Automated robotic platforms in design and development of formulations

Liwei Cao\textsuperscript{1,2}, Danilo Russo\textsuperscript{1}, Alexei A. Lapkin\textsuperscript{1,2}

\textsuperscript{1}Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK
\textsuperscript{2}Cambridge Centre for Advanced Research and Education in Singapore, CARES Ltd., Singapore

Correspondence
Alexei A. Lapkin, Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge CB3 0AS, UK.
Email: aal35@cam.ac.uk

Funding information
Engineering and Physical Sciences Research Council, Grant/Award Number: EP/R009902/1; Pharma Innovation Platform Singapore; Prime Minister’s Office, Singapore under its Campus for Research Excellence and Technological Enterprise; National Research Foundation

Abstract
Product design for formulations is an active and challenging area of research. The new challenges of a fast-paced market, products of increasing complexity, and practical translation of sustainability paradigms require to re-examine the existing theoretical frameworks to include the advantages deriving from the new reality of digitalization of business and research. In this work, we review the existing approaches, clearly stating the role of automation and machine-learning-guided optimization in the broader framework. Moving from this, we review the state of the art of automated hardware and software for formulated product design, and identify the open challenges for future research. Perspectives are given on the emerging fields of automated discovery, scale-up, and multistage optimization, and a unitary picture of the existing connections is provided, in the general context of a completely digital R&D workflow.

KEYWORDS
automation, formulated products, machine learning, product design, robotics

1 | INTRODUCTION

Formulated products consist of a blend of ingredients, processed to achieve a set of desired performance and appearance characteristics.\textsuperscript{1} The aim of formulated product design is to find a product that exhibits a behavior, corresponding to desired, customer-defined functional properties.\textsuperscript{1} Formulations are ubiquitous in daily life, ranging from medicines to cosmetic creams and gels, from detergent powders and liquids to processed foods, paints, adhesives, lubricants, pesticide granules, and many more. Because of the significance of formulations markets, the developments in formulation technologies is attracting attention of both academia and industry.\textsuperscript{2} In 2009, Chemical Product Engineering has been introduced as the third paradigm within the field of chemical engineering.\textsuperscript{3} The design of formulated products involves identification of target product attributes, determination of product form, selection of ingredients, development of processing steps, as well as economic and environmental analyses.\textsuperscript{4} As a result of this conceptual and empirical complexity, research has focused on identification of a theoretical framework for formulated products design, taking into consideration all of these interlined areas.\textsuperscript{5} Within this general framework, it is clear that access to: (i) a large number of reliable and repeatable data, and (ii) better models, would be key elements for faster and efficient formulated products development. The former challenge can be addressed by adopting robotic automated high-throughput experimentation, whereas the latter can be met by the adoption of data efficient statistical machine learning (ML) models.

The automation of chemical experiments and advances in ML algorithms to guide automated experiments has recently emerged as a new paradigm for chemical R&D\textsuperscript{6,7} in robotic experimental platform for nanomaterial discovery,\textsuperscript{8,9} design of experiments (DoE) for high-
dimensional statistical learning,\textsuperscript{10} synthesis planning,\textsuperscript{11} discovery of reactions,\textsuperscript{12} and optimization of process conditions through ML.\textsuperscript{13} Automation and digitalization of R&D are also offering significant advantages to discovery and optimal design of formulated products.\textsuperscript{14,15} The anticipated benefits stem from avoiding human bias and automating routine operations, while exploring highly complex multidimensional input space. As a result, these approaches are particularly suited to address the new challenges of a fast-paced market, especially with the emerging constraints of sustainability and ethics, for which rapid discovery and development are fundamental requirements.

The aim of this perspective article is to clearly state the role of this new approach within the already existing theoretical framework for formulated product design, and to discuss the different aspects of the hardware and the software to accomplish the full automation and digitalization in the field, with particular attention to the identification and discussion of the open challenges to guide future developments.

This article is structured as follows: in Section 2, we will briefly review the existing approaches to formulated product design, integrating the new approaches in the general theoretical framework; in Section 3, the state of art of automated closed-loop systems for discovery and optimization will be reviewed, identifying the key aspects and challenges for the extension of the methodology to formulation design; In Section 4, we discuss the perspectives in the field with a focus on the key challenges for discovery, scale-up, multistage design, and full digitalization.

2 | THEORETICAL FRAMEWORK FOR FORMULATED PRODUCTS DESIGN

During the past decades, enormous efforts have been made in order to develop methods and tools for product design and development in various disciplines, such as material science, chemical engineering, marketing, and management. Ng, Gani, and Dam–Johansen identified the following three approaches for product design, which are classified based on their solution strategies\textsuperscript{16}: (i) experiment-based trial-and-error approach, (ii) a physical model-based approach, and (iii) an integrated experiment-modeling approach. The three approaches are discussed in the following.

The experiment-based trial-and-error approach is the preferred and most common one for the design of formulated products. By performing experiments at all steps during the development of a formulation, the product with desired properties can be developed. One previously reported explicative example is the development of the inkjet formulation.\textsuperscript{17} In this case, with only few key ingredients, and a set of typical solvents and dispersants, it was possible for an experienced researcher to develop the optimal blend on a lab-scale, and eventually use the gathered experimental data to generate a model for future use.\textsuperscript{18} However, this approach suffers from two main drawbacks: (i) it requires a large amount of resources and is highly time-demanding, (ii) it is critically dependent on the level of expertise of an experimentalist, the past knowledge, both formal and tacit as identified by Chandrasegaran et al.\textsuperscript{19} In particular, tacit knowledge, consisting of subjective insights, intuition, and heuristic qualitative rules is not easily transferable and is usually lost with the loss of the experts in product development. Therefore, this approach would be beneficial if the number and the type of ingredients and processes conditions were limited a priori and skilled experts are involved in the process.

On the other hand, computational methods, that is, physical model-based design of formulations, were proposed in order to reduce the experimental cost and to speed up the R&D process. In the last few years, various attempts have been made to establish systematic methodologies. Computer-aided methods have been proposed for solvent design,\textsuperscript{20} mixture design,\textsuperscript{21} general molecular design,\textsuperscript{22} and so forth. A review of computer-aided molecular design (CAMD) methods for product and process design was published by Gani,\textsuperscript{4} while Ng et al reviewed significant developments, current challenges, and future opportunities in the field of chemical product design using the CAMD tools.\textsuperscript{23} A key concept in CAMD is to utilize different chemical property models for possible chemical species in the pool, formulated as a mixed-integer linear/nonlinear programming (MLP/MINLP) optimization model, then solved with numerical optimization techniques.\textsuperscript{24} The suitability of these structures for a particular task or process can be evaluated with respect to a chosen criterion (for instance, the solubility of the target compound), while considering physical and chemical constraints, as well as process constraints of varying complexity. From the solution of the optimization model, the optimal product is obtained. This approach has been first applied to the design of single molecule species with considerable success. The applications vary from the design of refrigerants\textsuperscript{25} to surfactants.\textsuperscript{26} The CAMD method was then extended to the design of mixtures and composite chemical products, and identified as computer-aided blend design (CAMBD)\textsuperscript{27} also reported as computer-aided mixture design (CAMbD).\textsuperscript{28,29} Typically, almost all CAMD/CAMbD methods use group contribution-based property prediction methods\textsuperscript{30,31} to evaluate the generated compound with respect to the specified set of desirable target properties. UNIFAC\textsuperscript{32,33} (universal quasichemical functional-group activity coefficients) and SAFT-\gamma\textsuperscript{24} (statistical associating fluid theory) demonstrated to be accurate and useful in calculating solubility, phase equilibrium, partition coefficients, and various other properties. However, one significant issue is that they rely on binary interaction parameters for every pair of groups in solution, often not available in thermodynamic properties databases.\textsuperscript{25} To address the shortcomings of the group contribution-based methods, topological indices have been introduced, which are descriptors of the chemical structure to predict the physical properties of a molecule.\textsuperscript{36} The obtained relationships are called quantitative structure property/activity relationships (QSPR/QSAR).\textsuperscript{37} These methods can take into account molecular information, such as the types of atoms and bonds, total number of atoms, and bonding between the atoms to predict physical properties; therefore, they play an important role in the design of large and complex molecules, such as pharmaceutical drugs, as they can capture the differences in conformations, isomers, or molecular structures. QSPR in surfactants studies and formulated product design has been reviewed in detail by Hu et al\textsuperscript{38} and McLeese et al.\textsuperscript{39}

An alternative is to use quantum chemistry calculation for thermodynamics estimation. The COSMO-RS (conductor like screening
model for real solvents) and COSMO-SAC (segment activity coefficient) are two of the most popular post-processing methods in COSMO solvation model, where the estimation of thermodynamics only relies on the composition-independent charge density distributions, also known as sigma profile, and molecular volume. Detailed review on those methods can be found in those references.22,29,40 Furthermore, a systematic review on available computer-aided methods and associated software tools for formulated product design can be found in Reference 36. Briefly, the model-based approaches are able to efficiently find feasible candidates within the application range of the available models. However, since the function-materials-structure-processing relations have not been developed for complex formulations, including the ones determined by nano or microscale, some target properties are hard to predict with computational tools only.41 With the increase in computational power, data-driven methods, such as ML-based models, provide an alternative way for establishing the required process-property models, in particular, when sufficient knowledge is not available.42 Compared to knowledge-based models, data-driven surrogate models do not require prior knowledge; therefore ML-based models are finding increasing use to extract structure-property relationships, particularly in the cases of complex chemical formulations and materials.36 The recent advances in molecular and material design using ML methods are summarized by Butler et al.43 Identifying suitable molecular descriptors for chemicals is still an open challenge for ML models, which may lead to further accuracy for chemical product property prediction.44 As a result, the third and final integrated experiment-modeling approach was proposed, which consists of combining the computer-aided model-based techniques with heuristic-based experimental testing and improvements of the formulation design. The integrated approach usually consists of three stages: the problem definition stage, the model-based design stage, and the experiment-based verification stage.45 In the problem definition stage, the targets are translated into a set of thermo-physical properties and into a list of categories of ingredients, which are to be included in the formulation via a knowledge base. In the model-based design stage, structured databases, dedicated algorithms, and property physical model libraries are employed for designing a candidate base-case formulation. Finally, in the experiment-based verification stage, the properties and performance of the proposed formulation are tested experimentally. Through this systematic sequence of actions, the formulation is developed.

