Multi-objective Bayesian optimization for experimental design in co-polymerization and revealing chemical mechanism of Pareto fronts

Hiromu Yamada,[a] Shogo Takasuka,[a] Shunto Oikawa,[a] Yosuke Harashima,[a,b] Tomoaki Takayama,[a,b] Aniruddha Nag,[a] Araki Wakiuchi,[c] Tsuyoshi Ando,[a] Tetsunori Sugawara,[d] Miho Hatanaka,[e] Tomoyuki Miyao,[a,b,f] Takamitsu Matsubara,[a] Yu-ya Ohnishi,[c] Hiroharu Ajiro [a,b,f] and Mikiya Fujii *[a,b,f]

a. Graduate School of Science and Technology, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara 630-0192, Japan.

b. Data Science Center, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara 630-0192, Japan.

c. Materials Informatics Initiative, RD technology and digital transformation center, JSR Corporation, 3-103-9 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa 210-0821, Japan.

d. Fine Chemical Process Dept., JSR Corporation, 100 Kawajiri-cho, Yokkaichi, Mie 510-8552, Japan.

e. Department of Chemistry, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan.

f. Center for Material Research Platform, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara 630-0192, Japan.

KEYWORDS Polymer, Flow synthesis, Radical polymerization, Bayesian optimization, Multi objective Bayesian optimization, Styrene, Methyl methacrylate

ABSTRACT: The chemical properties of copolymers are strongly influenced by a number of intrinsic characteristics such as their molecular weight and their monomer composition ratio. Identifying the optimal conditions for the copolymerization process that results in a synthesized copolymer with the desired characteristics is a major challenge. Optimization of the copolymerization process has traditionally been based on trial-and-error approaches by humans relying on empirical rules. Thus, the design space that can be explored experimentally is severely limited under time and economic constraints. In addition, solving problems such as nonuniformity of both temperature and the concentration of chemicals in the reaction field is also challenging. In this study, we established multi-objective Bayesian optimization and flow copolymerization systems to explore optimal copolymerization conditions for synthesizing copolymers that simultaneously exhibit multiple target characteristics. Finally, we visualized Pareto fronts representing trade-offs between polymer characteristics and employed quantum chemical calculations to reveal chemical origins of the Pareto fronts.

1. INTRODUCTION

Polymers, including copolymers, show distinctive chemical properties that cannot be achieved with low-molecular-weight compounds. They are used as important materials across various industrial fields, including the automotive, medical, and food industries. These chemical properties emerge from aggregation states of polymer chains, which are determined by molecular conformations that depend on the characteristics of the polymers, such as their molecular weight and their monomer composition ratio.[1-7] To achieve desired chemical properties, identifying the polymerization conditions that yield the appropriate polymer characteristics is critical, whereas relationships among polymer characteristics are complex as the characteristics are determined by multiple experimental conditions. This complexity makes it difficult to optimize the characteristics simultaneously.

The development of materials has traditionally been based on trial-and-error approaches by humans relying on empirical rules at each stage from molecular design to process optimization. Thus, this approach limits the number of materials that can be investigated because of time and economic constraints. This is an important issue in increasing development efficiency. Materials scientists have recently focused on efficient development methods using machine learning with data obtained from both theoretical calculations and experiments.[8-12] In polymer chemistry, polymer databases such as PoLyInfo,[13] Polymer Genome,[14] Khazana,[15] CoPolDB,[16] and CopDDB[17] have been established; machine learning using these databases and simulations has advanced the featurization and property prediction of polymers.[18-24] Recently, Schneider et al. proposed extending BigSMILES[25] to Generative BigSMILES, which is expected to serve as a novel tool supporting the data-driven development of materials by enabling the systemat-

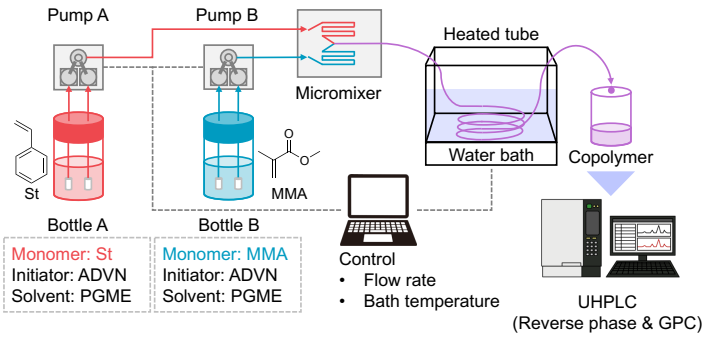

Figure 1 Diagram of the flow copolymerization system. Each of the two bottles contains one monomer, initiator, and solvent. These solutions in the two bottles were mixed using a micromixer, and a system was constructed to synthesize copolymers under specific conditions.

ic description and generation of polymer ensemble characteristics.[26] However, the amount of available data is often limited when optimization of experimental conditions, such as the optimization of polymerization processes, is necessary. Bayesian optimization is a method for effectively using limited datasets.[27-32] It is an experimental design method that enables efficient exploration by updating models while collecting data incrementally using acquisition functions.[33] For example, Wang et al. successfully optimized the copolymerization process of high-density polyethylene using Bayesian optimization, which resulted in increased productivity and profitability compared with the traditional polymerization process.[34] Bayesian optimization, which suggests experimental conditions instead of human discretion, plays a pivotal role in the automation[35-37] of development processes.

Bayesian optimization requires the acquisition of highly reproducible data to construct surrogate functions based on the data. Flow synthesis, in which reactions are conducted continuously in tubes, maintains high uniformity of reaction fields through efficient heat exchange and enables control of the residence time; by contrast, batch synthesis involves non-uniform reaction fields.[38-39] Therefore, flow synthesis is an effective method for obtaining highly reproducible experimental data. Coley et al. developed a platform integrating retrosynthesis prediction based on machine learning with robot-controlled flow synthesis, which resulted in the successful automated synthesis of 15 drug-candidate compounds.[40] Such integration of machine learning and flow synthesis is gaining attention as a new approach to efficiently optimize synthesis conditions and explore novel compounds for not only small molecules but also polymers.[41-45]

We have been establishing polymer synthesis processes that integrate flow synthesis and machine learning. In a previous study, we developed a flow copolymerization system incorporating micromixers.[46] Using this flow copolymerization system, we demonstrated that electronic states of transition states are critical factors in predicting reactivity for copolymerization with monomers not included in training data (i.e., prediction of the extrapolated region[47]) and established in situ prediction of residual monomer concentrations in binary copolymerization reactions from Fourier transform infrared spectroscopy (FTIR) spectra.[48] In a recent study, we conducted process optimization focused on monomer proportion in a polymer by combining Bayesian optimization with the flow copolymerization system. The previous study demonstrated the effectiveness of integrating flow synthesis and machine learning approaches.[27] However, even when the desired monomer proportion in a polymer was achieved, other polymer characteristics exhibited a range of values. Since these characteristics directly influence the chemical properties of the resultant polymers, simultaneous optimization of multiple polymer characteristics is necessary.

The aim of the present study is to establish a methodology for simultaneously optimizing multiple polymer characteristics by integrating multi-objective Bayesian optimization with a flow copolymerization system and to reveal the chemical factors contributing to Pareto fronts that exists between polymer characteristics. The establishment of the present method is expected to lead to efficient optimization of copolymerization processes, which should, in turn, lead to the creation of new guidelines of material design through the integration of machine learning and domain expertise.

2. MATERIALS AND METHODS

2.1 Materials. Styrene (St, FUJIFILM Wako, special grade) and methyl methacrylate (MMA, FUJIFILM Wako, special grade) were used as monomers. Both monomers were used as received without removal of stabilizers. Propylene glycol monomethyl ether (PGME) was used as the solvent, and 2,2'-azobis(2,4-dimethylvaleronitrile) (ADVN, FUJIFILM Wako, 1st grade) was used as the thermal initiator.

2.2 Characterization. Ultra-high-performance liquid chromatography (UHPLC) was carried out using a Shimadzu Nexera XS system (DGU-403, LC-40D XS, SIL-40C XS, CBM-40, RID-20A, SPD-M40, CTO-40C) equipped with a separation column (GL Science, ODS-3 2 μm, 3.0 × 50 mm^2). Gel permeation chromatography (GPC) was performed using a JASCO system (LC-NetII/ADC, AS-2055, PU-2080, CO-2065, UV-2075, RI-2031) equipped with columns (Tosoh, TSKGel SuperH3000, GMHR-L).

2.3 Flow synthesis system. A copolymer consisting of St and MMA was synthesized using the flow copolymerization system shown in Figure 1. This system demonstrates high experimental reproducibility because of the uniform mixing achieved by a micromixer and homogeneous heating with a tube.[46] Accordingly, the data obtained from this system are suitable for use as training data in machine learning applications. In preparation for the experiment, Bottle A contained a mixture of St, initiator, and solvent and Bottle B contained a mixture of MMA, initiator, and solvent. During the reaction, the bottles were cooled in ice water to prevent unintended polymerization.

In the present study, we classified five copolymerization conditions into two categories: soft copolymerization conditions and hard copolymerization conditions. Residence time and St proportion are defined as soft copolymerization conditions because these two copolymerization conditions can be easily changed by adjusting pump flow rates. By contrast, the Solvent/Monomer (SM) ratio, Initiator concentration, and Reaction temperature are defined as hard copolymerization conditions. These copolymerization conditions are difficult to adjust in real time. The SM and initiator concentration are determined by composition of solutions prepared in the bottles before the reaction. Therefore, once the bottles are set, these values cannot be altered by adjusting the pump flow rates. The Reaction temperature requires a certain time to reach and stabilize at new setpoints. Consequently, rapid temperature adjustments are difficult to implement during the reaction.

