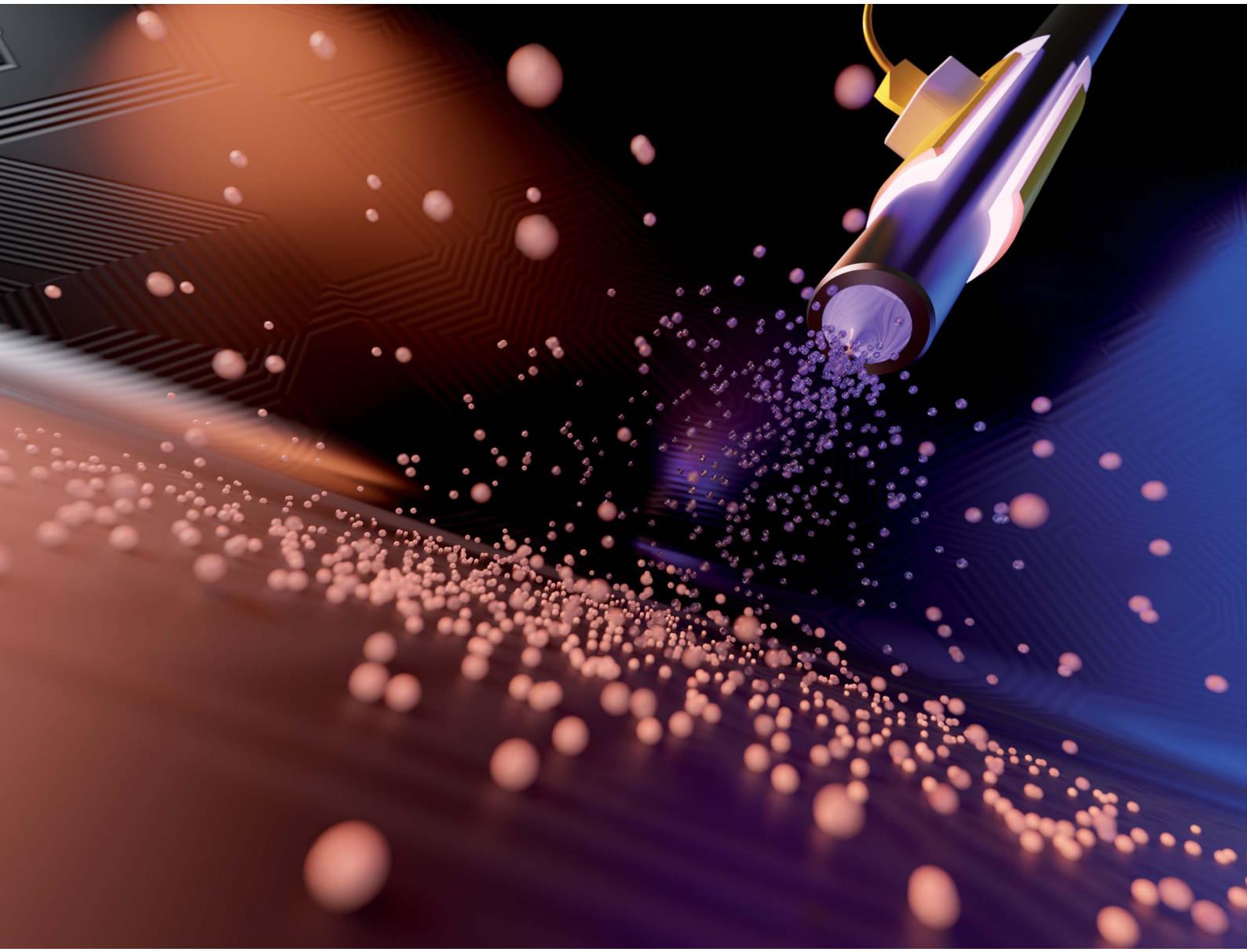
Volume 4
Number 11
November 2025
Pages 3055-3414

# Digital Discovery

rsc.li/digitaldiscovery



ISSN 2635-098X

ROYAL SOCIETY OF CHEMISTRY

**PAPER**
Fanjin Wang *et al.*
Constrained composite Bayesian optimization for rational synthesis of polymeric particles

Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue



Cite this: *Digital Discovery*, 2025, **4**, 3066

# Constrained composite Bayesian optimization for rational synthesis of polymeric particles

Fanjin Wang, *[ab] Maryam Parhizkar, [b] Anthony Harker[c] and Mohan Edirisinghe[a]

Polymeric nanoparticles have critical roles in tackling healthcare and energy challenges with miniature characteristics. However, tailoring synthesis processes to meet design targets has traditionally depended on domain expertise and trial-and-error. Modeling strategies, particularly Bayesian optimization, facilitate the discovery of materials with maximized/minimized properties. Based on practical demands, this study integrates constrained composite Bayesian optimization (CCBO) to perform target-value optimization under black-box feasibility constraints for by-design nanoparticle production. In a synthetic problem that simulates electrospraying, a representative nanomanufacturing process, CCBO avoided infeasible conditions and efficiently optimized towards predefined size targets, surpassing the baseline methods and state-of-the-art optimization pipelines. CCBO was also observed to provide decisions comparable to those of experienced experts in a human *vs.* BO campaign. Furthermore, laboratory experiments validated the use of CCBO for the guided synthesis of poly(lactic-*co*-glycolic acid) particles with diameters of 300 nm and 3.0 μm *via* electrospraying under minimal initial data. Overall, the CCBO approach represents a versatile and holistic optimization paradigm for next-generation target-driven particle synthesis empowered by artificial intelligence (AI).

Received 2nd June 2025
Accepted 1st August 2025

DOI: 10.1039/d5dd00243e

rsc.li/digitaldiscovery



## Introduction

Polymeric micro- and nano-particles have received great attention in pharmaceutical, catalysis, and energy applications due to their unique properties at a small scale.[1,2] Diverse design requirements for particles under the quality-by-design (QbD) framework have been put forward for specific usages.[3] For example, drug delivery platform particles span a wide size range from hundreds of nanometers for intravenous injection to micrometers for pulmonary administration.[4] However, the optimization of syntheses to meet these design requirements using various manufacturing technologies has mainly relied on human expertise and extensive trial-and-error experimentation. Modeling strategies could facilitate the optimization of parameters towards design targets.[5,6] Traditional design of experiment (DoE) strategies can identify dominating factors in the processing parameters and provide direction towards optimization, but the methodology becomes less effective in high-dimensional problems or complex relationships.[7] For example, orthogonal experiment designs such as the Plackett–Burman and Taguchi methods can typically accommodate up to three levels for each variable.[8] Moreover, it is also difficult to incorporate experiment feasibility into DoE optimization frameworks unless analytical descriptions of constraints are available. As a different approach, machine learning (ML) is powerful for modeling complicated relationships.[9,10] Using ML models as surrogates, adaptive sampling methods design sequential experiments for laboratory evaluation.[11] Bayesian optimization (BO) was developed for the efficient optimization of black-box functions and works well under small data regimes.[12–14] It employs a Gaussian process (GP) as a surrogate model, leveraging its ability to provide both mean and variance estimations for candidate selection. A carefully designed acquisition function is then used to score the candidates to explore uncertain points as well as to exploit promising optimal points.

More recently, BO has been investigated for materials and drug discovery to assist in the identification of optimal properties.[15–17] However, the application of BO in the targeted synthesis of materials presents two critical challenges. First, conventional BO was developed to seek a global maximum or minimum rather than to match a pre-defined target.[18,19] The latter is a common requirement in materials development tasks, *e.g.*, matching physiological mechanical properties for hydrogels and tailoring release profiles for drug delivery agents. Despite its relevance, the target-matching problem remains surprisingly underexplored in BO applications for materials discovery. This may be attributed to the prevailing emphasis on discovering materials with extreme or superior properties rather than materials that meet specific design criteria. Another issue

[a]*Department of Mechanical Engineering, University College London, London, WC1E 7JE, UK. E-mail: fanjin.wang.20@ucl.ac.uk*

[b]*School of Pharmacy, University College London, London, WC1N 1AX, UK*

[c]*Department of Physics and Astronomy, University College London, London, WC1E 6BT, UK*

is associated with experimental feasibility constraints. The majority of the current applications of BO within materials discovery and development do not incorporate feasibility.[17,20,21] Nevertheless, BO recommendations can present a myriad of practical concerns in laboratory experiments, such as impossible combinations of material compositions, incompatible processing parameters, and apparatus limitations. Shrinking the boundaries of variables to a practical region could be a direct solution at the cost of reduced search space. In special cases in which known constraints on the input variables are available (*e.g.*, as inequality equations), optimization can be performed subject to these constraints.[22,23] For example, Li *et al.* nested an active learning loop for constraint modelling to restrict the candidate space selectable by BO.[24] Low *et al.* suggested evolution-guided Bayesian optimization, which imposes known constraints on multi-objective optimization problems, for nanoparticle synthesis in microfluidics.[25] However, these strategies become impossible when the constraints are unknown *a priori* and must be evaluated through laboratory experiments.

