Faraday Discussions

Cite this: DOI: 10.1039/d5fd00066a

PAPER

Harnessing data and control with AI/ML-driven polymerization and copolymerization†

Rigoberto Advincula, * Ilia Ivanov, Rama Vasudevan, Rajeev Kumar, Panagiotis Christakopoulos, Marileta Tsakanika, Jihua Chen, Jan Michael Carillo, Qinyu Zhu and Bobby Sumpter

Received 3rd May 2025, Accepted 11th August 2025
DOI: 10.1039/d5fd00066a

Creating and curating new data to augment heuristics is a forthcoming approach to materials science in the future. Highly improved properties are advantageous even with "commodity polymers" that do not need to undergo new synthesis, high-temperature processes, or extensive reformulation. With artificial intelligence and machine learning (AI/ML), optimizing synthesis and manufacturing methods will enable higher throughput and innovative directed experiments. Simulation and modeling to create digital twins with statistical and logic-derived design, such as the design of experiments (DOE), will be superior to trial-and-error approaches when working with polymer materials. This paper describes and demonstrates protocols for understanding hierarchical approaches in optimizing the polymerization and copolymerization process *via* AI/ML to target specific properties, using model monomers such as styrene and acrylate. The key is self-driving continuous flow chemistry reactors with sensors (instruments) and real-time ML with an online monitoring set-up that allows a feedback loop mechanism. We provide initial results using ML refinement of the classical Mayo–Lewis equation (MLE), time-series data, and an autonomous flow reactor system build-up as a future data-generating station. More importantly, it lays the ground for precision control of the copolymerization process. In the future, it should be possible to undertake collaborative human–AI-guided protocols for the autonomous fabrication of new polymers guided by literature and available data sources targeting new properties.

Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. E-mail: advincularc@ornl.gov

† This manuscript has been co-authored by UT-Battelle, LLC, under contract DEAC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (**https://www.energy.gov/downloads/doe-public-access-plan**).

Introduction

Artificial Intelligence (AI) and machine learning (ML) have revolutionized many aspects of science, engineering, industry, and society.[1] Statistics and algorithms with physics-based equations have been developed by many scientists through decades of research ranging from Bayesian, regression, decision tree models, unsupervised clustering, dimensional reduction, regression, neural networks, and several deep learning (DL) or reinforced learning (RL) methods. In chemistry, the central science, relying on physical laws that govern the kinetics and thermodynamics of molecular design, intrinsic and extrinsic properties, energy, and mechanistic insights on various chemical reaction pathways is possible. With macromolecules or polymers, the high molecular weight, molecular weight distribution (MWD), linearity or branching microstructure, and end group functionality offer almost an infinitesimally large number of possible combinations that pose a challenge in defining structure–composition–processing–property (SCPP) relationships. This requires a new language beyond simple IUPAC nomenclature in chemistry, *e.g.*, simplified molecular line input systems (SMILES) with graphed-based definitions and context-free language codes.[2] Yet, the potential to make this high and hierarchical correlation is to link the atomic, nanoscale, and mesoscale phenomena to the macroscopic property. A key element is using advanced spectroscopic and microscopic characterization techniques to investigate phenomena at both time and space scale hierarchies. For example, real-time observation of the polymerization mechanism (step-growth or chain addition) can only be matched by the time scale of the probe phenomena (spectroscopic, microscopic, dielectric, *etc.*). With AI/ML, optimizing synthesis or characterization experiments can be made closer to heuristics. Using sensors capable of a feedback loop mechanism and the ability to use simulation to create digital twins, optimizing properties can be accomplished in close to real-time. Statistical and logic-derived designs are starting points for the design of experiments (DOE) or principal component analysis (PCA) protocols for optimization. This contrasts with many wasteful trial-and-error approaches when working with polymer materials (synthesis, mixing, formulation, *etc.*) in many industries.

This paper will give an example of ML-driven approaches in the synthesis, nanostructuring, and characterization of the polymerization process that is critical for any AI/ML-driven optimization and advanced manufacturing, using model monomers such as styrene and acrylate.[3–5] At the Center for Nanophase Materials Sciences (CNMS), Oak Ridge National Laboratory (ORNL), we have gained much insight into developing the methods and instrumentation for ML-driven reactions with advanced characterization methods and access to high-performance computing (HPC) resources.[6] This will have significant implications in using large language models (LLM) and other training/learning sets for optimizing specific properties and applications in soft matter. With continuous flow reaction chemistry,[7,8] and/or tandem with time-stamped batch or semi-batch reactions, it is possible to incorporate a new instrumentation design application programming interface (API) in mechatronics or robotics. *In situ* or *operando* hyphenated analytical tool automation to provide a continuous feedback loop for optimization. Digital twins developed with atomistic to coarse grain methods, including

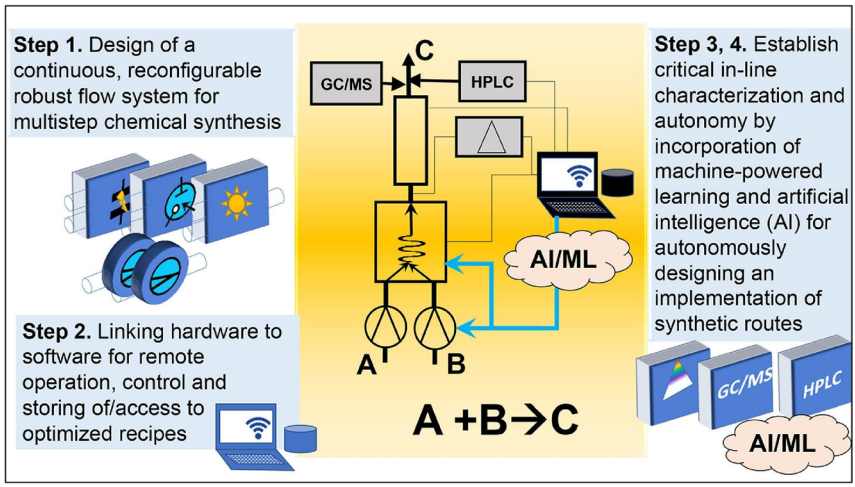

Fig. 1 Here, we describe the steps for developing an AI/ML-driven autonomous continuous flow reactor synthesis, using model monomers such as styrene and acrylate. (The figure is from ref. 10 without change, under creative commons license CC BY-NC-ND 4.0.[11]). More engineering parametrization of organic or polymerization reactions can be done in a plug flow and continuous flow reactor system with a built-in sensor and edge server that controls and monitors the reaction.

molecular dynamics (MD) and density functional theory (DFT), make it possible to translate the molecular or macromolecular level connectivity towards optimized synthesis protocols. The plan is to integrate (Fig. 1): (1) a controlled and well-designed continuous flow chemistry reactor or separator (plug flow, packed bed, membrane separated, *etc.*), (2) an ML-driven workflow for *in situ* multimode chemical/materials characterization[9] that can include instruments such as NMR, ESR, IR, Raman, UV-Vis, GC-MS, and HPLC, with flow adoptors; (3) use of an edge server or computer for integration of control and feedback/analysis/data storage of the control variables; (4) a core software stack that can be enabled for deep learning (DL) and reinforcement learning (RL); and (5) access to on-demand computing architectures that can parse calculations to compute resources needed. These computational resources include light-weight edge, mid-level edge (NVIDIA DGX-2), and connectivity to high-performance computing (HPC) or even exascale computing resources (Frontier). We demonstrate preliminary results and potential on how this ML-driven autonomous reactor system can enhance the ability to deliver homopolymers, copolymers, site-substituted molecules, and deuterated molecules.

