Modern Automation in Organic Synthesis Laboratories

Pascal Miéville[a] **and Florian de Nanteuil**[b], [a] Swiss CAT+ West Hub, EPFL SB ISIC SWISSCAT, Lausanne, Switzerland; and [b] DSM-Firmenich, Geneva, Switzerland

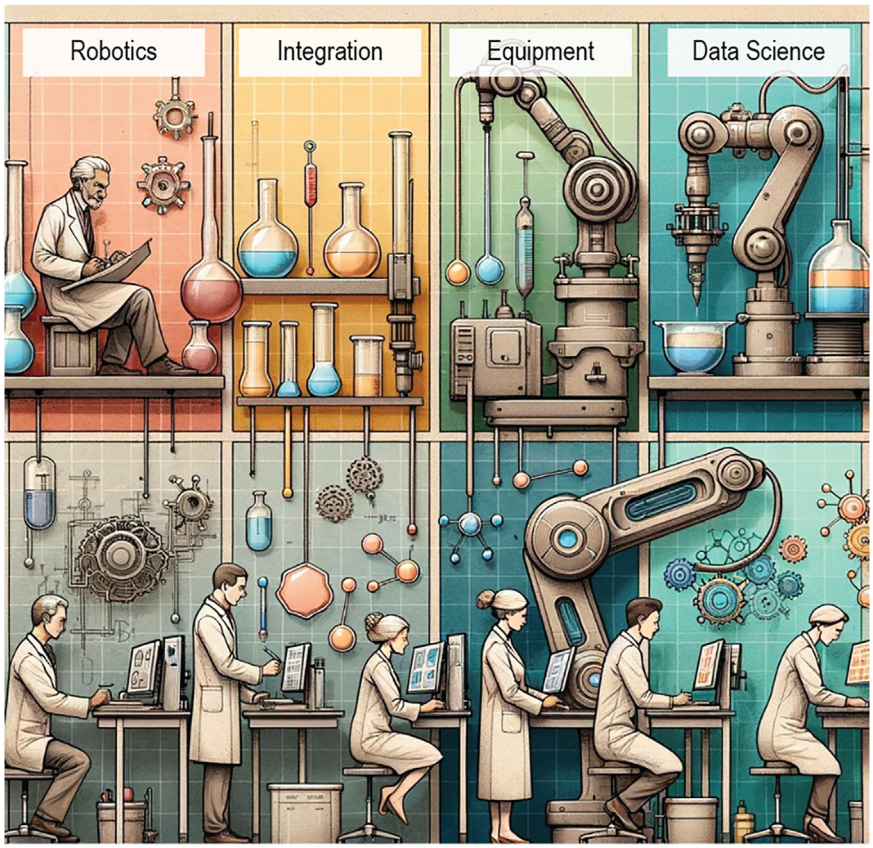
Key Points

- Evolution of automation in experimental sciences and chemistry laboratories
- State of the art of automation in organic synthesis and analysis
- Discussion of the strategies for chaining, integration and orchestration of samples and data

Abstract

The exponential growth of automation within organic chemistry has unlocked many opportunities for scientists worldwide. The chapter aims to provide a comprehensive understanding of the transformative potential and the complex landscape created by the evolution of automation in organic chemistry. It presents an overview dedicated to exploring the global evolution of automation in experimental chemistry, focusing on its pivotal concepts, cutting-edge technologies, and diverse applications that drive this new branch of experimental sciences. The integration of robotics, automation, global data management, and machine learning algorithms within the domain of organic synthesis signifies a transformative leap in laboratory practices. New tools harmonize to execute an extensive range of synthetic tasks, promising reproducibility, precision, and scalability in handling compounds and performing reactions. This approach optimizes molecule production, resulting in accelerated discovery processes, heightened efficiency, and a drastic reduction in errors. The chapter dives into various components underpinning this era of organic synthesis automation, encompassing the historical progression, current state-of-the-art in instrumentation, experimental coding, data processing automation, and the role of artificial intelligence in predictive modeling of reaction outcomes. Moreover, it explores the persistent challenges posed by these new technologies, addressing issues such as solid sampling, global laboratory integration, and examines implications touching on ethical considerations, safety concerns, and sustainability aspects.

Graphical abstract



Introduction

The recent growth of automation in organic chemistry is unleashing new potential for scientists around the world, offering a vision of unprecedented possibilities.[1–3] This chapter will introduce you to the global evolution of automation in experimental sciences, the key concepts, technologies, and applications that form the backbone of this modern revolution, and explore the latest instrumental developments and challenges at the heart of automated organic synthesis.

The automation of organic synthesis signifies the integration of robotics and automation, global data management, and machine learning algorithms into the chemical laboratory.[4] These tools, synergistically combined, can execute a wide array of synthetic tasks, offering reproducibility, precision, and the ability to handle vast libraries of compounds. The result is a process of molecule production that's faster, more efficient, and significantly less prone to error.[2,5] This contemporary approach to organic

synthesis has been a major driving force behind recent advancements in numerous areas of research and industry,[6–8] from pharmaceuticals[9–11] to materials science.[12–14]

In this chapter, we will explore several components that underpin this modern era of organic synthesis automation. We'll discuss the history of automation in experimental sciences and more specifically in chemistry, the development and current state of the art in automated synthetic and analytical chemistry instrumentation, the role of experimental coding and laboratory and data processing automation, and the use of artificial intelligence in predictive modeling of reaction outcomes. In addition, we will explore the real-world challenges still posed by these new technologies, considering the problems of solid sampling and global laboratory integration. We will also briefly discuss some implications such as ethics, safety concerns, and sustainability aspects.

Evolution of Automation and High-Throughput Methods in Experimental Sciences

The inception of automation in experimental sciences marks a significant turning point in the annals of scientific history. It is a convergence of disciplines, intertwining technology with the essence of empirical inquiry, that has transformed traditional research methods and propelled science to new frontiers. In this section, we will trace back the origins of this revolutionary shift, to understand how automation began to reshape the landscape of experimental sciences.

The advent of automation in experimental science can be traced back to the mid-20th century. The 1950s and 1960s saw the advent of computers[15–17] and early forms of digital technology.[18] Laboratories began to use these tools for rudimentary tasks such as data acquisition and basic calculations, then progressively for more complex tasks such as FFT,[19,20] sowing the seeds for future automation. However, as briefly mentioned above, laboratory automation is a large combination of different tools and methods, including robotics and automation, global data management, and machine learning algorithms. So it wasn't until the mid-1970s and 1990s that we saw the birth of true automation in the experimental sciences.[21,22]

In fact, although it was started in the 1950s with Elmer and Elise in the UK[23] and in the 1960s with SHAKEY at Stanford University[24] in dedicated laboratories, the development of a sufficient robotic system that could independently perform simple tasks in laboratories was only achieved in the mid 70s with a first example of automated remote sensing for agricultural development[25] and has been continuously expanded until now. These rudimentary robotic systems were the precursors of today's high-tech automated laboratory equipment, paving the way for the automation of more complex and demanding tasks.

The subsequent decades saw a flurry of advancements as the pace of technological evolution accelerated. Machine learning and artificial intelligence began to emerge as powerful tools in the late 1990s and early 2000s, facilitating the development of algorithms capable of more complex analysis and prediction.[26–28] This opened the door to automation of not just the physical tasks in the lab, but the intellectual tasks as well such as pattern recognition[29] or natural language understanding.[30] Software tools, integrated with powerful hardware, could now undertake experimental design,[31] data analysis,[32] and even hypothesis generation,[33] pushing the envelope of what was possible. At the same time, advancements in robotics technology led to the creation of automated systems capable of performing complex experimental procedures with high precision and consistency, revolutionizing fields like genomics, proteomics, and drug discovery.

Today, automation is deeply embedded in the experimental sciences, providing an array of tools that extend our capabilities far beyond the wildest dreams of our predecessors. We can now automate a wide range of different types of experimental work, from simple repetitive tasks to the design and execution of complex experiments, the analysis of massive data sets, and the interpretation of results. This not only increases efficiency, but also allows scientists to focus on higher-level problems, improving the overall quality of scientific research.

In this respect, physics, biology, drug discovery, and materials research are the sciences where automation has really taken off in recent decades. For a variety of reasons that we will briefly discuss below, these fields have benefited greatly from automation. In fact, some of them, such as genetic sequencing and metabolomics, would probably not have seen the light of day without automation. The current rapid development of automation in chemistry is largely based on existing strategies and tools developed in the above-mentioned fields. For this reason, we will briefly review some of their milestones and key tools in the following sections.

In Physics

Automation has seeped into all corners of physics, beginning with automated data collection[34,35] and remote sensing technologies.[36] Astronomers were among the first to automate telescopes[37] for sky surveys, capturing and processing vast amounts of data unaided, while geophysicists used remote sensing technologies to map the Earth's surface and interior.[38] These automated systems could operate round the clock, greatly enhancing the volume and variety of data collected.

Over time, physics experiments have grown increasingly complex, producing torrents of data that require sophisticated processing and analysis methods. Automation stepped in to meet this challenge, with advanced algorithms such as Fast Fourier Transform (FFT)[39] or Runge-Kutta[40] capable of processing data and signals at unprecedented scales and speeds or to study evolution of systems with methods such as Monte-Carlo.[41] Machine learning and artificial intelligence now play a crucial role, assisting in pattern recognition,[29] anomaly detection,[42] predictive modeling,[43] automated physical law analysis,[33] and data-driven developments,[44,45] making sense of the immense datasets that modern physics experiments generate.

Automation has also been instrumental in the realm of accelerators and particle physics experiments. The Large Hadron Collider (LHC) at CERN, for example, relies heavily on automated systems. These systems control the intricate processes involved in accelerating particles to near light speed and managing the enormous amount of data produced by particle collisions. Without automation, monitoring and controlling the countless parameters involved in these experiments would be practically impossible.

In Biology

In the landscape of biological research, the advent and evolution of automation has revolutionized our understanding of life in profound ways. From the cellular level to entire ecosystems, automation has become an indispensable tool driving breakthroughs across multiple disciplines.[46,47] This essential evolution was made necessary and possible by the very nature of biology. Indeed, given the complexity of the systems studied, biology relies heavily on statistics and thus on a large number of experimental replicates to generate significant results.[48] This need for repetition has been a strong motivation to develop automation, hence the necessity. Regarding the possibility aspect, in contrast to chemistry where a considerable variety of chemicals have to be handled,[49] in biology a large majority of systems, in complete logic with life conditions, are contained in aqueous solutions. Aqueous solutions have two major advantages in terms of automated handling. First, they are liquid at room temperature and therefore relatively homogeneous, and second, they are made with water as a solvent. Pure water has a high surface tension of interface with air in comparison to a majority of organic solvents (71.99 ± 0.05 mN m^{-1} at 25 °C[50]) and a relatively high boiling point (373.15 K at 101.325 kPa[51]) The first makes them easy to pipette and the second ensures that the solutions do not dry out during processes. This combination of elements makes biology an essential source for the development of automation, especially in the field of automated liquid handling with the emergence of robotized systems. These systems allow precise manipulation of fluids, a critical aspect in most biological experiments, thereby limiting variability and increasing reproducibility. They can pipette, dilute, and mix samples, facilitating numerous procedures such as PCR setup, cell culture maintenance, and plate replication. Another important development in automation in biology has been the rapid analysis of results using fluorescence or optical density methods.

The journey of automation in biology began with the advent of automated sequencing technologies in the late twentieth century. The first automated DNA sequencer, developed by Leroy Hood in the 1980s,[52] significantly accelerated the sequencing process and laid the groundwork for the Human Genome Project (HGP).[53] This ambitious endeavor, completed in 2003, was a monumental achievement made possible by automation. With the human genome sequenced, we entered the era of genomics,[54] where the emphasis shifted from single genes to the study of all genes and their interrelationships. It is interesting to note that, with the help of automation and the development of next-generation sequencing technologies (NGS),[55] the cost for a complete human genome sequencing dropped from ca. 300 M USD with the HGP to less than 1000 USD in 2016.[56]

Beyond genomics, the story of automation continues in other "omics" disciplines. In proteomics, the study of the complete set of proteins expressed by a genome, automation has enabled high-throughput protein analysis.[57] Automated liquid chromatography and mass spectrometry systems can analyze complex protein samples, revealing insights into protein structures, interactions, and functions. Similarly, in metabolomics,[58] the study of the complete set of small molecules or metabolites within a biological sample, automated systems can analyze and catalog thousands of metabolites, aiding in the understanding of metabolic pathways and disease mechanisms.

The integration of automation across the different "omics" disciplines is giving rise to systems biology,[59,60] an approach that seeks to understand biological systems in their entirety. Here, automated experimental systems generate vast amounts of data that feed into computational models, iteratively improving them. This approach aims to capture the complexity of biological systems, paving the way for breakthroughs in personalized medicine, synthetic biology, and more. Systems biology is in fact a part of the larger bioinformatics[61] development. A field born out of the need to make sense of this data deluge, heavily relies on automated algorithms for data processing, pattern recognition, and predictive modeling. Machine learning and artificial intelligence have proven to be invaluable tools in this regard, aiding in tasks ranging from gene annotation to the prediction of protein structures and functions.

In Drug Discovery

The evolution of automation has been pivotal to advances in high-throughput screening (HTS) within the field of drug discovery. Early screening methods were laborious and time-consuming, often limited to evaluating a handful of compounds per day. However, the advent of automation in the 1980s and 1990s transformed this process. With the integration of robotics and liquid handling systems, it became possible to systematically screen large libraries of chemical compounds, testing each for their biological activity against specific cellular, biochemical, or genetic targets.[62,63] The adoption of microplate technologies further streamlined this process, allowing for miniaturization and parallelization of assays, thereby significantly reducing cost and resources,[64] With the help of automated systems, HTS platforms can now rapidly evaluate up to millions of compounds, accelerating the early stages of drug discovery.[65] The integration of artificial intelligence and machine learning with automated HTS further enhances this process, providing improved data analysis and predictive models.[10,11,66–68] Consequently, automation has not only enhanced the speed and efficiency of drug discovery but has also opened new avenues for innovation and advancement in the field.

In Material Sciences

The advancement of automation technologies also has had a transformative impact on high-throughput screening (HTS) methods for new material formulation. This area, traditionally dependent on labor-intensive and time-consuming trial-and-error approaches, has been revolutionized by the integration of automated systems. Robots and automated liquid handling systems now perform systematic and efficient synthesis and testing of material libraries, dramatically accelerating the pace of discovery.[69] The introduction of microplate technologies has allowed for miniaturization and parallelization, which has significantly reduced costs and resources.[70] The use of machine learning algorithms such as Bayesian optimization, integrated with automation, has further enhanced HTS processes, helping to predict and interpret complex datasets, thus guiding the discovery and optimization

of novel materials.[71,72] Consequently, the evolution of automation in material HTS is not only accelerating the speed of new material formulation but is also opening up novel avenues for innovation in material science.

Evolution of Automation and High-Throughput Methods in Chemistry

The initial foray into automation in chemistry was driven by the need for high precision and reproducibility in experimental processes. The first automated systems were designed for simple tasks such as titration and colorimetry in the domain of clinical chemistry in the mid-20th century.[73] By reducing human error and enhancing reproducibility, these early systems significantly improved the quality of chemical analyses. In 1966, an automate for peptide synthesis was developed by Robert Merrifield and John Stewart[74] opening the way for automated synthesis.

Over time, automated systems became more sophisticated, particularly in the field of analytical chemistry. Automated spectrometers and chromatographs became critical tools for the precise and efficient identification and quantification of chemical substances.[75,76] These systems not only improved the accuracy and speed of chemical analysis, but also made it possible to perform complex analyses on a routine basis. With the combined development of automated analytical methods, liquid handling and dispensing systems (see Section **Evolution of Automation and High-Throughput Methods in Experimental Sciences**), the development of the first automated synthesizers, and the beginnings of artificial intelligence applied to chemistry, such as the LHASA software,[77] automation began to transform the field of chemical synthesis, including both organic and inorganic chemistry. At the end of the 1990s, the main issues in automation were to solve the problems of phase separation and to increase the level of system integration.[78.] For example, this period saw the development of several data standard formats such as JCAMP-DX for spectroscopy[79] or PDB for proteins.[80] At the same time, there was an intense discussion about the best automation strategy among flow, batch, and combined modes (see **Section Automation Strategie**).