Several prior works on constrained or composite BO have explored applications in hyperparameter tuning. In the area of constrained BO, Gramacy and Lee proposed weighting the expected improvement (EI) acquisition function with a modelled probability to enforce a preference for feasible candidates.[26] Gardner *et al.* extended this approach to inequality constraints, assuming that feasibility could be derived from a continuous-value constraint function.[27] More recently, Tian *et al.* proposed a boundary exploration method that relaxes the acquisition function weights to encourage exploration near the constraint boundaries.[28] In the area of composite BO, Uhrenholt and Jensen investigated target value optimization, specifically minimizing a 2-norm, by warping the GP to a noncentral chi-squared distribution.[29] As an improvement, Astudillo and Frazier approached a more general problem of composite BO for any arbitrary composite function over the objective function. They transformed the Gaussian posterior in the acquisition function directly with the composited function.[30] Although these strategies have been rigorously tested on synthetic benchmarks and hyperparameter optimization tasks, they have yet to be integrated into a combinatorial framework to facilitate guided laboratory experiments.

Here, we implement a constrained composite Bayesian optimization (CCBO) pipeline showcasing efficient identification of suitable processing parameters in the rational synthesis of polymeric particles. Through introducing a variational inference GP component, the black-box experiment feasibility was modeled and incorporated into the BO acquisition function. Composite BO handles the modeling of experimental parameters and targeting of particle size through a composite objective function. Amongst the various particle fabrication techniques, electrospraying was selected as the model technique based on its simplicity, versatility, and precision as a popular manufacturing method in drug delivery research.[31] This technique utilizes electric fields to deform the meniscus of a polymer solution to form fine jets, which eventually disintegrate into fine droplets. As these droplets travel towards a collector, they further shrink and solidify due to solvent evaporation. Various parameters in the electrospraying process, such as the flow rate, voltage, polymer concentration, and solvent can be adjusted to tailor the product characteristics, although the intertwined impact of these factors could lead to prolonged, if not infeasible, trial-and-error experiments.[32] We demonstrate the superior performance of CCBO in target parameter optimization compared to random baseline and conventional BO strategies through both synthetic data and wet-lab experiments for poly(lactic-*co*-glycolic acid) (PLGA) particle synthesis at multiple size targets.

# Methods

### Materials

PLGA (PURASORB PDLG 5004A, 50 : 50 ratio) was purchased from Corbion (Amsterdam, The Netherlands). Chloroform and *N*,*N*-dimethylacetamide (DMAc) were purchased from Sigma-Aldrich (Gillingham, UK).

### Electrospray production of particles

PLGA solutions were prepared by mixing PLGA granules with solvents at ambient temperature with magnetic stirring overnight. The solutions were fed to a 22-gauge needle (outer diameter 0.71 mm) through a capillary using a syringe pump (Harvard PHD Ultra, Edenbridge, UK). The positive output of a high-voltage power supply (Glassman High Voltage Inc., NJ, United States) was connected to the needle *via* a crocodile clamp, and the collection plate was connected to the ground. Prior to electrospraying, the flow rate and voltage were adjusted to the values recommended by BO. Experiments were conducted at atmospheric pressure. The temperature and humidity in the room were controlled to be 19–22 °C and 40–50%.

### Characterization of particles

Particles were collected on a glass slide placed on a collection plate for scanning electron microscopy (SEM) analysis. A Zeiss Gemini 360 SEM (Germany) instrument was used at an acceleration voltage of 1.0 kV with an SE2 detector. For each sample, three images were taken randomly at different locations. The images were further analyzed using ImageJ (National Institute of Health, USA). To obtain the mean particle size, diameters were randomly measured for 100 particles. For infeasible experiments, the diameters of the splashes from undried droplets on the collecting glass slides were recorded as a measurement of size.

### Constrained composite Bayesian optimization

Two components were incorporated in the BO pipeline and developed under the BoTorch[33] and GPyTorch[34] frameworks. The objective component, which tracked the distance (or particle size in the case of CCBO), followed the classical design of BO (see SI Note 1 for details of handling categorical inputs).[15] Notably, due to the difficulty in determining the noise level in the experiments, we assumed the input data from laboratory experiments, after averaging over triplicates, to be noiseless. In

terms of the acquisition function, the thoroughly investigated strategy of $q$-expected improvement ($q$EI, or batch EI) was selected to allow consideration of multiple candidates jointly in each iteration.[18] In its simplistic form in which $q$ equals 1, the EI acquisition function at a single point $x_0$ can be given by $\alpha_{\mathrm{EI}}(x_0) = \mathbb{E}[\max(y_o - f^*, 0)]$, where $y_o \sim \mathcal{N}(\mu(x_0), \sigma^2(x_0))$ and $\mu(x_0)$ and $\sigma^2(x_0)$ are the posterior mean and variance from the Gaussian process at $x_0$, and $f^*$ is the current best observation. As the calculation of expectation requires integrating over the posterior, it becomes analytically intractable under a batched scenario where $q > 1$. We followed the strategy in BoTorch in which Monte–Carlo sampling was used to approximate the expectation as:

$$\alpha_{\mathrm{qEI}}(\mathbf{X}) \approx \frac{1}{N}\sum_{i=1}^{N} \max_{j=1,\ \ldots,q}\left[\max\left(y_{o,ij} - g^*, 0\right)\right],\ \ y_{o,ij} \sim \mathbb{P}(\mathrm{GP}(\mathbf{X})|\mathfrak{D}) \tag{1}$$

where $N$ is the total number of Monte–Carlo samples, $q$ is the number of candidates to be evaluated in parallel, $y_{o,ij}$ is sampled through the reparameterization trick from the Gaussian process conditioned on data $\mathfrak{D}$, and $g^*$ represents the current closest distance (with respect to the target) achieved. Notably, the data $\mathfrak{D}$ consisted of $\{(x_i, y_{o,i})\}_{i=1}^{n}$ where $y_{o,i} = g(s_i) = -(s_i - s_o)^2$ with $s_o$ representing the target value (predefined as a constant). Under such configurations, this vanilla BO pipeline could help to identify suitable experiment variables $X$ that could maximize the negative distance measure $y_o$.

Furthermore, the feasibility component was introduced to learn the black-box constraints in the experiments. Here, a variational Gaussian process was implemented for the binary classification of experimental success or failure.[34] The details for variational inference in Gaussian classification have been described in previous publications.[35] Briefly, the latent Gaussian process is further warped with a Probit regression to limit the output to between 0 and 1 for the purpose of approximating a Bernoulli posterior. For our latent Gaussian process, it followed the same constant mean prior and kernel functions to incorporate mixed inputs. To incorporate feasibility modelling in the Bayesian optimization process, we followed the strategy proposed earlier[26] to extract the posterior probability as a scaling factor in the acquisition function: $\alpha_{\mathrm{qEIcon}}(\mathbf{X}) = \mathbb{P}(y_c = 1|\mathbf{X})^*\alpha_{\mathrm{qEI}}(\mathbf{X})$.Incorporating this factor in the acquisition function allowed the suppression of the values of experiments that are potentially infeasible, creating our constrained BO pipeline.

In both the vanilla and constrained BO pipelines, the Gaussian process modeled $y_o$ and attempted to maximize this negative distance. As a different strategy, the composite BO used a Gaussian process to directly model the particle size $s$. The composite part, namely, the negative squared distance function $g$, was separated from the input data. Instead, the distance function was directly applied to the Gaussian posterior in the acquisition function:[36]

$$\alpha_{\mathrm{qEICF}}(\mathbf{X}) \approx \frac{1}{N}\sum_{i=1}^{N} \max_{j=1,\ \ldots,q}\left[\max\left(g\left(s_{ij}\right) - g^*, 0\right)\right],\ \ s_{ij} \sim \mathbb{P}\left(\mathrm{GP}(\mathbf{X})|\mathfrak{D}'\right) \tag{2}$$

where $s_{ij}$ was sampled through the reparameterization trick and $\mathfrak{D}' = \{(x_i, s_i)\}_{i=1}^{n}$. By coupling the composite acquisition function $\alpha_{\mathrm{qEICF}}$ with the constraint probability, we obtain the acquisition function for CCBO: $\alpha_{\mathrm{qEICFcon}}(\mathbf{X}) = \mathbb{P}(y_c = 1|\mathbf{X})^*\alpha_{\mathrm{qEICF}}(\mathbf{X})$.