The microstructure and copolymer composition are essential in controlling the properties of polymer materials. The Mayo–Lewis equation (MLE) transformed copolymerization practice and theory with a correlation of co-monomer and copolymer composition as reactivity ratios (rr). Various forms of linear to nonlinear equations, all based upon the terminal model (TM) of copolymerization, have been developed to estimate rr values through the best fit of experimentally determined copolymer compositions with comonomer composition and/or monomer conversion, for monomers including styrene, acrylate, and others. Linear regression methodologies have been replaced by powerful

nonlinear numerical methods that provide a statistically refined estimation of the rr. However, the MLE is not suitable for some systems where the MLE fails to capture the influence of penultimate unit effects, depolymerization, and system (*e.g.*, solvent, concentration, pH) media. Discrepancies reported in the initial monomer/initiator and final copolymer composition can also be linked to kinetic fundamentals. Still, there is a lack of appreciation of reaction engineering and unit operations for copolymer synthesis. There is a need to demonstrate the use of ML and autonomous ML methods in continuous flow chemistry to update the MLE.[12]

Several key questions can be asked:

(1) How can the various mechanistic models for polymerization (step-growth, chain-addition, metathesis, *etc.*) be augmented by ML-driven control of the reaction in a unit operation?

(2) How can we harness theory and simulation (MD, DFT, SCFT, *etc.*) to create the polymerization process's effective and hierarchical digital twins?

(3) How can we utilize LLMs appropriately or effectively in defining the objectives or the training data for specific polymerization reactions or converting it to a ChatGPT experience? Is this necessary?

(4) What are the recent breakthroughs at the intersection of materials science, chemistry, and applied computing that can specifically accelerate the discovery of new polymer synthesis methods and properties?

(5) What are the critical challenges in developing self-driving autonomous reactors with homogeneous or heterogeneous platforms (with sensors and instruments) that are fundamentally relatable to physical and chemical principles in the polymerization process?

(6) How can we bridge the gap of decision-making timescales with the initially set ML-driven variables and, at the same time, take advantage of LLMs in polymer literature? How helpful is this exercise, and is the human-in-the-loop necessary?

In a recent ORNL-led workshop, "Shaping the Future of Self-Driving Autonomous Laboratories", the meeting brought together scientists, computer experts, chemists, and other AI/ML experts to discuss the need for more AI-driven autonomous laboratories.[13] In polymers, it is possible to revisit the high potential of self-driving laboratories to contribute much to science and manufacturing. In the future, there is a high potential for automated metadata collection systems for polymer science powered by AI and the creation of hybrid AI systems that combine data-driven learning with fundamental scientific principles. However, a big question in the future is the need for human oversight and expertise, the "expert-in-the-loop", while leveraging automation, ensuring that human scientific intuition and creativity are preserved and enhanced rather than eliminated in polymer science.

Methodologies and simulation

We first considered different correlational and causational methods, including chemistry and polymer science databases. There are several levels and granularity: molecular and polymer structure descriptors (SMILES, PSMILES, GRAPH, *etc.*), Quantitative Structure–Activity Relationship (QSAR), Polymer Handbook databases, polymer–property data tables and constants with various levels of quality and age, interpolation and extrapolation (not accurate). We also

considered the modeling of polymer properties that are application-relevant and scalable. Lastly, we looked at the Chat-GPT-like experience with the LLM tool. The bottom line is the need for new data acquisition and mining stations for polymer science, which will enable (1) rapid screening of material properties, (2) a feedback loop for synthesis to property correlation, (3) integration of ML into the workflow of empirical synthesis and characterizations, and (4) an autonomous-self-driving experimental set-up that serves as a closed-loop operation with theory-modeling feedback for real-time recalculation or first principles molecular design. In the process, we have identified several challenges and potential for efficiently accelerating time and effort toward optimization and discovery in polymer materials.

Despite being extensively adopted in the pharmaceutical industry, continuous-flow polymerization is just emerging as a powerful technology for precision in polymer synthesis.[10] The advantages of rapid mixing and good heat/matter transfer in a tubular reactor offer instantaneous initiation or termination of polymerizations with fast kinetics. One of the most critical parameters for polymerization underflow is the efficacy of mixing based on residence time distribution (RTD); a small RTD will result in well-controlled polymerization about monomer conversion, polymer molecular weight, and dispersity.[10]

Sampling and flow reactors

We have used several commercial flow reactors and instrumentations for sampling, including programmable gantry-type dosing and sample collection in the lab:

An automated reactor sampling system: Mettler-Toledo EasySampler for an aliquot sample and programmable unattended representative sampling and dosing beyond pipettes and syringes were used; a JKEM programmable robotic dosing and sample platform. These are used for time-sampling batch reactions. For continuous flow reaction chemistry, we have used the following: (1) a Phoenix Flow Reactor system with high temperature/high-pressure capabilities to enable synthesis in a vast parameter space not achievable with standard laboratory equipment (lower pressures, ambient temperature, less variable flow rates). This system is capable of temperatures up to 450 °C and pressures up to 200 bar. The H-Genie hydrogen generator is for on-demand hydrogenation and various modules: a gas module, mixer module, pressure module, and a systems controller and software. (2) Vapourtec reactor: a continuous flow or plug flow chemistry (R-series) with temperature, pump (pressure), and mixing control. (3) The Snapdragon Chemistry (Cambrex) is the most modular and integrated system with its laboratory operating system (LabOS). It has operability up to large-scale GMP commercial manufacturing. All the systems, beyond our demonstration for polymerization, are helpful for reaction kinetics screening, photochemistry and electrochemistry, and continuous – extraction, separation, and crystallization.

Python scripting and user-interface development in flow chemistry

In this case, a user interface can control the self-optimizing flow reaction, which includes controlling the equipment through a software interface, generating training data, reading sensor signals, evaluating optimization objectives, and performing further iterations in a feed-back loop protocol. The four variables

optimized are (1) the molar ratio of reagents or initiators, (2) flow rate, (3) temperature, and (4) pressure.

The pumps and temperature controllers were controlled *via* the existing LabOS API used by the vendor (SnapDragon), which utilizes a web server interface from which customized experimental workflows can be set up within the environment. However, this setup is somewhat limited in flexibility and features. To facilitate programming *via* python, we utilized simple http requests to interrogate sensor readings and adjust the necessary flow rates directly *via* python's requests module, which we ran from a separate machine that could communicate with the server running LabOS. This way, we could separate the 'action' commands from the central command and control server. We further connected several independent instruments to this server, including the FT/IR, Raman, UV-Vis, and a quartz crystal microbalance (QCM) system to enable characterization data to be available during the experiment. Although python APIs were available for both the UV-Vis and QCM systems, making integration simple, neither the FT/IR nor the Raman systems allowed for external control and API limitation. Instead, we relied on automated batch data collection where spectra are continually acquired every *N* minute and stored in a folder, which was interrogated based on time stamps to link with the currently active experiment. Data visualization was provided to the user through plotly[14] dash web applications, and data standardization was accomplished through the pycroscopy ecosystem's *sidpy* package.[15] The entire setup enabled the easy setup of customized experimental protocols, entirely written in python and utilizing one or several of the available characterization tools. A connection to a GPU-enabled server (Nvidia DGX-2 workstation) was available for advanced data analysis.

Based on the Mayo–Lewis model, we established a minimal coarse-grained molecular dynamics (CGMD) simulation protocol of batch copolymerization. Using the LAMMPS simulation package,[16] we implemented the fix bond/create command to facilitate the creation of bonds during the polymerization process.[17–19] The simulations used an implicit solvent model, assuming good solvent conditions for both polymer species. The initial conditions include a predetermined quantity of dimer initiators and free monomers randomly distributed in the simulation box. Instead of directly setting up the reaction rate constants in the kinetic equations, we assigned the probability of each bond formation event. We captured the varying reactivity ratios associated with the limiting cases outlined in Table 1. Specifically, the probabilities of the various bond-forming events are denoted as P_1, P_2, P_3, and P_4, respectively. Thus, these probabilities allow the calculation of reactivity ratios r_1 and r_2 as $r_1 = P_1/P_2$ and $r_2 = P_3/P_4$, respectively.

