CellPress
REVIEWS

Opinion

Universal Chemical Synthesis and Discovery with 'The Chemputer'

Piotr S. Gromski,[1] Jarosław M. Granda,[1] and Leroy Cronin[1,*]

There is a growing drive in the chemistry community to exploit rapidly growing robotic technologies along with artificial intelligence-based approaches. Applying this to chemistry requires a holistic approach to chemical synthesis design and execution. Here, we outline a universal approach to this problem beginning with an abstract representation of the practice of chemical synthesis that then informs the programming and automation required for its practical realization. Using this foundation to construct closed-loop robotic chemical search engines, we can generate new discoveries that may be verified, optimized, and repeated entirely automatically. These robots can perform chemical reactions and analyses much faster than can be done manually. As such, this leads to a road map whereby molecules can be discovered, optimized, and made on demand from a digital code.

Automation in Chemical Synthesis

Methodologies for the automation of chemical synthesis, optimization, and discovery have not generally been designed for the realities of laboratory-based research, tending instead to focus on engineering solutions to practical problems. We argue that the potential of rapidly developing technologies (e.g., machine learning and robotics) are more fully realized by operating seamlessly with the way that synthetic chemists currently work (Figure 1) [1]. This is because the organic chemist often works by thinking backwards as much as they do forwards when planning a synthetic procedure. To reproduce this fundamental mode of operation, a new universal approach to the automated exploration of chemical space is needed that combines an abstraction of chemical synthesis with robotic hardware and closed-loop programming [2,3]. However, this leads chemists to constantly test the reactions with different synthetic parameters and conditions. The alternative to this problem, as shown in this opinion article, is the development of an approach to universal chemistry using a programming language with automation in combination with machine learning and artificial intelligence (AI).

Chemists already benefit from algorithms in the field of chemometrics and, therefore, automation is one step forward that might help chemists to navigate and search chemical space more quickly, efficiently, and importantly, without bias. Chemometrics is a field that employs a broad range of algorithms to solve chemistry-related problems and has been well established over the past 50 years [4]. Figure 2 presents a standard chemometrics workflow for processing data. The process begins with data that may be of various formats that depend upon the experiment type and/or posed question. The next step is data preprocessing, which covers a variety of procedures depending on the type of data analyzed (e.g., peak detection, input of missing data, and/or normalization). This process is followed by statistical modeling, which is divided into supervised and unsupervised approaches. Probably one of the most well-known unsupervised approaches is **principal component analysis** (see Glossary), which allows summarizing large data sets into several components that capture most of the information. **Support vector machine**, along with **partial least-square discriminant analysis**, are probably the most well-known supervised approaches that allow samples to be classified into distinctive groups based on relevant information. If the produced outcomes are relevant, the next steps incorporate validation to ensure high quality conclusions are formed. Finally, each of the analyses ends with data interpretation. Further details on chemometrics and algorithms that enable exploration of chemical space are found elsewhere [4–7].

In the following paragraphs, we show how chemometrics can be synergistically combined with automation. As we will demonstrate through several examples, the process of automation allows

Highlights

Recent advances in chemical programming enable adoption and universal automation of chemical discovery and synthesis, combined with artificial intelligence, to efficiently perform laboratory tasks, including the closed-loop data exploration for new reactivity.

Robots can perform chemical reactions and analysis much faster than can be done manually, utilizing trial and error, as well as feedback to make autonomous decisions.

Chemists can actively seek out to explore chemical space, aiming for discovery of novel or new molecules and reactions using closed-loop robotic chemical search engines.