By limiting candidate formulations to be tested, and by verifying the design in the last stage, the integrated approach is convenient by saving the time and resources (compared to the experiment-based trial-and-error approach) and increasing the accuracy of the results (compared to the physical model-based approach).

In this framework for formulation design, the integration of robotic experiments and statistical ML models would be a further step in the improvement of the integrated approaches. In this sense, this approach would combine the time and resource efficiency of robotic platforms with the fact that predictions of statistical models are only based on data, with no need of extensive first principles physical knowledge.

The approaches reviewed so far are related to the core product design, and they are based on the assumption that the target properties and the final market destination have already been identified and analyzed. That is, the core design approaches are part of a broader theoretical framework, taking into account different co-existing levels within a complex decision-making hierarchy. These have been proposed by several papers and first reviewed by Gani and Ng in 2015, focusing on product conceptualization.2 Interesting hierarchical model integration in this broader framework was identified by Fung et al., who proposed a grand model for chemical product design, which indicates the relationships between different aspects.1 It consists of a process model, a property model, a quality model, a cost model, a pricing model, an economic model, as well as factors such as company strategy, government policies, and regulations. Further elements have been added by Seider et al, proposing a model considering issues such as sustainability, company strategy, aesthetics, human senses, and so on,46 while Zhang et al included supply chain analysis for optimization of selection of product ingredients.47 Despite the large number of problem aspects included in the grand structure, there are five common key elements included in these high-level integration approaches5: for the modeling part, (i) a physico-chemical model (material properties, product structure, process condition, etc.), (ii) a rule-based model (supply chain, economics analysis) and (iii) databases are involved; for the experimental part, (iv) experimental and (v) analytical tools are also considered within the superstructure for the validation of design.

In Figure 1, we illustrate the integration of the methodologies described in this article in the pre-existing theoretical framework reported by Zhang et al.36 Briefly, the market needs to define the product and its desired properties, that can be translated into quantifiable properties functions. Once identified, the next step in the general product design is to analyze the existing knowledge in terms of preliminary information, tacit knowledge in the form of operators' expertise, and formal knowledge, derived from first principles and the available models, to define the objective functions to optimize. It is important to stress that commercial formulated products are often complex mixtures for which no predictive models are available, and the complex interactions between different ingredients and process variables are not easily translatable into predictions of the final properties, even by experienced formulators. The most common situation would be the availability of a small preliminary data set, which can be used to define reasonable constraints in the input variables space. The preliminary data can be used in combination with DoE techniques to run a first batch of experiments to maximize the information gain and start the iterative process for the multi-optimization problem. In this regard, there is still an urgent need to have fast DoE algorithms to maximize the information on different conflicting continuous and discrete targets at the same time. The lack of predictive models for most of the complex interactions and properties of formulated products suggests the superiority of DoE algorithms based on the use of data-driven surrogate statistical models, as discussed in Section 3.3. Based on predictions of such models, trained on the available experimental data, robotic platforms can quickly and reliably run experiments to
generate samples. Automated analytics can then generate a new data set and use it to iteratively train the algorithms until satisfactory conditions are found. The underpinning assumption is that the adoption of automated machinery and statistical models would significantly speed up the discovery of new products and the optimization of the conditions leading to a faster release of the product in the market.

3 | CLOSED-LOOP SYSTEMS FOR FORMULATIONS: STATE-OF-THE-ART AND CHALLENGES

The recently reported combinations of process automation, artificial intelligence and statistical models in closed-loop optimization are potentially very promising in reducing the release time of new products in the market without the need for a detailed physical knowledge of new complex systems. However, at present, only a very few papers started to address the challenges and applications of computer-guided closed loop optimization in the field of formulated products and there is a general lack of discussion on the role of these new tools in the more general framework of product design, already outlined in the literature. In this section, we provide a brief overview of the existing techniques, identifying the challenges for future research. For a thorough review of closed loop optimization in the fields of chemistry and chemical engineering, the reader can refer to Mateos et al., Houben and Lapkin, and Horbaczewskyi et al.

The common features of automated closed loops reported in the literature are: (i) a robotic platform to run experiments in an automated fashion, (ii) automated on-line and in-line analytical tools to evaluate the outcome of experiments, (iii) an algorithm suggesting new experiments to carry out, based on the predictive results of surrogate models, ideally cheap to evaluate. The trade-off between exploitation and exploration is of crucial importance in reducing the time and the resources needed for product development.

3.1 | Robotic platforms

First automated hardware for chemistry can be dated back to the late 1960s. Since then, considerable advances have been made to expand the potentialities of such a tool. For a detailed historical excursion, the reader can refer to Dimitrov et al. Robotic platforms proposed in the academic literature are mostly developed in the field of chemical reaction and very little has been proposed for the generation of complex formulated products, for which physico-chemical interactions between the ingredients have less obvious outcomes compared to reactions between chemical species. The state-of-the-art hardware can be grouped into two main categories: (i) automated continuous microfluidic platforms, and (ii) modular batch systems. Advantages
and drawbacks of both configurations are widely discussed in the existing literature\textsuperscript{53,54} and are beyond the scope of the present article. However, it must be highlighted that, while the former continuous flow devices seem extremely promising for investigating reaction conditions in an efficient and resource-undemanding way, the latter possess great advantages for the mixing, the processing of emulsions, handling of solids, and for the investigation of thermodynamics-related properties,\textsuperscript{55} for example, stability. For most formulated products, the process determining the final structure, thermodynamic state, and properties, consists in a combination of rigorously controlled mixing of ingredients at a certain temperature, and stepwise addition of different ingredients at different stages of the process. Therefore, the main challenges for the application of automated hardware to formulated products design, can be identified as: (i) producing a large number of batch samples in a relatively short time, (ii) the stringent control of mixing and temperature, (iii) accurate handling of solid ingredients, (iv) transferring samples between different bays for different unit operations (dispensing, processing, analysis, etc.), and (v) standardized flexible robotic hardware that can be easily adapted to the specific workflow.

High-throughput production of batch samples has developed at different speed and with different purposes and philosophy in industry and academia. Interesting automated systems with applications for formulated products are represented by the robotic platform currently developed by Unilever and the University of Liverpool,\textsuperscript{56} and the FORMAX system, proposed by ChemSpeed Technologies.\textsuperscript{57} The latter seems to be one of the more flexible platforms on the market for formulation preparation; it consists of 30 exchangeable robotic tools and up to 36 formulation vessels. Each formulation vessel is equipped with a stirring system with speeds up to 6000 rpm and precise temperature control. They also include liquid and gravimetric solid handling, high viscous liquids dispensing, high shear homogenization, and other robotic features like capping, gripping, gripping. The former solution is the GEOFF automated formulation robot by LABMAN, with a productivity of 24 formulations per day.\textsuperscript{58} Despite great potential, the lack of academic papers describing in detail such systems or adopting them to exploit their full functionality suggests that their integration into existing laboratories is not always straightforward, affordable, or convenient. This poses a serious question about the “democratization” of such tools for their easy exploitation and wider impact in research.