The St proportion in a polymer, $\boldsymbol{CR_{St}}$, and the Sum conversion, $\boldsymbol{C_{sum}}$, were determined from the changes in the monomer concentrations before and after copolymerization, as evaluated using reverse-phase UHPLC. $\boldsymbol{CR_{St}}$ and $\boldsymbol{C_{sum}}$ were calculated using the following equations:

$$CR_{St} = \frac{(FR_{St} \times C_{St})}{(FR_{St} \times C_{St}) + (FR_{MMA} \times C_{MMA})} \times 100, \quad (1)$$

$$C_{sum} = C_{St} + C_{MMA}, \quad (2)$$

$$C_{St} = \left(1 - \frac{R_t}{R_0}\right) \times 100, \quad (3)$$

$$C_{MMA} = \left(1 - \frac{R_t}{R_0}\right) \times 100, \quad (4)$$

where $\boldsymbol{R_0}$ is the mass fraction of each monomer before the reaction and $\boldsymbol{R_t}$ is the mass fraction of each monomer at residence time *t*. $\boldsymbol{FR_{St}}$ and $\boldsymbol{FR_{MMA}}$ are the proportions of each monomer used in the raw materials mixture in the synthesis procedure, and $\boldsymbol{C_{St}}$ and $\boldsymbol{C_{MMA}}$ are the conversion rates of St and MMA, respectively. Sum conversion represents the sum of the St and MMA conversion rates, with values ranging from 0 to 200. The weight-average molecular weight, M_w, and number-average molecular weight, M_n, were obtained from GPC results.

Table 1 Design space of Latin hypercube sampling (LHS) and the multi-objective Bayesian optimization (MOBO) algorithm

Copolymerization conditions	Unit	Range
Initiator concentration	mol%	1–10
Solvent/Monomer (SM) ratio		1–10
St proportion	mol%	0–100
Reaction temperature	°C	50–90
Residence time	min	2–30

Table 2 Two target values considered to represent the difference in the degree of copolymerization progress

	St proportion in polymer [%]	Sum conversion [%]	M_w	M_n
Target 1	50	100	8000	4000
Target 2	50	20	5000	2500

2.4 Multi-Objective Bayesian Optimization (MOBO) Algorithm. In this study, we used a multi-objective Bayesian optimization (MOBO) algorithm to optimize the copolymerization conditions for St–MMA copolymers. This algorithm is used to efficiently explore combinations of explanatory variables that simultaneously minimize (or maximize) all objective functions. Five copolymerization conditions (Initiator concentration, SM ratio, St proportion, Reaction temperature, and Residence time) were selected as explanatory variables, and four polymer characteristics (St proportion in polymer, Sum conversion, M_w, and M_n) were targeted for optimization. Although Sum conversion does not directly affect polymer chemical properties, it was included as an optimization target because of its importance for preventing blockage of the flow tube and for the industrial productivity. The design space of the MOBO algorithm is shown in Table 1.

On the basis of our previous study,[27] when St proportion in polymer is approximately 50 mol%, the range of Sum conversion is from 10 to 100% and the range of M_w is from 1000 to 5000. An M_w exceeding 5000 was expected to be challenging to achieve.[27] Taking these insights into account, we set two different target values considering the degree of copolymerization progress shown in Table 2. This study identified optimal copolymerization conditions for simultaneously achieving these target values in copolymer synthesis.

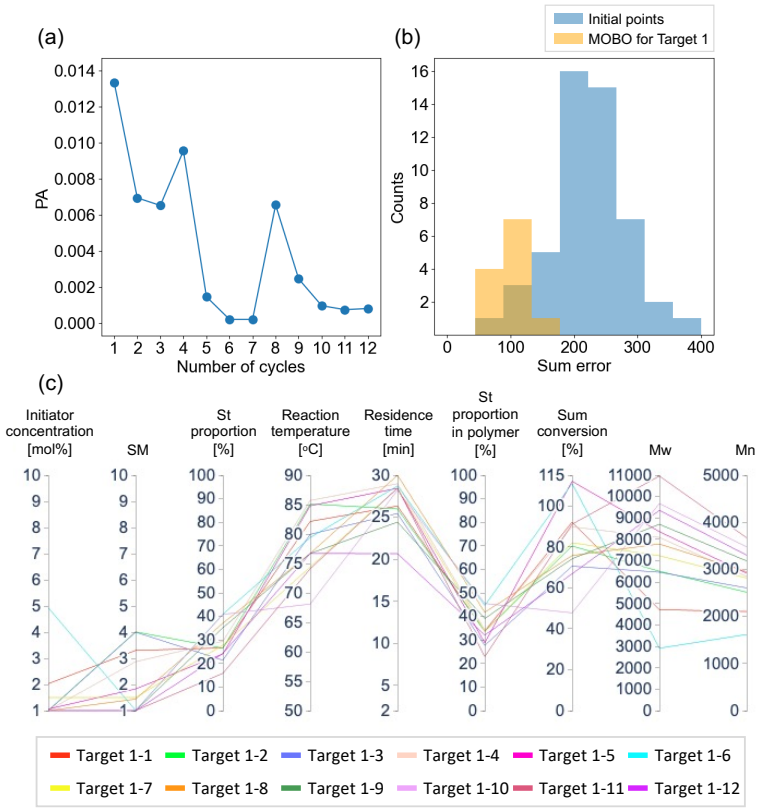

Figure 2 Multi-objective Bayesian optimization (MOBO) results for Target 1: (a) Probability of achievement (PA) of candidate points at the n-th cycle calculated by Gaussian process regression (GPR) trained on the initial points and all points up to the (n – 1)-th cycle; (b) distribution comparison of Sum error between experimental points selected by MOBO and Latin hypercube sampling (LHS); and (c) the parallel coordinates plot showing the proposed copolymerization conditions and resultant polymer characteristics for each optimization iteration. Each experimental point is shown in a different color.

This study adopted an approach that minimizes deviations between target and measured values. To treat both positive and negative deviations equally, the objective function was defined as

$$y = (T - A)^2, \tag{5}$$

where T represents the target value and A represents the actual value. Four Gaussian process regression (GPR) models for the four objectives were updated with new data obtained from evaluated experiments at each iteration. The Matérn 5/2 kernel function was used for the regression. Probability of achievement (PA) was adopted as the acquisition function and the PA was calculated as joint probability of the probability of improvement (PI) for each objective variable[32] as follows:

$$PI_i(\boldsymbol{x}) = \int_{-\infty}^{y_i^*} \frac{1}{\sqrt{2\pi\sigma_i^2(\boldsymbol{x})}} exp\left(-\frac{\left(y_i - \mu_i(\boldsymbol{x})\right)^2}{2\sigma_i^2(\boldsymbol{x})}\right) dy_i, \tag{6}$$

$$PA(\boldsymbol{x}) = \prod_{i=1}^{4} PI_i(\boldsymbol{x}), \tag{7}$$

where y^* represents the optimal point among the existing data, and i takes values ranging from 1 to 4, corresponding to the four polymer characteristics (St proportion in polymer, Sum conversion, M_w, and M_n); $\boldsymbol{x}$ is the five-dimensional vector representing the five copolymerization conditions; $\mu(\boldsymbol{x})$ is the predictive mean for $\boldsymbol{x}$ by GPR; and $\sigma(\boldsymbol{x})$ is the predictive standard deviation for $\boldsymbol{x}$ by GPR. The candidate of copolymerization conditions, $\boldsymbol{x}_{new}$, for the subsequent experiment were identified as variables at which the probability of achievement was maximized as follows:

$$\boldsymbol{x}_{new} = \underset{\boldsymbol{x}}{\operatorname{argmax}}\, PA(\boldsymbol{x}). \tag{8}$$

To calculate the first candidate points in MOBO, we used the same initial points employed in our previous single-objective Bayesian optimization study.[27] To make this paper self-contained, the preparation of the initial points is briefly described. The Latin hypercube sampling (LHS) method was employed hierarchically, and 50 points were