[1]WestCHEM, School of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK

*Correspondence: Lee.Cronin@glasgow.ac.uk

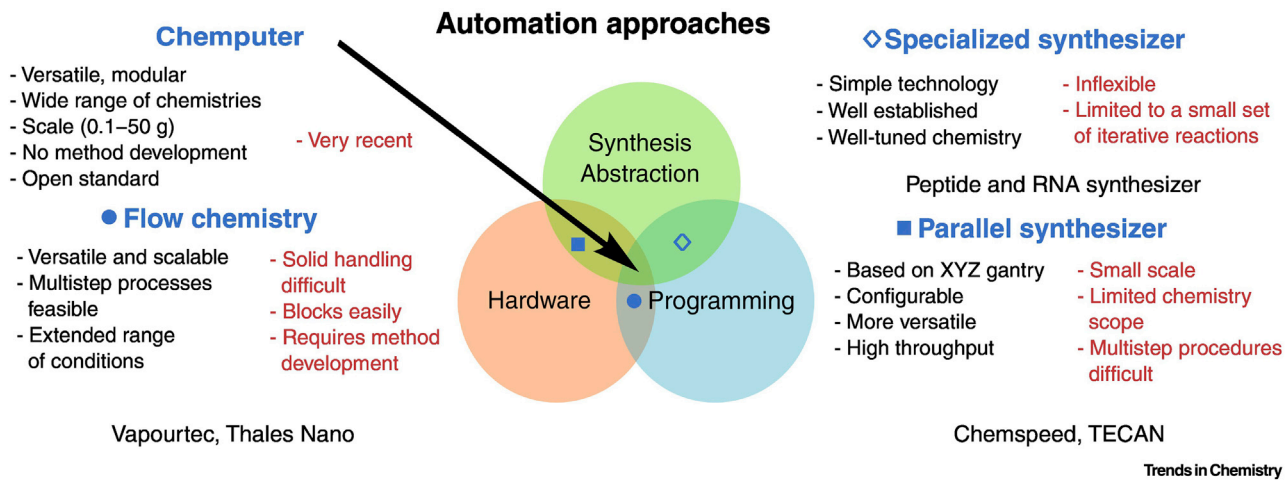


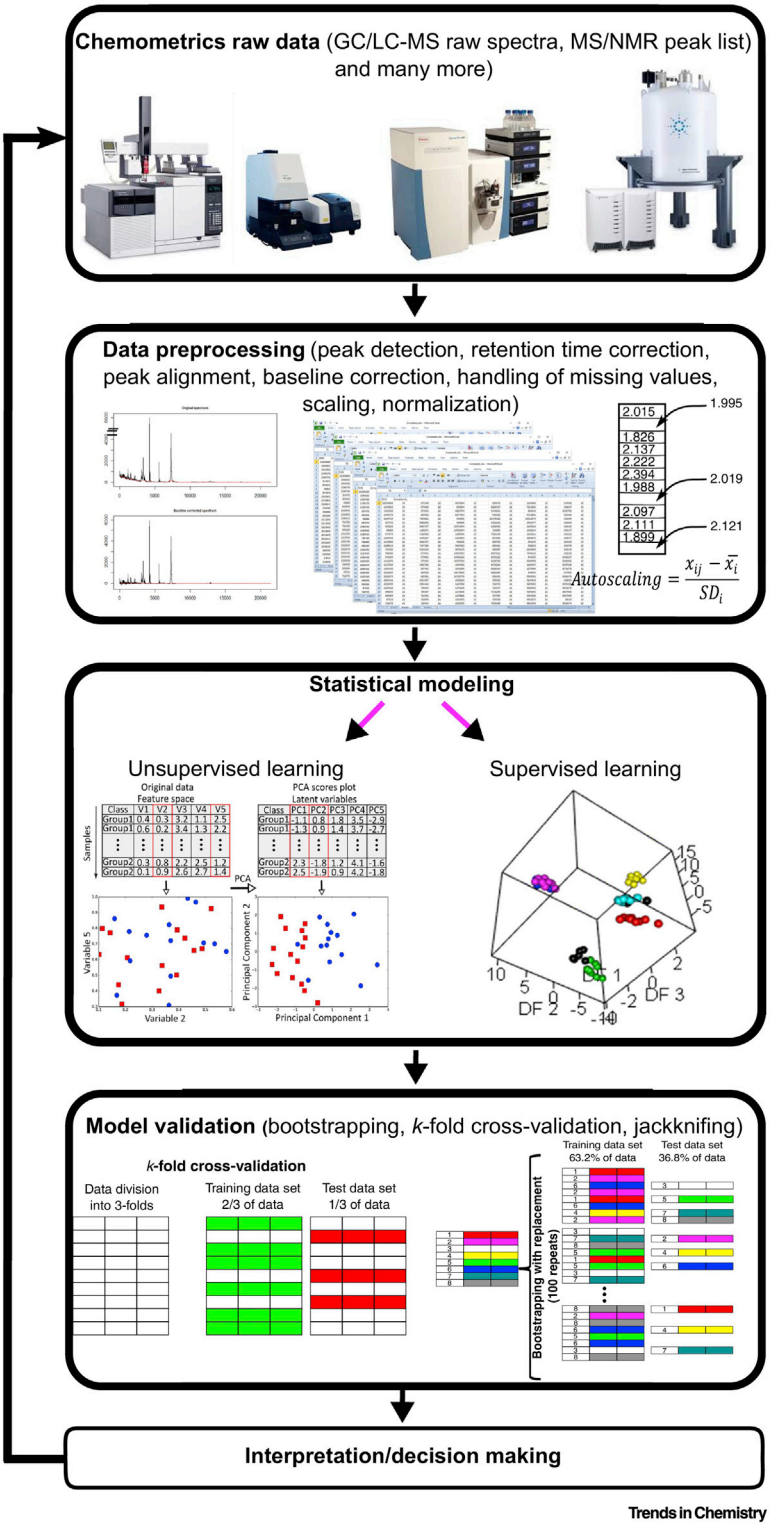
Figure 1. Approaches to Automate Organic Synthesis: Specialized Synthesizer, Parallel Synthesizer, Flow Chemistry, and 'the Chemputer'. The Venn diagram shows the concept of 'the Chemputer' joining together synthesis abstraction, chemical programming, and hardware control. Symbols indicate relative positions in the automation space within the Venn diagram.