Polymerization and reaction procedures

All the reactions were done in homogeneous solutions in batch-time series aliquot or plug flow continuous flow reaction chemistry, homopolymers, and copolymers to demonstrate principles and possibilities for ML-directed reactions. Batch-time series were done at various concentrations and monomer 1, 2, and initiator ratios at specific solvents and temperatures. The Easysampler (Mettler-Toledo) was used to both sample and quench the reaction for later analysis with NMR under a time-stamp. We used the Thalesnano and the Snapdragon

Table 1 Reactivity ratios, definition, and limiting cases defining sequence-morphology of copolymer. The terminal model (active polymer chain end may have different selectivity) was introduced along with the reactivity ratios[23–25]

Propagation steps	Definition	Limiting cases	Example
$M_1^* + M_1 \xrightarrow{k_{11}} M_1M_1^*$	Reactivity ratio indicates preference of M_1 to add M_1 vs. M_2 expressed as a ratio of corresponding propagation rate constant $r_1 = \frac{k_{11}}{k_{12}}$	1. r_1 and $r_2 \ll 1$ azeotrope with copolymer composition matching that of monomers	1.
$M_1^* + M_2 \xrightarrow{k_{12}} M_1M_2^*$	Reactivity ratio of M_2, presents probability of M_2 to add M_2 *vs.* M_1 $r_2 = \frac{k_{22}}{k_{21}}$	2. $r_1,\ r_1 \cong 0$ alternating copolymer (ideal)	2. …ABABAB……
$M_2^* + M_2 \xrightarrow{k_{22}} M_2M_2^*$		3. $r_1,\ r_1 \cong 1$ random copolymer	3. AABABAABB
$M_2^* + M_1 \xrightarrow{k_{21}} M_2M_1^*$		4. $r_1 \cong r_2 > 1$ mixture of homopolymers, block copolymer	4. AAAAA or BBBBB + AAAAAABBBBB
		5. $r_1 \cong r_2 \gg 1$ homopolymers	5. AAAAA or BBBBB
		6. $r_1 \gg r_2$ gradient copolymer, started with M_1 block and transitioning to M_2	6. AAAAAAABBABBB
		7. $k_{11} \gg k_{12}, k_{22} \gg k_{21}$ block copolymer	7. AAAAA or BBBBB

system for continuous flow reaction polymerization. The Snapdragon system was used for multi-modal monitoring and control, together with a separate edge server and computer for remote access to data.

ML algorithm and digital twins

Developing a Digital Twin for Copolymerization reaction is a virtual replica of an experimental polymerization system that enables predictive simulation, optimization, and data generation across various process parameters. The twin can serve as a mini computational laboratory for designing and understanding copolymerization reactions, allowing chemists and engineers to visualize multivariable space and explore multiple what-if scenarios without physical experiments. Digital twin combines reaction models (Mayo–Lewis, kinetic Monte Carlo), detailing initiation, propagation termination, and chain transfer steps, incorporating temperature feed strategy (continuously steering tank reactor, plug flow reactor, or mixed reactor scheme). The Mayo–Lewis copolymer equation predicts copolymer composition from reactivity ratios.[20–22] The mole fraction of each monomer in the copolymer can be related to the initial feed composition and the reactivity ratios. This correlation helps predict the copolymer composition under different feed ratios and is crucial in designing copolymers with specific monomer distributions. In this work, we use monomers including styrene and acrylate, as examples.

The reactivity ratios, *Q*–*e* scheme, developed by Alfrey and Price,[26] relates the reactivity ratios of monomers to their electronic and steric parameters. The *Q* (reactivity parameter) and *e* (polarity parameter) values are assigned to each monomer to capture electronic and steric influences on reactivity. Monomers with similar *Q*–*e* values tend to copolymerize more randomly, while those with disparate values lead to more blocky structures.[24,26] Mark–Houwink–Sakurada correlation links reactivity ratios with intrinsic viscosity parameters to understand how monomer choice and reactivity ratios affect the molecular weight distribution in copolymers. It's often used with the Flory–Huggins interaction parameter to predict polymer properties related to solubility and compatibility in copolymer blends.[27–29]

Cross-term reactivity correlation

Cross-term reactivity correlation in the copolymerization rate equations can be analyzed by looking at interactions between r_1 and r_2 that might predict compatibility, miscibility, or mechanical properties in block and random copolymers.

In principle, all existing approaches for predicting the morphology of the copolymer can be combined. The computational output of the digital twin comprises time-dependency of reactants–product dependency, reactivity ratios (calculated using the Kelen–Tüdös method if not known), copolymer composition (based on the reactivity ratios), molecular weight, and polydispersity.

For example, the changes in the reactivity ratio of methacrylic acid are commonly explained in terms of solvent–monomer association, radical stabilization, solvent electron and H-donating properties, as well as solvent dielectric properties.[30–33]

Analysis of the reactivity ratio values of monomers for participating in copolymerization reactions can be used to predict the copolymer sequence.[22–25,34] For instance, when the values of reactivity ratios of monomers M_1 and M_2 are around one, $r_1 = r_2 \sim 1$, one should expect a random sequence of copolymer. An alternating copolymer product is anticipated for r_1, $r_2 < 0.3$, M_1 preferred homopolymer ($r_1 > 1.0$, $r_2 < 0.1$), M_2 preferred homopolymer ($r_1 < 0.1$, $r_2 > 1.0$), ladder-copolymer ($-0.01 < r_1 < 0.05$, $r_2 = -0.01 < r_1 < 0.05$), gradient copolymer ($r_1 = 0.5$, $r_2 = 2.0$, $0.4 > r_1 < 0.7$, $0.2 < r_2 < 0.4$).

One would not expect a sharp transition between the regions with different sequences but a gradient transition from one sequence to another (producing a mixture of copolymer morphologies). Applying these extended conditions to selecting reactivity ratios for styrene copolymerization extracted from CopolyDB, we generated a graph of log-scale-reactivity ratios where the predicted copolymer sequence are color coded. Tuning the value of the reactivity ratios of monomers towards the sequence of copolymer may be challenging to achieve, especially in the case of gradient copolymers.

Ultimately, assuming copolymer production is the target of the research effort, the yield and the rate of copolymer production can be optimized using digital

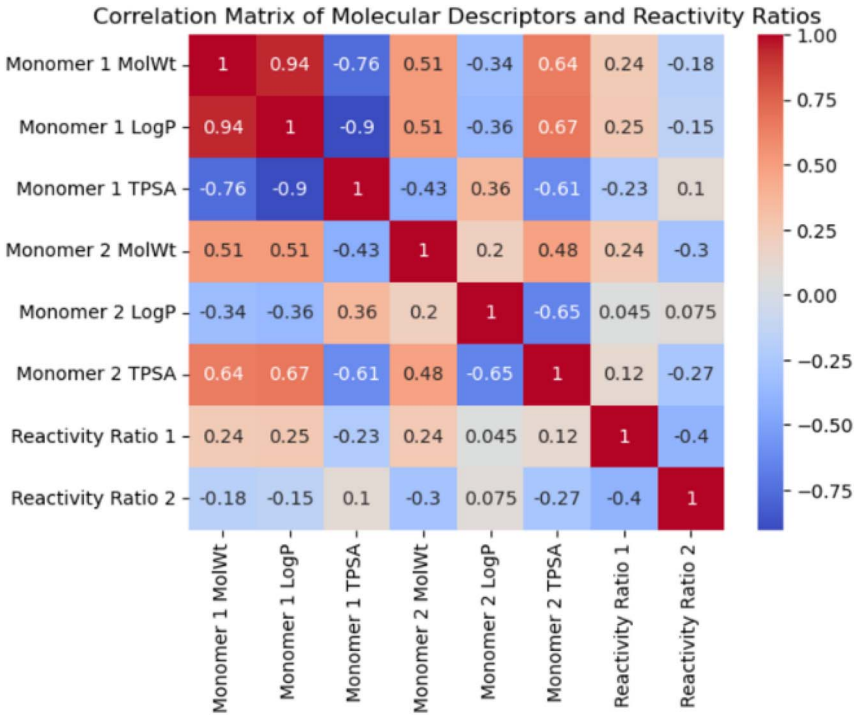

Fig. 2 ML modeling allows us to ascertain the effect of monomer descriptors on the monomer reactivity ratio, which defines the structure of the copolymer. It visualized molecular descriptors' positive and negative correlation on the reactivity ratio. Molecular Weight (MolWt) is the total mass of the molecule. Log *P* (Mol log *P*) molecular lipophilicity is measured as an octanol–water partition coefficient. Topological Polar Surface Area (TPSA) is an area of polar regions in the molecule linked to permeability and solubility.

twins. Simulated synthetic data generated by the digital twin can be used as a training dataset for supervised training targeting, such as predicting rate constants/reactivity ratios from monomer chemical descriptors and reaction conditions. Uncertainty quantification would allow analysis of the sensitivity of input reaction variables for a better understanding of reaction control options.