The 2000s and 2010s saw two distinct developments in chemistry. First, at the laboratory scale, automation made significant contributions to the fields of high-throughput screening and combinatorial chemistry. Through the use of automated platforms, it became possible to synthesize and test huge libraries of chemical compounds, dramatically accelerating the pace of drug discovery and materials science.[81] And the second set of developments was more industrial in nature, with the introduction of more and more automation (PID controllers, feedback loops, automated cell handling, etc …).[82]

During the same period and in the following years, the idea of general synthesis systems that could assemble a wide variety of molecules on demand emerged. Trobe and Burke compare these general systems to 3D printers for chemistry.1 The idea of generality was also present in the development of a general language for programming chemistry on Cronin's Chemputer,[83] Peplow's robot chemist[84] and latter the Cooper's,[85] and the graph analysis of laboratory automation workflows.[4] Following the development initiated in the 1990s about global system integration and support the idea of general synthesis systems, the 2010s and 2020s also experiment an important development of standardization with the apparition of general data standards such as AniML[86] and Allotrope,[87,88] and general instrument connection framework such as SiLa[89,90] and OPC-UA with the recent LADS—Laboratory and Analytical Device Standard[91] codex dedicated to laboratory instruments.

In recent years, machine learning and artificial intelligence have begun to play a key role in the automation of chemical research. These technologies are now being used to predict reaction outcomes,[92] design new synthetic routes,[93] and optimize experimental conditions,[94] paving the way for a new era in automated chemistry.

In the next section (**State of the art**), we will discuss in more detail the current state of development of automated methods in chemistry and, more specifically, in organic synthesis.

Automation Strategies

In this section, we will define some semantic and strategic elements that we consider important to clearly understand what an automated organic synthesis laboratory is and its inherent structure and limitations.

Intra Versus Inter Instrument Automation

The first distinction we want to make is between intra instrument automation (autosamplers, autocollectors, automated synthesis platforms, etc.) and inter instrument automation (sample transfer, global management of consumables, global management of chemical stocks, global management of waste, etc.).

The first is already well developed. Many instruments such as chromatographs and spectrometers (LC, GC, NMR), automated liquid handlers and synthesis platforms are equipped with or based on such intra instrument automation tools. These automated tools are generally well adapted for human use and are therefore not necessarily well designed for automated interfacing.

Inter instrument automation (global laboratory automation) is much less developed. This is undoubtedly due to a number of factors, such as the number of players involved, the variety of problems to be solved and the need to plan global automation from the outset of laboratory development.

However, some of the elements of global automation are currently undergoing an interesting evolution. The democratization of collaborative robotic arms (6-axis, scara) and open source micro-controllers such as Arduinos© are making automation more and more accessible to the chemical community. On the communication side, between the different instruments and the laboratory database and controllers, we can mention the development of some communication standards such as OPC-UA,[91] SiLa,[89,90] and Allotrope[88] for the data formats. The issue of sample transfer will be discussed in more detail in Section **Sample Transfer**.

Automated Versus Autonomous Versus Data-Driven Laboratory

We like to distinguish between an automated laboratory and an autonomous laboratory. The former is equipped with intra automated equipment and may be fully networked. This level of automation can be supported by a centralized information system, such as an Electronic Laboratory Notebook (ELN),[95] which is used to record process data and report a selection of results in a digital way. The automated laboratory can also be linked to a Laboratory Information Management System (LIMS)[96] where sample data is automatically collected and stored using bare-code readers and, to some extent, a direct interface between the instrument and the LIMS database. ELN and LIMS can work together to form a global data management environment. The combination of the ELN and LIMS define the sample and data ontology[97] within the laboratory.

The Autonomous Laboratory must also be fully automated, but the difference with the Automated Laboratory is that an Autonomous Laboratory should be able to run for a period of time (campaigns or batches) without human intervention. This means that in addition to complete intra automation of all equipment, a complete set of inter-automation is required, including sample transfer capabilities, laboratory logistics (chemicals, consumables, waste) and operation scheduling. The latter requires a complete connection of all instruments to the scheduler through dedicated APIs allowing in principle for a automated and direct data capture into a database. Therefore, data management and data ontology must be adapted to the direct scheduler-database approach.

In terms of the data-driven laboratory, potentially any laboratory can be data-driven. It does not necessarily have to be automated or autonomous. We can imagine a completely manual laboratory where all data is correctly and systematically entered into a database, and where optimization[98] or reaction generation[92,99] algorithms are used to generate new proposals based on the previously collected data, which are then tested by humans and collected again. Anyway, this manual data-driven laboratory would certainly present some limitations. First one is that to make interesting prediction, large sets of data are frequently required to train models. This require a certain throughput to be able to generates theses larges data sets in a reasonable amount of time. Second, data robustness and repeatability are essential in order to ensure a high quality in the datasets. Repeatability is undoubtedly a great automation strength. Finally, automation induces an important degree of reactivity to test rapidly new proposal generated by the algorithms. Therefore, it is interesting to couple intimately the data-driven approach with the highest degree of autonomy. A completely data-driven lab should ideally also be an autonomous laboratory where complete campaigns can be programmed with precise goals.

Automation in Organic Synthesis, Research Versus Production Approaches

Automation is already common at the production and pilot scale.[82] It is much less common at the research laboratory scale, although it is evolving rapidly, as evidenced by the ever-increasing number of articles devoted to this topic.[1–3,5,100] However, it is important to distinguish between chemical automation strategies in these two areas. In manufacturing, automation is adapted to a specifically optimized process. There is a long development process in which laboratory protocols are filtered and optimized to make them compatible with existing industrial unit operations.[101] During the development step, only the appropriate reactions are conserved and selected for further steps. In parallel, the automation line can be designed to meet the needs of a specific optimized protocol. In a research laboratory, where the goal is to discover new unknown molecules or to optimize reactions at an early stage, it is not possible to go through the development filter. In order to maintain the highest degree of flexibility and to handle the greatest number of different chemical sequences, the automation must be designed from the outset to be as generic as possible and thus as close as possible to the capabilities of a "manual" research laboratory.

It is probably not possible to maintain exactly the same degree of flexibility in an automated laboratory as in a manual laboratory, and some project selection must be made to ensure the best chances of getting results, and some specific conditions in a batch may not give results. But it is essential to design automated laboratories with the goal of mimicking human flexibility as closely as possible. It is also interesting to note that traditional laboratories are generally dedicated to a specific type of preparative chemistry.

This is not necessarily the best approach when designing an automated laboratory because of the high cost of building. It is therefore interesting to try to maximize the number of different types of chemistry that can be performed on the automated research line, either directly or with the least amount of modification.

In order to ensure a high degree of flexibility and also to consider a wide range of possible chemistries, a somewhat similar approach to industrial development must be carried out during an automated laboratory design. This analysis consists in converting as many known protocols as possible of the targeted field into unit operations based on the existing automation tools and in extracting from them types of generic chemical workflows that can be converted into machine workflows. Then, in order to generalize the automated line, a lateral analysis of nearby domains must be added to ensure that no interesting opportunities are missed.

This workflow analysis is usually performed using the typical workflow representation scheme[102] that can be elaborated using GUI tools or through the Workflow Description Language (WDL).[103.] The analysis must also include the data management and device interface strategy. Historically several approaches have been developed with respect to this specific question, the first of which was due to Aspuru-Guzik and Persson in 2018[104] and is called the platform-based approach. Since that time, several new approaches due to Leonard and coworkers[13] or Lapkin and coworkers[4] with knowledge-graphs, among others, allow for optimized automated data generation global schemes.

Batch Versus Flow

An automated synthesis laboratory can be based on the batch approach, the flow chemistry approach, or a combination of the two. As extensively discussed by Holze and Boehling in 2022,[105] the choice between these different approaches has to be made according to the objectives of the laboratory. A discovery laboratory is inherently combinatorial and will likely benefit from being batch-based.

Indeed, the batch approach, combined with high-throughput (HT) capabilities, allows testing a large number of very different combinations of chemicals and, ideally, conditions in parallel. On the other hand, an optimization laboratory could potentially benefit from being flow-based, as it allows for smooth and rapid variation of conditions and therefore coupling with closed-loop optimization algorithms.[94,106,107] It is interesting to note that with the development of data-driven approaches and automation tools, several authors[7,108–110] advocate the use of flow chemistry for discovery as well. An analysis similar to the one cited above by Holze and Boehling, specifically dedicated to the data-driven approach, would certainly be interesting in this regard for example questioning the logistic aspects, but is beyond the scope of this chapter.

Discovery and Optimization Versus Preparative Laboratories

When designing an automated laboratory, it is essential to distinguish between a discovery and optimization laboratory, where the goal is essentially to generate data, and a preparative laboratory, where the goal is to physically extract the resulting molecules. The former can work with a limited amount of chemicals, just enough to perform the analysis and collect the data. It is also not necessary to plan separation methods for such a laboratory, since the majority of typical analytical methods for organic synthesis (except X-ray methods) are performed in the solution phase. On the contrary, if the physical molecule has to be finally collected, it is necessary to plan separation methods, also to be able to work with larger amounts of molecules in order to compensate for the losses occurring in the successive steps. Separation can be a potential source of complexity in automation due to solubility issues. In fact, solubility is a crucial point in automation that is also interfering with the analytical methods (chromatography, choice of solvent for spectroscopy) and for initial mixing steps. Several authors published recently articles treating precisely of automated solubility prediction using, among others, Hansen parameters.[9,111,112.]

What's Next

In Section (**State of the Art**), we will discuss in detail the state of the art in terms of automated sample handling strategies (**Sample Handling**), focusing on the sampling aspects, and then we will discuss the automated synthesis aspects (**Synthesis**). In **Chemical Analysis and Characterization**, we will look at the current state of automated characterization strategies applied to organic synthesis and then finally will focus on the IT and data management strategies as briefly evoked in the previous section. In Section **Review of Concrete Automation Example**, the reader will find an overview of concrete and recent automation examples.

Automation has revolutionized modern organic chemistry, offering unprecedented capabilities in both synthesis and analysis. In the field of synthesis, automation has evolved into a cutting-edge tool, enabling chemists to streamline complex processes, rapidly screen reaction conditions, and efficiently produce novel compounds. In this chapter, we will explore the state of the art in automated synthesis, going into the latest advancements and technologies that have reshaped the landscape of organic chemistry. Simultaneously, automation has also made significant progress in the realm of analytical chemistry, where it plays a pivotal role in characterizing and validating the synthesized compounds. From high-throughput screening to sophisticated analytical instrumentation, this chapter will highlight the pivotal role that automation plays in the analytical facets of organic chemistry, enhancing the accuracy and efficiency of chemical analyses. Together, these advancements underscore the transformative impact of automation on modern organic chemistry, from synthesis to analysis.

State of the Art

When mentioning automated laboratories, it can be very quickly shortcut to highly intricate systems including expensive equipment and highly specialized technicians.[3,46] Additionally, the experimental landscape in organic synthesis is extremely complex to automatize due to the broad variety of manual operations and technical apparatus required. However, in order to optimize their time and focus more on productive tasks or conceptualization, chemists have started to automatize some scientific equipment such as pipettes, injectors, collectors, scale, etc.[78] With the development of informatic systems, automation has become more complex[7,8] and the opportunities for automation in organic synthesis laboratories grew quickly with various purpose:

Increased efficiency: high throughput experimentation and miniaturization enables the scientists to collect more datapoints in a shorter time and using less resources.[1,113,114]

Increased quality of data: Automation enables the use of negative datapoints with more confidence. Datapoints can be easily duplicated and quality control added along the protocols. Logs and metadata are also collected during experiments, increasing the density and quality of databases.[49]

Robustness: Basic tasks that are repeated often in a process can be automated to reduce the load on human resource.[115] Such automated system can also be scaled more easily if it is required to increase the output, for example with the use of continuous systems.

Safety: Automation reduces the exposure to toxic chemicals or dangerous operations such as inside radiochemical labs or production plants.[116]

Nowadays, scientists can use a vast array of automated equipment for batch or flow chemistry to optimize processes.[65] The goal of this chapter is to highlight the different tools and strategies that are commonly used in research labs dealing with organic synthesis.

Sample Handling

Sample Management

Sample management is crucial in an automated laboratory. It ensures that the right samples are processed in the right way, reducing the chances of errors and ensuring reliable results. It reduces the time and effort required to locate, process, and track samples. This results in higher throughput and faster results. In order to deliver high quality results, it is important to preserve integrity by proper labeling, storage, and tracking of laboratory samples. An efficient sample management strategy will maintain an adequate supply of samples and reagents while minimizing wastage by enabling efficient inventorying techniques.

In summary, sample management is the backbone of an automated laboratory, ensuring that processes run smoothly, results are reliable, and compliance is maintained. It is an essential component of maintaining the high standards and efficiency required in modern scientific research and analysis.

Labeling

In an automated laboratory, labeling of the samples is required for sample recognition, traceability and safety. There are different types of technology that can be used which depends on the level of automation of the workflow. If the sample handling is performed by humans, it is necessary to have human-readable labels. Barcodes and QR codes can also be used if visual interface and scanners are available during the process. If the sample handling is a mix between humans and robots, RFID technology and barcodes will enable the creation of an efficient interface for the communication between the two worlds.

Sample Transfer

In order to integrate different scientific tools together, for example a liquid handler and an HPCL, it is required that the samples are transferred from a machine to another. In this case the transfer can be done manually as the throughput is not high and the equipment not far from each other's. In more complex or intensive environments, it becomes relevant to automatize the sample transfer and different strategies exist.

Mobile Platforms

Mobile platforms consist of a robotic arm placed upon a mobile unit (**Fig. 1**). Such platforms have the advantage that they can be quickly adapted to an existing lab configuration and only one unit can be used at different stations.[85] However, they require extensive alignment sequences in order to deal with the movement of the system. It can become complicated in case of expensive laboratory equipment that requires precise handling. Additionally, they represent a non-negligible danger in the lab as, despite their safety features, they can create unexpected situations if scientists are also working in the same room. In term of cost, they also represent a substantial investment to purchase, configure and maintain.

Conveyors

Conveyors are an inexpensive way to bring automation for sample transfer in the lab. Simple systems are very fast and robust. However, as it is often the case in research laboratories, they are complex to reconfigure once a layout is implemented. This lack of flexibility is also reflected by the fact that in case of breakdown, there is no redundancy and the whole system is often blocked.

Shuttle Systems

Shuttle systems are hybrid combination of mobile units and robotic arms (**Fig. 2**). The samples are loaded onto moving robots that can independently reach the desired station. Once arrived, a robotic arm can take care of the interface with the next scientific equipment in the process. This strategy is interesting as it has the advantages of the conveyor, using inexpensive robotic tools, the

Fig. 1 Mobile platform. Andy Cooper lab.[13].

Fig. 2 Shuttle system. Credit @SwissCat+

flexibility of the mobile platform, without the safety issues as they can be located outside the human space. Additionally, the alignment between expensive scientific equipment and robotic arms is fixed and avoid the requirement for alignment. The alignment with mobile robots and robotic arm does not require an extreme precision and can be handled by basic vision techniques.

Sample Data

Managing sample data in automated laboratories is essential for ensuring data integrity, traceability, and efficient laboratory operations. For this purpose, different types of software have been developed and implemented with a specific purpose. Laboratory Information Management Systems (LIMS) are designed to manage and track samples and associated metadata throughout their lifecycle. They provide functionalities for sample registration, tracking, storage, analysis, and reporting. Having an LIMS is a requirement in an automated laboratory as it enhances the overall efficiency of operations. In order to promote collaboration and provide a digital record of all laboratory activities, Electronic Laboratory Notebooks (ELN) allow scientists and researchers to record electronically and manage experimental data, sample information, protocols, and results. It is also worth mentioning the importance of inventory management software which can track the inventory of samples, reagents, and lab supplies. It ensures that labs have an accurate record of available resources and can manage stock levels efficiently. These use of these specific tools and software, in combination with others, may vary depending on the type of laboratory, the nature of the research or analysis, and regulatory requirements. Implementing a combination of these tools ensures that sample data is accurately recorded, tracked, and made available for analysis, ultimately contributing to the success of automated laboratory operations.

