Emerging Trends in Multi-Objective Optimization of Organic Synthesis Leveraging High-throughput Tools and Machine Learning Methods

Pablo Quijano Velasco,[1] Kedar Hippalgaonkar*[1,2] and Balamurugan Ramalingam*[1,3]

*Correspondence to kedar@ntu.edu.sg; balamurugan_ramalingam@imre.a-star.edu.sg

[1] Institute of Materials Research and Engineering (IMRE), Agency for Science Technology and Research (A*STAR), 2 Fusionopolis Way, Singapore 138634, Republic of Singapore.

[2] Department of Materials Science and Engineering, Nanyang Technological University, Singapore 639798, Republic of Singapore

[3] Institute of Sustainability for Chemicals, Energy and Environment (ISCE^2), Agency for Science Technology and Research (A*STAR), 1 Pesek Road, Jurong Island, Singapore 627833, Republic of Singapore.

Abstract

The discovery of optimal conditions of chemical reactions is a labor-intensive, time-consuming task that requires exploring a high-dimensional parametric space. Historically the optimization of chemical reactions has been performed by manual experimentation guided by human intuition and Design of Experiments where one reaction variable is modified at any one time to find optimal conditions for one specific reaction outcome. Recently a change of paradigm in chemical reaction optimization procedures has been enabled by the advances in lab automation and the introduction of machine learning algorithms, where multiple reaction variables can be synchronously optimized to obtain optimal reaction conditions requiring a shorter amount of experimental time and minimal human intervention. Herein, we review the state-of-the-art of high-throughput automated chemical reaction platforms and machine learning algorithms that are currently used to drive the optimization of chemical reactions, highlighting the limitations and future opportunities that this new field of research encounters.

1. Introduction

Organic synthesis plays a crucial role in drug discovery, polymer synthesis, material science, agrochemicals, and specialty chemicals. Their synthesis and process optimization require substantial resources and are labour-intensive, often performed by exploring a single variable in search of optimal conditions while disregarding the intricate interactions among competing variables within the synthesis process. The complexity of the problem highly increases considering that process optimization often requires solutions that meet multiple targets, such as yield, selectivity, purity, cost, environmental impact, etc. In recent years, the advancement of artificial intelligence (AI), machine learning (ML) and automation has produced a shift in paradigm for chemical synthesis optimization techniques. By leveraging on ML models to predict reaction outcomes and ML optimization algorithms, this new framework has demonstrated the ability to navigate the complex relationships between reaction variables and to find global optimal conditions in fewer number of experiments than
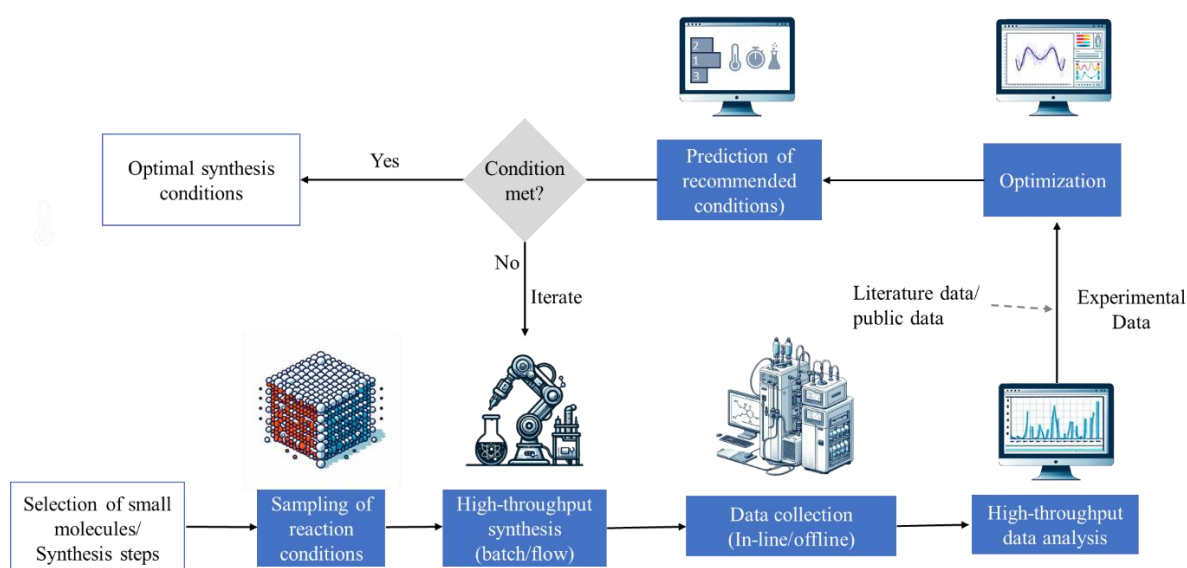
Figure 1: A high-level workflow for a typical multi-objective optimization (MOO) framework for the rapid optimization of small molecule synthesis.

traditional methods [1,2]. In addition, machine-guided multi-objective optimization (MOO) has emerged as a promising framework to obtain reaction conditions that perform optimally for multiple target objectives, enabling researchers to explore diverse solution spaces and uncover optimal conditions that strike a balance between consonant and/or conflicting targets. In addition, the incorporation of lab robotics into chemical synthesis has enabled the development of closed-loop optimization platforms capable of executing optimization campaigns rapidly with minimal human intervention, relieving experimenters from labour-intensive tasks and reducing the overall process development lead time [3,4].

A standard workflow and general methodology for organic reaction optimization through ML methods is shown in Figure 1. The workflow comprises (i) careful design of experiments; (ii) reaction execution with commercial high-throughput systems or in-house designed reaction modules; (iii) data collection by inline/offline analytical tools; (iv) applying optimization algorithms on collected data points to map with the target objectives; (v) prediction of optimal solutions and (vi) experimental testing of suggested optimization results. Through an examination of methodologies, algorithms, and various case studies, the article offers our perspective into the state-of-the-art techniques for optimizing synthesis of organic molecules highlighting both challenges and prospects. The structure of this perspective follows the steps presented in Figure 1 where in Section 2 we review the high-throughput platforms currently used to perform chemical reaction optimization. Section 3 discusses the techniques and developments on analytical tools and data processing algorithms. Section 4, discusses the latest trends in the selection of optimization algorithms for chemical synthesis. Finally, Section 5 we highlight the future directions and opportunities in the field. For an in-depth review on the topics of chemical reaction optimization, the readers can gain diverse insights from prominent reviews by Taylor et al. [5], Griffin et al. [6] and Sagmeister et al. [7]. The first two offer valuable perspectives on chemical reaction optimization, particularly focusing towards process scale-up, while the latter discusses

the potential of flow platforms for conducting self-optimization reactions. Additionally, we refer the readers to the following literature in other areas relevant to the application of ML to chemical synthesis that is not covered by this perspective, such as small molecule discovery [8], drug discovery [9,10], retrosynthesis [11,12], and catalyst selection/design [13,14].

2. High-throughput experimentation platforms

High-throughput experimentation (HTE) platforms were designed to accelerate the discovery and development of organic molecules by the rapid screening and analysis of large numbers of experimental conditions simultaneously. For the purpose of this article, we define HTE as a technique that leverages a combination of automation, parallelization of experiments, advanced analytics, and data processing methods to streamline repetitive experimental tasks, reduce manual intervention and increase the rate of experimental execution in comparison to traditional manual experimentation. In conventional chemical synthesis, several sequential steps are typically undertaken, involving the setup of the reaction, mixing of reactants, reaction workup, product analysis, and product purification. To perform all of these basic chemistry tasks adequately, customizable HTE platforms are available from various laboratory instrument manufacturers or can be assembled from a combination of commercial and in-house developed equipment. Normally a HTE for organic chemistry will include a liquid transfer module, a reactor stage, and analytical tools for product characterization. When the full experimental process is automated and coupled with a centralized control system performing ML optimization, the HTE can function as a self-driving platform where the next iteration of experiments is automatically selected by algorithm without human intervention. This section will highlight the key features of various HTE platforms, benefits, limitations, and applications to organic molecule synthesis.

2.1. HTE in batch modules

Batch reactions are defined as chemical reaction vessels where there is no flow of the reagents/products in or out of the reaction vessel until a target reaction conversion is obtained. HTE batch platforms leverage on parallelization of experiments to perform several reaction conditions simultaneously in order to increase the experimental throughput. Commonly, batch platforms that include a liquid handling system for setting up reactions based on a plunger pump (*e.g.* syringe, pipette), a reactor capable of heating and mixing, and inline/online analytical tools. Many HTE batch experiments have been performed in self-contained automated platforms developed by various instrument manufacturers (ChemSpeed, Zinsser Analytics, Metler-Toledo, Tecan, etc.). In these HTE platforms, microtiter well plates (MTP) and reaction blocks containing 96/48/24-well plates are widely used as reaction and characterization vessels [5]. UltraHTE configurations typically incorporate 1536-well plates, enabling the exploration of lager parametric spaces of reaction parameters. While UltraHTE was initially tailored for biological assays, the versatility of these modules has extended to optimizing chemistry-related processes [5]. The Chemspeed SWING robotic system equipped with two fluoropolymers and PFA mat-sealed 96-well metal blocks was used for the exploration of stereoselective Suzuki-Miyaura coupling,

offering precise control over both categorical and continuous variables (Figure 2a) [15]. The integrated robotic system containing a four-needle dispense head facilitated the delivery of reagents in low volumes and slurries, ensuring accuracy and throughout of the process. The entire experimental workflow was further optimized through parallelization, dividing reactions into loops of eight, which enabled them to complete 192 reactions within 24 loops, achieving significant throughput in four days. Other reports for various reactions include the Buchwald-Hartwig amination [16–19], Suzuki coupling [16,17,20], N-alkylation [21], hydroxylation [22], and photochemical reactions [23–29]. The adaptability and widespread availability of 96-well plates have facilitated their extensive adoption in optimizing chemical synthesis methodologies. The details of these developments have been reviewed recently [5].

In recent years research laboratories have deviated from the traditional commercial tools to custom-build HTE systems for the chemist's requirements and demands. Burger *et al.* [30]. have creatively developed a mobile robot equipped with sample-handling arms, tailored for the precise execution of photocatalytic water molecule
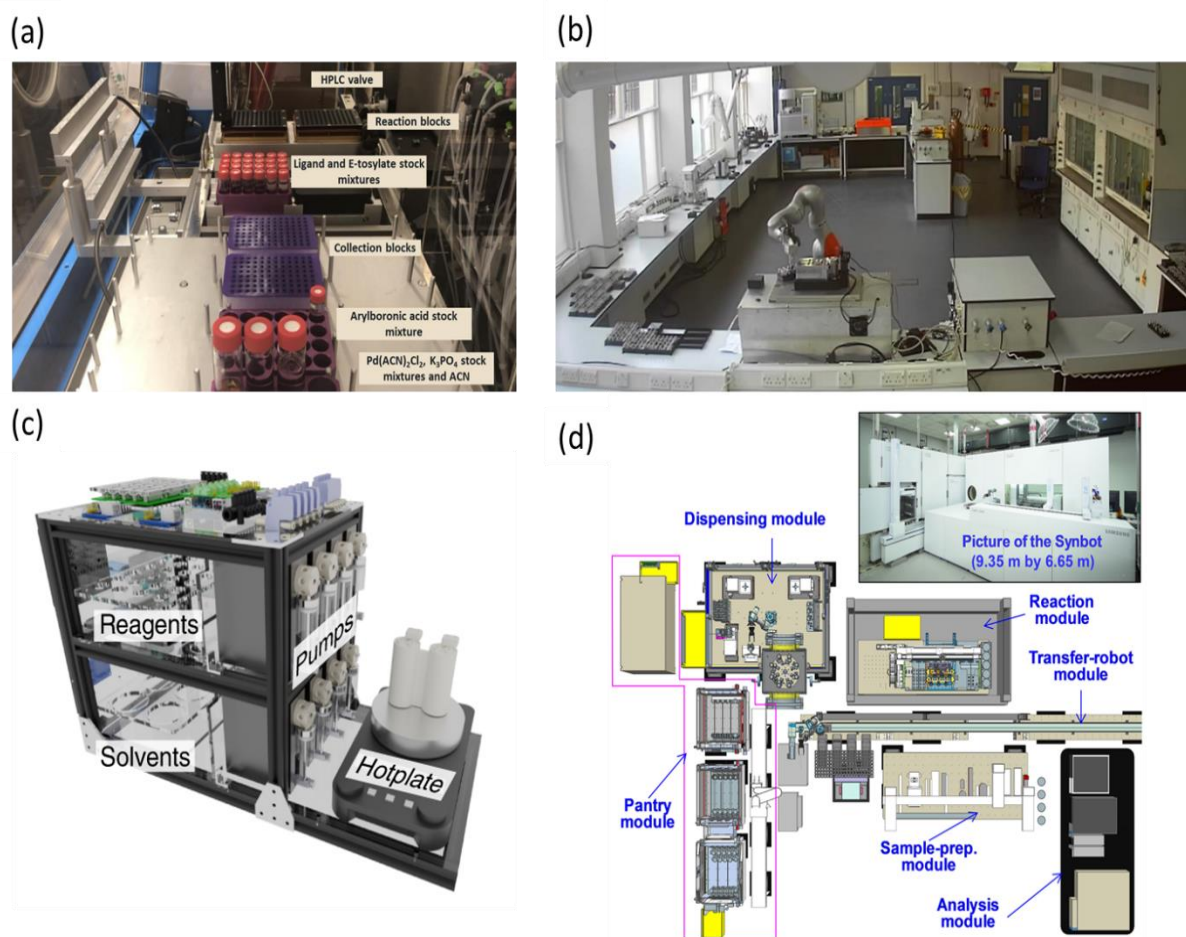

Figure 2: (a) Photograph showing a ChemSpeed HTE platform using 96 well reaction blocks (adapted from [15]). (b) Mobile robot equipment performing tasks normally executed by human experimenter for the photocatalytic conversion of water to hydrogen (adapted from [30]). (c) Small-footprint portable chemical synthesis platform (adapted from [31]). (d) Schematic of the Synbot platform developed by Samsung researchers showing each module used for chemical synthesis (adapted from [33]).

cleavage to produce hydrogen. The mobile robot (Figure 2b) acted as a substitute of a human experimenter by executing tasks and linking eight separate experimental stations, including solid and liquid dispensing, sonication, several characterization equipment and stations for consumables and sample storage. Remarkably, through a tedious ten-dimensional parameter search spanning 8 days, the robot achieved an impressive hydrogen evolution rate of approximately 21.05 μmol h^{-1}. Despite the initial investment and two-year development timeline, the versatility of this robotic system promises remarkable applications in materials, polymers, and chemical synthesis. Most automated synthesis platforms are based on expensive scientific equipment, require a large footprint and require extensive reconfiguration to adapt to new synthetic protocols. To address this issues Manzano *et al*. [31] have developed a small-footprint portable chemical synthesis platform able to perform liquid and solid phase organic reactions (Figure 2c). The platform utilizes 3D-printed reactors that can be generated on-demand depending the targeted reaction, and features liquid handling, stirring, heating, and cooling modules for enhanced versatility. In addition, the platform is capable of performing inert and low-pressure atmospheres, separation steps, and pressure sensing for reaction monitoring. Its efficacy and robustness was affirmed through the successful synthesis of five small organic molecules, four oligopeptides, and four oligonucleotides, yielding high purities and impressive yields. Although, in the current configuration the platform lacks characterization modules and has lower throughputs in comparison to other automated platforms, it does offer a low-cost alternative that can be adapted to perform chemical reaction optimization.

