Dimensional control over metal halide perovskite crystallization guided by active learning

Zhi Li,[§] Philip W. Nega,[§] Mansoor Ani Najeeb,[‡] Chaochao Dun,[§] Matthias Zeller,[#] Jeffrey J. Urban,[§] Wissam A. Saidi,[°] Joshua Schrier, [‡†] Alexander J. Norquist,[‡] Emory M. Chan*[§]

[§]Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA

[‡]Department of Chemistry, Haverford College, 370 Lancaster Avenue, Haverford, Pennsylvania 19041, USA

[†]Department of Chemistry, Fordham University, 441 E. Fordham Road, The Bronx, New York 10458, USA

[#]Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette IN 47907, USA

[°]Department of Mechanical Engineering and Materials Science, University of Pittsburgh, Pittsburgh, PA 15261, USA

*To whom correspondence should be addressed: EMChan@lbl.gov.

Abstract

Metal halide perovskite (MHP) derivatives, a promising class of optoelectronic materials, have been synthesized with a range of dimensionalities that govern their optoelectronic properties and determine their applications. We demonstrate a data-driven approach combining active learning and high-throughput experimentation to discover, control, and understand the formation of phases with different dimensionalities in the morpholinium (morph) lead iodide system. Using a robot-assisted workflow, we synthesized and characterized two novel MHP derivatives that have distinct optical properties: a one-dimensional (1D) morphPbI$_3$ phase ([C$_4$H$_{10}$NO][PbI$_3$]) and a 2D (morph)$_2$PbI$_4$ phase ([C$_4$H$_{10}$NO]$_2$[PbI$_4$]). To efficiently acquire the data needed to construct a machine learning (ML) model of the reaction conditions where the 1D and 2D phases are formed, data acquisition was guided by a diverse-mini-batch-sampling active learning algorithm, using prediction confidence as a stopping criterion. Querying the ML model uncovered the reaction parameters that have the most significant effects on dimensionality control. Based on these insights, we propose a reaction scheme that rationalizes the formation of different dimensional MHP derivatives in the morph-Pb-I system. The data-driven approach presented here, including the use of additives to manipulate dimensionality, will be valuable for controlling the crystallization of a range of materials over large reaction-composition spaces.

I. Introduction

Metal halide perovskite (MHP) derivatives are emerging optoelectronic materials with tunable physical properties[1–4] and applications in photovoltaic[5,6] and ferroelectric devices,[7] light-emitting diodes,[8,9] and lasers.[9,10] MHP derivatives have been synthesized with diverse crystal structures, with their metal halide frameworks exhibiting connectivities across one,[11] two,[12] or three dimensions.[13] MHP dimensionality (1D, 2D, and 3D, respectively) is a critical material parameter since it governs physical properties such as optical absorption, luminescence wavelength, charge transport, and stability.[14–16]

One strategy for controlling the dimensionality of MHP derivatives is tuning the inorganic or organoammonium A-cations that lie between metal halide $[BX_6]^{4-}$ octahedra, e.g., in the canonical ABX_3 perovskite unit cell. When A-cations are small, with effective radii of 1.7-2.6 Å (e.g., methylammonium),[17] 3D MHP derivatives form. In contrast, larger cations (> 2.6 Å) often give rise to lower dimensional MHP derivatives.[18,19] When the cross-sectional area of the cations is larger than 40 Å^2, MHP derivatives almost exclusively exhibit 0D or 1D connectivities rather than 2D, owing to steric hindrance.[20] For example, heterocyclic organoammonium cations with six or fewer ring members are able to form 2D MHP derivatives with lead iodide, while seven-member and larger rings rarely form in 2D.[20–23] Despite this limited ability to predict the dimensionalities of heterocyclic organoammonium lead halides, controlling dimensionalities for MHP derivatives incorporating a wide range of cations is still a major challenge.

One factor that complicates the prediction of MHP dimensionality is that different phases in an A-B-X reaction system (e.g., the 3D $FAPbBr_3$ and 2D FA_2PbBr_4[24], where FA = formamidinium) may form under different reaction conditions (FA:Pb ratio), given a reaction of A-cations, metal (B^{2+}) cations, and halide anions (X^-). Controlling dimensionality is a practical challenge for device applications — mixed phases[24] in MHP thin films can lead to low device performance.[25] Developing design rules for synthesizing MHPs with controllable dimensionalities will drive increases in performance of real-world MHP devices. Furthermore, understanding the kinetic and thermodynamic factors that influence dimensionality in a single cation-metal-halide (CMH) system will allow us to design functional materials and their reaction pathways. Nevertheless, it is still unclear how to control the dimensionalities of MHP derivatives by tuning the reaction parameters in a single CMH system. Multiple crystalline phases in a single CMH system are typically discovered through trial and error. This time- and resource-intensive approach is particularly inefficient for crystallizing new MHP phases since it can require simultaneous optimization of a large number of experimental variables in high-dimensional experimental parameter space. Insufficient sampling of reaction spaces risks missing rare MHP phases and precludes a comprehensive understanding of the formation of different phases.

Machine learning (ML) and high-throughput experimentation (HTE) have been recently leveraged to accelerate material discovery and design,[26,27] inspiring our efforts to apply these

tools to the crystallization of MHP derivatives.[28] Yet, even with HTE, it can be impractical to perform the large number of experiments needed to train common ML models. To overcome this challenge, researchers have utilized active learning (AL),[29,30] a sequential learning method in which a ML model is iteratively refined over repeated cycles of experimentation through an algorithm that selects new experiments based on the performance of the most recent model. An efficient AL sampling algorithm increases the learning speed of ML models and reduces the number of experiments needed. This efficiency is beneficial for constructing material phase diagrams,[31–33] – AL has been used to accelerate the acquisition of phase and composition diagrams of multicomponent materials including ferroelectric ceramics,[34] piezoelectric materials,[35] phase-change materials,[36] catalysts,[37] and metal halide perovskite thin films.[38] In these workflows, samples have been typically labeled using data acquired from simulations,[31–33] existing datasets,[37] or high-throughput characterization of existing material libraries.[36,37,38]

Controlling the dimensionality of MHP derivatives requires the analogous task of mapping the reaction conditions that produce specific phases in different regions of synthetic composition space. We hypothesized that an approach combining AL and HTE would be advantageous for building ML models that, given reaction conditions as inputs, predict the phase and dimensionality of MHP products. Solution-phase crystallization of materials presents a stringent test for AL since each reaction tends to be more costly than a simulation or measurement and can involve large numbers of reaction parameters, reagents, and additives. These syntheses are often governed by complex reaction networks[39] and stochastic processes (e.g., nucleation) that can lead to noisier and less predictable outcomes than observed with simulation and characterization. These problems are particularly acute when mapping phases across more than three dimensions. When exploring such high-dimensional space, determining when to stop AL experimentation is challenging[40] (and an ongoing subject of research[41]), since one cannot determine the actual accuracy of a model without synthesizing a large test set that is representative of the entire experimental space. To realize the potential of AL-guided materials synthesis, there is a strong need for robust AL workflows with clear stopping criteria and tolerance for noisy, high-dimensional data.

In this work, we used AL + HTE to discover, control, and understand the formation of MHP derivatives with different dimensionality in the morpholinium lead iodide (morph-Pb-I) system. We focused on this reaction system because morph$^+$ is a six-membered heterocyclic organoammonium that should theoretically form a 2D MHP derivative with PbI$_2$.[20–22] However, only 1D structures have been observed when morph is combined with different metal cations (e.g., Pb^{2+}, Sb^{4+}) and halides.[42–49] Using HTE, we successfully synthesized the 2D MHP derivative in the morph-Pb-I system. We adopted and modified an AL method to train ML classification models to predict the dimensionalities of phases formed in this chemical system. We established and validated a stopping criterion, based on the model prediction confidence, for terminating experimentation when the ML model was not improving significantly. We used a predictive ML model to uncover the reaction parameters that have the most significant effects on dimensionality control. These insights, combined with density functional theory calculations, allowed us to formulate a plausible reaction scheme that rationalized the formation of MHP derivatives with different dimensions in the morph-Pb-I system.