Standardization of Samples

In an automated organic chemistry laboratory, most of the materials are handled in glass or inert plastic vials. These vials can then be grouped on microplates. The advantages of the microplate is that its dimensions have been fixed by the Society of Biomolecular Screening (SBS) in 1995 and accepted as a standard in the field (Fig. 3). This standardization of solid handling is extremely useful to

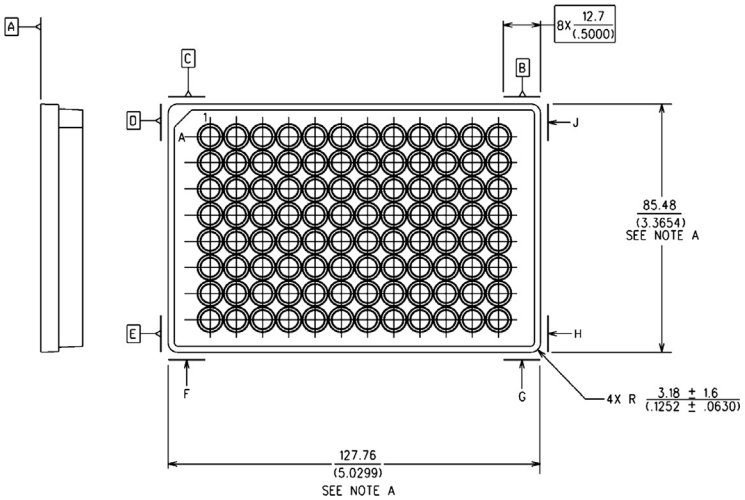


Fig. 3 Dimensions of the SBS plate. From ANSI SLAS 1–2004 (R2012).

integrate the workflows in the lab. The microplate can be configured for each tool but their outer dimensions are fixed, enabling the smooth transfer between scientific and robotic equipment.

As described in this first part, space and sample management is critical in an automated laboratory. They are the backbone on which the activity will grow and expand. When initializing an automation project, it is very important to think of various strategies that could be implemented and chose the most appropriated based on specific parameters (overall space, flexibility, budget etc …). Once this is set, the choice of scientific equipment can be considered. The next part will give a holistic overview of the automated equipment landscape.

Synthesis

The Laboratory

In automation, the laboratory space is part of the equation. It is not just a room; it is providing the environment required to support the functioning and connection of the different equipment. For this reason, it is important to implement as much as possible basic concepts in its design.

Hardware requirements play a fundamental role, as they dictate the type and quantity of equipment that needs to be accommodated within the space. This includes benches, fume hoods, and specialized machinery, all of which must be strategically placed to maximize efficiency. Sample flow is another crucial factor, determining how materials move through the lab. For this reason, space availability is a limiting factor that necessitates judicious allocation of resources. Energies, such as power and gas supplies, must be considered for the equipment's proper functioning. Ideally, they are located in strategic spots to be used by multiple equipment and are coming from the ceiling to give more flexibility for equipment positioning.

In order to help organizing laboratory space, layouts ensure a logical and efficient workflow, and technical drawings or CAD designs provide a visual representation of the planned setup, aiding in effective communication and decision-making. In combination, these elements enable the creation of a well-organized and functional laboratory space, facilitating smooth research and experimentation.

Equipment's for Automated Chemistry

In order to automatize experimental operations, different tools have been designed by equipment suppliers. Their classifications in term of material physical properties enables a comparison of the use cases of such equipment. For this reason, liquid and solid handlers represent the largest proportion of those tools. More specific instruments can also be found and will.

Liquid Handling

In chemical synthesis, most of the handling is done in the liquid state which is an advantage as liquids behave generally similarly. Some properties can be complicated to deal with, such as very viscous liquids or liquids with low boiling points, however, some strategies such as adjusting the temperature can help. For this reason, very simple systems exist for transferring liquids automatically.

Syringe Pumps

Syringe pumps are basic but precise and simple to use instruments. Once the size of the syringe is inputted and the flow required defined, the pump will provide a continuous dispense of liquid. They are useful in the case where precise slow additions over multiple hours are required.

Peristaltic Pumps

Peristaltic pumps are useful to transfer liquids from a container to another. They have a wide range of scale and speed to play with. Their mechanism is very simple, making them an inexpensive solution for liquid handling.

Multichannel Pipettes

Automatic pipettes with adjustable volumes are certainly the most represented tools in chemical laboratories. Their ease of use, precision and contamination-free tips make them a must have for repetitive liquid transfer tasks. They aspirate precisely liquids via air or positive displacement. This can influence the type of process and liquids that can be handled. They come with various sizes and channel numbers to improve throughput and precision. They can be used in combination with specific additional tools such as filters to adapt to more complex workflows.

Integrated Liquid Dispensing Stations

In order to increase drastically the output of liquid handling, platforms integrating robotics have been developed.[117] They are usually composed of an XYZ robotic system equipped with various handling capabilities to move microplates or samples, scan barcodes and adapt to complex workflows. They have been intensively used for analytical sample preparation, libraries creation or biological assay and allowed the manipulation of thousands of samples per day.

Some liquid handlers were designed with specific functions in mind. For example, with a small footprint to fit inside constrained laboratory space, to handle liquids in the nanoliter range[118–120] or inexpensive open-source equipment to lower the barrier to access automation. Such strategies enabled some developments of new methods in organic synthesis by being able to screen large

amount of reaction conditions. Interestingly, the chemistry had to be adapted to compatible solvents, which in the end, resulted in the discovery of catalysts compatible under such conditions.

Solid Sampling

Contrarily to liquids, solids behave in a much more complex way. They possess intrinsic physico-chemical properties such as flowability, compressibility, abrasiveness, size and shape … which make them hard to handle using automatized systems. For this reason, solid sampling has been limited to a few systems, standalone or integrated in larger platforms and depending on the technology can cover various range of scale and type of powders.[121] According to an article by Dr John Comley,[122] in 2009, compound storage and the supply of solid samples for retest or hit-to-lead confirmation represent the majority of the tasks undertaken, accounting for 43% and 37% of the activities, respectively. The next most explored area of focus is formulation development, which constitutes 7% of the undertaken tasks. The study was done in 2009 and in 2023, the use of solid dispensing has been largely extended to high throughput experimentation systems. Different equipment have been assessed in the literature and it can be very useful to refer to these studies when benchmark is required to decide a purchase for example.

In this context, automating the handling of small quantities, often in the sub-milligram range, of powders presents a significant challenge, and a universally fast and precise solution remains elusive. One of the primary reasons for this challenge is that most existing tools are primarily designed to dispense exact target masses, creating a trade-off between precision and the time required to achieve the desired result. Several robotic tools operate through a repetitive "collect and dispense" process to achieve precise powder quantities (Fig. 4).[117] However, this approach can be time-consuming, particularly when working with minute quantities of powders. Another category of robotic technologies employs gravimetric dispensing, where a container is positioned above the target vial on a balance. Through iterative "open and close" actions, these systems aim to reach the desired mass with a level of variable precision. The inherent complexity arises from the diverse properties of powders, which can range from being sticky, electrostatic, or hygroscopic. These properties further complicate the automation process, leading to slower or even infeasible handling of powders in a consistent and precise manner.

Some other types of solid handling technologies were developed but less frequently used, such as the electronic spatula, that uses electrostatics to collect solids in vials.

In order to improve the quality of automated solid dispensing, chemists have been using glass bead coated chemicals.[123,124] These techniques "dilute" the solid while reducing the variation of physical properties of the powders. In fine, lower scale and better precision are reached without loss of chemical activity.

Another strategy involves the encapsulation of reagents to dispense precisely and efficiently the desired chemicals in a very controlled fashion. This can be useful to then create libraries of compounds with important activity.[125]

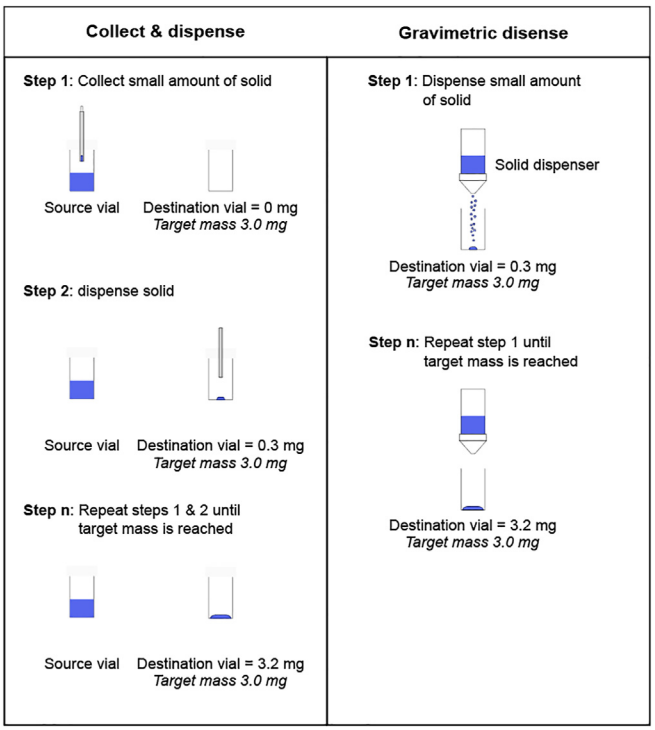


Fig. 4 Different mode of solid dispensing.

Other Applications

Whilst liquid and solid dispensing are the most represented techniques used in automation and high throughput experimentation, other more specific tools have been developed to complete workflows.[5,83,100,126,127] Automated evaporation or filtration systems, shakers, mixers, phase separators, etc., can be found depending on the application. They can also be integrated with sensors such as cameras or thermocouples to measure experimental conditions and provide additional metadata.

In the contemporary landscape, automation is progressively adopting a modular approach. The process of connecting equipment and constructing automated workflows has become more accessible, although not without its challenges. The cost of such automation equipment is rapidly declining, thanks to the emergence of novel tools from both academic research and industry suppliers, as well as the decreasing prices of robotics systems. However, the integration of these systems with IT infrastructure remains a high hurdle, albeit one that is actively being addressed through collaborative efforts with vendors.

Batch Versus Flow

In batch processing, chemicals are mixed together in a single vessel, and the entire reaction occurs as a single entity.[128] This approach offers simplicity and is well-suited for small-scale production. In contrast, flow chemistry involves a continuous stream of reactants flowing through a series of interconnected reactors, which provides precise control over reaction conditions.[110,129,130] Flow systems are highly efficient, enable rapid experimentation, and can facilitate the synthesis of complex molecules. They offer improved safety and scalability, and are particularly advantageous when handling hazardous or sensitive materials. The choice between batch and flow strategies ultimately depends on factors such as the desired level of automation, reaction complexity, safety requirements, and the scale of production.[131]

Flow chemistry relies on specialized equipment designed to facilitate continuous reactions within a controlled flow of reagents.[132] Key components of flow chemistry equipment typically include pumps to deliver reagents at precise flow rates, a mixer or reactor unit where reactions take place, and a residence time control mechanism to ensure precise reaction durations. Additionally, temperature control is critical, so many flow systems incorporate heating or cooling modules. Various types of reactors can be employed, such as microreactors, coil reactors, or packed-bed reactors, each offering specific advantages for different types of chemistry. In-line sensors and detectors monitor reaction progress and can trigger adjustments in real-time, providing greater control and safety.[131]

Outlook

The outlook for automation in chemistry labs is set for remarkable growth and transformation. However, it comes with notable challenges, primarily centered around the complexity of software integration and orchestration. As labs continue to adopt a diverse array of automated instruments, ensuring seamless communication and data exchange between these systems remains a critical hurdle. Additionally, many instruments have historically been designed with human operators in mind, necessitating further adaptation for complete automation. Yet, the potential is immense. The concept of self-driving labs, where robots manage experiments, data analysis, and decision-making autonomously, represents an attractive prospect. This could lead to unprecedented efficiency, reproducibility, and the accelerated discovery of novel compounds and materials.

Chemical Analysis and Characterization

In a recent paper, Hein and coworkers[133] have shed light on the importance of the analytical step in the data-driven approach. The authors distinguish the traditional single-point approach, where a specific data (e.g., HPLC peak area) is automatically acquired at a defined time, from the data-rich experimentation (DRE), where multiple points are repeatedly acquired with potentially an ensemble of analytical methods. The second method, thanks to the richness of the collected data, would definitely be the one of choice for any data-driven approach, since data limitation is probably the most limiting factor in the development of high-performance chemistry algorithms.[67,134,135] However, as the authors point out, such a method requires automatic chaining of all characterization methods, both for time reasons and for data robustness. Furthermore, in order to be used efficiently for optimization and data-driven scheduling of laboratory tasks, such a DRE approach generates a large amount of data that must be processed almost instantaneously. As illustrated by these two analytical approaches, the automation and integration of analytical and characterization steps into a more generally automated workflow poses significant challenges that we will explore in this section.

In order to fully understand the challenges, it is worth highlighting the functions expected of analytical steps in an automated synthesis workflow.

These functions are:

- qualitative identification of compounds,
- qualitative and quantitative analysis of mixtures,
- characterization of unknown compounds.

In the majority of automated systems described in the literature with some examples previously reported,[83,94,107,131] reactions and analyses are carried out in a restricted chemical domain, enabling the use of a limited number of analytical methods and potentially dispensing with characterization tools (resonance spectroscopies, spectrophotometries, XRD). In the discussion to follow, we will integrate all analytical dimensions by considering a global system capable of working on a large chemical space (e.g. all small organic molecules) without prior knowledge of the molecules used, and discuss their integration into an automated workflow.

More specifically, we will discuss the feasibility challenges of chaining multiple analytical methods in terms of workflow strategy, hardware and chemistry compatibility (subsection **Sample Handling and Method Chaining**), of automating different analytical instruments and methods (subsection **Automating Analytical Instruments and Methods**), of communication with instruments (subsection **Communication With Instruments**), of data formats (subsection **Analytical Data Formats**), and of data processing strategies (subsection **Data Processing**).

Sample Handling and Method Chaining

The complexity of chaining automatically various analytical methods comes from divers factors:

- compound solubility,
- suitability of solvent with analytical methods,
- concentration range with respect to the method sensitivity,
- potential use of internal standards.

Let's discuss these factors in detail.

Compound Solubility

It is no secret that not all molecules are soluble in all solvents. Solvent solubility is of central importance when it comes to dissolving molecules, but also when it comes to selecting the appropriate analytical method. In fact, when planning a workflow from synthesis to analysis of molecules, a real difficulty arises when it is certain that reaction products will have to be processed (work-up) and possibly dissolved in order to be analyzed. It is also necessary to plan the appropriate separation methods (LC, GC, SFC) to maximize the chances of obtaining a good separation of reagents and products in order to obtain an adequate description of the reaction mixtures obtained. It is also essential to know the solubility of the chemical compound of interest, if it is unknown, in order to choose the solvents used for its characterization (NMR, IR, UV, etc.).

To define the solubility of molecules as part of an automatic workflow, there are two main approaches that can be combined. The first approach consists of chemical solubility tests, possibly supported by visual recognition tools.[136,137] This empirical strategy is robust. However, it is limited to known molecules (starting materials or known products). Another approach to solubility prediction is based on chemical reaction propagation using either mechanistic[94,138,139] or LLM[140,141] algorithms. This propagation is used to predict the most likely products of the reaction and then be combined with DFT calculations to compute the LogP of molecules,[142] for example, or with machine learning algorithms to predict solubility in organic solvents.[143,144] Another complementary solubility estimation method is based on the Hansen Solubility Parameters.[145–147]

However, the procedures described above are complex and require considerable computing power, especially when applied to HTE, where a large number of molecules need to be evaluated on the fly. They are therefore not necessarily easy to implement in a fully automated way. automated way.

Suitability of Solvent With Analytical Methods

A concrete example of chaining analytical methods would be to start with a liquid chromatographic (LC) separation of reaction crudes to separate the components and collect the fractions obtained. These fractions are then analyzed by UV/VIS and FT-IR spectrophotometry, followed by NMR spectroscopy. In this case, compatibility problems are manifold.

First of all, the composition of the solvent resulting from reversed phased liquid chromatography (RPLC) is potentially evolving over time (gradient method). It is therefore not possible to record a background signal from the solvent in FT-IR or UV-VIS before acquiring the signal from the collected fraction. This produces optical spectra totally dominated by solvent signals. This effect is well known and is reflected in the baseline variation in UV detection applied to liquid chromatography.

Next comes the use of NMR spectroscopy. As is widely known, it is in principle desirable to use deuterated solvents in order to limit the saturating signals of protonated solvents.[148] It is potentially possible to suppress solvent signals either by presaturation or diffusion filtration.[149] However, these methods are fraught with risks, such as the loss of interesting signals in the event of overlapping, or the impossibility of suppressing all signals in the case of a multi-signal solvent (ethanol, toluene, etc.) or a mixture of solvents.

To overcome these difficulties, it is necessary to consider the strategy of solvent exchange by SPE. This can be done either online[150] or off-line.[151] But these solvent exchanges, particularly in an HTE context, are complex to implement, with problems such as lack of separative efficiency, yield and potential precipitation.

Concentration Range With Respect to the Method Sensitivity

Considering the same chain as above (synthesis to LCMS to UV-VIS and FTIR and then to NMR), the concentrations at the various stages can vary drastically. Let's start with a synthesis with a concentration of 1 mM and a yield of 50%. If we consider that after the work-up the volume is saved and all the product is recovered, the solution has a maximum concentration of 0.5 mM. 1 μL of this preparation solution is injected into the RPLCMS. According to experiments carried out in our laboratory, the fractions collected have an average volume of between 100 and 200 μL. Assuming almost complete recovery of the compound, we obtain a concentration of the order of 2–5 μM.