In addition to academia, industries are increasingly recognizing the value of investing in custom-built HTE setups to automate their synthesis workflows for enhancing the productivity. A fully integrated, cloud accessible, automated synthesis laboratory (ASL) was designed and built by Eli Lilly [32]. This state-of-the-art facility comprises heating, cryogenic, microwave, and high-pressure reaction, evaporation, and work-up modules, empowering researchers to conduct an extensive array of chemical reactions. The ASL comprises of three bench spaces dedicated to either high temperature reaction, cryogenic/microwave reactions, or reaction workup. In each bench a translational combination of robotic arms perform the specific experiments using the modular platforms, while consumables and samples are transferred between benches through a conveyor belt linking them together. According to their article, the ASL has facilitated over 16,350 reactions at gram scale across various case studies, showcasing their widespread capability. Researchers at Samsung have pioneered the development of Synbot an innovative autonomous synthesis robot by leveraging artificial intelligence (AI) and robotic technology to establish optimal synthetic recipes [33]. Similar to ASL, Synbot consists in five modules connected through a conveyor belt backbone with a robot arm in charge of transferring the samples between them. The modules include a pantry for chemical storage and selection, a dispensing module for solids and liquids, reaction module capable of heating and stirring, sample preparation module, and a liquid chromatography-mass spectrometer characterization module (Figure 2d). The efficiency of the system has been demonstrated through successful validations on three reactions namely, Suzuki-Miayura coupling, Buchwald-Hartwig amination, Ullmann coupling. These experiments showcased conversion rates that outperformed existing references and provided at least 6 times efficiency in

the experimentation besides synthesis planning, optimization, and downstream workup tasks. The throughput of Synbot is estimated to be an average of 12 reactions within 24 hours depends on the reaction time. IBM RoboRXN has developed a remotely accessible autonomous chemical laboratory that enables 100 times acceleration in chemical synthesis by leveraging cloud computing, AI and automation. The details of this pioneering project can be found on their website.

2.3. HTE using flow platforms

Flow reactions are described as the flow of reagents and products in and out of the reaction vessel continuously. The flow platform consists of a fluid delivery system, mixing tools, reactors, quenching units, pressure regulation units, and collection vessels. The fluid delivery is normally executed using either HPLC, syringe or peristatic pumps. A passive mixing stage where the reagents are introduced to the system through a Y or T connection is the most common approach observed for most flow reactions, more specialized mixing tools can be incorporated depending on the reaction requisites. The most common reactors used are either microfluidic chip or coil-based reactors for solution chemistry. Packed bed reactors are used when solid heterogenous catalysts are handled. Specialized reactors for electro [34–36] and photochemical [37–39] experiments have also been developed. Depending on the flow of the reaction mixture, flow reactions can be continuous or segmented (also known as slug). Segmented flow reactions present an efficient means to gather diverse data points by creating segmented or droplet flow within microfluidic reactors. Each droplet is carefully separated by either an antisolvent or an inert gas, thus providing every droplet with the functionality of an individual reactor. This segmentation ensures precise control over reactions and prevents interference between different reaction environments. Moreover, the ratios of reagents within these droplets are easily modulated using syringe pumps, providing users with a convenient means to collect data efficiently and coherently. This approach streamlines experimentation processes, enhances reproducibility, and facilitates the exploration of complex reaction spaces with unprecedented accuracy.

Droplet microfluidics has emerged as a powerful tool across diverse scientific disciplines, with dedicated literature offering concepts behind the droplet formation [40,41]. An example of a segmented flow droplet system was employed to screen a range of organic solvents for achieving optimal mono-alkylation of *trans*-1,2-diamino cyclohexane [42]. The HTE methodology in combination with feedback DoE facilitated the rapid identification of prime solvents, notably DMSO, DMF, and pyridine, leading to enhanced yields of the mono-alkylated product. An experimental setup was developed for single-droplet studies of visible-light photo redox catalysis in this case using an oscillatory flow strategy [43,44]. In an oscillatory reactor an alternating pressure gradient is applied within the reactor causing a back-and-forth oscillation of the reaction slugs, which leads to higher control in mixing and extended residence time of the reaction mixture. About 150 reaction conditions were explored using a total volume of 4.5 mL reaction mixture and the screening results are readily translated to continuous flow synthesis. The application of segmented flow or micro-slug reactor was demonstrated in the decarboxylative arylation cross-coupling reaction promoted by catalysts and light [39]. The design allows the screening to be more material and

time-efficient in the optimization of both continuous variables (e.g., temperature and residence time) and discrete variables (e.g., catalyst. base). Pieber *et al.* [45] reported the application of a segmental flow reactor for heterogenous solid-liquid reactions. In their report, they described the reaction slugs as serial micro-batch reactors (SMBRs) separated through gas segments that incorporated liquid reagents and solid photocatalysts in a continuous flow. The slugs were generated by establishing a stable gas–liquid segmented-flow pattern using a Y-shaped mixer, followed by the suspension of the catalyst via a T-mixer. This technology was utilized to develop selective and efficient decarboxylative fluorination reactions. Recently, a slug flow platform was developed (Figure 3a) by injecting segments of gas as a separating medium for enhancing the optimization of the Buchwald–Hartwig amination intermediate, which is crucial for synthesizing the drug olanzapine [46]. The reactor setup was integrated with spectroscopic and chromatographic in-line analytical tools, enabling the real-time monitoring of products and reaction intermediates. A detailed discussion on the optimization strategy is described in section 4 (Table 2, entry 19). Robochem a HTE platform was designed to streamline the screening of photochemical reactions, facilitating the rapid generation of diverse reaction mixtures each comprising 650 uL within a slug flow reactor [47]. This innovative system features precise monitoring of the reaction slug through a dedicated array of phase sensors and an algorithm designed for detecting its passage. As a result, the workflow delivers a notable boost in productivity, surpassing traditional batch reactions by over 500-fold and outperforming flow reactions with a five-fold improvement. A fully integrated automated multistep chemical synthesizer (AutoSyn) was reported to autonomously
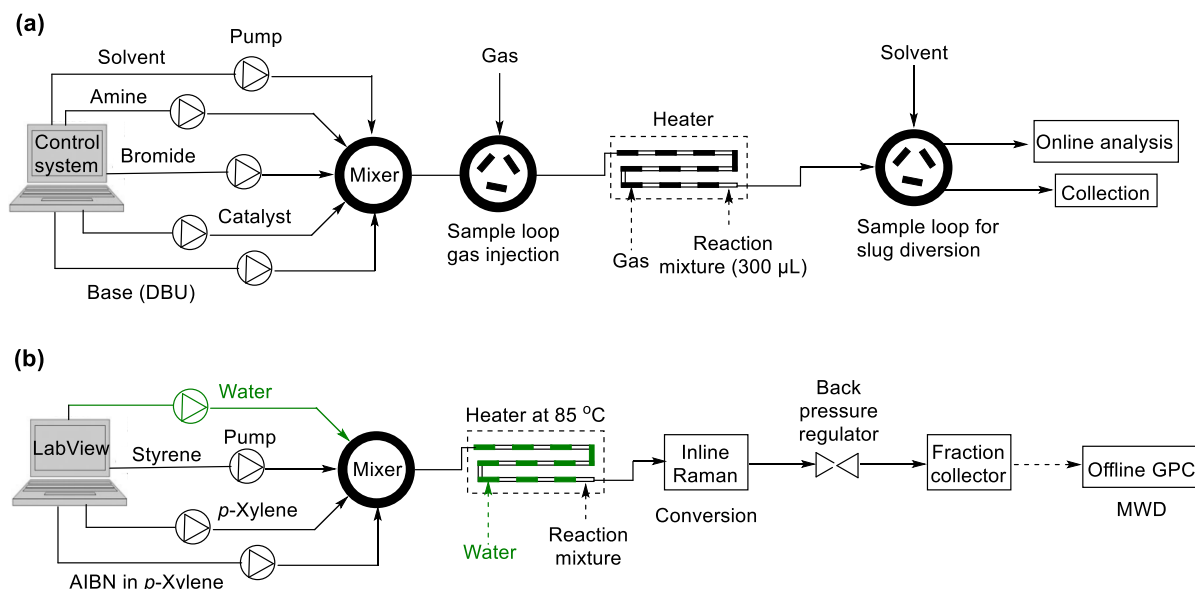

Figure 3: (a) Description of a slug flow platform developed using segments of gas as a separation medium for HT data collection Buchwald–Hartwig amination. 6-way mixer was used to mix the solvents and reagents. (b) Schematic representation a computer-controlled segmented flow pattern developed using degassed water as an antisolvent for the HT polymerization of styrene in *p*-Xylene. A staggered infusion of organic and aqueous phases resulted in the exploration of wider parameter space.

synthesis milligram-to-gram-scale amounts of any organic molecules or drug-like molecules [48]. The system comprised a flow chemistry synthesis platform, a reagent delivery system, a packed bed reactor, and process analytical tools. An integrated software control system that automates end-to-end process operations and monitoring. The system has been used to demonstrate the synthesis of at least 10 drug molecules autonomously and there is no closed-loop optimization framework is embedded.

In addition to organic synthesis, the slug-flow methodology (SFM) has found application in polymer synthesis and is worth citing to understand SFM's wider utility. A flow platform capable of polymerizing 397 unique copolymer compositions was developed by Reis *et al*. [49] using a droplet flow reactor. The methodology and high-fidelity data enable them to discover >10 copolymer compositions of promising ^{19}F MRI agents that outperformed state-of-the-art materials. A rapid generation of copolymer libraries was achieved by forming droplet-flow in an automated HTE flow setup [50]. This approach not only assists in overcoming the challenges of viscosity in conventional photopolymerisation reaction, but also generates structure-property relationships for co-polymer libraries. We have generated a segmented flow pattern (Figure 3b) by alternating the infusion of organic components and degassed water to create 9 different compositions [51]. The organic components consisting of styrene, AIBN (α,α′-Azobisisobutyronitrile), and *p*-Xylene were infused using a computer-controlled segmented-flow platform. These approaches allow the compartmentalization of reaction mixtures without cross-contamination and enhance experimental throughput significantly.

The major bottleneck in HTE synthesis lies in the challenge of isolating and purifying reaction products once experiments are performed. Despite this bottleneck, the landscape is evolving, with various practical tools emerging to streamline purification processes. From prepacked silica gel tubes to the precision semi-preparative liquid chromatography, and the versatile capabilities of various scavenger resins, laboratories are witnessing a surge in options for efficient high-throughput purification, particularly in chemical synthesis on a modest scale. A change in thinking beyond conventional purification methods presents an opportunity to revolutionize HTE flow platforms. A completely novel design, differing from established isolation and separation techniques, holds the promise of not only enhancing the efficiency of HTE flow synthesis but also paving the way for more sustainable growth in this research area.

2.4. Autonomous self-optimizing flow reactor

Autonomous self-optimizing flow reactors (ASFRs) represent a promising advancement in high-throughput chemical synthesis, combining principles of automation, artificial intelligence, in-line analytics, and robotics to streamline and accelerate the process of creating molecules. ASTR increases the throughput of experimentations in flow reactions by engaging inline/online analytics, ML-based optimizations and subsequent prediction of the next experimental conditions. The schematic representation of ASTR is given in figure 4a.
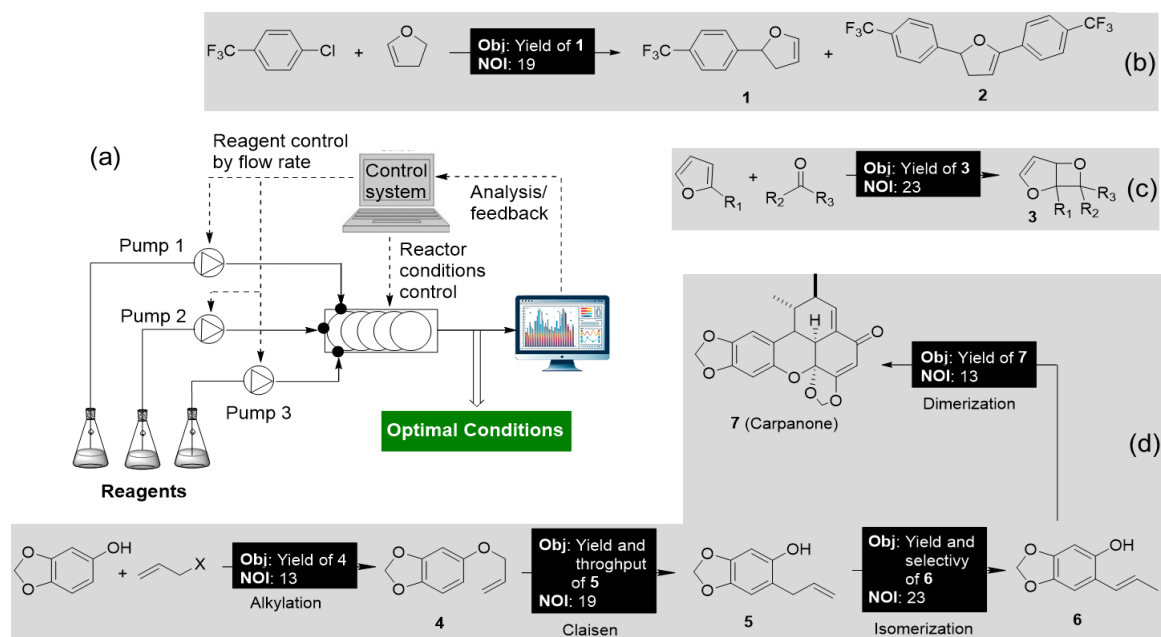
Figure 4: (a) Schematic representation of autonomous self-optimizing flow reactor (ASFR) for obtaining optimal solution with minimal human intervention. Selected case studies (b-d) with closed-loop optimization are provided. Abbreviations 'Obj' and 'NOI' represent 'objective functions' and 'number of interactions' respectively.

