard data structure to a wide range of affordable, modular synthetic, and analytical hardware (Figure 2).^{4,14} This architecture must also work in tandem with feedback algorithms to provide the key features required for process optimization and the acquisition of large reaction data sets.²⁴

2. DIGITIZING CHEMICAL REACTIONS

To describe a reaction, chemists typically draw a reaction scheme made up of 2D graphs of the molecular structures

comprising the reagents, any catalysts, and the products, see Figure 3A. Reaction conditions may or may not be indicated by inclusion of parameters in the scheme. While this may indicate key variables, much knowledge is assumed or omitted in these “schemes”. In addition, the actual description of how to perform the reaction in detail is often recorded in continuous prose in a different supplementary document to where the graph is documented, see Figure 3B.

In the context of reaction optimization, experiments are often carried out using a one factor at a time (OFAT) methodology and may be represented in a parameter table, listing the values of continuous (e.g., temperature, time) and categorical (e.g., solvent type) variables as well as the reaction outcome, see Figure 3C. Alternatively, design of experiments screening,² spider plots,²⁵ and traffic light systems²⁶ have all been used to identify the specific parameters, which disproportionately influence reaction yield or any other optimizable metric. However, for all these methods, the decision regarding which parameters are varied or recorded during an optimization campaign is the choice of the chemist and is, therefore, somewhat subjective. In addition, an arbitrary threshold may be set for the optimal output value, and the process operation described by the variables recorded can also be ambiguous.

For cheminformatic applications, the simplified molecular-input line-entry system (SMILES) was developed as a simple single line notation representing the chemical structure of a molecule in ASCII strings.²⁷ SMILES representations have now been widely adopted for digital representations of molecules and can be easily converted into graph-based representations of the structures as to enable easy interpretation by the chemist. In this context reactions are often represented as reaction SMILES where reactants are separated by a period (.) and the reaction arrow is indicated by “ \gg ” with the option to insert “arrow conditions”, see Figure 3D. Notable alternative representations of reactions include RInChI,²⁸ which extend the idea of the IUPAC International Chemical Identifier (InChI)²⁹ toward reactions, reaction-data files (RDfiles), CSRML,³⁰ Reaction-MQL (an extension of the molecular query language (MQL)),³¹ difference vectors based on molecular maps of atom-level properties (MOLMAPS),³² and condensed graph of reaction (CGR),³³ which were

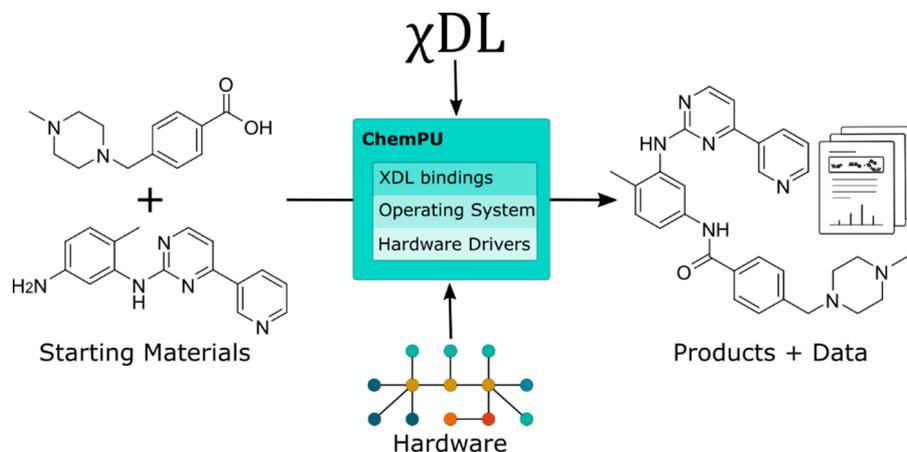


Figure 2. Chemputation: The process of converting a chemical code (χ DL) to a reaction outcome with low error in any compatible robotic architecture, c.f. computation—running code on a digital computer. A chemical processing unit converts the inputs (starting materials, χ DL code, and synthetic hardware) into outputs (reaction products, analytical data).

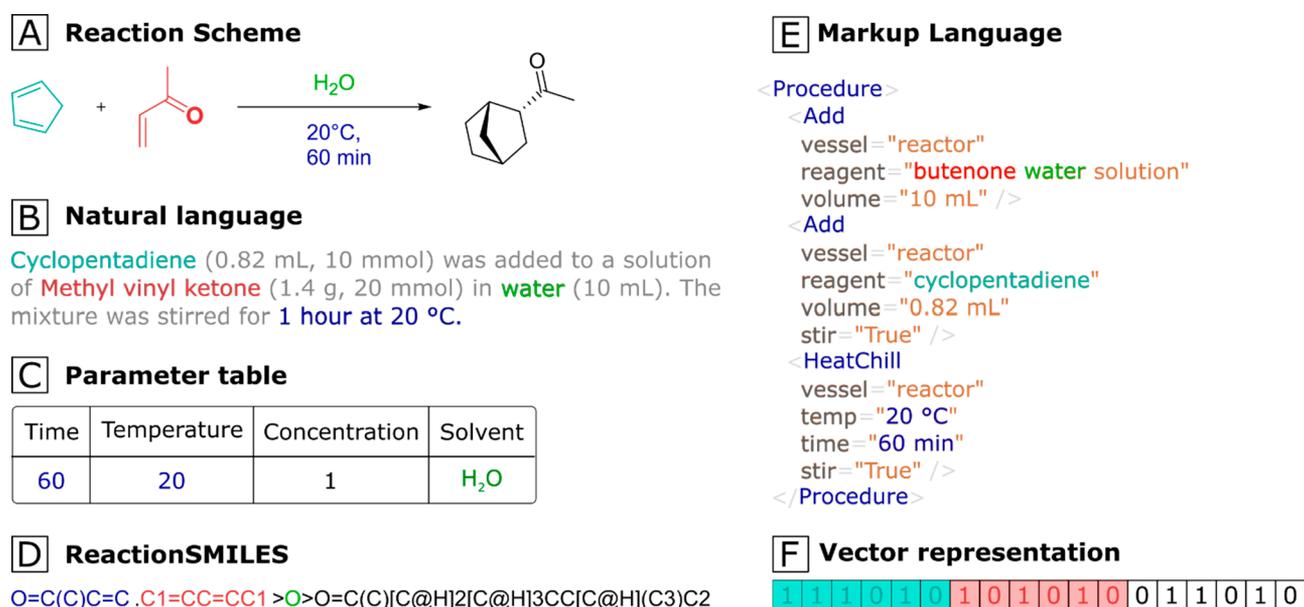


Figure 3. Different representations of the same reaction. Chemists commonly use Kekulé type reaction graphs (A) to depict reactions however these assume many procedural variables. More variables may be captured in prose (B), but this can be challenging for machines to interpret. Parameter tables are commonly used to capture optimization data (C). However, for digital chemistry representations, such as ReactionSMILES (D), vectors (F), or a markup language (E), which may or may not utilize the former, are favored.

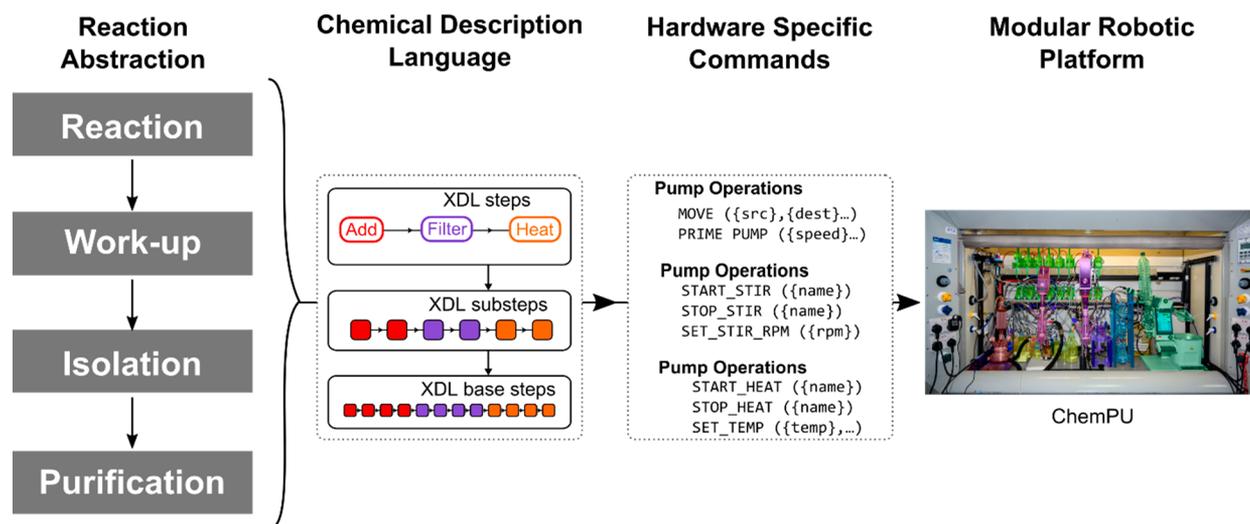


Figure 4. χ DL files act as a bridge between the generalized abstraction of a chemical reaction process and the specific commands required for each platform or hardware.

developed as a pseudomolecular object that represents the reaction by indicating the bonds formed and broken.³⁴

Most machine learning algorithms are designed to work with fixed-length vector representations such as concatenated structural fingerprints, a set of calculated descriptors, or learned representations, see Figure 3F.^{35,21} Graph-neural networks have been employed for reaction prediction tasks,³⁶ but simple text representations in combination with advanced natural language processing (NLP) techniques have achieved a similar performance.³⁷ String representations such as SELFIES or DeepSMILES have recently been developed to increase reliability and ease for data-heavy cheminformatics applications.³⁸

A number of tools have been developed to translate the full synthetic protocols, written in natural language into a machine-readable format.³⁹ Chemical tagger was developed as a rule-

based text-mining tool and eventually led to the US patent and trademark office (USPTO) data set where the procedure is stored as an action sequence in the XML-based chemical markup language (CML), which was developed to allow both humans and machines to disseminate chemical data without information loss.^{40–42} In addition, data-driven approaches for converting text into action sequences have also been demonstrated.⁴³ Despite this most procedures currently in the literature do not contain all of the information required to fully automate a procedure with factors such as stirring rate and the addition rate. Furthermore, other data, such as the actual temperature, pressure, humidity, etc., the reaction was conducted (rather than room temperature) could be critical for successful automation.

In 2019, we introduced our invention of the Universal Chemical Description Language (χ DL, see Figure 3E),^{4,14,44} in

Table 1. Chemical Reaction Databases and Their Respective Data Classes Provided

	Reaxys	USPTO	Pistachio	Open Reaction Database	CAS Reactions
curator	Elsevier	D. Lowe/NextMove	NextMove	ORD Consortium	CAS
source	Gmelin, Beilstein, patents, papers	text-mined US patent grants and applications from 1976 to 2016	text-mined US and EPO Patents from all available years	public data sets + contributed data sets	curated from journals, patents, dissertations, etc.
size	>55 M	3.7 M	>9 M	NA	>136 M
failed reactions	N	N	N	Y	~7 k
classification	Y	N	Y	Y	N
text	Y	Y	Y	Y	Y
conditions	Y	Y	Y	Y	Y
machine-readable actions	N	Y	Y	N	N
open access	N	Y	N	Y	N

which all synthetic procedures can be encoded without ambiguity, executed on any compatible robotic platform, or manually, and exchanged using a standard format, see Figure 4.^{4,45} The aim was to expand on previous work by directly linking the sequences of actions to robotic execution on an automated synthesis machine.¹⁴ This was achieved by generating the universal abstraction of a batch chemical reaction. In general, all batch reactions follow the same procedure: (i) reaction, (ii) workup, (iii) isolation, and (iv) purification. Even complex stepwise synthetic procedures just follow the same abstraction in loops whereby one set of reagents is transformed into a set of products, and this is carried on to the next step and the process occurs again in a loop as the output products from step n become the reagents in step $n + 1$. This key recognition of the modular nature of the abstraction means it should be highly amenable to coding. The key is that many of the operations, once perfected, could be reused again. Furthermore, the abstraction being based on batch has parallels with computational architectures since the hardware could be set into specific states around the batch synthesis, whereas starting the abstraction in flow is harder. This is because flow processes are continuous which makes mathematical analysis difficult to make universal. However, the transformation from batch into flow is much easier because the frequency of the operations can be increased until they are continuous, and since these could have a discrete time stamp it will be easier to globally synchronize them. The instantiation of the abstraction requires the hardware modules to be capable of carrying out the operations described such that the hardware can be reset to the ready state (e.g., by cleaning) after each operation.