These latter concentrations are far too low to be potentially injected into high-resolution UV-VIS, high-resolution FT-IR or, even worse, NMR. Depending on the model of equipment and the configuration of the measuring cell, a concentration of 50–200 μM is

desirable in UV-VIS, and 0.5−1 mM in FT-IR (except in the case of an ATR detector, in which case a solution of the order of 500 mM should at least be considered).

Thus, in addition to the previously discussed compatibility problem, the direct use of LC solutions in spectrophotometry is not possible under common experimental conditions due to concentration limitations. At this point, we could imagine several strategies to overcome this new problem. One would be to increase the concentration of the starting material during the reaction. With a factor of 10 at the beginning, we could expect to reach the concentrations required for spectrophotometry. However, it is clear that reaction conditions can affect both the yield and the outcome of the reaction. Therefore, it may not be the best strategy. Another or complementary strategy would be to use preconcentration by SPE. This can be combined with the solvent exchange step described above.

In fact, the latter strategy makes particular sense when it comes to NMR. Indeed, the concentrations generally used are of the order of 1−10 mM, a far cry from the concentrations obtained at the end of liquid chromatography. However, as we have seen, solvent exchange is desirable in order to work with deuterated solvents. It is therefore rational to couple this exchange with preconcentration to have a chance of reaching sufficient concentrations and hence acquisition times compatible with the HTE approach.

That said, for preconcentration to take place, the mass ratios of products synthesized, separated and extracted in SPE and the volume ratios must be compatible with the hardware chain (reactors, analytical vs. preparative LC, optical cell volumes for spectroscopy, NMR tube volumes and diameters). In the case of a project to automate a laboratory and set up a complete analytical chain, it is essential to start by modeling a chain that takes precise account of the concentrations, volumes and masses of all the elements in the chain, in order to quickly establish the feasibility of the project and to correctly design intermediate steps such as dilutions and preconcentrations, or solvent exchanges.

Potential Use of Internal Standards

Analytical methods can be used for both qualitative and quantitative applications, as discussed in the introduction to this chapter. In both cases, it may be interesting or even necessary to add internal references and standards, particularly in view of the automated processing of results.

For example, in the case of automated NMR structural analysis, the alignment of 1D and 2D spectra is very important in order to identify correlations with a good degree of certainty. This often requires the addition of TMS or another reference. Furthermore, as NMR is absolute in terms of quantification,[152−154] once it has been correctly parameterized (pulses, relaxation time, field homogeneity), it is particularly tempting to use it to measure the concentration of solutions. For example, NMR can be used to quantify solutions that have already been measured using UV-VIS and FT-IR. Once "c" is defined and the optical path length "l" is known, it becomes possible to extract EL, a characteristic quantity of molecules, using Beer−Lambert's law.

The question that arises here is whether TMS should be added from the outset to the deuterated solvent used to prepare the solution, thereby potentially contaminating FT-IR and UV-VIS measurements, or whether it is preferable to add TMS between spectrophotometry and NMR?

Another problem is the chromatographic quantification of reaction mixtures. We know that the method is not absolute, and that the area of the peaks obtained for the different molecules depends, in addition to the concentration, on numerous factors including recovery rate, the sensitivity of the molecule to the detection wavelength, ionization in the case of detection by mass spectrometry, and so on. So, if you want to make a quantitative chromatographic measurement, you need to be able to calibrate either externally or internally. In the case of the HTE approach, it is not necessarily possible to know all the products in advance and therefore to have the standards required for calibration available. And even if these molecules are available, the choice of sequence for inserting the internal standard into the chain of methods remains complicated.

In conclusion to this subsection, we have seen that establishing the complete chaining of a series of analytical methods directly with the synthesis and work-up is intrinsically complicated. In fact, there are still other difficulties to be discovered in the process, some of which are linked to the choice of equipment used. We'll look at these in the next few subsections.

Automating Analytical Instruments and Methods

We will treat this subsection by discussing the different families of methods independently, and by looking at some specific difficulties in automating them. We will begin with chromatographic methods.

Chromatographic Methods

With regard to the automation of chromatographies, there are some hardware difficulties to be mentioned first. Among them, we can mention the frequent lack of compatibility of sample and holder shapes between different instruments of different or the same brand, and sometimes even between different parts of a single instrument (e.g. the autosampler and the fraction collector).

A second difficulty in automating chromatography instruments is the complexity of the kinematics used to operate them. They are mostly designed for human operation and often require the use of both hands in parallel. This makes automation unnecessarily complex, whereas quite simple adaptations such as automated trays for SBS[155] plates would work perfectly and be compatible with traditional laboratory SCARA collaborative robots (type Brooks/Precise Automation or PAA KX).

Part of this automation problem can be overcome in specific cases by using an online connection (flow mode) between the synthesis setup and the chromatographic device,[109,131] but this requires specific types of flow injectors and is obviously not compatible with a batch approach or, as discussed in the previous subsection, when other analytical methods and sample preparation steps need to be coupled to the chromatographic separation.

Another difficulty is the choice of appropriate methods. Indeed, to cover even a reasonably small chemical space of molecules, different structures, due to their differences in polarity, acidity and volatility, will require different conditions for optimal separation. A first strategy could be to define a set of different methods covering at best the expected chemical space and then to systematically run all samples by these different methods. Apart from being very information-rich, this strategy will quickly prove to be extremely costly in terms of solvent, time and columns. It is probably not applicable in preparative mode since the solvent cost and waste would quickly become prohibitive.

A perhaps more useful strategy is to use the same approach as described earlier in the solubility evaluation of molecules (subsection **Compound Solubility**). More specifically, it is interesting to use algorithms to predict the most likely products of the reactions and then, using a classifier algorithm[30] trained on the data sets generated by the chromatographic methods available in the laboratory, to suggest the most appropriate separation methods. Let us now turn to the characterization methods.

Characterization Methods

We consider here the following families of methods:

- optical spectrophotometry,
- mass spectrometry,
- resonance spectroscopy.

Let's take them one at a time. Optical spectrophotometers are generally well behaved when it comes to automation. They can be easily converted to flow mode using commercially available flow cells and, as mentioned above, they are quite solvent tolerant and sensitive. This last point makes them fast at obtaining good quality results, which is necessary for tracking reactions or measuring large arrays of samples. They can be an excellent choice for identifying already known compounds, especially IR methods that can be coupled with large databases and automated pattern recognition algorithms.[29,156,157] It can also be used to quantify the sample if ε_λ is known.

Despite the remarkable qualities described above, the information contained in optical spectrophotometry does not allow a complete structural elucidation of unknown molecules. For this, it is necessary to combine it with mass spectrometry and resonance spectroscopy.

The automation of mass spectrometry is generally quite straightforward, since it is usually connected to a GC or LC and does not raise any specific issues other than those mentioned above.

For resonance spectroscopy, mainly NMR, a distinction should be made between the low-frequency, low-resolution benchtop version with permanent magnets and the high-resolution version with superconducting cryomagnets. The former has a very small magnetic stray field and a small footprint (about 1 m^2), which allows it to be easily integrated into an instrumentation chain, possibly using flow probes. Several examples of reaction monitoring using benchtop NMR have been published recently.[158–160] However, due to its limited resolution and sensitivity, as well as important strong coupling effects, this technique cannot be used to characterize completely new unknown complex molecules. High-resolution NMR, on the other hand, has a large magnetic stray field and a large footprint (about 10 m^2) with drastic requirements for mechanical vibration control and thermal control. It usually must be integrated in batch mode using existing automation equipment in combination with laboratory automation. This makes it complex to integrate such instruments into a chain of other analytical instruments. Flow probes for high-field NMR are available, but due to the very long tubing required to connect such a probe to other instruments, high-resolution flow or stop-flow NMR techniques have never become standard, despite decades of development.

Another limitation of NMR is, of course, its inherent lack of sensitivity, which results in long acquisition times. This brings us to another issue in chaining, namely the time of acquisition in relation to the concentration of the sample and the interest of the acquired data in an HTE setup.

Time of Acquisition Versus Concentration Versus Data Interest

Coming back to high-field NMR, on-flow analysis is not to be expected due to the required acquisition times (>1 min.), which force to work either in batch mode or in stop-flow mode. The long acquisition times, especially in 2D NMR (COSY, HSQC, HMBC) or with low sensitivity nuclei (^{13}C, ^{31}P, ^{15}N), make it difficult to plan a systematic analysis of samples in the context of an HTE setup. However, this method is indispensable for structure elucidation and can hardly be avoided. Let us discuss some time related issues.

Solution concentration has a drastic effect on the acquisition time, especially for Fourier transformed spectroscopic methods (NMR, FT-IR). In fact, the signal-to-noise ratio (SNR) increases linearly with concentration, but only with the square root of the number of scans. It is therefore more efficient to increase the concentration of the solution to be measured instead of adding more scans to conserve short times of acquisition. But, as previously discussed, the concentration cannot easily be tuned since it is defined by the complete sample preparation chain. A fine tuning of the concentrations with respect to the sample volume and sample preparation steps must definitely be performed to find the best compromised possible.

Long acquisition times (GC, NMR) also make it critical to carefully evaluate the need for a specific analysis for a sample. In a discovery sequence, not all compounds are necessarily unknowns and require full characterization with long or complex sample preparation steps. Therefore, when automating an analytical chain, it is interesting to first perform a fast screening step (e.g. LC-DAD-MS or FT-IR), which allows to identify already known compounds (unreacted starting material, known products …) by comparing spectra. With this strategy it is possible to characterize only unknown compounds in detail and to reduce significantly the load of time-consuming equipment for data of low interest.

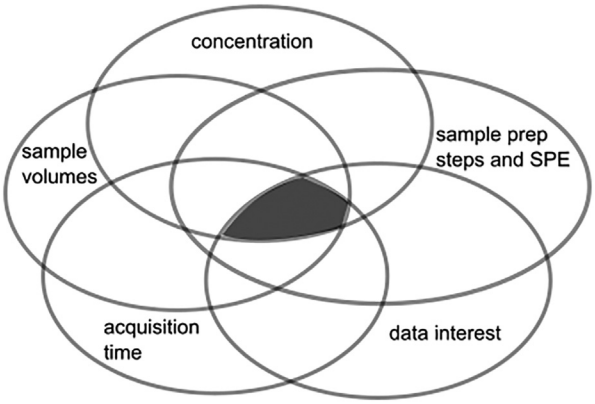


Fig. 5 To chain analytical instruments and methods in a more globally automated organic synthesis project, it is necessary to find the dark gray spot in the center that matches all parameters.

When planning a fully automated analytical chain, it is therefore necessary to make a critical analysis of all the parameters described above (concentration, sample volumes, sample preparation steps and SPE, acquisition time and data interest). As is often the case in chemistry automation, the game is to find the small common space that fits all the parameters, as shown in Fig. 5.

Communication With Instruments

A serious problem in automating a complete chain of analytical instruments is the scheduling and orchestration of tasks and samples on different instruments of potentially different brands. There is currently no commonly accepted communication standard for instruments, as has long been the case in industrial automation.

Most of the time, laboratories are obliged either to work with many different interfaces (middlewares, libraries) or even more complex, to code their own API using potentially a locally developed hardware interface with adequate I/O connections.

Some initiatives are currently going in the right direction with SiLa[89,90] and OPC-UA—LADS.[91] Both are based on a client-server structure where multiple devices (servers), regardless of make and model, can communicate with a centralized scheduler (client). We can only pray that one of these initiatives will reach a sufficient level of usage and recognition to become a true standard.

Analytical Data Formats

Another major problem faced by all automation chemists is the lack of non-proprietary, open and reasonably well-structured data formats generated by analytical instruments. Each vendor generates its own data format, if not several for the same types of data. These proprietary formats are generally compiled files that are not human readable and require specific commercial software to read. They also often produce PDF files with smoothed images of spectra and chromatograms, not ideal for automated post-processing and analysis.

The need is not for yet another new "standard" format; there are already many high-quality formats available in the literature.[79,80,161] Rather, there is a need for some general and basic rules, similar to the Asimov rules for robotics,[162] that all vendors should agree to in order to ensure data interoperability. Data format must:

- Follow FAIR principles,[163]
- Be human and machine readable without proprietary software,
- Be as much as possible numerical (e.g. xy tables of values instead of pictures of spectra, peaks lists, numerical peaks properties),
- Have a clear format and content description at the beginning of the file,
- Contain all but only necessary metadata (method, main parameters) and data (results) needed to repeat and validate the experiment on any other similar equipment from any brand.

To be fair, it is important to mention that several initiatives are moving in this direction. We can mention here the Allotrope Foundation,[88] which is working on such interoperable formats.

Data Processing

First, data processing must be clearly distinguished from data analysis, although they are closely related. Data analysis uses the processed data to extract chemical meaning. Whereas data processing is directly related to signal processing.

It starts very early in the instrument, since the initial signal is potentially analog (analog voltage or current generated by sensor per unit of time)[164] and must be sampled to become digital (digital-to-analog conversion, DAC).[165] Depending on the spectrometer's hardware (memory, DAC converter quality), the DAC introduces a more or less significant degree of pixelation of the analog signal (bin size in the time domain with respect to sampling frequency, sensitivity in the intensity domain with respect to the number of bits used by the DAC to encode the signal).

Then, potentially automated digitized intensity per time signal (time domain) treatments are applied to filter electronic noise or to further improve the appearance of the chromatograms and spectra and, at least for Fourier Transform (FT) spectroscopy, Fast Fourier Transform (FFT) algorithms[166,167] are applied to provide the typical intensity per frequency spectra (frequency domain). In principle, the FFT is a reversible calculation (up to numerical accuracy). This reversibility is crucial for many applications. For example, in digital communications, data is often transformed into the frequency domain for transmission and then back into the time domain at the receiver. This reversibility means that, in principle, we should be able to systematically recover the original time domain signal (e.g. FID in NMR or interferogram in FT-IR) from the FFT-generated spectra. However, reality is a bit different. In fact, the full reversibility can be altered by different FFT parameters, pre-FFT or post-FFT signal manipulation. Among others, alterations can be:

FFT parameters—Precision: Real-world implementations of FFTs and IFFTs (like those in computers) use floating-point arithmetic, which can introduce rounding errors. In principle, these are typically very small and do not prevent the operations from being effectively reversible.

Pre FFT—Windowing: In some applications, the time domain signal is windowed (multiplied by some function) or zeros are added at the beginning before the FFT is applied. This can also make the process irreversible, depending on how it's done and whether the windowing function is reversible.

Post-FFT data loss: If you modify the frequency-domain data (for example, by filtering, which can zero out certain frequencies), or correct the baseline, or apply a background spectrum difference, then the IFFT will not recover the exact original time-domain signal. The process is reversible only if the frequency domain data remains unchanged.

Post-FFT and post storage data loss: After spectrum output and final report generation, the image resolution in the.pdf file (possibly extracted from an article or external database) may be greatly reduced for storage reasons. Digitizing back such a spectrum image and applying IFFT to it will certainly produce a very low resolution time domain signal of poorer quality than the originally acquired time domain signal.

The automated data processing described above must force us to think seriously about the quality of the data we extract from analytical equipment (degree of alteration by hidden or overt signal processing, format and degree of compression used to generate images …). As much as possible, we should extract and store raw data at the lowest possible level. This raw data should be stored in a secure way (read-only) to ensure that it will not be altered by any post-processing.

A second level of data processing needs to be discussed here. Data extraction from processed chromatograms or spectra. For example, signal alignment,[168] deconvolution,[169–171] peak picking,[172–174] peak fitting[175,176] are critical steps that can be partially or fully automated. This step is essential for the upcoming data analysis step, but it is not really part of this step since it is essentially a pure treatment of the measured signals. As mentioned above, many groups are working to propose automated algorithms for extracting chromatogram and spectrum data. This step is necessary to generate useable numerical values for automated structure assignment or to relate spectral features to other types of observables such as yield or enantiomeric excess (ee). But we have to be careful at the time of reporting the extracted values (peak lists, peak properties, coupling constants …) because these algorithms are prone to errors and misinterpretations. They should therefore be stored together with the raw data mentioned above, with clear information about the extraction algorithm used. Ideally, it is also desirable to store the fit error and a level of confidence in the extracted data. In this way, it will always be possible in the future to reprocess the same raw data with a new generation of algorithm and to properly inform the data analysis tools with the degree of confidence of the data.

