Supporting Information

Accelerated AI Development for Autonomous Materials Synthesis in Flow

Robert W. Epps,[1] Amanda A. Volk,[1] Kristofer G. Reyes,[2] and Milad Abolhasani*[1]

[1] *Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina, 27606, United States E-mail: abolhasani@ncsu.edu*

[2] *Department of Materials Design and Innovation, University at Buffalo, Buffalo, New York, 14260, United States*

Table of Contents

SI Section S.1 – Surrogate Model Design 2
SI Section S.2 – Validation of the Surrogate Model-Based Optimization 4
SI Section S.3 – Gaussian Process Regression Training Parameters..... 5
SI Section S.4 – Ensemble Neural Network Architecture 6
SI Section S.5 – Neural Network Structure Tuning..... 8
SI Section S.6 – UCB Parameter Tuning..... 9
SI Section S.7 – Evolutionary Algorithm Tuning..... 10

SI Section S.1 – Surrogate Model Design

All four components of the surrogate model are derived from real-world experimental data. The global failure rate was estimated from experiments reported in prior work.[1] Shown in **Figure S1**, 130 experiments of the same reaction conditions were conducted in series. Among those tests, 2 experiments resulted in no detected emission. Subsequently, a global failure rate of 1.5% was assigned to every simulated experiment. The regions of the input space which produce non-emitting samples were estimated by training a Gaussian-naïve Bayes classifier on 85% of the full experimental data set for either emitting (1) or non-emitting (0) samples, where the remaining 15% was used for model validation. The ground truth model was built through three GPRs trained on 85% of the full experimental data set, with 15% for validation, for each of the three output variables. The GPRs were trained with a zero-prior mean estimate and a squared exponential kernel

$$\boldsymbol{\Sigma^0(x, x')} = \boldsymbol{\sigma_0^2}\, \boldsymbol{exp}\left[-\frac{\| \boldsymbol{x} - \boldsymbol{x'} \|_2^2}{\boldsymbol{2l^2}}\right]$$

The prior uncertainty σ_0 and length scale l hyperparameters were optimized through five-fold cross-validation. Given experimental training data, the surrogate model predictions for the responses are then given by the posterior predictive distribution

$$f(x) = \Sigma^0(x, X)[\Sigma^0(X, X) + \sigma_W^2 I]^{-1}\mathrm{y}$$

where σ_W is the noise level, y is the vector of experimental data responses, X is the matrix of experimental data inputs, and $\Sigma^0(X, X)$ is the covariance matrix for every pair of experimental data inputs in X, and $\Sigma^0(x, X)$ is the vector of covariances between the test in put x and the experimental data inputs in X. Sampling noise for each of these parameters was derived from the estimated variance as predicted by the GPRs used in the ground truth models. A constant value was used for each, as statistical analysis indicates homoscedastic noise.

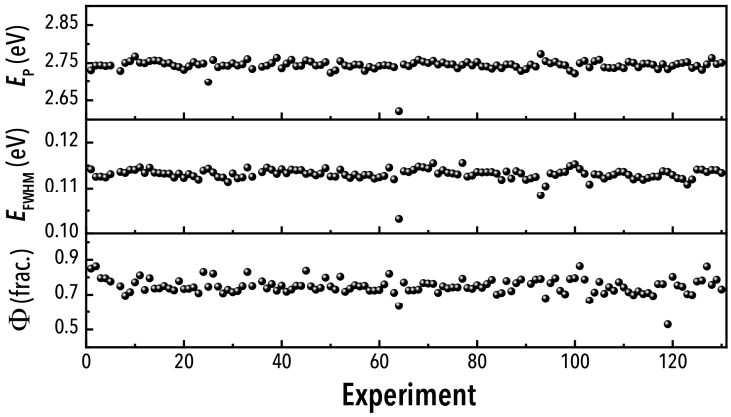

Figure S1. Sampling Stability Analysis. *Output experimental data – peak emission energy (E_P), emission linewidth (E_{FWHM}), and photoluminescence quantum yield (**Φ**) – collected across 130 replicates of the same experimental conditions on the real-world microfluidic anion exchange reaction platform, reported in Epps et al.*[1]