The structure of the monomers participating in copolymerization determines the value of the reactivity ratio, which in turn defines the structure of the copolymer (using the Mayo–Lewis reaction mechanism). We compared the extracted monomer rdkit molecular descriptors (MolWt, log *P*, and TPSA) and the monomer reactivity ratios summarized in the CopolDB.[35] The results are summarized in the form of a correlation matrix (red color indicates positive and blue color negative correlations) in Fig. 2.

It was determined that the solvent has a significant effect on kinetics, thermodynamics, and the mechanism of the copolymerization reaction while explicitly not accounted for in Mayo–Lewis's reaction mechanism. The mechanism of the solvent effect on the copolymerization involves changes in material solubility, reaction thermodynamic parameters, and variations in the propagation and termination rates (particularly in viscous solvents with high dielectric constants). H-bonding interactions with the monomer may also influence copolymerization. Fig. 3 illustrates the variation in the values of the reactivity ratios for copolymerization of methacrylic acid and styrene. In protic solvents, the carbonyl group of methacrylic acid will act as an H-bond acceptor, changing the stability of the radical, reducing electron deficiency of the double bond, and decreasing the reactivity ratio of the methacrylic acid. This effect is minimal in aprotic solvents, and the reactivity ratio of methacrylic acid shows larger values. The tunability of the rr, using the solvent effect, enabled control over polymer morphology without changing the structure of the monomer. Fig. 3b shows variation in the copolymer morphologies anticipated in the reaction of methacrylic acid and styrene carried out in different solvents.

The changes in rr of methacrylic acid are explained in terms of solvent–monomer association, radical stabilization, solvent electron, and H-donating properties, as well as solvent dielectric properties.[30–33] As solvent electron

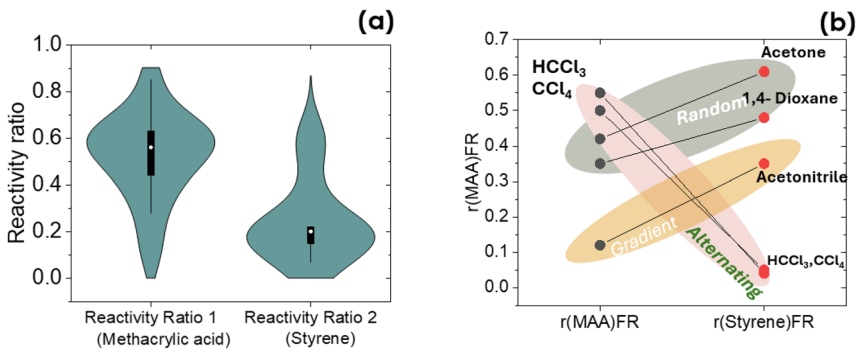

Fig. 3 (a) The distribution of methacrylic acid and styrene reactivity ratios reported for copolymerization reactions. (b) Reactivity ratio of methacrylic acid and styrene pairs for various solvents. The assignment of possible copolymer architecture (random, gradient, alternating) is based on the conventional correlation between the reactivity ratios.

donating properties increases, and methacrylic acid's reactivity ratio decreases by ~23% from 0.56 (solvent-free) to 0.47 (DMSO). In contrast, the reactivity ratio of styrene steadily increases by a factor of 2.8 (0.224 in solvent-free to 0.63 in DMSO).[30] A line between corresponding reactivity ratios of monomers provides a visual reference that while solvents affect methacrylic acid reactivity ratio, in four reactions, the reactivity ratio of styrene is affected as well and exceeds that of methacrylic acid.[33]

Using previously established rr and copolymer sequence correlation, we anticipate that changes in the solvent could vary copolymer sequence from random in acetone, 1,4-dioxane (balanced reactivity ratios) to alternating in tetrachloromethane and chloroform (MAA moderately self-reactive, and styrene avoids homopolymerization) and gradient in acetonitrile (strong reactivity imbalance leading to MAA rich backbone). Based on the correlation between polarity, hydrogen bonding of solvents, and reactivity ratio, we identified DMF (good polar aprotic solvent for stabilization of MAA through dipole–dipole interaction without the adverse effects of hydrogen bonding, while increasing the reactivity ratio of styrene, it also has good solubility for monomers) as a good candidate for the synthesis of gradient copolymer. Furthermore, the integration of green solvent, whose polarity can be tuned while avoiding distillation cost, could be considered as a medium for copolymerization.[36]

Results and discussions

We aimed to ultimately develop an autonomous continuous flow chemistry capability to translate high-quality lead molecules and materials to quantities that meet scalability demands. The first step was to design a continuous flow system that is both reconfigurable and robust to pursue different multistep chemical synthesis outcomes. The second step involved putting critical in-line characterization capabilities into this platform. The third step was to link the hardware to user key software to remotely control the setup and digitally store known optimized recipes that can be easily rerun. Finally, the fourth step is to move beyond the automated platforms to one powered by AI/ML for autonomously designing synthetic routes and carrying them out. This was accomplished as described in the methodology section, where we configured the Snapdragon continuous flow reactor system, and the LabOS augmented it with python scripting and rerouted the flow on various instruments as described. This successful protocol enabled us to put as many as 5 instruments at a single flow and use the edge computer to control the flow chemistry parameters and monitor the reactions using the data from the independent instruments in time series. The data are shown in the individual experiments.

All the reactions were done in homogeneous solutions in batch-time series aliquot or plug flow continuous flow reaction chemistry. The polymerization reactions were done as follows:

(1) For the polymerization of poly(sulfobetaine methacrylate) (PSBMA) in batch and flow, we used potassium persulfate as the initiator ($K_2S_2O_8$). The polymerization was conducted in H_2O at 70 °C.

(2) For the polymerization of poly(2-vinyl pyridine propane sulfone) (P2VPPS), the initiator was again $K_2S_2O_8$ in H_2O at 80 °C, and we run the polymerization for 24 hours in a batch-time series aliquot sampling.

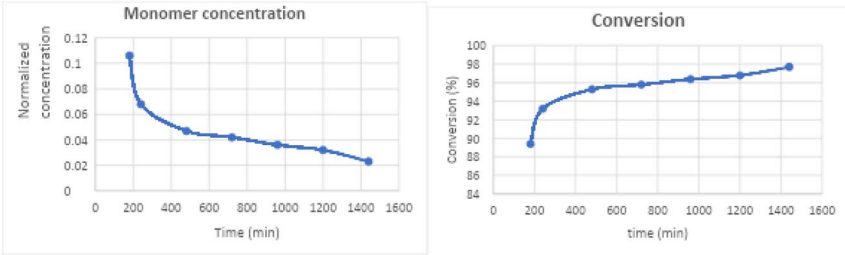

Fig. 4 Homopolymerization of P2VVPS in a continuous flow reactor set-up: flow 0.2 ml min^{-1} up to 1 ml min^{-1} using a step of 0.1 ml min^{-1} showed the correlation of monomer concentration and the conversion percentage in a quiescent experiment under ambient time and pressures.

(3) For the copolymerization of styrene and butyl methacrylate, we have used 10 mixtures of the monomers starting from St : BuMA 1 : 9 and up to 9 : 1 ratio. We used AIBN as the initiator for this polymerization and BTC as the RAFT agent. The polymerization was conducted in bulk at 60 °C in a batch-time series aliquot sampling.

(4) For the copolymerization of 2-vinylpyridine propane sulfone (2VPPS) and hydroxy ethyl methacrylate (HEMA) we used 50 : 50 molar ratio with $K_2S_2O_8$ as the initiator in H_2O at 60 °C, 70 °C and 80 °C in both batch-time series aliquot or plug flow continuous flow reaction.

(5) For the homopolymers P2VPPS and PSBMA, we collected samples using the Easysampler in batch-time series aliquot and plug flow continuous flow reactions. The experiment in the flow reactor was conducted to synthesize PSBMA, where we collected samples at 0.2 ml min^{-1} up to 1 ml min^{-1} using a step of 0.1 ml min^{-1}.