II. Results and discussion

Benchtop synthesis of morpholinium lead iodide.

A range of synthetic methods is used to grow MHP derivative crystals, such as seeded crystal growth,[50] slow evaporation,[51] and inverse temperature crystallization.[52] In this work, to synthesize morpholinium lead iodide (morph-Pb-I) structures across a range of dimensionalities, we used Anti-Solvent Vapor-assisted Crystallization (ASVC), a straightforward, room-temperature approach known to produce high-quality MHP crystals suitable for structure determination using single-crystal X-ray diffraction (sXRD)[53]. In initial experiments, we manually performed ASVC reactions by exposing a solution of morpholinium iodide and lead iodide (1:1 morph:Pb mole ratio) in dimethylformamide (DMF) to a saturated vapor of dichloromethane (DCM), the anti-solvent (scheme illustrated in Figure S1). These benchtop syntheses yielded yellow crystals (Figure 1a). Structural determination based on sXRD confirmed a new MHP derivative—morphPbI$_3$, crystallizing in the orthorhombic space group $P2_12_12_1$ (no. 19). Full crystallographic details of this new phase are given in Table S1; bond length and angles are listed in Table S2a and S2b. In morphPbI$_3$, [PbI$_6$]$^{4-}$ units are arranged in 1D chains of face-sharing octahedra (Figure 1b). The powder XRD (pXRD) pattern of morphPbI$_3$ matches the pXRD pattern simulated for the sXRD-derived crystal structure (Figure S2a). Tauc analysis of the absorption spectrum of the ground powder of morphPbI$_3$ indicates a direct bandgap of 2.69 eV (Figure 1e). No photoluminescence (PL) is detected for this compound. Despite the large bandgap and the absence of PL, the 1D yellow phase could have possible applications in second-order nonlinear optics[54] and piezoelectric devices[55] owing to its non-centrosymmetric space group (only ~18% of inorganic crystal structures reported are non-centrosymmetric)[56].
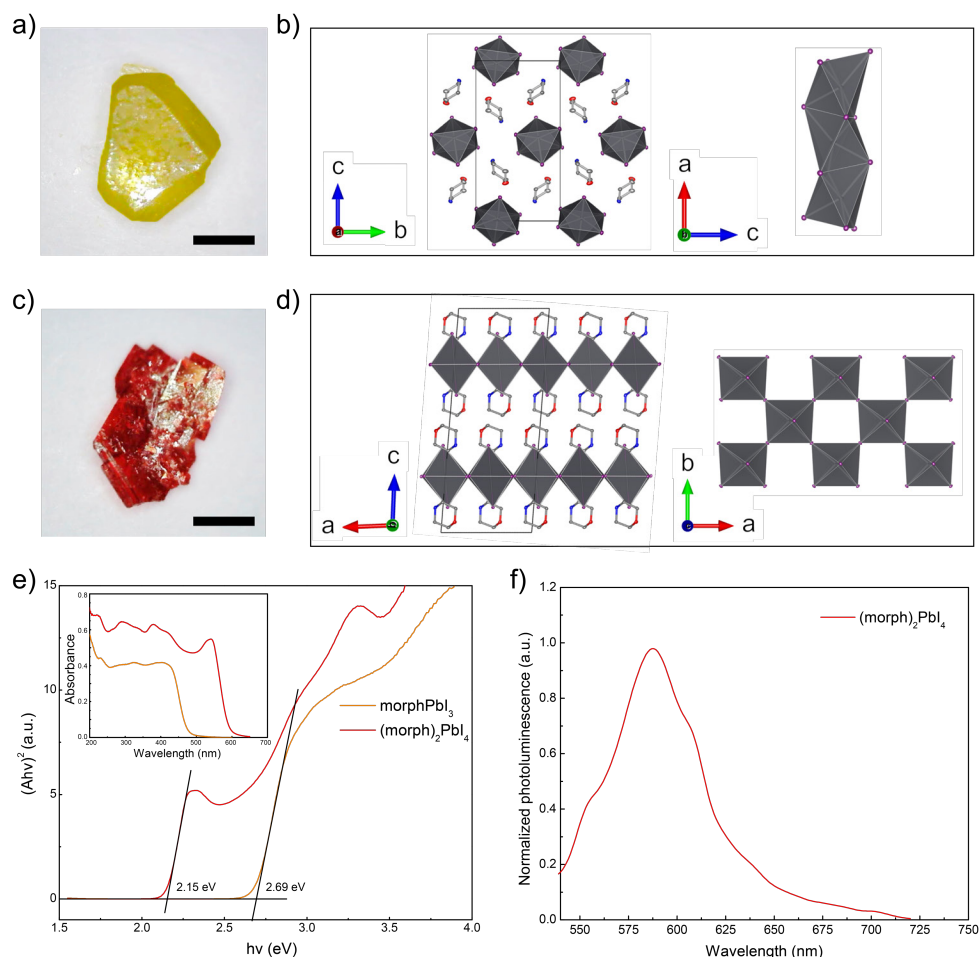
Figure 1. (a) Optical micrograph of a large crystal of the 1D "yellow" phase (scale bar: 1 mm). (b) The packing of $[PbI_{6/2}]^-$ chain structures in morphPbI$_3$ (1D yellow phase). (c) Optical micrograph of a large crystal of the 2D "red" phase (scale bar: 1 mm). (d) The packing of $[PbI_{2/1}I_{4/2}]^{2-}$ layer structures in (morph)$_2$PbI$_4$ (2D red phase). For (b) and (d), dark gray polyhedra represent [PbI$_6$] octahedra, while purple, light gray, red, and blue atoms correspond to iodine, carbon, oxygen, and nitrogen, respectively. Hydrogen atoms have been removed for clarity. (e) Diffuse reflectance spectra of ground powders of morphPbI$_3$ and (morph)$_2$PbI$_4$ in Tauc plots and absorbance units (in the inset). (f) Photoluminescence spectrum of (morph)$_2$PbI$_4$ (λ_{ex} = 470 nm).

Robot-accelerated ASVC perovskite workflow.

Although we only observed a single 1D morph-Pb-I structure with our isolated benchtop syntheses, 2D derivatives have been predicted theoretically.[20] To more comprehensively search for morph-Pb-I phases of different dimensionalities (especially 2D), we developed a robot-accelerated perovskite workflow based on high-throughput (HT) ASVC to explore a much larger reaction-composition space (see Figure S3 for the workflow). Similar to our previous robot-assisted perovskite investigation and discovery workflow,[28] our HT-ASVC workflow utilizes a liquid handling robot to dispense perovskite precursor solutions into reaction vials. We designed a custom, multi-well ASVC microplate (see Figure S3; CAD file

available[57]) that allowed our liquid-handling robot to prepare 24 parallel ASVC reactions on the 500 µl scale. Additional details for the HT-ASVC process are described in the SI. After crystallization, we photographed the reaction vials and recorded the morphologies and colors of the solid products. Then we characterized the solid products using HT optical microscopy, absorption and PL spectroscopy, and pXRD (HT characterization workflow illustrated in Figure S4). Representative optical micrographs, absorption/PL spectra, and pXRD patterns are shown in Figure S5-8.

Primary screening of 3D reaction-composition space.