The modules that carry out the operations can then be represented on a graph as the nodes and these are given a network (IP) address, whereas the material transfer pathways (e.g., pipes moving liquids or path of a robot arm) are shown as the edges connecting the nodes together. The hardware set detailed in the graph has a defined set of unit operations they are each able to perform, while the χ DL details all the unit operations required to undertake a given procedure. Thus, any robotic system with the required hardware modules organized appropriately on a graph can execute a given procedure by generating a platform specific executable file (XDLEXE) from the universal χ DL. Importantly, χ DL is simple and easily readable by both human and machine but can be produced automatically from natural language by extraction of action

sequences from the procedures which allows the incorporation of key parameters such as order of addition.

Despite the vast wealth of chemical data contained in the literature, failed experiments are commonly omitted from the data published with articles in many journals⁴⁶ even though this data has been shown to be valuable in predicting reaction outcomes.⁴⁷ Most of this “failed” reaction data is currently stored, unindexed, in paper-based or electronic laboratory notebooks in academia or industry and is time-consuming to search even to those with access. With the advent of deep learning for chemistry, machine-readable data sets become increasingly important for computer-aided synthesis planning (CASP), automated synthesis and machine learning applications.⁴⁸ Publicly available databases of chemical reactions remain rare with the most detailed and information rich databases, such as Reaxys, SciFinder, or Spresi, held behind paywalls (Table 1). For this reason, a consortium of academic groups and industry representatives was formed to develop the open reaction database (ORD) - an open-access repository of chemical reactions to fuel research in machine learning applied to chemistry.⁴⁹ Despite this, the ORD does not intend to include digital action sequences geared toward automatic execution of reactions. In addition, analytical data (e.g., from LCMS, IR, NMR) is rarely available in an usable format from any of these databases. One vital component that is currently absent from the databases is the ability to store any run-time or current real-time data collected by sensors attached to the reaction hardware, for example, temperature probe on a stirrer hot plate to name a very simple example. We believe that the ability to collect real time data, or reaction telemetry, will be incredibly useful to both fingerprint successful reactions, record failure, and generate new insights and understanding. The so-called reaction-telemetry fingerprints that will be generated should have great potential to increase the robustness of automated workflows, validate robotically generated data for reaction optimization, and also aid in teaching and training of chemistry at all levels.

Laboratory automation equipment is often expensive, highly specialized for certain tasks and only accessible to a small group of expert users.⁵⁰ Open software and open hardware movements contribute to widening access but with an ever-increasing number of custom robotic platforms, cheminformatic workflows and file formats, there is an increasing need for standardization.⁵¹ The standardization in laboratory automation (SiLA) consortium was formed to develop principles that would enable plug-and-play laboratory

automation.⁵² SiLA 2 was released in 2019 and is based on a microservice architecture and built to connect with laboratory information management systems (LIMS), electronic lab notebooks (ELNs), and other common laboratory software.⁵³ For chemistry, ESCALATE (experiment specification, capture, and laboratory automation technology) was developed as a software pipeline to specify automated experiments and capture generated data in a structured manner.⁵⁴ Research data should be findable, accessible, interoperable, and reproducible (i.e., FAIR) but since the inception of these principles by Mons et al. in 2016⁵⁵ confusion on how to implement them has led to slow adoption.⁵⁶ As HTE workflows become more commonly used to generate large amounts of data, there is an increasing need to validate and verify the data quality, for example, a simple hardware failure could lead to a large amount of false negative reaction data.

In our experience, best practices for digital chemistry research include describing the experiment in a human- and machine-readable mark-up language which provides a high-level user interface, capturing exact versions of control software and dependencies, declaring the platform as a graph with all relevant metadata (such as type and volume of vessels, devices used etc.).^{57,58} The disconnect between process variables and actions, the suggestion of missing values and the removal of ambiguity in reaction data are currently bridged by expert chemists and need to be addressed by future standards. Importantly, the digitization of chemistry remains an ultimately human endeavor which is dependent on a change of the research culture. This means that university curricula must be adapted to provide education and training in digital technologies.⁵⁹ Similarly, existing publication standards must be revised with new applications such as data mining in mind.

3. AUTOMATION OF CHEMISTRY

The synthesis of organic small molecules is still largely performed by hand in a laboratory setting that has barely changed in decades, but experts see the digitization of synthesis fast approaching.^{13,60} Merrifield pioneered the field by introducing the concept of automated solid phase peptide synthesis in 1965.⁶¹ The robust chemistry, simple purification procedure and iterative reaction cycles made the process amenable to automation. Smart laboratory automation holds promise to accelerate chemical research, eliminate tedious tasks, improve safety, and reliability.⁶² Indeed, the automated synthesis of oligopeptides, oligonucleotides,⁶³ oligosaccharides,⁶⁴ and metal oxides^{22,65–67} provided unprecedented access to these compound classes fueling exciting research in the areas of protein biochemistry,⁶⁸ synthetic biology,⁶⁹ chemical glycomics,⁷⁰ and materials. Within the chemistry domain, automation may prove as a valuable tool for studying kinetics, optimization, discovery, and reaction telescoping.^{10,71–73} Over the last few decades, a range of synthetic hardware has become available to handle the throughput and procedural needs of autonomous optimization of chemical reactions (Figure 5).

Flow chemistry is one of the key enabling technologies for automation in chemistry.^{74–76} Ideally suited for solution-based reactions or those with immobilized catalysts, flow systems are easily automated as they rely on the simplest liquid handling robotics, i.e. syringe pumps. The ability to carry out multiple synthetic steps in tandem by facile addition of flow loops, to improve heat or light transfer and to append in-line analytics are all significant advantages for flow chemistry.⁷⁷ These advantages were leveraged for the on-demand synthesis of

		Execution Mode	
		Batch	Flow
Scale	Gram	Isolable Product  Scalable Low Throughput	 Solubility Key Startup Complex High Throughput
	Microgram	 High throughput Low Cost Surface Dynamics	 Solubility Key Low Cost Unique Dynamics

Figure 5. Automation hardware can be categorized by execution mode or by scale. Each has its advantages and disadvantages.

multiple active pharmaceutical ingredients in a single robotic flow chemistry platform.⁷⁸ While this approach required manually adjusting the platform to switch between different processes, a “radial synthesizer” was capable of performing multistep syntheses and optimizations for specific target molecules, as well as derivative libraries without instrument reconfiguration.⁷⁹ The utility of automated flow for the optimization of cross-coupling reactions was demonstrated by performing over 5700 reactions on a micromolar scale in flow with inline HPLC analysis.⁸⁰ Interestingly, it was possible to overcome the common inability of flow systems to meaningfully vary the reaction solvent by injecting a 9:1 ratio of diluent solvent to reagent stock solutions, achieving homogeneous mixing and screening the effect of solvent on the process. Jensen et al. elegantly demonstrated the advantages of flow automation for optimization with their “plug-and-play” modular flow reactor which features discrete loops for heating, cooling and photoirradiation.⁷ Despite these examples, current flow system architectures are not able to access full gamut of reactions available to batch chemists and parallelization of unit operations is required for continuous processing,^{81,82} meaning systems are only reproducible under equivalent flow conditions. In addition, in systems with metal-based flow loops, the role of the hardware in the chemistry may be noninnocent, adding significant complications for reproducibility.⁸³ Finally startup and shutdown of these systems presents additional complications over beginning batch processes adding to the method development required for these syntheses. With the majority of reactions still performed in batch, significant method development is required to adapt known chemistry to flow platforms.

An alternative approach to automation is the combinatorial use of multiparallel batch reactors for high-throughput experimentation (HTE), which became popular in the early 1990s for the synthesis of large diverse screening libraries for drug discovery. This combinatorial chemistry and high-throughput screening paradigm was often blamed for a decline in productivity in the pharmaceutical industry,⁸⁴ and yet it became a valuable research tool whose relevance is underlined by the commercial availability of mature XYZ Gantry (Chemspeed, Tecan) and ultralow-volume pipetting platforms (Mosquito). In recent years, we have observed a renaissance in the use of massively parallel, miniaturized ultrahigh-throughput

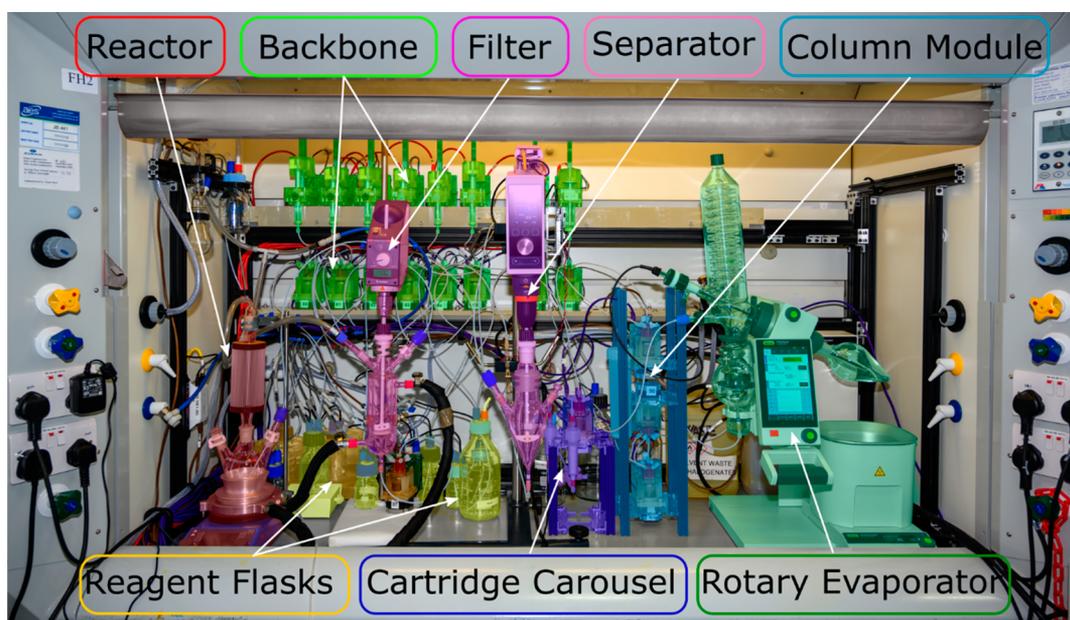


Figure 6. Photograph of the modular, universal ChemPU (Previously Chemputer) platform.