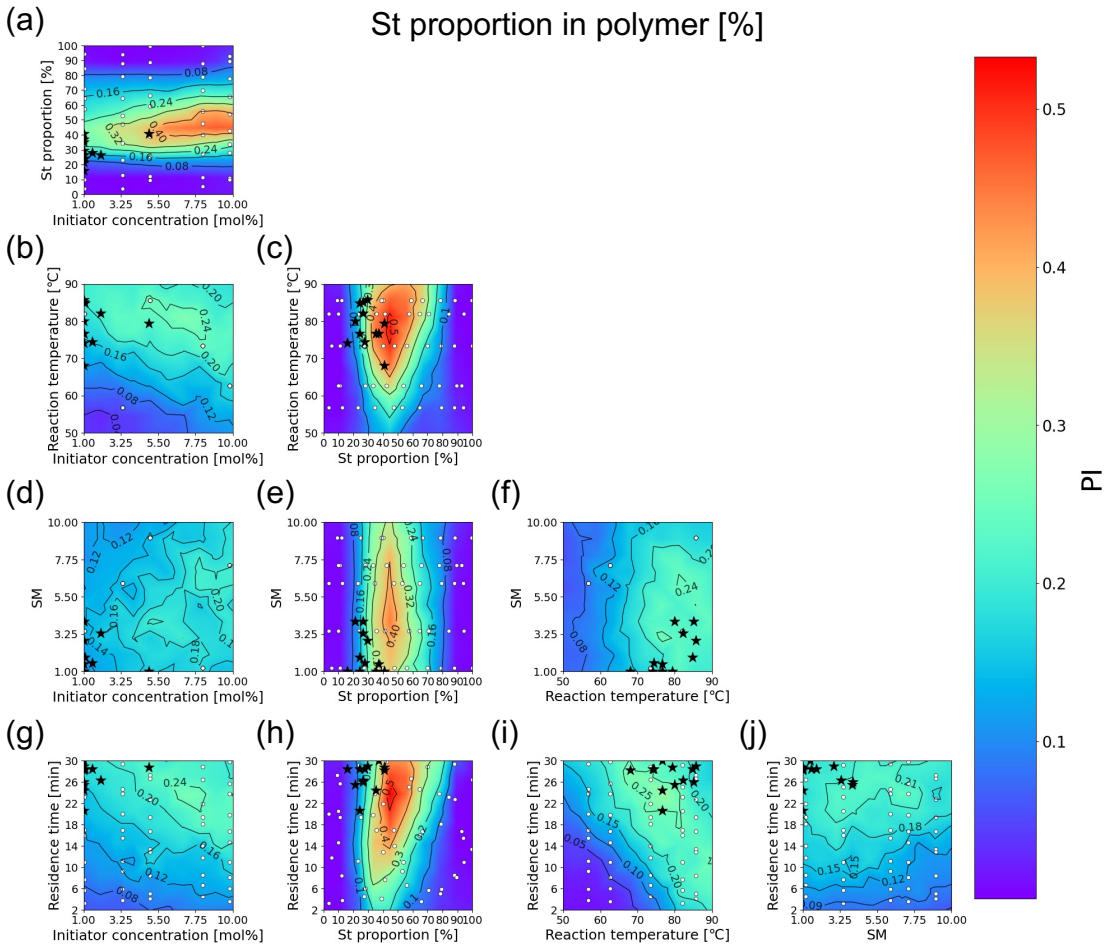

Figure 3 Visualization of the partial dependence of the PI of St proportion in polymer on the copolymerization conditions. The color-coded background is calculated from the PI values obtained using GPR trained on data from the initial points to 12 cycles for Target 1. Red indicates high PI values, and blue indicates low PI values. Stars represent the experimental points proposed by MOBO, and white circles indicate initial points.

selected on the basis of the maximin criterion. The exploration ranges in the design space for LHS are shown in Table 1. We initially conducted LHS in the hard copolymerization conditions space to enhance efficiency and obtained five points. We subsequently performed additional LHS in the soft copolymerization conditions space on the basis of these five points and obtained 10 points for each hard copolymerization condition, resulting in 50 initial points. The MOBO implementation was primarily conducted using the BoTorch library in a Python 3.8 environment (Python = 3.8 series, BoTorch = 0.6.2, GPyTorch = 1.6.0, SciPy = 1.10.1, scikit-learn = 1.2.2).

3. RESULTS AND DISCUSSION

3.1 MOBO for Target 1. Initially, MOBO was conducted for Target 1. Figure 2a shows the PA of candidate points at the n-th cycle by GPR trained on the initial and all points up to the (n − 1)-th cycle. As the number of cycles increased, the PA converged to 0. Therefore, we concluded that further improvements would unlikely be achieved through continued exploration. Figure 2b shows the distribution of sum of relative errors of the experimental points obtained through LHS and MOBO (hereafter referred to as "Sum error"). Sum error was defined as

$$\text{Sum error} = \sum_{i=1}^{4} \left| \frac{(T_i - A_i)}{T_i} \right|, \tag{9}$$

where i takes values ranging from 1 to 4, corresponding to the four polymer characteristics (St proportion in polymer, Sum conversion, M_w, and M_n). A comparison of both distributions, by LHS and MOBO, in Figure 2b reveals that the experiments conducted through MOBO led to more results

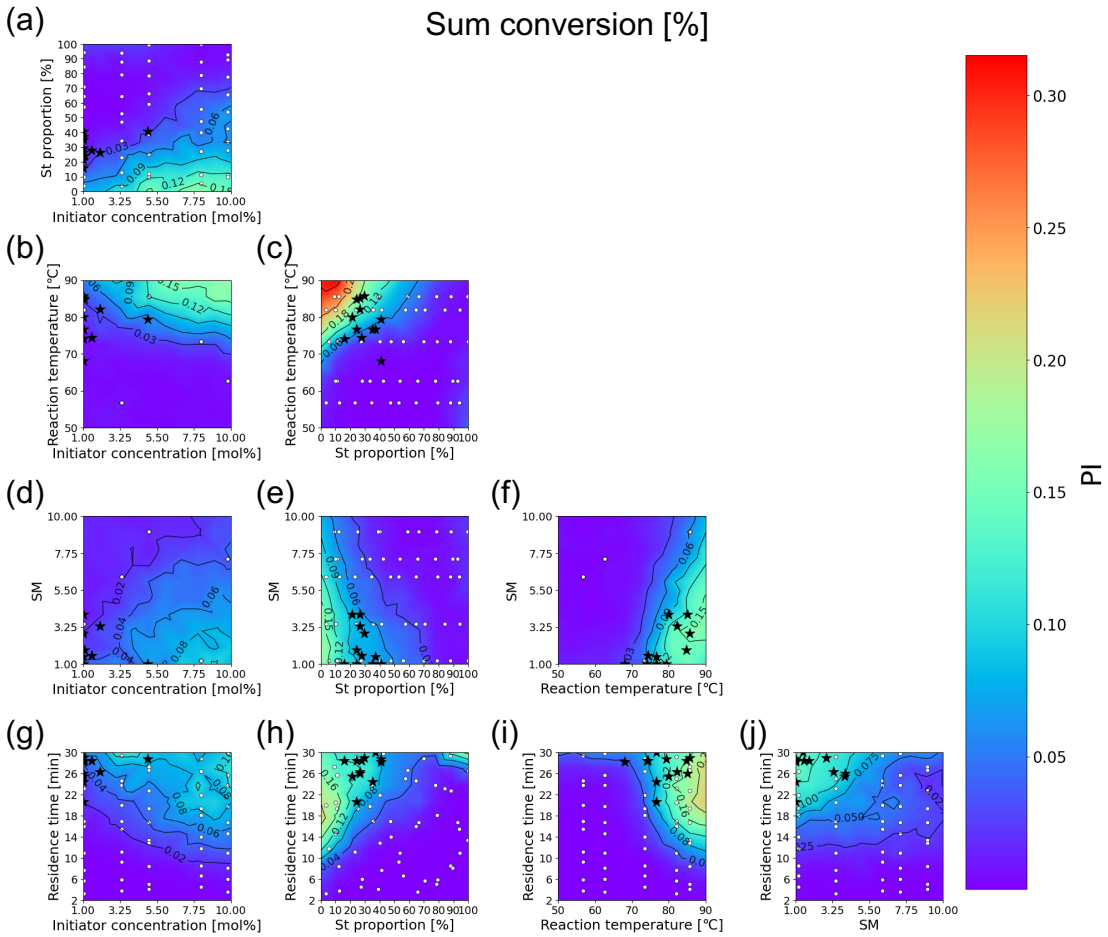

Figure 4 Visualization of the partial dependence of the PI of Sum conversion on the copolymerization conditions. The color-coded background is calculated from the PI obtained using GPR trained on data from the initial points to 12 cycles for Target 1. Red indicates high PI values, and blue indicates low PI values. Stars represent the experimental points proposed by MOBO, and white circles indicate initial points.

with lower Sum error than the initial points. These results demonstrate that MOBO can efficiently explore the vicinity of optimal solutions and identify copolymerization conditions that closely approach the target values.

Figure 2c shows a parallel plot with connected lines indicating the copolymerization conditions proposed by MOBO and the corresponding polymer characteristics obtained under these conditions. Samples are named according to the formula "Target 1-*cycle no.*" for MOBO for Target 1. For example, "Target 1-5" represents the 5th cycle of MOBO for Target 1. The copolymerization conditions proposed by MOBO show consistent trends (Figure 2c). These trends differ from our previous results of single-objective Bayesian optimization for St proportion in polymer.[27] In the single-objective optimization, even when the target composition was achieved, copolymerization conditions indicated a range of values. This difference in the trends confirms the effectiveness of the proposed method for multi-objective optimization problems. However, the resultant polymer characteristics did not achieve Target 1. All experimental results obtained using MOBO for Target 1 are presented in Table S1. A detailed analysis of Figure 2c reveals that both the initiator concentration and SM ratio converged to their lower boundary value of 1.0 after Target 1-9. This result indicates that Target 1 is challenging to achieve within the current exploration range. However, the exploration ranges for the Initiator concentration and SM ratio were not set lower than these values because of the possibility that the reaction may become too slow or that the solvents may be ineffective.