for: (i) increased productivity through design of complex experiments that are entirely automated; (ii) increased reliability by reducing human error and increased confidence in the outcomes (i.e., experiments are directly linked with software/algorithms that can produce graphical representation of the results directly); (iii) improved safety when experiments must be performed in a closed environment (e.g., fume hood, glove box, or sealed reactor) equipped with software to indicate when there is a risk or incident; (iv) increased efficiency of processes performed due to increased productivity, reduced waste, and improved outcome quality; and finally (v) valuable walk-away time where the chemist may focus on research [7,8].

Robotics for Automation and Optimization in Chemistry

The ability to make small molecules autonomously and automatically will be fundamental to many applications, including searching for new drugs and materials. So far, automation of small molecule synthesis has relied on a single reaction class limiting its overall universality (e.g., in iterative *N*-methyliminodiacetic (MIDA) boronate synthesis [9] or enzyme-assisted carbohydrate synthesis [10]). Additionally, automated synthesis requires (in many cases) optimization of reaction yields; following optimization, the best conditions can be fed to the synthesis robot to increase the overall yield. There are many approaches to automated yield optimization, some of which are described below. As optimization of reaction conditions requires live feedback from the robotic system, many different detectors have been introduced to monitor progress of the reactions, including benchtop nuclear magnetic resonance spectroscopy [2], infrared spectroscopy [11], mass spectrometry [12], Raman spectroscopy [13], UV-Vis spectroscopy [14], and high-performance liquid chromatography [15]. Harvested data are then fed to optimization algorithms to explore the often multidimensional parameter space. For example, Bédard and colleagues showed an automated-flow system for the optimization of many different types of chemical reactions, including Buchwald-Hartwig amination, Suzuki-Miyaura cross couplings, nucleophilic aromatic substitution (S_NAr), Horner-Wadsworth-Emmons olefination, and photoredox catalysis. The platform could be easily reconfigured to the desired task in a plug-and-play fashion, by attaching different modules (e.g., a photo light-emitting diode or cooled reactor) to the platform core [16].

Robotic approaches also promise to speed up chemical space exploration. To this point, high-throughput experimentation (HTE) appears particularly promising because it can perform thousands of nanomole-scale reactions per day. These HTE approaches could deliver the vast amount of information necessary to train machine learning and AI models, yielding chemical 'big data'. Perera and



Glossary

Curiosity algorithm: developed to replicate curiosity-driven learning in humans that can accurately analyze an unknown and complex chemical system. The knowledge for the search is designed in such a way that the algorithm autonomously chooses the experiments that maximize the number of new and reproducible observations.

Partial least-square discriminant analysis: a supervised approach used for a separation between two or more different groups of samples. The process is achieved through maximization of covariance between the independent variables *X* (matrix readings) and corresponding dependent variables *Y* (sample names/classes).

Phoenics: a probabilistic global optimization algorithm that helps to identify a set of conditions of an experimental or computational procedure, which satisfy desired targets.

Principal component analysis: unsupervised dimensionality reduction techniques that transforms data into a space that allows retention as much of relevant information as possible.

Random forest: a supervised approach that belongs to a group of classification trees; the method is based on a nonlinear approach that allows generation of many decision trees, which organize by using randomly selected input from the original data set.

Support vector machine: a machine learning method that can be applied for both classification and regression tasks. The approach maps data into high-dimensional space in order to identify any differences between processed groups.

(Figure legend at the bottom of the next page)

colleagues demonstrated a flow platform for nanomole screening of Suzuki-Miyaura reactions allowing for screening of greater than 1500 reactions per 24 h [17]. Despite the high-throughput capability, the search of chemical space is not guided by a specific objective. Therefore, many different machine learning algorithms have been developed to explore chemical space.

Machine Learning towards Chemical Space Exploration

Machine learning approaches are fundamental to scientific investigation in many disciplines. In chemistry, many of these methods are well-covered within chemometrics. These methods, linked with chemistry and automation, are rapidly changing the face of chemical research and discovery. Here, we explore how chemometrics and robotics/automation are helping to progress discovery through exploring chemical space and beyond.