experimentation^{35,85–87} combined with design of experiments (DoE) and other screening techniques applied to these reactors for discovering and optimizing novel reactivity, properties⁸⁸ and even bioactivity,⁸⁹ though not necessarily isolation procedures.⁹⁰ These reactors, usually based on 96-, 384-, or 1536-well plate type designs, and allow hundreds of reactions to be run at once under the same process conditions.⁸⁰ Despite this, the material consumption and waste were minimized because of the small reaction scale. While extremely powerful, this setup is far from general as there are significant experimental constraints imposed by the hardware (scale, compatible solvents, feasible temperature range) and such approaches appear to miss the flexibility required to automate multistep organic syntheses. Progress in recent years has led to significant progress toward a universal batch synthesis platform.^{21,72} The synthesis of many different types of small molecules in one automated process using *N*-methyliminodiacetic acid (MIDA) boronate building blocks could be accomplished by applying iterative synthesis similar to peptide synthesis enabled by a general MIDA catch-and-release purification protocol.⁹¹

In 2019, we reported the development of a new approach to chemical synthesis architectures we first embodied in the “Chemputer” (now known as the ChemPU, Figure 6) providing standard software and hardware for complete automated synthesis and workup of a range of organic compounds based on a universal liquid handling backbone and modular additions for filtration, extraction, solvent evaporation, etc., see Figure 4.¹⁴ This platform emulates the traditional process operations, which would be carried out manually by a laboratory chemist and since the vast majority of the literature is based on batch chemistry, automation of these syntheses requires robotics founded in batch. Extension of the modularity of the original platform allowed for the execution of a wide range of chemistries on a single platform, including cross coupling, amide bond formation via peptide synthesis and diazirine formation.⁵⁷ Importantly, the hardware modules, which make up each platform in the ChemPU family, are represented graphically and flexibly modified for each

procedure using an online GUI. Both the hardware graph and the procedure are required for the generation of a platform specific executable to run the procedure. The procedure files for the ChemPU use the simple human and machine-readable chemical description language (χ DL) which is hardware independent and therefore represents a universal chemical coding format. These χ DL files describing synthetic procedures can be executed directly on the platform with minimal human intervention recording accurately every unit operation undertaken by the ChemPU.⁴ χ DL files and hardware graphs can then be shared via Github upon publication allowing for others to reproduce these procedures exactly as performed in Glasgow even on different hardware as long as it meets the minimum standards for χ DL compatibility.⁴

3.1. Control Systems

Simple experimental tasks can easily be automated using microcontrollers or single board computers with only a few lines of Python or Arduino (C++) code.⁹² While short linear scripts and hardcoded variables may suffice in some applications, robust and reusable software is needed to meet the reproducibility expectations of work that is to be classed as scientific research. Contrary to the expectation that digitization and automation should lead to increased reproducibility, code from many academic software projects does not execute, even just a few years after initial publication.⁹³ This phenomenon is commonly known as “bit rot”, a term describing the apparent decay of software over time, can only be prevented by good software development practices and active maintenance by highly skilled programmers. This is hard to achieve in fast-paced research environments and with the high turnover of junior researchers in academia. Therefore, it cannot be expected that the end-user, most likely a chemist, is proficient in programming for broad adoption, but at the same time a well-designed user experience (UX) is of key importance.⁹⁴ A simple standardized, flexible digital framework for chemical operations, interfacing with hardware, and controlled by a procedure code that links the operations to the framework is required for a digital revolution in the field because it could be

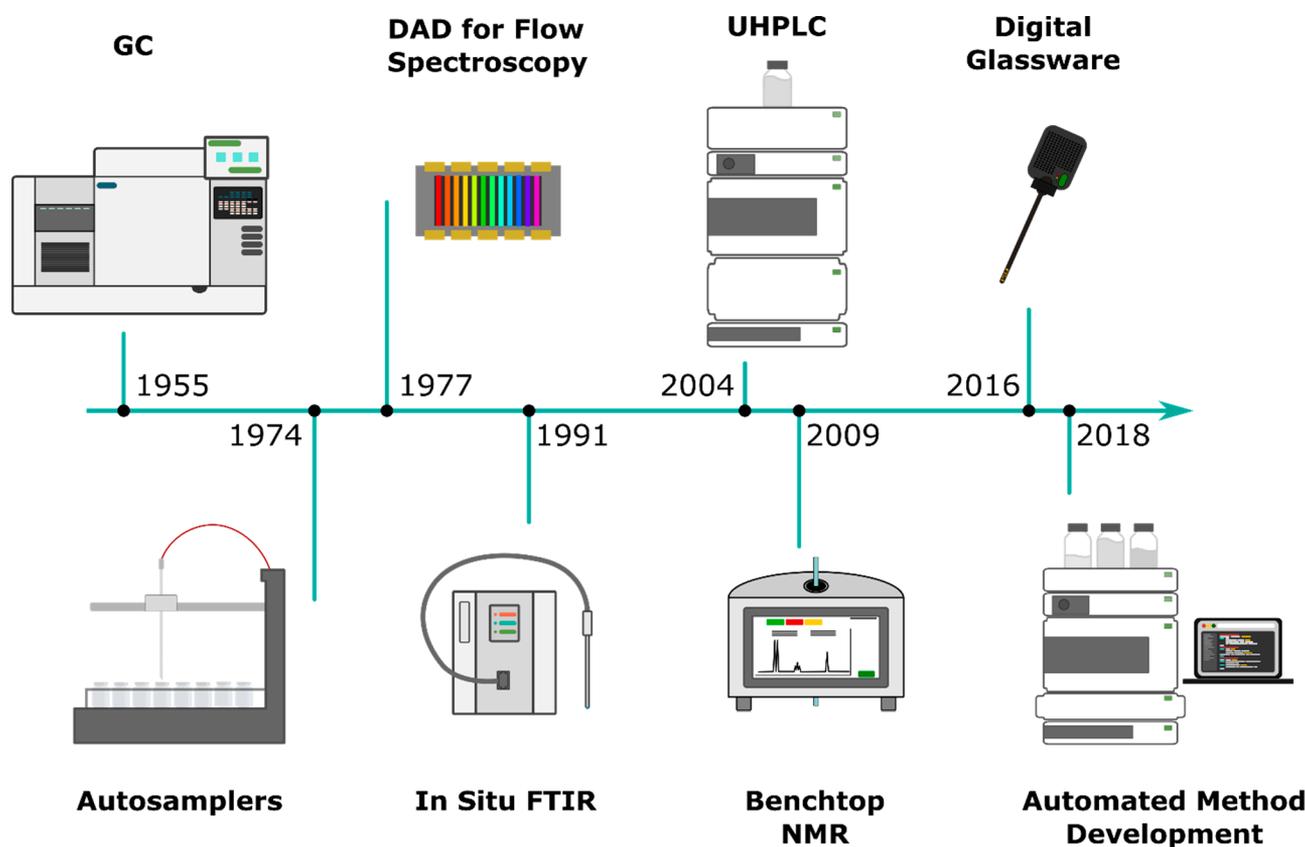


Figure 7. Advances in enabling analytics for automated reaction optimization. The earliest automation system was devised for gas chromatography (GC) with automatic flow rate and temperature control however automation really drove advances in science with the advent of autosamplers which allowed for hundreds of samples to be analyzed with minimum labor, freeing up scientists for data analysis. The ability to conduct analysis in flow has also been a strong driver of automation with diode array detectors (DAD) for UV–vis being among the first advances to facilitate this. In situ analytics, such as Fourier transform infrared spectroscopy (FTIR) and Digital Glassware, also allow for rapid or continuous monitoring of reactions in real time while UHPLC and benchtop NMR both cut the time and effort required for more advanced analytics. Currently method development for chromatographic separations is still a key bottleneck but new automated method development systems are becoming available to optimize separations autonomously.

designed with UX in mind, and in such a way that it could be maintained and developed collaboratively, c.f., Linux.

Commercial equipment often has proprietary software that comes with a user-friendly interface and, in some cases, basic scripting capabilities. However, closed platforms can lead to a vendor lock-in and impose barriers to innovation in research laboratories. For custom robotic workflows, open application programming interfaces (APIs) are therefore of great importance to integrate third party equipment and software.⁹⁵ We believe independent bodies, such as the SiLA consortium, could and should establish standard APIs to interface with a wide variety of commercial laboratory equipment. LabVIEW by National Instruments (NI) is a generalized proprietary control software for a range of automation workflows integrating equipment from different third-party vendors.⁹⁶ Its graphical programming approach allows users without programming experience to develop workflows via drag-and-drop. Furthermore, it is inherently concurrent, allowing for parallel execution and provides advanced signal processing capabilities. Since LabVIEW is limited in the areas of scientific programming including optimization, signal processing, statistics, and machine learning it is commonly coupled with Matlab.⁷ However, the ecosystem is considerably smaller than Python's, for example, which is a general-purpose programming language and can be used to build sophisticated software

solutions for automation-enabled or “self-driving” laboratories. Such solutions may combine chemical robotics with AI planning, database-management systems but also chatbots frameworks integrated in social media platforms for interaction with the human researchers.^{22,97,98}

Hardware drivers, a platform operating system (OS), and bindings between this OS and the hardware-independent χ DL form the software stack needed to run a ChemPU (Figure 2).⁴ This layered approach allows for simultaneous low-level access to the hardware for debugging and development purposes as well as high-level scripting capabilities for synthetic chemists with no programming experience using the χ DL language. Importantly, this also allows the χ DL to be run on different hardware for example using different drivers and OS, so long as equivalent bindings are present linking the unit operations detailed in the χ DL to those performed by the hardware, see Figure 4. A web application, ChemIDE, was also developed as a human-friendly graphical user interface to allow chemists to directly develop chemical programs with little or no programming experience.

4. DATA COLLECTION

Rapid real-time analytics are fundamental to optimization and a range of techniques have been developed to facilitate this alongside hardware advances which allow their incorporation

into automated systems (Figure 7).⁹⁹ Perhaps even more so than for synthesis, standardization of analytical hardware and data formats is vital and improving data standards means, increasingly, analytical data is available alongside publications, however, the facile machine readability of this data and interpretation in context of its source hardware still present challenges in developing standards. Inline Raman and IR spectrometers, in particular React-IR, have significantly increased the impact of this low energy spectroscopic method for real-time reaction monitoring and thus rapid feedback for optimization.^{100–102} The short time scale of the IR experiment means a vast number of data points can be collected for each reaction and the ability to home-in on particular functional groups of interest, to the absence of noise created by other reagents, makes this a useful technique for use in autonomous digital optimization. UV–vis methods also have the advantage of allowing one to focus on a specific wavelength; however, the resolution of UV–vis versus for example, IR makes analysis of this data more challenging. These methods may thus find more use in the optimization of inorganic compounds where a small difference in coordination environment leads to significant changes in the UV–vis spectrum.¹⁰³

However, analyzing the shape of a time-dependent UV–vis absorption plot at a fixed wavelength has been shown to be useful for peptide synthesis.¹⁰⁴ The inclusion of the time dimension allows a range of additional parameters be elucidated including reaction rate and reagent diffusion with different deprotection reagents and thus optimized using deep learning data analysis approaches. Gas and high-pressure liquid chromatograph methods are both well suited for automation, in-line analytics and subsequent reaction optimization and can be easily paired with other analytic techniques (e.g., UV–vis, MS).⁷ However, several key factors limit their use. Both methods require long experiment times, preventing high-throughput analysis, and they may require method development to ensure valid data is achieved when varying reaction parameters. In addition proprietary software and equipment means that the experimental methods and hardware modules are significantly less standardized across different systems than for IR and UV–vis spectroscopies.¹⁰⁵