Using our robotic workflow, we performed primary screening of the reaction-composition space of PbI_2 concentration ([Pb]), morphI concentration ([morph]), and formic acid concentration ([FAH]). We chose this reaction space because our previous HTE work has demonstrated that tuning the analogous concentrations in different ammonium lead halide systems resulted in the successful synthesis of 19 MHP derivatives from 45 A-cation candidates.[28] We used the Kennard-Stone (KS) algorithm[58,59] to uniformly sample 48 primary reactions from grid points generated in the allowed reaction-composition space (concentration constraints shown in Table S4). KS and grid-point generation algorithms are implemented in the ESCALATE software pipeline we developed to manage HT experiments and capture data.[60]
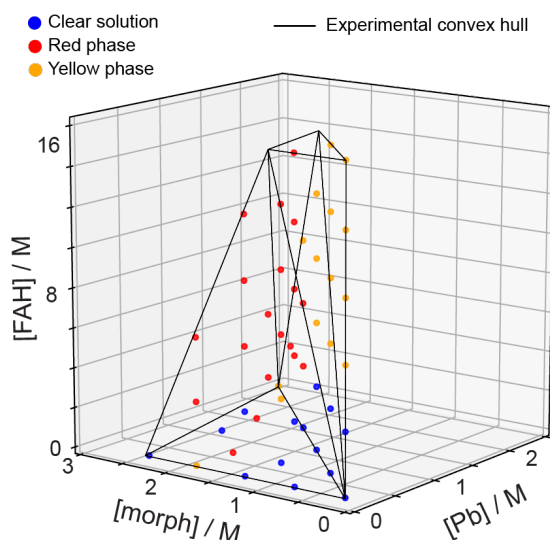

Figure 2. Convex hull of the allowed reaction-composition space (black lines) and the primary screening experiments (colored circles) contained within it, as a function of [Pb], [morph], and [FAH]. The blue circles indicate a clear solution with no crystals. The red and yellow circles indicate reaction outcomes of 2D (red) and 1D (yellow) phases, respectively.

Unlike benchtop syntheses, primary HT-ASVC screening produced both yellow-colored solids and red-colored solids with an unknown structure. Structural determination based on sXRD confirmed that the red crystals (Figure 1c) are a new MHP derivative with formula (morph)$_2$PbI$_4$ in the monoclinic space group $C2/c$ (no. 15, see Table S1 and Table S3a and b for details). In this crystal structure, [PbI$_6$]$^{4-}$ octahedral units are corner-connected to form 2D layers (Figure 1d). Therefore, we successfully demonstrated the efficacy of the robotic workflow in the discovery of a new 2D phase in the morph-Pb-I system, which was found neither in our benchtop synthesis nor experimentally reported by the literature. Compared to the 1D yellow phase, the 2D red phase has distinct optical properties. The absorption spectrum of the 2D phase indicates a direct bandgap of 2.15 eV, which is 0.54 eV lower than the 1D phase (Figure 1e). Under 470 nm excitation, the 2D phase exhibits a broad emission peak at 588 nm (Figure 1f), which could be utilized in the application of light-emitting diodes.[61] Unlike the 1D phase, the 2D phase has a centrosymmetric crystal structure.

Visualizing the distribution of reaction outcomes across the [morph]-[Pb]-[FAH] composition space (Figure 2) illustrates how reaction conditions determine the crystallization of the 1D and 2D (red and yellow) phases. For each HT-AVSC experiment in this space, we assigned a reaction outcome from one of three classes: (1) clear solution, (2) red phase, and (3) yellow phase (see "Product Scoring Rubric Based on Human Inspection" in the SI). For mixtures of yellow and red products, we labeled the reaction outcomes based on the major product. In general, clear solutions (no solids) were observed below 6 M FAH. Above 6 M FAH, the yellow phase formed at lower morph:Pb ratios (~1), while the red phase was more likely to form at higher morph:Pb. This dependence on the morph-to-Pb ratio is due to the different chemical stoichiometries of the two phases (morph:Pb = 1 in yellow phase, and 2 in red phase). Tuning the reactant ratio per target-compound stoichiometry has been utilized previously in several CMH systems, such as FA-Pb-Br[24]. Therefore, not only did primary HT-ASVC screening of the morph-Pb-I system identify a 2D phase, but the resulting dataset also provided guidance for controlling the dimensionality of morph-Pb-I in 3D reaction-composition space.

A modified workflow to screen additives in 6D reaction-composition space.

A prevailing trend in the fabrication of MHP devices is to use mixed solvents and additives to modify the crystallinity and morphology of MHP thin films, which can improve device performance.[62] However, the effect of these solvents and additives on the dimensionality of MHP products is still unclear. To understand such effects, we simultaneously incorporated four additional solvents and additives into our HT-ASVC reactions. These included three common solvents for MHP syntheses: DMF, dimethyl sulfoxide (DMSO), and γ-butyrolactone (GBL). These solvents were selected due to their distinct physical properties, such as their polarity[63] and their affinity for coordinating metal ions[62] and accepting hydrogen bonds.[64] We also included water as an additive since our previous study demonstrated that water content in perovskite precursor solutions affects the crystallinity of MHP single crystals and thin films.[65] Unlike DMSO or DMF, water and formic acid are both hydrogen bond donors and acceptors.[66]

When coexisting in solution, DMSO (or DMF) and water (or formic acid) are likely to form hydrogen bonds. Our new reaction-composition space was thus composed of six reaction parameters: [morph]; [Pb]; the volume fraction of DMSO, $V_{f,\,DMSO}$; the volume fraction of GBL, $V_{f,\,GBL}$; [FAH]; and the concentration of water, [H_2O]. A seventh parameter, $V_{f,\,DMF}$, was not included in this reaction space since it is calculated as 1- $V_{f,\,DMSO}$ - $V_{f,\,GBL}$. The modified robotic-synthesis procedure is described in the SI, and the constraints of all six parameters are given in Table S5. In the constrained 6D space, we generated a pool of 469,326 possible reaction compositions located on a fixed grid.

To rapidly characterize the outcomes of reactions performed in this extended 6D space, we acquired absorption spectra on products with a multifunction plate reader. Automated scripts classified reaction products with absorption edges ≥ 2.3 eV as the "yellow phase" and the reactions with absorption edges ≤ 2.15 eV as the "red phase" (see Figure S6a). If no solids formed, we labeled the reaction as a "clear solution." Representative pXRD patterns (Figure S7 and Figure S8) verified that the diffraction peaks of reaction products correspond to the phases predicted using absorption edges.

Exploring the role of MHP additives using AL.

To understand how combinations of additives contribute to the dimensionality of crystals in the morph-Pb-I system, we sought to train a ML model to predict the phase and dimensionality (i.e., 1D yellow phase or 2D red phase) for each combination of reagents in our 6D reaction pool. To train such a model efficiently in such high-dimensional space, we developed an uncertainty-based active learning method to perform repeated cycles of HT-ASVC microplate reactions that iteratively refine the ML model. For each cycle in our method, an AL algorithm selects the next batch of reactions to perform by identifying the regions of the 6D reaction-composition space where the ML model has the highest prediction uncertainty.[29,67]

To initiate AL, we performed a uniform sampling of 48 reactions using the KS algorithm. Since visualizing 6D data graphically is difficult, we projected reaction outcomes onto a lower-dimensional space using t-distributed stochastic neighborhood embedding (t-SNE)[68] (Step ❶ in Figure 3). In this initial sampling, we observed all three classes of reaction outcomes. We tested Random Forest (RF) models and Pearson VII Universal Function Kernel-based Supporting Vector Machine (SVM_PUFK) models, optimizing their hyperparameters on the initial dataset using five-fold cross-validation (CV). We selected the RF model as the basis for our AL algorithm since it exhibited the highest CV accuracy of 0.80 ± 0.09 (Table S6). The RF model, trained by the initial dataset, was used to predict reaction outcomes (yellow, red, or no crystals) and calculate the prediction probabilities P for each member of the reaction pool (Steps ❷ and ❸ in Figure 3). P is defined as the probability of the reaction outcome predicted to have the highest likelihood of forming (e.g., $P = 0.6$ if prediction likelihoods are 0.2 for clear solution, 0.6 for red phase, and 0.2 for yellow phase). Then the prediction uncertainty (U) for each point is calculated as $1 - P$. The distribution of U (Step ❹ in Figure 3) shows

that the regions with high uncertainty ($U > 0.5$) are found where multiple compounds form under similar conditions.