To gain further insight into these results, we conducted an analysis of the partial dependence of PI on copolymerization conditions for each polymer characteristic (St proportion in polymer, Sum conversion, M_w, and M_n).[49] Partial

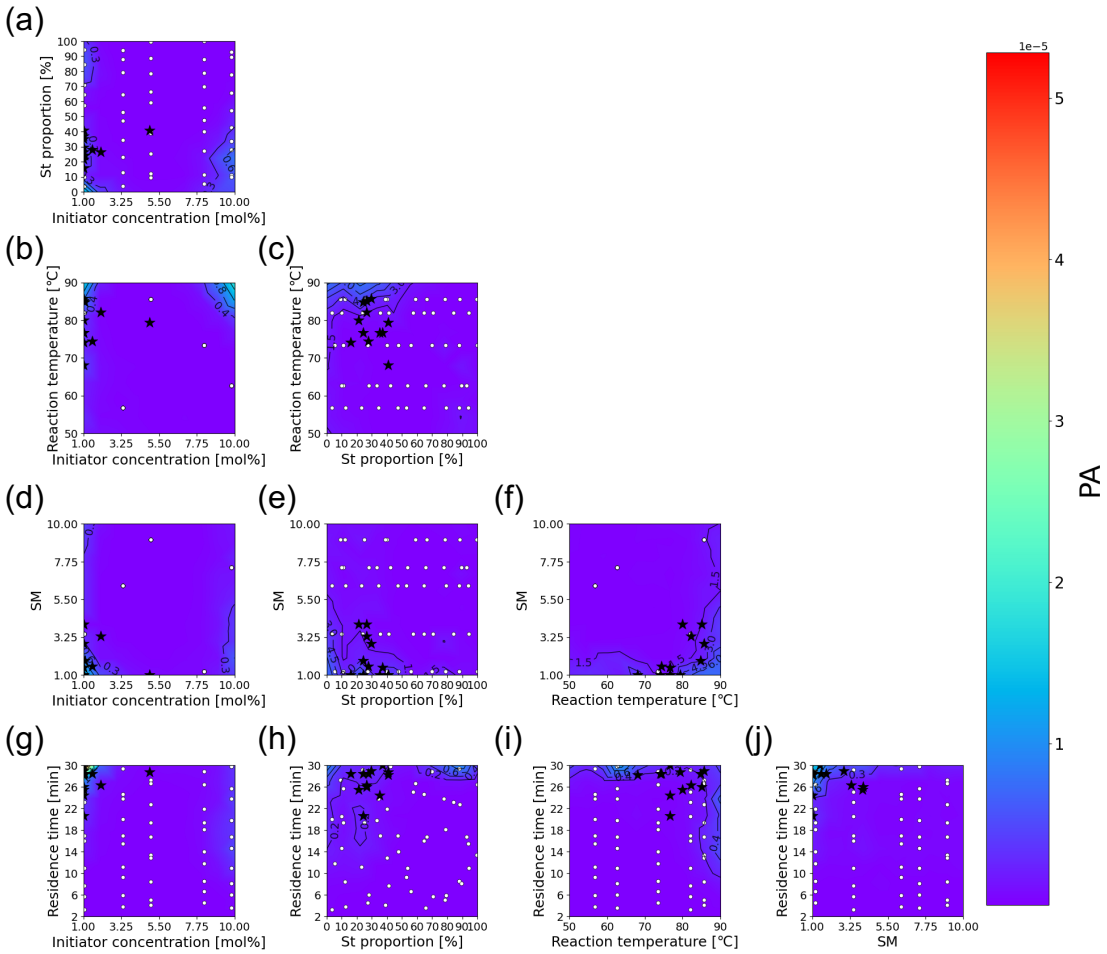

Figure 5 Visualization of the partial dependence of the PA on the copolymerization conditions. The color-coded background is calculated from the PA values obtained using GPR trained on data from the initial points to 12 cycles for Target 1. Red indicates high PA values, and blue indicates low PA values. Stars represent the experimental points proposed by MOBO, and white circles indicate initial points.

dependence plots (PDPs) of the St proportion in polymer and Sum conversion are shown in Figure 3 and 4, respectively. PDPs of the remaining two polymer characteristics, Mw and Mn, are shown in Figure S1 and Figure S2 in the Supporting Information, respectively. In the PDPs, the average of PI values is plotted as the Monte Carlo average for each coordinate point by keeping the two copolymerization conditions corresponding to the axes constant and varying the remaining conditions. The PI values were calculated from the GPR model constructed using the data from the initial points and 12 optimization cycles. The partial dependences are shown as color maps, where red and blue indicate high and low PI values, respectively. The higher PI values suggest a greater probability of improvement over the current best value by conducting experiments under the corresponding copolymerization conditions. Two types of data points are shown on the plots: stars represent the experimental points proposed by MOBO, and white circles indicate initial points. A crucial insight obtained from the PDPs is that regions showing the highest PI for one objective characteristic do not necessarily correspond to regions where high PI values for other characteristics are expected. This result indicates extreme difficulty in simultaneously achieving all four target values. Specifically, a comparison of Figure 3c and 4c reveals that, although the St proportion in polymer shows the optimal PI in the regions corresponding to an St proportion of about 40%, Sum conversion shows the optimal PI in the regions corresponding to about 10%. This conflicting tendency demonstrates the challenges of simultaneously optimizing multiple polymer characteristics. These analytical results highlight the complexity of multi-objective optimi-

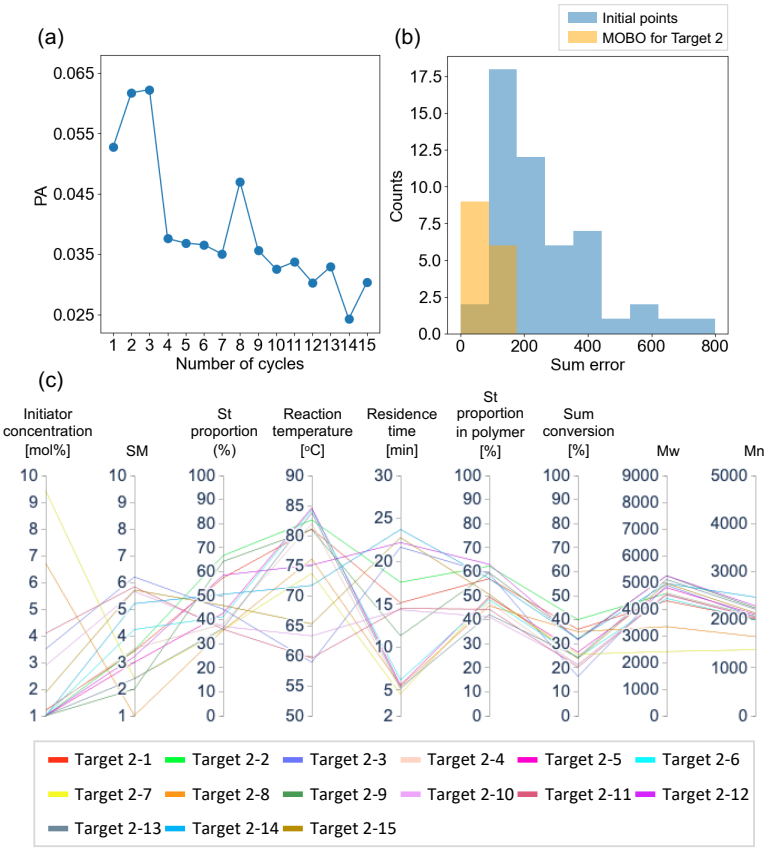

Figure 6 MOBO results for Target 2: (a) PA of candidate points at the n-th cycle calculated by GPR trained on the initial and all points up to the (n – 1)-th cycle; (b) distribution comparison of Sum error between experimental points selected by MOBO and LHS; and (c) parallel coordinates plot showing proposed copolymerization conditions and resultant polymer characteristics for each optimization iteration. Each experimental point is shown in a different color.

zation problems and the existence of the Pareto front between different polymer characteristics. Figure 5 shows the PDP of the PA for all combination of the copolymerization conditions. The partial dependencies of PA were calculated from the GPR model constructed with data from the initial points and 12 optimization cycles. Similarly, Figure 5 demonstrates the significant challenge in simultaneously achieving Target 1, as evidenced by the minimal presence of red regions indicating high PA values.

This design space did not achieve the Target 1 polymer characteristics. However, this result should be seen as revealing that Target 1 is realistically unattainable rather than invalidating the proposed method as an effective method for multi-objective optimization. A significant advantage of this method is its capacity to facilitate decision-making regarding research directions through the visualization of PDPs, even when confronted with challenging target values. PDPs effectively illustrate the trade-offs between polymer characteristics and visually represent the limitations of optimization within the current design space. Such visual representation enables researchers to make more objective and rapid decisions about whether to continue experiments or reconsider the target values and design space.