In this direction, pioneering work in the development of batch modular systems has been described by Cronin and co-workers,\textsuperscript{54,59,60} that could be adapted to the specific requirements of this type of products. To date, the developed hardware has only been used to investigate chemical reactions, with the only exception of the study of physical interactions determined by thermodynamics, which then manifests itself in complex dynamic behavior of oil droplets in a continuous water phase.\textsuperscript{55,61} Being developed for different purposes, these platforms are only able to produce one sample at a time with interstage automated cleaning of the reactionware/containers. An attempt to overcome this limitation can be found in the recent studies,\textsuperscript{62} where an automated rotating wheel, coupled with a 3D-printed dispensing element and automated syringe pumps, can allocate batches of 24 vials per run. The potential of using 3D printing technologies to build inexpensive hardware was also highlighted.\textsuperscript{63}

Systems for carrying out reactions in parallel under different conditions have been implemented by Chemspeed, Hel Ltd., and other vendors.\textsuperscript{64-81} Specifically, high- and medium-throughput facilities have been implemented in the pharmaceutical industry, where large numbers of compounds and reactions should be screened. Some examples are represented by high-capacity storage facilities handled by moving robots and/or robotic arms and miniaturized prototypes for synthesis and testing.\textsuperscript{82} Researchers at Merck have recently proposed 96-well metal microtiter plates to screen large numbers of reactions on a small scale, still highlighting that automated liquid handling requires significant investment and training, whereas solid handling is both slow and inaccurate.\textsuperscript{83} Novartis Pharma developed a high-throughput robotic system based on the use of deep well microplates with up to 480 samples.\textsuperscript{84}

Dispensing of ingredients is followed by or is simultaneous with processing of the mixtures. In most academic papers, mixing seems to be efficiently automated using magnetic stirrers activated by software-controlled magnets.\textsuperscript{54,63} However, in all the presented solutions, temperature control is not ensured and, in some cases, mixing appears to be far from ideal, stressing the need for standard thermostated mixing devices to explore different effects of mixing and shear stress on the final product. In this sense, continuous flow microdevices might be advantageous due to the fine control of the shear stress and the flexibility in attaining different mixing regimes on a small-scale. However, also in this case, the challenge of reaching similarly high mechanical stresses typical of a shaking process is still an issue, also considering that most formulated products are processed on a long time scale not easily achievable in continuous flow micro-devices. Some continuous flow high-shear mixers have been proposed,\textsuperscript{85-87} also for formulated products.

To the best of our knowledge, the only reported example in the literature of closed-loop robotic optimization for formulated products involves the use of an off-line non-automated incubator to efficiently process samples,\textsuperscript{88,89} once again demonstrating that this remains an open challenge in academia. Integration of all the main aspects of sample preparation, that is, liquid and solid dispensing, mixing, heating, and weighing, in the same platform has been achieved in some commercial systems, the most notable platforms being Chemspeed and Symyx.\textsuperscript{90} GEOFF automated formulation robot by LABMAN.

Assuming that different operations cannot be easily carried out in the same part of an automated platform, the next problem to address is the automated transfer of samples from a bay/station to another. Li et al\textsuperscript{91} proposed the use of automated guided vehicles for transferring operations in automated labs. Another common solution, robotic arms with multiple degrees of freedom, is not without its own challenges, since these are often expensive and designed to carry out only specific repetitive operations with a reduced flexibility. The problem was
effectively solved in the work by Steiner et al.\textsuperscript{54} for reactive mixtures: reactive mixtures were transferred from a batch unit operation to another (reaction, separation, purification, etc.) by pumping through automatically controlled connection channels. In the field of synthetic chemistry another industrial example of automated laboratory can be found in the system developed by Aventis Pharma,\textsuperscript{52} based on the use of robotic transferring shuttles on a rail system between different work stations. The automated lab could efficiently carry out synthesis, mixing and temperature control, and auxiliary operations such as weighing, capping and uncapping, as well as separation operations, like liquid-liquid extraction, evaporation, filtration, and drying. However, formulated products pose new challenges, such as the flow of highly viscous products, the presence of dispersed solids, multiple heterogeneous phases, and the formation of inter-molecular structures (core-shell particles, entangled molecular chains, micelles, etc.) that might be dramatically affected by the shear effects in continuous flow.

Solids handling is of primary importance in automated platforms for the optimization of formulated products, considering that solid ingredients are often dissolved in a liquid matrix or present in the final product in the form of solid dispersions or solid blends.\textsuperscript{93,94} However, none of the reviewed papers in the field of robotic platforms for the optimization, discovery, and development of chemical processes has efficiently addressed the issue, despite the fact that it would be beneficial also for chemical reactions. Several technical solutions commercially available in drug discovery had been described.\textsuperscript{95} Trap-door mechanisms with holes of different diameter are described as surprisingly accurate methods of solid dispensing, although inflexible and dependent on the packing of the material in the holes, which makes it unreliable for powders of different size distributions. Solid handling pipettes rely on vacuum,\textsuperscript{96} or electrostatic forces,\textsuperscript{97} which makes them difficult to install and limits their applicability, whereas more traditional devices are based on the use of Archimedes screws.\textsuperscript{94} Other commercially available solutions are included in the Chemsped's gravimetric dispensing unit,\textsuperscript{98} the Mettler-Toledo's dispensing stations,\textsuperscript{99} and the REDI Zinsser Analytic Calli robotic powder handling\textsuperscript{100}; at present, there are no publications reporting their integration in robotic platforms for reaction and formulations development and was recently stressed that no general automated solid handling solution currently exists.\textsuperscript{83}

One final remark is the need for standardized unit platforms to be combined together in different ways for a faster and cheaper exploitation of robotic laboratory technology. Standardization is a fundamental requirement for commercialization and application of technologies on a large scale. Moreover, standardization of robotic hardware and software would enable faster implementation of communication between different platforms, efficient collaboration between researchers with different backgrounds, and creations of networks of different platforms working on the same task, even remotely. The need of modularity\textsuperscript{53,54} and standardization\textsuperscript{60} has been already highlighted and partially tackled in the field of reaction development. However, future research is fundamental to extend the range of applications and to make this relatively new tool accessible, available, and usable for scholars with different expertise.

### 3.2 Analytics

Automated analytical tools are of crucial importance for the fast and efficient adoption of robotic platforms for formulation development and they represent the ongoing bottleneck to the wide adoption of such systems in academia and industry. Once again, most of the analytics automated in the literature has been used for reaction development. Through reviews can be found in Mateos et al\textsuperscript{18} and Houben et al.\textsuperscript{13} The most common adopted techniques for reactive systems are UPLC and HPLC,\textsuperscript{59,111,112} GC,\textsuperscript{37} and MS.\textsuperscript{37} Turbidimetric (NTU), conductivity, zeta potential, and droplets size distribution, in the case of emulsions. Therefore, more complex, analytical sensors need to be identified and integrated in the robotic platforms to acquire data about different properties at the same time. Other important properties are functional performances, that is, for example, UV protection of solar creams, or other sensory properties like odor, stickiness, and so forth. A first step in the automation of sensory properties measurement is represented by the new robotic tactile systems SynTouch.\textsuperscript{111}

One key property of several commercial formulated products, ranging from detergents to personal care products, is their external appearance, which can be quantified using discrete and continuous variables. The former can be defined as “stability” which is related to the capability of the system to not show phase separation, whereas the latter can be quantified considering their absorbance spectra in the visible range and their turbidity value, measured in nephelometric turbidity unit (NTU). Phase separation can be evaluated through automated image processing from automated cameras. Automated cameras and image processing coupled to robotic platforms have already been proposed in other contexts.\textsuperscript{55,61}

Very recently, Cao et al \textsuperscript{88,89} have adapted a robotic platform to measure turbidity of a commercial detergent, based on the adoption of a cheap LED and a light sensor on a moving 3D-printed supporting frame. The LED and the light sensor are fixed on the opposite sides of a vial containing the samples, on a rotating wheel accommodating up to 24 vials. The electrical signal can be converted to the turbidity value in NTU, based on a calibration with turbidity standards. Following an analogous protocol, moving probes have been proposed to automatically measure pH values\textsuperscript{62} and the same can be applied to conductivity measurements. pH measurements can be also carried out by the FORMEX platform by Chemspeed Technologies and the automated GEOFF Labman platform. At present, there are a few
examples of closed-loop optimization optimizing particle size\textsuperscript{120} and viscosity,\textsuperscript{88} whereas no example of zeta potential and surface tension in such loops have been found to date. However, in the few above-mentioned examples, both dynamic light scattering (DLS) and viscosity measurements were manually carried out offline. A promising alternative to carry out DLS and Zeta potential analysis in an automated fashion seems to be represented by the new Malvern Zetasizer coupled with an automated autosampler, however, there are still no examples of integration of such pieces of equipment in the automated platforms at present. Automation of DLS analysis was reported by Zhao et al.\textsuperscript{84} using a DynaPro light scattering system (Wyatt Santa Barbara, CA), with an average throughput of 30 samples per hour.

Automated viscosity measurements can be a challenging task, especially for high viscosity and non-Newtonian fluids. One example of a semi-automated capillary viscometer can be found in Neumann et al.\textsuperscript{118} In this case, viscosity is calculated from the measurement of pressure drop in a capillary in which the fluid is flowing. An automated syringe pump can dispense the fluid sample and the system can perform cleaning cycles with a solvent in between the measurement runs. The described device has been successfully adopted to characterize rheology of both Newtonian and non-Newtonian low-viscosity fluids. Desmukh et al.\textsuperscript{121} also proposed a similar system, based on the analysis of the mass flow behavior or modeling of the pressure profile along the tips of multiple pipettes. In this case, it is claimed that the system is rapid and parallelized, allowing analysis of more than 100 samples in less than an hour, although accurate testing was only shown for Newtonian fluids. The main challenges associated with this type of devices are the narrow pressure range and solvent compatibility of pressure sensors, the accurate temperature control, and the need for smooth and pulseless dispensing to have accurate measurements. Further research will have to extend the range of usability of these devices and to integrate them in more comprehensive workflow processes, for the autonomous production and analysis of liquid formulations. An alternative solution is represented by the coupling of robotic arms with traditional rotational viscometers.\textsuperscript{122} Commercially available systems for automating viscosity measurements are illustrated by the GEOFF robot and the Phil CUP/BOB rheometer by LABMAN automation, and the high-throughput rheometer Anton Parr HTR 502.