In the present work, the Monte–Carlo sampling number $N$ was 512 and $q$ was fixed to 2 throughout all BO pipelines. All inputs $\mathbf{X}$ were normalized to unit cubes, and the flow rate variable was transformed to a logarithm before normalization. The outcomes of the objective component, including the distance variable $y_o$ in vanilla BO and constrained BO, as well as the particle size variable $s$ in CCBO, were standardized to zero mean and unit variance. The outcomes of the feasibility component $y_c$ were rescaled to $\{-1,1\}$.

### Synthetic electrospray data generation

The synthetic electrospray data was generated through the following functions:

$$s = 2 \times \frac{\sqrt{QC}}{\log(U)} + \alpha + 0.4 \tag{3}$$

$$y_c = \begin{cases} 1, & \text{if } \log(Q) \times (\alpha - 0.5) + 1.4 > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

where $s$ is the particle size (μm), $Q$ is the flow rate (μL min$^{-1}$), $c$ is the concentration of the polymer solution (% w/v), and $U$ is the applied voltage (kV). The parameter $\alpha$ is a constant that depends on the solvent ($CHCl_3$: 1, DMAc: 0).

### Validating BO with synthetic data

The target particle size $s_o$ was arbitrarily set to 0.6, 3.0, 6.0 or 18.0 μm to validate BO performance. In each run, three BO

Table 1 Boundaries of experiment variables for BO and the starting data for synthetic experiments

| Label | Polymer concentration (% w/v) | Flow rate (μL min$^{-1}$) | Voltage (kV) | Solvent |
|---|---|---|---|---|
| Bounds | [0.05–5.00] | [0.01–60.00] | [10.0–18.0] | {$CHCl_3$, DMAc} |
| S-1 | 0.50 | 15.00 | 10.0 | DMAc |
| S-2 | 0.50 | 0.10 | 10.0 | $CHCl_3$ |
| S-3 | 3.00 | 20.00 | 15.0 | DMAc |
| S-4 | 1.00 | 20.00 | 10.0 | $CHCl_3$ |
| S-5 | 0.20 | 0.02 | 10.0 | $CHCl_3$ |

pipelines and the random baseline were performed for 10 iterations with the starting data listed in Table 1. The outcomes of experiments were calculated using synthetic eqn (3) and (4) from the corresponding experimental variables. Each run was repeated 20 times to account for variations. The comparison between human and BO followed settings similar to those used previously. The starting data (Table 1) were first shown to the participants ($N = 14$), who had varying experience with electrospraying, including advanced users with more than 3 years of experience ($N = 4$), intermediate users with 1 to 3 years of experience ($N = 4$), and beginners with less than 1 year of experience ($N = 6$). In each iteration, two experiments were recommended by participants to optimize towards a 3.0 μm target, and the experimental results calculated using the synthetic equations were then revealed. In total, five iterations were performed for the human *vs.* BO campaign considering that CCBO achieved significant reduction within a few rounds. During the campaign, the participants were not allowed to access each other's results except for the five initial data provided. They were asked to work out the recommendations as if they were dealing with a new electrospray setup. No strict time constraints were imposed on the participants to provide answers, but a typical time of 2 minutes was observed for participants to suggest experiments for one iteration.

The regret, defined as the closest distance to the target particle size, was plotted at each iteration. The experimental variables proposed in a typical run were visualized on 3D plots with symbols representing solvent and feasibility, and colors encoding the iteration. The area under the curve (AUC) for each strategy and human participant was calculated based on trapezoid rules for quantitative comparison. One-tailed Mann–Whitey *U*-tests were performed with the alternative hypothesis being that CCBO had smaller AUC/regret compared to the BO baseline or human groups, respectively.

### Guiding laboratory experiments with CCBO

The boundaries of the experimental variables remained the same as for the validation with synthetic data. The starting eight experiments were generated through a Sobol sequence within the boundaries for each variable. The targeted particle sizes were 300 nm and 3.0 μm based on domain expertise in drug delivery. The two experiments in each iteration were performed in triplicate. The results were fed back into the BO pipeline to obtain the next recommendations. The stopping criterion was set as achieving ±10% the target size.

## Results

### Validating CCBO through synthetic data

The performance of CCBO was first validated with synthetic experimental data. Before introducing the benchmark results, the three configurations of the BO pipelines tested in this study are presented (Fig. 1a). More details of the implementation can be found in the Methods section. Briefly, the vanilla BO pipeline followed a traditional BO design in which the target to be maximized was the negative squared distance $y_o$. The feasibility component, which leveraged a variational GP for classification, was added to track experimental feasibility. Through factoring a probability term into the acquisition function, the constrained BO pipeline was able to pick candidates with a higher chance of success. CCBO adopted the same feasibility modelling whilst modifying the objective component. It utilized GP to model the fundamental relationship between the processing variables $x$ with the size $s$ in the experiments. The negative squared distance function was incorporated in the acquisition function to prioritize candidates for minimizing the distance to the preset target. In terms of the synthetic problem, the data was produced using equations that simulated electrospray processing. Specifically, the function for determining the size of the electrosprayed particles (see eqn (3)) was inspired by scaling laws proposed for electrospray and experimental observations, in which the flow rate and polymer concentration (through affecting the viscosity) are both positively correlated to the diameter, with voltage having a negative impact.[32,37] The logarithm and power transformations in the function were intended to add complexity to the modeling process to simulate the nonlinear nature of the electrospraying process. The constant alpha was added to account for the impact of the solvents considered in the process. In addition, the feasibility zone, as visualized in Fig. 1b, was set to be highly related to the flow rate and the solvent. This rationale was based on practical considerations, as chloroform, a highly volatile solvent, would result in a clogged nozzle at lower flow rates, while higher flow rates would lead to insufficient evaporation of the solvent DMAc and produce splashes of droplets on the collector instead of solid particles.

As a benchmark, CCBO, together with random baseline, vanilla BO, and constrained BO only, was performed for 10 iterations. Five initial experiments were included, accounting for successful and failed cases for both solvents. The optimization target was set to 18 μm. Results for other target sizes, including 0.6, 3 and 6 μm, can be found in SI Fig. 1. In each iteration, two sets of processing parameters were proposed and subjected to simulation functions to retrieve the synthetic experimental result as well as the feasibility. The regret, defined as the difference between the target and the closest candidate, was recorded after each iteration as a measurement of performance (Fig. 1c). After 10 iterations, the random baseline reached 0.8 μm regret. Similarly, the vanilla BO and constrained BO both achieved around 0.4 μm regret. In contrast, the CCBO algorithm rapidly converged to the targeted diameter after only two iterations with minimal regret. Moreover, the AUC of each strategy was calculated using a trapezoidal method to quantify the optimization efficiency. CCBO achieved a minimal AUC of $2.47 \pm 0.85$, which was significantly lower than that of the random method ($19.48 \pm 8.12$, $p < 0.0001$), vanilla BO ($18.35 \pm 3.86$, $p < 0.0001$) and constrained BO ($16.26 \pm 3.73$, $p < 0.0001$) under the one-tailed Mann–Whitney *U*-test.