3.2 MOBO for Target 2. In this subsection, MOBO was conducted for Target 2. Figure 6a shows the PA of candidate points at the *n*-th cycle by GPR trained on the initial points and all points up to the (*n* – 1)-th cycle. The PA values show a steady decrease with increasing number of cycles. However, PA values increased in the 8th and 15th cycles. These increases can be attributed to changes in the GPR model following the experiments conducted in the 7th and 14th cycles. Table S2 shows PA values for experimental points at the *j*-th cycle, as calculated by the GPR model after completion of the *i*-th cycle. Comparing PA values for experimental points at the 8th cycle calculated by models after the 6th and 7th cycles shows an improvement from 0.02 to 0.047. This improvement indicates that the addition of the experimental point at the 7th cycle resulted in a substantial change in the GPR model for the

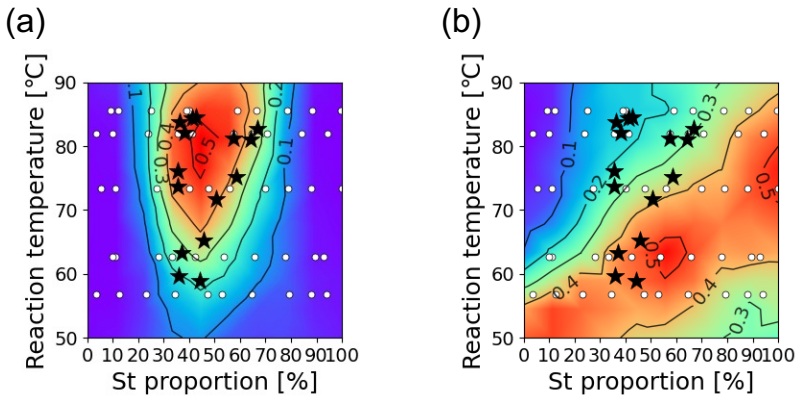

Figure 7 Visualization of the partial dependence of PI on St proportion and Reaction temperature showing (a) St proportion in polymer and (b) Sum conversion. The color-coded background is calculated from the PI values obtained using GPR trained on data from the initial points to 15 cycles for Target 2. Red indicates high PI values, and blue indicates low PI values. Stars represent the experimental points proposed by MOBO, and white circles indicate initial points.

selection of a candidate point at the 8th cycle. Similar improvements were observed between cycles 14 and 15. These improvements indicate that the proposed method not only exploits known promising regions but also explores new domains, thereby demonstrating global optimization progress through model refinement. Specifically, Table S3 presents all experimental results for Target 2, showing that experimental points at the 8th cycle explored a new region in terms of Initiator concentration.

Figure 6b shows the distribution of Sum error of the experimental points obtained through LHS and MOBO. A comparison of these distributions reveals that the experiments conducted through MOBO yielded more results with lower Sum error than experiments conducted through LHS, which is consistent with the results in Section 3.1.

Figure 6c shows a parallel plot with connected lines showing the copolymerization conditions proposed by MOBO and the corresponding polymer characteristics obtained under these conditions. Samples are named using the formula "Target 2-*cycle no*." for MOBO for Target 2. Analysis of Figure 6c reveals that copolymerization conditions of high Reaction temperature and short Residence time were initially proposed. Following the discovery of a promising sample at Target 2-5, the proposed conditions gradually shifted to copolymerization conditions of low Reaction temperature and long Residence time, ultimately leading to the discovery of a sample with polymer properties closer to Target 2 at Target 2-15. These results indicate that MOBO has effectively used the information obtained during the exploration process and shifted its strategy from local to global optimization. The experimental results for Target 2-5 and Target 2-15 are shown in Table 3.

Table 3 Results for Target 2-5 and Target 2-15 with values close to Target 2

Copolymerization conditions and polymer characteristic	Unit	Target 2-5	Target 2-15
Initiator concentration	mol%	1.0	1.87
SM ratio		3.0	5.71
St proportion	mol%	42.64	45.88
Reaction temperature	°C	84.56	65.28
Residence time	min	5.57	22.76
St proportion in polymer	%	49.49	50.52
Sum conversion	%	26.48	23.81
M_w		4799.0	4961.0
M_n		2109.0	2139.0

Figure 7 shows the PDPs of PI to St proportion and Reaction temperature for St proportion in polymer and Sum conversion. All PDPs are provided in the Supporting Information as Figure S3, S4, S5, and S6. The PI shown in the plots was calculated from a GPR model constructed using data from the initial points and 15 optimization cycles. The PDPs shown in Figure 7 indicate that Target 2 is more achievable than Target 1. A comparison of Figure 7a and Figure 7b reveals overlapping regions of high PI indicated by red areas. This overlap suggests a high probability of simultaneously achieving the target values for St proportion in polymer and Sum conversion. This result contrasts with the mutually exclusive tendencies observed when comparing PDPs focused on the same copolymerization conditions in Figure 3c and 4c. Figure 8 shows the PDP of

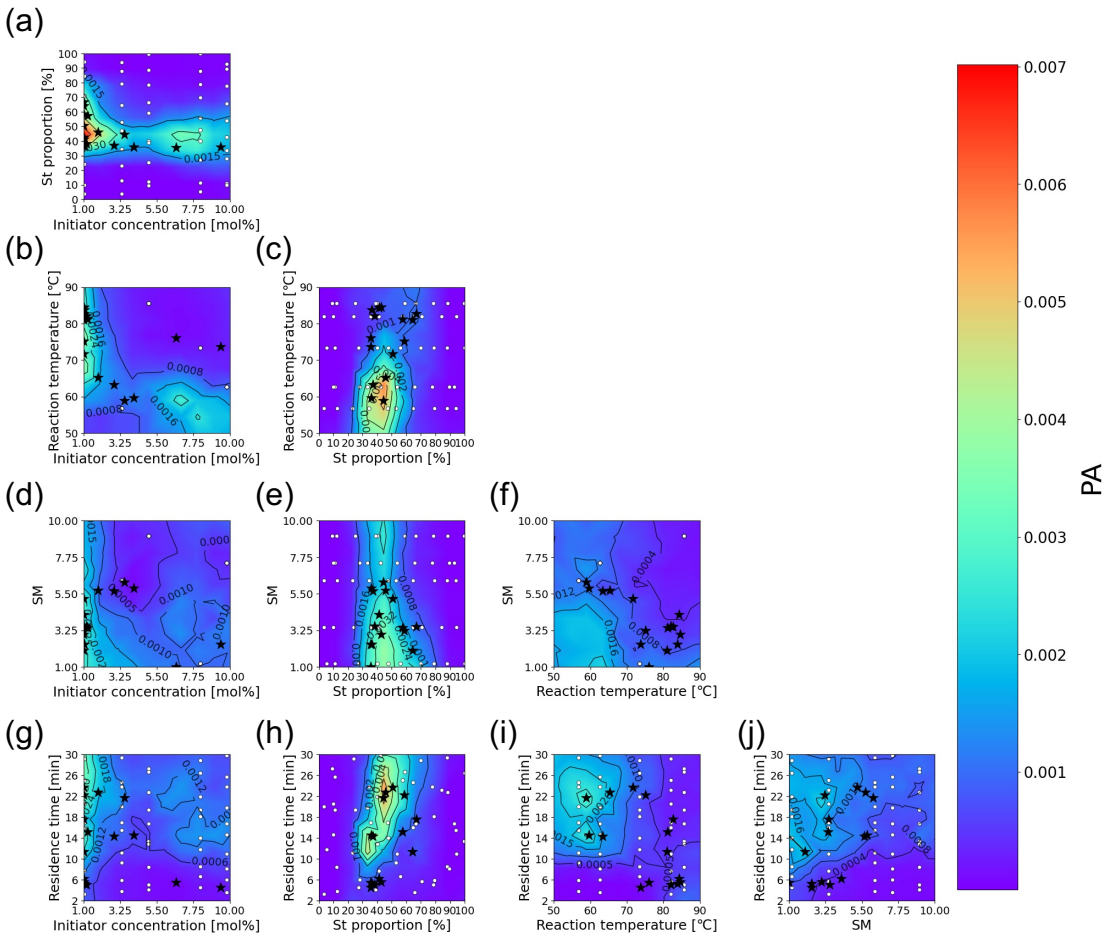

Figure 8 Visualization of the partial dependence of the PA on the copolymerization conditions. The color-coded background is calculated from the PA values obtained using GPR trained on data from the initial points to 15 cycles for Target 2. Red indicates high PA values, and blue indicates low PA values. Stars represent the experimental points proposed by MOBO, and white circles indicate initial points.

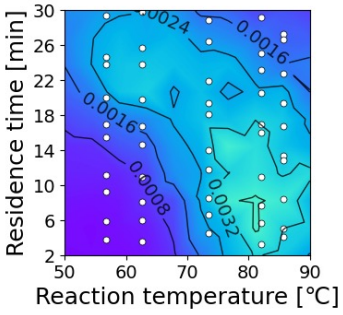

Figure 9 Visualization of the partial dependence of the PA on Reaction temperature and Residence time. The color-coded background is calculated from the PA values obtained using GPR trained on the initial points. As with Figure 3 to 8, red indicates high PA values, and blue indicates low PA values. White circles indicate initial points.

the PA to copolymerization conditions. PA values shown in the plots were calculated from the GPR model constructed with data from the initial points and 15 optimization cycles. Similarly, Figure 8 reveals regions of high PA indicated by red areas, suggesting that Target 2 is more achievable than Target 1. Figure 9 shows the PDPs of PA to Reaction temperature and Residence time, where the PA values in the plots were calculated from the GPR model constructed using only initial points (pre-MOBO implementation). PDPs for other combinations of copolymerization conditions are provided in the Supporting Information as Figure S7. A comparison of Figure 8i and Figure 9 demonstrates that the optimal region transitioned from high Reaction temperature and short Residence time before MOBO to low Reaction temperature and long Residence time after MOBO. The transition in the optimal regions is consistent with the changes in the copolymerization conditions proposed by