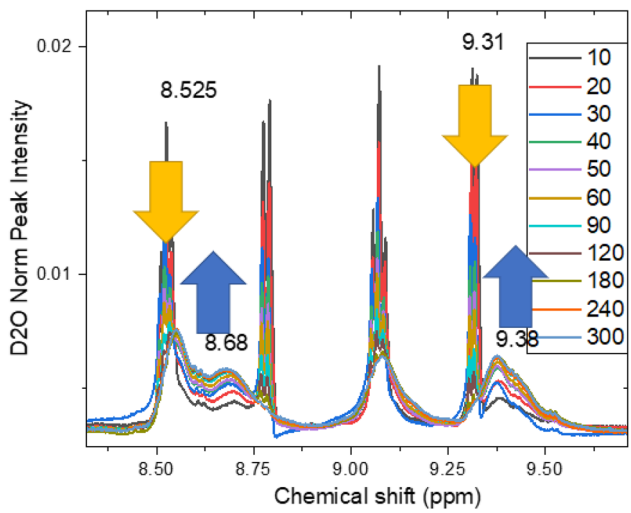

Fig. 5 NMR data correlation of copolymerization in progress in D_2O in time (min). Arrows indicate a decrease in the intensity of monomer (yellow-filled) and growth of polymer (blue-filled) NMR peaks. It is possible to obtain an Arrhenius plot of the natural logarithm of the rate constant against the inverse temperature of the copolymerization.

Representative data for the homopolymerization of P2VVPS under flow 0.2 ml min^{-1} up to 1 ml min^{-1} using a step of 0.1 ml min^{-1} showed the correlation of monomer concentration and the percentage of conversion, which confirms a plug flow environment in solution (using a Thalesnano setup) is near similar to a batch-time series in a quiescent experiment under ambient time and pressures (Fig. 4).

We considered the design of the experiment protocol for both homo- and co-polymerization variables (T, ratio of monomers, and flow rate), and this was also demonstrated in the case of HEMA and VSB. We used stochastic and chemical master equation (CME) modeling of polymerization (limited to low concentration, hard calc.). In principle, we can determine the activation energy of copolymerization (batch) and do a Kinetic Monte Carlo simulation of the probability of polymer MWD from different monomer ratios and their reactivities. An example of NMR data that distinguishes copolymerization and the corresponding change in the monomer concentration is shown in Fig. 5. It is possible to obtain an Arrhenius plot of the natural logarithm of the rate constant against the inverse temperature of the copolymerization.

Copolymerization phenomena

The polymerization process is classically divided into homopolymerization and copolymerization.[21] Beyond constitutional classification, microstructure control and pseudo-chirality are other considerations that define the method. The architecting of polymer materials requires careful planning of these factors and the appropriate mechanistic routes that lead to a topological construct.

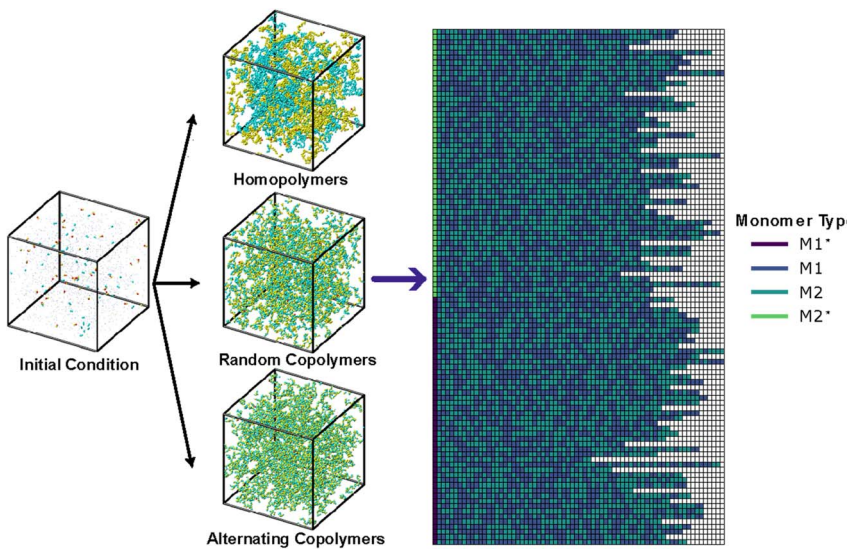

Fig. 6 Simulation of the copolymerization processes and resulting snapshots for various limiting cases. The system is initialized with a predetermined number of initiators and free monomers. Three representative limiting cases are shown: homopolymerization ($r_1 = r_2 = 1000$), random copolymerization ($r_1 = r_2 = 1$), and alternating copolymerization ($r_1 = r_2 = 0.001$). The chain sequences from the simulations are also presented for the random copolymer case.

In Fig. 6, the capability of CGMD simulations to reproduce the different limiting cases is described in Table 1. The left panel displays a representative snapshot of the initial condition, where initiators are shown as larger beads within the simulation box. We highlight three distinct cases: homopolymers ($r_1 = r_2 = 1000$), random copolymers ($r_1 = r_2 = 1$), and alternating copolymers ($r_1 = r_2 = 0.001$). A key advantage of CGMD is that it allows not only tracking of polymerization kinetics but also access to microscopic details such as chain sequences, chain length distributions, polydispersity, *etc.* As a demonstration, we present the chain sequences obtained for the random copolymer case, showcasing the ability of our approach to capture sequence-level information. Direct comparisons with experiments and setting up a digital twin using the CGMD require introducing effects of fluid flow, mapping the time scales from the simulations to the experiments, and introducing termination and transfer processes during the copolymerization.

We revisited the classical Mayo–Lewis Equation (MLE) as a predictor of the type of copolymers obtained based on relative reactivity in chain-addition reactions: random copolymers (statistical), alternating copolymers (regular pattern), block copolymers (long sequences of blocks) – each monomer type arranged linearly (diblock, triblock) or in more complex architectures. The copolymer will have several controllable properties, including solubility, glass transition temperature (T_g), crystallinity, amphiphilicity, microphase separation, and copolymer synthesis route. For example, graft copolymers allow for combining properties from the backbone and graft polymers, *e.g.*, high-impact polystyrene (HIPS) *via* polybutadiene onto the polystyrene backbone. Thus, the ultimate correlation is that chain architecture and composition influence their properties, including their chemical, mechanical, thermal, and rheological behavior. Without defining the distinct mechanistic routes, *e.g.*, living polymerization, grafting, sequential addition, macro initiator or macromonomer, *etc.*, we focused on the classical monomer reactivity ratios (rr) of the MLE to determine the ability of a growing polymer chain to add its monomer *vs.* another monomer to influence the final copolymer composition and sequence distribution. This is very useful in predicting and controlling copolymer properties. The MLE or copolymerization equation in chain-addition reactions defines the instantaneous copolymer composition as the monomer feed composition and rr. Reactivity ratio diagrams visually represent copolymer composition as a function of monomer feed composition.

Chain addition copolymerizations may be performed using heterogeneous and homogeneous methods, *e.g.*, solution, bulk, suspension, emulsion, slurry, and gas phase processes. This can also be done in batch, semi-batch, and continuous-flow reactors. It should be noted, however, that the monomer concentrations local to the active centers control the copolymerization and not the average concentrations in the reactor. For example, this can affect the multisite nature of the catalysts in heterogeneous copolymerization reactions where the propagation constants vary from site to site, affecting polydispersity. However, control of the microenvironment of local concentration can bias the MWD. Copolymerization is still a chemical process subject to unit operation with controlled temperature, pressure, solubility, and, in the case of flow, flow rate, and dosing. The goal is to maintain and eventually obtain novel material properties.