SI Section S.2 – Validation of the Surrogate Model-Based Optimization

The same experiment selection algorithm applied in our prior study[1] was integrated into the simulated reaction environment. The resulting optimization predictions were compared to the corresponding real-world experimental data. A sub-set of optimization runs was evaluated in this manner (**Figure S2**), and the simulation environment produced reasonably high predictability with 69% of experimentally measured best objective function values falling inside the predicted inter-quartile range.

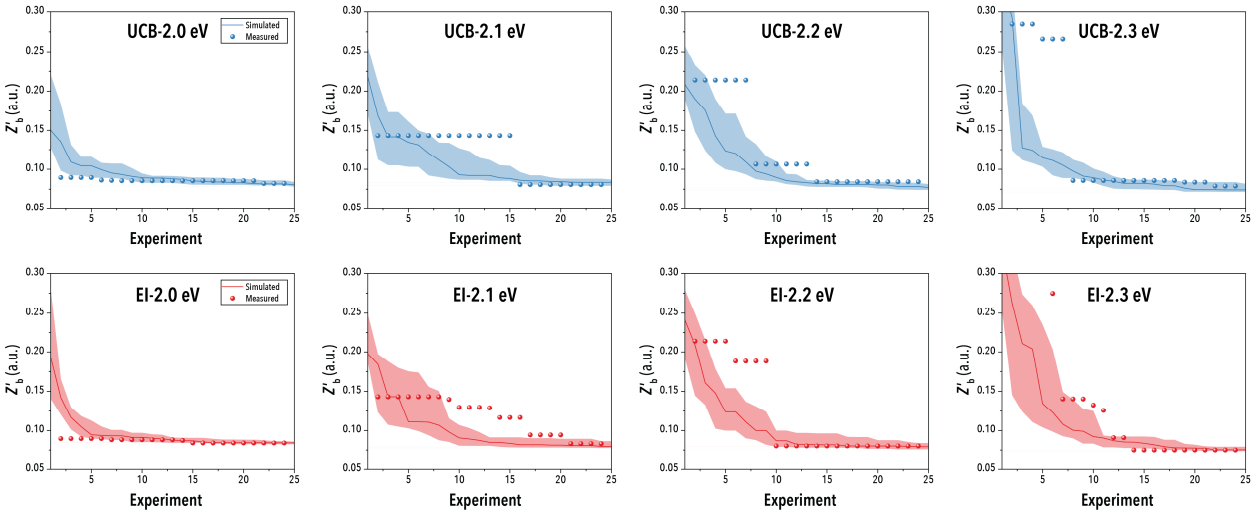

Figure S2. Validation of the Surrogate Model-Based Optimization Strategy. *Outputs of real-world collected experimental data reported in Epps et al.[1], compared to the median best simulated objective function value predictions (Z'_b) using the corresponding decision-making algorithms, where the shaded regions correspond to the 25th and 75th percentiles. Experimental results were compared for peak emission energies of 2.0 eV, 2.1 eV, 2.2 eV, and 2.3 eV with upper confidence bound (UCB) and expected improvement (EI) decision policies.*

SI Section S.3 – Gaussian Process Regression Training Parameters

All GP-based simulated campaigns began with three independent GP prior beliefs, for each objective and each using an independent, isotropic, squared exponential kernel:

$$\boldsymbol{\Sigma^0(X, X') = \sigma_0^2 \exp\left[-\frac{\| X - X' \|_2^2}{2l^2}\right]}$$

where $\boldsymbol{\sigma_0 = 0.1}$ and $\boldsymbol{l = 0.25}$. Upon receiving noisy data, the hyper-parameters were allowed to be optimized using by maximum likelihood methods, and the hyper-parameters assigned to the beliefs of each of the three different objectives were allowed be optimized to different values. That is, the three GP beliefs did not share any hyper-parameters information between them. The GP models were trained in this way for 50 consecutive experiments with the surrogate model. This process was repeated 100 times for each experimental campaign.