A self-optimizing microreactor system has been devised specifically for closed-loop optimization of the Heck reaction, employing a "black-box" optimization strategy directed by Nelder–Mead Simplex method algorithm [52]. In-line HPLC analysis was performed to determine the product yield in real time and feedback to the control system to direct the input conditions to achieve optimum product yield in 19 automated experiments. The optimum conditions identified in microfluidics for the formation of monoarylated product (**1**, Figure 4b) are successfully translated to a meso-fluidics system at a 50-fold scale to afford the product **1** in 26.9 g yield. LeyLab, a modular software system developed by Fitzpatrick *et al.* [53], allows researchers to oversee chemical reactions online. The hydration of 3-cyanopyridine to its amide was monitored by online mass spectrometer offering real-time conversion. Through 30 experiments within 10 hours, five key reaction parameters were finely tuned for optimal conditions. Photochemical reactions require uniform light penetration to the reaction mixture and flow setups with uniform path lengths would be ideal for such reactions. A self-optimizing continuous-flow reactor was designed by Poscharny *et al.* [49] for [2 + 2] cycloaddition reaction promoted by light. The optimization (modified simplex) algorithm arrives the optimal conditions in 25 iterative experiments to afford compound **3** (Figure 3c) in good yield. A modular autonomous flow reactor controlled via MATLAB was designed for the carpanone (**7**, Figure 4d) synthesis using a modified Nelder−Mead algorithm [50]. The four-step process involves allylation, Claisen rearrangement, isomerization and oxidative dimerization. Each reaction steps were optimized independently by using either online HPLC or in-line benchtop NMR spectroscopy to afford an overall yield of 67 % in 66 iterative experiments over four

linear reaction steps. Nadiwale *et al.* [56] reported the autonomous optimization of three multiphase catalytic reactions involving the handling of solid substrates, photoreactor, feeding of slurries, catalysts, and inorganic bases in an automated flow platform comprising a CSTR cascade. The platform allows them to showcase the autonomous optimization of ideal reaction conditions for Suzuki–Miyaura and photoredox-catalysed coupling reactions.

A plug-and-play, continuous-flow chemical synthesis system (Figure 5a) was intelligently designed by Bédard *et al.* [57] to mitigate the challenges in organic synthesis by the integration of hardware, software, and analytics. Comprising an array of modular components, including units for heating, cooling, LED light exposure, and packed bed reactors, it provides a flexible platform for various reaction categories. The system consists of a liquid-liquid separator and an inline/online analytical tool to facilitate closed-loop autonomous optimization. The capability of the system was demonstrated in the optimization of C-C and C-N cross-coupling, olefination, reductive amination, photoredox catalysis, nucleophilic aromatic substitution, and a two-step synthesis of cyclobutanone. The molecules synthesized under optimal conditions are presented in Figure 5b, employing the stable noisy optimization by branch and fit (SNOBFIT) algorithm. SNOBFIT offers a convenient methodology for global optimization, eliminating the necessity of a theoretical model. A reconfigurable automated flow platform integrating online HPLC monitoring was used for the cobalt-catalyzed aerobic oxidative dimerization of desmethoxycarpacine to carpanone in the presence of oxygen as an oxidant [58]. A gas−liquid segmented, or tube-in-tube strategy adopted for achieving higher yield in a shorter residence time. Substantial further developments have been made in applying ASFR in multi-objective optimizations which will be discussed under the section 4 in detail.
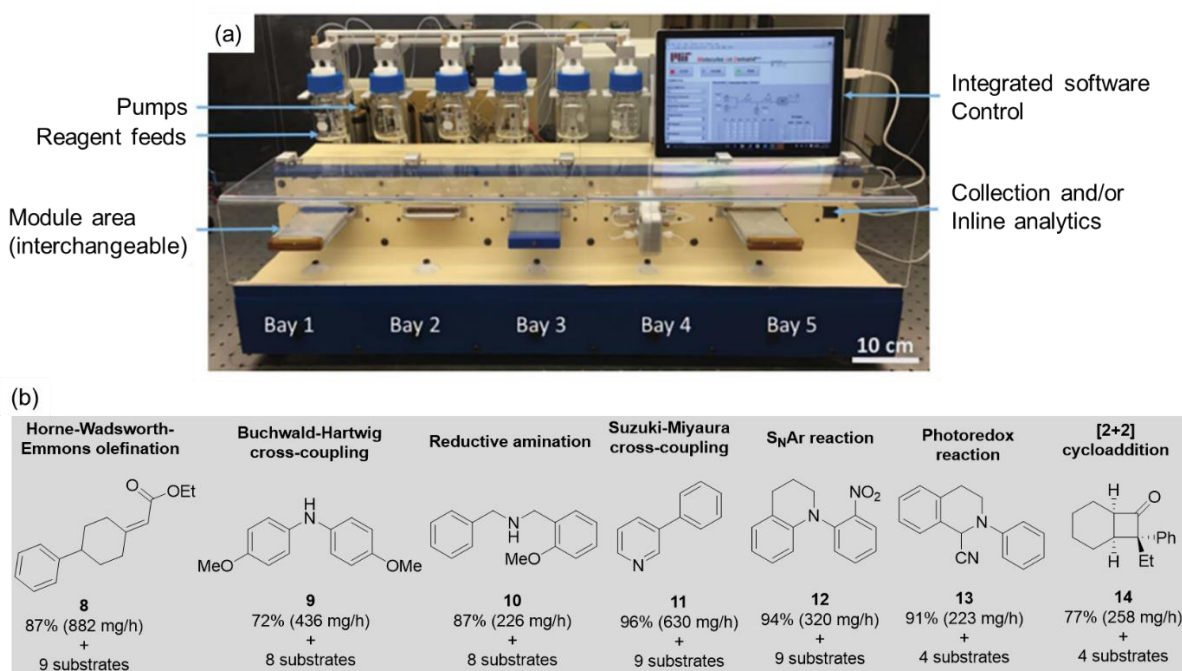

Figure 5: (a) A modular flow platform developed for wider variety of chemical synthesis (adapted from [57]). (b) Various categories of chemical reactions optimized and molecules synthesized in a continuous flow system are given.

3. Real-time analytics and HT data processing

Real-time analytics plays a critical role in the optimization of chemical reactions via high-throughput synthesis and ML algorithms. Process analytical technology (PAT) tools empower researchers to obtain chemical insights from large number of experiments, facilitating the precise measurement of optimization targets. The integration of real-time analysis within high-throughput experiments presents a multitude of advantages over traditional, one-time final product evaluations outlined below.

(i)   Real-time analysis facilitates rapid decision-making, enabling researchers to continuously monitor and analyze data as it is generated and allowing for immediate adjustments to process parameters during experiments.

(ii) Early detection of trends or anomalies are made possible through real-time analysis, providing valuable insights that can guide subsequent experiments and inform iterative improvements and optimizations in experimental protocols.

(iii) By optimizing experimental workflows and minimizing waste through real-time analysis, researchers can allocate resources more efficiently, ensuring that resources are utilized effectively to maximize experimental outcomes.

(iv) Enhanced experimental control on the process to deliver constant product quality to meet desired specifications and standards.

(v)   By providing instantaneous feedback, real-time analysis accelerates the optimization process, reducing the experimental time and expediting the discovery of optimal reaction conditions with minimum material use.

Analytical tools are integral components of high-throughput platforms and are found in various configurations such as inline, online, atline, and offline, contingent upon their placement within the experimental workflow. In Table 1 we describe the subtle disparities for clarity and reference.

Table 1: Different analysis methods depending on their placement on the experimental workflow.

Analysis methods	Description
Inline	Analysed in real-time directly during the reaction or production process by integrating appropriate devices directly.
Online	Sampling and analysis take place while the reaction or process is running. The analysis will be done on a device located nearby. Online analyses can be carried out continuously or at set intervals.
Atline	Like online analysis, the samples are analyzed usually within a manufacturing facility. Atline analysis still provides relatively rapid results compared to offline methods, offering a balance between real-time monitoring and convenience.
Offline	Analysis conducted outside of the process environment and separate from ongoing operations. Provide a more detailed and comprehensive analyse compared to real-time monitoring.

Self-optimizing HTE throughput platforms require inline and/or online characterization, data analysis and processing for rapid optimization of organic reactions. Chromatographic (*i.e.* HPLC, GC) and spectroscopic (*e.g.* NMR, FTIR, UV-Vis, Raman) characterization methods are commonly used in real-time reaction monitoring. To quantify the products of a chemical reaction a calibration curve is required before the optimization campaign. The following sequential steps are typically employed to refine raw data into actionable inputs for building ML models for optimization. (i) Extraction and categorization of appropriate spectra; (ii) Fitting of spectral peaks utilizing predefined functional models, alongside deconvolution of overlapping signals; (iii) Consolidation of extracted peak information and generation of relevant data plots and (iv) extracting the relevant information and formatting into input data for ML models. A recent review from Felpin *et al.* [59] highlighted the selection of in-line/online analytical tools that can be integrated into flow reactors for the monitoring of chemical reactions. In the current review, we focus on the HT data processing that complements the HTE platforms for rapid optimization of organic reactions. Although multivariate data analysis has been adopted in analytical chemistry frequently for rapid data processing, the availability of relevant open-source code is relatively low [60,61]. Consequently, the development of open-source code for data processing become interest to the scientific community extensively.

Jansen *et al.* [62] have developed a tool for the analysis of high-pressure liquid chromatography (HPLC) measurements termed as 'HappyTools', able to calibrate retention times, perform peak quantitation, and use various quality criteria to curate the compiled data. For the quantitation and calibration of the chromatographic peaks the user can either input a peak list containing the time retention time and retention time window of the target chemicals, or it can use an automated peak detection algorithm removing the need of user input. The peak detection algorithm was developed using a loop to attain the user-specified cut-off value of the highest intensity peak. A new univariate spline is fitted for each iteration, from which the local maxima and minima are determined. Overall, HappyTools showed similar or better performance in comparison to existing commercial software. In particular, HappyTools showed an enhanced throughput demonstrating up to a 10-fold reduction in the total processing time for biopharmaceutical samples. The authors have released the source code and an executable program in an online repository to be employed freely for research purposes.

In addition to HappyTools, there are other available open source python packages to analyse chromatographic and spectroscopic data. A cross-platform python package named Aston can be used to process both UV-Vis and mass spectrometry chromatographic data. The open-source library is written using Python, Numpy, and Scipy is openly hosted in an online repository [63]. Similarly, for processing chromatographic data from GC-FID, HPLC-UV, or HPLC-FD packages also available open source. Embedding these codes into HTE and ML workflow dramatically improves the efficiency and speed of the optimization processes significantly. Liu et al., [64] developed a custom-built Python script to study the kinetics of carbonyldiimidazole (CDI) mediated amide formation by analysing data from an online HPLC and inline FT-IR spectroscopic measurements. Their algorithm was able to automatically detect peaks from chromatographic spectra and to automatically assign the peaks as reagents or products depending if the peak intensity decreased or increased respectively over time. In addition to monitoring the evolution of the

reaction, the IR spectral data was processed in real-time to ensure the complete consumption of acid reactant and feedback to the pump for immediate quenching of CDI to prevent any side reactions. The entire process allows to control the acid activation and amide formation precisely to afford the desired final product in quantitative yield.

Recently Sagmeister et al. [65] assembled four complementary PAT including inline NMR, UV/Vis, IR, and online UHPLC, to meticulously monitor the intricate three-step linear synthesis of the mesalazine drug (Figure 6) with a 1.6 g h^{-1} throughput. In the first step, the nitration reaction was monitored by inline NMR. The overlapping peaks were resolved for accurate quantification by building a chemometric model. The model also allowed for flexibility to small changes in peak positions and shapes in repetitive analysis. An in-house designed flow cell equipped with a reflectance probe was employed for real-time monitoring of hydrolysis by inline UV-Vis spectroscopy. The raw data were processed using a sophisticated neural network (NN) algorithm, yielding rapid quantification with an impressive processing time of 1.4 ms per spectrum. This streamlined approach ensured efficient and timely data analysis, facilitating seamless real-time monitoring of the hydrolysis of **16**. The final step hydrogenation progress was monitoroted by inline IR probe. The spectral data was processed using a partial least squares (PLS) regression model and quantified. An online ultra-high performance liquid chromatography (UHPLC) was used to analyze the final composition of the reaction mixture after three reaction steps. The integration of all the PAT tools into the three-step reaction was carefully executed with an open platform communication unified architecture (OPCUA) platform for inter-platform equipment communication. The adoption of the OPCUA platform, ensured seamless communication among different equipment platforms for enhanced efficiency and accuracy in data analysis.