Looking forward in the direction of data analysis, the next step will be to use the previously extracted data for automated peak assignment (database comparison, comparison with simulations), structure elucidation, and possibly molecular property calculation (DFT, AI). In our vision, this last step is still part of the data processing, although it is quite far away from the original raw data with all the embedded risks of errors due to algorithmic inaccuracies. This last step should be considered as a third level of processing and, like the second, should be added to the raw data without affecting it, and should be clearly labeled with the algorithms used, the confidence levels, and the error estimates.

Finally, the processed data can be used for data analysis, which consists of observing series of molecules and reactions and finding significant trends and correlations between reaction conditions, relevant extracted molecular properties, and observables such as yield and ee.

The global data processing strategy described above is summarized in a more global data analysis scheme in the next Fig. 6.

Outlook and Perspectives

The analytical and characterization segment of the overall automated synthesis workflow is becoming increasingly important. A good automation chemist must at least have a good understanding of chromatography, spectroscopy and sample preparation methods, in order to find suitable strategies for linking them effectively to the synthesis steps and between analytical methods.

It is interesting to note that, while the theory behind these different parts of the overall synthesis process is clearly different, automated synthetic chemistry methods increasingly resemble, in programming terms, what has been done for a relatively long time in analytical chemistry. We're talking more and more in terms of methods with parameters and variables that are encoded in automata. This similarity could be an interesting source of global coherence in chemistry, particularly when it comes to data and metadata formats.

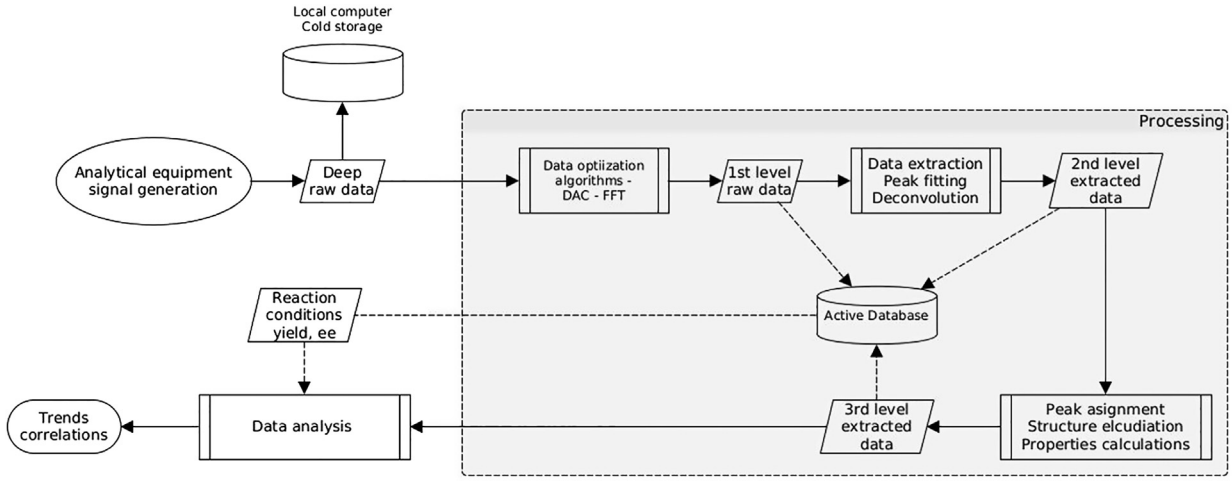


Fig. 6 Data processing structure inside a global data generation and analysis scheme.

Review of Concrete Automation Example

Use of Automation for Method Development and Process Optimization

In this section, we will present a number of case studies to highlight the diverse and impactful applications of automation in the field of chemical synthesis. By exploring these case studies, we will gain valuable insights into the practical benefits, challenges and transformative potential of automation in the rapidly evolving field of digital chemistry.

Case study 1: In a recent paper, Doyle, Sigman and co-workers make use of automated reaction set up with liquid handlers to collect data around Ni− and Pd-catalyzed cross coupling.[177] The catalyst are pre-plated in solution and the solvent evaporated. Once ready, those plates can be used with substrate mix and shaken under reaction conditions. This strategy uses a Tecan liquid handler, tumble stirrer from Core Module and plates from Analytical Sales. Analysis is performed using UPLC-MS runs of 2.5 minutes. This setup enabled the testing of various metal−ligand combinations to feed machine-learning algorithms. Eventually, a strategy for the classification of monodentate phosphine ligation and reactivity in cross coupling catalysis was developed.

Case study 2: In 2015, Cernak and co-workers used a Mosquito form Labtech to perform nanomolar experiments.[118] It is very frequent in industry that starting material is scarce and in order to optimize its use, running reactions at very small scales is crucial. In this context, it was found that high-boiling points polar solvents could be used with the right combination of Pd-complexes after thousands of conditions tested. The workflow enabled the generation of heat maps of organic reactions to identify quickly successful conditions and optimization.

Case study 3: Automation of on-demand synthesis using flow chemistry was also investigated.[131] A robotic platform was designed and operated in a modular fashion in order to perform multistep synthesis. In combination with artificial intelligence, the platform is able to autonomously decide the synthetic route and configure itself using a robotic arm. One of the highlighted constraints was that the AI model relied on historical experimental data which are not sufficiently clean to train efficient models. With the development of new automated platforms, it is clear that these problems will be solved in a near future. An interesting example of such developments can be seen in the recent work of Slattery et al.[178] In this study, the authors propose the RoboChem, which is designed to facilitate the self-optimization, intensification and scale-up of photocatalytic transformations. This is achieved through the integration of readily available hardware, customized software and a Bayesian optimization (BO) algorithm. We can also mention here the development of several large-scale automated infrastructures such as, just for academia, ROAR[179] and ATLAS,[180] both at Imperial College, the Bristol Automated Synthesis Facility at the University of Bristol,[181] the A-Lab at Berkley,[182] the Acceleration Consortium at the University of Toronto[183] or the Catalysis Hub Swiss Cat null+ at ETHZ and EPFL,[184] which are dedicated to the production of high quality large data sets worldwide. Similar developments are taking place in industry.

Still in relation to case study 3, a parallel approach is described by Gromski et al.[83] with the development of the Chemputer. Here the authors developed a solution by transforming the manual tools into robotic modules, thanks to automated pumps and drains. The instructions are coded in a programming language (χDL) that standardizes communication between chemists and hardware, between computers and hardware, and between chemists themselves, so that experiments can be easily reproduced, thus producing more reproducible datasets.

Case study 4: The optimization of catalytic processes can be very time and resource consuming. Usually, it is performed by changing one parameter at a time and observing the consequences on the reaction outcome. By using closed loop optimization, the optimization process learns from the experimental results by a machine learning model and propose the next experiment to run. Ideally, optimal conditions are reached more rapidly. Along the way, data is collected and can be used to improve prediction algorithms. In 2022, the team of Burke developed an automated flow platform to optimize Suzuki−Miyaura couplings autonomously.[94] It was shown that the system was able to predict general conditions for a complex C−C coupling event, enabling the

synthesis of small molecules. Another recent example of such closed loop optimization setup is proposed by the team of L. Cronin[185] and is an evolution of the already presented Chemputer.

Use of Algorithms to Automate Synthesis

Case study 5: To fully automate a laboratory, it is necessary to be able to automatically generate protocols, synthetic routes and chemical properties predictions that can potentially be fed into the laboratory controller or scheduler. We can identify two main strategies, the expert systems, for example, used to generate retrosynthesis trees based on a large ensemble of chemical rules, as is typically done by the Chematica software developed in the Grzybowski group.[186] This strategy is already largely used in industry with proven results to rationalize the starting materials.[139] Or the Large Language Models (LLM) approaches, such as ChemCrow developed by Schwaller and co-workers.[99] An interesting application which combine directly such kind of LLM predictive algorithm with an automated synthetic platform has been presented by the Laino group at IBM Research.[187] Also, an important development is taking place in this field with the recent publication of several papers reporting an efficient use of chatbots for predictions in chemistry.[188,189] As mentioned above, one of the limitations of these LLM methods lies in the quality of the datasets available for training the models.[190−192] The current trend towards high throughput and laboratory automation should help to generate high quality and widely available datasets, thus contributing to the improvement of these automated models.

Case study 6: Optimization of reaction conditions can be performed either by systematic condition exploration or by design of experiments (DOE),[193] or more recently by using Bayesian Optimization (BO) algorithms.[98,194] Several recent developments need to be highlighted, including an application example for heterogeneous catalysis developed at the Swiss Cat null+ East Hub at ETHZ by Paco Laveille and co-workers,[195] or the integration of a BO optimizer into an automated robotic flow platform for synthetic chemistry by Jensen and co-workers.[196] This strategy seems promising for this type of applications and, unlike other new generation algorithms, does not require excessively large datasets to train.[197]

Automation of Analysis and Characterization

As described in Section (State of the Art), the automation of analysis and characterization is essential for the development of a complete automated organic synthesis setup. Thurow and Fleisher in 2017,[198] wrote a reference book that specifically addresses the automation of analytical measurements and presents several interesting use cases. In 2021, Garcia and co-workers,[199] wrote a review presenting different potential applications of AI in analytical chemistry. The combination of automation and AI creates many new possibilities for the full integration of analytical steps in a globally automated synthesis workflow. To conclude the case study section, let us take a look at a specific topic in analytical chemistry that has been particularly active recently.

Case study 7: Data processing is critical to the global HTE and data-driven chemistry process. Due to the required throughput, automated advanced processing of the acquired data is mandatory. For example, automated processing of HPLC data as proposed by Haas and al.[200] in 2023 is essential. In this registry, automated NMR processing and ideally automated and robust structure elucidation is a must. Several authors propose open source data processors. Some, such as the DP4-AI by Goodman and co-workers in 2020[201] are oriented towards automated processing and assignment of ^{13}C and ^{1}H NMR data. It is optimized to assign candidate structures to raw NMR data with a high degree of efficiency and to perform a complete analysis including stereochemistry. Another interesting tool in the same field, called Schmarnica and proposed by Pesek, Gazvoda and co-workers in 2021,[202] is based on a graph theory approach. This tool can automatically generate possible structures from IR, ^{1}H and ^{13}C NMR and MS data. Some other tools, such as the one proposed by Kwon et al. in 2020,[203] are more focused on predicting chemical shifts using trained neural networks. This is an active field and new studies are presented regularly. In the context of the Swiss Cat null+ West Hub, we tend to combine these different tools and compare the predictions to improve their respective robustness.

Conclusion

In this comprehensive exploration of modern automation in organic synthesis laboratories, we have described the fascinating evolution of automation and high-throughput methods in experimental science and, more specifically, in organic synthesis. From its inception to the specialized and indispensable role it plays in chemistry, we've traced the evolution of automation as a pioneering force in the laboratory. We have shown how automation has moved seamlessly into the laboratory, enabling researchers to tackle complex chemical challenges with unprecedented efficiency. We have shown how automation has continually reshaped the way chemists conduct experiments and synthesize compounds, from the earliest automated titrators to today's robotic platforms. From this analysis, it is clear that automation and data science are closely intertwined and mutually dependent. Data enables chemists to optimize reaction conditions, improve reproducibility and ensure consistent production of high-quality compounds. By providing a detailed record of the synthesis process, data not only supports the validation of results, but also plays a critical role in quality control by confirming the identity and purity of synthesized compounds. In addition, data helps troubleshoot unexpected results and provides the basis for scaling up processes to industrial levels. In the age of big data and machine learning, the wealth of information generated from organic synthesis opens the door to predictive modeling, ultimately advancing the field and driving innovation. As the field of automation continues to evolve, this chapter has provided a foundation for understanding its past, its dynamic present, and its potential for the future.

References

1. Trobe, M.; Burke, M. D. The Molecular Industrial Revolution: Automated Synthesis of Small Molecules. *Angew. Chem. Int. Ed.* **2018,** *57* (16), 4192—4214. https://doi.org/10.1002/anie.201710482.
2. Shi, Y.; Prieto, P. L.; Zepel, T.; Grunert, S.; Hein, J. E. Automated Experimentation Powers Data Science in Chemistry. *Acc. Chem. Res.* **2021,** *54* (3), 546—555. https://doi.org/10.1021/acs.accounts.0c00736.
3. Sanderson, K. Automation: Chemistry Shoots for the Moon. *Nature* **2019,** *568* (7753), 577—579. https://doi.org/10.1038/d41586-019-01246-y.
4. Bai, J.; Cao, L.; Mosbach, S.; Akroyd, J.; Lapkin, A. A.; Kraft, M. From Platform to Knowledge Graph: Evolution of Laboratory Automation. *JACS Au.* **2022.** https://doi.org/10.1021/jacsau.1c00438. jacsau.1c00438.
5. Holland, I.; Davies, J. A. Automation in the Life Science Research Laboratory. *Front. Bioeng. Biotechnol.* **2020,** *8*.
6. 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. https://doi.org/10.1126/science.aar5169.
7. Coley, C. W.; Eyke, N. S.; Jensen, K. F. Autonomous Discovery in the Chemical Sciences Part I: Progress. *Angew. Chem. Int. Ed.* **2020,** *59* (51), 22858—22893. https://doi.org/10.1002/anie.201909987.
8. Coley, C. W.; Eyke, N. S.; Jensen, K. F. Autonomous Discovery in the Chemical Sciences Part II: Outlook. *Angew. Chem. Int. Ed.* **2020,** *59* (52), 23414—23436. https://doi.org/10.1002/anie.201909989.
9. Alsenz, J.; Kansy, M. High Throughput Solubility Measurement in Drug Discovery and Development. *Adv. Drug Deliv. Rev.* **2007,** *59* (7), 546—567. https://doi.org/10.1016/j.addr.2007.05.007.
10. Bajorath, J. *Deep Machine Learning for Computer-Aided Drug Design*. https://doi.org/10.3389/fddsv.2022.829043, 2022.
11. Bogdan, A.; Gesmundo, N.; Webster, M. P.; Wang, Y.; Vasudevan, A. Recent Advances in Chemistry Technologies and Applications to Medicinal Chemistry. *Burger's Med. Chem. Drug Discov.* **2021,** 1—78. https://doi.org/10.1002/0471266949.BMC293.
12. Abolhasani, M.; Kumacheva, E. The Rise of Self-Driving Labs in Chemical and Materials Sciences. *Nat. Synth.* **2023.** https://doi.org/10.1038/s44160-022-00231-0. null, null.
13. Flores-Leonar, M. M.; Mejía-Mendoza, L. M.; Aguilar-Granda, A.; Sanchez-Lengeling, B.; Tribukait, H.; Amador-Bedolla, C.; Aspuru-Guzik, A. Materials Acceleration Platforms: On the Way to Autonomous Experimentation. *Curr. Opin. Green Sustain. Chem.* **2020,** *25,* 100370. https://doi.org/10.1016/j.cogsc.2020.100370.
14. Gomez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; MacLaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D.-G.; Wu, T.; Markopoulos, G.; Jeon, S.; Kang, H.; Miyazaki, H.; Numata, M.; Kim, S.; Huang, W.; Hong, S. I.; Baldo, M.; Adams, R. P.; Aspuru-Guzik, A. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016,** *15* (10), 1120—1127. https://doi.org/10.1038/nmat4717.
15. Turign, A. M. I. —Computing Machinery and Intelligence. *Mind* **1950,** *LIX* (236), 433—460. https://doi.org/10.1093/mind/LIX.236.433.
16. Amdahl, G. M.; Blaauw, G. A.; Brooks, F. P., Jr. Architecture of the IBM System/360. *IBM J. Res. Dev.* **1964,** *8* (2).
17. Knuth, D. E. On the Translation of Languages from Left to Right. *Inf. Control* **1965,** *8* (6), 607—639. https://doi.org/10.1016/S0019-9958(65)90426-2.
18. Tardieu, H.; Daly, D.; Esteban-Lauzán, J.; Hall, J.; Miller, G. The Birth of Digital—A Brief History of Digital Technologies. In *Deliberately Digital: Rewriting Enterprise DNA for Enduring Success;* Tardieu, H., Daly, D., Esteban-Lauzán, J., Hall, J., Miller, G., Eds.; *Future of Business and Finance*, Springer International Publishing: Cham, 2020; pp 3—7. https://doi.org/10.1007/978-3-030-37955-1_1.
19. Perkel, J. M. Ten Computer Codes That Transformed Science. *Nature* **2021,** *589* (7842), 344—348. https://doi.org/10.1038/d41586-021-00075-2.
20. Meltzer, B.; Michie, D. Machine Intelligence 4. *Br. J. Philos. Sci.* **1970,** *21* (2), 212—214.
21. Langley, P. *Scientific Discovery: Computational Explorations of the Creative Processes*, MIT Press, 1987.
22. Sparkes, A.; Aubrey, W.; Byrne, E.; Clare, A.; Khan, M. N.; Liakata, M.; Markham, M.; Rowland, J.; Soldatova, L. N.; Whelan, K. E.; Young, M.; King, R. D. Towards Robot Scientists for Autonomous Scientific Discovery. *Autom. Exp.* **2010,** *2*. https://doi.org/10.1186/1759-4499-2-1 (Copyright © 2023 U.S. National Library of Medicine.), 1.
23. Grey Walter, W. An Imitation of Life. *Sci. Am.* **1950,** 42—45.
24. Kuipers, B.; Feigenbaum, E. A.; Hart, P. E.; Nilsson, N. J. Shakey: From Conception to History. *AI Mag.* **2017,** *38* (1), 88—103. https://doi.org/10.1609/aimag.v38i1.2716.
25. Macdonald, R. B. A Summary of the History of the Development of Automated Remote Sensing for Agricultural Applications. *IEEE Trans. Geosci. Remote Sens.* **1984,** *GE-22* (6), 473—482. https://doi.org/10.1109/TGRS.1984.6499157.
26. Feigenbaum, E. A. *Expert Systems in the 1980s*, Stanford University, 1980.
27. Campbell, M.; Hoane, A. J.; Hsu, F. Deep Blue. *Artif. Intell.* **2002,** *134* (1), 57—83. https://doi.org/10.1016/S0004-3702(01)00129-1.
28. Simon, H. A. Artificial Intelligence: An Empirical Science. *Artif. Intell.* **1995,** *77* (1), 95—127. https://doi.org/10.1016/0004-3702(95)00039-H.
29. Bishop, C. M. *Pattern Recognition and Machine Learning. Information Science and Statistics,* 1st ed.; Springer: New York, NY, 2006.
30. Bates, M. Models of Natural Language Understanding. | PNAS. *PNAS* **1995,** *92* (22), 9977—9982. https://doi.org/10.1073/pnas.92.22.9977.
31. Mokhbery, M. Dexpert: The Application of AI to Design of Experiments. *IFAC Proc.* **1992,** *25* (28), 362—365. https://doi.org/10.1016/S1474-6670(17)49526-7.
32. O'Leary, D. E. Artificial Intelligence and Big Data. *IEEE Intell. Syst.* **2013,** *28* (2), 96—99. https://doi.org/10.1109/MIS.2013.39.
33. Iten, R.; Metger, T.; Wilming, H.; del Rio, L.; Renner, R. Discovering Physical Concepts with Neural Networks. *Phys. Rev. Lett.* **2020,** *124* (1), 010508. https://doi.org/10.1103/PhysRevLett.124.010508.
34. Tward, E.; Armstrong, R. L. The Application of an Automated Data Acquisition System to T1 Measurements in Compressed Hydrogen Gas. *Can. J. Phys.* **1968,** *46* (5), 331—336. https://doi.org/10.1139/p68-049.
35. Gladstone, D. J.; Chin, L. M. Automated Data Collection and Analysis System for MOSFET Radiation Detectors. *Med. Phys.* **1991,** *18* (3), 542—548. https://doi.org/10.1118/1.596731.
36. Megie, G. Laser Remote Sensing: Fundamentals and Applications. *EOS Trans. Am. Geophys. Union* **1985,** *66* (40). https://doi.org/10.1029/EO066i040p00686-05, 686—686.
37. Genet, R. M. *History of Robotic and Remotely Operated Telescopes*, California Polytechnic State University, 2011.
38. Watson, K. Remote Sensing—A Geophysical Perspective. *Geophysics* **1985,** *50* (12), 2595—2610. https://doi.org/10.1190/1.1441885.
39. Kumar, G. G.; Sahoo, S. K.; Meher, P. K. 50 Years of FFT Algorithms and Applications. *Circuits Syst. Signal Process.* **2019,** *38* (12), 5665—5698. https://doi.org/10.1007/s00034-019-01136-8.
40. Butcher, J. C.; Wanner, G. Runge-kutta Methods: Some Historical Notes. *Appl. Numer. Math.* **1996,** *22* (1), 113—151. https://doi.org/10.1016/S0168-9274(96)00048-7.
41. Hammersley, J. M.; Handscomb, D. C. The General Nature of Monte Carlo Methods. In Monte Carlo Methods*; Monographs on Applied Probability and Statistics Book Series (MSAP);* 1964.
42. Nassif, A. B.; Talib, M. A.; Nasir, Q.; Dakalbab, F. M. Machine Learning for Anomaly Detection: A Systematic Review. *IEEE Access* **2021,** *9,* 78658—78700. https://doi.org/10.1109/ACCESS.2021.3083060.
43. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*, Springer: New York: New York, NY, 2013. https://doi.org/10.1007/978-1-4614-6849-3.
44. Zhang, S.; Jacobs, R.; Ghosh, S.; Kulkarni, A.; Wang, L. *Automated Data-Driven Physics Discovery of Turbine Component Damage*, American Society of Mechanical Engineers Digital Collection, 2022. https://doi.org/10.1115/GT2022-83372.