Our AL algorithm is designed to generate the next batch of 24 robot reactions based on the points in reaction-composition space with highest U. Simply selecting the 24 reactions with the highest U, however, would result in reactions with very similar conditions. To avoid such over-sampling in a small region, we implemented the diverse-mini-batch-sampling algorithm,[69] which divides high-uncertainty reactions ($U > 0.5$) into 24 mini-batches using k-Mean Clustering,[70] weighted by U. Then the centers of mass of the 24 mini-batches are selected as the set of reactions to perform in the next AL cycle. As shown in Step ❺ in Figure 3, the selected reactions have diverse reaction conditions and are located in the regions of high U. After this batch of reactions is performed and characterized, the reaction outcomes are collected (❻) and added to the dataset. The updated dataset is then used to retrain the RF model, at which point, the next active learning cycle (❷-❻) commences.
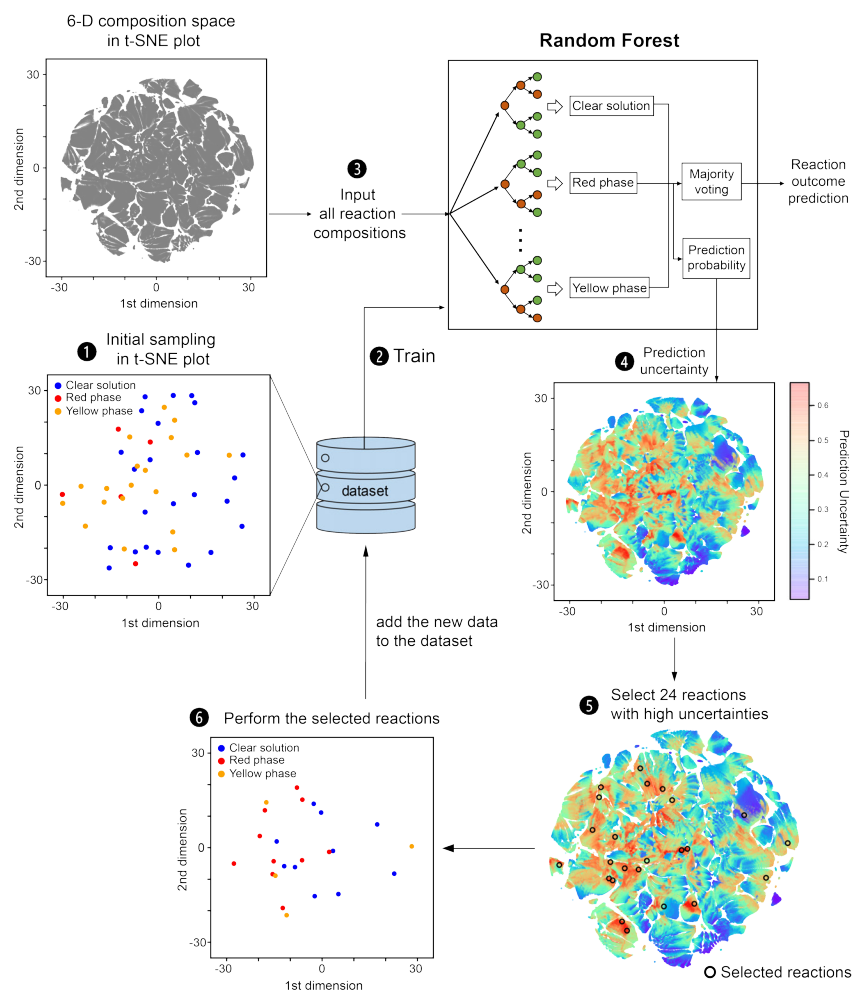

Figure 3. Illustration of the diverse-mini-batch-sampling AL loop. Uniformly sampled seed reactions (Step 1) are used to train a Random Forest model (2), which is then used to calculate prediction uncertainties (4)

for the entire pool of potential reactions previously generated in 6D experimental space (3). In subsequent cycles of AL, reactions are selected using a diverse-mini-batch-sampling algorithm that prioritizes reaction conditions where model predictions have high uncertainty (5). These reactions are performed using a synthesis robot (6), and the data are used to retrain the model and perform additional cycles of active learning (2-6). All potential reactions (gray), AL-sampled reactions (colored scattered circles), and prediction uncertainty distribution (colored map) are projected onto 2D space using t-distributed stochastic neighborhood embedding (t-SNE).

A critical outstanding question in AL-guided materials synthesis is when to terminate the AL loop. In principle, the RF model performance should be evaluated during AL to terminate the AL process when the model ceases to improve, but such evaluation is practically challenging since a large test set is not available to evaluate the RF model. To determine the stopping point for our AL runs, we monitored the average uncertainty (AU) and prediction confidence (PC) of the RF model after each AL iteration.[71] AU and PC are defined in equations (1) and (2):

$$AU = \frac{\sum_{k=1}^{N} U_k}{N} \tag{1}$$

$$PC = \frac{\sum_{k=1}^{N} (P_k - P'_k)}{N} \tag{2}$$

Here, $N = 469{,}326$ is the number of potential reactions in the reaction pool ; U_k is the prediction uncertainty for the kth reaction in the reaction pool; P_k and P'_k are the prediction probabilities of the most likely class and second most likely class for the kth reaction. A rising PC (decreasing AU) over AL cycles indicates that AL is still improving the RF model. A decreasing PC (rising AU) suggests that the RF model has ceased to improve, and AL should be stopped. Heuristically, PC usually increases at the beginning of AL and then decreases, which indicates that ML models were often improved by AL in the first several cycles and then remained little changed.[71,72] In general, adding new data to the training set does not reduce model performance, so decreasing PC implies a lack of improvement rather than a reduction in prediction quality.

After the 1st AL iteration of 24 reactions, the PC of our RF model increased from 0.41 to 0.45, and AU decreased from 0.34 to 0.32 (Figure 4a), which indicates an improvement of the RF model. Surprisingly, the 2nd AL iteration reduced PC and increased AU, suggesting that that experimentation should stop, as the model is no longer improving with the added data points. To confirm the downward trend of PC (and upward trend of AU), we performed three more AL iterations after the 2nd iteration. PC continued to decrease, so we ceased the AL experiments after the 5th iteration (the practical stopping point). The dataset from AL iterations 2-5 (shown in Figure S9) is included in the training set because expanding a training set rarely reduces the prediction accuracy for the test set.
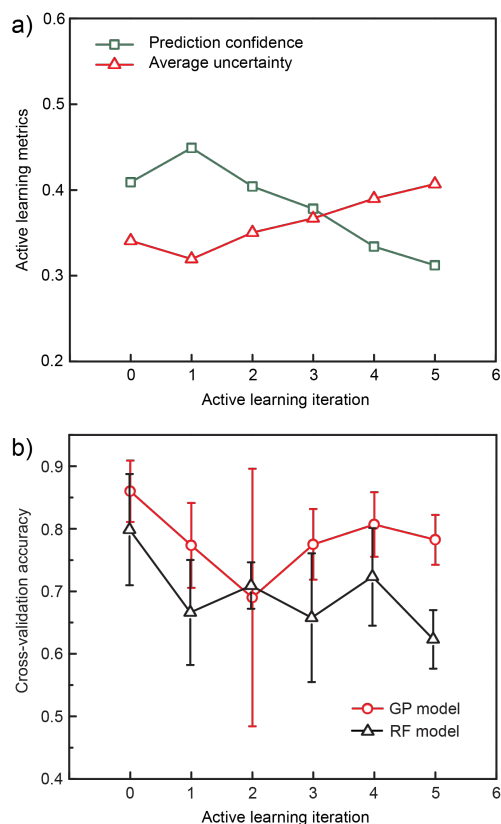
Figure 4. (a) Prediction confidence and average uncertainty of the RF model in each iteration of AL. (b) CV accuracies of the Gaussian process and RF models on the dataset collected after each iteration of AL (error bars represent the standard deviation across 5-fold CV).

To justify the practical stopping point, one could in principle calculate the prediction accuracy of the RF model (trained by initial sampling + 5 AL iterations) over a large test set representative of the overall reaction pool. However, the time and labor required to collect this test set defeats the purpose of AL,[71] which is to minimize the number of experiments. Here, we performed three analyses to support the conclusion that we have reached a reasonable stopping point after five AL iterations.