Our attempts to use simulation and time-series copolymerization showed the possibilities for control and an ML-driven experiment. We first designed a digital twin for copolymerization using MLE, as it allows us to explore an extensive range of variable conditions. Digital twin provides visualization of reactant/product concentration–time profiles, a correlation matrix, and effects of reactivity ratio first to maximize the polymer yield as a function of temperature. Note that an extensive range of reactivity ratios and monomer concentrations can be compared to LLMs. It is possible to compare this with QSAR/QSPR modeling, clustering, and similar analysis of reactivity ratio and the structure–functional properties of monomer to the total mass or MW of the polymer. We employed ML modeling to ascertain the effect of monomer descriptors on the monomer property (reactivity ratio), which ultimately defines the copolymer structure of the polymer. A correlation matrix depicts the correlation where the higher MW is less mobile for diffusion in a solution. The positive correlation between monomer MW and the molecular weight varies based on the pseudo-kinetic rate constant method. This demonstrates how the polymerization process evolves at different reactivity ratios and temperatures even after 1000 seconds of reaction. This applies not only to the systems described by the terminal copolymerization model but to the higher-order Markov chain statistics such as the penultimate model. Thus, ML modeling allows us to generate kinetic profiles for specific structures of the copolymer… random, block ladder, and gradient. The kinetics of monomer consumption is another factor. Modeling the effect of reaction variables to optimize polymer yield and polydispersity modeling can take advantage of the Markov chain statistic penultimate model.

The MLE equation considers a monomer mixture of two components and the four reactions that can occur at the reactive chain end, terminating in either monomer placement. Each possibility can be defined by its reaction rate constants. The rr for each propagating chain end is defined as the ratio of the rate constant for adding a monomer of the species already at the chain end to the rate constant for adding the other monomer, with the concentrations of the components defined. The equation then gives the two monomers' relative instantaneous rates of incorporation. The ratio of active center concentrations can be found using the steady state approximation, meaning that the concentration of each type of active center remains constant. It is essential to distinguish the rate of formation of active centers of each monomer *vs.* the rate of their destruction. The ratio of monomer consumption rates is essentially a descriptor for the MLE. Using mole fractions to express the concentrations makes it easier to define the composition of the copolymer formed at each instant. However, the feed and copolymer compositions can change as polymerization proceeds. The limiting cases establish the type of copolymer that will be obtained (random, alternating, "blocky," or just a mixture of two homopolymers). The composition drift is not easily accounted for, which is less ideal than an azeotrope system, where feed and copolymer composition are the same. Still, it is helpful to form databases of calculated reactivity ratios (plotting the copolymer equation and using nonlinear least squares analysis). This generally involves several polymerizations at varying monomer ratios, and the copolymer composition is determined by appropriate analytical methods (NMR, IR, or GC-MS). The polymerizations can also be carried out at low conversions, so monomer concentrations can be assumed constant. Refinements with Kelen–Tüdős or Fineman–Ross methods by linearization of the

MLE can bias the results closer to the model data. Another method is the *Q–e* scheme, a semi-empirical method for predicting reactivity ratios that involves introducing proportionality constants and considering the monomer reactivity based on resonance stability and polarity. At this point, there is no discussion on how the kinetics bias or thermodynamics of the mechanism can be shifted by solubility, temperature, pressure, flow (turbulent or laminar), and dosing rate can affect the MLE. It is this possibility that an ML-driven autonomous flow chemistry set-up has the potential to control the copolymerization phenomena further and augment the MLE protocol.

Can we use the rr alone to model copolymerization via ML? Literature-reported data in databases such as the Polymer Handbooks is often insufficient. There is a need to extract this information from publications and LLMs to introduce it as a descriptor. A more detailed look at the "reactivity ratio" data summarized in the databases will allow some interesting insights into reaction recipes that may have been previously overlooked. Another is the solvent or solubility-controlled sequence, and the microstructure of copolymers from random to gradient can be biased. The value of the monomer rr is not constant, as evidenced by the composition drift. We demonstrate the dependency of the reactivity ratio of two monomers on their solvent environment. We have also illustrated the effect of polarity and hydrogen bonding on the reactivity ratio. Predicting the copolymer sequence and microstructure may be possible based on these values. We have already demonstrated two of three key components, and our next effort will focus on training the LLM on synthesis protocols.

We used a retrieval augmented generation (RAG) framework with a FORGE LLM to differentiate and provide selection protocols from the prompts to demonstrate the use of LLMs.[37] The RAG-DB training was explicitly done for copolymerization. The FORGE must be integrated with a curated database. FORGE is an open foundation model for LLMs retrieval with up to 26B parameters using 257B tokens from over 200M scientific articles, with performance either on par or superior to other state-of-the-art comparable models. It was initially developed at ORNL. Using a Chat GPT 4.0 prompting on the embedded copolymerization database, we used a RAG protocol that includes reactants, products, specific *T*, *P*, and solvent and searched for references and patents. The following results yielded 5823 copolymerization items; 247 referred to butyl acrylate–styrene reactions, for example, derived from publications, patents, abstracts, *etc.*

Copolymerization experiments with a universal monitoring set-up

To augment the flow reactors, we have developed a universal spectroscopic *in situ* diagnostics protocol for feedback loop monitoring and reaction control. This is a critical part of the autonomous experiment, whether in batch, semi-batch, or continuous flow.[38] Very few online monitoring and feedback loop set-ups are available for commercial flow synthesis systems. The main goals are to generate the training data set for ML and identify the unique signature of the material of interest. *In situ* and real-time diagnostics can be designed for spectroscopic, electrical, gravimetric, viscoelastic, and optical detection. We developed a match with the trigger and the "observation window" to be synchronous with the reaction coordinate (time series) or the sequence of addition or dosing (space series). This allows us not only to get direct information on the course of the reaction but

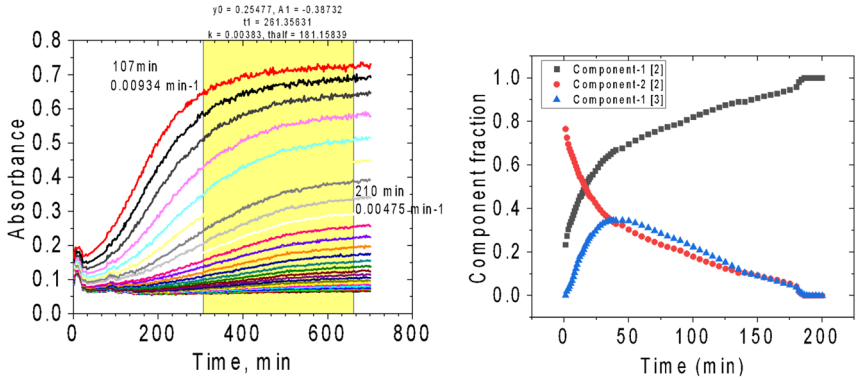

Fig. 7 Optical spectra of the copolymerization of VSB and HEMA in $K_2S_2O_8$ as initiators. The left figure shows the evolution of absorbance over time. The UV-vis absorption spectra of homopolymerizing with VSB with noticeable isosbestic points with kinetic traces taken at 10 nm intervals indicate simultaneous processes within the reaction. We differentiated component fraction traces to follow the kinetic evolution of the copolymerization process.

also to identify fundamental parameters (kinetic and thermodynamic) of the reaction based on simulation and LLMs. The diagnostics or online monitoring that is carried out becomes a multimodal regime that allows us to visualize a reaction from various viewpoints or sensitivity.

Fig. 7 is a representative of the type of data that we obtain from this set-up: Optical diagnostics follow the thermal copolymerization of VSB and HEMA in $K_2S_2O_8$ as initiators using thermally controlled cell block by measuring UV-vis absorption spectra as a function of increased concentration of monomer and time with copolymerization. The UV-vis absorption spectra of homopolymerizing with VSB with noticeable isosbestic points with kinetic traces taken at 10 nm intervals indicate simultaneous processes within the reaction. Principal component analysis (PCA) of the absorption spectra enables the choice of essential components of the kinetics experiment to follow. We differentiated a component fraction trace to follow the kinetic evolution of the copolymerization process.

In summary, the fine control and copolymerization for two types of monomers have the potential to produce a family of new functional polymers by AI/ML. Developing AI/ML-centric synthesis and characterization tools will open new scientific opportunities to generate high-quality, unique data sets and ML models that further progress in cross-disciplinary discovery and innovation.