SI Section S.4 – Ensemble Neural Network Architecture

The ensemble neural network models were constructed using equally weighted cascade- or feed- forward ANNs. Data sets were randomly subsampled for each model in the ensemble using a training-testing-validation split of 0.7-0.15-0.15. ANNs were trained using the Levenberg-Marquardt backpropagation function with a maximum epoch number of 1000, an initial μ of 0.001, a μ decrease factor of 0.1, a μ increase factor of 10, a maximum μ of $1x10^{10}$, a minimum performance gradient of $1x10^{-9}$, a maximum number of validation failures of 100, a maximum training time of 2 s, a hyperbolic tangent sigmoid transfer function, and with memory reduction by a factor of 2. ENNs were fully reconstructed after each experiment. Model uncertainty was determined through the variance of predictions across the ensemble.

Subsampling of the available experimental data has the potential to normalize the influence of individual data points, resulting in either more robust or less precise models. However, shown in **Figure S3**, data subsampling within the ensemble neural network demonstrated no detectable impact on the efficiency of the AI-guided optimization algorithm. This observation is likely due to sparse data availability in a five-dimensional space. The limited available information results in similar model predictions regardless of subsampling rate.

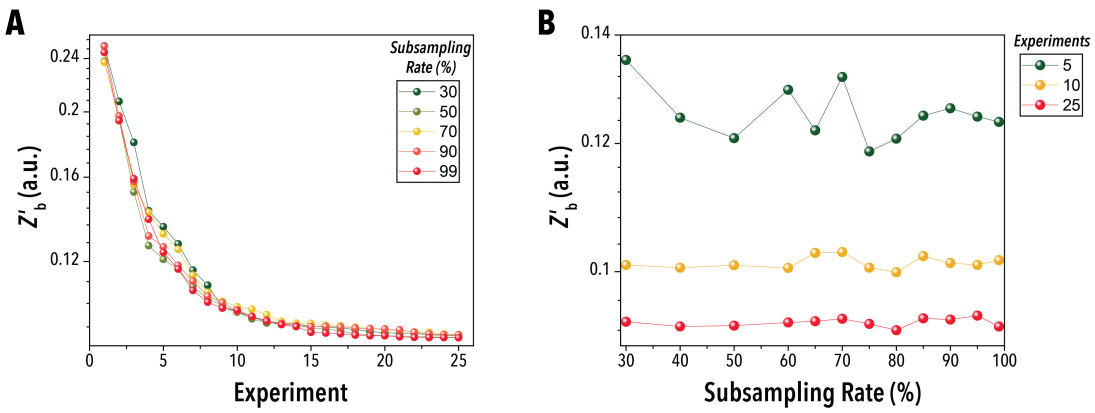

Figure S3. Data Subsampling. *(A) Best median objective function (Z'_b) for objective function weights of [0.8,0.1,0.1] with an expected improvement policy across 25 experiments and (B) after 5, 10, and 25 experiments for data subsampling rates for each model in the 200-model ensemble ranging from 30% to 99%. A training-testing split of 4.667:1 was maintained for all subsampled data sets.*

SI Section S.5 – Neural Network Structure Tuning

While the highest performing ensemble neural network model featured individual neural networks of randomized structure, ensembles containing constant architecture neural networks were also systematically studied to provide further insights into the predictive capability of various systems. Shown in **Figure S4**, the feed forward neural network ensemble significantly underperformed relative to cascade forward for all tested structures. Furthermore, the feed forward structure ensemble demonstrated decreasing formulation optimization performance with increasing structural complexity, suggesting that the optimal feedforward system is unlikely to match the performance of cascade forward ensembles. The cascade forward structures showed little distinction in performance once a sufficient level of complexity was attained.