Scientists have begun to embrace the power of machine learning coupled with statistically driven design in their research to predict the performance of synthetic reactions. For example, the yield of a Pd-catalyzed Buchwald-Hartwig reaction was predicted using **random forest** in the multidimensional chemical space obtained via HTE [18]. Furthermore, Nielsen and colleagues applied random forest to map the yield landscape of intricate deoxyfluorination with sulfonyl fluoride allowing improved prediction of high-yielding conditions for untested substrates [19]. More recently, **Phoenics** was developed, which combines a concept from Bayesian optimization with ideas from Bayesian kernel density estimation to solve optimization problems and afford efficient exploitation of the search space [20]. Meanwhile, our emphasis is on automation of discovery, which is controlled by robots/computers rather than by humans. Discovery through automation offers far better efficiency and accuracy, as recently shown by Duros and colleagues, where the authors compared human- and robot-based discovery of gigantic polyoxometalates. Specifically, it was shown that algorithm-based search covered approximately nine times more crystallization space than a random search and approximately six times more than human-based discovery. Perhaps even more importantly is that the rate of successful crystallization also increased by ~5% [21]. In addition, the algorithm explored a wider range of space that would need to be performed either by human or purely random search. Recently, the same researchers observed that collaboration between smart robotics and humans may be even more efficient than either alone [22]. Grizou and colleagues described a chemical robotic discovery assistant equipped with a **curiosity algorithm** that can efficiently explore a complex chemical system in search of complex emergent phenomena exhibited by proto-cell droplets [23]. This brings the development of automation, optimization, and discovery very close, a topic widely described in the work by Aspuru-Guzik and Henson, where the authors highlight the fact that self-driven laboratories/robots lead the way forward to fast-track discovery by boosting automated experimentation platforms with machine learning to explore chemical space [7,24].

The automated synthesis could make also use of retrosynthetic analysis for planning the synthesis routes to the target molecules. There are many approaches to automated retrosynthesis, and the most recent one by Segler and colleagues seems to be particularly promising [25]. It used Monte Carlo tree search and symbolic AI to discover retrosynthetic routes. The neural networks were trained on all reactions published in organic chemistry. The system allowed cracking for nearly twice as many molecules, 30 times faster than the traditional computer-aided search method, which is based on extracted rules and hand-designed heuristics. In general, this approach allowed for faster and more efficient retrosynthetic analysis than any other well-known method. Figure 3 shows a workflow for joining

Figure 2. Graphical Representation of a Standard Chemometrics Workflow.
The process begins with chemometrics raw data that could be represented by different inputs depending on the experiment. In the first step, data are preprocessed into well-organized data matrices, followed by statistical modeling that can be solved either by application of unsupervised (e.g., principal component analysis) or supervised (e.g., separation into distinguished groups achieved through discriminant function analysis) models. These processes are usually followed by a validation process (e.g., cross-validation and bootstrapping) that allows assessing validity or accuracy of the process. This is followed by interpretation/decision making that may lead either to experiment modification or final recommendation/decision. Abbreviations: GC, Gas chromatography; LC, liquid chromatography; MS, mass spectrometry.