Finally, it is worth mentioning other examples of high-throughput automated assessment of less obvious and easily quantifiable properties of formulated products reported in the literature: among these: the dirt removal efficacy of different cleaning systems,\textsuperscript{123} and color, glossiness, homogeneity, friction, and other mechanical properties of coatings proposed by the FORMAX platform by Chemspeed Technologies (https://www.chemspeed.com/formax/).

3.3 | Algorithms

Robotic platforms can iteratively provide data points to train DoE algorithms, suggesting new conditions in order to optimize the input variables with respect to one or more target functions.
In contrast, methods which formulate optimization problem to place new samples were also reported. Cozad et al proposed the automated learning of algebraic models for optimization, which is a surrogate modeling tool where a derivative-free optimization problem is solved to maximize deviation of the surrogate model prediction error in order to place the next sampling point.\(^{151}\) Detailed review of the existing algorithms has been published\(^ {152}\) and is beyond the scope of this article. However, it is worth outlining the general needs of the formulated product development and identify the key aspects for future research.

i. Formulations can be complex physico-chemical systems for which no existing physical models are readily available. The use of cheap-to-evaluate black-box surrogate statistical models is particularly suited to model the responses of the products to variations in the input space.

ii. In formulation design, it is extremely important to consider multiple, often conflicting, targets and performance criteria. There is no commercially relevant formulation which does not have to meet several targets at the same time in terms of final properties (aspect, fragrance, touch, viscosity, stability, etc), costs, and environmental impact. In this sense, several of the proposed single-objective optimization algorithms are completely inadequate. Combining several targets in one single objective, that is, scalarization, is a possibility as shown, for example, in the multi-objective active learner\(^ {120}\) methods. However, this is not ideal, since it requires prior knowledge, introduces bias, and often is not straightforward.\(^ {153}\) Successful implementation of multitarget optimization has been so far achieved for continuous variables using the Thompson sampling efficient multiobjective algorithm (TS-EMO).\(^ {10,154}\)

iii. The sustainability challenge imposes targets for rapid development of new formulations or substitutions of some ingredients with others, as environmental legal requirements and consumers’ ethics become more and more stringent.\(^ {155}\) As a result, algorithms need to be fast and models cheap to evaluate, also in exploring a high-dimensional combinatorial space. In addition, both discrete and continuous variables and target performances need to be efficiently taken into account at the same time.

iv. The main drawback of black-box surrogate models is that they generally do not provide any information about the physics underpinning product’s functional performance. In this sense, the use of data collected from closed-loop optimization procedure for generation of physical knowledge is crucial in gaining a better understanding of the processes to rapidly adapt and transfer the results to similar systems. Very preliminary results in this sense can be identified in the physical interpretation of models hyperparameters,\(^ {10}\) the manual interpretation of Pareto fronts by human experts,\(^ {88,89}\) and, more recently, the automated capture of chemical intuition transferred between similar systems,\(^ {156}\) and the automated generation of physical laws from data.\(^ {118}\)

v. As in the case of hardware, there will be a general need for user-friendly open-source software interfaces, to enable experimentalists to apply the developed techniques regardless of their specific field of expertise and democratize the use of such tools.

4 | PERSPECTIVES AND OTHER CHALLENGES IN FORMULATION DESIGN

4.1 | Product discovery and prediction of scale-up

Although laboratory automation has already demonstrated a remarkable increase in experimental throughput, discovery of new phenomenon and/or product is still challenging. Automation alone is insufficient, as the relative rate of discovery does not change with the increase in experimental throughput. An appealing alternative is to implement the process of curious and knowledge-based inquiry inherent to human scientific research, within a reliable and high-throughput robotic system.\(^ {157}\) Active searching and pooling strategies were proposed and applied in automated discovery of new chemical reactions. A detailed review of screening approaches in chemical reaction discovery and development can be found in Collins et al,\(^ {158}\) Coley et al,\(^ {159}\) Henson et al.\(^ {52}\)

A pioneering work in the study of multicomponent chemical formulations discovery is the investigation of the self-propelled droplet system. Grizou et al describe an experimental method complemented by a curiosity algorithm, which enables to observe more variety of droplet behaviors than the random parameter search under the same experimental budget.\(^ {160}\) This approach may lead to new discoveries with potential applications in formulation chemistry.

One of the main open challenges in product design by using automated approaches is the translation of the acquired knowledge to full predictive scalability. At present, in the field of chemistry and chemical engineering, most of the data collected in lab-scale robotic platforms is used to build statistical reaction models,\(^ {161}\) not taking into account scale-dependent or process-dependent interactions. At a process scale, however, mass, heat, and momentum transport almost always become the most relevant controlling mechanisms, and very little information about them can be inferred from small-scale black-box optimizations. Specifically, in the field of formulated products, this can critically determine the thermodynamic state of the final product, that is, for example, its stability, shelf life, physical properties, and so forth. One promising way to overcome these limitations is to use robotic experiments to build generalizable physical knowledge and to learn physical models that can be integrated at a later stage to predict the behavior of a chemical system at scale. However, self-optimizing loops do not usually provide useful information to develop physical models,\(^ {48}\) and model-based design of experiments (MBDoE) is usually applied to closed-loop systems for the generation of more informative data.\(^ {114}\)

A number of studies, representing the first steps to overcome these limitations, have been published, addressing interpretability of surrogate models and the automated generation of fundamental knowledge in the form of physical models. This approach has been
successfully applied mainly for the identification of predictive kinetic models in the field of reaction engineering and chemical reaction development.\textsuperscript{138} Moreover, the current state of art focused on the problems of model selection out of a library of pre-derived models\textsuperscript{162-165} and parameter estimation\textsuperscript{163,166} and there are no examples in the literature of automated ex novo generation of kinetic models from data. The identification of the physical models in formulated products is even more challenging as predictive process modelling has to take into account thermodynamic properties, for example, solubility, mass transfer, and heat transfer between phases and within a phase, and fluid dynamics.

Robotic and automated platforms have the potential of running large numbers of experiments to explore transport phenomena and thermodynamic properties. On the other hand, a correct definition of MBDoE to properly design experiments for predictive models development is still an open challenge. Nevertheless, some examples of automated hardware for measurement of physical properties have recently been published, such as measurements of gas solubility in liquids.\textsuperscript{167} Also, several pioneering works\textsuperscript{55,61,168} have shown advances in both the hardware and the software for the exploration of physics-controlled phenomena in microfluidics, for example, multiphase droplets generation. A future challenge will be to use algorithms to find predictive correlations between the behaviors observed at the lab scale and the final properties of the processed products at the production scale. The underlying hypothesis, still to be proven or rejected, is that a direct correlation can be found between the final properties of a processed sample, for example, stability of oil-in-water emulsion, and the observable behavior of the two phases in a robotic platform, like the one proposed by Henson and Points.\textsuperscript{55}

Finally, an interesting emerging approach to infer physical knowledge from the increased availability of good quality data is the adoption of non-parametric form-free algorithmic search. In particular, the ML method of symbolic regression seems to be particularly promising, allowing to discover analytical equations describing a set of data combining a set of basic operators. Recent developments consisted in the formulation of symbolic regression as a mixed-integer nonlinear programming to ensure solving to the global optimality. The methodology has been successfully applied to simulated data to re-discover the equations governing different physical systems.\textsuperscript{169} More recently, some successful examples of application to experimental noisy data have been proposed. Specifically, Neumann and co-workers used this approach to derive simple kinetic laws and constitutive rheology equations for Newtonian and non-Newtonian fluids, showing the potential and the limitations of the present formulation of the algorithm.\textsuperscript{138} Despite the promising results, this methodology is still at its early stage and future research will be needed to reduce the computational time required to derive physical models from noisy data in the presence of a large number of variables and operators, allowing the data-driven search for complex physical models.

The challenge of discovery and scale-up are intimately interconnected and of fundamental importance for speeding up product design and development under the constraints of the fast-changing market. Ideally, given a product of interest with certain properties, an algorithm would suggest libraries of possible combinations of ingredients and process conditions for its manufacture and rank them according to different criteria, such as practical feasibility, environmental impact, cost, and so forth. This may be done based on approaches similar to those currently being developed for retro- and forward synthesis in synthetic chemistry, and/or through their integration with high-throughput experimentation for discovery. The top-ranked pathways may then be explored from a mechanistic point of view to derive different candidate physical models. Finally, MBDoE closed-loop systems may be used to discriminate between different candidate models and estimate the parameters for the most reliable one. Also, black-box optimization on a lab-scale can provide useful data and a first insight into favorable process conditions. This paradigm would generate detailed predictive models and information to be used on a later stage of the process development for the efficient and fast scale up of the process (Figure 2).