Finally, the information content of chromatographic methods alone can be low, particularly if they are not utilized in combination with analytical standards. Ultrahigh pressure (UHPLC) and techniques, such as flow injection analysis (FIA-MS),¹⁰⁶ multiple injections in a single experiment (MISER),¹⁰⁵ and 2D-LC analysis, have been developed to reduce experiment times for HTE.¹⁰⁷ Agilent have also recently begun producing an automethod development system, InfinityLab, incorporating up to 8 different columns and 15 mobile phases for rapid autonomous chromatographic method development.¹⁰⁸

Mass spectrometry for optimization is most often coupled with a chromatographic method allowing for more accurate quantification of relative yields and clearer identification of byproducts. For MS alone the high sensitivity and low sample mass requirements are a significant advantage, however, long experiment times can also be a factor here. Mass spectrometry has also been applied in flow systems for optimization of reactions, which allows the incorporation of feedback loops in the synthesis and, thus, more autonomy in the optimization process.¹⁰⁹ Benchtop APCI MS has been utilized to optimize the formation of nicotinamide in such a system, using only 18

experiments, with yield calculated by normalization of the $[M + H]^+$ adducts.¹¹⁰

Traditional NMR spectroscopy is based on batch analysis, with autosamplers allowing for more rapid throughput of samples; however, the recent inception of flow-NMR has truly allowed the technique to proliferate for automated reaction optimization.¹¹¹ Flow NMR has the advantage of having highly transferable data analysis, high information content, and moderate acquisition times (~ 0.5 s scan⁻¹).¹¹² The variety of nuclei, techniques, and ongoing advances also makes this method widely applicable to different chemistries, which may not be suitable for MS or chromatography.¹¹³ For example, insights gained from mechanistic analysis can inform reaction optimization and flow NMR has been utilized for the in situ study of reactive organometallics representing catalytic intermediates.¹¹⁴ Even common challenges in NMR such as overlapping peaks have been overcome utilizing state of the art algorithms.¹¹⁵

A prerequisite for autonomous operation of any laboratory equipment is dynamic feedback. While condition monitoring and process control are routine tasks in chemical engineering, it is much less common in academic chemistry research laboratories.⁷³ Tasks such as measuring the pH value or ensuring a flask is empty are trivial for a human researcher but challenging and crucial for the safe operation of automated laboratory equipment. Data from “human monitoring operations”, such as the visual inspection of a reaction (e.g., to determine if precipitation or a color change has occurred), are not consistently captured. Emerging technologies enable chemists to acquire, share, and analyze digital data sets of their chemical experiments (Table 2).^{92,116,117} Recent

Table 2. Sensors Which Can Contribute to Automation of Standard Laboratory Procedures

sensor	example use case
pH	“Adjust the pH to 7.0 with NaOH (1 M).”
conductivity	“The solution was extracted with ethyl acetate.”
bubble	“Check if line is clogged.”
viscosity	“Heat until gelation occurs.”
turbidity	“Add hexane until a fine precipitate forms.”
liquid level	“Notify if waste receptacle is full.”
temperature	“Add dropwise, maintaining the temperature below 0 °C.”
color	“Stir until the color changes from red to blue.”

examples of this development are the “Smart Stirrer” capable of measuring reaction conditions, such as temperature, conductivity, visible spectrum, opaqueness, stirring rate, and viscosity in situ, and the use of low-cost optical bubble sensors to control a hydrogenation reaction. Also, the open-source software package Heinsight, which uses webcams and computer vision algorithms to monitor liquid levels, is being used for diverse applications, such as continuous preferential crystallization, slurry filtration, and solvent swap distillation.^{118–120}

One important point which arises for all of these analytical systems is the need for robust data sharing processes upon publication.¹²¹ X-ray crystallography represents a case study in best practice with regards to analytical data sharing. Currently, an article containing an X-ray diffraction crystal structure must include a crystallographic information file (cif) incorporating all of the details of the equipment used, experimental parameters, raw data and processed output.¹²² All crystallog-

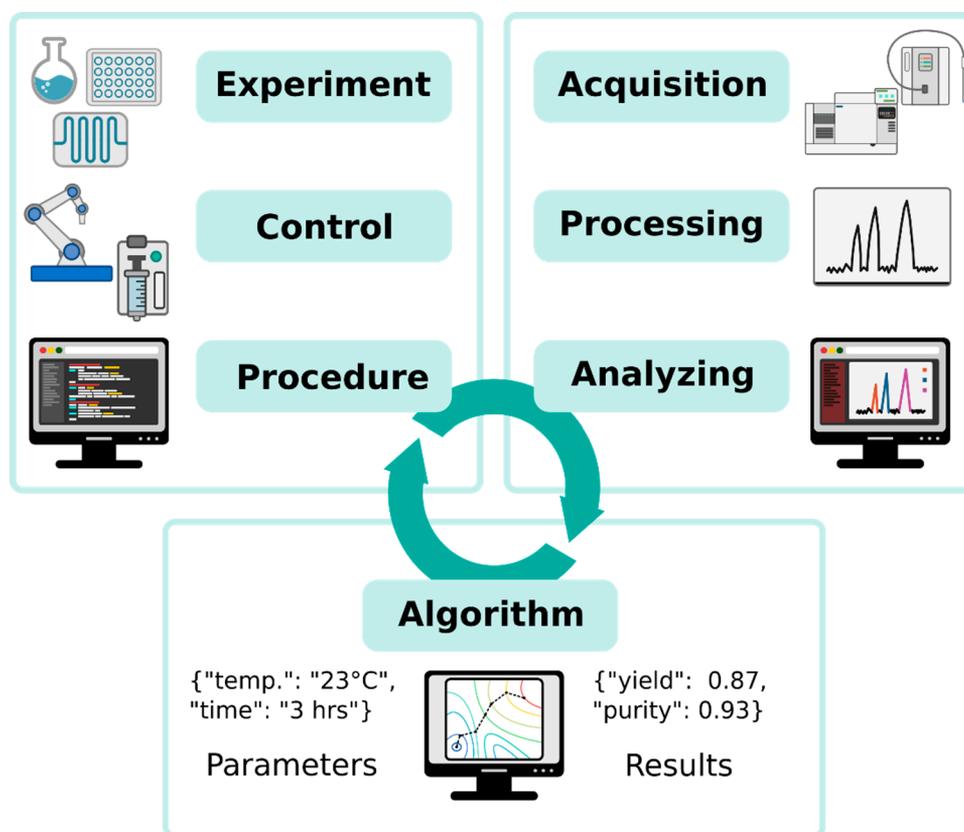


Figure 8. Workflow for the automated chemical reaction optimization: the given procedure is automatically executed on an experimental platform using a corresponding control system; the results of the process are evaluated using an analytical instrument with the resulting data processed and analyzed to yield an outcome that can be understood by an optimization algorithm; the algorithm's output is used to update the procedure parameters and thus start the next iteration of the process.

raphy software is unified in its ability to produce identical output files and read identical input files and this data is stored in the easily visualizable and searchable Cambridge Structural Database as well as alongside the corresponding publications.¹²³ At publication this file is subject to CheckCIF standards checks, a report of which must be provided to reviewers and editors.¹²⁴ For other analytical data types, no such requirement exists but universal data standards for other methods, the quality of which can be validated, would be beneficial to the whole community. FIDs from NMR spectrometers,¹²⁵ CSV files from spectrophotometers, mzML files from mass spectrometers,¹²⁶ and similar data from GC/HPLC systems should all be a minimum requirement for publication of results, which rely upon this data.

Only through the peer review system can we effectively ensure data standards are maintained across the discipline. One feasible example format for sharing data at publication could be the JCAMP-DX format which has been demonstrated for use in all these analytical techniques and more, with these files detailing both spectral parameters and metadata.¹²⁷ An alternative format could be the Analytical Information Markup Language (AnIML)—a XML-based solution for storing analytical data from a variety of instruments and techniques, offering a validation process via strictly defined schema.¹²⁸ Our vision includes the sharing of analytical (e.g., in aforementioned format) and process (in tabular format, e.g., a csv file, with the timestamps for each measurement) data alongside the χ DL process file. Thereby each “version” of data created by a new experiment run includes all data for that run, including

any failed runs, and specifications/parameters on all the hardware used to collect such data alongside the procedural information contained in the χ DL.

5. OPTIMIZATION

Process optimization is among the most tedious and labor-intensive tasks within scientific and engineering disciplines, and chemistry is no exception. Even achieving satisfactory reaction conditions is nontrivial and can take weeks of experimentation for a human researcher.^{20,23,129} Most recent developments in reaction automation have been applied to facilitate optimal parameters discovery, however the vast majority of the published platforms are bespoke systems, capable of performing single reaction optimization in flow.¹³⁰ These platforms, despite demonstrating proof-of-concept results, are typically limited to a specific task, while creating a universal, fully automated framework remains challenging (Figure 8). In 2018, a reconfigurable system was presented with modules to perform flow chemistry, including temperature control and a photoreactor, that can run closed-loop optimization in conjunction with various analytical instruments.⁷ The same year OpenFlowChem—a platform for flow chemistry automation, providing communication protocols for analytical instruments and control systems—was demonstrated to run the process of self-optimization using a PID algorithm, as well as predicting the optimal reaction conditions for a semi-hydrogenation reaction.¹³¹ A significant disadvantage of this approach is the usage of commercial software such as LabVIEW for instruments control, and experiment manage-

ment accompanied by tools for data processing and analysis, most commonly Matlab.^{96,132} While this software suite can create closed-loop experiments with analytical feedback and an optimization algorithm generating new parameters, the setup is not easily transferrable across automation platforms due to proprietary licenses which are not widely accessible.

Recently two new approaches to chemical optimization were reported, Summit¹³³ and Olympus,¹³⁴ benchmarking frameworks offer a large set of optimization strategies together with virtual benchmarks, and an experimental planning toolkit. In contrast to the aforementioned software, these frameworks are fully open-sourced and due to their open interface, it is possible to plug in any given algorithm, and this could be universally adapted to any typical automation platform.

5.1. Data Processing/Treatment

The foundation of a closed-loop reaction optimization system needs to be the automated processing and analysis of the analytical data in real time.⁹⁵ Any reaction outcome (i.e., product spectrum) should be translated into distinct descriptors (i.e., product yield) for the optimization algorithm to process. Initial processing, for example, noise reduction or baseline correction, could be performed within an automated system by the instrument operating software through either manually written macros or an application programming interface (API). Given that the data format specification is provided by the manufacturer, the resulting output may be further analyzed using third party software which may be included in the automation workflow. The LabVIEW/Matlab suite is a common choice to process the analytical data and utilize the output for the mathematical optimization.¹³⁵ With the necessary drivers provided to control the instruments, it can create a robust environment for the reaction optimization, although only available under a proprietary license. With the emergence of open-source software, several packages were developed to process data from a variety of analytical instruments, including HPLC¹³⁶ and NMR.¹³⁷ Such programs often include the graphical user interface (GUI) and a programming interface for seamless integration into any chemical automation platform. With the source code open to the community, the software can be extended with the novel processing algorithms and accommodate new instruments on the market, if this is permitted by the license agreement for the latter.

Manual product assignment is a common approach to selecting the reaction outcome where the reference sample is analyzed by the human expert, and the respective descriptors are “hard coded” into the experiment workflow. However, manual analysis is not suitable for a fully autonomous systems, and full assignment of data on a novel compound or reaction can be time-consuming for either those discovered experimentally or predicted using a retrosynthetic analysis. Recently, supervised and unsupervised machine learning techniques were proposed to aid with identifying signals of interest from raw spectra or direct assignment of ¹³C and ¹H spectra to proposed structure;^{138–141} however, these approaches were not applied in context of reaction automation. This is, thus, a very attractive direction in the creation of a fully autonomous system for discovery of new materials and synthetic routes and subsequent optimization within the same system.