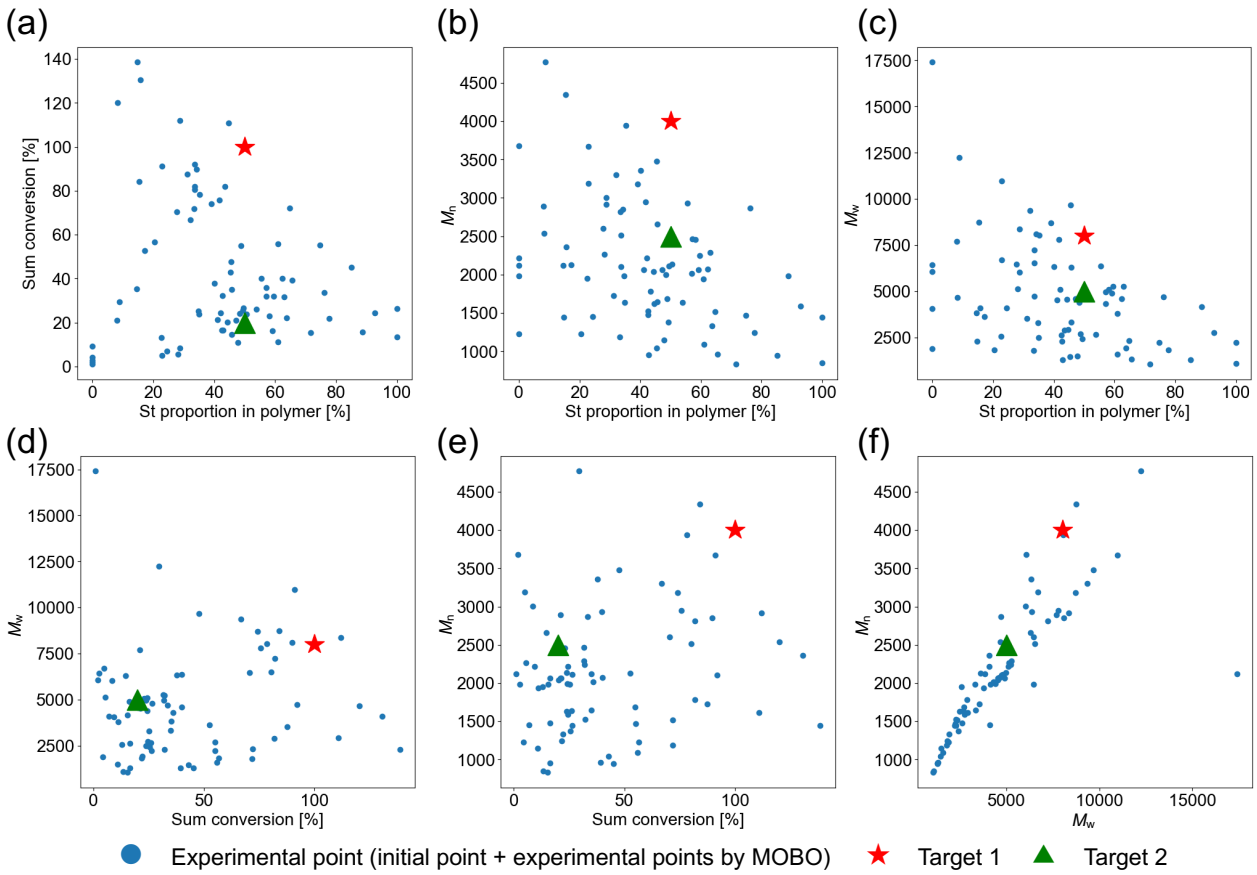

Figure 10 Two-dimensional scatter plots for all combinations of polymer characteristics. Blue circles represent initial points and MOBO experimental points for Target 1 and Target 2. Red stars and green triangles represent Target 1 and Target 2, respectively.

MOBO. This correspondence observed through analysis of PDPs indicates that MOBO successfully transitioned its strategy from local to global optimization.

The accuracy of the GPR model constructed through 15 optimization cycles was evaluated by cross-validation. Figure S8a, Figure S8b, Figure S8c, and Figure S8d show results of cross-validation for each polymer characteristic (St proportion in polymer, Sum conversion, M_w, and M_n, respectively). In Figure S8d, the coefficient of determination for M_w is low: −0.076. This low accuracy is attributable to the influence of outliers in the initial points. These outliers are copolymerization conditions under which the reaction was assumed to have not sufficiently progressed because of low reaction temperatures or a short residence time. Details of outliers are described in Table S4. The results of cross-validation performed after these outliers were excluded are shown in Figure S9a–d. Figure S9d shows that the coefficient of determination for M_w substantially improved to 0.85, indicating that the outliers significantly reduce the prediction accuracy. However, in the present study, we prioritized the automated progression of optimization. Therefore, we did not remove the outliers. Moreover, the primary objective of MOBO was not to construct a high-accuracy prediction model but rather to discover combinations of explanatory variables that simultaneously optimize all objective functions with minimal experimental iterations. Because Target 2 achieved this objective, we did not pursue further improvements in prediction accuracy.

3.3 Analysis of Pareto Fronts by Quantum Chemical Calculations. The results in Sections 3.1 and 3.2 reveal that the existence of a Pareto front between polymer characteristics makes the simultaneous achievement of target values possible or impossible in some cases. This section examines the chemical factors that lead to the existence of Pareto fronts. Figure 10 shows a two-dimensional scatter plot of polymer characteristics for all combinations of the polymer characteristics. The scatter plots confirm the existence of Pareto fronts among the properties except for Figure 10f. Target 1 was clearly confirmed to be situated

outside the Pareto front in the relationship between the St proportion in the copolymer and M_n. For all other property combinations, Target 1 was situated on Pareto front boundaries. These results suggest that achieving Target 1 presents greater challenges than achieving Target 2. In addition, from the analysis of Figure 10a and 10b, we found that Sum conversion and M_n tend to decrease with increasing St proportion in polymer. This observation means that the increase in St proportion is a factor that prevents the copolymerization reaction from progressing. To clarify the chemical factors responsible for this phenomenon, we conducted an analysis using the results of quantum chemical calculations. Table 4 shows the activation barriers during the approach of St or MMA monomers to St or MMA radicals. Table 4 quotes the values from Table S2 in Ref. 17. These values show that the activation barrier for the monomer reactive with the St radical is as much as ~10 kJ/mol higher than that for the monomer reacting with the MMA radical. Applying the Arrhenius equation, this difference in activation barriers corresponds to an approximately 31-fold difference in reaction rate constant at 350 K. This difference is considered significant.

From these results, we concluded that the existence of Pareto fronts among polymer characteristics is due to the higher activation energy in the propagation reaction when the polymer end is St compared to when the polymer end is MMA.

Table 4 Dependence of activation energies on the radical species at polymer ends and on the monomer approaching the radical. The values of activation energies are extracted from Ref. 17.

Radical species	Monomer	Activation energy [kJ/mol] a)	Average of activation energy [kJ/mol]
Me-MMA·	MMA	53.2	
Me-MMA-MMA·	MMA	52.5	51.63
Me-St-MMA·	MMA	49.2	
Me-MMA·	St	52.1	
Me-MMA-MMA·	St	52.5	52.73
Me-St-MMA·	St	53.6	
Me-St ·	MMA	60.5	
Me-MMA-St·	MMA	55.6	56.77
Me-St-St·	MMA	54.2	
Me-St ·	St	63.9	
Me-MMA-St·	St	61.0	60.83
Me-St-St·	St	57.6	

a) All the calculations were at the B3LYP-D3/def2-SVP level of theory.

4. CONCLUSIONS

In this study, we established a method to search for copolymerization conditions that simultaneously optimize multiple polymer characteristics by combining MOBO and a flow copolymerization system. In addition, the analysis of multidimensional data by PDP revealed the existence of the Pareto front between polymer characteristics and quantum chemical calculations demonstrated that this Pareto front is due to the high activation energy of the propagation reaction at the St-terminus. The advantage of the proposed method is that it does not simply search for the optimal solution but objectively evaluates the validity of the problem setting itself through visualization by PDP. It can therefore function as a powerful tool to promote appropriate modification of a research strategy.

On the other hand, in the present study, initial points were obtained for copolymers of St and MMA. However, if other monomers are used, new initial points will need to be obtained. This experimental burden is a barrier to the practical application of the proposed method in terms of time and economic costs. To address this issue, it is necessary to establish a MOBO method that eliminates the need to reacquire initial points by effectively using past experimental data. Another important issue is to expand the design space to increase the number of copolymers that can be synthesized. Specifically, the design space can be expanded by introducing new copolymerization conditions such as solvent species. Controlling activation energy through solvent effects is expected to increase the flexibility of copolymerization reactions and enable the synthesis of various copolymers. In the future, our goal is to solve these problems and establish a new polymer design method that is both practical and efficient.

DATA AVAILABILITY

The source code and data for our model are publicly available at https://github.com/MI-Lab-NAIST/Multi-objective_Bayesian_optimization_for_experimental_design_in_copolymerization.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Additional partial dependence plots; prediction accuracy of GPR models constructed through cycle of MOBO for Target 2; copolymerization conditions and results proposed by MOBO for Target 1 and Target 2; and outliers in the initial points.