A good ML model should show modest variance during the last few AL iterations before the practical stopping point (from the 2nd to 5th iteration),[73] shown as small changes in its CV accuracy when including the last few AL runs in the training set. It is worth noting that since the training set has been collected, we are not limited to using RF models in the search for the best ML model. Our first analysis calculated the CV accuracies of multiple ML models after different AL iterations (Tables S7 and Figure S10). After the 5th AL iteration, the Gaussian process (GP) model shows the highest CV accuracy of 0.78 ± 0.04 (Figure 4b) and close to the theoretical limit of 0.78 (calculated by overfitting the dataset with deep neural networks, shown in Figure S11).[74] The CV accuracy is well above the random-classification (control) accuracy of 0.34 and the majority-class vote accuracy of 0.41. As the most accurate ML model, the GP model shows only slight variation in CV accuracy from the 2nd to 5th AL iteration, which

indicates an insignificant variance in the GP model and suggests that the AL experiments have reached the stopping point.

If the most accurate model, the GP model (trained on the dataset after the 5th AL iteration), can accurately predict unseen reaction conditions, AL has likely reached its stopping point. Therefore, in our second analysis, we performed 24 reactions located far from the tested reactions (initial sampling + 5 AL runs) in the reaction-composition space (Figure S12a). The 24 reactions were selected by the KS algorithm. The GP model, trained on the tested reactions, shows a prediction accuracy of 0.92 for the unseen 24 reactions (Figure S12b), which leaves little room for improvement and indicates that AL has reached its stopping point.

Testing the AL algorithm on a synthetic dataset (with labels) allows us to monitor the model prediction accuracy for the whole dataset and to investigate whether the stopping criterion based on prediction confidence is reasonable. As the final analysis, we ran the AL algorithm on a synthetic dataset with a similar structure to our experimental dataset (see Figure S13 and SI for details). After AL commences, the RF model's PC increases until the 4th AL cycle, after which PC decreases (Figure S14a). Thus, the 4th iteration is considered the theoretical stopping point. Like our AL experiment, the decreasing trend can be confirmed with a few additional iterations. Meanwhile, the prediction accuracy for the whole dataset rapidly increases from 0.48 (after initial sampling) to 0.82 (after four AL iterations) (Figure S14b). After the 4th cycle, the accuracy plateaus, eventually stabilizing at 0.85 after 100 iterations. Therefore, the model performance has only a 4% improvement with additional 96 AL iterations after the theoretical stopping point, confirming that the stopping criterion based on prediction confidence is reasonable. Clear stopping criteria, such as that demonstrated here, will benefit AL-assisted materials and chemistry research.

To summarize, we determined the AL stopping point based on the changes in prediction confidence and performed three tractable analyses to support the stopping point. Using diverse-mini-batch AL and a stopping criterion based on prediction confidence, we only needed to explore at most 0.035% of the reaction pool to successfully build and confirm the stopping point of ML models that accurately predict the formation of the 1D and 2D MHP derivatives in the morph-Pb-I system.

Importance of features and their effect on dimensionality.

A predictive ML model can be used to understand the physicochemical process of morpholinium lead iodide crystallization. To uncover the reaction parameters that have the most significant influence on morph-Pb-I dimensionality, we performed a permutation-feature-importance analysis[75] (see details in SI) on the RF model trained by portion of the dataset that includes only yellow and red phase outcomes. Feature-importance analysis revealed that [Pb], [morph], [FAH], and [H_2O] are important for controlling the formation of 1D and 2D phases (feature importance > 0.1) while the compositions of solvents (i.e., $V_{f,\ DMSO}$ and $V_{f,\ GBL}$) are much less important (feature importance < 0.05) and can be ignored (Figure 5a). To validate the feature downselection, we retrained the RF model with the yellow/red dataset and used

only [Pb], [morph], [FAH], and [H$_2$O] as features. Similar CV accuracies were observed compared to the model trained by the full set of features (Figure 5b), suggesting that the feature selection is effective. This analysis is corroborated by the visually distinguishable boundary between yellow and red phases in the parameter space of ln([morph]/[Pb]), [FAH], and [H$_2$O] (Figure 5c).
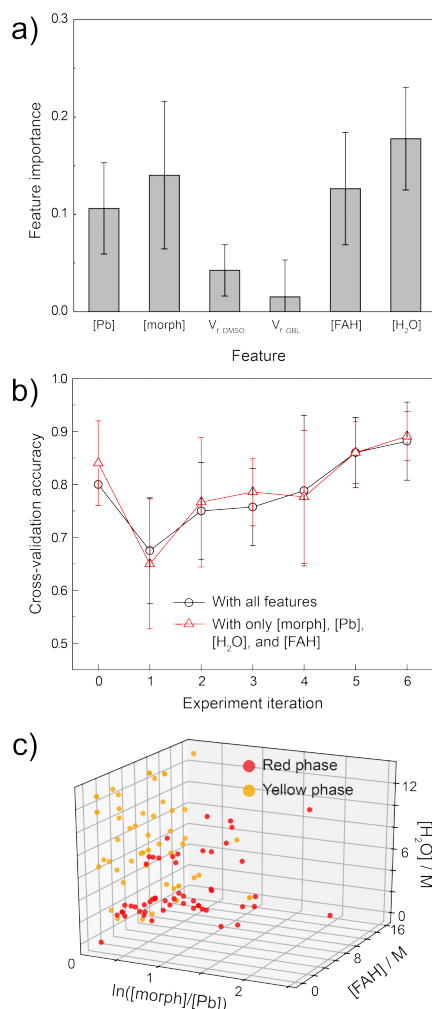

Figure 5. (a) Permutation-feature-importance of the RF model trained by the yellow and red phase reactions. (b) CV accuracies of the RF model on predicting yellow phase vs. red phase, with all features (black line) and only [morph], [Pb], [H$_2$O], and [FAH] (red line). (c) Outcomes of the yellow and red phase reactions as a function of ln([morph]/[Pb]), [FAH], and [H$_2$O].

To determine whether the important features have positive or negative effects on the formation of the yellow and red phases, we developed a data-driven approach that combines logistic regression (LG) modeling and statistical hypothesis testing. Using a "crystals only" dataset containing only outcomes that produced solids, we performed LG using only [Pb], [morph], [FAH], and [H$_2$O] as inputs, and with the outcomes labeled as "1" for the yellow phase and

"0" for the red phase. Since the LG model demonstrated a reasonable CV accuracy of 0.81 ± 0.07, the sign of the slope can be treated as the direction of the correlation between the corresponding feature and yellow phase formation. To test the hypotheses with statistical significance, we used bootstrapping[76] to sample the "crystals only" dataset 1000 times. We fit the 1000 samples into the LG model and obtained a distribution of slopes for each important feature (Figure S16a). The slopes of [morph], [Pb], [FAH], and [H_2O] are -2.33 ± 0.24, 2.37 ± 0.16, 1.09 ± 0.24, and 1.51 ± 0.22, respectively. Based on this LG analysis, we hypothesized that when solid is formed, [morph] has a negative effect on yellow phase formation while [Pb], [FAH], and [H_2O] have positive effects.

To further validate the correlations between reagent concentrations and crystal phases, we performed one-tailed statistical hypothesis testing.[77] We defined one null hypothesis for each feature: for the yellow phase formation, [morph] is hypothesized to have positive or no effect while [Pb], [FAH], and [H_2O] are hypothesized to have negative or no effect. Given the mean values and standard errors of the slopes, we rejected all four null hypotheses with a confidence level of 99%. Therefore, all four features have the hypothesized effects on the yellow phase formation. Using the same approach, we discovered that when solid is formed, [morph] has a positive effect on red phase formation while [Pb], [FAH], and [H_2O] have negative effects.

Understanding the underlying physicochemical process of dimensionality control.