5.2. Algorithms for Decision Making

An extremely important part of the reaction optimization is minimizing the total number of time-consuming experiments

that often involve expensive reagents by maximizing the information learned at each iteration. Design of experiments (DoE) was one of the first techniques to formalize the screening process and is used to build a model that describes the relationship between experimental inputs (e.g., reaction temperature or catalyst loading) and outputs (e.g., yield or product purity).^{1,2} This approach can guide optimization by mapping the optimal conditions or initiate exploration of a search space for more sophisticated algorithms to exploit.

Traditional algorithms for a function optimization, such as Nelder–Mead Simplex (and modifications thereof) or gradient-based methods, were among the first used in chemical self-optimization tasks.¹⁴² The significant disadvantage of these local optimizers is their inability to tackle experimental noise which can lead to the premature halting of the optimization far from the true system optimum. Furthermore, the overall growth the complexity of a given process, and an increase in the amount of input parameters, may lead to multiple optima, where such local optimization algorithms are not applicable, despite their robustness and small computation times. The SNOBFIT algorithm was one of the first techniques created to target global optimization of noisy and expensive-to-evaluate functions (i.e., experiments).¹⁴³ Despite the limitation on number of input parameters and their continuous nature, the overall performance and availability made SNOBFIT the go-to method for running self-optimization experiments, and a reference for future single-objective optimizers. More recently, a different category of algorithms has gained increased attention in machine learning community. Bayesian optimization utilizes a surrogate model as an approximation of an experiment or simulation that is updated with more sample data in conjunction with an acquisition function (Figure 9).¹⁴⁴ These methods were designed for the global optimization of noisy “black box” functions that are expensive to evaluate and have been adapted for use within an experimental environment.¹⁴⁵ It was also shown that this algorithm can guide an automated flow system to find optimal conditions between multiple competing outcomes.¹⁴⁶ Discrete and categorical parameters, such as solvent or catalyst choice, represent a significant portion of the input variables for chemical processes and are major challenge in developing optimization algorithms. Previously only available using the DoE approach and the response surface methodology,¹⁴⁷ such variables can be easily incorporated into Bayesian optimization due to flexibility of surrogate models.¹⁴⁸ Another advantage of this approach is reduced number of iterations needed to achieve the best outcome. It was reported that optimal conditions can be found after exploring only tiny fraction of parameter space.⁵ It is worth noting, that despite current trends and recent developments, one should not be biased when selecting a strategy to solve the optimization task. Traditional machine learning methods and classical mathematical algorithms for function minimization should also be considered, if they can provide an efficient solution. A more detailed description and general overview of algorithms used in chemical optimization are presented in reviews by Bourne and co-workers,¹⁴⁹ Houben and Lapkin,¹³⁵ and Cronin et al.¹⁸

One major challenge for current chemical optimization is incorporating chemical knowledge (e.g., reagent structure, solvent, catalyst nature, etc.) into the overall workflow. Despite significant advances in retrosynthetic planning,¹⁵⁰ machine learning methods that encode the chemical structure for predictions have not found a wide application in optimization

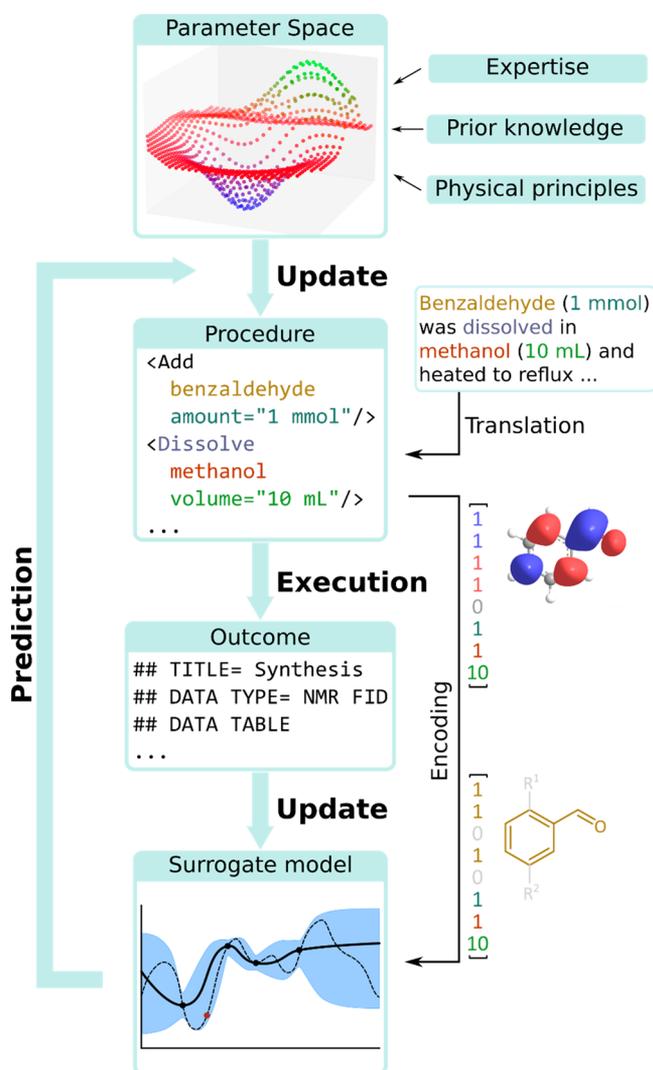


Figure 9. Workflow of the Bayesian reaction optimization: the reaction code, predefined by the parameter space which may be partially informed by prior knowledge and expertise, is executed to result an outcome (illustrated as JCAMP-DX NMR spectrum header); the outcome is then used to update the surrogate model, built using the encoded chemical and conditions data, to predict the next set of reaction conditions.

tasks. Most algorithms treat the chemical reaction as a “black-box function”, not only ignoring the physical principles of the reaction, but disregarding data from previously published results or chemical databases in the initial modeling. Previous attempts to incorporate structural knowledge for predicting reaction conditions were based on nearest-neighbor approach, which recommended similar conditions for similar substrates.¹⁵¹ The neural network model for encoding structure information has however demonstrated *in silico* efficiency in predicting suitable reaction conditions when trained on large chemistry databases.¹⁵² When combined with Bayesian optimization methodology, chemical encoding (achieved using DFT-descriptors) shows excellent performance across several chemical reactions, compared to traditional DoE approach or human expertise, when applied to mixed categorical-continuous parameter domains.⁶ Studies with deep reinforcement learning have shown that a model trained on a specific reaction can also be transferred to a similar or different reaction class to improve the performance and reduce the number of experiments to identify the best conditions.¹⁵³ Overall, this paradigm for the chemical optimization opens a new perspective for predicting the optimal parameters for known reactions but also suggests the conditions for yet unknown reactions, designed using algorithmic retrosynthetic analysis.¹⁵⁴

6. CONCLUSIONS

It is clear from the work described in this Perspective that automation is already ubiquitous in the chemical lab environment. From autosamplers to flow reactors, chemists around the world are taking advantage of the labor-saving benefits and increased reproducibility of robotic systems. Thus, the digitization of chemistry should not merely be considered a problem for the future but a challenge of the present. However, there are still strides to be made to reach full automation of chemical optimization or even fully autonomous laboratories, which can discover, optimize the synthesis of and analyze the mode of action of a new lead compound. Comparisons can be drawn with other technological advances, such as cars: when first steam prototypes were built, they were rejected by society as noisy, dangerous and destructive to roadways. With early developments of gasoline engines, cars were still considered to be less practical than common horse wagons, because of

Table 3. Challenges and Potential Solutions for the Main Factors in Successful Automation of Optimization of Chemistry

topic	challenge	solution
optimization	algorithm implementations not suited for laboratory use	unified ask-and-tell interface
	chemistry often treated as a black box	big data analysis to gain new insights into reactivity
execution	high experimental cost, optimization only out of necessity	incorporate prior knowledge for optimizations with lower experimental budgets
	transparency and reproducibility	specify device capabilities and limitations
	no plug-and-play automation	establish an open standard
	reliability	control software with error correction
collaboration	interoperability	learn the “delta” between different platforms from empirical data
	slow knowledge transfer	share digital code so that optimized procedures can be directly used in a different lab
data management	scalability	distributed clients working together via a central server
	programming	no code, GUI, webapps, slack integration
	no platform fits all needs	modular platforms, cross-platform code portability
	disconnect between variables and actual actions	record action sequences, map variables
	different research cultures, formats, etc.	human and machine-readable format, FAIR principles
	missing values and procedural ambiguity reduce reproducibility	synthetic chemistry experts needed to bridge the gap

regular stalling and high demand in servicing. A significant stream of technological advances (e.g., four-wheel brakes, independent suspension and three-point seat belts), as well as standardization of mass production led to cars that are robust, cheap, and safe enough to dominate the field of private transport. Several decades ago, car owners required basic mechanical skills for everyday maintenance, while nowadays cars can self-diagnose a range of internal faults. With the current rate of development of self-driving cars, in few years one might not even need a license to drive. Chemistry is poised to undergo a similar “great leap forward” by drawing upon the automation and computational advantages described above but there are still several key hurdles to be overcome to transition the remaining areas of manual input to fully autonomous systems (Table 3). Integration and interoperability of hardware is vital, including the presence of standardized hardware interfaces for control and data transfer as well as open access API to empower flexible use of hardware. Another major hurdle for the more elaborate automated workflows is the reliability of current systems, and here we expect real-time monitoring for error detection and correction to play a vital role in guaranteeing the quality of robotically generated reaction data sets. We also urgently need data standards integrated into our publication and output systems, whereby data is open access, acquisition parameters are fully detailed, and the data is in a simple machine-readable format, subject to verification checks and easily searchable using nonproprietary databases. Importantly, this data must include details of automated hardware used for synthesis and the precise action sequence followed in a universally readable format. Missing values and ambiguities (e.g., vigorous stirring versus 1000 rpm) in published data lead to unnecessary barriers when adopting literature procedures to automated platforms. Sharing digital code, such as χ DL, is likely to enhance knowledge transfer and scientific collaboration especially when these digital tools are accessible for researchers without programming skills. This can be facilitated by developing web applications and GUIs or interfacing laboratory automation control software with popular, existing workplace tools. Finally, since research is ultimately a human endeavor researchers must have access to training in digital skills and clear directions on how automation tools are used in cutting edge research such that the barrier to adoption of these tools is lowered. Only when the expertise of chemists is combined with the advances allowed by digital tools will the promise of the digital chemistry revolution be fulfilled.

To fully leverage the time saving ability of digitization, in the future chemical processes must be able to be optimized autonomously by integration of closed-loop feedback processes in concert with state-of-the-art algorithms. Given the expensive and complex nature of chemical experiments, algorithms should provide an interface that allows control over the optimization loop and frameworks, such as Summit and Olympus, are expected to prove highly useful for closed-loop systems. When combined these advances will revolutionize scientific collaboration and innovation not only within chemistry but in a wide range of downstream applications such as pharmacy, materials, food technology and energy ultimately bringing the central science into the 21st century.

AUTHOR INFORMATION

Corresponding Author

Leroy Cronin – School of Chemistry, University of Glasgow, Glasgow G12 8QQ, United Kingdom; orcid.org/0000-0001-8035-5757; Email: Lee.Cronin@Glasgow.ac.uk

Authors

Alexander J. S. Hammer – School of Chemistry, University of Glasgow, Glasgow G12 8QQ, United Kingdom

Artem I. Leonov – School of Chemistry, University of Glasgow, Glasgow G12 8QQ, United Kingdom

Nicola L. Bell – School of Chemistry, University of Glasgow, Glasgow G12 8QQ, United Kingdom; orcid.org/0000-0002-7497-9667

Complete contact information is available at:
<https://pubs.acs.org/10.1021/jacsau.1c00303>

Funding

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). A.J.S.H. acknowledges a scholarship from the Studienstiftung des deutschen Volkes.