A recent study introduced a novel approach for processing and analyzing HPLC−DAD raw data directly using Python [66]. This method leverages the Multivariate Online Contextual Chromatographic Analysis (MOCCA) package, designed for integration for both automated and manual workflows. MOCCA offers a range of benefits, including automated management of internal standards for precise relative quantification, reliable peak assignments, accelerated sample processing, and efficient deconvolution of overlapping peaks. Its versatility was showcased through the successful completion of four comprehensive case studies, demonstrating its broad applicability across diverse analytical scenarios. Recently, we implemented inline Raman spectroscopy to monitor the real-time conversion of styrene to polystyrene, utilizing a custom Python package developed in-house [51]. This approach enabled us to track the conversion process across different residence times. Specifically, we quantified the conversion by analysing the area under the curve of Raman-active vibrational modes associated with the styrene vinyl C=C stretch (~ 1630 cm^{-1}), which
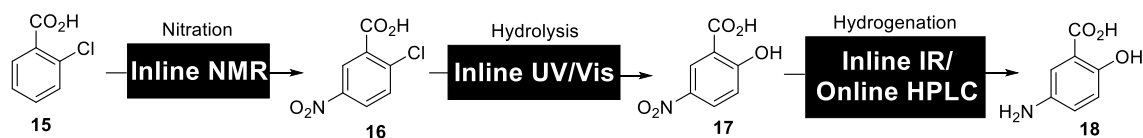

Figure 8: Engagement of four complementary PATs in a three-step synthesis optimization process.

we calibrated against signals from p-xylene (~ 830 cm^{-1}). To resolve overlapping peaks, we employed curve-fitting techniques utilizing Lorentzian functional forms, facilitated by the lmfit Python package. This methodology (Figure 7) allowed us to accurately calculate conversion rates and to make precise predictions using ML models.
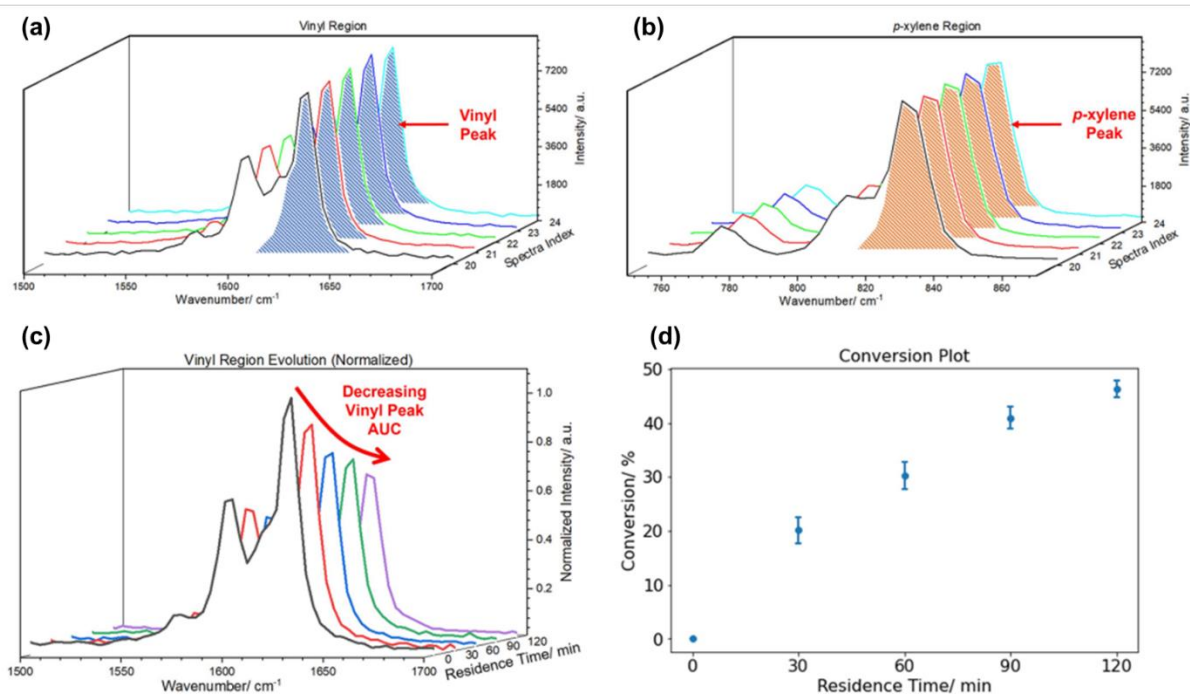

Figure 7: Overlay of several Raman spectra of a single condition featuring the styrene vinyl region (a). *p*-xylene (b). (c) waterfall plot depicting the decrease in the vinyl peak area under the curve (AUC) over time. (d) a representative conversion plot shows an increasing conversion with residence time (adapted from [51]).

Traditionally, the optimization of a chemical reaction, the development kinetic models and optimization of analytical characterization parameters are undertaken independently. With this approach many overlapping tasks are performed in parallel, thus leading to long lead-times and inefficient manpower allocation. To overcome these redundancies, Sagmeister et al. [67] developed a dual modelling approach using a single platform that seamlessly integrates the calibration of PAT, reaction optimization, kinetic modelling, and parametrizes a process model for scale-up in approximately 8 hours. Their platform consisted of a flow reactor connected to a in-line FTIR. In addition, the platform has two valves that allows a stream of reagents or target product to bypass the reactor coil directly into the in-line FTIR to perform. Using this configuration the platform can perform a calibration of the reagent and product concentration through a standard addition method. Once the PAT is calibrated the platform performs dynamic experiments where the concentration of the reagents are ramped to explore the parametric space. Finally, using scientific programming language called Julia they use the collected data to fit the kinetic model parameters and perform in-silico optimization of the reaction parameters. The efficacy of the dual

modeling approach was validated through the attainment of Pareto fronts for both amidation and alkylation reactions.

4. Machine learning driven optimization of chemical reactions

Historically, optimization of chemical reactions has been performed using methodologies derived from the design of experiments (DOE) with the objective of maximizing the yield of the reaction product. However, these techniques are not well-suited to find the global optimal conditions and scale exponentially with the number of variables. Computational approaches that rely on optimization algorithms offer more efficient methods to obtain optimal conditions without requiring an exponential number of experiments per optimization variable. Early examples of optimization of chemical reaction conditions through computational approaches focused on the application in "black-box" optimization algorithms such as steepest descent, SNOBFIT and Nelder-Mead Simplex demonstrated positive results and the ability to perform self-optimizing automated workflows with little human intervention [52,52,53,55,57,68–70]. In recent years, the incorporation of ML optimization methods has demonstrated the ability to obtain optimal reaction conditions in a reduced number of experiments in comparison to human intuition, traditional DOE and other "black-box" optimization algorithms [2,71,72]. Opposed to previous examples of optimization algorithms, the ML approach is characterized by the construction of predictive models that map relationships between the reaction conditions and the target optimization objectives. In this section we review the latest developments of ML optimization strategies for the optimization of chemical reactions.

Figure 8a outlines the basic steps for the optimization of chemical reactions using ML methods. The workflow requires an initial set of experimental data that contains different variables for reaction conditions (*i.e.* temperature, time, solvent, catalyst, etc.) and the corresponding outcome values for the target optimization objectives (*e.g.* yield, purity, cost, etc.). The initial data set is commonly obtained by sampling a combination of reaction variables from the parametric space, performing the synthesis experiments under the selected reaction conditions, and measuring the values for the target optimization objectives. The sampling of initial reaction variables often is performed through near-random statistical methods such as Latin hypercube sampling (LHS), Sobol sampling, full factorial sampling, and centre point sampling methods. Alternatively, the initial data set can be obtained from values previously reported in the literature. After, one or various predictive models are fitted to the initial data set to predict the expected values of the optimization objectives, the number of models that are fitted depends on the number of optimization objectives, and normally one model is constructed for each optimization objective. The next step involves the application of an optimization algorithm to find the most likely parameters that would lead to optimal outcomes for the target optimization objectives. Finally, a set of the most promising suggestions is selected and tested experimentally. The data set is then updated with the outcomes of the latest experimental parameters and the process is repeated until the optimal conditions have been found. Depending on the number of objectives, optimization campaigns are classified as single-objective (Figure 8b) or multi-objective optimizations (Figure 8c). In single-objective optimizations, the algorithm will explore the parametric space to find the optimal conditions by finding the

variables that either maximize or minimize the target objective function. In multi-objective optimization, the algorithms will search for optimal conditions that either maximize or minimize each objective. On the other hand, when competing objectives are optimized the algorithm aims to discover the set of solutions where the improvement in one of the objectives results in the deterioration of the other objective. This set of solutions is called the Pareto front of the system (also known as non-dominated solutions), all other solutions that are not part of the Pareto front are not optimal for any of the objectives and are referred to as dominated solutions. Since all solutions in the Pareto front are optimal the users hold the responsibility to choose the set of conditions for their specific application.
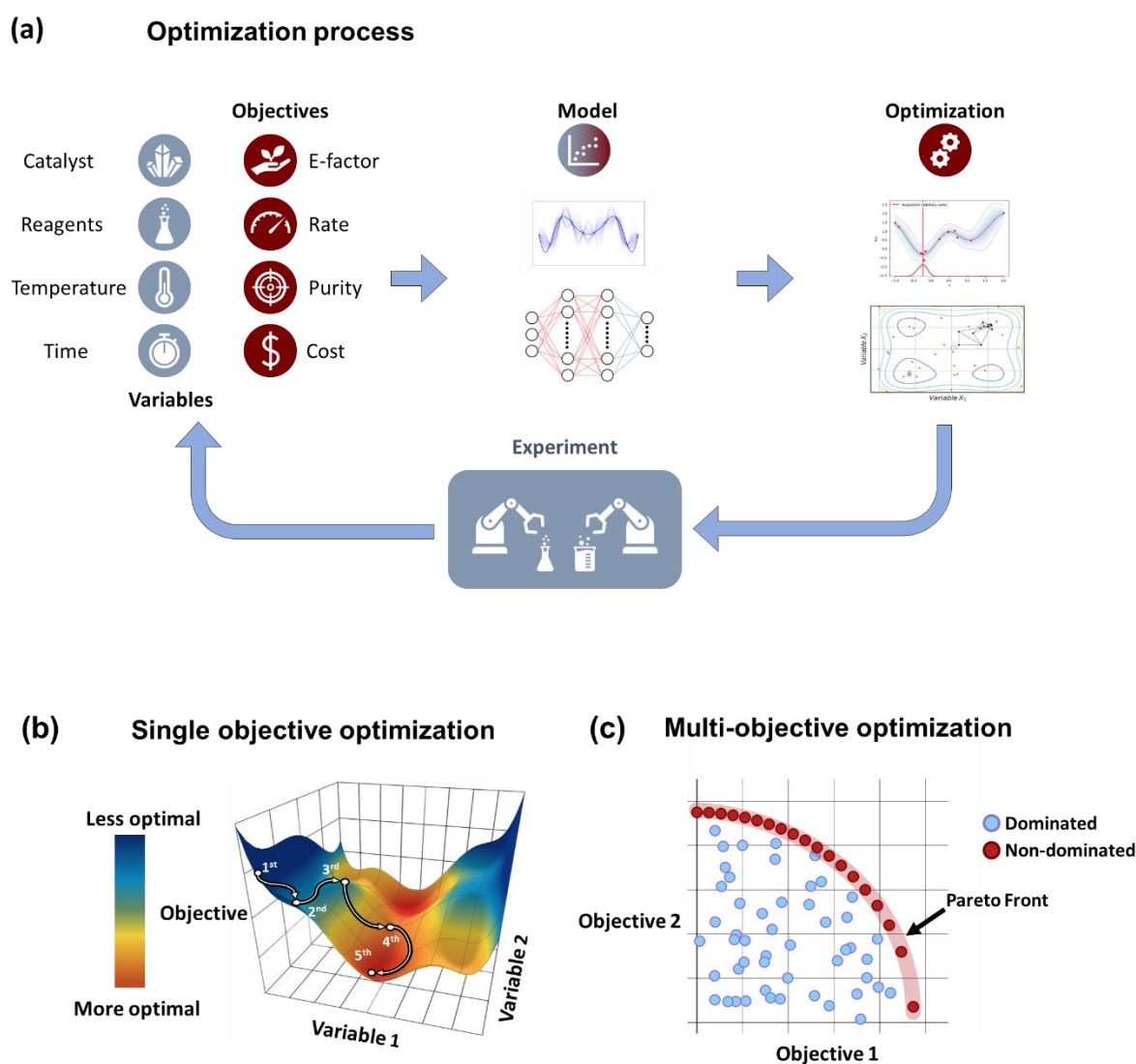

Figure 8: (a) Schematic describing the process of chemical reaction optimization through machine learning methods. (b) 3D Representation of the objective function in relationship to two variables showing the path of 5 optimization iterations that look to minimize the value of the objective function. (c) Representation of the outcomes of a multi-objective optimization campaign. Each data point represents one experimental reaction condition. The Pareto front of the system where the improvement of one of objective leads to the deterioration of the other is highlighted in red.
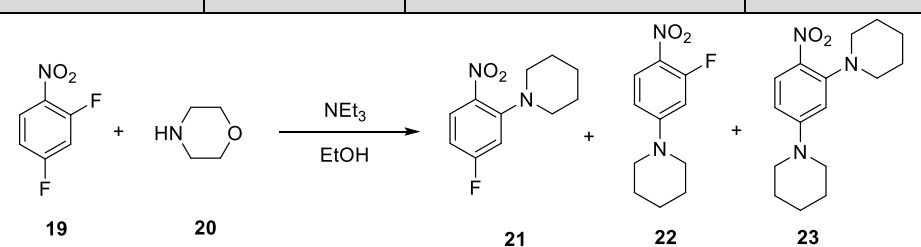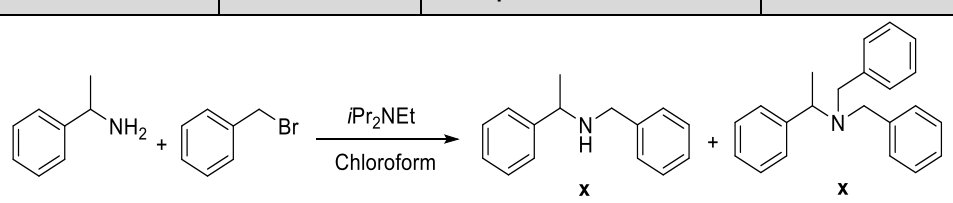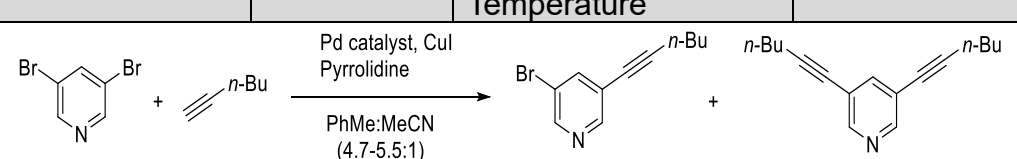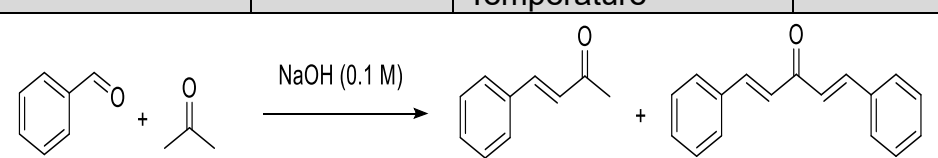
The first reports of the application of ML to the optimization of chemical reactions can be traced over 20 years ago. A handful of studies used ML algorithms such as neural networks and support vector machines to fit models to chemical reaction data that then were optimized by a genetic algorithms [73–75]. However, the use of ML for chemical reaction optimization did not popularize until the introduction of Bayesian Optimization (BO) techniques by Lapkin and Bourne's research group [76]. BO is a global optimization method that fits a probabilistic function to model the optimization objective and utilizes it to search for parameters that will likely lead to optimal objective values. Commonly, BO will use Gaussian process (GP) to create surrogate models that maps the relationships between the variables and objectives (Figure 8a). Then the surrogate model is sampled and the output values are passed to an acquisition function that balances the surrogate model predictions and uncertainties to find variable combinations that are likely to lead to optimal solutions (Figure 8a). The application of GP and BO to optimization of chemical reactions has the advantages of being able to model complex non-linear relationships between multiple variables and to incorporate uncertainty in their predictions, making them suitable to optimize noisy and expensive evaluation functions.