The effects of [Pb] and [morph] on dimensionality control can be explained by the chemical stoichiometries in the two phases. The physicochemical process through which additives (i.e., water and formic acid) influence the dimensionality is still unclear. To understand this physicochemical process, we studied both the thermodynamics and kinetics of the ASVC reaction. First, we investigated whether the reaction is under thermodynamic control. We calculated the total energy of the yellow phase and red phase using density functional theory (DFT).[78] DFT calculations show that the yellow phase is slightly more stable than the red phase. However, the formation energy difference between these two phases is negligible (ΔE_{form} = 6.8 kJ per mole of Pb) and within the typical intrinsic error of DFT (~0.1 eV or ~10 kJ/mole)[79]. Thermogravimetric analysis (TGA) shows that the thermal decomposition temperatures for the yellow and red phases are close (240 °C and 200 °C respectively, shown in Figure S17), which agrees with DFT results. DFT calculations also show that only a small amount of energy (33.77 kJ/mol) is needed to convert the yellow phase to the red phase in the solid state (Scheme 1). Theoretically, this conversion can be achieved by mechanochemical grinding, which provides energy of 95-112 kJ/mol.[80] UV-vis absorption spectra show that at room temperature, the yellow phase can be converted to the red phase in the solid state with one equivalent morphI and manual grinding (Figure 6). The room-temperature synthesis of ASVC and the small energy difference between the 1D and 2D phases suggest that the formation of different dimensional phases in the morph-Pb-I system is not under thermodynamic control.

$$morphPbI_3 + morphI \longrightarrow (morph)_2PbI_4$$

$$\Delta E = 33.77 \; kJ \, / \, mol$$
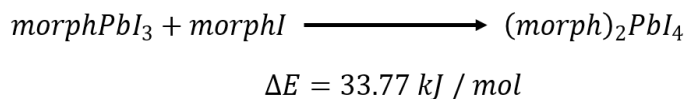
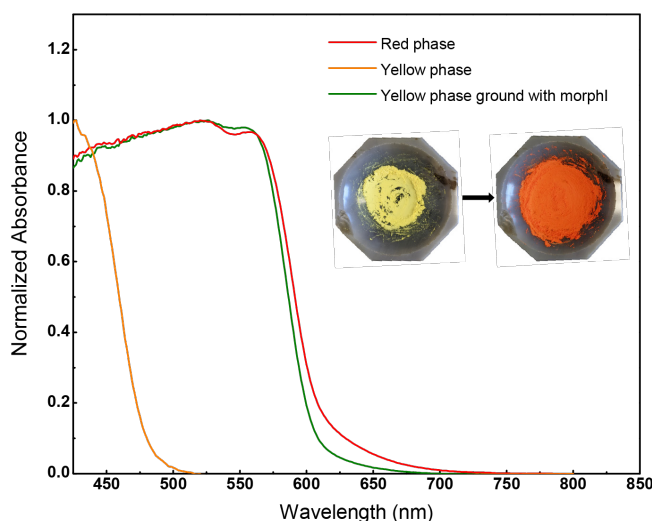**Scheme 1.** Solid-state conversion of yellow phase to red phase.

Figure 6. UV-vis absorption spectra of red phase, yellow phase, and yellow phase ground with additional morphI. The inset photos are powders of yellow phase (left) and yellow phase ground with morphI (right).

Based on the additives' effects on dimensionality control and intermediate species previously identified in MHP syntheses, we propose a reaction scheme that rationalizes how water and formic acid control the dimensionality by influencing the reaction kinetics of morpholinium lead iodide syntheses (Scheme 2). When morphI and PbI_2 are dissolved in the solution, solvent (i.e., DMF and DMSO) and additive molecules (i.e., formic acid and H_2O) coordinate with PbI_2 to form 1D complexes (INT-1 and INT-2 in Scheme 2) as prenucleation intermediates.[81–83] When the additives' concentrations are low, the 1D PbI_2 chains are coordinated mainly by solvent molecules to form intermediates similar to INT-2. When the additives' concentrations increase, more solvent molecules on the 1D PbI_2 chains are substituted by water and formic acid to form intermediates similar to INT-1.[82] In INT-1 and INT-2, additive molecules on Pb can form hydrogen bonds with the axial ligands (additives and solvents) on the adjacent Pb atoms at the same side of the 1D chain. Such intra-chain hydrogen bonds labilize the additive and solvent molecules on Pb by reducing the density of lone-pair electrons (from N and O) in the metal-ligand bonds and allocating part of the electron density to forming the hydrogen bonds.

In INT-1, the abundant intra-chain hydrogen bonds reduce the distance between adjacent axial ligands and facilitate the simultaneous dissociation of these ligands, which are substituted by the bridging-I⁻ from morphI. Thus, INT-1 is converted to INT-4—the 1D PbI_2 chain that

forms the framework of the yellow phase (see crystal structure in Figure 1b). Compared to INT-1 → INT-4, the I$^-$ ligand substitution reaction that polymerizes INT-1 to INT-3 (the 2D polymeric intermediate) could be much slower, making the formation of 1D yellow phase dominant process at higher additive concentrations. For INT-2, the lower density of intra-chain hydrogen bonds make the reaction rate of INT-2 → INT-4 slower than INT-2 → INT-5, which then becomes the dominant pathway. With additional I$^-$, INT-5 can be converted to INT-6, the 2D Pb-I plane that composes the red phase. For this reason, the 2D red phase becomes the main product at lower [H$_2$O] and [FAH].

These results suggest that water and formic acid influence the dimensionality of the products in morpholinium lead iodide syntheses by accelerating the formation of the 1D phase through intra-chain hydrogen bonding. Controlling dimensionality of MHP derivatives through the use of hydrogen-bonding additives should be applicable to other MHP systems.
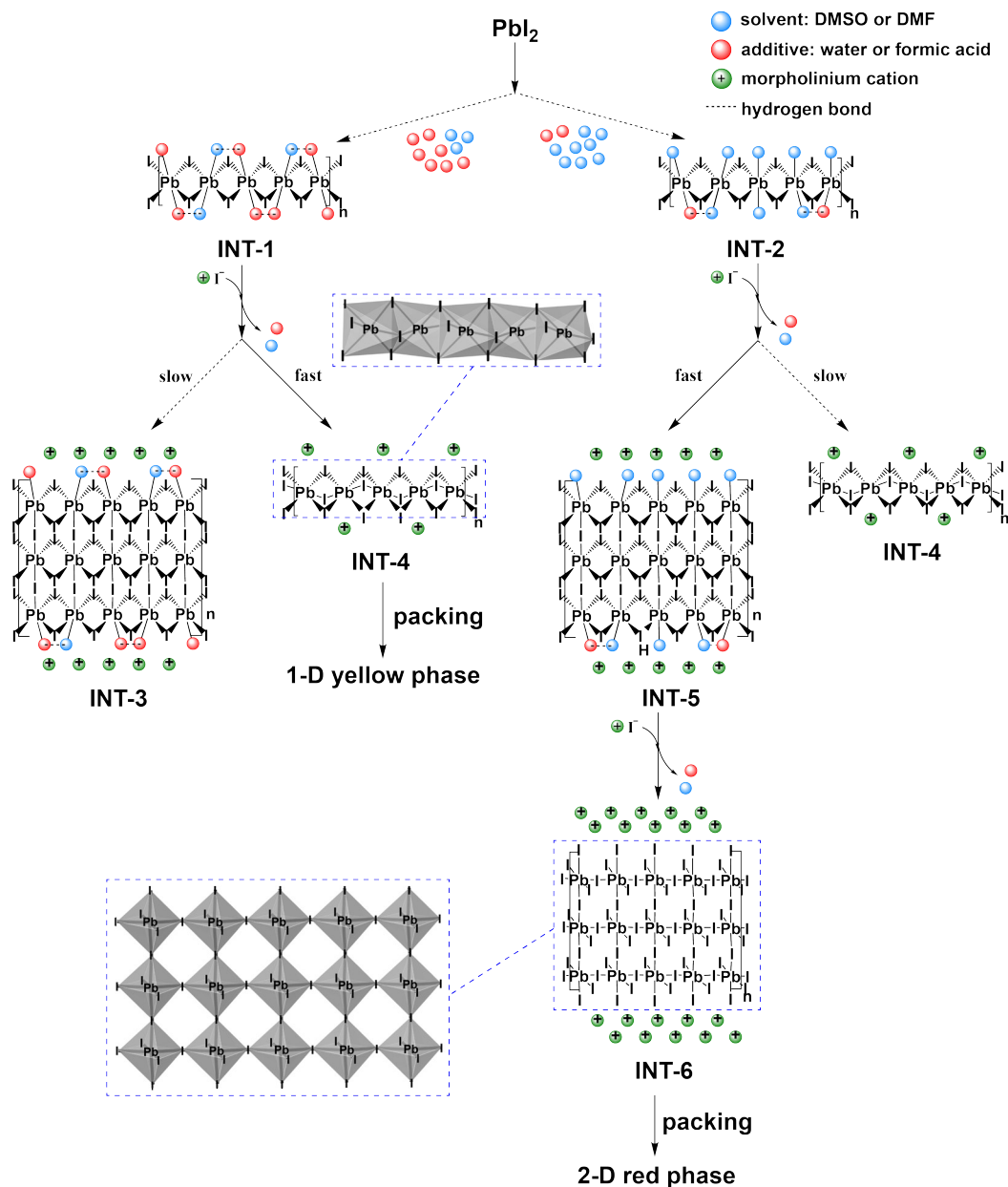
Scheme 2. Proposed reaction scheme for the formation of the 1D yellow phase and 2D red phase.