The benchmark for synthetic electrospray was extended to compare CCBO with state-of-the-art optimization methods such as Summit,[38] Dragonfly,[39] EDBO+,[40] and Atlas.[41] Notably, the implementations in Summit and Dragonfly did not support optimization under unknown constraints. Therefore, their

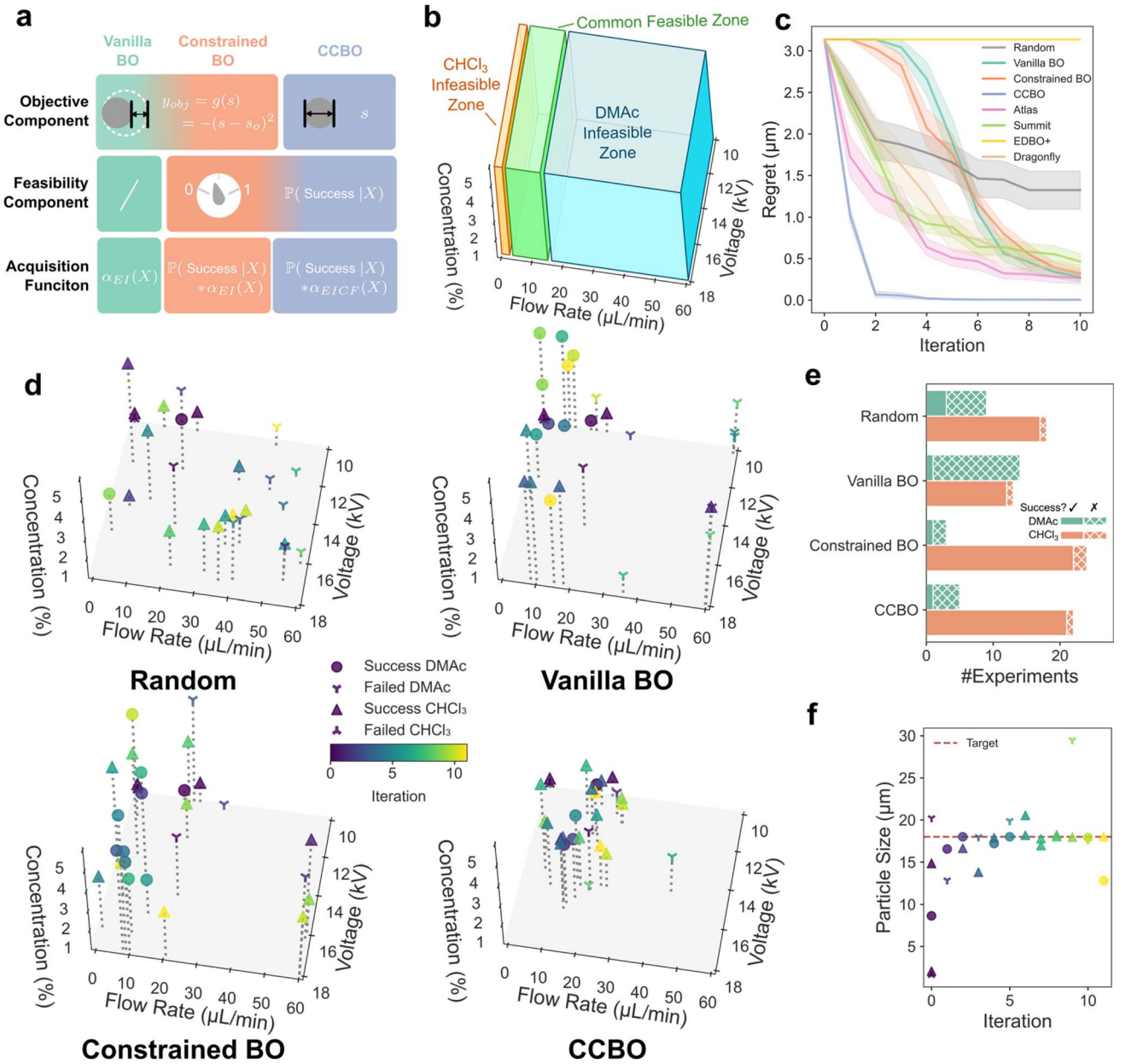



Fig. 1 Results for CCBO validation with synthetic data. (a) An illustration of the configurations for vanilla BO, constrained BO, and CCBO. (b) Parameter space visualization for the synthetic data with the feasibility zone highlighted for each solvent. (c) Benchmark results of target value optimization with random baseline, vanilla BO, constrained BO, Atlas (constrained BO), Summit (single objective BO), EDBO+ (pool-based BO), Dragonfly (bandit BO), and CCBO. The regret is calculated as the closest distance with respect to the design target achieved at different iterations of BO. Each benchmark experiment was performed for 10 iterations. Shaded areas indicate standard error from 20 repetitions. (d) Visualization of the suggested experimental parameters. Each data point represents one synthetic experiment. The corresponding iterations are coded by color. Symbols represent the solvent used and the feasibility of the experiment. (e) Comparison of total number of successful (filled bars) and failed experiments (hatched bars) in a typical run of 10 iterations with the four strategies. (f) Particle sizes produced using the parameters chosen by CCBO. The data points are color-coded by iteration and the symbols represent the solvent and feasibility. The target (18 μm) is highlighted as a dashed line.

performance was similar to that of the vanilla BO baseline. EDBO+ was developed as a pool-based active learning optimization platform. The EDBO+ algorithm did not provide improvement in regret, potentially due to its lack of support for constrained optimization and the limitation of a pool-based search space compared to other strategies. Finally, the most recently developed approach, Atlas, which is a framework library for self-driving libraries by Hickman *et al.*, utilized a variational GP to model unknown constraints for experimental feasibility.[42] The optimization by Altas with *a priori* unknown constraints showed better performance than other existing strategies in the benchmark. However, as none of the

state-of-the-art libraries support target-value optimization natively, none of these strategies outperformed the CCBO algorithm proposed herep for electrospray optimization. More detailed results can be found in SI Fig. 1. These results highlight the importance of incorporating both a constrained and composite optimization scheme for target-driven design problems in experiments.

To understand the recommendation process, the proposed experiments were visualized in Fig. 1d. The random baseline sampled uniformly across the experiment space with both solvents, resulting in many failed DMAc experiments due to the flow rate feasibility constraints. Vanilla BO started exploring the boundary conditions in the first few rounds. With an additional model to account for feasibility, the constrained BO algorithm managed to learn the feasible region for DMAc, as reflected by most DMAc experiments being recommended with lower flow rates. This corresponded well to the initial feasible zone visualized in Fig. 1b. In addition, the number of failed and successful attempts of each algorithm from the results were plotted in Fig. 1e, highlighting the reduction in infeasible experimental conditions with the help of the additional constraint model.

Furthermore, the CCBO strategy was observed to show highly efficient searching in a localized experiment space (Fig. 1d). The performance of CCBO could be explained by its design. The routes taken by vanilla BO and constrained BO were directly minimizing the distance, whereas the surrogate GP was forced to model more complicated results from both the experiment and the superimposed distance function. On the contrary, GP was solely used for modeling the black-box experiment results for CCBO. Our observations with CCBO echoed the findings in the composite BO literature: extracting the analytically trackable part from the black-box function can drastically benefit the optimization efficiency.[30] In standard BO, the EI acquisition function assumes a Gaussian posterior distribution. However, the posterior of the composite function becomes non-Gaussian after transformation with a non-linear function. To address this, Astudillo and Frazier suggested leaving the GP to model the black-box function. The composite part was instead incorporated into the acquisition function to transform the Gaussian posterior of the black-box function. This allows more efficient optimization through a closer approximation of posterior distribution in a composite scenario.[36] In our implementation, the composite acquisition function was optimized in the CCBO pipeline with Monte Carlo sampling. Through the benchmark validation, we have shown that vanilla BO or constrained BO alone would not be able to efficiently optimize our design problem, highlighting the importance of the integration of CCBO.

Finally, we compared CCBO to human electrospray users with varying levels of expertise in this synthetic campaign. More experienced users were believed to approach the target more