### 4.2 Multistage design

Multistage experiments and processes are prevalent in both industry and academic research. Here we adopt the definition of a multistage experiment\textsuperscript{170} as an experiment that uses the same or different experimental units in multiple stages with distinct responses measured at each stage. Apart from the model for each stage, a “general” model which describes the response from the n-th stage to the factors from 1, 2, ..., n − 1, n stages. The use of the cumulative models indicates that different stages of experiments are related and have to be designed in conjunction with each other. A common example in formulated product is the design of cleaning products such as shampoo, washing liquid and so forth. In stage 1, key ingredients such as surfactants, polymers and so forth will be mixed under certain conditions, achieving a stable product within a specified viscosity range. In stage 2, the mixture will be diluted 10 times, and the foaming and cleaning ability will be measured. The factors applied in stage 1 will have influence on stage 2. There are three different types of optimal designs for multistage experiments; (i) multistage completely randomized, (ii) multistage split-plot, and (iii) multistage strip-plot designs. Matthew et al\textsuperscript{170} introduced a two-stage split-plot design for the formulation of a pharmaceutical product, the modeling of the similarity factor from dissolution testing, and the prediction of points which maximize the probability of passing specification. However, this work was limited to computational study without experimental validation. Also, only single objective was considered in the case study. Due to the complex nature of formulated products, future work could therefore focus on extending and adapting existing methodology to exploring outside the current experimental region. For example, Pareto optimality can be used for multiobjective optimization, Bayesian reliabilities can be applied to identify compromise directions for exploration of design spaces for multiple output, and so forth.

In other applications, such as healthcare\textsuperscript{171} traffic management\textsuperscript{172} smart grid management\textsuperscript{173,174} and others\textsuperscript{175}, multistage design problem is often treated as multistage stochastic optimization problem (two-stage, linear models as an example):
Min \( c_1 x_1 + c_2 x_2 \) subject to \( A_1 x_1 \geq b_1, \ E_1 x_1 + A_2 x_2 \geq b_2 \). \hspace{1cm} (1)

It can be further interpreted as a two-stage decision process, exemplified in Scheme 1, which means in the first stage, trial feasible values \( x_1 \) will be given for \( x_1 \), then the optimal solution of the second stage function can be found based on Equation (2)

\[
\text{Min } c_2 x_2 \text{ subject to } A_2 x_2 \geq b_2 - E_1 x_1
\]

Dynamic programming (DP) algorithms can then be used to solve the sequential decision process described above. The DP approach has many attractive features, such as extendibility to multistage problems, accommodation of discrete values and nonlinearities, and so on. In the DP approach, the first-stage problem is defined as:

\[
\text{Min } c_1 x_1 + \alpha_1(x_1) \text{ subject to } A_1 x_1 \geq b_1
\]

where \( c_1 x_1 \) is the immediate cost, and \( \alpha_1(x_1) \) is defined as:

\[
\alpha_1(x_1) = \text{Min } c_2 x_2 \text{ subject to } A_2 x_2 \geq b_2 - E_1 x_1
\]

\( \alpha_1(x_1) \) represents the future cost of decision \( x_1 \), that is, the consequences of this decision for the second-stage problem. The DP algorithm constructs the future cost function \( \alpha_1(x_1) \) by discretizing \( x_1 \) into a set of trial values \( \{x_{1i}, i=1, \ldots, n\} \) and solving Equation (4) for each of the trial values. Intermediate values of \( \alpha_1(x_1) \) are obtained by interpolation from the neighboring discretized states. One of the main drawbacks for this approach is the need to discretize the decisions \( x_1 \) for the calculation of the future cost \( \alpha_1(x_1) \), as it may lead to a large number of combinations to enumerate even for a modest number of variables. Possible ways of addressing the dimensionality problem were proposed in the literature, such as approximate the future cost function by an analytical function rather than a set of discrete values.
More detailed explanations of the corresponded methodologies can be found in the recent review.177

4.3  |  A complete digital workflow

Further development of the formulation technology is likely to follow the general trend of complete digitalization and automation of chemical R&D and manufacture. As previous sections have shown, introduction of robotics and high-throughput equipment in formulations screening and development results in significant gains in performance. But this also results in the increased pressure on process analytical technologies, which frequently become the bottlenecks, on data management and knowledge abstraction, as these tasks can no longer be done manually. To fully digitize the workflow of design and development of formulations a similar set of challenges as in the general field of chemistry should be addressed:

- standardization of data representation and data exchange formats,
- creation of data repositories,
- standardization of knowledge representation and exchange formats, and
- standardization of equipment and software interfaces.

All of these aspects are currently under active discussion and development. Here we shall point to only few most significant developments that are relevant to the field of formulations:

The core of the fully digital workflow consists of the mechanisms for exchange of data and of knowledge models. There exist a number of formats for representing molecules, such as SMILES,178 InChI,179 MDL molfiles,180 as well as of crystallographic and spectroscopic data. None of these formats allow a complete representation of a molecule or a material and several representations may be required to cover all the necessary information. For this, a data exchange repository must exist. There are commercial (e.g., Reaxys,181 CAS,182 etc) and public (e.g., ChemSpider183) chemical databases, however, they do not yet support the fully digital workflow. Very recent proposal to establish the new repository Chemotion offers a potential platform for a fully digital workflow of chemical R&D.184 Lab information management systems (LIMS) are another powerful tool for automatically collected data; originally conceived for automated data tracking and exchange, they have been further integrated with data mining, analysis, and translation tools, allowing for much wider range of applications.185 One example is the large LIMS systems for formulated products developed by Syngenta, operated at their unique formulation robot at the Jeallott’s Hill R&D site.

To navigate from data to research hypothesis, data must be transferred to knowledge and knowledge structures be accessible to algorithms and to scientists. This requires establishment of ontologies—relationship models, and mechanisms of access to different knowledge domains, for example, via the semantic web technology.186 This infrastructure then enables to exploit the power of algorithms and of automation, linking laboratories located anywhere in the world with computational resources, algorithm developers, material and process scientists, and end-user product specialists. An example of such infrastructure, a virtual world of chemical processes, filled with data, knowledge models, and AI agents, was recently demonstrated by Kraft et al.187 Then the outcome is the connected world of machines, algorithms, data, knowledge and scientists. The benefits to the scientific community and to society are clear: faster sharing of knowledge, a much wider access to talent, much better utilization of resources, both material and computational, and ability to pose much more challenging problems.

5  |  CONCLUSIONS

In this review, we highlight the role of the recent developments in the fields of robotic automated platform and optimization for formulated product design, identifying the challenges for future research efforts. Fully robotically automated closed-loop optimization is configured as a promising hybrid approach for fast and efficient product development, combining the advantages of fast and reliable data collection and targeted experimentation guided by surrogate models, without the need for human expertise.

The main challenges associated to the hardware of such systems can be identified in the need of high-throughput platforms that can safely and efficiently handle solid ingredients, process all the components in a controlled and precise fashion and transfer them to other platforms. There is also a general lack of automated analytics for non-traditional properties assessment. It is clear in the literature the need for more flexible, modular, and standardized platforms to allow access to a large number of combinations from users with different levels of expertise. Models should focus on efficiently treating high-dimensional spaces, also allowing for the choice between different candidate ingredients for the optimization of both continuous and discrete competing outputs.

Discovery of new products in an automated fashion is also a promising research area that would allow for new unbiased approaches to efficiently identify and select new products to be optimized. In this sense, it also highlighted the need to apply the field of multistage optimization to experimental case studies in the field of chemical engineering and product development, to improve the overall process. Finally, automated interpretation of surrogate models and targeted DoE approaches need to be developed to allow for the translation of the information gain at full scale. The perspective is a full digitalization of the complete workflow and democratization of such tools, with consequent benefits for both academia and industry.

ACKNOWLEDGMENTS

This study was co-funded by EPSRC grant “Combining Chemical Robotics and Statistical Methods to Discover Complex Functional Products” (EP/R009902/1) (Danilo Russo, Alexei A. Lapkin), National Research Foundation (NRF), Prime Minister’s Office, Singapore under...
its Campus for Research Excellence and Technological Enterprise (CREATE) program as a part of the Cambridge Centre for Advanced Research and Education in Singapore Ltd (CARES), (Liwei Cao and Alexei A. Lapkin), and Pharma Innovation Platform Singapore (PIPS, Projects “Development of multi-step processes in pharma” and “Data2Knowledge”) (Alexei A. Lapkin).