In the future, for continuous flow reactor copolymerization, single objective optimization can be achieved with algorithms with Simplex and Stable Noisy Optimization by Branch and Fit (SNOBFIT), which seeks a global optimum of the objective function by iteratively exploring the search space and adaptively refining the sampling points. Bayesian Optimization (BO) algorithms can be used to improve future highly complex and non-linear chemical and reaction spaces for new microstructures, including branching to process even sparse datasets. Bayesian methods utilize probabilistic models to strike a balance between exploration and exploitation, making them well suited for handling noisy and

computationally expensive objective functions. The goal is to develop ML algorithms for "self-optimizing" flow reactors integrated with online reaction monitoring. Multi-objective optimizations are standard for flow chemistry and other chemical engineering processes; therefore, Pareto Efficient Global Optimization (ParEGO) methods can be instrumental in addition to genetic algorithms. Thompson Sampling Efficient Multi-Objective (TS-EMO) can also tackle multi-objective optimization in automated flow-chemistry processes. It uses Gaussian processes as surrogates. It uses Thompson sampling with the hypervolume quality indicator and Non-Dominated Sorting Genetic Algorithm II (NSGA-II) to decide the next evaluation point after each iteration.

Flow chemistry approaches have been applied to many monomers and polymerization mechanisms for polymer synthesis, including anionic, cationic, radical, and ring-opening.[7,39,40] Compared to traditional batch processing, chemical synthesis performed in continuous flow gives better reproducibility and minor time scales due to efficient mixing, heat- and mass transfer, and variable flow rates that can accelerate reaction rates. Other desirable attributes include real-time reaction monitoring, storage of intermediates, increased product quality, and enhanced safety due to flowing material within channels from stock containers to reactors. Some of the present drawbacks of the standard continuous flow approach, such as the rate at which reactions occur, can be optimized. Finally, this approach is well suited for automation. It can potentially enable intelligent reaction systems to operate safely, efficiently, and for production 24 hours per day.

Conclusions

We have demonstrated ML modeling and rationalization of the copolymerization process in batch time series and continuous flow synthesis. The aim is a more autonomous discovery and innovative ecosystem for materials synthesis. It will be a data-driven approach that seeks to go beyond today's conventional "make, measure, model" linear R&D discovery to one that uses *in situ* interconnected instruments and measurements coupled with data science approaches to accelerate the materials discovery for polymers. This opens many chemical reactions and mechanistic studies, directly comparing responses under different conditions. Noted advantages include: (1) The possibility of discovering new kinetic and thermodynamic biased reaction pathways. (2) The production of new and known target polymers that can be made considerably faster and more efficiently, with optimal synthesis protocols stored as digital files. (3) Increased safety with highly sensitive reactions that can be affected by uncontrolled heat and mass transfer in batches. (4) Modularization of reactions and separations, in principle, can be deployed or transported to deploy synthesis in a manufacturing ecosystem rapidly. In the current work, we have demonstrated its advantage in copolymerization. LLMs on "continuous flow synthesis" can be used to identify key publications and training data with high *H*-index specifications. There are a lot of opportunities for the proposed autonomous flow chemistry setup for polymer science with a high utility of a 'real-time' data-driven approach. *In situ* and real-time interconnected instruments and measurements will accelerate the pace of precision polymer synthesis and materials discovery.

Author contributions

Rigoberto Advincula: led the project and authored the manuscript; Ilia Ivanov: project design, writing the manuscript and performed experiments; Rama Vasudevan: project design, writing the manuscript and performed the experiments; Rajeev Kumar: writing the manuscript and performed the simulation experiments; Panagiotis Christakopoulos: writing the manuscript and performed the experiments; Marileta Tsakanika: performed the experiments and contributed to the data; Jihua Chen: writing the manuscript and performed experiments; Jan Michael Carillo: writing of the manuscript and performed simulations; Qinyu Zhu: writing of the manuscript and performed simulations; Bobby Sumpter: design of the project and the manuscript.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All the materials and data available with correspondence to the authors and are also part of referenced publications when indicated. The availability of data and access is governed by fair use. We will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (**http://energy.gov/downloads/doe-public-access-plan**) as available.

Acknowledgements

This work was performed at the Center for Nanophase Materials Sciences (CNMS), a US Department of Energy Office of Science User Facility at Oak Ridge National Laboratory (ORNL). The INTERSECT Initiative supported the development of the autonomous flow reactor system as part of the Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725. We also acknowledge support from the Governor's Chair Professor funds, University of Tennessee for R. C. Advincula.

References

1 J. Chen, Y. Yuan, A. K. Ziabari, X. Xu, H. Zhang, P. Christakopoulos, P. V. Bonnesen, I. N. Ivanov, P. Ganesh, C. Wang, K. P. Jaimes, G. Yang, R. Kumar, B. G. Sumpter and R. Advincula, AI for Manufacturing and Healthcare: A Chemistry and Engineering Perspective, *arXiv*, 2024, preprint, arXiv:2405.01520, DOI: **10.48550/arXiv.2405.01520**.

2 T.-S. Lin, C. W. Coley, H. Mochigase, H. K. Beech, W. Wang, Z. Wang, E. Woods, S. L. Craig, J. A. Johnson, J. A. Kalow, K. F. Jensen and B. D. Olsen, BigSMILES: A Structurally-Based Line Notation for Describing

Macromolecules, *ACS Cent. Sci.*, 2019, **5**(9), 1523–1531, DOI: **10.1021/acscentsci.9b00476**.

3 W. Choi, R. C. Advincula, H. F. Wu and Y. Jiang, Artificial Intelligence and Machine Learning in the Design and Additive Manufacturing of Responsive Composites, *MRS Commun.*, 2023, **13**(5), 714–724, DOI: **10.1557/s43579-023-00473-9**.

4 S. Ferdousi, R. Advincula, A. P. Sokolov, W. Choi and Y. Jiang, Investigation of 3D Printed Lightweight Hybrid Composites *via* Theoretical Modeling and Machine Learning, *Composites, Part B*, 2023, **265**, 110958, DOI: **10.1016/j.compositesb.2023.110958**.

5 Y. Jiang, Md. N. Islam, R. He, X. Huang, P.-F. Cao, R. C. Advincula, N. Dahotre, P. Dong, H. F. Wu and W. Choi, Recent Advances in 3D Printed Sensors: Materials, Design, and Manufacturing, *Adv. Mater. Technol.*, 2023, **8**(2), 2200492, DOI: **10.1002/admt.202200492**.

6 Creating the lab of the future, ORNL, **https://www.ornl.gov/labsofthefuture**, accessed 2025-05-02.

7 L.-H. Rong, E. B. Caldona and R. C. Advincula, PET-RAFT Polymerization under Flow Chemistry and Surface-Initiated Reactions, *Polym. Int.*, 2023, **72**(2), 145–157, DOI: **10.1002/pi.6475**.

8 J. Ge, Y. Yin, E. B. Caldona and R. C. Advincula, Hyperbranched PDMAEMA-Functionalized SiO_2 Microparticles: ATRP Polymerization and Grafting in a Continuous Flow Reactor, *MRS Commun.*, 2022, **12**(6), 1147–1153, DOI: **10.1557/s43579-022-00263-9**.

9 E. S. Muckley, R. Vasudevan, B. G. Sumpter, R. C. Advincula and I. N. Ivanov, Machine Intelligence-Centered System for Automated Characterization of Functional Materials and Interfaces, *ACS Appl. Mater. Interfaces*, 2023, **15**(1), 2329–2340, DOI: **10.1021/acsami.2c16088**.

10 B. G. Sumpter, K. Hong, R. K. Vasudevan, I. Ivanov and R. Advincula, Autonomous Continuous Flow Reactor Synthesis for Scalable Atom-Precision, *Carbon Trends*, 2023, **10**, 100234, DOI: **10.1016/j.cartre.2022.100234**.

11 Deed – Attribution-NonCommercial-NoDerivatives 4.0 International, Creative Commons, **https://creativecommons.org/licenses/by-nc-nd/4.0/**, accessed 2025-03-19.

12 I. Zapata-González, E. Saldívar-Guerra and R. A. Hutchinson, 80 Years of the Mayo Lewis Equation. A Comprehensive Review on the Numerical Estimation Techniques for the Reactivity Ratios in Typical and Emerging Copolymerizations, *Prog. Polym. Sci.*, 2025, **163**, 101956, DOI: **10.1016/j.progpolymsci.2025.101956**.