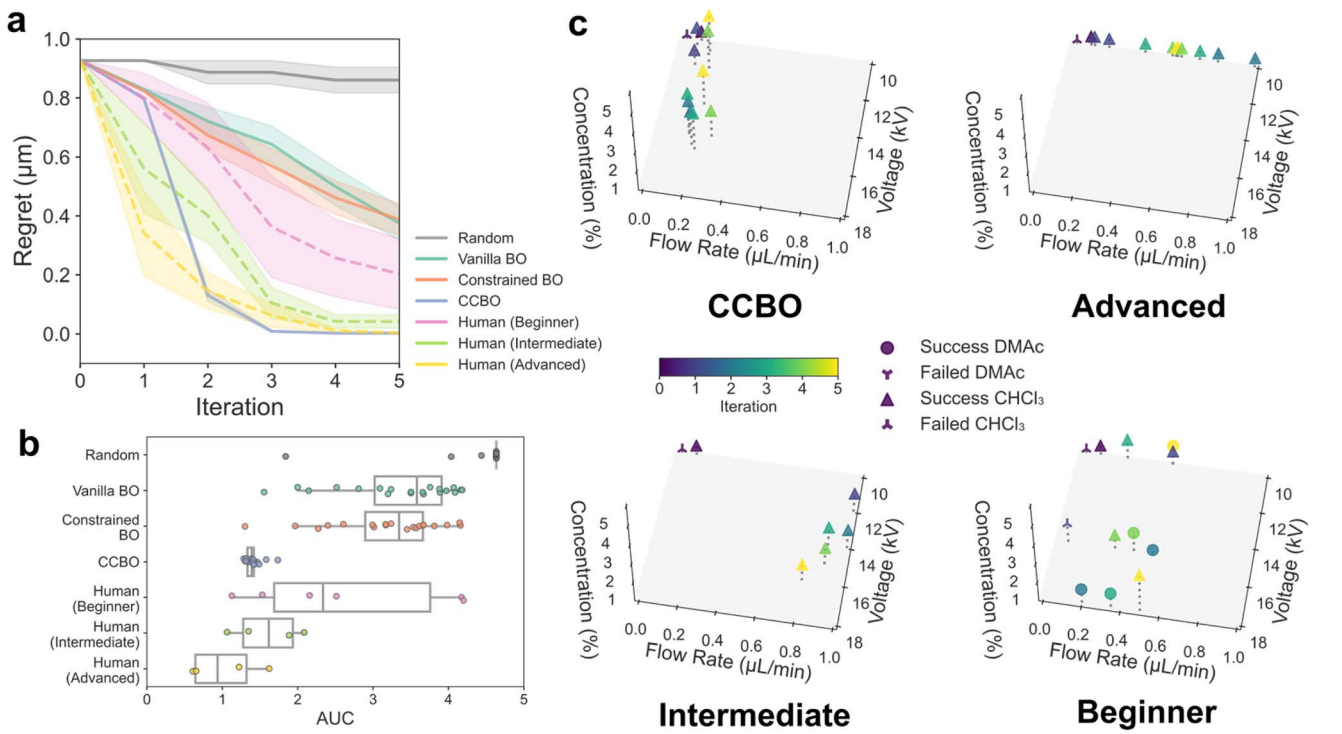



Fig. 2 Comparing human and BO performance with synthetic data. (a) Benchmark results with BO pipelines (solid lines) for 3.0 μm target in comparison with human users (dashed lines) on synthetic data. The experiment was performed for 5 iterations. Shaded areas for random, vanilla BO, constrained BO, and CCBO indicate standard error from 20 repetitions. Shaded areas for human performance, including beginner ($N = 6$), intermediate ($N = 4$), and advanced ($N = 4$) groups, represent the standard error for the respective participants. (b) Box plot with scatters of the area-under-the-curve (AUC) as calculated using the trapezoid rule of the benchmark results from each strategy and human group. (c) Visualization of experiments selected by CCBO and human participants with various experience levels. Each data point represents one 'synthetic' experiment. The iteration is coded by color. Symbols represent the solvent used and feasibility of the experiment.

Table 2 Processing parameters generated through a Sobol sequence and the resulting particle sizes and feasibility ($N = 3$)

| Label | Polymer concentration (% w/v) | Flow rate (μL min$^{-1}$) | Voltage (kV) | Solvent | Mean size (μm) | Feasible? |
|---|---|---|---|---|---|---|
| 0-1 | 2.40 | 1.73 | 14.0 | DMAc | 0.56 | 1 |
| 0-2 | 4.06 | 0.44 | 15.7 | $CHCl_3$ | 1.00 | 0 |
| 0-3 | 2.88 | 49.11 | 11.8 | DMAc | 15.00 | 0 |
| 0-4 | 0.76 | 0.01 | 17.6 | $CHCl_3$ | 1.20 | 0 |
| 0-5 | 0.11 | 10.43 | 14.5 | $CHCl_3$ | 6.26 | 1 |
| 0-6 | 3.55 | 0.06 | 12.8 | DMAc | 0.15 | 1 |
| 0-7 | 4.55 | 2.39 | 16.7 | $CHCl_3$ | 5.24 | 1 |
| 0-8 | 1.88 | 0.21 | 11.0 | DMAc | 1.12 | 1 |

efficiently, as they were equipped with prior knowledge of the influence of the parameters and the selection of solvent. All participants ($N = 14$) evaluated the same initial experimental data and suggested experiments to achieve a target particle size of 3 μm. The comparative results are plotted in Fig. 2a. More detailed human performance results are available in SI Table 1 and Fig. 2. In the first iteration, the CCBO strategy was behind intermediate (1–3 years of experience, $N = 4$) and advanced users (≥3 years of experience, $N = 4$), and performed similarly to beginners (<1 year of experience, $N = 6$). However, CCBO soon overtook intermediate users from the second iteration onwards and surpassed advanced users on later iterations. Quantitatively, the AUC was calculated and plotted (Fig. 2b) with respect to each strategy or human group, and CCBO (1.40 ± 0.10) exhibited a significantly smaller ($p = 0.01$) AUC than beginners (2.62 ± 1.19) under the one-tailed Mann–Whitney $U$-test. There were no significant reductions in AUC for CCBO compared to intermediate (1.60 ± 0.41, $p = 0.34$) or advanced (1.03 ± 0.42, $p = 0.95$) users. When focusing on overall performance (regret at final iteration), the regret of the CCBO strategy was significantly lower than that of intermediate ($p = 0.02$) and beginner ($p < 0.0001$) users. Further analysis of parameter selection strategies revealed that advanced users predominantly followed a one-factor-at-a-time (OFAT) approach, resulting in linear adjustment patterns (Fig. 2c). Most beginner users and intermediate users attempted to adjust multiple parameters simultaneously. Unlike human participants, CCBO employed more strategic exploration and exploitation, effectively reducing experimental regret by targeting promising regions in the parameter space. Taken together, these findings demonstrated that CCBO could achieve performance comparable to highly experienced participants and navigate complex experimental spaces more effectively than human users. In addition, the performance differences among users with various levels of expertise reflected the successful development of the synthetic problem simulating electrospraying, consolidating our confidence in proceeding to laboratory validation.

## Guiding laboratory electrospraying with CCBO for targeted particle production

Following the validation of CCBO with synthetic data, it was applied to guide real-world experiments for the electrospraying production of micro- and nanoparticles. The initial experiments, which were generated through a Sobol sequence, were performed to accumulate starting data for BO pipelines (Table 2).

Two particle sizes, 300 nm and 3.0 μm, were set as the design targets based on pharmaceutical interest as drug carriers for intravenous injection and pulmonary delivery.[4] Based on previous reports, the production of PLGA particles with these two particle sizes require distinct processing parameters involving different solvents and flow rates.[32,43] Thus, the setting of these targets could simulate distinct experimental scenarios to challenge BO pipelines. The workflow of targeted particle production under CCBO guidance is illustrated in Fig. 3a. With the initial data gathered, a CCBO pipeline was implemented to propose two experiments in parallel for laboratory investigation. The selection of two experiments was based on the capacity for laboratory work and to avoid wasting materials and preparation time. After collecting samples and characterization, the results from triplicate experiments were evaluated and compared with the target. The next iteration of BO was performed based on the addition of the new data.

The parameters proposed by CCBO are visualized using heatmaps in Fig. 3b. The heatmap of the initial experiments reflects the selection of diverse parameters in the Sobol sequence. In total, three iterations of BO were performed for the target of 300 nm and four iterations for the 3.0 μm target. The selection of solvent was the most obvious difference for these two targets. Indeed, in previous reports of PLGA particle synthesis, DMAc is a popular solvent due to its high boiling point.[44] From a mechanistic viewpoint, droplets will experience fission due to the competition between coulombic repulsion and liquid surface tension in an electrospraying process.[45] At the same time, the evaporation of solvents increases the concentration and viscosity of the droplet. As a non-volatile solvent, DMAc allows this fission process to fully develop and thus generates sub-micrometer particles.[32] Chloroform, on the contrary, is preferred in the literature to produce larger particles within the tens of micrometers range.[46] These practical considerations, which are normally accumulated through experience and trial-and-error, were also picked up by the BO pipeline. The recommendations provided by CCBO clearly showed a trend of adopting DMAc for the 300 nm target and chloroform for the 3.0 μm target.