45. Karniadakis, G. E.; Kevrekidis, I. G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-Informed Machine Learning. *Nat. Rev. Phys.* **2021,** *3* (6), 422—440. https://doi.org/10.1038/s42254-021-00314-5.
46. Heather, J. M.; Chain, B. The Sequence of Sequencers: The History of Sequencing DNA. *Genomics* **2016,** *107* (1), 1—8. https://doi.org/10.1016/j.ygeno.2015.11.003.
47. Shendure, J.; Balasubramanian, S.; Church, G. M.; Gilbert, W.; Rogers, J.; Schloss, J. A.; Waterston, R. H. DNA Sequencing at 40: Past, Present and Future. *Nature* **2017,** *550* (7676), 345—353. https://doi.org/10.1038/nature24286.
48. Forestiero, S. The Historical Nature of Biological Complexity and the Ineffectiveness of the Mathematical Approach to it. *Theor. Biosci.* **2022,** *141* (2), 213—231. https://doi.org/10.1007/s12064-022-00369-7.
49. Kearnes, S. M.; Maser, M. R.; Wleklinski, M.; Kast, A.; Doyle, A. G.; Dreher, S. D.; Hawkins, J. M.; Jensen, K. F.; Coley, C. W. The Open Reaction Database. *J. Am. Chem. Soc.* **2021,** *143* (45), 18820—18826. https://doi.org/10.1021/jacs.1c09820.
50. Pallas, N. R.; Harrison, Y. An Automated Drop Shape Apparatus and the Surface Tension of Pure Water. *Colloids Surf.* **1990,** *43* (2), 169—194. https://doi.org/10.1016/0166-6622(90)80287-E.
51. Handbook of Chemistry and Physics. https://hbcp.chemnetbase.com/contents/ContentsResults.xhtml?dswid=7245 (accessed 2023-07-20).
52. Hood, L. E.; Hunkapiller, M. W.; Smith, L. M. Automated DNA Sequencing and Analysis of the Human Genome. *Genomics* **1987,** *1* (3), 201—212. https://doi.org/10.1016/0888-7543(87)90046-2.
53. Human Genome Project Fact Sheet. Genome.gov. https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genome-project (Accessed 2023-07-20).
54. Seehausen, O.; Butlin, R. K.; Keller, I.; Wagner, C. E.; Boughman, J. W.; Hohenlohe, P. A.; Peichel, C. L.; Saetre, G.-P.; Bank, C.; Brännström, Å.; Brelsford, A.; Clarkson, C. S.; Eroukhmanoff, F.; Feder, J. L.; Fischer, M. C.; Foote, A. D.; Franchini, P.; Jiggins, C. D.; Jones, F. C.; Lindholm, A. K.; Lucek, K.; Maan, M. E.; Marques, D. A.; Martin, S. H.; Matthews, B.; Meier, J. I.; Möst, M.; Nachman, M. W.; Nonaka, E.; Rennison, D. J.; Schwarzer, J.; Watson, E. T.; Westram, A. M.; Widmer, A. Genomics and the Origin of Species. *Nat. Rev. Genet.* **2014,** *15* (3), 176—192. https://doi.org/10.1038/nrg3644.
55. Goodwin, S.; McPherson, J. D.; McCombie, W. R. Coming of Age: Ten Years of Next-Generation Sequencing Technologies. *Nat. Rev. Genet.* **2016,** *17* (6), 333—351. https://doi.org/10.1038/nrg.2016.49.
56. The Cost of Sequencing a Human Genome. Genome.gov. https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost (Accessed 2023-07-20).
57. Cui, M.; Cheng, C.; Zhang, L. High-Throughput Proteomics: A Methodological Mini-Review. *Lab. Invest.* **2022,** *102* (11), 1170—1181. https://doi.org/10.1038/s41374-022-00830-7.
58. D'Alessandro, A., Ed.; High-Throughput Metabolomics: Methods and Protocols; *Methods in Molecular Biology*vol. 1978; Springer: New York, NY, 2019. https://doi.org/10.1007/978-1-4939-9236-2.
59. Kitano, H. Systems Biology: A Brief Overview. *Science* **2002,** *295* (5560), 1662—1664. https://doi.org/10.1126/science.1069492.
60. Systems biology—Latest research and news | Nature. https://www.nature.com/subjects/systems-biology (Accessed 2023-07-20).
61. Bioinformatics—Latest research and news | Nature. https://www.nature.com/subjects/bioinformatics (Accessed 2023-07-20).
62. Bleicher, K. H.; Böhm, H.-J.; Müller, K.; Alanine, A. I. Hit and Lead Generation: Beyond High-Throughput Screening. *Nat. Rev. Drug Discov.* **2003,** *2* (5), 369—378. https://doi.org/10.1038/nrd1086.
63. Drews, J. Drug Discovery: A Historical Perspective. *Science* **2000,** *287* (5460), 1960—1964. https://doi.org/10.1126/science.287.5460.1960.
64. Inglese, J.; Johnson, R. L.; Simeonov, A.; Xia, M.; Zheng, W.; Austin, C. P.; Auld, D. S. High-Throughput Screening Assays for the Identification of Chemical Probes. *Nat. Chem. Biol.* **2007,** *3* (8), 466—479. https://doi.org/10.1038/nchembio.2007.17.
65. Mayr, L. M.; Bojanic, D. Novel Trends in High-Throughput Screening. *Curr. Opin. Pharmacol.* **2009,** *9* (5), 580—588. https://doi.org/10.1016/j.coph.2009.08.004.
66. Zhang, L.; Tan, J.; Han, D.; Zhu, H. From Machine Learning to Deep Learning: Progress in Machine Intelligence for Rational Drug Discovery. *Drug Discov. Today* **2017,** *22* (11), 1680—1685. https://doi.org/10.1016/j.drudis.2017.08.010.
67. Bender, A.; Cortes-Ciriano, I. Artificial Intelligence in Drug Discovery: What Is Realistic, what Are Illusions Part 2: A Discussion of Chemical and Biological Data. *Drug Discov. Today* **2021,** *26* (4), 1040—1052. https://doi.org/10.1016/j.drudis.2020.11.037.
68. Meyers, J.; Fabian, B.; Brown, N. De Novo Molecular Design and Generative Models. *Drug Discov. Today* **2021,** *26* (11), 2707—2715. https://doi.org/10.1016/j.drudis.2021.05.019.
69. Green, M. L.; Takeuchi, I.; Hattrick-Simpers, J. R. Applications of High Throughput (Combinatorial) Methodologies to Electronic, Magnetic, Optical, and Energy-Related Materials. *J. Appl. Phys.* **2013,** *113* (23), 231101. https://doi.org/10.1063/1.4803530.
70. Potyrailo, R.; Rajan, K.; Stoewe, K.; Takeuchi, I.; Chisholm, B.; Lam, H. Combinatorial and High-Throughput Screening of Materials Libraries: Review of State of the Art. *ACS Comb. Sci.* **2011,** *13* (6), 579—633. https://doi.org/10.1021/co200007w.
71. Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018,** *559* (7715), 547—555. https://doi.org/10.1038/s41586-018-0337-2.
72. Zhang, Y.; Apley, D. W.; Chen, W. Bayesian Optimization for Materials Design with Mixed Quantitative and Qualitative Variables. *Sci. Rep.* **2020,** *10* (1), 4924. https://doi.org/10.1038/s41598-020-60652-9.
73. Armbruster, D. A.; Overcash, D. R.; Reyes, J. Clinical Chemistry Laboratory Automation in the 21st Century—Amat Victoria Curam (Victory Loves Careful Preparation). *Clin. Biochem. Rev.* **2014,** *35* (3), 143—153.
74. Merrifield, R. B.; Stewart, J. M.; Jernberg, N. Instrument for Automated Synthesis of Peptides. *Anal. Chem.* **1966,** *38* (13), 1905—1914. https://doi.org/10.1021/ac50155a057.
75. Handbook of Modern Pharmaceutical Analysis, vol. 10, 2nd ed.
76. de Galan, L.; Kornblum, G. R.; de Loos-Vollebregt, M. T. C. Automated Atomic Spectrometric Analysis. In *Recent Advances in Analytical Spectroscopy;* Fuwa, K., Ed., Pergamon, 1982; pp 33—50. https://doi.org/10.1016/B978-0-08-026221-5.50009-X.
77. Long, A. K.; Others, A. A Computer Program for Organic Synthesis. *Chem. Eng. News* **1983,** *61* (19), 22—30.
78. Lindsey, J. S. A Retrospective on the Automation of Laboratory Synthetic Chemistry. *Chemom. Intell. Lab. Syst.* **1992,** *17* (1), 15—45. https://doi.org/10.1016/0169-7439(92)90025-B.
79. Baumbach, J. I.; Davies, A. N.; Lampen, P.; Schmidt, H. J. C. A. M. P.-D. X. A Standard Format for the Exchange of Ion Mobility Spectrometry Data (IUPAC Recommendations 2001). *Pure Appl. Chem.* **2001,** *73* (11), 1765—1782. https://doi.org/10.1351/pac200173111765.
80. Bhat, T. N.; Bourne, P.; Feng, Z.; Gilliland, G.; Jain, S.; Ravichandran, V.; Schneider, B.; Schneider, K.; Thanki, N.; Weissig, H.; Westbrook, J.; Berman, H. M. The PDB Data Uniformity Project. *Nucleic Acids Res.* **2001,** *29* (1), 214—218. https://doi.org/10.1093/nar/29.1.214.
81. 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 Discov.* **2011,** *10* (3), 188—195. https://doi.org/10.1038/nrd3368.
82. Schuler, H. Automation in Chemical Industry (Automatisierung in der Chemischen Industrie). *at—Automatisierungstechnik* **2006,** *54* (8), 363—371. https://doi.org/10.1524/auto.2006.54.8.363.
83. Gromski, P. S.; Granda, J. M.; Cronin, L. Universal Chemical Synthesis and Discovery with 'The Chemputer.'. *Trends Chem.* **2020,** *2* (1), 4—12 https://doi.org/10.1016/j.trechm.2019.07.004.
84. Peplow, M. Organic Synthesis: The Robo-Chemist. *Nature* **2014,** *512* (7512), 20—22. https://doi.org/10.1038/512020a.
85. Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I. A Mobile Robotic Chemist. *Nature* **2020,** *583* (7815), 237—241. https://doi.org/10.1038/s41586-020-2442-2.