DATA AVAILABILITY STATEMENT
Data sharing not applicable to this article as no datasets were generated or analysed during the current study

ORCID
Liwei Cao https://orcid.org/0000-0002-7639-8022
Alexei A. Lapkin https://orcid.org/0000-0001-7621-0889

REFERENCES
1. Fung KY, Ng KM, Zhang L, Gani R. A grand model for chemical product design. Comput Chem Eng. 2016;91:15-27. https://doi.org/10.1016/j.compchemeng.2016.03.009.
2. Gani R, Ng KM. Product design—molecules, devices, functional products, and formulated products. Comput Chem Eng. 2015;81:70-79. https://doi.org/10.1016/j.compchemeng.2015.04.013.
3. Hill M. Chemical product engineering—the third paradigm. Comput Chem Eng. 2009;33(5):947-953. https://doi.org/10.1016/j.compchemeng.2008.11.013.
4. Gani R. Chemical product design: challenges and opportunities. Comput Chem Eng. 2004;28(12):2441-2457. https://doi.org/10.1016/j.compchemeng.2004.08.010.
5. Ng KM, Gani R. Chemical product design: advances in and proposed directions for research and teaching. Comput Chem Eng. 2019;126:147-156. https://doi.org/10.1016/j.compchemeng.2019.04.008.
6. Sanderson K. Automation: chemistry shoots for the moon. Nature. 2019;568(7753):577-579. https://doi.org/10.1038/d41586-019-01246-y.
7. Lindsey JS. A retrospective on the automation of laboratory synthetic chemistry. Chemom Intel Lab Syst. 1992;17(1):15-45. https://doi.org/10.1016/0169-7439(92)90025-B.
8. Miras HN, Cooper GJT, Long DL, et al. Unveiling the transient template in the self-assembly of a molecular oxide nanowheel. Science. 2010;327(5961):72-74. https://doi.org/10.1126/science.1181735.
9. Sans V, Cronin L. Towards dial-a-molecule by integrating continuous flow, analytics and self-optimisation. Chem Soc Rev. 2016;45(8):2032-2043. https://doi.org/10.1039/c5cs00793c.
10. Schweidtmann AM, Clayton AD, Holmes N, Bradford E, Bourne RA, Lapkin AA. Machine learning meets continuous flow chemistry: automated optimization towards the Pareto front of multiple objectives. Chem Eng J. 2018;352:277-282. https://doi.org/10.1016/j.cej.2018.07.031.
11. Coley CW, Thomas DA, Lummis JAM, et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. Science. 2019;365(6453):eaax1566. https://doi.org/10.1126/science.aax1566.
12. Mcnally A, Haffemeyer B, Collins BSL, Gaunt MJ. Palladium-catalysed C–H activation of aliphatic amines to give strained nitrogen heterocycles. Nature. 2014;510(7503):129-133. https://doi.org/10.1038/nature13389.
13. Houben C, Lapkin AA. Automatic discovery and optimization of chemical processes. Curr Opin Chem Eng. 2015;9:1-7. https://doi.org/10.1016/j.coche.2015.07.001.
14. Clomburg JM, Crumbley AM, Gonzalez R. Industrial biomanufacturing: the future of chemical production. Science. 2017;355(6320):aaq0804. https://doi.org/10.1126/science.aaq0804.
15. Patrick JF, Robb MJ, Sottos NR, Moore JS, White SR. Polymers with autonomous life-cycle control. Nature. 2016;540(7633):363-370. https://doi.org/10.1038/nature21002.
16. Ng KM, Gani R, Dam-Johansen K. Chemical Product Design: Towards a Perspective Through Case Studies. Vol 23. Amsterdam, Netherlands: Elsevier; 2006.
17. Cummins G, Desmulliez MPY. Inkjet printing of conductive materials: a review. Circuit World. 2012;38(4):193-213. https://doi.org/10.1108/03056121212280413.
18. Tam SK, Fung KY, Poon GSH, Ng KM. Product design: metal nanoparticle-based conductive inkjet inks. AIChE J. 2016;62(8):2740-2753. https://doi.org/10.1002/aic.15271.
19. Chandrasegaran SK, Ramani K, Srimad RD, et al. The evolution, challenges, and future of knowledge representation in product design systems. Comput Des. 2013;45(2):204-228. https://doi.org/10.1016/j.cad.2012.08.006.
20. Struebing H, Ganase Z, Karamertzian PG, et al. Computer-aided molecular design of solvents for accelerated reaction kinetics. Nat Chem. 2013;5(11):952-957. https://doi.org/10.1038/nchem.1755.
21. Mattei M, Conte E, Kontogeorgis GM, Gani R. Prediction of thermophysical properties of liquid formulated products. Prod Des Eng Formul Gels Pastes [published online]. 2013;121-151. https://doi.org/10.1002/9783527654741.ch5.
22. Austin ND, Sahinidis NV, Trahan DW. Computer-aided molecular design: an introduction and review of tools, applications, and solution techniques. Chem Eng Res Des. 2016;116:2-26. https://doi.org/10.1016/j.cherd.2016.10.014.
23. Ng LY, Chong FK, Chemmangattuvalappil NG. Challenges and opportunities in computer-aided molecular design. Comput Chem Eng. 2015;81:115-129. https://doi.org/10.1016/j.compchemeng.2015.03.009.
24. Cignitti S, Zhang L, Gani R. Computer-aided framework for design of pure, mixed and blended products. Comput Aided Des Chem Eng. 2015;37:2093-2098. https://doi.org/10.1016/B978-0-444-63576-1.50043-1.
25. Sahinidis NV, Tawarmalani M, Yu M. Design of alternative refrigerants via global optimization. AIChE J. 2003;49(7):1761-1775. https://doi.org/10.1002/aic.690490714.
26. Zhang L, Cignitti S, Gani R. Generic mathematical programming formulation and solution for computer-aided molecular design. Comput Chem Eng. 2015;78:79-84. https://doi.org/10.1016/j.compchemeng.2015.04.022.
27. Achenie L, Venkatasubraman V, Gani R. Computer Aided Molecular Design: Theory and Practice. Amsterdam: Elsevier; 2003:121-393.
28. Austin ND, Samudra AP, Sahinidis NV, Trahan DW. Mixture design using derivative-free optimization in the space of individual component properties. AIChE J. 2016;62(5):1514-1530. https://doi.org/10.1002/aic.15142.
29. Austin ND, Sahinidis NV, Konstantinov IA, Trahan DW. COSMO-based computer-aided molecular/mixture design: a focus on reaction solvents. AIChE J. 2018;64(1):104-122. https://doi.org/10.1002/aic.15871.
30. Franklin JL. Prediction of heat and free energies of organic compounds. Ind Eng Chem. 1949;41(5):1070-1076. https://doi.org/10.1021/i1ei00473a041.
31. Hukkerikar AS, Sarup B, Abildskov J, Sin G, Gani R. Group-contribution + (GC +) based estimation of properties of pure components: improved property estimation and uncertainty analysis. Fluid Phase Equilib. 2012;321:25-43. https://doi.org/10.1016/j.fluid.2012.02.010.
32. Fredenslund A, Jones RL, Prausnitz JM. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. AIChE J. 1975;21(6):1086-1099. https://doi.org/10.1002/aic.690210607.
33. Gani R, Brignole EA. Molecular design of solvents for liquid extraction based on UNIFAC. Fluid Phase Equilib. 1983;13:331-340.
34. Chapmanis WG, Gubbins KE, Jackson G, Radosz M. SAFT: equation-of-state solution model for associating fluids. Fluid Phase Equilibria. 1989;52:31-38.

35. Austin ND, Sahinidis NV, Trahan DW. A COSMO-based approach to computer-aided mixture design. Chem Eng Sci. 2017;159:93-105. https://doi.org/10.1016/j.ces.2016.05.025.

36. Zhang L, Mao H, Liu Q, Gani R. Chemical product design—recent advances and perspectives. Curr Opin Chem Eng. 2020;27:22-34. https://doi.org/10.1016/j.coche.2019.10.005.

37. Kier LB, Hall LH. Molecular Connectivity in Structure-Activity Analysis. New York: Wiley; 1986.

38. Hu J, Zhang X, Wang Z. A review on progress in QSAR studies for surfactants. Int J Mol Sci. 2010;11(3):1020-1047. https://doi.org/10.3390/ijms11031020.

39. McLeese SE, Eislick JC, Hoffmann NJ, Scorto AM, Camarda KV. Design of ionic liquids via computational molecular design. Comput Chem Eng. 2010;34(9):1476-1480. https://doi.org/10.1016/j.compchemeng.2010.02.017.

40. Gertig C, Leonhard K, Baradow A. Computer-aided molecular and processes design based on quantum chemistry: current status and future prospects. Curr Opin Chem Eng. 2020;27:89-97. https://doi.org/10.1016/j.coche.2019.11.007.

41. Bernardo FP, Saraiva PM. A conceptual model for chemical product design. AIChE J. 2015;61(3):802-815. https://doi.org/10.1002/acid.24681.

42. Zhang L, Mao H, Liu L, Du J, Gani R. A machine learning based computer-aided molecular design/screening methodology for fragrance molecules. Comput Chem Eng. 2018;115:295-308. https://doi.org/10.1016/j.compchemeng.2018.04.018.

43. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559 (7715):547-555. https://doi.org/10.1038/s41586-018-0337-2.

44. Amar Y, Schleidtman AM, Deutsch P, Cao L, Lapkin A. Machine learning and molecular descriptors enable rational solvent selection in asymmetric catalysis. Chem Sci. 2019;10(27):6697-6706. https://doi.org/10.1039/C9SC01844A.