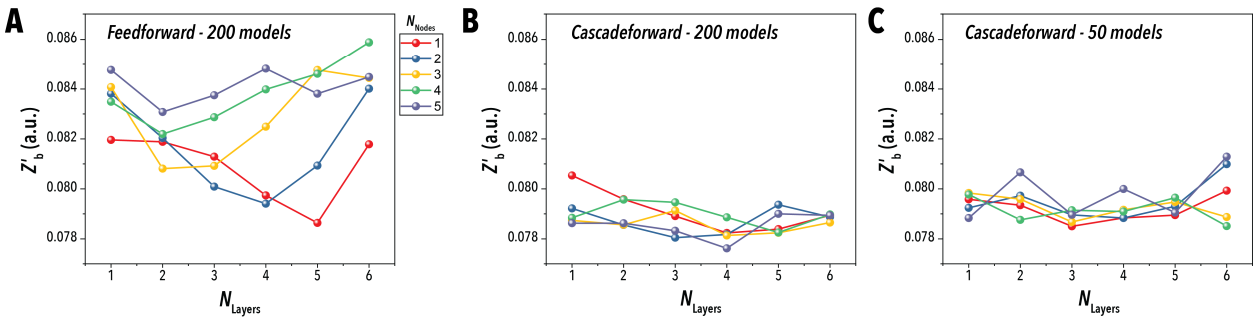

Figure S4. Ensemble Neural Network Structure Analysis*. Best median objective function value after 25 experiments ($Z'_{25,b}$) for ensemble neural networks of constant structure with an expected improvement policy as a function of the number of layers (N_{Layers}) and nodes (N_{Nodes}) for (A) a 200 feed forward, (B) a 200 cascade forward, and (C) a 50 cascade forward ensemble neural network model. Objective function weights of [0.85,0.05,0.1] were used.*

SI Section S.6 – UCB Parameter Tuning

The upper confidence bound (UCB) decision policy on an ensemble model operates with the following equation,

$$X_{n+1}^{UCB} = argmax(f_n(X) + \varepsilon\sigma_n(X))$$

where X_{n+1}^{UCB} is the next set of experimental conditions to text, $f_n(X)$ and $\sigma_n(X)$ are the mean estimate and variance for reaction conditions X at time n, and ε is a tunable parameter representative of the weight given to exploration over exploitation of the model. Systematic analysis of the UCB tunable parameter through the simulation environment indicates that the optimal value for this system is weighted more heavily towards exploration over the value assigned in our prior reported algorithm. As shown in **Figure S5**, UCB with ε value of 0.8 performed the best in terms of the learning and optimization rates of the metal halide perovskite quantum dot formulations.