Linking the recommendations to the experimental results (Fig. 3c) could provide a more holistic viewpoint of the selection

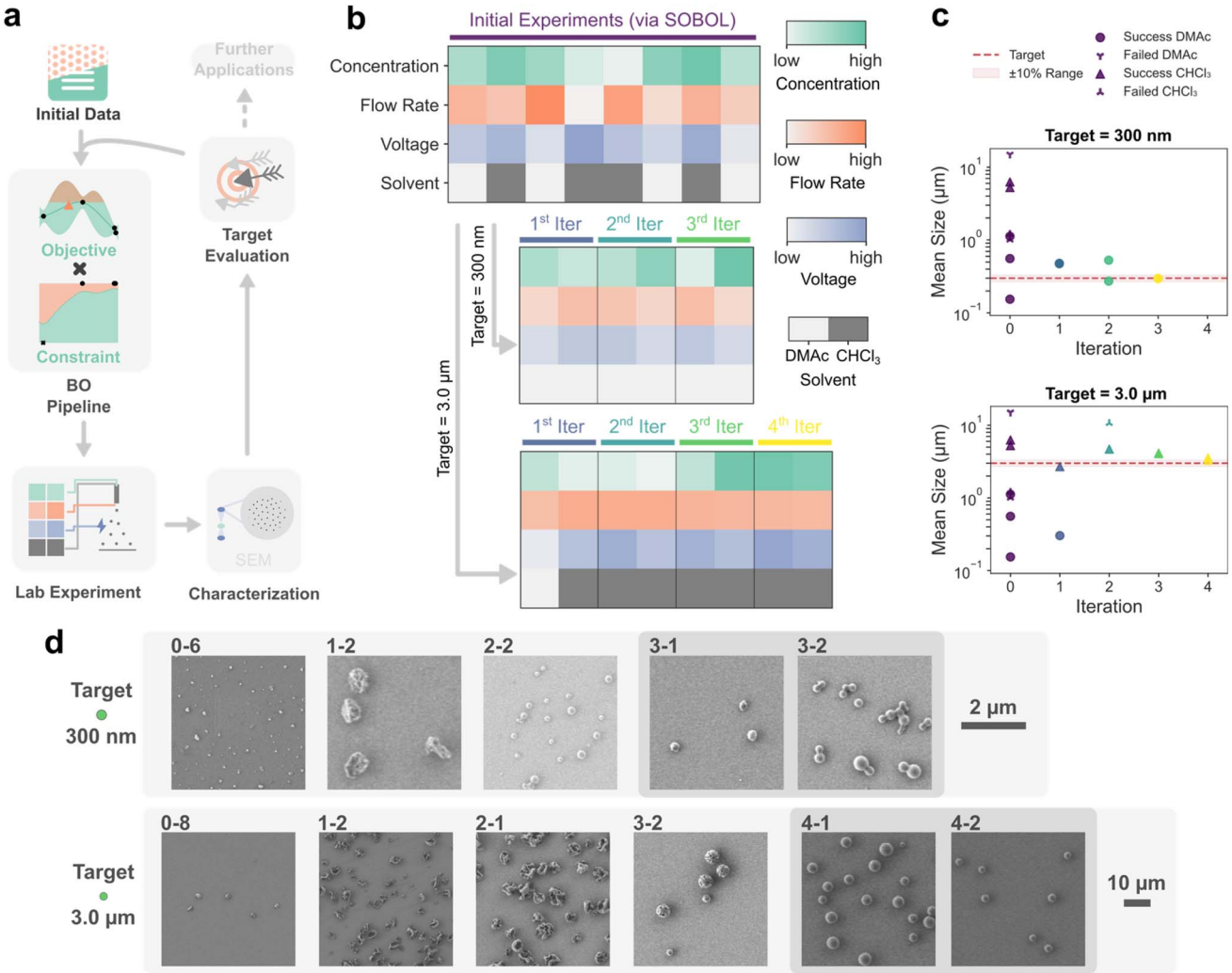



Fig. 3 Guiding electrospray experiments with CCBO. (a) Schematic diagram representing the experiment process with the integration of CCBO. (b) Heatmaps visualizing the processing parameters used for the (top) initial experiments, (middle) 300 nm target, and (bottom) 3.0 µm target. The initial experiments were generated with a Sobol sequence and the targeted experiment series were suggested by the CCBO pipeline. (c) Experimental results of particles generated *via* electrospraying under the parameters proposed for the (top) 300 nm and (bottom) 3.0 µm target. Each data point represents the mean of triplicate laboratory experiments. Symbols represent the solvent used and the feasibility of the experiment. (d) SEM images of particles produced at different iterations for the (top) 300 nm and (bottom) 3.0 µm target.

strategy of CCBO. For the 300 nm target, the best candidate in the initial experiments (0-8 on Table 2) used DMAc with a low polymer concentration, flow rate and voltage to obtain 0.15 µm particles. The recommendations from the CCBO pipeline showed exploration of higher concentrations and fine-tuning of the flow rate parameter (SI Table 2). Interestingly, the 3-1 and 3-2 experiments both achieved a 300 nm particle size with distinct processing parameters, suggesting that the impact of the less-concentrated polymer solution was compensated by the higher flow rate used for 3-1. Furthermore, the balance of exploration–exploitation from the EI acquisition function was further demonstrated through the experiment series for the 3.0 µm target. In the first iteration, CCBO attempted the use of both DMAc and chloroform as the solvent (SI Table 3). The second iteration tested the lowest polymer concentration (0.05% w/v), which is shown as the lightest green in the heatmap (Fig. 3b). Finally, the recommendations settled at higher concentrations with reduced flow rates to approach the target based on the fine-tuning from exploitation. It was also observed from the SEM images (Fig. 3d) that experiment 1-2 for the 3.0 µm target managed to produce 2.69 µm particles with rough and polydisperse characteristics using a low polymer concentration (0.36% w/v) sprayed at a high flow rate of 3.65 µL $min^{-1}$. The final experiment 4-2 suggested a 4.02% w/v solution sprayed at 1.08 µL $min^{-1}$ (SI Table 3) to obtain 3.29 µm diameter particles. This result again highlighted the ability to achieve similar particle size through balancing polymer concentration and flow rate, together with adjusting other parameters. The SEM images of the final iteration experiments show satisfactory particle production at the targeted sizes.

Overall, we have verified the performance of CCBO in the automatic identification of the experiment feasibility region

and rapid convergence to design targets through synthetic data validation. Comparison with human experts demonstrated the competitive performance of CCBO. The rational exploration of the experiment space outperformed the instinct-driven OFAT trial-and-error approach of humans. As a further step, wet-lab experiments consolidated the potential of CCBO in real-world applications for guided particle synthesis within a few iterations.

## Discussion

The present work demonstrated the application of an efficient CCBO pipeline for target value optimization under black-box constraints. The two components in CCBO worked cohesively to address the need for guiding particle synthesis. For target optimization, the composite BO demonstrated strong capacity in modeling under the composited distance function over the underlying black-boxed electrospray relationship function. On the other hand, the constraint compartment managed to learn and regulate the suggested experiments using a variational Gaussian process. To deal with unknown feasibility boundaries, many current strategies choose to apply active learning for the identification of unknown feasibility regions, followed by running BO pipelines under the established boundaries.[47,48] As an improvement, CCBO was designed to integrate these two individual processes and focus on identifying the feasibility regions around the design target. This can be seen from the initial experiments in which the infeasibility caused by the mismatch of high flow rate with the less-volatile solvent DMAc (experiment 0-3 on SI Table 2) was not further explored, because the target only requires experiments in the lower-flow-rate region. In comparison, with an active learning pipeline, extra experiments would be needed to determine the possible range for DMAc. Thus, the design of the CCBO pipeline allows an efficient reduction in the number of experiments to save laboratory resources.

In addition, the innate exploration–exploitation trade-off of BO made possible the identification of multiple possible experimental parameters that can achieve the same design target. This is especially helpful when other design considerations coexist. For example, in the validation with the synthetic problem (Fig. 1f), CCBO attempted to use both DMAc and chloroform and paired them with a wide range of other processing parameters to hit the design target in iterations 6 to 10. From the perspective of production rate, a higher flow rate and polymer concentration might be preferred. Similarly, if the sustainability of the solvent is considered, DMAc would be selected over chloroform as a less harsh solvent. Besides the synthetic data, laboratory experiments also managed to find multiple parameters to produce particles with 300 nm or 3.0 μm diameter. These particles exhibited distinctive morphology and polydispersity, demonstrating varying characteristics for their applications. Although not explicitly coded as a multiple-objective optimization problem, these sets of experimental parameters could be presented to the user as alternative choices. In practice, such flexibility allows the researcher to consider product properties, manufacturing metrics, or other aspects in production without changing the main design target.