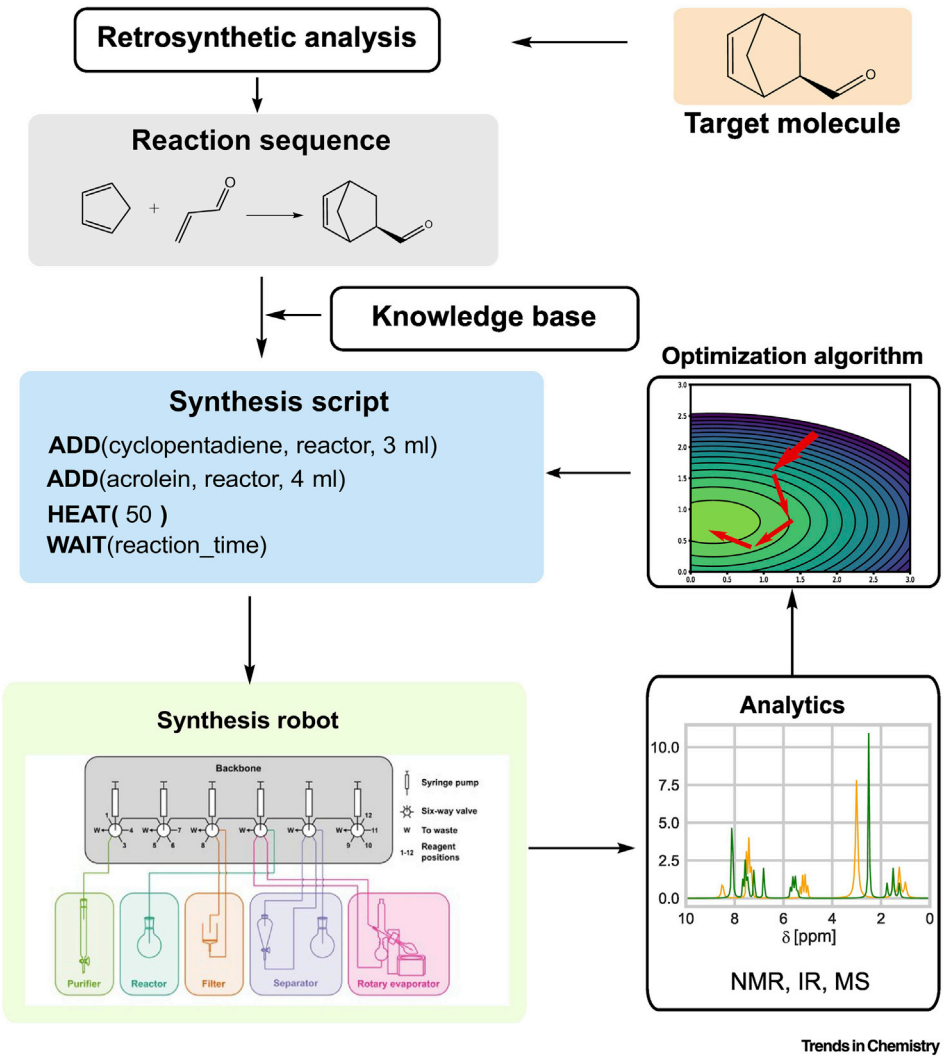


Figure 3. Towards a Universal Chemical Synthesizer: Automated Retrosynthesis, Synthesis, and Optimization.

Reproduced, with permission, from [1]. Abbreviations: NMR, Nuclear magnetic resonance; IR, infrared; MS, mass spectrometry.

automated retrosynthesis with a synthesis robot and reaction optimization. The retrosynthetic module will generate a valid synthesis of the target that can then be transferred into synthesis code that can be executed in a robotic platform. The optimization module can optimize the whole sequence, getting the feedback from the robot.

Chemistry and Discovery via Programmable Modular System: 'The Chemputer'

We recently showed a modular platform for automating batch organic synthesis, which embodies our abstraction in 'the Chemputer' (Figure 4) [1]. Our abstraction of organic synthesis (Figure 4A) contains the key four stages of synthetic protocols: reaction, workup, isolation, and purification, that can be linked to the physical operations of an automated robotic platform. Software control over hardware allowed combination of individual unit operations into multistep organic synthesis. A Chempiler was created to program the platform (Figure 4B); the Chempiler creates low-level instructions for the