Notes

The authors declare the following competing financial interest(s): L.C. is the founder and a shareholder in DeepMatter, the company which produces Digital Glassware (DG) mentioned here, and is also the founder of Chemify. The XDL standard code is freely available from gitlab: <https://gitlab.com/croningroup/chemputer/xdl> and the documentation is available at www.xdl-standard.com.

REFERENCES

- (1) Carlson, J. E.; Carlson, R. *Design and Optimization in Organic Synthesis*; Elsevier: Amsterdam, 2005.
- (2) Weissman, S. A.; Anderson, N. G. Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications. *Org. Process Res. Dev.* **2015**, *19* (11), 1605–1633.
- (3) Bergman, R. G.; Danheiser, R. L. Reproducibility in Chemical Research. *Angew. Chem., Int. Ed.* **2016**, *55* (41), 12548–12549.
- (4) 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* (6512), 101–108.
- (5) Reker, D.; Hoyt, E. A.; Bernardes, G. J. L.; Rodrigues, T. Adaptive Optimization of Chemical Reactions with Minimal Experimental Information. *Cell. Rep. Phys. Sci.* **2020**, *1* (11), 100247.
- (6) 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* (7844), 89–96.
- (7) 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* (6408), 1220–1225.
- (8) Gromski, P. S.; Granda, J. M.; Cronin, L. Universal Chemical Synthesis and Discovery with ‘The Chemputer’. *Trends Chem.* **2020**, *2* (1), 4–12.
- (9) Zalesskiy, S. S.; Kitson, P. J.; Frei, P.; Bubliauskas, A.; Cronin, L. 3D designed and printed chemical generators for on demand reagent synthesis. *Nat. Commun.* **2019**, *10* (1), 5496.

- (10) Kitson, P. J.; Glatzel, S.; Chen, W.; Lin, C.-G.; Song, Y.-F.; Cronin, L. 3D printing of versatile reactionware for chemical synthesis. *Nat. Protoc.* **2016**, *11* (5), 920–936.
- (11) Symes, M. D.; Kitson, P. J.; Yan, J.; Richmond, C. J.; Cooper, G. J. T.; Bowman, R. W.; Vilbrandt, T.; Cronin, L. Integrated 3D-printed reactionware for chemical synthesis and analysis. *Nat. Chem.* **2012**, *4* (5), 349–354.
- (12) Cronin, L.; Mehr, S. H. M.; Granda, J. M. Catalyst: The Metaphysics of Chemical Reactivity. *Chem.* **2018**, *4* (8), 1759–1761.
- (13) Davies, I. W. The digitization of organic synthesis. *Nature* **2019**, *570* (7760), 175–181.
- (14) 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* (6423), eaav2211.
- (15) Hou, W.; Bubliskas, A.; Kitson, P. J.; Francoia, J.-P.; Powell-Davies, H.; Gutierrez, J. M. P.; Frei, P.; Manzano, J. S.; Cronin, L. Automatic Generation of 3D-Printed Reactionware for Chemical Synthesis Digitization using ChemSCAD. *ACS Cent. Sci.* **2021**, *7* (2), 212–218.
- (16) 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* (6373), 314–319.
- (17) 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* (2), 119–128.
- (18) Henson, A. B.; Gromski, P. S.; Cronin, L. Designing Algorithms To Aid Discovery by Chemical Robots. *ACS Cent. Sci.* **2018**, *4* (7), 793–804.
- (19) Herres-Pawlis, S.; Koepler, O.; Steinbeck, C. NFDI4Chem: Shaping a Digital and Cultural Change in Chemistry. *Angew. Chem., Int. Ed.* **2019**, *58* (32), 10766–10768.
- (20) Salley, D. S.; Keenan, G. A.; Long, D.-L.; Bell, N. L.; Cronin, L. A Modular Programmable Inorganic Cluster Discovery Robot for the Discovery and Synthesis of Polyoxometalates. *ACS Cent. Sci.* **2020**, *6* (9), 1587–1593.
- (21) 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* (7714), 377–381.
- (22) Duros, V.; Grizou, J.; Sharma, A.; Mehr, S. H. M.; Bubliskas, A.; Frei, P.; Miras, H. N.; Cronin, L. Intuition-Enabled Machine Learning Beats the Competition When Joint Human-Robot Teams Perform Inorganic Chemical Experiments. *J. Chem. Inf. Model.* **2019**, *59* (6), 2664–2671.
- (23) Porwol, L.; Kowalski, D. J.; Henson, A.; Long, D.-L.; Bell, N. L.; Cronin, L. An Autonomous Chemical Robot Discovers the Rules of Inorganic Coordination Chemistry without Prior Knowledge. *Angew. Chem., Int. Ed.* **2020**, *59* (28), 11256–11261.
- (24) Wilbraham, L.; Mehr, S. H. M.; Cronin, L. Digitizing Chemistry Using the Chemical Processing Unit: From Synthesis to Discovery. *Acc. Chem. Res.* **2021**, *54* (2), 253–262.
- (25) Pitzer, L.; Schäfers, F.; Glorius, F. Rapid Assessment of the Reaction-Condition-Based Sensitivity of Chemical Transformations. *Angew. Chem., Int. Ed.* **2019**, *58* (25), 8572–8576.
- (26) Gensch, T.; Teders, M.; Glorius, F. Approach to Comparing the Functional Group Tolerance of Reactions. *J. Org. Chem.* **2017**, *82* (17), 9154–9159.
- (27) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* **1988**, *28* (1), 31–36.
- (28) Grethe, G.; Blanke, G.; Kraut, H.; Goodman, J. M. International chemical identifier for reactions (RInChI). *J. Cheminf.* **2018**, *10* (1), 22.
- (29) Heller, S.; McNaught, A.; Stein, S.; Tchekhovskoi, D.; Pletnev, I. InChI - the worldwide chemical structure identifier standard. *J. Cheminf.* **2013**, *5* (1), 7.
- (30) Yang, C.; Tarkhov, A.; Maruszczyk, J.; Bienfait, B.; Gasteiger, J.; Kleinoeder, T.; Magdziarz, T.; Sacher, O.; Schwab, C. H.; Schwoebel, J.; Terfloth, L.; Arvidson, K.; Richard, A.; Worth, A.; Rathman, J. New Publicly Available Chemical Query Language, CSRML, To Support Chemotype Representations for Application to Data Mining and Modeling. *J. Chem. Inf. Model.* **2015**, *55* (3), 510–528.
- (31) Reisen, F. H.; Schneider, G.; Proschak, E. Reaction-MQL: Line Notation for Functional Transformation. *J. Chem. Inf. Model.* **2009**, *49* (1), 6–12.
- (32) Zhang, Q.-Y.; Aires-de-Sousa, J. Structure-Based Classification of Chemical Reactions without Assignment of Reaction Centers. *J. Chem. Inf. Model.* **2005**, *45* (6), 1775–1783.
- (33) Hoonakker, F.; Lachiche, N.; Varnek, A.; Wagner, A. A Representation to apply usual data mining techniques to chemical reactions - Illustration on the rate constant of SN2 reactions in water. *Int. J. Art. Intel. Tools* **2011**, *20* (02), 253–270.
- (34) Nugmanov, R. I.; Mukhametgaleev, R. N.; Akhmetshin, T.; Gimadiev, T. R.; Afonina, V. A.; Madzhidov, T. I.; Varnek, A. CGRtools: Python Library for Molecule, Reaction, and Condensed Graph of Reaction Processing. *J. Chem. Inf. Model.* **2019**, *59* (6), 2516–2521.
- (35) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting reaction performance in C–N cross-coupling using machine learning. *Science* **2018**, *360* (6385), 186–190.
- (36) Coley, C. W.; Jin, W.; Rogers, L.; Jamison, T. F.; Jaakkola, T. S.; Green, W. H.; Barzilay, R.; Jensen, K. F. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* **2019**, *10* (2), 370–377.
- (37) Tetko, I. V.; Karpov, P.; Van Deursen, R.; Godin, G. State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. *Nat. Commun.* **2020**, *11* (1), 5575.
- (38) 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* (4), 045024.
- (39) Schwaller, P.; Probst, D.; Vaucher, A. C.; Nair, V. H.; Kreutter, D.; Laino, T.; Reymond, J.-L. Mapping the space of chemical reactions using attention-based neural networks. *Nat. Mach. Intel.* **2021**, *3* (2), 144–152.
- (40) Hawizy, L.; Jessop, D. M.; Adams, N.; Murray-Rust, P. ChemicalTagger: A tool for semantic text-mining in chemistry. *J. Cheminf.* **2011**, *3* (1), 17.
- (41) Daniel, L. Chemical Reactions from US Patents (1976–Sep 2016). *Figshare*, 2017. https://figshare.com/articles/dataset/Chemical_reactions_from_US_patents_1976-Sep2016_/5104873.
- (42) Murray-Rust, P.; Rzepa, H. S. CML: Evolution and design. *J. Cheminf.* **2011**, *3* (1), 44.
- (43) 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* (1), 3601.
- (44) Cronin, L. XDL *Open Source Repository*, 2021. <https://gitlab.com/croningroup/chemputer/xdl>.
- (45) Note that the encoding of the procedure without ambiguity does not preclude error in the synthesis as uncertainty associated with measurements is inevitable. Instead the XDL contains the same information included in the prose procedure but with additional parameters such as hotplate stirring rate which are required for accurate reproduction of the code on another compatible system.
- (46) Gibney, E.; Van Noorden, R. Scientists losing data at a rapid rate. *Nature* **2013**, DOI: 10.1038/nature.2013.14416.
- (47) 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* (7601), 73–76.
- (48) Thakkar, A.; Kogej, T.; Reymond, J.-L.; Engkvist, O.; Bjerrum, E. J. Datasets and their influence on the development of computer assisted synthesis planning tools in the pharmaceutical domain. *Chem. Sci.* **2020**, *11* (1), 154–168.