45. Kontogeorgis GM, Mattei M, Ng KM, Gani R. An integrated approach for the design of emulsified products. AIChE J. 2019;65(1):75-86. https://doi.org/10.1002/aiic.16363.

46. Seider WD, Lewin DR, Seader JD, Widagdo S, Gani R, Ng KM. Product and Process Design Principles, Synthesis, Analysis and Evaluation. 4th ed. New York: Wiley; 2017.

47. Zhang L, Fung KY, Zhang X, Fung HK, Ng KM. An integrated framework for designing formulated products. Chem Eng J. 2017;107:61-76. https://doi.org/10.1016/j.cej.2017.05.014.

48. Mateos C, Nieves-Romacha MJ, Rincón JA. Automated platforms for reaction self-optimization in flow. React Chem Eng. 2019;4(9):1536-1544. https://doi.org/10.1039/c9re00011f.

49. Jeral MI, Holmes N, Akien GR, Bourne RA. Enhanced process development using automated continuous reactors by self-optimisation algorithms and statistical empirical modelling. Tetrahedron. 2018;74 (25):3158-3164. https://doi.org/10.1016/j.tet.2018.02.061.

50. Hrbaczewskiy JS, Willans CE, Lapkin AA, Bourne RA. An introduction to closed-loop process optimization and online analysis. Green Chemical Engineering, Vol 12. 1st ed. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA; 2018. 329-369.

51. Dimitrov D, Kreisbeck C, Becker JS, Aspuru-Guzik A, Saikin SK. Autonomous molecular design: then and now. ACS Appl Mater Interfaces. 2019;11(28):24825-24836. https://doi.org/10.1021/acsami.9b01226.

52. Henson AB, Gromski PS, Cronin L. Designing algorithms to aid discovery of chemical robots. ACS Cent Sci. 2018;4(7):793-804. https://doi.org/10.1021/acscentsci.8b00176.

53. Bédard AC, Adamo A, Aroh KC, et al. Reconfigurable system for automated optimization of diverse chemical reactions. Science. 2018;361 (6408):1220-1225. https://doi.org/10.1126/science.aat0650.
73. Buitrago E, Lundberg H, Andersson H, Ryberg P, Adolfsson H. High throughput screening of a catalyst library for the asymmetric transfer hydrogenation of heteroaromatic ketones: formal syntheses of (R)-fluoxetine and (S)-duloxetine. ChemCatChem. 2012;4(12):2082-2089. https://doi.org/10.1002/ccct.201200308.

74. Majoros Li, Dekeyser B, Hoogenboom R, Fijten MWM, Haucourt N, Schubert US. Solution prepolymerization as a new route for preparing aliphatic polyurethane prepolymers using high-throughput experimentation. J Polym Sci Part A Polym Chem. 2009;47(15):3729-3739. https://doi.org/10.1002/pola.23382.

75. Fenster E, Long TR, Zang Q, et al. Automated synthesis of a 184-member library of thiadiazepan-1,1-dioxide-4-ones. ACS Comb Sci. 2011;13(3):244-250. https://doi.org/10.1021/co100606x.

76. Stoll N, Allwardt A, Dingerdissen U, Thurov K. An 8-fold parallel reactor system for combinatorial catalysis research. J Autom Methods Manag Chem. 2006;2006:1-9. https://doi.org/10.1115/JAMMC.2006/56329.

77. Nasrullah MJ, Bahr JA, Gallagher-Lein C, Webster DC, Roessler RR, Schmitt P. Automated parallel polyurethane dispersion synthesis and characterization. J Coatings Technol Res. 2009;6(1):1-10. https://doi.org/10.1007/s11998-008-9102-z.

78. Jallia I, Beauvaine C, Beauvière S, et al. Automated synthesis of a 96 product-sized library of triazole derivatives using a solid phase supported copper catalyst. Molecules. 2010;15(5):3087-3120. https://doi.org/10.3390/molecules15053087.

79. Buchstaller H-P, Anlauf U. Parallel solution-phase synthesis of a 2-aminothiazole library including fully automated work-up. Comb Chem High Throughput Screen. 2011;14(2):104-108. https://doi.org/10.2139/astj.v14i2.4740.

80. Anthoni J, Chebil L, Lionneton F, Magdalou J, Humeau C, Ghoul M. Automated analysis of synthesized oligorutin and oligoesculin by HPLC. J Chromatogr A. 2011;1218(9):964-970. https://doi.org/10.10113/j.cn11-v11-072.

81. Tse NMK, Kennedy DF, Moffat BA, Kirby N, Caruso RA, Drummond CJ. High-throughput preparation of hexagonally ordered mesoporous silica and gadolinioisilicate nanoparticles for use as MRI contrast agents. ACS Comb Sci. 2012;14(8):443-450. https://doi.org/10.1021/acs.combchem.100035q.

82. Schneider G. Automated drug discovery. Nat Rev Drug Discov. 2018;17(2):97-113. https://doi.org/10.1038/nrd.2017.232.

83. Shevlin M. Practical high-throughput experimentation for chemists. ACS Med Chem Lett. 2017;8(6):601-607. https://doi.org/10.1021/acsmedchemlett.7b00165.

84. Zhao H, Graf O, Milovic N, et al. Formulation development of antibodies using robotic system and high-throughput laboratory (HTL). J Pharm Sci. 2010;99(5):2279-2294. https://doi.org/10.1002/jps.22008.

85. Tamminen J, Koiranen T. Mixing performance comparison of milliscal continuous high-shear mixers. Can J Chem Eng. 2015;93(12):2245-2252. https://doi.org/10.1002/cjce.22327.

86. Vanarase AU, Osorio JG, Muzzio FJ. Effects of powder flow properties and shear environment on the performance of continuous mixing of pharmaceutical powders. Powder Technol. 2013;246:63-72. https://doi.org/10.1016/j.powtec.2013.05.002.

87. Håkansson A. Rotor-stator mixers: from batch to continuous mode of operation—a review. Processes. 2018;6(4):32. https://doi.org/10.3390/pr6040032.

88. Cao L, Russo D, Felton KC, et al. Optimisation of formulations using robotic experiments driven by machine learning DoE. Cell Rep Phys Sci. 2021;2:100295.

89. Cao L, Russo D, Mauw W, Gao HH, Lapkin AA. Machine learning-aided process design for formulated products. Comput Aided Chem Eng. 2020;48:1789-1794. https://doi.org/10.1016/B978-0-12-823377-1.50299-8.

90. Vegas AJ, Anderson DG. High-throughput approaches. Polymer Science: A Comprehensive Reference, 10 Volume Set. Vol 9. Amsterdam: Elsevier; 2012:457-484. https://doi.org/10.1016/B978-0-44-53349-4.00231-4.

91. Li S, Yan J, Li L. Automated guided vehicle: the direction of intelligent logistics. Proceedings of the 2018 IEEE International Conference on Service Operations and Logistics, and Informatics. SOLI 2018. Singapore: Institute of Electrical and Electronics Engineers Inc.; 2018: 250-255. https://doi.org/10.1109/SOLI.2018.8476726.

92. Weber A, Von Roedern E, Sliz HJ. SynCar: an approach to automated synthesis. J Comb Chem. 2005;7(2):178-184. https://doi.org/10.1021/cj0498382.

93. Tran P, Pyo Y-C, Kim D-H, Lee S-E, Kim J-K. Overview of the manufacturing methods of solid dispersion technology for improving the solubility of poorly water-soluble drugs and application to anticancer drugs. Pharmaceuticals. 2019;11(3):132. https://doi.org/10.3390/pharmaceutics11030132.

94. Couillaud BM, Espeau P, Mignet N, Corvis Y. State of the art of pharmaceutical solid forms: from crystal property issues to nanocrystals formulation. ChemMedChem. 2019;14(1):8-23. https://doi.org/10.1002/cmdc.201800612.

95. Reader J. Automation in medicinal chemistry. Curr Top Med Chem. 2005;4(7):671-686. https://doi.org/10.2174/1568026043451069.

96. Fermier AM, Troisi J, Heritage EC, Drexel MA, Gallea P, Swinney KA. Powder dispensing robot for sample preparation. Analyst. 2003;128(6):790-795. https://doi.org/10.1039/b300274h.

97. Wong SK-F, Lu Y, Heineman W, Palmer J, Courtney C. Fully automated solid weighing workstation. J Biomol Screen. 2005;10(5):524-531. https://doi.org/10.1177/108705705275457.

98. https://www.chemspeed.com/video/flex-powder-l/. Accessed February 11, 2021.

99. https://www.mt.com/ca/en/home/products/Laboratory_Weighing_Solutions/Automated_Sample_Preparation/Automated_Powder_Liquid_Dispensing.html. Accessed February 11, 2021.

100. http://www.zinsensa.com/tools_powderhandling.htm. Accessed February 11, 2021.

101. Baumgartner LM, Dennis JM, White NA, Buchwald SL, Jensen KF. Use of a droplet platform to optimize Pd-catalyzed C-N coupling reactions promoted by organic bases. Org Process Res Dev. 2019;23(8):1594-1601. https://doi.org/10.1021/acs.oprd.9b00236.