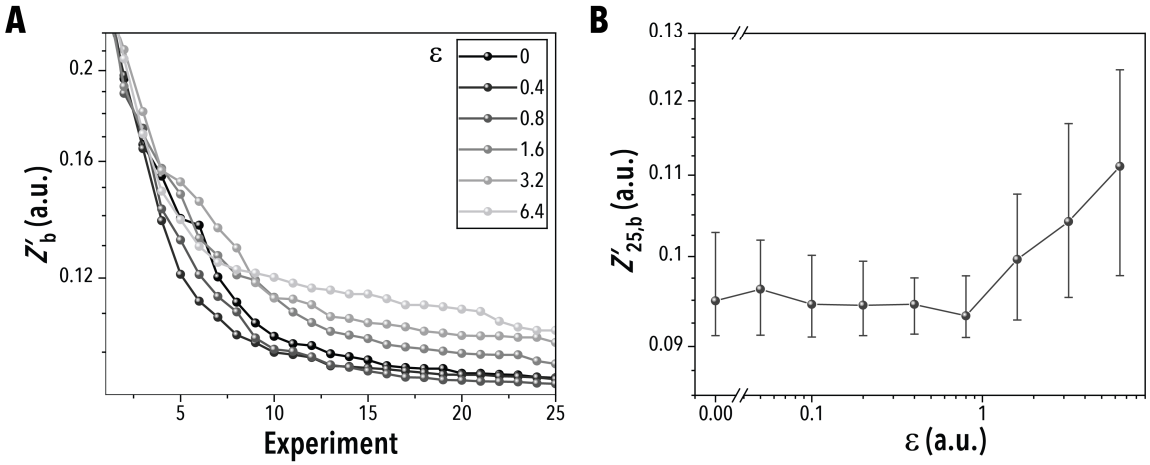

Figure S5. Tuning of UCB Exploration Parameter*. (A) Best median objective function (Z'_b) for objective function weights of [0.8,0.1,0.1] for varied UCB tuning parameter (ε) across 25 experiments.* ***(B)*** *Best median objective function value after 25 experiments ($Z'_{25,b}$) for nine different values of ε, where error bars correspond to the 25th and 75th percentiles.*

SI Section S.7 – Evolutionary Algorithm Tuning

Two evolutionary algorithms were evaluated in the simulated environment: Covariance Matrix Adaptation Evolution Strategy (CMA-ES) and Non-dominated Sorting Genetic Algorithm II (NSGA-II). Both of which feature many tunable parameters that may be adjusted for a given optimization problem. The primary variable optimized in this work is the generation population size ($N_{Generation}$). As presented in **Figure S6**, fine tuning of $N_{Generation}$ revealed optimal population sizes of 4 and 6 for CMA-ES and NSGA-II respectively.

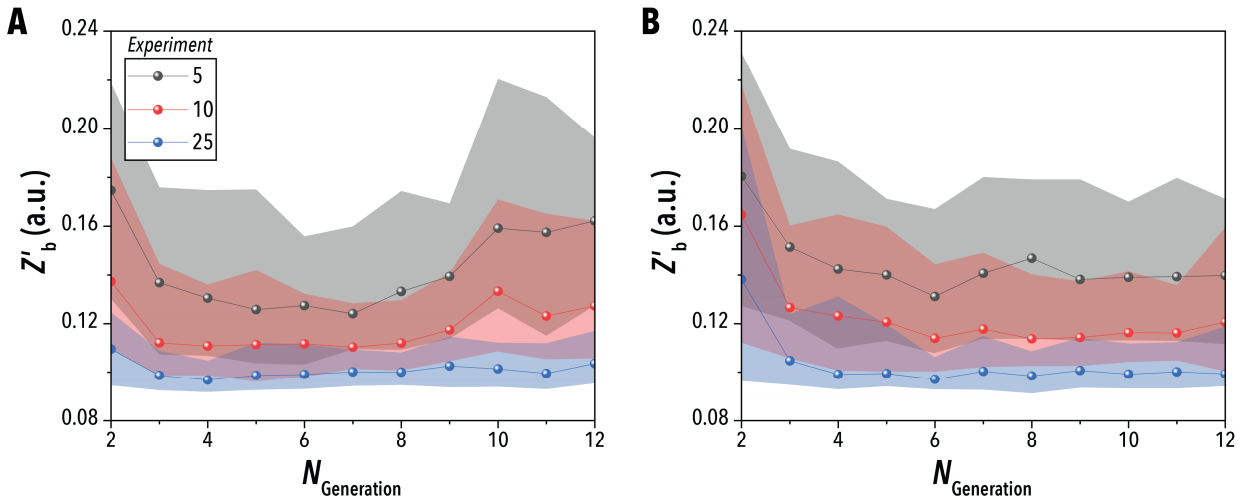

Figure S6. Tuning of Evolutionary Algorithms*. (A) Best median objective function value (Z'_b) after 5, 10, and 25 experiments for eleven generation sizes ($N_{Generation}$) for Covariance Matrix Adaptation Evolution Strategy (CMA-ES) and (B) Non-dominated Sorting Genetic Algorithm II (NSGA-II), where the shaded regions correspond to the 25th and 75th percentiles.*

Reference

[1] R. W. Epps, M. S. Bowen, A. A. Volk, K. Abdel-Latif, S. Han, K. G. Reyes, A. Amassian, M. Abolhasani, *Adv. Mater.* **2020**, 2001626.