AUTHOR INFORMATION

Corresponding Author

Mikiya Fujii - Graduate School of Science and Technology, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara, 630-0192, Japan; Data Science Center, Nara

Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara, 630-0192, Japan; Center for Material Research Platform, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara 630-0192, Japan; orcid.org/0000-0002-3728-3097; E-mail: fujii.mikiya@ms.naist.jp

Author Contributions

H. Yamada: Investigation (ML, EXP), Writing - Original Draft. S. Takasuka: Investigation (EXP). S. Oikawa: Software. Y. Harashima: Resources. T. Takayama: Formal analysis (EXP). Aniruddha Nag: Investigation (EXP). Araki Wakiuchi: Investigation (EXP), Software. T. Sugawara: Conceptualization (EXP), Formal analysis (EXP). M. Hatanaka: Formal analysis (ML). T. Miyao: Formal analysis (ML). T. Matsubara: Supervision (ML). Y. Ohnishi: Formal analysis (ML). H. Ajiro: Investigation (EXP). M. Fujii: Project administration, where ML and EXP represent machine learning and experiments.

CONFLICTS OF INTEREST

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This article is based on results obtained from a project, JPNP14004, subsidized by the New Energy and Industrial Technology Development Organization (NEDO) and JSPS KAKENHI Grant Number JP21K20537.

REFERENCES

(1) Ma, P.; Wen, S.; Wang, C.; Guo, W.; Shen, L.; Dong, W.; Lu, J.; Ruan, S. Optimization of PDTS-DTffBT-Based Solar Cell Performance through Control of Polymer Molecular Weight. *J. Phys. Chem. C* **2016**, *120* (35), 19513–19520.

(2) Intemann, J. J.; Yao, K.; Yip, H.-L.; Xu, Y.-X.; Li, Y.-X.; Liang, P.-W.; Ding, F.-Z.; Li, X.; Jen, A. K.-Y. Molecular Weight Effect on the Absorption, Charge Carrier Mobility, and Photovoltaic Performance of an Indacenodiselenophene-Based Ladder-Type Polymer. *Chem. Mater.* **2013**, *25* (15), 3188–3195.

(3) Guo, J.; Liu, X.; Liu, M.; Han, M.; Liu, Y.; Ji, S. Effect of Molecular Weight of Poly(Ethylene Glycol) on Plasticization of Poly(L-Lactic Acid). *Polymer* **2021**, *223*, 123720.

(4) Liu, C.; Wang, K.; Hu, X.; Yang, Y.; Hsu, C.-H.; Zhang, W.; Xiao, S.; Gong, X.; Cao, Y. Molecular Weight Effect on the Efficiency of Polymer Solar Cells. *ACS Appl. Mater. Interfaces* **2013**, *5* (22), 12163–12167.

(5) Hester, H. G.; Abel, B. A.; Coates, G. W. Ultra-High-Molecular-Weight Poly(Dioxolane): Enhancing the Mechanical Performance of a Chemically Recyclable Polymer. *J. Am. Chem. Soc.* **2023**, *145* (16), 8800–8804.

(6) Scheidelaar, S.; Koorengevel, M. C.; van Walree, C. A.; Dominguez, J. J.; Dörr, J. M.; Killian, J. A. Effect of Polymer Composition and PH on Membrane Solubilization by Styrene-Maleic Acid Copolymers. *Biophys. J.* **2016**, *111* (9), 1974–1986.

(7) Shibata, M.; Sugane, K.; Yanagisawa, Y. Biobased Polymer Networks by the Thiol-Ene Photopolymerization of Allylated *p*-Coumaric and Caffeic Acids. *Polym. J.* **2019**, *51* (5), 461–470.

(8) Wei, J.; Chu, X.; Sun, X.-Y.; Xu, K.; Deng, H.-X.; Chen, J.; Wei, Z.; Lei, M. Machine Learning in Materials Science. *InfoMat* **2019**, *1* (3), 338–358.

(9) Gubernatis, J. E.; Lookman, T. Machine Learning in Materials Design and Discovery: Examples from the Present and Suggestions for the Future. *Phys. Rev. Mater.* **2018**, *2* (12), 120301. https://doi.org/10.1103/physrevmaterials.2.120301

(10) Morgan, D.; Jacobs, R. Opportunities and Challenges for Machine Learning in Materials Science. *Annu. Rev. Mater. Res.* **2020**, *50* (1), 71–103.

(11) Liu, Y.; Zhao, T.; Ju, W.; Shi, S. Materials Discovery and Design Using Machine Learning. *J. Mater.* **2017**, *3* (3), 159–177.

(12) Rajan, K. Materials Informatics. *Mater. Today (Kidlington)* **2005**, *8* (10), 38–45.

(13) Otsuka, S.; Kuwajima, I.; Hosoya, J.; Xu, Y.; Yamazaki, M. PoLyInfo: Polymer Database for Polymeric Materials Design. In 2011 International Conference on Emerging Intelligent Data and Web Technologies; IEEE, 2011, pp 22– 29.

(14) Kim, C.; Chandrasekaran, A.; Huan, T. D.; Das, D.; Ramprasad, R. Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions. *J. Phys. Chem. C* **2018**, *122* (31), 17575–17585.

(15) Huan, T. D.; Mannodi-Kanakkithodi, A.; Kim, C.; Sharma, V.; Pilania, G.; Ramprasad, R. A Polymer Dataset for Accelerated Property Prediction and Design. *Sci. Data* **2016**, *3* (1), 160012.

(16) Takahashi, K.-I.; Mamitsuka, H.; Tosaka, M.; Zhu, N.; Yamago, S. CoPolDB: A Copolymerization Database for Radical Polymerization. *Polym. Chem*. **2024**, *15* (10), 965–971.

(17) Yoshimura, T.; Kato, H.; Oikawa, S.; Inagaki, T.; Asano, S.; Sugawara, T.; Miyao, T.; Matsubara, T.; Ajiro, H.; Fujii, M.; Ohnishi, Y.-Y.; Hatanaka, M. CopDDB: A Descriptor Database for Copolymers and Its Applications to Machine Learning. Digit. Discov. 2025, 4 (1), 195–203. https://doi.org/10.1039/d4dd00266k.

(18) Ishii, M.; Ito, T.; Sado, H.; Kuwajima, I. NIMS Polymer Database PoLyInfo (I): An Overarching View of Half a Million Data Points. *Sci. Technol. Adv. Mater.: Methods* **2024**, *4* (1), 2354649. https://doi.org/10.1080/27660400.2024.2354649.

(19) Ishii, M.; Ito, T.; Sakamoto, K. NIMS Polymer Database PoLyInfo (II): Machine-Readable Standardization of Polymer Knowledge Expression. *Sci. Technol. Adv. Mater.: Methods* **2024**, *4* (1), 2354651. https://doi.org/10.1080/27660400.2024.2354651.

(20) Audus, D. J.; de Pablo, J. J. Polymer Informatics: Opportunities and Challenges. *ACS Macro Lett.* **2017**, *6* (10), 1078–1082.

(21) Chen, L.; Pilania, G.; Batra, R.; Huan, T. D.; Kim, C.; Kuenneth, C.; Ramprasad, R. Polymer Informatics: Current Status and Critical next Steps. *Mater. Sci. Eng. R Rep.* **2021**, *144*, 100595.

(22) Cencer, M. M.; Moore, J. S.; Assary, R. S. Machine Learning for Polymeric Materials: An Introduction. *Polym. Int.* **2022**, *71* (5), 537–542.

(23) Ma, R.; Zhang, H.; Xu, J.; Sun, L.; Hayashi, Y.; Yoshida, R.; Shiomi, J.; Wang, J.-X.; Luo, T. Machine Learning-Assisted Exploration of Thermally Conductive Polymers Based on High-Throughput Molecular Dynamics Simulations. *Mater. Today Phys.* **2022**, *28*, 100850.

(24) Hayashi, Y.; Shiomi, J.; Morikawa, J.; Yoshida, R. RadonPy: Automated Physical Property Calculation Using All-Atom Classical Molecular Dynamics Simulations for Polymer Informatics. *npj Comput. Mater.* **2022**, *8* (1), 1–15.

(25) Lin, T.-S.; Coley, C. W.; Mochigase, H.; Beech, H. K.; Wang, W.; Wang, Z.; Woods, E.; Craig, S. L.; Johnson, J. A.; Kalow, J. A.; Jensen, K. F.; Olsen, B. D. BigSMILES: A Structurally-Based Line Notation for Describing Macromolecules. *ACS Cent. Sci.* **2019**, *5* (9), 1523–1531.