4.1. Multi-objective optimization of chemical synthesis There are different versions of BO algorithms depending on the acquisition function used to evaluate the surrogate model and the strategies used to suggest the most likely optimal values, for chemical reaction optimization the TSEMO (Thompson sampling efficient multi-objective optimization) algorithm has been the most widely used. Table 2 summarizes the use of various ML algorithms for the optimization of chemical synthesis with multiple objectives function. The use of TSEMO for optimization of a chemical reaction was reported by Schweidtmann *et al.* [76]. In this study the multi-objective Bayesian optimization (MOBO) was used to optimize a nucleophilic aromatic substitution (S_NAr) reaction (Table 2, entry 1) and a *N*-benzylation reaction (Table 2, entry 2) using an automated flow reactor. The objectives of the optimization were to maximize the space-time yield (STY) of **22** and **24**, while minimize either the *E*-factor for the S_NAr reaction or the impurity concentration for the *N*-benzylation reaction. For both reactions there were four variables to optimize that included metrics for reaction time, reagent concentrations and temperature. After an initialization of 20 experimental conditions sampled by LHS, the choice of reaction conditions was left to the TSEMO algorithm to optimize the S_NAr for a total of 48 iterations and the N-benzylation reaction for a total of 58 iterations. Both optimizations resulted in the discovery of a dense Pareto front with approximately 30-50% of the total suggested conditions resulting in non-dominated solutions.
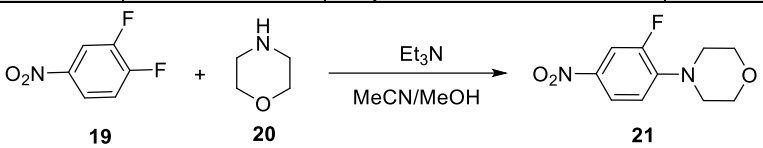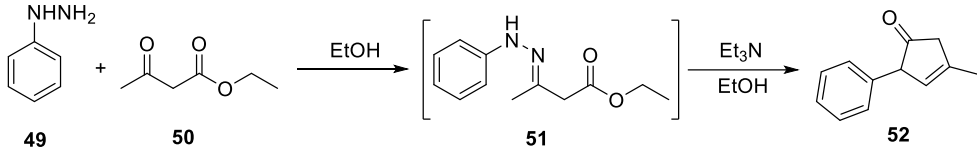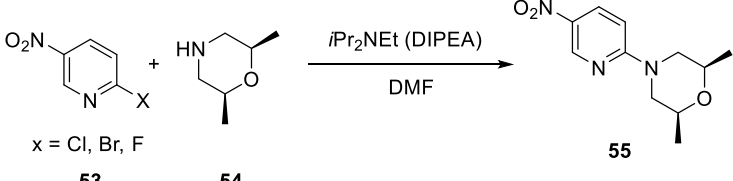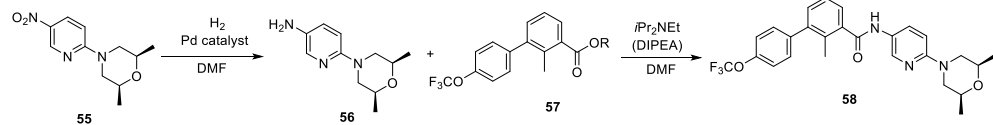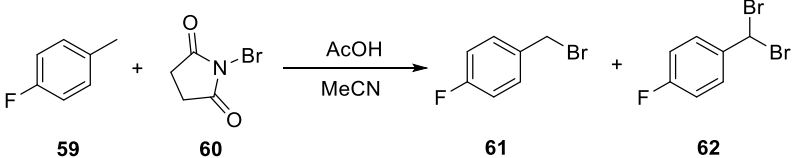
Since, multiple reports have demonstrated the ability of TSEMO to optimize multi-objective optimizations for the synthesis of organic molecules (see examples in Table 2, entries 3-5, 6-8). A particular noteworthy development is the application of TSEMO for the optimization of synthetic routes composed of two and more successive reaction steps or telescoped reactions [77–79]. Sagmeister et al. [77] reported the optimization of a two-step telescoped synthesis of the active pharmaceutical ingredient edaravone (Table 2, entry 9). In this study, a self-optimizing flow reactor was used to run the optimization of seven continuous variables including three variables for the first step

and four variables for the second step reaction. The optimization had the objective of maximizing the yield of the imine intermediate (**51**) obtained after the first reaction, the STY of **52** and minimizing the overall used equivalents of the reagents. After 85 iterations, a maximum yield of 95% for the synthesis of **51**, and a maximum STY of 5.42 kg/h for the synthesis of edaravone (**52**) were achieved. Setting an objective that limited the quantities of reagents led to the discovery of unexpected reaction conditions where the sub-stoichiometric amount of triethyl amine was sufficient to promote the second reaction step decreasing the waste produced during synthesis. Although no global solution that provided optimal reaction conditions for all three objectives was found, a distinct set of reaction conditions was identified that presented high yields and low overall equivalents. These conditions can be consolidated to meet user-defined manufacturing requirements effectively.

Table 2: Multi-objective optimization of organic synthesis case studies using machine learning (ML) methods and single-objective optimization of telescoped reactions.

Entry	Platform	Algorithm	Variables	Objectives	Reference
1					[76]
	Flow	BO (TSEMO)	Residence time Equiv. of **20** Conc. of **19** Temperature	↑STY[a] **21** ↓E-Factor	
2					[76]
	Flow	BO (TSEMO)	Flow rate Ratio of **24**:**25** Solvent Temperature	↑STY **26** ↓Yield **27**	
3					[80]
	Flow (CSTR)	TSEMO	Residence time Equiv. of **29** Temperature	↑STY **30** ↓Yield **31**	
4					

	Mode	Optimizer	Variables	Objectives	Ref
	Flow (CSTR)[b]	TSEMO	Flow of **32** Equiv. of **33** Equiv. of NaOH Temperature	↑STY **34** ↓Yield **35** ↑RME[c] **32**	[80]
5	NaOH (0.1 M) scheme				[81]
	Flow	TSEMO	Equiv. **33** Equiv. of NaOH Temperature Residence time	↑Yield **34** ↓Cost ↓E-factor	
6	Pd Catalyst, Ligand; K₃PO₄ (eq) (1 M); MeCN scheme (**37** + **38** → (E)-**39** + (Z)-**39**)				[15]
	Batch	Phoenics Gryffin	Ligand Ligand:Pd Pd cat. Loading Eq reagent 2 Temperature	↑Yield of (E)-**39** ↓Yield of (Z)-**39** ↓Pd catalyst loading ↓Equiv. of **x**	
6	H₂SO₄ aq scheme (**40** + **41** + **42** + **43** → **44** + **45** + **46** + **47**)				
	Batch	TSEMO	Temperature Conc. of H₂SO₄ Aq:organic ratio Time Equiv. of **40** Equiv. of **41** Equiv. of **42** Equiv. of **43**	↑Conversion **40-43** ↑Yield **44-47**	[82]
7	Dehydrogenation scheme (**44** + **45** + **46** + **47** → **48**)				
	Flow	TSEMO	Temperature Air flow Liquid flow Time Equiv. of **44**	↑Conversion **44-47** ↑Yield **48**	[82]

			Equiv. of **45** Equiv. of **46** Equiv. of **47**		
8					[77]
	Flow	TSEMO	Temperature Residence time Concentration of **19** Equiv. of **20** Et₃N	↑Conversion **1** ↑STY **21** ↓E-factor	
9					[77]
	Flow	TSEMO	Equiv. of **49** Conc. of step 1 Residence time of step 2 Temperature of step 1 Temperature of step 2 Equiv. of Et₃N	↑Yield of **51** ↑STY of **52** ↓Eq. **49** + Et₃N	
10					[78]
	Flow	Dragonfly	Temperature Time of residence Equiv. of **53** Equiv. of DIPEA Leaving group X	↓Cost ↑Productivity **55** ↑Yield **55**	
11					[78]
	Flow	Dragonfly	Activation time Equiv. ratio of **55:57** Temperature step **2** Volume reactor Activated group R	↑Yield **58** ↑Productivity **58**	
12					

	Flow	TSEMO	Equiv. of **59** Temperature Conc. of **60** Equiv. of AcOH Light intensity Residence time	↑STY **61** ↑Conversion of **60** ↑Selectivity	[83]
13					[79]
	Flow	BOAEI	Residence time Equiv. of **64** Temperature Equiv. of TsOH	↑Yield **66**	
14					[84]
	Flow	MVMOO	Solvents Residence time Conc. of **19** Equiv. of **20** Temperature	↑Yield of **21** ↑Yield of **22**	
15					[84]
	Flow	MVMOO	Ligands Residence time Equiv. of **68** Temperature	↑RME ↑STY **69**	
16					[47]
	Photo flow reactor (Robochem)	Bayesian Optimization	Conc. of **70** Catalyst loading Conc of CF_3SO_2Na $(NH_4)_2S_2O_8$ loading Residence time Light intensity	↑Yield ↑Throughput	

17					[47]
	Photo flow reactor (Robochem)	Bayesian Optimization	Conc. of **73** Conc. of **74** Catalyst loading Residence time Light intensity	↑Yield ↑Throughput	
18					[47]
	Photo flow reactor (Robochem)	Bayesian Optimization	Conc. of **76** R-H loading TBADT loading Residence time Light intensity	↑Yield ↑Throughput	
19					[46]
	Flow Slug reactor	TSEMO	Residence time Concentration Equiv. of **80** Temperature Equiv. of DBU Catalyst loading	↑Yield **81** ↑ STY **81** ↓Cost	

[a]Space-time-yield; [b]Continuous Stirred Tank Reactor; [c]Reaction mass efficiency, ↑ Maximization, ↓ Minimization.

4.2. Accelerating optimization campaigns

Shortening optimization times is desirable, especially when manufacturing active pharmaceutical ingredients (API) where only small amounts of materials available in each development step. Currently, optimization methods require an initialization step where reaction conditions are sampled and executed to train the surrogate models used during the optimization (Figure 8a). Sagmesiter et al. [77] performed a multi-objective optimization of a S_NAr reaction in an automated flow reactor platform and compared initialization sampling methods to understand how different methods affect the final number of experiments required to find optimal conditions (Table 2, entry 9). They compared LHS (20 experiments), full factorial DoE (17 experiments) and centre point only (1 experiment) as the starting data points. They found that LHS and full factorial DoE required a smaller number of optimization iterations after the initial set of experiments were conducted due to the better predictive capability of GPRs trained with larger amounts of data. However, when the total number of experiments including

the initialization set was considered, the number of experiments required to obtain optimal reaction values were larger or equal than the case where only one starting point was used as the only initial sample of reaction conditions. Thus, the authors concluded that is beneficial to start the algorithm driven optimization as soon as possible instead of performing an initial thorough exploration of the parametric space. However, they did not fully explore if there was a trade-off between a reduced number of initialization sampling and a total number of experiments to achieve optimal reaction conditions.

Recently, Taylor *et al.* [85] introduced the idea of multi-task learning Bayesian optimization (MTBO) for chemical reaction optimization. Analogous to transfer learning in ML models, the idea behind multi-task learning is to pre-train the surrogate GP models with data that has been previously collected from similar reactions to eliminate the need of an initial sampling step and reduce the overall number of experiments required to obtain optimal reaction conditions. In MTBO, the standard GPR surrogate models are replaced with multitask GPs that use kernels able to create correlations between multiple GPs. The GP that models the experimental conditions that are being optimized is called the main task while any other GP trained on previous data are called auxiliary tasks (Figure 9a). The authors benchmarked MTBO in-silico for a single objective optimization for Suzuki-Miyaura reaction and discovered that pretraining the multi-task GPs using a single data set as an auxiliary task resulted in fewer iterations to achieve the optimal conditions in comparison to standard BO in most cases. Moreover, the authors observed that when four auxiliary tasks were used instead of one, the number of iterations to obtain optimal reaction conditions was reduced from 15 to fewer than 5 experiments (Figure 9b). Finally, the authors tested the performance of MTBO by performing the optimization in an automated flow reactor of a series of palladium catalysed C-H activation reactions of chloroacetanilides to produce their corresponding oxindoles (Figure 9c). For all reactions, three continuous and one categorical variable were optimized to maximize reaction yield. The authors first performed a standard single-objective BO of reaction (i) in Figure 9c. The optimization was initialized with a set of 16 distinct reaction conditions sampled by LHS, reaching optimal reaction conditions in 7 further BO iterations. Subsequently, a reaction (ii, Figure 9c) yielding a similar oxindole product was optimized using MTBO where the data gathered from the previous optimization was used to train the auxiliary GP, obtaining the optimal conditions in only 11 iterations in comparison to 18 required for the first reaction. A third reaction (iii, Figure 9c) yielding a similar oxindole product was optimized using the previous data from the first two optimization campaigns to train the auxiliary task GP. The authors found the optimal conditions by the algorithm in 5 iterations. Futher, the authors tested the ability of MTBO to learn from previous experiments by performing the optimization of two other C-H activation reactions where the structure of the C-H activation substrate (**88**) had a different structure in comparison to the first three optimizations. Thus, for the fourth campaign they tested the optimization of reaction that produced an a 6-membered nitrogen ring quinolinone instead of the 5-membered ring present in oxindoles. The MTBO was able to find optimal reaction conditions in 10 iterations, demonstrating the capability of the algorithm to handle the optimization of reactions that show small structural deviations
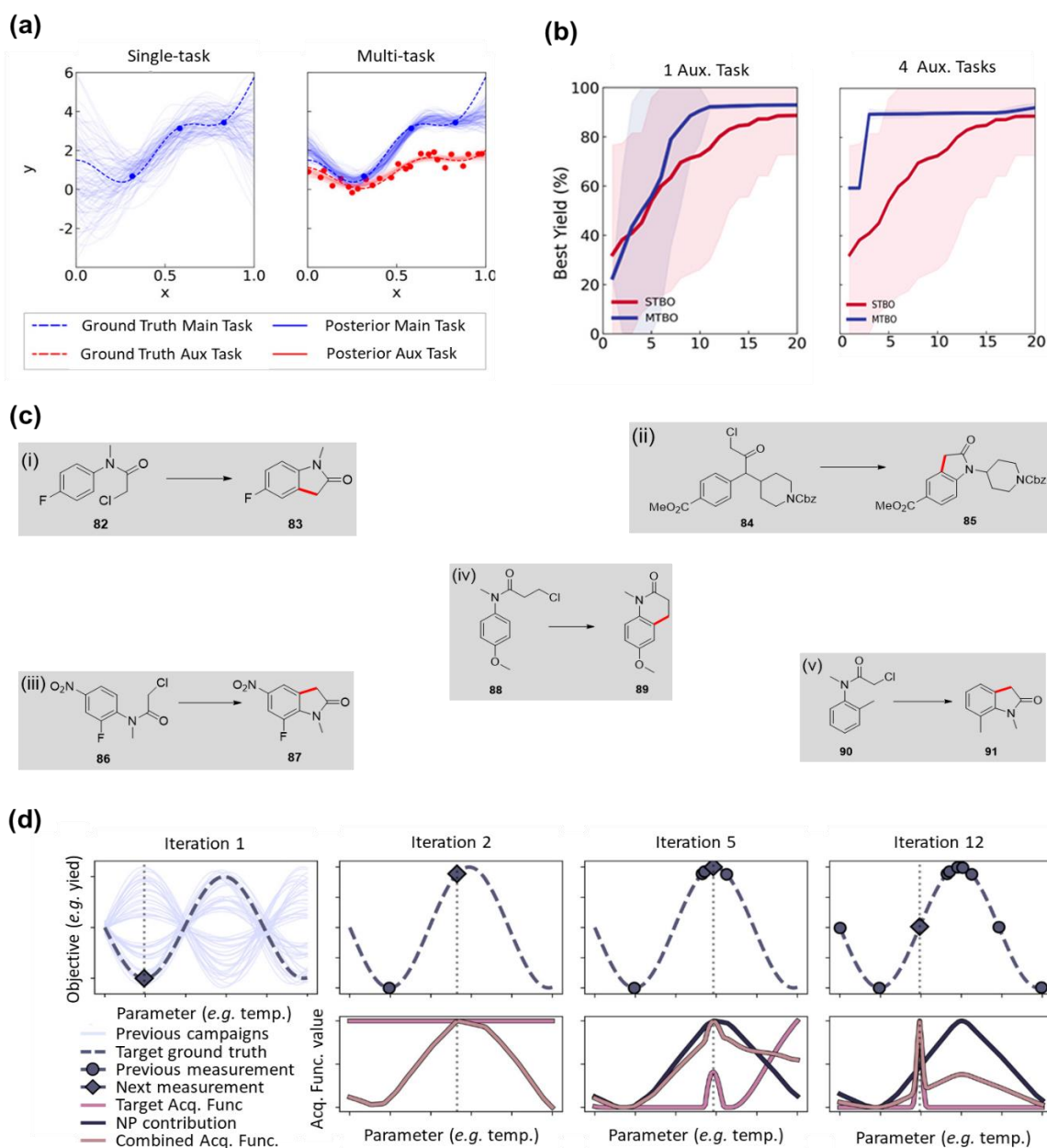
Figure 9: (a) Comparison to standard GP (single-task) and multi-task GP, it can be observed that by training an auxiliary task using data collected from a similar reaction the uncertainties associated with the GP predictions are greatly reduced (adapted from [85]). (b) Comparison of reaction optimizations performed in silico for single-task and multi-task BO. It can be observed that multi-task BO requires a reduced number of iterations to find optimal parameters that maximize reaction yield. The performance is further increased by incorporating a larger number of auxiliary tasks (adapted from [85]). (c) Reactions used to test multi-task BO in experimental conditions. Reaction (i) was performed using standard single-task BO, each subsequent reaction incorporated the previously collected data to train auxiliary tasks. (d) Example of SeMOpt maximizing a sine function. The upper row shows the ground truth function with the sampled points and best suggested candidate by the BO algorithm. The bottom row shows the values from the acquisition function from the surrogate of the target objective, the NP process and their combination (adapted from [86]).