Since only two solvents have been investigated in the present work, categorial representations of the solvent variable were used instead of applying molecular featurization. Many modern BO libraries designed for chemistry and materials research support molecular featurization, such as Atlas and GAUCHE.[41,49] Featurizing molecules with their physicochemical properties could incorporate chemistry knowledge in the optimization process and benefit molecular structure optimization and discovery tasks. For example, Griffiths *et al.* managed to leverage BO to optimize molecular design in a latent space generated from variational autoencoders.[50] Although optimizing the solvent molecule *per se* was not necessarily a focus in particle synthesis applications, leveraging molecular fingerprints to represent solvents would equip the optimization process with chemically meaningful knowledge (*via* representing similar solvents with close descriptors).[51] In addition, extending the present single-objective optimization paradigm to multiple objectives could benefit more complicated particle design tasks, including the control of both particle size and size distributions, or morphological features. Our implementation of constrained optimization was through feasibility-weighting of the acquisition function. This should be extensible to multi-objective optimization seamlessly, considering that the feasibility modelling is irrelevant to the type of acquisition functions. Notably, Li *et al.* recently proposed a new method to balance (unknown) constraint modelling and multi-objective optimization through unifying constraint violation with hypervolume regret.[52] They demonstrated improved efficiency compared to baseline scalarization-based methods such as *q*ParEGO.[53] On the other hand, composite optimization is expected to be transferable to multi-objective optimization scenarios, in which objectives are scalarized. However, the implementation of these extensions is beyond the scope of the current manuscript and thus left as a future direction.

Finally, we highlight that CCBO could potentially be extended to other particle synthesis systems, such as batch methods and microfluidics, to facilitate the guided design and production of particles. In the past, the resource-demanding nature of experimentation and scarcity of data have posed significant challenges and prolonged the workflow of particle synthesis. We expect CCBO to empower nanotechnology with a smarter and more efficient paradigm for target-driven design.

## Conclusions

Achieving the rational synthesis of nanoparticles often relies on extensive domain expertise and trial-and-error experimentation to navigate within the feasibility space with target product specifications. In the present work, we introduced CCBO as a unified framework to address constraint-aware and target-value optimization in nanoparticle production. It was evaluated in a synthetic electrospray problem and presented superior performance compared to baseline and state-of-the-art strategies. Benchmarking against human electrospray users further demonstrated that CCBO could match experts with at least

three years of experience. Laboratory experiments validated its practical ability to guide the electrospray synthesis of PLGA particles with biomedically meaningful target diameters of 300 nm and 3.0 μm. These findings highlight CCBO as a powerful and efficient strategy for materials development tasks characterized by complex, black-box constraints and precise design objectives, thus contributing to the next generation of AI-driven nanomanufacturing.

## Author contributions

F. W. – conceptualization, data curation, methodology and software, validation, formal analysis, investigation, writing – original draft, writing – review & editing; M. P. – methodology, resources, supervision, writing – review & editing; A. H. – methodology, writing – review & editing; M. E. – methodology, resources, supervision, and writing – review & editing.

## Conflicts of interest

The authors declare no competing interests.

## Data availability

The code to implement CCBO presented in this study has been made available at **https://github.com/FrankWanger/CCBO.git**. The specific code and data for reproducing CCBO and the baselines is also archived on Zenodo: **https://doi.org/10.5281/zenodo.16614771**. The raw experimental data, including the human participant results, electrospraying details and raw SEM images, has been included in the SI.

Supplementary information is available. See DOI: **https://doi.org/10.1039/d5dd00243e**.

## Acknowledgements

The author Fanjin Wang would like to thank the Engineering and Physical Sciences Research Council (EPSRC) for supporting his PhD research (EP/R513143/1 and EP/W524335/1). Dr Jakob Zeitler from Matterhorn Studio is thanked for initial discussions. The participants of the human *vs.* BO campaign are gratefully acknowledged for their time and expertise.

## References

1 M. J. Mitchell, M. M. Billingsley, R. M. Haley, M. E. Wechsler, N. A. Peppas and R. Langer, Engineering precision nanoparticles for drug delivery, *Nat. Rev. Drug Discov.*, 2020, **20**(2), 101–124.
2 M. A. C. Stuart, W. T. S. Huck, J. Genzer, M. Müller, C. Ober, M. Stamm, *et al.*, Emerging applications of stimuli-responsive polymer materials, *Nat. Mater.*, 2010, **9**(2), 101–113.
3 S. Colombo, M. Beck-Broichsitter, J. P. Bøtker, M. Malmsten, J. Rantanen and A. Bohr, Transforming nanomedicine manufacturing toward Quality by Design and microfluidics, *Adv. Drug Deliv. Rev.*, 2018, **128**, 115–131.
4 W. Poon, B. R. Kingston, B. Ouyang, W. Ngo and W. C. W. Chan, A framework for designing delivery systems, *Nat. Nanotechnol.*, 2020, **15**(10), 819–829.
5 L. Rao, Y. Yuan, X. Shen, G. Yu and X. Chen, Designing nanotheranostics with machine learning, *Nat. Nanotechnol.*, 2024, **3**, 1–13.
6 H. Tao, T. Wu, M. Aldeghi, T. C. Wu, A. Aspuru-Guzik and E. Kumacheva, Nanoparticle synthesis assisted by machine learning, *Nat. Rev. Mater.*, 2021, **6**(8), 701–716.
7 J. Antony, *Design of experiments for engineers and scientists*, Elsevier, London, 2nd edn, 2014, p. 208.
8 D. C. Montgomery, *Design and analysis of experiments*, John Wiley & Sons, Inc., Hoboken, 9th edn, 2017, p. 1.
9 R. Batra, L. Song and R. Ramprasad, Emerging materials intelligence ecosystems propelled by machine learning, *Nat. Rev. Mater.*, 2020, **6**(8), 655–678.
10 S. Back, A. Aspuru-Guzik, M. Ceriotti, G. Gryn'ova, B. Grzybowski, G. Ho Gu, *et al.*, Accelerated chemical science with AI, *Digital Discovery*, 2024, **3**(1), 23–33.
11 P. Xu, X. Ji, M. Li and W. Lu, Small data machine learning in materials science, *npj Comput. Mater.*, 2023, **9**(1), 1–15.
12 S. Greenhill, S. Rana, S. Gupta, P. Vellanki and S. Venkatesh, Bayesian Optimization for Adaptive Experimental Design: A Review, *IEEE Access*, 2020, **8**, 13937–13948.
13 C. E. Rasmussen, *Gaussian processes for machine learning. 3. print*, MIT Press, Cambridge, Mass, 2006, p. 272.
14 Y. Wu, A. Walsh and A. M. Ganose, Race to the bottom: Bayesian optimisation for chemical problems, *Digital Discovery*, 2024, **3**(6), 1086–1100.
15 B. J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J. I. M. Alvarado, *et al.*, Bayesian reaction optimization as a tool for chemical synthesis, *Nature*, 2021, **590**(7844), 89–96.
16 N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, *et al.*, An autonomous laboratory for the accelerated synthesis of novel materials, *Nature*, 2023, **624**(7990), 86–91.
17 T. Lookman, P. V. Balachandran, D. Xue and R. Yuan, Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design, *npj Comput. Mater.*, 2019, **5**(1), 1–17.
18 D. R. Jones, M. Schonlau and W. J. Welch, Efficient Global Optimization of Expensive Black-Box Functions, *J. Global Optim.*, 1998, **13**(4), 455–492.
19 J. G. Hoffer, S. Ranftl and B. C. Geiger, Robust Bayesian target value optimization, *Comput. Ind. Eng.*, 2023, **180**, 109279.
20 O. Borkowski, M. Koch, A. Zettor, A. Pandi, A. C. Batista, P. Soudier, *et al.*, Large scale active-learning-guided exploration for in vitro protein production optimization, *Nat. Commun.*, 2020, **11**(1), 1872.
21 A. Ortiz-Perez, D. van Tilborg, R. van der Meel, F. Grisoni and L. Albertazzi, Machine learning-guided high throughput nanoparticle design, *Digital Discovery*, 2024, **3**(7), 1280–1291.
22 C. Antonio, Sequential model based optimization of partially defined functions under unknown constraints, *J. Global Optim.*, 2021, **79**(2), 281–303.