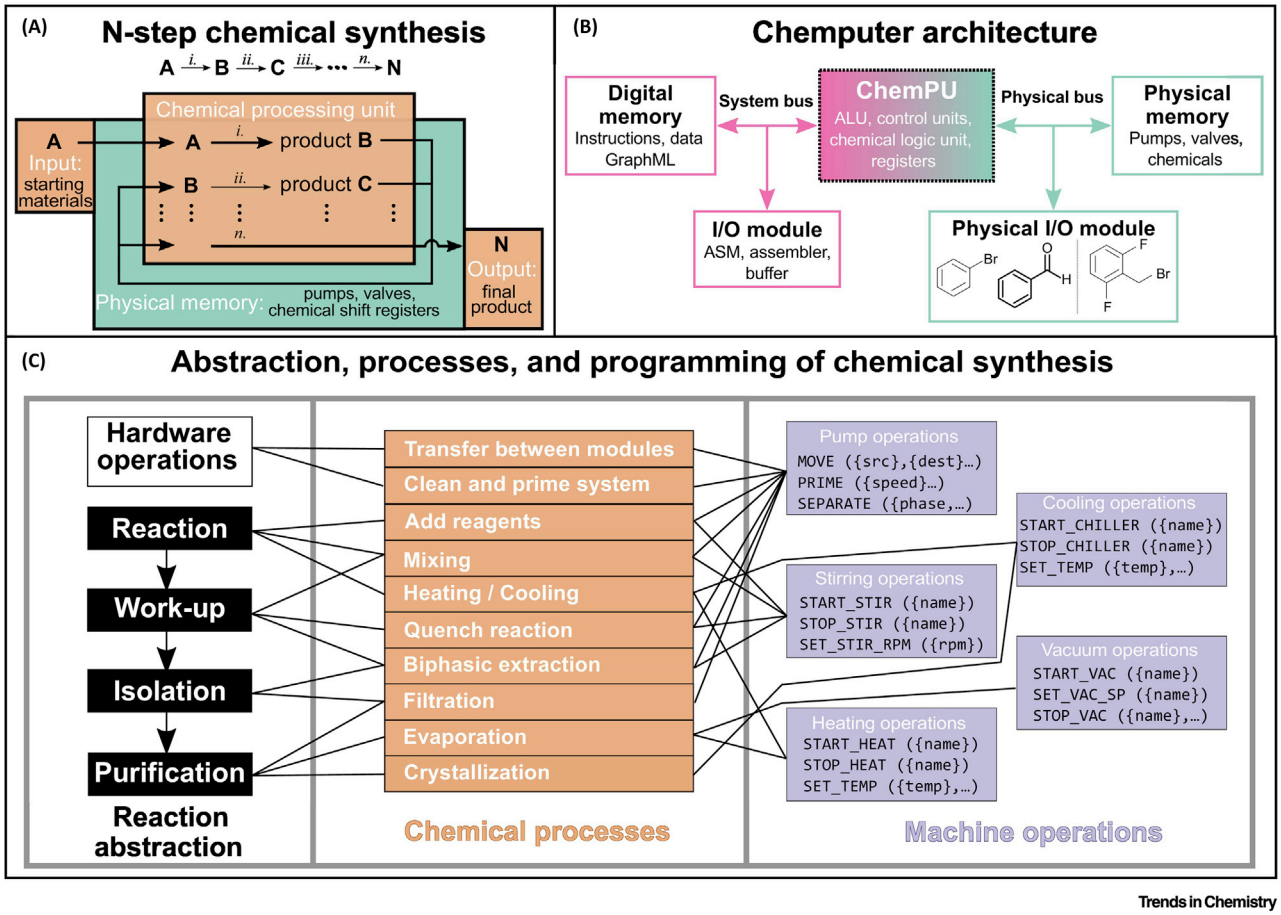


Figure 4. 'Chemputer' Operational Codes.

(A) Iterative representation of organic synthesis treating starting materials as inputs and product as output. (B) Architecture of 'the Chemputer'. (C) Abstraction of organic synthesis (reproduced, with permission, from [1]). Abbreviations: ALU, arithmetic logic unit; ASM, assembly language; I/O, input/output.

hardware taking graph representation of the platform and abstraction representing organic synthesis (Figure 4C). In this way, it is possible to script and run published syntheses without reconfiguration of the platform, providing that necessary modules are present in the system. The synthesis of three small drug molecules was successfully scripted and performed automatically with yields comparable to manual [1].

Finally, by combining robotic systems with AI, it is possible to build autonomous systems working in closed loop, making decisions based on prior experiments. We recently demonstrated a flow system for navigating a network of organic reactions utilizing an infrared spectrometer as the sensor for data feedback. The system was able to select the most reactive starting materials autonomously on the basis of change in the infrared spectra between starting materials and products [11]. Building on that work, we built a robotic platform for autonomous searching of chemical space with three benchtop analytical instruments (infrared spectroscopy, nuclear magnetic resonance spectrometry, and mass spectrometry) for on-line analysis. The search of chemical space is summarized in Figure 5.

The platform operated in a closed loop with a machine learning algorithm; the machine learning algorithm suggested the most promising reactions that were then conducted and analyzed automatically within the platform. The results of each experiment were automatically interpreted and the

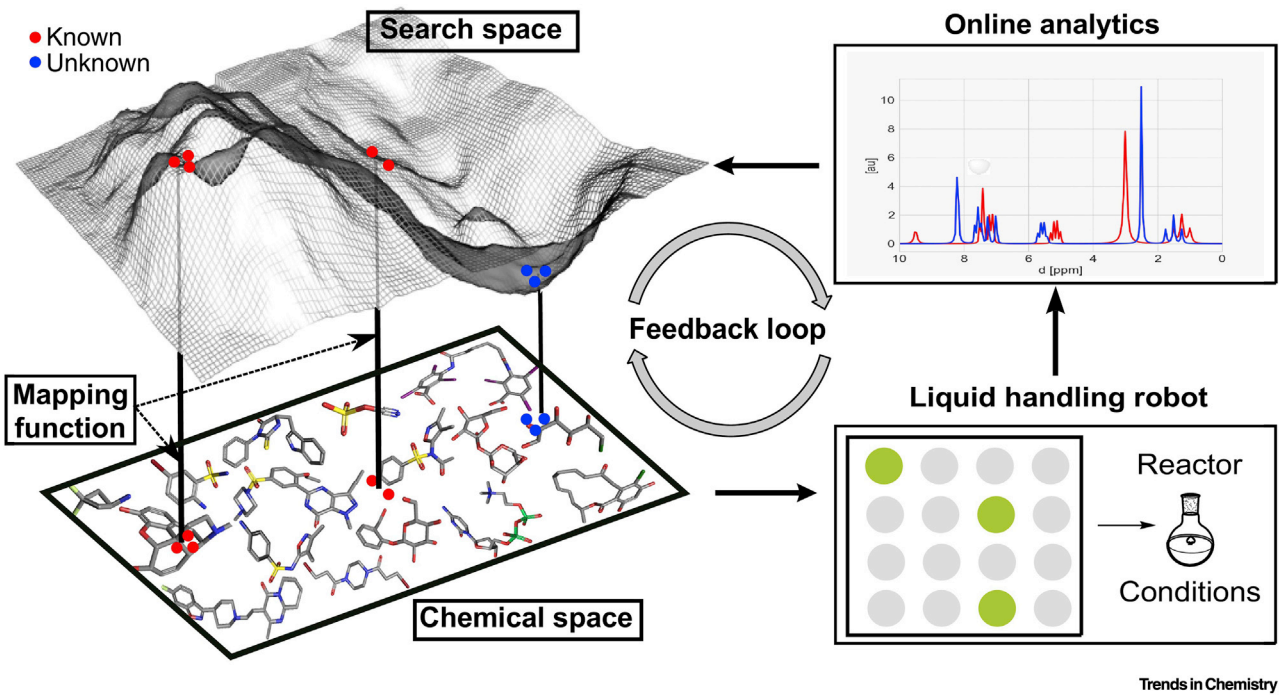


Figure 5. A Closed-Loop Framework for Exploring the Space of Experiments with Machine Learning.
The system explores the space of experiments with feedback from the sensors and then this is mapped onto the chemical space from the analytics that can update the search space (reproduced, with permission, from [8,26]).