from the auxiliary task. Finally, the limits of the MTBO were tested by using achloroacetanilide that presented an electron-rich aromatic ring (**90**), in this case the MTBO was unable to discover good performing reaction conditions.

Recently researchers from Atinary Technologies reported the development of SeMOpt, a BO framework that similarly to MTBO aims to transfer knowledge obtained from previous optimization campaigns to accelerate chemical reaction optimization [86]. In comparison to MTBO, SeMOpt has the advantage of being model agnostic thus can be applied with any combination of surrogate model and acquisition function used during the BO campaign. In addition to the surrogate model used for BO (see Figure 8a), SeMOpt introduces a surrogate neural process (NP) to model and make predictions based on previously gathered data. Then an acquisition function is used to select likely candidates by evaluating both the surrogate model and NP predictions. SeMOpt introduces the knowledge learnt by biasing the acquisition function of the surrogate model for the target optimization with the acquisition function evaluated on the NP model (Figure 9d). In addition, the biased introduced to the acquisition function by the NP is continuously updated and decrease as the number of optimization iterations increase. In this way, the optimization surrogate will eventually disregard the bias introduce by the NP whenever it becomes uninformative. The authors benchmarked the performance of the SeMOpt framework by performing an in-silico single-objective optimization of a simulated cross-coupling reaction and a Buchwald Hartwig cross-coupling of aryl halides. For the benchmarking the authors used several different BO algorithms and compared their performance when paired with SeMOpt. The authors observed that for all cases the application of SeMOpt outperformed the single-task implementation of the same BO algorithm. In addition, they compared the performance of SeMOpt against other algorithms that include some knowledge transfer into the optimization workflow including MTBO. The authors observed that SeMOpt outperformed most of the other algorithms, with MTBO closely matching SemMOpt performance.

4.3. Mixed-variable optimizations

A challenge in BO is to include categorical variables (i.e. non-continuous) into the optimization procedures due to inherent limitations of standard GP regressors (GPR) to include discrete variables into their predictions. Categorical variables such as choice of solvent, catalyst, ligands, additives, etc. are crucial for many chemical reactions. For this purpose, new algorithms have been developed to include categorical variables into MOBO. Kershaw *et al.* [84] utilized an in-house developed mixed-variable multi-objective optimization (MVMOO) algorithm, employing GP regression surrogate models tailored for predictions with discrete variable inputs. Their study employed a self-driving flow reactor to optimize the synthesis of ortho (**21**) and para (**22**) isomers of SNAr reaction, leveraging four continuous variables alongside a single discrete variable representing solvent (Table 2, entry 14). After 99 sequential reactions (25 LHS ad 74 optimzation iterations), the optimization found 20 non-dominated solutions that mapped the Pareto front from a highly dominant ortho product to a 50-50% split between the isomers. In addition, the researchers explored the optimization of a Sonogashira cross-coupling to optimize the STY and reaction mass efficiency (RME) for the synthesis of **69** (Table 2, entry 15). In this case the optimization involved three

continuous variables and the selection of a ligand for the catalyst as a discrete variable. After 69 sequential experiments (25 LHS, 44 optimization) the platform was able to identify 12 non-dominated solutions that demonstrated the trade-off between RME and STY. In general, most Pareto solutions were obtained when triphenyl phosphine was used as the catalyst ligand. Interestingly, triphenyl phosphine was the least sterically hindering ligand, which is counterintuitive to expert intuition which identifies sterically demanding ligands as more favourable choices for cross-coupling reactions.

Another noteworthy approach for the optimization of both continuous and categorical variables for a Suzuki-Mayura coupling reaction was reported by Christensen *et al.* [15] using in-house developed BO algorithms called Phoenics and Gryffin (Table 2 entry 6). The Gryffin algorithm uses Bayesian Neural Networks to construct the surrogate model, circumventing the limitations of GPR to fit categorical variables. The authors selected a total of four continuous reaction variables and the selection of a catalyst ligand as the unique categorical variable for the optimization. The algorithm targeted to find optimal reaction variables for four objectives including the maximization of the targeted stereoisomers (*E*-**39**, Table 2) and minimization of undesirable ones (*Z*-**39**, Table 2), catalyst loading, and reagent equivalents. Twelve ligands were initially selected based on domain expert knowledge and after 120 trials the best-performing conditions were similar to those previously reported in the literature. To further improve the performance of the reaction, the authors used DFT simulations to compute the chemical properties of 365 commercially available phosphine ligands, and by using k-means clustering they grouped the ligands in 24 distinct regions. Through the strategic selection of a representative ligand from each distinct region, the researchers identified a novel set of ligands, differing from conventional recommendations based on domain expertise. These ligands were discovered through ML clustering techniques, sampling distinct groups of molecules. Following optimization of reaction conditions integrating these 23 new ligands, the authors observed enhanced performance, surpassing that of previously reported ligands (Figure 10). This study showcased how data science, ML algorithms and chemical reaction optimization can be used to optimize and discover reaction conditions that would have been missed by sampling of ligands using human intuition. Another great example of a combination of ML/AI chemoinformatic tools and reaction optimization was reported by Nambiar et al. [78], where a computer-aided synthesis planning (CASP) tool was used to find a 3-step reaction pathway for the synthesis of the API Sonidegib (**58**). After the generation of multiple reaction pathways by the CASP tool, the author manually selected a highly ranked route based on synthetic feasibility. This three-step reaction comprised a $S_N Ar$, hydrogenation reduction of a nitro group and an amide coupling as presented (Table 2, entry 10 and 11). Using an automated flow reactor, the researchers attempted to perform the optimization of the fully telescoped reaction. However, the optimization campaign had to be restructured into two independent optimizations due to the subproducts of the $S_N Ar$ reaction poisoned the Pd catalyst used in the hydrogenation reaction. Thus, the MOBO of the

S$_N$Ar reaction was performed to maximize yield of **55**, productivity and minimize cost of reagents per mole of product by optimizing 4 continuous and a categorical variable. The second optimization campaign was performed on the telescoped reaction that included the hydrogenation step and the amide coupling. In this case the objectives for the optimization were maximizing yield and productivity by optimizing two categorical and three continuous variables. Dragonfly, an open-sourced Bayesain optimization package was used to optimize both categorical and continuous reaction variables. An increase in yield and productivity was observed as the optimization progressed. The authors found that the selection of F as a leaving group led to the highest yield (98.3%) and productivity (5.97 g/h) for the synthesis of **58**. However, if Cl was selected as the leaving group only a marginal reduction in yield and productivity was observed (93.8%, 5.70 g/h) but a 33% reduction in the cost. In the second reaction, both high yields and productivity were achieved concurrently. Because these objectives were positively correlated, no trade-offs were observed in the optimization suggestions.
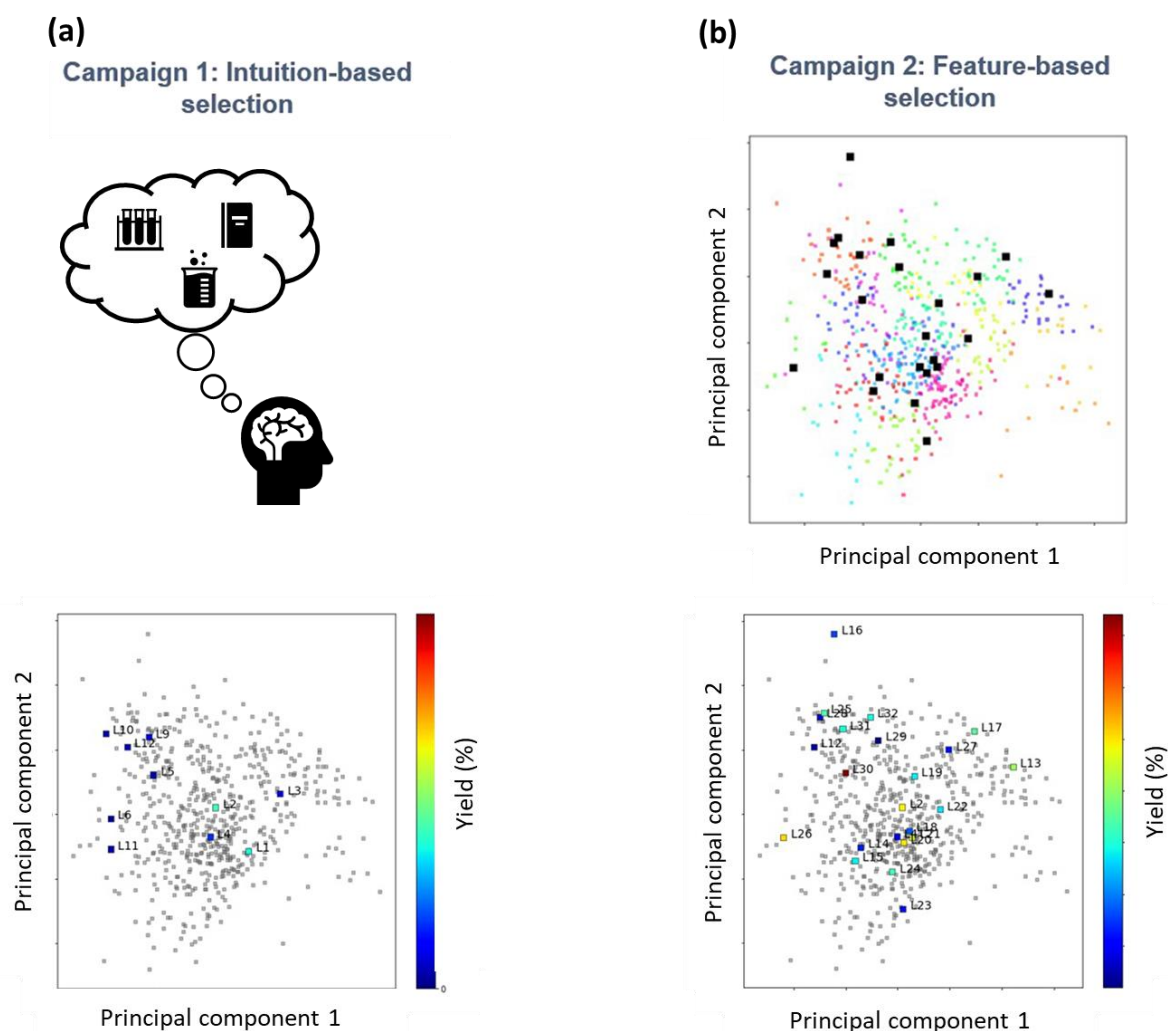

Figure 10: Comparison of reaction yield between (a) optimization campaign where catalyst ligand selection as based on expert's intuition and (b) optimization campaign where ligand selection was derived from sampling the distinct ligand clusters obtained from k-means clustering of ligands in the chemical space calculated (adapted from [15]).