- (49) Open Reaction Database. <https://docs.open-reaction-database.org/en/latest/> (accessed 2021-06-17).
- (50) Pearce, J. M. Return on investment for open source scientific hardware development. *Sci. Pub. Policy* **2016**, *43* (2), 192–195.
- (51) Pearce, J. M. Building Research Equipment with Free, Open-Source Hardware. *Science* **2012**, *337* (6100), 1303–1304.
- (52) Bär, H.; Hochstrasser, R.; Papenfuß, B. SiLA: Basic Standards for Rapid Integration in Laboratory Automation. *J. Lab. Automat.* **2012**, *17* (2), 86–95.
- (53) SiLa SiLa 2 Standard. https://sil2.gitlab.io/sila_base/ (accessed 2021-07-06).
- (54) Pendleton, I. M.; Cattabriga, G.; Li, Z.; Najeeb, M. A.; Friedler, S. A.; Norquist, A. J.; Chan, E. M.; Schrier, J. Experiment Specification, Capture and Laboratory Automation Technology (ESCALATE): a software pipeline for automated chemical experimentation and data management. *MRS Commun.* **2019**, *9* (3), 846–859.
- (55) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3* (1), 160018.
- (56) Jacobsen, A.; de Miranda Azevedo, R.; Juty, N.; Batista, D.; Coles, S.; Cornet, R.; Courtot, M.; Crosas, M.; Dumontier, M.; Evelo, C. T.; Goble, C.; Guizzardi, G.; Hansen, K. K.; Hasnain, A.; Hettne, K.; Heringa, J.; Hooft, R. W. W.; Imming, M.; Jeffery, K. G.; Kaliyaperumal, R.; Kersloot, M. G.; Kirkpatrick, C. R.; Kuhn, T.; Labastida, I.; Magagna, B.; McQuilton, P.; Meyers, N.; Montesanti, A.; van Reisen, M.; Rocca-Serra, P.; Pergl, R.; Sansone, S.-A.; da Silva Santos, L. O. B.; Schneider, J.; Strawn, G.; Thompson, M.; Waagmeester, A.; Weigel, T.; Wilkinson, M. D.; Willighagen, E. L.; Wittenburg, P.; Roos, M.; Mons, B.; Schultes, E. FAIR Principles: Interpretations and Implementation Considerations. *Dat. Intell.* **2020**, *2* (1–2), 10–29.
- (57) 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 programmable chemical synthesis machine. *Nat. Chem.* **2021**, *13* (1), 63–69.
- (58) Kitson, P. J.; Glatzel, S.; Cronin, L. The digital code driven autonomous synthesis of ibuprofen automated in a 3D-printer-based robot. *Beilstein J. Org. Chem.* **2016**, *12*, 2776–2783.
- (59) Cronin, L.; Points, L.; Grizou, J. Robotic chemistry sets for the classroom. *Education in Chemistry*, 2016. <https://edu.rsc.org/feature/robotic-chemistry-sets-for-the-classroom/2000110.article>.
- (60) Peplow, M. Organic synthesis: The robo-chemist. *Nature* **2014**, *512* (7512), 20–22.
- (61) Merrifield, R. B. Automated Synthesis of Peptides. *Science* **1965**, *150* (3693), 178–185.
- (62) Trobe, M.; Burke, M. D. The Molecular Industrial Revolution: Automated Synthesis of Small Molecules. *Angew. Chem., Int. Ed.* **2018**, *57* (16), 4192–4214.
- (63) Alvarado-Urbina, G.; Sathe, G.; Liu, W.; Gillen, M.; Duck, P.; Bender, R.; Ogilvie, K. Automated synthesis of gene fragments. *Science* **1981**, *214* (4518), 270–274.
- (64) Plante, O. J.; Palmacci, E. R.; Seeberger, P. H. Automated Solid-Phase Synthesis of Oligosaccharides. *Science* **2001**, *291* (5508), 1523–1527.
- (65) 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* (36), 10815–10820.
- (66) Ruiz de la Oliva, A.; Sans, V.; Miras, H. N.; Long, D.-L.; Cronin, L. Coding the Assembly of Polyoxotungstates with a Programmable Reaction System. *Inorg. Chem.* **2017**, *56* (9), 5089–5095.
- (67) Miras, H. N.; Cooper, G. J. T.; Long, D.-L.; Bögge, H.; Müller, A.; Streb, C.; Cronin, L. Unveiling the Transient Template in the Self-Assembly of a Molecular Oxide Nanowheel. *Science* **2010**, *327* (5961), 72–74.
- (68) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B.; Baldwin, E.; Weber, I.; Selk, L.; Clawson, L.; Schneider, J.; Kent, S. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **1989**, *245* (4918), 616–621.
- (69) Gibson, D. G.; Glass, J. I.; Lartigue, C.; Noskov, V. N.; Chuang, R. Y.; Algire, M. A.; Benders, G. A.; Montague, M. G.; Ma, L.; Moodie, M. M.; Merryman, C.; Vashee, S.; Krishnakumar, R.; Assad-Garcia, N.; Andrews-Pfannkoch, C.; Denisova, E. A.; Young, L.; Qi, Z. Q.; Segall-Shapiro, T. H.; Calvey, C. H.; Parmar, P. P.; Hutchison, C. A.; Smith, H. O.; Venter, J. C. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **2010**, *329* (5987), 52–56.
- (70) Seeberger, P. H.; Werz, D. B. Automated synthesis of oligosaccharides as a basis for drug discovery. *Nat. Rev. Drug Discovery* **2005**, *4* (9), 751–763.
- (71) Fitzpatrick, D. E.; Ley, S. V. Engineering chemistry for the future of chemical synthesis. *Tetrahedron* **2018**, *74* (25), 3087–3100.
- (72) Dragone, V.; Sans, V.; Henson, A. B.; Granda, J. M.; Cronin, L. An autonomous organic reaction search engine for chemical reactivity. *Nat. Commun.* **2017**, *8* (1), 15733.
- (73) Ley, S. V.; Fitzpatrick, D. E.; Ingham, R. J.; Myers, R. M. Organic Synthesis: March of the Machines. *Angew. Chem., Int. Ed.* **2015**, *54* (11), 3449–3464.
- (74) Wegner, J.; Ceylan, S.; Kirschning, A. Flow Chemistry – A Key Enabling Technology for (Multistep) Organic Synthesis. *Adv. Synth. Catal.* **2012**, *354* (1), 17–57.
- (75) Richmond, C. J.; Miras, H. N.; de la Oliva, A. R.; Zang, H.; Sans, V.; Paramonov, L.; Makatsoris, C.; Inglis, R.; Brechin, E. K.; Long, D.-L.; Cronin, L. A flow-system array for the discovery and scale up of inorganic clusters. *Nat. Chem.* **2012**, *4* (12), 1037–1043.
- (76) Porwol, L.; Henson, A.; Kitson, P. J.; Long, D.-L.; Cronin, L. On the fly multi-modal observation of ligand synthesis and complexation of Cu complexes in flow with 'benchtop' NMR and mass spectrometry. *Inorg. Chem. Front.* **2016**, *3* (7), 919–923.
- (77) Reizman, B. J.; Jensen, K. F. Feedback in Flow for Accelerated Reaction Development. *Acc. Chem. Res.* **2016**, *49* (9), 1786–1796.
- (78) Adamo, A.; Beingsner, 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.; Stelzer, T.; Weeranoppanant, N.; Wong, S. Y.; Zhang, P. On-demand continuous-flow production of pharmaceuticals in a compact, reconfigurable system. *Science* **2016**, *352* (6281), 61–67.
- (79) Chatterjee, S.; Guidi, M.; Seeberger, P. H.; Gilmore, K. Automated radial synthesis of organic molecules. *Nature* **2020**, *579* (7799), 379–384.
- (80) Perera, D.; Tucker, J. W.; Brahmabhatt, S.; Helal, C. J.; Chong, A.; Farrell, W.; Richardson, P.; Sach, N. W. A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science* **2018**, *359* (6374), 429–434.
- (81) Nishiyama, Y.; Fujii, A.; Mori, H. Selective synthesis of azoxybenzenes from nitrobenzenes by visible light irradiation under continuous flow conditions. *React. Chem. Eng.* **2019**, *4* (12), 2055–2059.
- (82) Koo, H.; Kim, H. Y.; Oh, K. (E)-Selective Friedel–Crafts acylation of alkynes to β -chlorovinyl ketones: defying isomerizations in batch reactions by flow chemistry approaches. *Org. Chem. Front.* **2019**, *6* (11), 1868–1872.