data were then used to update the machine learning model. The use of machine learning allowed for autonomous exploration of reaction space allowing for discovery of four new chemical transformations [26]. In another example, Moosavi and colleagues developed a framework for using failed and successful experiments to improve synthesis strategies [27]. This has been achieved through application of automation and machine learning to capture chemical intuition in the synthesis of metal-organic frameworks.

The exploration of chemical space by autonomous robots requires them to assess the novelty of the obtained results [8]. To achieve this, we proposed a framework for assessing novelty and newness of the experimental results (Figure 6). First, the experiment must be repeatable to be valid and exclude experimental and measurement noise. Following confirmation of result repeatability, the next step is to check if this result has a precedent. This can be achieved simply by querying a given database containing knowledge of a given subject. If the search confirms that similar observation has been reported, the experiment can be classified as not new, not contributing added information to our knowledge. However, if the result has not been observed previously, we need to consider if it could be predicted using all the current knowledge. The predictability implies that this result is not novel but new to some extent. Unpredictability implies that result obtained is novel, for example, a reaction mechanism that cannot be predicted can be classified as novel, opening a new branch of research. In the future, this framework will enable automatic assessment of the experimental results by autonomous robots [8].

Concluding Remarks

With the ability to incorporate enhanced hardware and AI/machine learning to carry out many everyday jobs, smart automation now enables the discovery of new molecules and improvements to existing chemical synthesis [1]. In addition, AI/machine learning coupled with 'big data'-

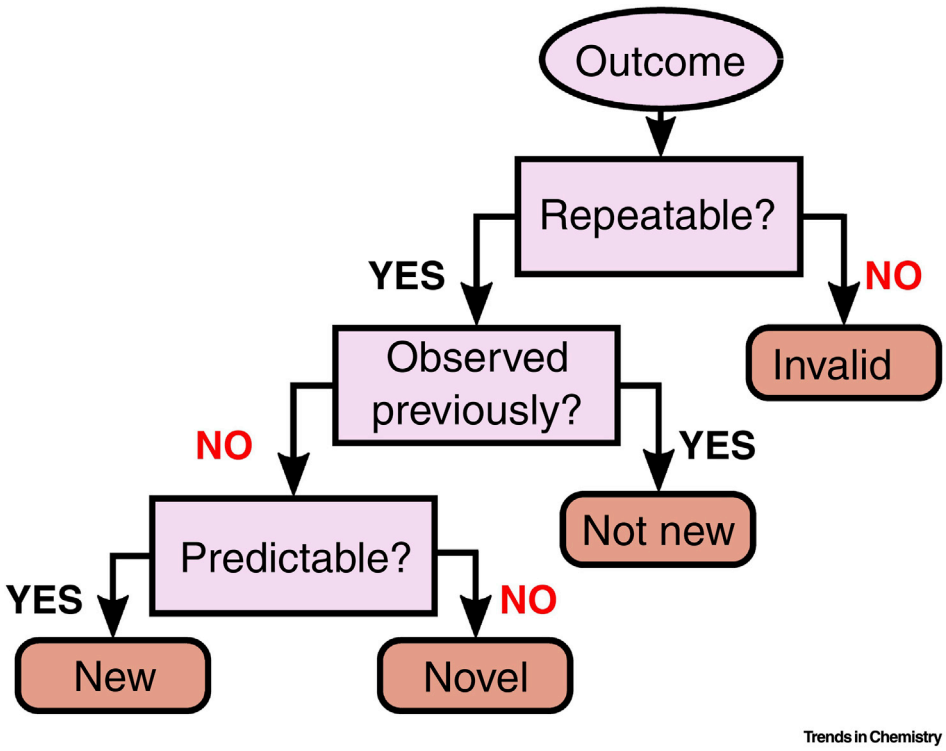


Figure 6. Novelty, Newness, and Validity Diagram.
The process starts with classifying an outcome; if repeatable, further steps are considered (observed previously), else invalid. This repeats until new or novel is defined (reproduced, with permission, from [8]).

generating systems will, and already can in some cases, directly provide outputs for many purposes in the various fields of chemistry. This is because the fundamental nature of AI/machine learning permits the model to be updated and continuously refreshed as new data is produced, leading to more discoveries that cover a larger area of chemical space and eliminating negative confounding factors. In our view, the enthusiasm of the field should now be focusing on the potential of developments for chemical discovery with emphasis on automation coupled with machine learning (see Outstanding Questions), harnessing the powerful capabilities of these approaches shown throughout this opinion article. Here, we have shown how automation and machine learning can improve efficiency and accuracy and therefore are a universal combination for synthesis, optimization, and discovery in the chemistry laboratory.