4.4. Benchmarking of optimization algorithms

With an increasing number of optimization algorithms, an effort to benchmark their performance is required. Felton *et al.* [71] have highlighted the fact that the ability of an algorithm to perform well in a specific task may not translate universally to other problems, thus a specific optimization algorithm for chemical reaction optimization may have different performances depending on the nature of the target variables, objectives and chemical reaction. Also, the computational time required to execute an algorithm varies and it should be taken in consideration to select the most appropriate variation for each case study. To benchmark different optimization algorithms Felton et al. released Summit, a Python module containing several optimization algorithms and two benchmark in-silico models to compare the performance of algorithms. Initially, the benchmarking models included in Summit were a kinetic model for the S_NAr reaction of difluronitrobenzene with pyrrolidine, and neural network forward model for the prediction of yield of diphenylamine in a Pd-catalysed C-N cross coupling reaction trained on a previously published data set containing 96 unique set of reaction conditions. The optimization for the S_NAr reaction included four continuous variables and two optimization objectives, while the C-N cross coupling included three continuous variables, two categorical variables and two optimization objectives. The optimization algorithms used during the optimization included non-ML algorithms (Nelder Mead, SNOBFIT), BO algorithms (Gryffin, SOBO and TSEMO) and DRO a pretrained reinforcement learning agent algorithm. For the optimization of S_NAr reaction, BO methods were superior to any other of the algorithms, reaching a higher HV in a smaller amount of iterations. When the BO algorithms were compared, TSEMO outperformed by a significant margin against Gryffin and SOBO. For the C-N cross-coupling, all models had a similar hypervolume (HV) performance including a random search of reaction conditions, due to the small parametric space for the selected categorical variable. Müller *et al.* [72] also conducted benchmarking in-silico study for 6 different chemical reactions using previously reported kinetic models. In this case three distinct BO algorithms (TSEMO, ParEGO, EIM-EGO) and a genetic algorithm (NSGA-II) were compared. They demonstrated that BO methods outperformed non-BO methods such as NSGA-II, which is consistent with the Felton et al. studies earlier.

5. Summary ad Outlook

In this article, we outlined the latest advances in machine learning-driven multi-objective optimization for chemical synthesis, in addition to breakthroughs in high throughput experimentation and analytical techniques. The recent developments of ML algorithms, high throughput experimentation tools, data processing techniques, and self-optimizing reactors has been a transformative force for chemical optimization processes. However, there are still plenty of research opportunities to continue the transformation of the field and to accelerate the execution of chemical reaction optimization. Recent work has had substantial progress in optimizing multiple continuous variables, yet the utilization of categorical variables in chemical synthesis optimization has predominantly been confined to single-step reactions with one or two

optimization objectives. The development of ML algorithms that can efficiently optimize a larger number of categorical variables will be crucial to unlocking the full potential of optimization methods. This is particularly true when objective functions that go beyond direct measurements of the reaction product outputs (*e.g.* yield, throughput, selectivity, etc.) are targeted. For example, optimizations that target to minimize the environmental impact of chemical synthesis is becoming a priority in the pharmaceutical industry. The environmental impact of a reaction not only depends on the efficiency of the process (*i.e.* yield and throughput) but will be highly affected by the nature of the solvent, catalyst, choice of reagents, downstream workup, etc. used during synthesis. To obtain optimal reaction conditions that minimize the environmental impact, the exploration of a large number of different reagents may be required, which is not possible to perform through traditional optimization methodologies. Nonetheless, ML algorithms could offer an efficient approach to navigate the parametric space and to reduce the experimentation time to find the conditions that minimize environmental impact of a particular manufacturing process. However, the state-of-the-art optimization algorithms still fall off from the target.

Manufacturing of pharmaceutical and specialty chemicals commonly involves multiple reaction steps to transform the starting reagents into the final product. So far, optimization algorithms have been mostly applied to single-step reactions or applied step by step to each reaction of a multi-step procedure. Few examples in the literature have demonstrated the ability of ML methods to optimize telescoped reactions in automated flow reactors, the positive results should encourage further research in this field. However, there is bound to occur situations where the telescoped reactions are not feasible due to competing chemical interactions of the reagents in the reaction mixture. Thus, more research should investigate optimization strategies in multi-step reaction procedures in which the final objective function has input variables from multiple steps of the synthetic route.

The application of ML algorithms to aid the discovery of new chemistry knowledge is flourishing, from generative design to property prediction and reaction planning. Further work should incorporate the diverse applications of ML in chemistry into chemical reaction optimization campaigns to open new avenues of research and discoveries. In particular, ML tools have shown to have a large potential in the planning of reaction optimization campaigns to assist in the selection of categorical chemical variables (*e.g.* catalysts, ligands, additives, etc.). Reviewed work by Christensen *et al.* [15] have already demonstrated the advantages of applying ML clustering methods to discover new catalyst ligands that would have been missed if the selection of test ligands only relied on human chemical insights. Taylor *et al.* [85] also highlighted the use of DFT or ML alternatives to find similarities between reaction models to apply efficient multi-task learning to chemical reaction optimization. Another potential application of ML tools is the use of CASP to discover alternative reaction routes with the potential to improve the efficiency of current manufacturing methods. Finally, leveraging on the large quantities of data generated from self-optimizing chemical platforms and their experimental versatility, we envision the incorporation of reaction optimization methods with generative design to create full-driving laboratories that

both tackle the discovery of new molecules and the search for optimal synthesis conditions to satisfy the production constraints for a chemical commodity.

In addition to ML algorithms, future research on the optimization of organic chemistry reactions should leverage advanced AI models. In particular, we highlight large language models (LLMs) as a promising technology to enable the extraction of chemical knowledge from previously written literature. LLMs can be used to generate synthesis protocols for target materials through data mining of peer-reviewed literature [87,88]. Bran et al. [89] recently demonstrated an advanced LLM-powered chemistry engine called ChemCrow that is capable of planning and executing the synthesis of organic molecules. The LLM integrated 18 cheminformatic tools and performed the reasoning steps based on the information supplied by these tools to accomplish specific chemistry tasks. Along these lines, we envision that the integration of CASP tools and LLMs could accelerate the optimization of organic reactions by providing viable reaction routes with starting conditions that are close to the reaction optimum based on previous studies. Finally, LLMs could assist researchers with limited coding experience in writing the code required to automate their experimental workflows and execute their reaction optimizations.

Standardizing benchmarking methods for ML optimization algorithms will be crucial as the number of optimization methodologies increases. Foundational work has been laid by Lapkin research group with the release of the Summit open-source software package [71]. Given the vast spectrum of chemical reactions, there is a necessity to develop a diverse array of reaction models to comprehensively assess the suitability of optimization methods for various scenarios. The field should leverage the ability of high throughput experimentation to produce large amounts of data to create reliable forward models that can be incorporated into an online repository. Thus, researchers could access this online repository to benchmark new optimization algorithms by performing in-silico optimization campaigns of the chemical reaction models.

For the continued advancement of this research, democratizing access to proprietary autonomous platforms, algorithms, and fostering collaboration to share expertise within academia is paramount. While significant advances have been made particularly in addressing immediate challenges, we are convinced that the full potential of machine learning and artificial intelligence is yet to be fully realized. This highlights the importance of raising cross-functional expertise both within universities and at pre-university levels, thereby nurturing a broader knowledge base. Such an approach empowers young researchers to tackle research challenges holistically right from the outset, thereby unlocking new possibilities for innovation and advancement.

Acknowledgements

6. References

(1)   Meuwly, M. *Chem. Rev.* **2021**, *121*, 10218–10239. doi:10.1021/acs.chemrev.1c00033

(2)   Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. *Nature* **2021**, *590*, 89–96. doi:10.1038/s41586-021-03213-y

(3)   Gao, W.; Raghavan, P.; Coley, C. W. *Nat Commun* **2022**, *13*, 1075. doi:10.1038/s41467-022-28736-4

(4)   Seifrid, M.; Pollice, R.; Aguilar-Granda, A.; Morgan Chan, Z.; Hotta, K.; Ser, C. T.; Vestfrid, J.; Wu, T. C.; Aspuru-Guzik, A. *Acc. Chem. Res.* **2022**, *55*, 2454–2466. doi:10.1021/acs.accounts.2c00220

(5)   Taylor, C. J.; Pomberger, A.; Felton, K. C.; Grainger, R.; Barecka, M.; Chamberlain, T. W.; Bourne, R. A.; Johnson, C. N.; Lapkin, A. A. *Chem. Rev.* **2023**, *123*, 3089–3126. doi:10.1021/acs.chemrev.2c00798

(6)   Griffin, D. J.; Coley, C. W.; Frank, S. A.; Hawkins, J. M.; Jensen, K. F. *Org. Process Res. Dev.* **2023**, *27*, 1868–1879. doi:10.1021/acs.oprd.3c00229

(7)   Sagmeister, P.; Williams, J. D.; Kappe, C. O. *Chimia* **2023**, *77*, 300. doi:10.2533/chimia.2023.300

(8)   Fromer, J. C.; Coley, C. W. *Patterns* **2023**, *4*, 100678. doi:10.1016/j.patter.2023.100678

(9)   Askr, H.; Elgeldawi, E.; Aboul Ella, H.; Elshaier, Y. A. M. M.; Gomaa, M. M.; Hassanien, A. E. *Artif Intell Rev* **2023**, *56*, 5975–6037. doi:10.1007/s10462-022-10306-1

(10)   Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. *Nat Rev Drug Discov* **2019**, *18*, 463–477. doi:10.1038/s41573-019-0024-5

(11)   Kreutter, D.; Reymond, J.-L. *Chem. Sci.* **2023**, *14*, 9959–9969. doi:10.1039/D3SC01604H

(12)   Mo, Y.; Guan, Y.; Verma, P.; Guo, J.; Fortunato, M. E.; Lu, Z.; Coley, C. W.; Jensen, K. F. *Chem. Sci.* **2021**, *12*, 1469–1478. doi:10.1039/D0SC05078D

(13)   Bennett, J. A.; Orouji, N.; Khan, M.; Sadeghi, S.; Rodgers, J.; Abolhasani, M. *Nat Chem Eng* **2024**, *1*, 240–250. doi:10.1038/s44286-024-00033-5

(14)   Yang, W.; Fidelis, T. T.; Sun, W.-H. *ACS Omega* **2020**, *5*, 83–88. doi:10.1021/acsomega.9b03673

(15)   Christensen, M.; Yunker, L. P. E.; Adedeji, F.; Häse, F.; Roch, L. M.; Gensch, T.; Dos Passos Gomes, G.; Zepel, T.; Sigman, M. S.; Aspuru-Guzik, A.; Hein, J. E. *Commun Chem* **2021**, *4*, 112. doi:10.1038/s42004-021-00550-x

(16)   Buitrago Santanilla, A.; Christensen, M.; Campeau, L.-C.; Davies, I. W.; Dreher, S. D. *Org. Lett.* **2015**, *17*, 3370–3373. doi:10.1021/acs.orglett.5b01648

(17)   Brocklehurst, C. E.; Gallou, F.; Hartwieg, J. C. D.; Palmieri, M.; Rufle, D. *Org. Process Res. Dev.* **2018**, *22*, 1453–1457. doi:10.1021/acs.oprd.8b00200

(18)   Grainger, R.; Heightman, T. D.; Ley, S. V.; Lima, F.; Johnson, C. N. *Chem. Sci.* **2019**, *10*, 2264–2271. doi:10.1039/C8SC04789H

(19)   Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. *Science* **2018**, *360*, 186–190. doi:10.1126/science.aar5169

(20)   Gesmundo, N.; Dykstra, K.; Douthwaite, J. L.; Kao, Y.-T.; Zhao, R.; Mahjour, B.; Ferguson, R.; Dreher, S.; Sauvagnat, B.; Saurí, J.; Cernak, T. *Nat. Synth* **2023**, *2*, 1082–1091. doi:10.1038/s44160-023-00351-1

(21)   Zarate, C.; Ardolino, M.; Morriello, G. J.; Logan, K. M.; Kaplan, W. P.; Torres, L.; Li, D.; Chen, M.; Li, H.; Su, J.; Fuller, P.; Maddess, M. L.; Song, Z. J. *Org. Process Res. Dev.* **2021**, *25*, 642–647. doi:10.1021/acs.oprd.0c00446

(22)   Fier, P. S.; Maloney, K. M. *Org. Lett.* **2017**, *19*, 3033–3036. doi:10.1021/acs.orglett.7b01403

(23)   Huff, C. A.; Cohen, R. D.; Dykstra, K. D.; Streckfuss, E.; DiRocco, D. A.; Krska, S. W. *J. Org. Chem.* **2016**, *81*, 6980–6987. doi:10.1021/acs.joc.6b00811

(24)   DiRocco, D. A.; Dykstra, K.; Krska, S.; Vachal, P.; Conway, D. V.; Tudge, M. *Angewandte Chemie International Edition* **2014**, *53*, 4802–4806. doi:10.1002/anie.201402023

(25) Primer, D. N.; Molander, G. A. *J. Am. Chem. Soc.* **2017**, *139*, 9847–9850. doi:10.1021/jacs.7b06288

(26) Nicastri, M. C.; Lehnherr, D.; Lam, Y.; DiRocco, D. A.; Rovis, T. *J. Am. Chem. Soc.* **2020**, *142*, 987–998. doi:10.1021/jacs.9b10871

(27) Corcoran, E. B.; Pirnot, M. T.; Lin, S.; Dreher, S. D.; DiRocco, D. A.; Davies, I. W.; Buchwald, S. L.; MacMillan, D. W. C. *Science* **2016**, *353*, 279–283. doi:10.1126/science.aag0209

(28) Lin, S.; Dikler, S.; Blincoe, W. D.; Ferguson, R. D.; Sheridan, R. P.; Peng, Z.; Conway, D. V.; Zawatzky, K.; Wang, H.; Cernak, T.; Davies, I. W.; DiRocco, D. A.; Sheng, H.; Welch, C. J.; Dreher, S. D. *Science* **2018**, *361*, eaar6236. doi:10.1126/science.aar6236

(29) Lee, H.; Boyer, N. C.; Deng, Q.; Kim, H.-Y.; Sawyer, T. K.; Sciammetta, N. *Chem. Sci.* **2019**, *10*, 5073–5078. doi:10.1039/C9SC00694J

(30) 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. *Nature* **2020**, *583*, 237–241. doi:10.1038/s41586-020-2442-2

(31) Manzano, J. S.; Hou, W.; Zalesskiy, S. S.; Frei, P.; Wang, H.; Kitson, P. J.; Cronin, L. *Nat. Chem.* **2022**, *14*, 1311–1318. doi:10.1038/s41557-022-01016-w

(32) Godfrey, A. G.; Masquelin, T.; Hemmerle, H. *Drug Discovery Today* **2013**, *18*, 795–802. doi:10.1016/j.drudis.2013.03.001

(33) Ha, T.; Lee, D.; Kwon, Y.; Park, M. S.; Lee, S.; Jang, J.; Choi, B.; Jeon, H.; Kim, J.; Choi, H.; Seo, H.-T.; Choi, W.; Hong, W.; Park, Y. J.; Jang, J.; Cho, J.; Kim, B.; Kwon, H.; Kim, G.; Oh, W. S.; Kim, J. W.; Choi, J.; Min, M.; Jeon, A.; Jung, Y.; Kim, E.; Lee, H.; Choi, Y.-S. *SCIENCE ADVANCES* **2023**