23 R. J. Hickman, M. Aldeghi, F. Häse and A. Aspuru-Guzik, Bayesian optimization with known experimental and design constraints for chemistry applications, *Digital Discovery*, 2022, **1**(5), 732–744.
24 G. Li, Y. Wang, S. Kar and X. Jin, Bayesian Optimization with Active Constraint Learning for Advanced Manufacturing Process Design, *IISE Trans.*, 2025, 1–15.
25 A. K. Y. Low, F. Mekki-Berrada, A. Gupta, A. Ostudin, J. Xie, E. Vissol-Gaudin, *et al.*, Evolution-guided Bayesian optimization for constrained multi-objective optimization in self-driving labs, *npj Comput. Mater.*, 2024, **10**(1), 104.
26 R. B. Gramacy and H. K. H. Lee, Optimization Under Unknown Constraints, *Bayesian Statistics 9*, ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger, Oxford University Press, 2011, DOI: **10.1093/acprof:oso/9780199694587.003.0008**.
27 J. R. Gardner, M. J. Kusner, Z. Xu, K. Q. Weinberger and J. P. Cunningham, Bayesian optimization with inequality constraints, in *Proceedings of the 31st International Conference on International Conference on Machine Learning*, Beijing, China, 2014, pp. 937–945.
28 Y. Tian, A. Zuniga, J. P. Dürholt, P. Das, J. Chen, W. Matusik and M. K. Luković, Boundary exploration for Bayesian optimization with unknown physical constraints, in *Proceedings of the 41st International Conference on Machine Learning (ICML'24)*, 2024, vol. 235, pp. 48295–48320, DOI: **10.48550/arXiv.2402.07692**.
29 A. K. Uhrenholt and B. S. Jensen, Efficient Bayesian Optimization for Target Vector Estimation, in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, PMLR, 2019, pp. 2661–2670, **https://proceedings.mlr.press/v89/uhrenholt19a.html**.
30 R. Astudillo and P. I. Frazier, Thinking Inside the Box: A Tutorial on Grey-Box Bayesian Optimization, in *2021 Winter Simulation Conference (WSC)*, IEEE, Phoenix, AZ, USA, 2021, pp. 1–15, **https://ieeexplore.ieee.org/document/9715343/**.
31 A. Ali, A. Zaman, E. Sayed, D. Evans, S. Morgan, C. Samwell, *et al.*, Electrohydrodynamic atomisation driven design and engineering of opportunistic particulate systems for applications in drug delivery, therapeutics and pharmaceutics, *Adv. Drug Deliv. Rev.*, 2021, **176**, 113788.
32 J. Xie, J. Jiang, P. Davoodi, M. P. Srinivasan and C. H. Wang, Electrohydrodynamic atomization: A two-decade effort to produce and process micro-/nanoparticulate materials, *Chem. Eng. Sci.*, 2015, **125**, 32–57.
33 M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, *et al.*, BOTORCH: a framework for efficient monte-carlo Bayesian optimization, in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Curran Associates Inc., 2020. pp. 21524–21538.
34 J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger and A. G. Wilson. GPyTorch: blackbox matrix-matrix Gaussian process inference with GPU acceleration, in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Curran Associates Inc., Red Hook, NY, USA, 2018, pp. 7587–7597.
35 J. Hensman, A. Matthews and Z. Ghahramani, Scalable Variational Gaussian Process Classification, in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, PMLR, San Diego, California, USA, 2015, pp. 351–360, **https://proceedings.mlr.press/v38/hensman15.html**.
36 R. Astudillo and P. Frazier, Bayesian Optimization of Composite Functions, in *Proceedings of the 36th International Conference on Machine Learning*, PMLR, 2019, pp. 354–363, **https://proceedings.mlr.press/v97/astudillo19a.html**.
37 H. B. Zhang, M. J. Edirisinghe and S. N. Jayasinghe, Flow behaviour of dielectric liquids in an electric field, *J. Fluid Mech.*, 2006, **558**, 103.
38 K. C. Felton, J. G. Rittig and A. A. Lapkin, Summit: Benchmarking Machine Learning Methods for Reaction Optimisation, *Chem.: Methods*, 2021, **1**(2), 116–122.
39 K. Kandasamy, K. R. Vysyaraju, W. Neiswanger, B. Paria, C. R. Collins, J. Schneider, *et al.*, Tuning hyperparameters without grad students: scalable and robust Bayesian optimisation with dragonfly, *J. Mach. Learn. Res.*, 2020, **21**(1), 3098–3124.
40 J. A. G. Torres, S. H. Lau, P. Anchuri, J. M. Stevens, J. E. Tabora, J. Li, *et al.*, A Multi-Objective Active Learning Platform and Web App for Reaction Optimization, *J. Am. Chem. Soc.*, 2022, **144**(43), 19999–20007.
41 R. J. Hickman, M. Sim, S. Pablo-García, G. Tom, I. Woolhouse, H. Hao, *et al.*, Atlas: a brain for self-driving laboratories, *Digital Discovery*, 2025, **4**(4), 1006–1029.
42 R. J. Hickman, G. Tom, Y. Zou, M. Aldeghi and A. Aspuru-Guzik, Anubis: Bayesian optimization with unknown feasibility constraints for scientific experimentation, *Digital Discovery*, 2025, **4**, 2104–2122, DOI: **10.1039/D5DD00018A**, **https://pubs.rsc.org/en/content/articlelanding/2025/dd/d5dd00018a**.
43 F. Wang, M. Elbadawi, S. L. Tsilova, S. Gaisford, A. W. Basit and M. Parhizkar, Machine learning predicts electrospray particle size, *Mater. Des.*, 2022, **219**, 110735.
44 M. Parhizkar, P. J. T. Reardon, J. C. Knowles, R. J. Browning, E. Stride, R. B. Pedley, *et al.*, Performance of novel high throughput multi electrospray systems for forming of polymeric micro/nanoparticles, *Mater. Des.*, 2017, **126**, 73–84.
45 R. P. A. Hartman, D. J. Brunner, D. M. A. Camelot, J. C. M. Marijnissen and B. Scarlett, Jet break-up in electrohydrodynamic atomization in the cone-jet mode, *J. Aerosol Sci.*, 2000, **31**(1), 65–95.
46 J. Xu, K. Li, M. Liu, X. Gu, P. Li and Y. Fan, Studies on preparation and formation mechanism of poly(lactide-co-glycolide) microrods via one-step electrospray and an application for drug delivery system, *Eur. Polym. J.*, 2021, **148**, 110372.
47 D. Khatamsaz, B. Vela, P. Singh, D. D. Johnson, D. Allaire and R. Arróyave, Bayesian optimization with active

learning of design constraints using an entropy-based approach, *npj Comput. Mater.*, 2023, **9**(1), 1–14.

48 R. Arróyave, D. Khatamsaz, B. Vela, R. Couperthwaite, A. Molkeri, P. Singh, *et al.*, A perspective on Bayesian methods applied to materials discovery and design, *MRS Commun.*, 2022, **12**(6), 1037–1049.

49 R. R. Griffiths, L. Klarner, H. B. Moss, A. Ravuri, S. Truong, S. Stanton, *et al.*, GAUCHE: A Library for Gaussian Processes in Chemistry, in *Proceedings of the 37th International Conference on Neural Information Processing Systems (NIPS '23)*, Curran Associates Inc., Red Hook, NY, USA, 2023, pp. 76923–76946, DOI: **10.48550/arXiv.2212.04450**.

50 R. R. Griffiths and J. M. Hernández-Lobato, Constrained Bayesian optimization for automatic chemical design using variational autoencoders, *Chem. Sci.*, 2020, **11**(2), 577–586.

51 F. Wang, A. Harker, M. Edirisinghe and M. Parhizkar, Tackling Data Scarcity Challenge through Active Learning in Materials Processing with Electrospray, *Adv Intell Syst.*, 2024, **6**(7), 2300798.

52 D. Li, F. Zhang, C. Liu and Y. Chen, Constrained Multi-objective Bayesian Optimization through Optimistic Constraints Estimation, *arXiv*, 2025, preprint, arXiv:2411.03641, DOI: **10.48550/arXiv.2411.03641**.

53 S. Daulton, M. Balandat and E. Bakshy, Expected Hypervolume Improvement for Parallel Multi-Objective Bayesian Optimization, in *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS '20)*, Curran Associates Inc., Red Hook, NY, USA, 2020, pp. 9851–9864, DOI: **10.48550/arXiv.2006.05078**.