Outstanding Questions

How can we enable synthetic chemists to operate chemputers without having to know how to code?

How much of current chemistry can be done with the Q12019 chemputer?

Can we drive adoption of the chemputer via development of a new way to write synthesis protocols?

References

1. Steiner, S. *et al.* (2019) Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* 363, eaav2211
2. Sans, V. *et al.* (2015) A self optimizing synthetic organic reactor system using real-time in-line NMR spectroscopy. *Chem. Sci.* 6, 1258–1264
3. Kitson, P.J. *et al.* (2018) Digitization of multistep organic synthesis in reactionware for on-demand pharmaceuticals. *Science* 359, 314–319
4. Brereton, R.G. (2014) A short history of chemometrics: a personal view. *J. Chemom.* 28, 749–760
5. Brereton, R.G. *et al.* (2017) Chemometrics in analytical chemistry–part I: history, experimental design and data analysis tools. *Anal. Bioanal. Chem.* 409, 5891–5899
6. Brereton, R.G. *et al.* (2018) Chemometrics in analytical chemistry–part II: modeling, validation, and applications. *Anal. Bioanal. Chem.* 410, 6691–6704
7. Henson, A.B. *et al.* (2018) Designing algorithms to aid discovery by chemical robots. *ACS Cent. Sci.* 4, 793–804
8. Gromski, P.S. *et al.* (2019) How to explore chemical space using algorithms and automation. *Nat. Rev. Chem.* 3, 119–128
9. Li, J. *et al.* (2015) Synthesis of many different types of organic small molecules using one automated process. *Science* 347, 1221–1226

10. Li, T. *et al.* (2019) An automated platform for the enzyme-mediated assembly of complex oligosaccharides. *Nat. Chem.* 11, 229–236
11. Dragone, V. *et al.* (2017) An autonomous organic reaction search engine for chemical reactivity. *Nat. Commun.* 8, 15733
12. Yunker, L.P.E. *et al.* (2014) Practical approaches to the ESI-MS analysis of catalytic reactions. *J. Mass Spectrom.* 49, 1–8
13. Svensson, O. *et al.* (1999) Reaction monitoring using Raman spectroscopy and chemometrics. *Chemom. Intell. Lab. Syst.* 49, 49–66
14. Yue, J. *et al.* (2013) Microreactors with integrated UV/Vis spectroscopic detection for online process analysis under segmented flow. *Lab Chip* 13, 4855–4863
15. Malig, T.C. *et al.* (2017) Real-time HPLC-MS reaction progress monitoring using an automated analytical platform. *React. Chem. Eng.* 2, 309–314
16. Bédard, A.-C. *et al.* (2018) Reconfigurable system for automated optimization of diverse chemical reactions. *Science* 361, 1220–1225
17. Perera, D. *et al.* (2018) A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science* 359, 429–434
18. Ahneman, D.T. *et al.* (2018) Predicting reaction performance in C-N cross-coupling using machine learning. *Science* 360, 186–190
19. Nielsen, M.K. *et al.* (2018) Deoxyfluorination with sulfonyl fluorides: navigating reaction space with machine learning. *J. Am. Chem. Soc.* 140, 5004–5008
20. Hase, F. *et al.* (2018) Phoenics: a Bayesian optimizer for chemistry. *ACS Cent. Sci.* 4, 1134–1145
21. Duros, V. *et al.* (2017) Human versus robots in the discovery and crystallization of gigantic polyoxometalates. *Angew. Chem.* 56, 10815–10820
22. Duros, V. *et al.* (2019) Intuition-enabled machine learning beats the competition when joint human-robot teams perform inorganic chemical experiments. *J. Chem. Inf. Model.* 59, 2664–2671
23. Grizou, J. *et al.* (2018) A closed loop discovery robot driven by a curiosity algorithm discovers proto-cells that show complex and emergent behaviours. *ChemRxiv*. Published online 13 February, 2019. https://doi.org/10.26434/chemrxiv.6958334.v1
24. Häse, F. *et al.* (2019) Next-generation experimentation with self-driving laboratories. *Trends Chem.* 1, 282–291
25. Segler, M.H.S. *et al.* (2018) Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* 555, 604–610
26. Granda, J.M. *et al.* (2018) Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* 559, 377–381
27. Moosavi, S.M. *et al.* (2019) Capturing chemical intuition in synthesis of metal-organic frameworks. *Nat. Commun.* 10, 539