86. Roth, A.; Jopp, R.; Schaefer, R.; Kramer, G. W. Automated Generation of AnIML Documents by Analytical Instruments. *NIST* **2006,** *11* (4), 247—253.
87. Jones, P.-J.; Roberts, J.; Vanderwall, D.; Vergis, J. Unlocking the Power of Data. *LCGC N. Am.* **2015,** *33* (4), 270—281.
88. Data Standard | Allotrope Foundation. allotropefoundation. https://www.allotrope.org (Accessed 2023-07-21).
89. Standards | SiLA Rapid Integration. https://sila-standard.com/standards/ (Accessed 2023-07-21).
90. Porr, M.; Schwarz, S.; Lange, F.; Niemeyer, L.; Hentrop, T.; Marquard, D.; Lindner, P.; Scheper, T.; Beutel, S. Bringing IoT to the Lab: SiLA2 and Open-Source-Powered Gateway Module for Integrating Legacy Devices into the Digital Laboratory. *HardwareX* **2020,** *8,* e00118. https://doi.org/10.1016/j.ohx.2020.e00118.
91. LADS—Laboratory and Analytical Device Standard. OPC Foundation. https://opcfoundation.org/markets-collaboration/lads/ (Accessed 2023-07-21).
92. Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. *Nature* **2018,** *555* (7698), 604—610. https://doi.org/10.1038/nature25978.
93. Schwaller, P.; Vaucher, A. C.; Laplaza, R.; Bunne, C.; Krause, A.; Corminboeuf, C.; Laino, T. Machine Intelligence for Chemical Reaction Space. *WIREs Comput. Mol. Sci.* **2022,** *12* (5), e1604. https://doi.org/10.1002/wcms.1604.
94. Angello, N. H.; Rathore, V.; Beker, W.; Wołos, A.; Jira, E. R.; Roszak, R.; Wu, T. C.; Schroeder, C. M.; Aspuru-Guzik, A.; Grzybowski, B. A.; Burke, M. D. Closed-Loop Optimization of General Reaction Conditions for Heteroaryl Suzuki-Miyaura Coupling. *Science* **2022,** *378* (6618), 399—405. https://doi.org/10.1126/science.adc8743.
95. Higgins, S. G.; Nogiwa-Valdez, A. A.; Stevens, M. M. Considerations for Implementing Electronic Laboratory Notebooks in an Academic Research Environment. *Nat. Protoc.* **2022,** *17* (2), 179—189. https://doi.org/10.1038/s41596-021-00645-8.
96. Herres-Pawlis, S.; Bach, F.; Bruno, I. J.; Chalk, S. J.; Jung, N.; Liermann, J. C.; McEwen, L. R.; Neumann, S.; Steinbeck, C.; Razum, M.; Koepler, O. Minimum Information Standards in Chemistry: A Call for Better Research Data Management Practices. *Angew. Chem. Int. Ed.* **2022,** *61* (51), e202203038. https://doi.org/10.1002/anie.202203038.
97. Strömert, P.; Hunold, J.; Castro, A.; Neumann, S.; Koepler, O. Ontologies4Chem: The Landscape of Ontologies in Chemistry. *Pure Appl. Chem.* **2022,** *94* (6), 605—622. https://doi.org/10.1515/pac-2021-2007.
98. Häse, F.; Roch, L. M.; Kreisbeck, C.; Aspuru-Guzik, A. Phoenics: A Bayesian Optimizer for Chemistry. *ACS Cent. Sci.* **2018.** https://doi.org/10.1021/acscentsci.8b00307.
99. Bran, A. M.; Cox, S.; Schilter, O.; Baldassari, C.; White, A. D.; Schwaller, P. ChemCrow: Augmenting Large-Language Models with Chemistry Tools. *arXiv* **June 16, 2023**. http://arxiv.org/abs/2304.05376.
100. Christensen, M.; Yunker, L. P. E.; Shiri, P.; Zepel, T.; Prieto, P. L.; Grunert, S.; Bork, F.; Hein, J. E. Automation Isn't Automatic. *Chem. Sci.* **2021,** *12* (47), 15473—15490. https://doi.org/10.1039/D1SC04588A.
101. Rathoure, D. A.; Ram, B.; Goyal, S. Unit Operations in Chemical Industries. *Int. J. Environ. Chem.* **2019,** *5,* 11—29.
102. van der Aalst, W. M. P. Workflow Model Analysis. In *Encyclopedia of Database Systems;* Liu, L., Özsu, M. T., Eds., Springer US: Boston, MA, 2009; p 3551. https://doi.org/10.1007/978-0-387-39940-9_1476.
103. Workflow Description Language (WDL), 2023. https://github.com/openwdl/wdl (Accessed 2023-08-20).
104. Aspuru-Guzik, A.; Persson, K. Aterials Acceleration Platform: Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods and Artificial Intelligence. Harvard University's DASH repository.
105. Holtze, C.; Boehling, R. Batch or Flow Chemistry?—A Current Industrial Opinion on Process Selection. *Curr. Opin. Chem. Eng.* **2022,** *36,* 100798. https://doi.org/10.1016/j.coche.2022.100798.
106. Taylor, C. J.; Pomberger, A.; Felton, K. C.; Grainger, R.; Barecka, M.; Chamberlain, T. W.; Bourne, R. A.; Johnson, C. N.; Lapkin, A. A. A Brief Introduction to Chemical Reaction Optimization. *Chem. Rev.* **2023.** https://doi.org/10.1021/acs.chemrev.2c00798.
107. González-Esguevillas, M.; Fernández, D. F.; Rincón, J. A.; Barberis, M.; de Frutos, O.; Mateos, C.; García-Cerrada, S.; Agejas, J.; MacMillan, D. W. C. Rapid Optimization of Photoredox Reactions for Continuous-Flow Systems Using Microscale Batch Technology. *ACS Cent. Sci.* **2021.** https://doi.org/10.1021/acscentsci.1c00303.
108. Klucznik, T.; Mikulak-Klucznik, B.; McCormack, M. P.; Lima, H.; Szymkuć, S.; Bhowmick, M.; Molga, K.; Zhou, Y.; Rickershauser, L.; Gajewska, E. P.; Toutchkine, A.; Dittwald, P.; Startek, M. P.; Kirkovits, G. J.; Roszak, R.; Adamski, A.; Sieredzińska, B.; Mrksich, M.; Trice, S. L. J.; Grzybowski, B. A. Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory. *Chem* **2018,** *4* (3), 522—532. https://doi.org/10.1016/j.chempr.2018.02.002.
109. Jensen, K. F. Flow Chemistry—Microreaction Technology Comes of Age. *AIChE J.* **2017,** *63* (3), 858—869. https://doi.org/10.1002/aic.15642.
110. Cherkasov, N.; Bai, Y.; Expósito, A. J.; Rebrov, E. V. OpenFlowChem—a Platform for Quick, Robust and Flexible Automation and Self-Optimisation of Flow Chemistry. *React. Chem. Eng.* **2018,** *3* (5), 769—780. https://doi.org/10.1039/C8RE00046H.
111. Qiu, J.; Albrecht, J. Solubility Correlations of Common Organic Solvents. *Org. Process. Res. Dev.* **2018,** *22* (7), 829—835. https://doi.org/10.1021/acs.oprd.8b00117.
112. Bjelobrk, Z.; Mendels, D.; Karmakar, T.; Parrinello, M.; Mazzotti, M. Solubility Prediction of Organic Molecules with Molecular Dynamics Simulations. *Cryst. Growth Des.* **2021,** *21* (9), 5198—5205. https://doi.org/10.1021/acs.cgd.1c00546.
113. Morato, N. M.; Le, M.; Holden, D. T.; Cooks, R. G. Automated High-Throughput System Combining Small-Scale Synthesis with Bioassays and Reaction Screening. *SLAS Technol.* **2021,** *26,* 555—571. https://doi.org/10.1177/24726303211047839.
114. Schneider, G. Automating Drug Discovery. *Nat. Rev. Drug Discov.* **2018,** *17* (2), 97—113. https://doi.org/10.1038/nrd.2017.232.
115. Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M.; Damon, D. B.; Eugenio de Diego, J.; García-Cerrada, S.; García-Losada, P.; Haro, R.; Janey, J.; Leitch, D. C.; Li, L.; Liu, F.; Lobben, P. C.; MacMillan, D. W. C.; Magano, J.; McInturff, E.; Monfette, S.; Post, R. J.; Schultz, D.; Sitter, B. J.; Stevens, J. M.; Strambeanu, I. I.; Twilton, J.; Wang, K.; Zajac, M. A. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process. Res. Dev.* **2019,** *23* (6), 1213—1242. https://doi.org/10.1021/acs.oprd.9b00140.
116. 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. https://doi.org/10.1146/annurev-chembioeng-060816-101411.
117. Fermier, A. M.; Troisi, J.; Heritage, E. C.; Drexel, M. A.; Gallea, P.; Swinney, K. A. Powder Dispensing Robot for Sample Preparation. *Analyst* **2003,** *128* (6), 790. https://doi.org/10.1039/b300274h.
118. 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. https://doi.org/10.1126/science.1259203.
119. Cernak, T.; Gesmundo, N. J.; Dykstra, K.; Yu, Y.; Wu, Z.; Shi, Z.-C.; Vachal, P.; Sperbeck, D.; He, S.; Murphy, B. A.; Sonatore, L.; Williams, S.; Madeira, M.; Verras, A.; Reiter, M.; Lee, C. H.; Cuff, J.; Sherer, E. C.; Kuethe, J.; Goble, S.; Perrotto, N.; Pinto, S.; Shen, D.-M.; Nargund, R.; Balkovec, J.; DeVita, R. J.; Dreher, S. D. Microscale High-Throughput Experimentation as an Enabling Technology in Drug Discovery: Application in the Discovery of (Piperidinyl)Pyridinyl-1 *H* -Benzimidazole Diacylglycerol Acyltransferase 1 Inhibitors. *J. Med. Chem.* **2017,** *60* (9), 3594—3605. https://doi.org/10.1021/acs.jmedchem.6b01543.
120. Mahjour, B.; Zhang, R.; Shen, Y.; McGrath, A.; Zhao, R.; Mohamed, G.; Lin, Y.; Zhang, Z.; Douthwaite, J. L.; Tripathi, A.; Cernak, T. Rapid Planning and Analysis of High-Throughput Experiment Arrays for Reaction Discovery **2023,** *14.*
121. Bahr, M. N.; Morris, M. A.; Tu, N. P.; Nandkeolyar, A. Recent Advances in High-Throughput Automated Powder Dispensing Platforms for Pharmaceutical Applications. *Org. Process. Res. Dev.* **2020,** *24* (11), 2752—2761. https://doi.org/10.1021/acs.oprd.0c00411.
122. Team, D. D. W. Automation Of Solid/Powder Dispensing: Much needed, but cautiously used!. *Drug Discov. World (DDW)* **2009**. https://www.ddw-online.com/automation-of-solid-powder-dispensing-much-needed-but-cautiously-used-917-200908/.
123. Gensch, T.; Smith, S.; Colacot, T.; Timsina, Y.; Xu, G.; Glasspoole, B.; Sigman, M. Design and Application of a Screening Set for Monophosphine Lig-Ands in Metal Catalysis. preprint *Chemistry* **2021.** https://doi.org/10.33774/chemrxiv-2021-fgm7v.

124. Tu, N. P.; Dombrowski, A. W.; Goshu, G. M.; Vasudevan, A.; Djuric, S. W.; Wang, Y. High-Throughput Reaction Screening with Nanomoles of Solid Reagents Coated on Glass Beads. *Angew. Chem.* **2019,** *131* (24), 8071−8075. https://doi.org/10.1002/ange.201900536.
125. Jiang, T.; Bordi, S.; McMillan, A. E.; Chen, K.-Y.; Saito, F.; Nichols, P.; Wanner, B.; Bode, J. An Integrated Console for Capsule-Based, Fully Automated Organic. *Synthesis* **2019.** https://doi.org/10.26434/chemrxiv.7882799.v1.
126. Wang, G.; Ang, H. T.; Dubbaka, S. R.; O'Neill, P.; Wu, J. Multistep Automated Synthesis of Pharmaceuticals. *Trends Chem.* **2023,** *5* (6), 432−445. https://doi.org/10.1016/j.trechm.2023.03.008.
127. Fasano, V.; Mykura, R. C.; Fordham, J. M.; Rogers, J. J.; Banecki, B.; Noble, A.; Aggarwal, V. K. Automated Stereocontrolled Assembly-Line Synthesis of Organic Molecules. *Nat. Synth.* **2022,** 1−6. https://doi.org/10.1038/s44160-022-00158-6.
128. Hird, N. W. Automated Synthesis: New Tools for the Organic Chemist. *Drug Discov. Today* **1999,** *4* (6), 265−274. https://doi.org/10.1016/S1359-6446(99)01337-9.
129. Sun, A. C.; Steyer, D. J.; Allen, A. R.; Payne, E. M.; Kennedy, R. T.; Stephenson, C. R. J. A Droplet Microfluidic Platform for High-Throughput Photochemical Reaction Discovery. *Nat. Commun.* **2020,** *11* (1), 6202. https://doi.org/10.1038/s41467-020-19926-z.
130. 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. https://doi.org/10.1002/anie.201410744.
131. 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. https://doi.org/10.1126/science.aax1566.
132. Sugimoto, A.; Fukuyama, T.; Rahman, T. M.; Ryu, I. An Automated-Flow Microreactor System for Quick Optimization and Production: Application of 10- and 100-Gram Order Productions of a Matrix Metalloproteinase Inhibitor Using a Sonogashira Coupling Reaction. *Tetrahedron Lett.* **2009,** *50* (46), 6364−6367. https://doi.org/10.1016/j.tetlet.2009.08.089.
133. Liu, J.; Hein, J. E. Automation, Analytics and Artificial Intelligence for Chemical Synthesis. *Nat. Synth.* **2023,** *2* (6), 464−466. https://doi.org/10.1038/s44160-023-00335-1.
134. Chiang, L.; Lu, B.; Castillo, I. Big Data Analytics in Chemical Engineering. *Annu. Rev. Chem. Biomol. Eng.* **2017,** *8* (1), 63−85. https://doi.org/10.1146/annurev-chembioeng-060816-101555.
135. Himanen, L.; Geurts, A.; Foster, A. S.; Rinke, P. Data-Driven Materials Science: Status, Challenges, and Perspectives. *Adv. Sci.* **2019,** *6* (21), 1900808. https://doi.org/10.1002/advs.201900808.
136. Shiri, P.; Lai, V.; Zepel, T.; Griffin, D.; Reifman, J.; Clark, S.; Grunert, S.; Yunker, L. P. E.; Steiner, S.; Situ, H.; Yang, F.; Prieto, P. L.; Hein, J. E. Automated Solubility Screening Platform Using Computer Vision. *iScience* **2021,** *24* (3), 102176. https://doi.org/10.1016/j.isci.2021.102176.
137. Janey, J. M. Measuring Solubility Automatically with Vision. *Chem* **2021,** *7* (5), 1151−1153. https://doi.org/10.1016/j.chempr.2021.04.007.
138. Beker, W.; Gajewska, E. P.; Badowski, T.; Grzybowski, B. A. Prediction of Major Regio-, Site-, and Diastereoisomers in Diels−Alder Reactions by Using Machine-Learning: The Importance of Physically Meaningful Descriptors. *Angew. Chem. Int. Ed.* **2019,** *58* (14), 4515−4519. https://doi.org/10.1002/anie.201806920.
139. Szymkuc, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem. Int. Ed.* **2016,** *55* (20), 5904−5937. https://doi.org/10.1002/anie.201506101.
140. Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C. A.; Bekas, C.; Lee, A. A. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Cent. Sci.* **2019,** *5* (9), 1572−1583. https://doi.org/10.1021/acscentsci.9b00576.
141. Prieto Kullmer, C. N.; Kautzky, J. A.; Krska, S. W.; Nowak, T.; Dreher, S. D.; MacMillan, D. W. C. An Expedited Phenotypic Approach towards Organic Reaction Generality. preprint *Chemistry* **2021.** https://doi.org/10.33774/chemrxiv-2021-8f5ws.
142. Nedyalkova, M. A.; Madurga, S.; Tobiszewski, M.; Simeonov, V. Calculating the Partition Coefficients of Organic Solvents in Octanol/Water and Octanol/Air. *J. Chem. Inf. Model.* **2019,** *59* (5), 2257−2263. https://doi.org/10.1021/acs.jcim.9b00212.
143. Boobier, S.; Hose, D. R. J.; Blacker, A. J.; Nguyen, B. N. Machine Learning with Physicochemical Relationships: Solubility Prediction in Organic Solvents and Water. *Nat. Commun.* **2020,** *11* (1), 5753. https://doi.org/10.1038/s41467-020-19594-z.
144. Lee, S.; Lee, M.; Gyak, K.-W.; Kim, S. D.; Kim, M.-J.; Min, K. Novel Solubility Prediction Models: Molecular Fingerprints and Physicochemical Features vs Graph Convolutional Neural Networks. *ACS Omega* **2022,** *7* (14), 12268−12277. https://doi.org/10.1021/acsomega.2c00697.
145. Hansen, C. M. *Hansen Solubility Parameters: A User's Handbook,* 2nd ed.; CRC Press, 2007.
146. Hansen, C. M. The Universality of the Solubility Parameter. *Product. R&D* **1969,** *8* (1), 2−11. https://doi.org/10.1021/i360029a002.
147. Wojeicchowski, J. P.; Ferreira, A. M.; Okura, T.; Pinheiro Rolemberg, M.; Mafra, M. R.; Coutinho, J. A. P. Using COSMO-RS to Predict Hansen Solubility Parameters. *Ind. Eng. Chem. Res.* **2022,** *61* (42), 15631−15638. https://doi.org/10.1021/acs.iecr.2c01592.
148. Keeler, J. Understanding NMR Spectroscopy.
149. Zheng, G.; Price, W. S. Solvent Signal Suppression in NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010,** *56* (3), 267−288. https://doi.org/10.1016/j.pnmrs.2010.01.001.
150. Ahmad, I. A. H.; Losacco, G. L.; Shchurik, V.; Wang, X.; Cohen, R. D.; Herron, A. N.; Aiken, S.; Fiorito, D.; Wang, H.; Reibarkh, M.; Nowak, T.; Makarov, A. A.; Stoll, D. R.; Guillarme, D.; Mangion, I.; Aggarwal, V. K.; Yu, J.; Regalado, E. L. Trapping-Enrichment Multi-dimensional Liquid Chromatography with On-Line Deuterated Solvent Exchange for Streamlined Structure Elucidation at the Microgram Scale. *Angew. Chem. Int. Ed.* **2022.** https://doi.org/10.1002/anie.202117655.
151. Liska, I. On-Line versus Off-Line Solid-phase Extraction in the Determination of Organic Contaminants in Water: Advantages and Limitations. *J. Chromatogr. A* **1993,** *655* (2), 163−176. https://doi.org/10.1016/0021-9673(93)83220-M.
152. Bharti, S.; Roy, R. Quantitative 1H NMR Spectroscopy. *TrAC Trends Anal. Chem.* **2012,** *35,* 5−26. https://doi.org/10.1016/j.trac.2012.02.007.
153. Cullen, C. H.; Ray, G. J.; Szabo, C. M. A. Comparison of Quantitative Nuclear Magnetic Resonance Methods: Internal, External, and Electronic Referencing. *Magn. Reson. Chem.* **2013,** *51* (11), 705−713. https://doi.org/10.1002/mrc.4004.
154. Muhamadejev, R.; Melngaile, R.; Paegle, P.; Zibarte, I.; Petrova, M.; Jaudzems, K.; Veliks, J. Residual Solvent Signal of CDCl3 as a qNMR Internal Standard for Application in Organic Chemistry Laboratory. *J. Org. Chem.* **2021,** *86* (5), 3890−3896. https://doi.org/10.1021/acs.joc.0c02744.
155. ANSI_SLAS_1-2004_FootprintDimensions.Pdf. https://www.slas.org/SLAS/assets/File/public/standards/ANSI_SLAS_1-2004_FootprintDimensions.pdf (Accessed 2024-02-12).
156. Guo, J.; Chen, C.; Chen, C.; Zuo, E.; Dong, B.; Lv, X.; Yang, W. Near-Infrared Spectroscopy Combined with Pattern Recognition Algorithms to Quickly Classify Raisins. *Sci. Rep.* **2022,** *12* (1), 7928. https://doi.org/10.1038/s41598-022-12001-1.
157. Fadlelmoula, A.; Pinho, D.; Carvalho, V. H.; Catarino, S. O.; Minas, G. Fourier Transform Infrared (FTIR) Spectroscopy to Analyse Human Blood over the Last 20 Years: A Review towards Lab-On-A-Chip Devices. *Micromachines* **2022,** *13* (2), 187. https://doi.org/10.3390/mi13020187.
158. Maschmeyer, T.; Yunker, L. P. E.; Hein, J. E. Quantitative and Convenient Real-Time Reaction Monitoring Using Stopped-Flow Benchtop NMR. *React. Chem. Eng.* **2022,** *10,* 1039. https://doi.org/10.1039/D2RE00048B.
159. Silva Elipe, M. V.; Milburn, R. R. Monitoring Chemical Reactions by Low-Field Benchtop NMR at 45 MHz: Pros and Cons. *Magn. Reson. Chem.* **2016,** *54* (6), 437−443. https://doi.org/10.1002/mrc.4189.
160. Dalitz, F.; Cudaj, M.; Maiwald, M.; Guthausen, G. Process and Reaction Monitoring by Low-Field NMR Spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **2012,** *60,* 52−70. https://doi.org/10.1016/j.pnmrs.2011.11.003.