102. Waldron C, Pankajakshan A, Quaglio M, Cao E, Galvanin F, Gavrilidis A. An autonomous microreactor platform for the rapid identification of kinetic models. React Chem Eng. 2019;4(9):1623-1636. https://doi.org/10.1039/c8re00345a.

103. Hone CA, Boyd A, O’Kearney-McMullan A, Bourne RA, Muller FL. Definitive screening designs for multistep kinetic models in flow. React Chem Eng. 2019;4(9):1565-1570. https://doi.org/10.1039/c9re00180h.

104. Hone CA, Holmes N, Aikin GR, Bourne RA, Muller FL. Rapid multi-step kinetic model generation from transient flow data. React Chem Eng. 2017;2(2):103-108. https://doi.org/10.1039/c7re00109b.

105. Rodriguez-Garcia M, Surman AJ, Cooper GJT, et al. Formation of oligopeptides in high yield under simple programmable conditions. Nat Commun. 2015;6(11):1-7. https://doi.org/10.1038/ncomms9385.

106. Zhu C, Raghuvanshi K, Coley CW, et al. Powder mixing device for one-step kinetic model generation. React Chem Eng. 2018;5(4):8567-8570. https://doi.org/10.1039/c8re00465f.

107. Reizman BJ, Jensen KF. Simultaneous solvent screening and reaction optimization in microfluidic slugs. Chem Commun. 2015;51(68):13290-13293. https://doi.org/10.1039/C5CC03651H.

108. Holmes N, Aikin GR, Blacker AJ, Woodward RL, Meadows RE, Bourne RA. Self-optimisation of the final stage in the synthesis of...
computer experiments. SIAM J Sci Comput. 2011;33(4):1948-1974. https://doi.org/10.1137/090761811.

145. Zhou Q, Shao X, Jiang P, Zhou H, Shu L. An adaptive global variable fidelity metamodeling strategy using a support vector regression based scaling function. Simul Model Pract Theory. 2015;59:18-35. https://doi.org/10.1016/j.simpat.2015.08.002.

146. Zhou Q, Shao X, Jiang P, Gao Z, Zhou H, Shu L. An active learning variable-fidelity metamodeling approach based on ensemble of metamodels and objective-oriented sequential sampling. J Eng Des. 2016;27(4-6):205-231. https://doi.org/10.1080/09544828.2015.1135236.

147. Eason J, Cremaschi S. Adaptive sequential sampling for surrogate model generation with artificial neural networks. Comput Chem Eng. 2014;68:220-232. https://doi.org/10.1016/j.compchemeng.2014.05.021.

148. Vu KK, D’Ambrosio C, Hamadi Y, Liberti L. Surrogate-based methods for black-box optimization. Int Trans Oper Res. 2017;24(3):393-424. https://doi.org/10.1111/itor.12292.

149. Regis RG. Constrained optimization by radial basis function interpolation for high-dimensional expensive black-box problems with infeasible initial points. Eng Optim. 2014;46(2):218-243. https://doi.org/10.1080/0305215X.2013.765000.

150. Boukouvala F, Ierapetritou MG. Surrogate-based optimization of expensive workflow modeling for continuous pharmaceutical manufacturing. J Pharm Innov. 2013;8(2):131-145. https://doi.org/10.1007/s12247-013-9154-1.

151. Wilson ZT, Sahinidis NV. The ALAMO approach to machine learning. Comput Chem Eng. 2017;106:785-795. https://doi.org/10.1016/j.compchemeng.2017.02.010.

152. Garud SS, Karimi IA, Kraft M. Design of computer experiments: a review. Comput Chem Eng. 2017;106:71-95. https://doi.org/10.1016/j.compchemeng.2017.05.010.

153. Abraham A, Jain L. Evolutionary multiobjective optimization. Evolutionary Multiobjective Optimization. London: Springer-Verlag; 2005:1-6. https://doi.org/10.1007/1-84628-137-7_1.

154. Bradford E, Schweidtmann AM, Lapkin A. Efficient multiobjective optimization employing Gaussian processes, spectral sampling and a genetic algorithm. J Glob Optim. 2018;71(2):407-438. https://doi.org/10.1007/s10898-018-0609-2.

155. Kleinhekote J, Flettmann L, Bachmann M, et al. Life cycle assessment of an active learning variable-fidelity metamodeling approach based on ensemble of metamodels and objective-oriented sequential sampling. J Eng Des. 2015;27(04-06):205-231. https://doi.org/10.1080/09544828.2015.1135236.

156. Gil Y, Greaves M, Hendler J, Hirsh H. Amplify scientific discovery with artificial intelligence: many human activities are a bottleneck in progress. Science (80- ). 2014;346(6206):171-172. https://doi.org/10.1126/science.1259439.

157. Collins KD, Gensch T, Glorius F. Contemporary screening approaches to reaction discovery and development. Nat Chem. 2014;6(10):859-871. https://doi.org/10.1038/nchem.2062.

158. Coley CW, Eyke NS, Jensen KF. Autonomous discovery in the chemical sciences part II: outlook. Angew Chem Int Ed. [published online]. 2019;59:23414-23436. https://doi.org/10.1002/anie.201909989.

159. Grizou J, Points LJ, Sharma A, Cronin L. A curious formulation robot enables the discovery of a novel protein-bacteriophage. Sci Adv. 2020;6(5):eaay4237. https://doi.org/10.1126/sciadv.aay4237.

160. Ehly M, Gemperline PJ, Nordon A, Littlejohn D, Basford JK, De Cecco M. Scale-up of batch kinetic models. Anal Chim Acta. 2007;595(1-2):80-88. https://doi.org/10.1016/j.aca.2007.02.040.

161. Galvanin F, Cao E, Al-Rifai N, Gavrilliadis A, Dua V. Model-based design of experiments for the identification of kinetic models in microreactor platforms. Comput Aided Chem Eng. 2015;37:323-328. https://doi.org/10.1016/B978-0-444-63578-5.S0049-9.

162. Galvanin F, Cao E, Al-Rifai N, Gavrilliadis A, Dua V. A joint model-based experimental design approach for the identification of kinetic models in continuous flow laboratory reactors. Comput Chem Eng. 2016;95:202-215. https://doi.org/10.1016/j.compchemeng.2016.05.009.

163. Galvanin F, Marchesini R, Barolo M, Bezzo F, Fidaleo M. Optimal design of experiments for parameter identification in electrolysis models. Chem Eng Res Des. 2016;105:107-119. https://doi.org/10.1016/j.cherd.2015.10.048.

164. Zhou Q, Shao X, Jiang P, Zhou H, Shu L. An adaptive global variable fidelity metamodeling strategy using a support vector regression based scaling function. Simul Model Pract Theory. 2015;59:18-35. https://doi.org/10.1016/j.simpat.2015.08.002.

165. McMillen JP, Jensen KF. Rapid determination of reaction kinetics with an automated microfluidic system. Org Process Res Dev. 2011;15(2):398-407. https://doi.org/10.1021/op100300p.

166. Galvanin F, Cao E, Al-Rifai N, Gavrilliadis A, Dua V. Model-based design of experiments for the identification of kinetic models in microreactor platforms. Comput Aided Chem Eng. 2015;37:323-328. https://doi.org/10.1016/B978-0-444-63578-5.S0049-9.
180. Dalby A, Nourse JG, Hounshell WD, et al. Description of several chemical structure file formats used by computer programs developed at molecular design limited. *J Chem Inf Comput Sci*. 1992;32(3):244-255. https://doi.org/10.1021/ci00007a012.

181. RELX Intellectual Properties SA, Reaxys—Reaxys is a trademark, copyright owned by RELX Intellectual Properties SA and used under licence. https://www.reaxys.com/, https://www.elsevier.com/solutions/reaxys. Accessed December 9, 2020.

182. CAS Databases | CAS. https://www.cas.org/support/documentation/cas-databases. Accessed December 9, 2020.

183. ChemSpider | Search and share chemistry. http://www.chemspider.com/. Accessed December 9, 2020.

184. Tremouilhac P, Lin C, Huang P, et al. The repository chemotion: infrastructure for sustainable research in chemistry. *Angew Chem Int Ed* [published online]. 2020;59:22771-22778. https://doi.org/10.1002/anie.202007702.

185. Skobelev DO, Zaytseva TM, Kozlov AD, Perepelitsa VL, Makarova AS. Laboratory information management systems in the work of the analytic laboratory. *Meas Tech*. 2011;53(10):1182-1189. https://doi.org/10.1007/s11018-011-9638-7.

186. Inderwildi O, Zhang C, Wang X, Kraft M. The impact of intelligent cyber-physical systems on the decarbonization of energy. *Energ Environ Sci*. 2020;13(3):744-771. https://doi.org/10.1039/c9ee01919g.

187. Elbeck A, Lim MQ, Kraft M. J-Park simulator: an ontology-based platform for cross-domain scenarios in process industry. *Comput Chem Eng*. 2019;131:106586. https://doi.org/10.1016/j.compchemeng.2019.106586.

How to cite this article: Cao L, Russo D, Lapkin AA. Automated robotic platforms in design and development of formulations. *AIChE J*. 2021;67:e17248. https://doi.org/10.1002/aic.17248