- (83) Hatit, M. Z. C.; Reichenbach, L. F.; Tobin, J. M.; Vilela, F.; Burley, G. A.; Watson, A. J. B. A flow platform for degradation-free CuAAC bioconjugation. *Nat. Commun.* **2018**, *9* (1), 4021.
- (84) Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discovery* **2011**, *10* (3), 188–195.
- (85) Selekman, J. A.; Qiu, J.; Tran, K.; Stevens, J.; Rosso, V.; Simmons, E.; Xiao, Y.; Janey, J. High-Throughput Automation in Chemical Process Development. *Annu. Rev. Chem. Biomol. Eng.* **2017**, *8* (1), 525–547.
- (86) Mahjour, B.; Shen, Y.; Cernak, T. Ultrahigh-Throughput Experimentation for Information-Rich Chemical Synthesis. *Acc. Chem. Res.* **2021**, *54* (10), 2337–2346.
- (87) Buitrago Santanilla, A.; Regalado, E. L.; Pereira, T.; Shevlin, M.; Bateman, K.; Campeau, L.-C.; Schneeweis, J.; Berritt, S.; Shi, Z.-C.; Nantermet, P.; Liu, Y.; Helmy, R.; Welch, C. J.; Vachal, P.; Davies, I. W.; Cernak, T.; Dreher, S. D. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science* **2015**, *347* (6217), 49–53.
- (88) 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* (1), 100295.
- (89) Gesmundo, N. J.; Sauvagnat, B.; Curran, P. J.; Richards, M. P.; Andrews, C. L.; Dandliker, P. J.; Cernak, T. Nanoscale synthesis and affinity ranking. *Nature* **2018**, *557* (7704), 228–232.
- (90) Isbrandt, E. S.; Sullivan, R. J.; Newman, S. G. High Throughput Strategies for the Discovery and Optimization of Catalytic Reactions. *Angew. Chem., Int. Ed.* **2019**, *58* (22), 7180–7191.
- (91) 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* (6227), 1221–1226.
- (92) Prabhu, G. R. D.; Yang, T.-H.; Hsu, C.-Y.; Shih, C.-P.; Chang, C.-M.; Liao, P.-H.; Ni, H.-T.; Urban, P. L. Facilitating chemical and biochemical experiments with electronic microcontrollers and single-board computers. *Nat. Protoc.* **2020**, *15* (3), 925–990.
- (93) Cerf, V. G. Avoiding “Bit Rot”: Long-Term Preservation of Digital Information [Point of View]. *Proc. IEEE* **2011**, *99* (6), 915–916.
- (94) Wilson, G. V.; Landau, R. H.; McConnell, S. What should computer scientists teach to physical scientists and engineers? 1. *IEEE Comput. Sci. Eng.* **1996**, *3* (2), 46–65.
- (95) Dryden, M. D. M.; Fobel, R.; Fobel, C.; Wheeler, A. R. Upon the Shoulders of Giants: Open-Source Hardware and Software in Analytical Chemistry. *Anal. Chem.* **2017**, *89* (8), 4330–4338.
- (96) Bitter, R.; Mohiuddin, T.; Nawrocki, M. *LabVIEW: Advanced Programming Techniques*; CRC Press, 2006.
- (97) Roch, L. M.; Häse, F.; Kreisbeck, C.; Tamayo-Mendoza, T.; Yunker, L. P. E.; Hein, J. E.; Aspuru-Guzik, A. ChemOS: Orchestrating autonomous experimentation. *Sci. Robot.* **2018**, *3* (19), eaat5559.
- (98) Roch, L. M.; Häse, F.; Kreisbeck, C.; Tamayo-Mendoza, T.; Yunker, L. P. E.; Hein, J. E.; Aspuru-Guzik, A. ChemOS: An orchestration software to democratize autonomous discovery. *PLoS One* **2020**, *15* (4), e0229862.
- (99) Fabry, D. C.; Sugiono, E.; Rueping, M. Online monitoring and analysis for autonomous continuous flow self-optimizing reactor systems. *React. Chem. Eng.* **2016**, *1* (2), 129–133.
- (100) Carter, C. F.; Lange, H.; Ley, S. V.; Baxendale, I. R.; Wittkamp, B.; Goode, J. G.; Gaunt, N. L. ReactIR Flow Cell: A New Analytical Tool for Continuous Flow Chemical Processing. *Org. Process Res. Dev.* **2010**, *14* (2), 393–404.
- (101) Fath, V.; Kockmann, N.; Otto, J.; Röder, T. Self-optimizing processes and real-time-optimization of organic syntheses in a microreactor system using Nelder–Mead and design of experiments. *React. Chem. Eng.* **2020**, *5* (7), 1281–1299.
- (102) Leadbeater, N. E.; Smith, R. J.; Barnard, T. M. Using in situ Raman monitoring as a tool for rapid optimization and scale-up of microwave-promoted organic synthesis: esterification as an example. *Org. Biomol. Chem.* **2007**, *5* (5), 822–825.
- (103) Meier, M. A. R.; Wouters, D.; Ott, C.; Guillet, P.; Fustin, C.-A.; Gohy, J.-F.; Schubert, U. S. Supramolecular ABA Triblock Copolymers via a Polycondensation Approach: Synthesis, Characterization, and Micelle Formation. *Macromolecules* **2006**, *39* (4), 1569–1576.
- (104) Mohapatra, S.; Hartrampf, N.; Poskus, M.; Loas, A.; Gómez-Bombarelli, R.; Pentelute, B. L. Deep Learning for Prediction and Optimization of Fast-Flow Peptide Synthesis. *ACS Cent. Sci.* **2020**, *6* (12), 2277–2286.
- (105) Zawatzky, K.; Barhate, C. L.; Regalado, E. L.; Mann, B. F.; Marshall, N.; Moore, J. C.; Welch, C. J. Overcoming “speed limits” in high throughput chromatographic analysis. *J. Chromatogr. A* **2017**, *1499*, 211–216.
- (106) Nanita, S. C.; Kaldon, L. G. Emerging flow injection mass spectrometry methods for high-throughput quantitative analysis. *Anal. Bioanal. Chem.* **2016**, *408* (1), 23–33.
- (107) Stoll, D. R.; Carr, P. W. Two-Dimensional Liquid Chromatography: A State of the Art Tutorial. *Anal. Chem.* **2017**, *89* (1), 519–531.
- (108) Naegele, E.; Schneider, S. *Automated Scouting of Stationary and Mobile Phases Using the Agilent 1290 Infinity II Method*; Development Solution, Inc., 2017.
- (109) Mathieson, J. S.; Rosnes, M. H.; Sans, V.; Kitson, P. J.; Cronin, L. Continuous parallel ESI-MS analysis of reactions carried out in a bespoke 3D printed device. *Beilstein J. Nanotechnol.* **2013**, *4*, 285–291.
- (110) 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 optimization of automated flow reactors. *React. Chem. Eng.* **2016**, *1* (1), 96–100.
- (111) 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* (2), 1258–1264.
- (112) Gomez, M. V.; de la Hoz, A. NMR reaction monitoring in flow synthesis. *Beilstein J. Org. Chem.* **2017**, *13*, 285–300.
- (113) Kunjir, S.; Rodriguez-Zubiri, M.; Coeffard, V.; Felpin, F.-X.; Giraudeau, P.; Farjon, J. Merging Gradient-Based Methods to Improve Benchtop NMR Spectroscopy: A New Tool for Flow Reaction Optimization. *ChemPhysChem* **2020**, *21* (20), 2311–2319.
- (114) Berry, D. B. G.; Codina, A.; Clegg, I.; Lyall, C. L.; Lowe, J. P.; Hintermair, U. Insight into catalyst speciation and hydrogen evolution during enantioselective formic acid-driven transfer hydrogenation with bifunctional ruthenium complexes from multi-technique operando reaction monitoring. *Faraday Discuss.* **2019**, *220* (0), 45–57.
- (115) Sagmeister, P.; Poms, J.; Williams, J. D.; Kappe, C. O. Multivariate analysis of inline benchtop NMR data enables rapid optimization of a complex nitration in flow. *React. Chem. Eng.* **2020**, *5* (4), 677–684.
- (116) Prabhu, G. R. D.; Witek, H. A.; Urban, P. L. Telechemistry: monitoring chemical reactions via the cloud using the Particle Photon Wi-Fi module. *React. Chem. Eng.* **2019**, *4* (9), 1616–1622.
- (117) Fitzpatrick, D. E.; Battilocchio, C.; Ley, S. V. Enabling Technologies for the Future of Chemical Synthesis. *ACS Cent. Sci.* **2016**, *2* (3), 131–138.
- (118) Cherkasov, N.; Baldwin, S.; Gibbons, G. J.; Isakov, D. Monitoring Chemistry In Situ with a Smart Stirrer: A Magnetic Stirrer Bar with an Integrated Process Monitoring System. *ACS Sens.* **2020**, *5* (8), 2497–2502.
- (119) Cherkasov, N.; Expósito, A. J.; Bai, Y.; Rebrov, E. V. Counting bubbles: precision process control of gas–liquid reactions in flow with an optical inline sensor. *React. Chem. Eng.* **2019**, *4* (1), 112–121.
- (120) Zepel, T.; Lai, V.; Yunker, L. P. E.; Hein, J. E. Automated Liquid-Level Monitoring and Control using Computer Vision. *ChemRxiv*, 2020. DOI: 10.26434/chemrxiv.12798143.v1.

- (121) Bird, C. L.; Frey, J. G. Chemical information matters: an e-Research perspective on information and data sharing in the chemical sciences. *Chem. Soc. Rev.* **2013**, *42* (16), 6754–6776.
- (122) Hall, S. R.; Allen, F. H.; Brown, I. D. The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1991**, *47* (6), 655–685.
- (123) Kitson, P. J.; Macdonell, A.; Tsuda, S.; Zang, H.; Long, D.-L.; Cronin, L. Bringing Crystal Structures to Reality by Three-Dimensional Printing. *Cryst. Growth Des.* **2014**, *14* (6), 2720–2724.
- (124) Spek, A. L. What makes a crystal structure report valid? *Inorg. Chim. Acta* **2018**, *470*, 232–237.
- (125) Schober, D.; Jacob, D.; Wilson, M.; Cruz, J. A.; Marcu, A.; Grant, J. R.; Moing, A.; Deborde, C.; de Figueiredo, L. F.; Haug, K.; Rocca-Serra, P.; Easton, J.; Ebbels, T. M. D.; Hao, J.; Ludwig, C.; Günther, U. L.; Rosato, A.; Klein, M. S.; Lewis, I. A.; Luchinat, C.; Jones, A. R.; Grauslys, A.; Larralde, M.; Yokochi, M.; Kobayashi, N.; Porzel, A.; Griffin, J. L.; Viant, M. R.; Wishart, D. S.; Steinbeck, C.; Salek, R. M.; Neumann, S. nmrML: A Community Supported Open Data Standard for the Description, Storage, and Exchange of NMR Data. *Anal. Chem.* **2018**, *90* (1), 649–656.
- (126) Bhamber, R. S.; Jankevics, A.; Deutsch, E. W.; Jones, A. R.; Dowsey, A. W. mzMLb: A Future-Proof Raw Mass Spectrometry Data Format Based on Standards-Compliant mzML and Optimized for Speed and Storage Requirements. *J. Proteome Res.* **2021**, *20* (1), 172–183.
- (127) Gasteiger, J.; Hendriks, B. M. P.; Hoefer, P.; Jochum, C.; Somberg, H. JCAMP-CS: A Standard Exchange Format for Chemical Structure Information in Computer-Readable Form. *Appl. Spectrosc.* **1991**, *45* (1), 4–11.
- (128) Schäfer, B. A.; Poetz, D.; Kramer, G. W. Documenting Laboratory Workflows Using the Analytical Information Markup Language. *JALA* **2004**, *9* (6), 375–381.
- (129) 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* (1), 2771.
- (130) Mateos, C.; Nieves-Remacha, M. J.; Rincón, J. A. Automated platforms for reaction self-optimization in flow. *React. Chem. Eng.* **2019**, *4* (9), 1536–1544.
- (131) 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* (5), 769–780.
- (132) MATLAB, ver. 7.10.0 (R2010a); The MathWorks, Inc., 2010.
- (133) Felton, K. C.; Rittig, J. G.; Lapkin, A. A. Summit: Benchmarking Machine Learning Methods for Reaction Optimisation. *Chem. Methods* **2021**, *1* (2), 116–122.
- (134) 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* (3), 035021.
- (135) Houben, C.; Lapkin, A. A. Automatic discovery and optimization of chemical processes. *Curr. Opin. Chem. Eng.* **2015**, *9*, 1–7.
- (136) Wenig, P.; Odermatt, J. OpenChrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data. *BMC Bioinf.* **2010**, *11* (1), 405.
- (137) Helmus, J. J.; Jaroniec, C. P. NmrGlue: an open source Python package for the analysis of multidimensional NMR data. *J. Biomol. NMR* **2013**, *55* (4), 355–367.
- (138) Cobas, C. NMR signal processing, prediction, and structure verification with machine learning techniques. *Magn. Reson. Chem.* **2020**, *58* (6), 512–519.
- (139) Ament, S. E.; Stein, H. S.; Guevarra, D.; Zhou, L.; Haber, J. A.; Boyd, D. A.; Umehara, M.; Gregoire, J. M.; Gomes, C. P. Multi-component background learning automates signal detection for spectroscopic data. *NPJ. Comput. Mater.* **2019**, *5* (1), 77.
- (140) Fine, J.; Rajasekar, A. A.; Jethava, K. P.; Chopra, G. Spectral deep learning for prediction and prospective validation of functional groups. *Chem. Sci.* **2020**, *11* (18), 4618–4630.
- (141) Howarth, A.; Ermanis, K.; Goodman, J. M. DP4-AI automated NMR data analysis: straight from spectrometer to structure. *Chem. Sci.* **2020**, *11* (17), 4351–4359.
- (142) 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* (39), 7076–7080.
- (143) Huyer, W.; Neumaier, A. SNOBFIT—Stable Noisy Optimization by Branch and Fit. *ACM Trans. Math. Softw.* **2008**, *35* (2), No. 9.
- (144) Brochu, E.; Cora, V. M.; Freitas, N. D., A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. *arXiv*, 2010, 1012.2599. <https://arxiv.org/abs/1012.2599>.
- (145) Häse, F.; Roch, L. M.; Kreisbeck, C.; Aspuru-Guzik, A. Phoenix: A Bayesian Optimizer for Chemistry. *ACS Cent. Sci.* **2018**, *4* (9), 1134–1145.
- (146) 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.
- (147) 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* (3), 301–311.
- (148) Häse, F.; Roch, L. M.; Aspuru-Guzik, A., Gryffin: An algorithm for Bayesian optimization for categorical variables informed by physical intuition with applications to chemistry. *arXiv*, 2020, 2003.12127. <https://arxiv.org/abs/2003.12127>.
- (149) Clayton, A. D.; Manson, J. A.; Taylor, C. J.; Chamberlain, T. W.; Taylor, B. A.; Clemens, G.; Bourne, R. A. Algorithms for the self-optimization of chemical reactions. *React. Chem. Eng.* **2019**, *4* (9), 1545–1554.
- (150) Coley, C. W.; Barzilay, R.; Jaakkola, T. S.; Green, W. H.; Jensen, K. F. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **2017**, *3* (5), 434–443.
- (151) Segler, M. H. S.; Waller, M. P. Modelling Chemical Reasoning to Predict and Invent Reactions. *Chem. - Eur. J.* **2017**, *23* (25), 6118–6128.
- (152) Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F. Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **2018**, *4* (11), 1465–1476.
- (153) Zhou, Z.; Li, X.; Zare, R. N. Optimizing Chemical Reactions with Deep Reinforcement Learning. *ACS Cent. Sci.* **2017**, *3* (12), 1337–1344.
- (154) 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* (6453), eaax1566.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on August 31, 2021, without the revised figures for the TOC and Figure 1. The corrected version was reposted on September 2, 2021.