(34) Kuleshova, J.; Hill-Cousins, J. T.; Birkin, P. R.; Brown, R. C. D.; Pletcher, D.; Underwood, T. J. *Electrochimica Acta* **2011**, *56*, 4322–4326. doi:10.1016/j.electacta.2011.01.036

(35) Green, R. A.; Brown, R. C. D.; Pletcher, D.; Harji, B. *Electrochemistry Communications* **2016**, *73*, 63–66. doi:10.1016/j.elecom.2016.11.004

(36) IBM RXN for Chemistry https://rxn.res.ibm.com/rxn/robo-rxn/welcome (accessed Apr 9, 2024)

(37) Sun, A. C.; Steyer, D. J.; Robinson, R. I.; Ginsburg-Moraff, C.; Plummer, S.; Gao, J.; Tucker, J. W.; Alpers, D.; Stephenson, C. R. J.; Kennedy, R. T. *Angewandte Chemie International Edition* **2023**, *62*, e202301664. doi:10.1002/anie.202301664

(38) 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. *ACS Cent. Sci.* **2021**, *7*, 1126–1134. doi:10.1021/acscentsci.1c00303

(39) Hsieh, H.-W.; Coley, C. W.; Baumgartner, L. M.; Jensen, K. F.; Robinson, R. I. *Org. Process Res. Dev.* **2018**, *22*, 542–550. doi:10.1021/acs.oprd.8b00018

(40) Churski, K.; Korczyk, P.; Garstecki, P. *Lab Chip* **2010**, *10*, 816. doi:10.1039/b925500a

(41) Kaminski, T. S.; Jakiela, S.; Czekalska, M. A.; Postek, W.; Garstecki, P. *Lab Chip* **2012**, *12*, 3995. doi:10.1039/c2lc40540g

(42) Reizman, B. J.; Jensen, K. F. *Chem. Commun.* **2015**, *51*, 13290–13293. doi:10.1039/C5CC03651H

(43) Abolhasani, M.; Jensen, K. F. *Lab Chip* **2016**, *16*, 2775–2784. doi:10.1039/C6LC00728G

(44) Coley, C. W.; Abolhasani, M.; Lin, H.; Jensen, K. F. *Angewandte Chemie International Edition* **2017**, *56*, 9847–9850. doi:10.1002/anie.201705148

(45) Pieber, B.; Shalom, M.; Antonietti, M.; Seeberger, P. H.; Gilmore, K. *Angewandte Chemie International Edition* **2018**, *57*, 9976–9979. doi:10.1002/anie.201712568

(46) Wagner, F.; Sagmeister, P.; Jusner, C. E.; Tampone, T. G.; Manee, V.; Buono, F. G.; Williams, J. D.; Kappe, C. O. *Advanced Science* **2024**, 2308034. doi:10.1002/advs.202308034

(47) Slattery, A.; Wen, Z.; Tenblad, P.; Pintossi, D.; Sanjose-Orduna, J.; Den Hartog, T.; Noel, T. An All-in-One Multipurpose Robotic Platform for the Self-Optimization, Intensification and Scale-up of Photocatalysis in Flow. June 8, 2023. doi:10.26434/chemrxiv-2023-r0drq

(48) Collins, N.; Stout, D.; Lim, J.-P.; Malerich, J. P.; White, J. D.; Madrid, P. B.; Latendresse, M.; Krieger, D.; Szeto, J.; Vu, V.-A.; Rucker, K.; Deleo, M.; Gorfu, Y.; Krummenacker, M.; Hokama, L. A.; Karp, P.; Mallya, S. *Org. Process Res. Dev.* **2020**, *24*, 2064–2077. doi:10.1021/acs.oprd.0c00143

(49) Reis, M.; Gusev, F.; Taylor, N. G.; Chung, S. H.; Verber, M. D.; Lee, Y. Z.; Isayev, O.; Leibfarth, F. A. *J. Am. Chem. Soc.* **2021**, *143*, 17677–17689. doi:10.1021/jacs.1c08181

(50) Zhou, Y.; Gu, Y.; Jiang, K.; Chen, M. *Macromolecules* **2019**, *52*, 5611–5617. doi:10.1021/acs.macromol.9b00846

(51) Tan, J. D.; Ramalingam, B.; Wong, S. L.; Cheng, J.; Lim, Y.-F.; Chellappan, V.; Khan, S. A.; Kumar, J.; Hippalgaonkar, K. Machine Learning Predicts Conversion and Molecular Weight Distributions in Computer Controlled Polymerization. ChemRxiv July 14, 2022. doi:10.26434/chemrxiv-2022-tlz53

(52) McMullen, J. P.; Stone, M. T.; Buchwald, S. L.; Jensen, K. F. *Angewandte Chemie International Edition* **2010**, *49*, 7076–7080. doi:10.1002/anie.201002590

(53) Fitzpatrick, D. E.; Battilocchio, C.; Ley, S. V. *Org. Process Res. Dev.* **2016**, *20*, 386–394. doi:10.1021/acs.oprd.5b00313

(54) Poscharny, K.; Fabry, D. C.; Heddrich, S.; Sugiono, E.; Liauw, M. A.; Rueping, M. *Tetrahedron* **2018**, *74*, 3171–3175. doi:10.1016/j.tet.2018.04.019

(55) Cortés-Borda, D.; Wimmer, E.; Gouilleux, B.; Barré, E.; Oger, N.; Goulamaly, L.; Peault, L.; Charrier, B.; Truchet, C.; Giraudeau, P.; Rodriguez-Zubiri, M.; Le Grognec, E.; Felpin, F.-X. *J. Org. Chem.* **2018**, *83*, 14286–14299. doi:10.1021/acs.joc.8b01821

(56) Nandiwale, K. Y.; Hart, T.; Zahrt, A. F.; Nambiar, A. M. K.; Mahesh, P. T.; Mo, Y.; Nieves-Remacha, M. J.; Johnson, M. D.; García-Losada, P.; Mateos, C.; Rincón, J. A.; Jensen, K. F. *React. Chem. Eng.* **2022**, *7*, 1315–1327. doi:10.1039/D2RE00054G

(57) Bédard, A.-C.; Adamo, A.; Aroh, K. C.; Russell, M. G.; Bedermann, A. A.; Torosian, J.; Yue, B.; Jensen, K. F.; Jamison, T. F. *Science* **2018**, *361*, 1220–1225. doi:10.1126/science.aat0650

(58) Aka, E. C.; Wimmer, E.; Barré, E.; Cortés-Borda, D.; Ekou, T.; Ekou, L.; Rodriguez-Zubiri, M.; Felpin, F.-X. *Org. Process Res. Dev.* **2020**, *24*, 745–751. doi:10.1021/acs.oprd.9b00525

(59) Rodriguez-Zubiri, M.; Felpin, F.-X. *Org. Process Res. Dev.* **2022**, *26*, 1766–1793. doi:10.1021/acs.oprd.2c00102

(60) Roos, G. R., Christoph; Büter, Karin Berger; Simmen, Urs. *Planta Med* **2004**, *70*, 771–777. doi:10.1055/s-2004-827210

(61) Schmidt, B.; Jaroszewski, J. W.; Bro, R.; Witt, M. *Anal. Chem.* **2008**, *80*, 1978–1987. doi:10.1021/ac702064p

(62) Jansen, B. C.; Hafkenscheid, L.; Bondt, A.; Gardner, R. A.; Hendel, J. L.; Wuhrer, M.; Spencer, D. I. R. *PLoS ONE* **2018**, *13*, e0200280. doi:10.1371/journal.pone.0200280

(63) Bovee, R. Bovee/Aston, 2024

(64) Liu, J.; Sato, Y.; Yang, F.; Kukor, A. J.; Hein, J. E. *Chemistry - Methods* **2022**, *2*, e202200009. doi:10.1002/cmtd.202200009

(65) Sagmeister, P.; Lebl, R.; Castillo, I.; Rehrl, J.; Kruisz, J.; Sipek, M.; Horn, M.; Sacher, S.; Cantillo, D.; Williams, J. D.; Kappe, C. O. *Angewandte Chemie International Edition* **2021**, *60*, 8139–8148. doi:10.1002/anie.202016007

(66) 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. *ACS Cent. Sci.* **2023**, *9*, 307–317. doi:10.1021/acscentsci.2c01042

(67) Sagmeister, P.; Melnizky, L.; Williams, J.; Kappe, C. O. An Automated Dual Modeling Approach to Accelerate Reaction Analysis and Optimization. March 5, 2024. doi:10.26434/chemrxiv-2024-mj3h8

(68) Holmes, N.; Akien, G. R.; Savage, R. J. D.; Stanetty, C.; Baxendale, I. R.; Blacker, A. J.; Taylor, B. A.; Woodward, R. L.; Meadows, R. E.; Bourne, R. A. *React. Chem. Eng.* **2016**, *1*, 96–100. doi:10.1039/C5RE00083A

(69) Parrott, A. J.; Bourne, R. A.; Akien, G. R.; Irvine, D. J.; Poliakoff, M. *Angewandte Chemie International Edition* **2011**, *50*, 3788–3792. doi:10.1002/anie.201100412

(70) Sans, V.; Porwol, L.; Dragone, V.; Cronin, L. *Chem. Sci.* **2015**, *6*, 1258–1264. doi:10.1039/C4SC03075C

(71) Felton, K. C.; Rittig, J. G.; Lapkin, A. A. *Chemistry Methods* **2021**, *1*, 116–122. doi:10.1002/cmtd.202000051

(72) Müller, P.; Clayton, A. D.; Manson, J.; Riley, S.; May, O. S.; Govan, N.; Notman, S.; Ley, S. V.; Chamberlain, T. W.; Bourne, R. A. *React. Chem. Eng.* **2022**, *7*, 987–993. doi:10.1039/D1RE00549A

(73) Nandi, S.; Badhe, Y.; Lonari, J.; Sridevi, U.; Rao, B. S.; Tambe, S. S.; Kulkarni, B. D. *Chemical Engineering Journal* **2004**, *97*, 115–129. doi:10.1016/S1385-8947(03)00150-5

(74) Lahiri, S. K.; Khalfe, N. *Chemical Product and Process Modeling* **2008**, *3*. doi:10.2202/1934-2659.1261

(75) Lahiri, S. K.; Khalfe, N. *The Canadian Journal of Chemical Engineering* **2009**, *87*, 118–128. doi:10.1002/cjce.20123

(76) Schweidtmann, A. M.; Clayton, A. D.; Holmes, N.; Bradford, E.; Bourne, R. A.; Lapkin, A. A. *Chemical Engineering Journal* **2018**, *352*, 277–282. doi:10.1016/j.cej.2018.07.031

(77) Sagmeister, P.; Ort, F. F.; Jusner, C. E.; Hebrault, D.; Tampone, T.; Buono, F. G.; Williams, J. D.; Kappe, C. O. *Advanced Science* **2022**, *9*, 2105547. doi:10.1002/advs.202105547

(78) Nambiar, A. M. K.; Breen, C. P.; Hart, T.; Kulesza, T.; Jamison, T. F.; Jensen, K. F. *ACS Cent. Sci.* **2022**, *8*, 825–836. doi:10.1021/acscentsci.2c00207

(79) Clayton, A. D.; Pyzer-Knapp, E. O.; Purdie, M.; Jones, M. F.; Barthelme, A.; Pavey, J.; Kapur, N.; Chamberlain, T. W.; Blacker, A. J.; Bourne, R. A. *Angewandte Chemie International Edition* **2023**, *62*, e202214511. doi:10.1002/anie.202214511

(80) Clayton, A. D.; Schweidtmann, A. M.; Clemens, G.; Manson, J. A.; Taylor, C. J.; Niño, C. G.; Chamberlain, T. W.; Kapur, N.; Blacker, A. J.; Lapkin, A. A.; Bourne, R. A. *Chemical Engineering Journal* **2020**, *384*, 123340. doi:10.1016/j.cej.2019.123340

(81) Jeraal, M. I.; Sung, S.; Lapkin, A. A. *Chemistry Methods* **2021**, *1*, 71–77. doi:10.1002/cmtd.202000044

(82) Jorayev, P.; Russo, D.; Tibbetts, J. D.; Schweidtmann, A. M.; Deutsch, P.; Bull, S. D.; Lapkin, A. A. *Chemical Engineering Science* **2022**, *247*, 116938. doi:10.1016/j.ces.2021.116938

(83) Knoll, S.; Jusner, C. E.; Sagmeister, P.; Williams, J. D.; Hone, C. A.; Horn, M.; Kappe, C. O. *React. Chem. Eng.* **2022**, *7*, 2375–2384. doi:10.1039/D2RE00208F

(84) Kershaw, O. J.; Clayton, A. D.; Manson, J. A.; Barthelme, A.; Pavey, J.; Peach, P.; Mustakis, J.; Howard, R. M.; Chamberlain, T. W.; Warren, N. J.; Bourne, R. A. *Chemical Engineering Journal* **2023**, *451*, 138443. doi:10.1016/j.cej.2022.138443

(85) Taylor, C. J.; Felton, K. C.; Wigh, D.; Jeraal, M. I.; Grainger, R.; Chessari, G.; Johnson, C. N.; Lapkin, A. A. *ACS Cent. Sci.* **2023**, acscentsci.3c00050. doi:10.1021/acscentsci.3c00050

(86) Hickman, R. J.; Ruža, J.; Tribukait, H.; Roch, L. M.; García-Durán, A. *React. Chem. Eng.* **2023**, *8*, 2284–2296. doi:10.1039/D3RE00008G

(87) Thway, M.; Low, A. K. Y.; Khetan, S.; Dai, H.; Recatala-Gomez, J.; Chen, A. P.; Hippalgaonkar, K. *Digital Discovery* **2024**, *3*, 328–336. doi:10.1039/D3DD00202K

(88) Zhang, W.; Wang, Q.; Kong, X.; Xiong, J.; Ni, S.; Cao, D.; Niu, B.; Chen, M.; Li, Y.; Zhang, R.; Wang, Y.; Zhang, L.; Li, X.; Xiong, Z.; Shi, Q.; Huang, Z.; Fu, Z.; Zheng, M. *Chemical Science* **2024**. doi:10.1039/D4SC00924J

(89) M. Bran, A.; Cox, S.; Schilter, O.; Baldassari, C.; White, A. D.; Schwaller, P. *Nat Mach Intell* **2024**, *6*, 525–535. doi:10.1038/s42256-024-00832-8