III.   Conclusion

Using a robotic workflow based on ASVC, we synthesized two novel MHP derivatives [1D morphPbI$_3$ and 2D (morph)$_2$PbI$_4$] with distinct optical properties. Although the existence (but not crystal structures) of 2D MHP derivatives based on morpholinium has been postulated theoretically, the synthesis and characterization of a 2D derivative has not been reported until this work. We demonstrated the efficacy of the KS sampling algorithm + robotic workflow in finding rare MHP derivatives. Using the uncertainty-based AL method with decreased prediction confidence as a stopping criterion, we sampled only 0.035% of the reaction-

composition space to build a predictive ML model to classify the reaction conditions where 1D and 2D phases are formed. By analyzing the feature importance of the predictive ML model, we elucidated that [Pb], [morph], [FAH], and [H_2O] have significant influence on the dimensionality control in the morph-Pb-I system. Using this data, along with DFT calculations, thermogravimetric measurements and mechanochemistry observations, we proposed a reaction scheme to explain the formation of the 1D and 2D phases. In the hypothesized scheme, water and formic acid accelerate the formation of the 1D phase via intra-chain hydrogen bonding, which is likely observed with other A-cations and MHP systems. Our strategy of using additives to control dimensionality has the potential to be applied in many other CMH systems. With the AL stopping criterion developed and tested in this work, the AL + HTE approach will be valuable for any materials research that benefits from predicting and controlling different phases/compounds in a vast reaction-composition space.

V.   Crystal Structure

The CIF files of morphPbI$_3$ and (morph)$_2$PbI$_4$ crystal structures have been submitted to the Cambridge Crystallographic Data Centre under Deposition Numbers 2110021 and 2110022, respectively.

VI.   References

(1)   Herz, L. M. Charge-Carrier Mobilities in Metal Halide Perovskites: Fundamental Mechanisms and Limits. *ACS Energy Lett.* **2017**, *2* (7), 1539–1548. https://doi.org/10.1021/acsenergylett.7b00276.

(2)   Zhao, Y.; Zhu, K. Organic–Inorganic Hybrid Lead Halide Perovskites for Optoelectronic and Electronic Applications. *Chem. Soc. Rev.* **2016**, *45* (3), 655–689. https://doi.org/10.1039/C4CS00458B.

(3)   Takahashi, Y.; Obara, R.; Nakagawa, K.; Nakano, M.; Tokita, J.; Inabe, T. Tunable Charge Transport in Soluble Organic–Inorganic Hybrid Semiconductors. *Chem. Mater.* **2007**, *19* (25), 6312–6316. https://doi.org/10.1021/cm702405c.

(4)   Leveillee, J.; Katan, C.; Even, J.; Ghosh, D.; Nie, W.; Mohite, A. D.; Tretiak, S.; Schleife, A.; Neukirch, A. J. Tuning Electronic Structure in Layered Hybrid Perovskites with Organic Spacer Substitution. *Nano Lett.* **2019**, *19* (12), 8732–8740. https://doi.org/10.1021/acs.nanolett.9b03427.

(5)   Correa-Baena, J.-P.; Abate, A.; Saliba, M.; Tress, W.; Jacobsson, T. J.; Grätzel, M.; Hagfeldt, A. The Rapid Evolution of Highly Efficient Perovskite Solar Cells. *Energy Environ. Sci.* **2017**, *10* (3), 710–727. https://doi.org/10.1039/C6EE03397K.

(6)   Correa-Baena, J.-P.; Saliba, M.; Buonassisi, T.; Grätzel, M.; Abate, A.; Tress, W.; Hagfeldt, A. Promises and Challenges of Perovskite Solar Cells. *Science* **2017**, *358* (6364), 739–744. https://doi.org/10.1126/science.aam6323.

(7)   Rational design of one-dimensional hybrid organic–inorganic perovskites with room-temperature ferroelectricity and strong piezoelectricity - Materials Horizons (RSC Publishing) https://pubs.rsc.org/en/content/articlelanding/2019/mh/c9mh00230h#!divAbstract (accessed 2020 -01 -21).

(8)   Wang, H.-C.; Bao, Z.; Tsai, H.-Y.; Tang, A.-C.; Liu, R.-S. Perovskite Quantum Dots and Their Application in Light-Emitting Diodes. *Small* **2018**, *14* (1), 1702433–1702455. https://doi.org/10.1002/smll.201702433.

(9)   Veldhuis, S. A.; Boix, P. P.; Yantara, N.; Li, M.; Sum, T. C.; Mathews, N.; Mhaisalkar, S. G. Perovskite Materials for Light-Emitting Diodes and Lasers. *Adv. Mater.* **2016**, *28* (32), 6804–6834. https://doi.org/10.1002/adma.201600669.

(10)   Kao, T. S.; Hong, Y.-H.; Hong, K.-B.; Lu, T.-C. Perovskite Random Lasers: A Tunable Coherent Light Source for Emerging Applications. *Nanotechnology* **2021**, *32* (28), 282001. https://doi.org/10.1088/1361-6528/abe907.

(11)   Seth, C.; Jana, D.; Jindal, V.; Khushalani, D.; Ghosh, S. One-Dimensional Behavior of Imidazolium Lead Iodide. *J. Phys. Chem. C* **2019**, *123* (26), 16449–16455. https://doi.org/10.1021/acs.jpcc.9b03894.

(12)   Stoumpos, C. C.; Cao, D. H.; Clark, D. J.; Young, J.; Rondinelli, J. M.; Jang, J. I.; Hupp, J. T.; Kanatzidis, M. G. Ruddlesden–Popper Hybrid Lead Iodide Perovskite 2D Homologous Semiconductors. *Chem. Mater.* **2016**, *28* (8), 2852–2867. https://doi.org/10.1021/acs.chemmater.6b00847.

(13)   Poglitsch, A.; Weber, D. Dynamic Disorder in Methylammoniumtrihalogenoplumbates (II) Observed by Millimeter-wave Spectroscopy. *J. Chem. Phys.* **1987**, *87* (11), 6373–6378. https://doi.org/10.1063/1.453467.

(14)   Gao, P.; Bin Mohd Yusoff, A. R.; Nazeeruddin, M. K. Dimensionality Engineering of Hybrid Halide Perovskite Light Absorbers. *Nat. Commun.* **2018**, *9* (1), 5028. https://doi.org/10.1038/s41467-018-07382-9.

(15)   Hong, K.; Le, Q. V.; Kim, S. Y.; Jang, H. W. Low-Dimensional Halide Perovskites: Review and Issues. *J. Mater. Chem. C* **2018**, *6* (9), 2189–2209. https://doi.org/10.1039/C7TC05658C.

(16)   Yusoff, A. R. bin M.; Nazeeruddin, M. K. Low-Dimensional Perovskites: From Synthesis to Stability in Perovskite Solar Cells. *Adv. Energy Mater.* **2018**, *8* (26), 1702073. https://doi.org/10.1002/aenm.201702073.