13 R. Ferreira Da Silva, R. Moore, B. Mintz, R. Advincula, A. Al Najjar, L. Baldwin, C. Bridges, R. Coffee, E. Deelman, C. Engelmann, B. Etz, M. Firestone, I. Foster, P. Ganesh, L. Hamilton, D. Huber, I. Ivanov, S. Jha, Y. Li, Y. Liu, J. Lofstead, A. Mandal, H. Martin, T. Mayer, M. McDonnell, V. Murugesan, S. Nimer, N. Rao, M. Seifrid, M. Taheri, M. Taufer and K. Vogiatzis, *Shaping the Future of Self-Driving Autonomous Laboratories Workshop*, ORNL/TM-2024/3714, 2481197, 2024, DOI: **10.2172/2481197**.

14 Dash Documentation & User Guide, Plotly, **http://dash.plotly.com/**, accessed 2025-05-02.

15 Pycroscopy/Sidpy, 2025, **https://github.com/pycroscopy/sidpy**, accessed 2025-05-02.

16 A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott and S. J. Plimpton, LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales, *Comput. Phys. Commun.*, 2022, **271**, 108171, DOI: **10.1016/j.cpc.2021.108171**.

17 J. R. Gissinger, B. D. Jensen and K. E. Wise, Modeling Chemical Reactions in Classical Molecular Dynamics Simulations, *Polymer*, 2017, **128**, 211–217, DOI: **10.1016/j.polymer.2017.09.038**.

18 J. R. Gissinger, B. D. Jensen and K. E. Wise, REACTER: A Heuristic Method for Reactive Molecular Dynamics, *Macromolecules*, 2020, **53**(22), 9953–9961, DOI: **10.1021/acs.macromol.0c02012**.

19 J. R. Gissinger, B. D. Jensen and K. E. Wise, Molecular Modeling of Reactive Systems with REACTER, *Comput. Phys. Commun.*, 2024, **304**, 109287, DOI: **10.1016/j.cpc.2024.109287**.

20 I. Zapata-González, R. A. Hutchinson, K. Matyjaszewski, E. Saldívar-Guerra and J. Ortiz-Cisneros, Copolymer Composition Deviations from Mayo–Lewis Conventional Free Radical Behavior in Nitroxide Mediated Copolymerization, *Macromol. Theory Simul.*, 2014, **23**(4), 245–265, DOI: **10.1002/mats.201300137**.

21 F. R. Mayo and C. Walling, Copolymerization, *Chem. Rev.*, 1950, **46**(2), 191–287, DOI: **10.1021/cr60144a001**.

22 F. M. Lewis, F. R. Mayo and W. F. Hulse, Copolymerization. II. The Copolymerization of Acrylonitrile, Methyl Methacrylate, Styrene and Vinylidene Chloride, *J. Am. Chem. Soc.*, 1945, **67**(10), 1701–1705, DOI: **10.1021/ja01226a025**.

23 F. T. Wall, The Structure of Copolymers. II[1], *J. Am. Chem. Soc.*, 1944, **66**(12), 2050–2057, DOI: **10.1021/ja01240a014**.

24 T. Alfrey, Jr. and G. Goldfinger, The Mechanism of Copolymerization, *J. Chem. Phys.*, 1944, **12**(6), 205–209, DOI: **10.1063/1.1723934**.

25 F. T. Wall, The Structure of Vinyl Copolymers, *J. Am. Chem. Soc.*, 1941, **63**(7), 1862–1866, DOI: **10.1021/ja01852a016**.

26 T. Alfrey Jr. and C. C. Price, Relative Reactivities in Vinyl Copolymerization, *J. Polym. Sci.*, 1947, **2**(1), 101–106, DOI: **10.1002/pol.1947.120020112**.

27 M. N. Nguyen, C. Bressy and A. Margaillan, Controlled Radical Polymerization of a Trialkylsilyl Methacrylate by Reversible Addition–Fragmentation Chain Transfer Polymerization, *J. Polym. Sci. Part Polym. Chem.*, 2005, **43**(22), 5680–5689, DOI: **10.1002/pola.21063**.

28 C. Zhang, Y. Zhou, Q. Liu, S. Li, S. Perrier and Y. Zhao, Facile Synthesis of Hyperbranched and Star-Shaped Polymers by RAFT Polymerization Based on a Polymerizable Trithiocarbonate, *Macromolecules*, 2011, **44**(7), 2034–2049, DOI: **10.1021/ma1024736**.

29 M. Zhang, H. Liu, W. Shao, K. Miao and Y. Zhao, Synthesis and Properties of Multicleavable Amphiphilic Dendritic Comblike and Toothbrushlike Copolymers Comprising Alternating PEG and PCL Grafts, *Macromolecules*, 2013, **46**(4), 1325–1336, DOI: **10.1021/ma3025283**.

30 A. F. Nikolayev and V. M. Gal'perin, Copolymerization of Styrene with Methacrylic Acid in the Presence of Lewis Bases, *Polym. Sci. USSR*, 1967, **9**(11), 2793–2796, DOI: **10.1016/0032-3950(67)90383-8**.

31 Yu. D. Semchikov, L. A. Smirnov, T. Y. Knyazeva, S. A. Bulgakova, G. A. Voskoboinik and V. I. Sherstyanykh, General Nature of the Effect of Molecular Weight on the Composition of a Copolymer on Homogeneous Radical Copolymerization, *Polym. Sci. USSR*, 1984, **26**(4), 780–788, DOI: **10.1016/0032-3950(84)90244-2**.

32 S. I. Gusev, S. D. Zaitsev and Yu. D. Semchikov, Effect of Cobalt(III) 1-Nitroso-2-Naphtholate on Free-Radical Polymerization to Vinyl Monomers, *Polym. Sci. Ser. B*, 2008, **50**(3), 78–82, DOI: **10.1134/S1560090408030093**.

33 H. Boudevska, S. Platchkova and O. Todorova, The Influence of Molecular Interaction on Polymerization, 7. Copolymerization of Monomethyl Itaconate with Styrene, *Makromol. Chem.*, 1982, **183**(10), 2583–2591, DOI: **10.1002/macp.1982.021831028**.

34 K. Farajzadehahary, X. Telleria-Allika, J. M. Asua and N. Ballard, An Artificial Neural Network to Predict Reactivity Ratios in Radical Copolymerization, *Polym. Chem.*, 2023, **14**(23), 2779–2787, DOI: **10.1039/D3PY00246B**.

35 K. Takahashi, H. Mamitsuka, M. Tosaka, N. Zhu and S. Yamago, CoPolDB: A Copolymerization Database for Radical Polymerization, *Polym. Chem.*, 2024, **15**(10), 965–971, DOI: **10.1039/D3PY01372C**.

36 S. Abou-Shehada, J. H. Clark, G. Paggiola and J. Sherwood, Tunable Solvents: Shades of Green, *Chem. Eng. Process. Process Intensif*, 2016, **99**, 88–96, DOI: **10.1016/j.cep.2015.07.005**.

37 J. Yin, S. Dash, F. Wang and M. Shankar, FORGE: Pre-Training Open Foundation Models for Science, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis; SC '23*, Association for Computing Machinery, New York, NY, USA, 2023, pp. 1–13, DOI: **10.1145/3581784.3613215**.

38 L.-H. Rong, X. Cheng, J. Ge, E. B. Caldona and R. C. Advincula, Synthesis of Hyperbranched Polymers *via* PET-RAFT Self-Condensing Vinyl Polymerization in a Flow Reactor, *Macromol. Chem. Phys.*, 2022, **223**(1), 2100342, DOI: **10.1002/macp.202100342**.

39 P. Ye, P.-F. Cao, Z. Su and R. Advincula, Highly Efficient Reversible Addition–Fragmentation Chain-Transfer Polymerization in Ethanol/Water *via* Flow Chemistry, *Polym. Int.*, 2017, **66**(9), 1252–1258, DOI: **10.1002/pi.5374**.

40 P. Ye, P.-F. Cao, Q. Chen and R. Advincula, Continuous Flow Fabrication of Block Copolymer–Grafted Silica Micro-Particles in Environmentally Friendly Water/Ethanol Media, *Macromol. Mater. Eng.*, 2019, **304**(2), 1800451, DOI: **10.1002/mame.201800451**.