161. Moore, J.; Basurto-Lozada, D.; Besson, S.; Bogovic, J.; Brown, E. M.; Burel, J.-M.; Medeiros, G. de; Diel, E. E.; Gault, D.; Ghosh, S. S.; Gold, I.; Halchenko, Y. O.; Hartley, M.; Horsfall, D.; Keller, M. S.; Kittisopikul, M.; Kovacs, G.; Yoldaş, A. K.; Villegeorges, A.; le, T. de la; Li, T.; Liberali, P.; Linkert, M.; Lindner, D.; Lüthi, J.; Maitin-Shepard, J.; Manz, T.; McCormick, M.; Mohamed, K.; Moore, W.; Özdemir, B.; Pape, C.; Pelkmans, L.; Prete, M.; Pietzsch, T.; Preibisch, S.; Rzepka, N.; Stirling, D. R.; Striebel, J.; Tischer, C.; Toloudis, D.; Walczysko, P.; Watson, A. M.; Wong, F.; Yamauchi, K. A.; Bayraktar, O.; Haniffa, M.; Saalfeld, S.; Swedlow, J. R. OME-zarr: A Cloud-Optimized Bioimaging File Format with International Community Support. *bioRxiv* **February 25, 2023.** https://doi.org/10.1101/2023.02.17.528834.
162. *Three Laws of Robotics*, Wikipedia, 2023.
163. 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,* 160018. https://doi.org/10.1038/sdata.2016.18.
164. Fukushima, E. *Experimental Pulse NMR: A Nuts and Bolts Approach*, CRC Press: Boca Raton, 2019. https://doi.org/10.1201/9780429493867.
165. Digital to Analog Conversion—an overview | ScienceDirect Topics. https://www.sciencedirect.com/topics/engineering/digital-to-analog-conversion (Accessed 2023-10-14).
166. Johnson, S. G. *Fast Fourier Transform Algorithms*, MIT IAP, 2008.
167. Fast Fourier Transform Algorithm - an overview | Science Direct Topics. https://www.sciencedirect.com/topics/engineering/fast-fourier-transform-algorithm (Accessed 2023-10-14).
168. Huang, Y.-C.; Tremouilhac, P.; Nguyen, A.; Jung, N.; Bräse, S. ChemSpectra: A Web-Based Spectra Editor for Analytical Data. *J. Cheminformatics* **2021,** *13* (1), 8. https://doi.org/10.1186/s13321-020-00481-0.
169. Lacey, R. F. Deconvolution of Overlapping Chromatographic Peaks. *Anal. Chem.* **1986,** *58* (7), 1404—1410. https://doi.org/10.1021/ac00298a029.
170. Häckl, M.; Tauber, P.; Schweda, F.; Zacharias, H. U.; Altenbuchinger, M.; Oefner, P. J.; Gronwald, W. An R-Package for the Deconvolution and Integration of 1D NMR Data: MetaboDecon1D. *Metabolites* **2021,** *11* (7), 452. https://doi.org/10.3390/metabo11070452.
171. Li, D.-W.; Hansen, A. L.; Yuan, C.; Bruschweiler-Li, L.; Brüschweiler, R. DEEP Picker Is a Deep Neural Network for Accurate Deconvolution of Complex Two-Dimensional NMR Spectra. *Nat. Commun.* **2021,** *12* (1), 5229. https://doi.org/10.1038/s41467-021-25496-5.
172. Li, D.-W.; Hansen, A. L.; Bruschweiler-Li, L.; Yuan, C.; Brüschweiler, R. Fundamental and Practical Aspects of Machine Learning for the Peak Picking of Biomolecular NMR Spectra. *J. Biomol. NMR* **2022,** *76* (3), 49—57. https://doi.org/10.1007/s10858-022-00393-1.
173. Asnin, L. D. Peak Measurement and Calibration in Chromatographic Analysis. *TrAC Trends Anal. Chem.* **2016,** *81,* 51—62. https://doi.org/10.1016/j.trac.2016.01.006.
174. Melnikov, A. D.; Tsentalovich, Y. P.; Yanshole, V. V. Deep Learning for the Precise Peak Detection in High-Resolution LC—MS Data. *Anal. Chem.* **2020,** *92* (1), 588—592. https://doi.org/10.1021/acs.analchem.9b04811.
175. Isaacman-VanWertz, G.; Sueper, D. T.; Aikin, K. C.; Lerner, B. M.; Gilman, J. B.; de Gouw, J. A.; Worsnop, D. R.; Goldstein, A. H. Automated Single-Ion Peak Fitting as an Efficient Approach for Analyzing Complex Chromatographic Data. *J. Chromatogr. A* **2017,** *1529,* 81—92. https://doi.org/10.1016/j.chroma.2017.11.005.
176. Lipfert, M.; Rout, M. K.; Berjanskii, M.; Wishart, D. S. Automated Tools for the Analysis of 1D-NMR and 2D-NMR Spectra. In *NMR-based Metabolomics: Methods and Protocols;* Gowda, G. A. N., Raftery, D., Eds.; *Methods in Molecular Biology*, Springer: New York, NY, 2019; pp 429—449. https://doi.org/10.1007/978-1-4939-9690-2_24.
177. Newman-Stonebraker, S. H.; Smith, S. R.; Borowski, J. E.; Peters, E.; Gensch, T.; Johnson, H. C.; Sigman, M. S.; Doyle, A. G. Univariate Classification of Phosphine Ligation State and Reactivity in Cross-Coupling Catalysis. *Science* **2021,** *374* (6565), 301—308. https://doi.org/10.1126/science.abj4213.
178. Slattery, A.; Wen, Z.; Tenblad, P.; Sanjosé-Orduna, J.; Pintossi, D.; den Hartog, T.; Noël, T. Automated Self-Optimization, Intensification, and Scale-Up of Photocatalysis in Flow. *Science* **2024,** *383* (6681), eadj1817. https://doi.org/10.1126/science.adj1817.
179. Centre for Rapid Online Analysis of Reactions (ROAR). Imperial College London. https://www.imperial.ac.uk/a-z-research/rapid-online-analysis-of-reactions/ (Accessed 2024-02-12).
180. Automated high-throughput platform suite (ATLAS) | Research groups | Imperial College London. https://www.imperial.ac.uk/automated-high-throughput-platform-suite/ (Accessed 2024-02-12).
181. Bristol, U. of Bristol Automated Synthesis Facility. https://www.bristol.ac.uk/chemistry/facilities/basf/ (Accessed 2024-02-12).
182. Szymanski, N. J.; Rendy, B.; Fei, Y.; Kumar, R. E.; He, T.; Milsted, D.; McDermott, M. J.; Gallant, M.; Cubuk, E. D.; Merchant, A.; Kim, H.; Jain, A.; Bartel, C. J.; Persson, K.; Zeng, Y.; Ceder, G. An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials. *Nature* **2023,** *624* (7990), 86—91. https://doi.org/10.1038/s41586-023-06734-w.
183. Acceleration Consortium. https://acceleration.utoronto.ca/ (Accessed 2024-02-12).
184. Laveille, P.; Miéville, P.; Chatterjee, S.; Clerc, E.; Cousty, J.-C.; Nanteuil, F. de; Lam, E.; Mariano, E.; Ramirez, A.; Randrianarisoa, U.; Villat, K.; Copéret, C.; Cramer, N. Swiss CAT+, a Data-Driven Infrastructure for Accelerated Catalysts Discovery and Optimization. *CHIMIA* **2023,** *77* (3), 154. https://doi.org/10.2533/chimia.2023.154.
185. Leonov, A. I.; Hammer, A. J. S.; Lach, S.; Mehr, S. H. M.; Caramelli, D.; Angelone, D.; Khan, A.; O'Sullivan, S.; Craven, M.; Wilbraham, L.; Cronin, L. An Integrated Self-Optimizing Programmable Chemical Synthesis and Reaction Engine. *Nat. Commun.* **2024,** *15* (1), 1240. https://doi.org/10.1038/s41467-024-45444-3.
186. Grzybowski, B. A.; Szymkuć, S.; Gajewska, E. P.; Molga, K.; Dittwald, P.; Wołos, A.; Klucznik, T. Chematica: A Story of Computer Code that Started to Think like a Chemist. *Chem* **2018,** *4* (3), 390—398. https://doi.org/10.1016/j.chempr.2018.02.024.
187. IBM RoboRXN | Science | IBM Research IBM Research—Project Photoresist. https://research.ibm.com/science/ibm-roborxn/ (Accessed 2024-02-13).
188. Jablonka, K. M.; Schwaller, P.; Ortega-Guerrero, A.; Smit, B. Leveraging Large Language Models for Predictive Chemistry. *Nat. Mach. Intell.* **2024.** https://doi.org/10.1038/s42256-023-00788-1.
189. Boiko, D. A.; MacKnight, R.; Kline, B.; Gomes, G. Autonomous Chemical Research with Large Language Models. *Nature* **2023,** *624* (7992), 570—578. https://doi.org/10.1038/s41586-023-06792-0.
190. Maloney, M. P.; Coley, C. W.; Genheden, S.; Carson, N.; Helquist, P.; Norrby, P.-O.; Wiest, O. Negative Data in Data Sets for Machine Learning Training. *Org. Lett.* **2023,** *25* (17), 2945—2947. https://doi.org/10.1021/acs.orglett.3c01282.
191. Taniike, T.; Takahashi, K. The Value of Negative Results in Data-Driven Catalysis Research. *Nat. Catal.* **2023,** *6* (2), 108—111. https://doi.org/10.1038/s41929-023-00920-9.
192. Nandy, A.; Duan, C.; Kulik, H. J. Audacity of Huge: Overcoming Challenges of Data Scarcity and Data Quality for Machine Learning in Computational Materials Discovery. *Curr. Opin. Chem. Eng.* **2022,** *36,* 100778. https://doi.org/10.1016/j.coche.2021.100778.
193. Murray, P. M.; Bellany, F.; Benhamou, L.; Bučar, D.-K.; B. Tabor, A.; Sheppard, D. T. The Application of Design of Experiments (DoE) Reaction Optimisation and Solvent Selection in the Development of New Synthetic Chemistry. *Org. Biomol. Chem.* **2016,** *14* (8), 2373—2384. https://doi.org/10.1039/C5OB01892G.
194. Shields, B. J. Bayesian Reaction Optimization as a Tool for Chemical Synthesis **2021,** *12.*
195. Ramirez, A.; Lam, E.; Gutierrez, D. P.; Hou, Y.; Tribukait, H.; Roch, L. M.; Copéret, C.; Laveille, P. Accelerated Exploration of Heterogeneous CO2 Hydrogenation Catalysts by Bayesian-Optimized High-Throughput and Automated Experimentation. *Chem. Catal.* **2024,** 100888. https://doi.org/10.1016/j.checat.2023.100888.
196. Nambiar, A. M. K.; Breen, C. P.; Hart, T.; Kulesza, T.; Jamison, T. F.; Jensen, K. F. Bayesian Optimization of Computer-Proposed Multistep Synthetic Routes on an Automated Robotic Flow Platform. *ACS Cent. Sci.* **2022,** *8* (6), 825—836. https://doi.org/10.1021/acscentsci.2c00207.
197. Hein, J. E. Machine Learning Made Easy for Optimizing Chemical Reactions. *Nature* **2021,** *590* (7844), 40—41. https://doi.org/10.1038/d41586-021-00209-6.
198. Fleischer, H.; Thurow, K. *Automation Solutions for Analytical Measurements: Concepts and Applications*, John Wiley & Sons, 2017.

199. Ayres, L. B.; Gomez, F. J. V.; Linton, J. R.; Silva, M. F.; Garcia, C. D. Taking the Leap between Analytical Chemistry and Artificial Intelligence: A Tutorial Review. *Anal. Chim. Acta* **2021,** *1161,* 338403. https://doi.org/10.1016/j.aca.2021.338403.
200. Haas, C. P.; Lübbesmeyer, M.; Jin, E. H.; McDonald, M. A.; Koscher, B. A.; Guimond, N.; Di Rocco, L.; Kayser, H.; Leweke, S.; Niedenführ, S.; Nicholls, R.; Greeves, E.; Barber, D. M.; Hillenbrand, J.; Volpin, G.; Jensen, K. F. Open-Source Chromatographic Data Analysis for Reaction Optimization and Screening. *ACS Cent. Sci.* **2023,** *9* (2), 307−317. https://doi.org/10.1021/acscentsci.2c01042.
201. 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. https://doi.org/10.1039/D0SC00442A.
202. Pesek, M.; Juvan, A.; Jakoš, J.; Košmrlj, J.; Marolt, M.; Gazvoda, M. Database Independent Automated Structure Elucidation of Organic Molecules Based on IR, 1H NMR, 13C NMR, and MS Data. *J. Chem. Inf. Model.* **2021,** *61* (2), 756−763. https://doi.org/10.1021/acs.jcim.0c01332.
203. Kwon, Y.; Lee, D.; Choi, Y.-S.; Kang, M.; Kang, S. Neural Message Passing for NMR Chemical Shift Prediction. *J. Chem. Inf. Model.* **2020,** *60* (4), 2024−2030. https://doi.org/10.1021/acs.jcim.0c00195.