(26) Schneider, L.; Walsh, D.; Olsen, B.; de Pablo, J. Generative BigSMILES: An Extension for Polymer Informatics, Computer Simulations & ML/AI. *Digital Discovery* **2024**, *3* (1), 51–61.
(27) Takasuka, S.; Ito, S.; Oikawa, S.; Harashima, Y.; Takayama, T.; Nag, A.; Wakiuchi, A.; Ando, T.; Sugawara, T.; Hatanaka, M.; Miyao, T.; Matsubara, T.; Ohnishi, Y.; Ajiro, H.; Fujii, M. Bayesian Optimization of Radical Polymerization Reactions in a Flow Synthesis System. *Sci. Technol. Adv. Mater.: Methods* **2024**, *4* (1), 2425178 https://doi.org/10.1080/27660400.2024.2425178.
(28) Kershaw, O. J.; Clayton, A. D.; Manson, J. A.; Barthelme, A.; Pavey, J.; Peach, P.; Mustakis, J.; Howard, R. M.; Chamberlain, T. W.; Warren, N. J.; Bourne, R. A. Machine Learning Directed Multi-Objective Optimization of Mixed Variable Chemical Systems. *Chem. Eng. J.* **2023**, *451*, 138443.
(29) Sattari, K.; Wu, Y.; Chen, Z.; Mahjoubnia, A.; Su, C.; Lin, J. Physics-Constrained Multi-Objective Bayesian Optimization to Accelerate 3d Printing of Thermoplastics. *Addit. Manuf.* **2024**, *86*, 104204.
(30) Albuquerque, R. Q.; Rothenhäusler, F.; Gröbel, P.; Ruckdäschel, H. Multi-Objective Optimization of Sustainable Epoxy Resin Systems through Bayesian Optimization and Machine Learning. *ACS Appl. Eng. Mater.* **2023**, *1* (12), 3298–3308.
(31) Pruksawan, S.; Lambard, G.; Samitsu, S.; Sodeyama, K.; Naito, M. Prediction and Optimization of Epoxy Adhesive Strength from a Small Dataset through Active Learning. *Sci. Technol. Adv. Mater.* **2019**, *20* (1), 1010–1021.
(32) Hanaoka, K. Bayesian Optimization for Goal-Oriented Multi-Objective Inverse Material Design. *iScience* **2021**, *24* (7), 102781.
(33) Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R. P.; de Freitas, N. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* **2016**, *104* (1), 148–175.
(34) Wang, A.; Ye, H.; Yang, Y.; Dong, H. Bayesian Optimization of HDPE Copolymerization Process Based on Polymer Product-Process Integration. *Polymer* **2024**, *292*, 126554.
(35) Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I. A Mobile Robotic Chemist. *Nature* **2020**, *583* (7815), 237–241.
(36) Kodera, M.; Sayama, K. An Automatic Robot System for Machine Learning–Assisted High-Throughput Screening of Composite Electrocatalysts. *Digital Discovery* **2023**, *2* (6), 1683–1687.
(37) Nishikawa, C.; Nishikubo, R.; Ishiwari, F.; Saeki, A. Exploration of Solution-Processed Bi/Sb Solar Cells by Automated Robotic Experiments Equipped with Microwave Conductivity. *JACS Au* **2023**, *3* (11), 3194–3203.
(38) Hartman, R. L.; McMullen, J. P.; Jensen, K. F. Deciding Whether to Go with the Flow: Evaluating the Merits of Flow Reactors for Synthesis. *Angew. Chem. Int. Ed. Engl.* **2011**, *50* (33), 7502–7519.
(39) Reis, M. H.; Leibfarth, F. A.; Pitet, L. M. Polymerizations in Continuous Flow: Recent Advances in the Synthesis of Diverse Polymeric Materials. *ACS Macro Lett.* **2020**, *9* (1), 123–133.
(40) Coley, C. W.; Thomas, D. A.; Lummiss, J. A. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; Hicklin, R. W.; Plehiers, P. P.; Byington, J.; Piotti, J. S.; Green, W. H.; Hart, A. J.; Jamison, T. F.; Jensen, K. F. A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning. *Science* **2019**, *365* (6453), eaax1566.
(41) Knox, S. T.; Parkinson, S. J.; Wilding, C. Y. P.; Bourne, R. A.; Warren, N. J. Autonomous Polymer Synthesis Delivered by Multi-Objective Closed-Loop Optimisation. *Polym. Chem.* **2022**, *13* (11), 1576–1585.
(42) Rong, L.-H.; Cheng, X.; Ge, J.; Caldona, E. B.; Advincula, R. C. Synthesis of Hyperbranched Polymers via PET-RAFT Self-condensing Vinyl Polymerization in a Flow Reactor. *Macromol. Chem. Phys.* **2022**, *223* (1), 2100342.
(43) Senthil Vel, A.; Konan, K. E.; Cortés-Borda, D.; Felpin, F.-X. Enhancing Optimization of Mixed Variables on a Robotic Flow Platform: Integrating Statistical Filtering with Nelder–Mead and Bayesian Methods. *Org. Process Res. Dev.* **2023**, 28 (5), 1597–1606 https://doi.org/10.1021/acs.oprd.3c00238.
(44) Dunlap, J. H.; Ethier, J. G.; Putnam-Neeb, A. A.; Iyer, S.; Luo, S.-X. L.; Feng, H.; Torres, J. A. G.; Doyle, A. G.; Swager, T. M.; Vaia, R. A.; Mirau, P.; Crouse, C. A.; Baldwin, L. A. Continuous Flow Synthesis of Pyridinium Salts Accelerated by Multi-Objective Bayesian Optimization with Active Learning. *Chem. Sci.* **2023**, *14* (30), 8061–8069.
(45) Sugisawa, N.; Sugisawa, H.; Otake, Y.; Krems, R. V.; Nakamura, H.; Fuse, S. Rapid and Mild One-flow Synthetic Approach to Unsymmetrical Sulfamides Guided by Bayesian Optimization. *Chem. Methods* **2021**, *1* (11), 484–490.
(46) Wakiuchi, A.; Takasuka, S.; Asano, S.; Hashizume, R.; Nag, A.; Hatanaka, M.; Miyao, T.; Ohnishi, Y.; Matsubara, T.; Ando, T.; Sugawara, T.; Fujii, M.; Ajiro, H. Composition Regulation by Flow Copolymerization of Methyl Methacrylate and Glycidyl Methacrylate with Free Radical Method. *Macromol. Mater. Eng.* **2023**, *308* (6), 2200626.
(47) Takasuka, S.; Oikawa, S.; Yoshimura, T.; Ito, S.; Harashima, Y.; Takayama, T.; Asano, S.; Kurosawa, A.; Sugawara, T.; Hatanaka, M.; Miyao, T.; Matsubara, T.; Ohnishi, Y.; Ajiro, H.; Fujii, M. Extrapolation Performance Improvement by Quantum Chemical Calculations for Machine-Learning-Based Predictions of Flow-Synthesized Binary Copolymers. *Digital Discovery* **2023**, *2* (3), 809–818.
(48) Wakiuchi, A.; Jasial, S.; Asano, S.; Hashizume, R.; Hatanaka, M.; Ohnishi, Y.; Matsubara, T.; Ajiro, H.; Sugawara, T.; Fujii, M.; Miyao, T. Chemometrics Approach Based on Wavelet Transforms for the Estimation of Monomer Concentrations from FTIR Spectra. *ACS Omega* **2023**, *8* (22), 19781–19788.
(49) Friedman JH. Greedy function approximation: a gradient boosting machine. Ann. Stat. 2001, 29 (5), 1189–1232.

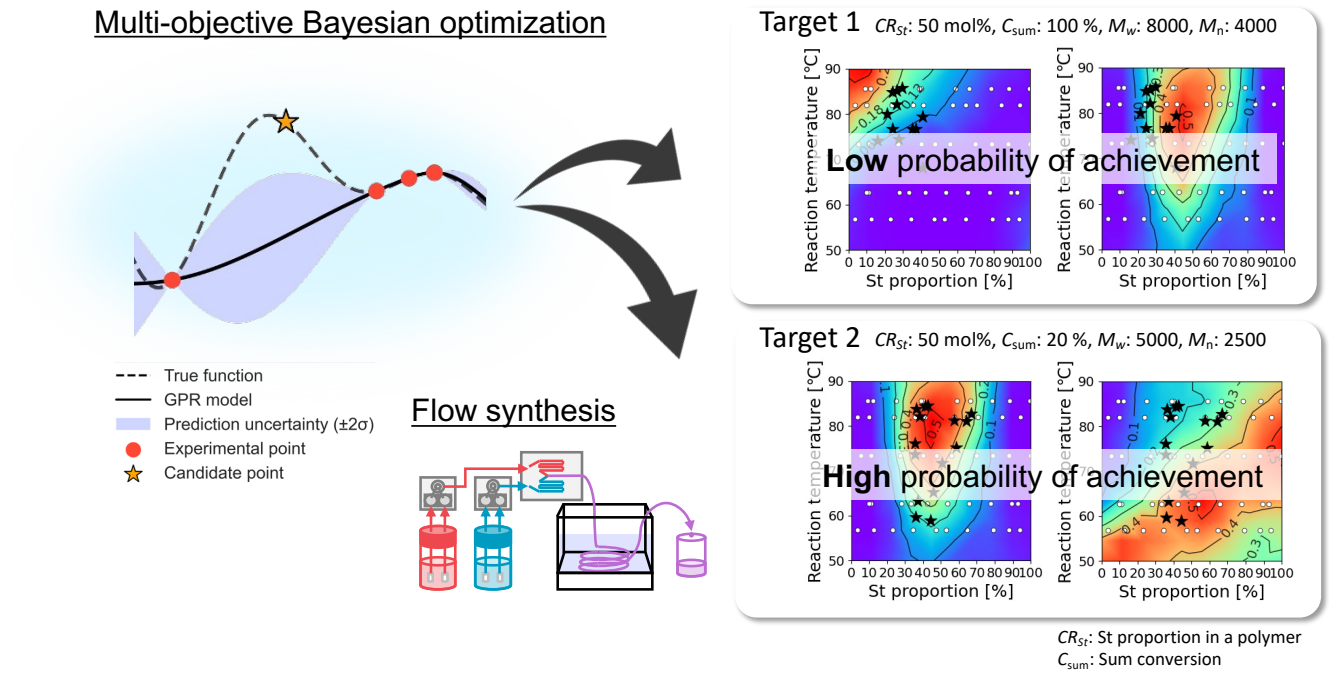

Table of Contents