(17)   Kieslich, G.; Sun, S.; Cheetham, A. K. Solid-State Principles Applied to Organic–Inorganic Perovskites: New Tricks for an Old Dog. *Chem. Sci.* **2014**, *5* (12), 4712–4715. https://doi.org/10.1039/C4SC02211D.

(18)   Billing, D. G.; Lemmerer, A. Synthesis, Characterization and Phase Transitions in the Inorganic–Organic Layered Perovskite-Type Hybrids [(CnH2n + 1NH3)2PbI4], n = 4, 5 and 6. *Acta Crystallogr. B* **2007**, *63* (5), 735–747. https://doi.org/10.1107/S0108768107031758.

(19)   Billing, D. G.; Lemmerer, A. Synthesis, Characterization and Phase Transitions of the Inorganic–Organic Layered Perovskite-Type Hybrids [(CnH2n+1NH3)2PbI4] (n = 12, 14, 16 and 18). *New J. Chem.* **2008**, *32* (10), 1736–1746. https://doi.org/10.1039/B805417G.

(20)   Saparov, B.; Mitzi, D. B. Organic–Inorganic Perovskites: Structural Versatility for Functional Materials Design. *Chem. Rev.* **2016**, *116* (7), 4558–4596. https://doi.org/10.1021/acs.chemrev.5b00715.

(21)   Billing, D. G.; Lemmerer, A. Inorganic–Organic Hybrid Materials Incorporating Primary Cyclic Ammonium Cations: The Lead Iodide Series. *CrystEngComm* **2007**, *9* (3), 236–244. https://doi.org/10.1039/B618196A.

(22)   Billing, D. G.; Lemmerer, A. Inorganic–Organic Hybrid Materials Incorporating Primary Cyclic Ammonium Cations: The Lead Bromide and Chloride Series. *CrystEngComm* **2009**, *11* (8), 1549–1562. https://doi.org/10.1039/B819455F.

(23)   Chen, X.-G.; Song, X.-J.; Zhang, Z.-X.; Zhang, H.-Y.; Pan, Q.; Yao, J.; You, Y.-M.; Xiong, R.-G. Confinement-Driven Ferroelectricity in a Two-Dimensional Hybrid Lead Iodide Perovskite. *J. Am. Chem. Soc.* **2020**. https://doi.org/10.1021/jacs.0c03710.

(24)   Fateev, S. A.; Petrov, A. A.; Marchenko, E. I.; Zubavichus, Y. V.; Khrustalev, V. N.; Petrov, A. V.; Aksenov, S. M.; Goodilin, E. A.; Tarasov, A. B. FA2PbBr4: Synthesis, Structure, and Unusual Optical Properties of Two Polymorphs of Formamidinium-Based Layered (110) Hybrid Perovskite. *Chem. Mater.* **2021**, *33* (5), 1900–1907. https://doi.org/10.1021/acs.chemmater.1c00382.

(25)   Huang, J.; Yuan, Y.; Shao, Y.; Yan, Y. Understanding the Physical Properties of Hybrid Perovskites for Photovoltaic Applications. *Nat. Rev. Mater.* **2017**, *2* (7), 17042. https://doi.org/10.1038/natrevmats.2017.42.

(26)   Stach, E.; DeCost, B.; Kusne, A. G.; Hattrick-Simpers, J.; Brown, K. A.; Reyes, K. G.; Schrier, J.; Billinge, S.; Buonassisi, T.; Foster, I.; Gomes, C. P.; Gregoire, J. M.; Mehta, A.; Montoya, J.; Olivetti, E.; Park, C.; Rotenberg, E.; Saikin, S. K.; Smullin, S.; Stanev, V.; Maruyama, B. Autonomous Experimentation Systems for Materials Development: A Community Perspective. *Matter* **2021**, *4* (9), 2702–2726. https://doi.org/10.1016/j.matt.2021.06.036.

(27)   Liu, Y.; Niu, C.; Wang, Z.; Gan, Y.; Zhu, Y.; Sun, S.; Shen, T. Machine Learning in Materials Genome Initiative: A Review. *J. Mater. Sci. Technol.* **2020**, *57* (0), 113. https://doi.org/10.1016/j.jmst.2020.01.067.

(28)   Li, Z.; Najeeb, M. A.; Alves, L.; Sherman, A. Z.; Shekar, V.; Cruz Parrilla, P.; Pendleton, I. M.; Wang, W.; Nega, P. W.; Zeller, M.; Schrier, J.; Norquist, A. J.; Chan, E. M. Robot-Accelerated Perovskite Investigation and Discovery. *Chem. Mater.* **2020**, *32* (13), 5650–5663. https://doi.org/10.1021/acs.chemmater.0c01153.

(29)   Lookman, T.; Balachandran, P. V.; Xue, D.; Yuan, R. Active Learning in Materials Science with Emphasis on Adaptive Sampling Using Uncertainties for Targeted Design. *Npj Comput. Mater.* **2019**, *5* (1), 1–17. https://doi.org/10.1038/s41524-019-0153-8.

(30)   del Rosario, Z.; Rupp, M.; Kim, Y.; Antono, E.; Ling, J. Assessing the Frontier: Active Learning, Model Accuracy, and Multi-Objective Candidate Discovery and Optimization. *J. Chem. Phys.* **2020**, *153* (2), 024112. https://doi.org/10.1063/5.0006124.

(31)   Dai, C.; Glotzer, S. C. Efficient Phase Diagram Sampling by Active Learning. *J. Phys. Chem. B* **2020**, *124* (7), 1275–1284. https://doi.org/10.1021/acs.jpcb.9b09202.

(32)   Zhao, S.; Cai, T.; Zhang, L.; Li, W.; Lin, J. Autonomous Construction of Phase Diagrams of Block Copolymers by Theory-Assisted Active Machine Learning. *ACS Macro Lett.* **2021**, *10* (5), 598–602. https://doi.org/10.1021/acsmacrolett.1c00133.

(33)   Tian, Y.; Yuan, R.; Xue, D.; Zhou, Y.; Wang, Y.; Ding, X.; Sun, J.; Lookman, T. Determining Multi-Component Phase Diagrams with Desired Characteristics Using Active Learning. *Adv. Sci.* *n/a* (n/a), 2003165. https://doi.org/10.1002/advs.202003165.

(34)   Balachandran, P. V.; Kowalski, B.; Sehirlioglu, A.; Lookman, T. Experimental Search for High-Temperature Ferroelectric Perovskites Guided by Two-Step Machine Learning. *Nat Commun* **2018**, *9* (1), 1668. https://doi.org/10.1038/s41467-018-03821-9.

(35)   Xue, D.; Balachandran, P. V.; Yuan, R.; Hu, T.; Qian, X.; Dougherty, E. R.; Lookman, T. Accelerated Search for BaTiO3-Based Piezoelectrics with Vertical Morphotropic Phase Boundary Using Bayesian Learning. *Proc. Natl. Acad. Sci.* **2016**, *113* (47), 13301–13306. https://doi.org/10.1073/pnas.1607412113.

(36)   Kusne, A. G.; Yu, H.; Wu, C.; Zhang, H.; Hattrick-Simpers, J.; DeCost, B.; Sarker, S.; Oses, C.; Toher, C.; Curtarolo, S.; Davydov, A. V.; Agarwal, R.; Bendersky, L. A.; Li, M.; Mehta, A.; Takeuchi, I. On-the-Fly Closed-Loop Materials Discovery via Bayesian Active Learning. *Nat. Commun.* **2020**, *11* (1), 5966. https://doi.org/10.1038/s41467-020-19597-w.

(37)   Rohr, B.; Stein, H. S.; Guevarra, D.; Wang, Y.; Haber, J. A.; Aykol, M.; Suram, S. K.; Gregoire, J. M. Benchmarking the Acceleration of Materials Discovery by Sequential Learning. *Chem. Sci.* **2020**, *11* (10), 2696–2706. https://doi.org/10.1039/C9SC05999G.

(38)   Sun, S.; Tiihonen, A.; Oviedo, F.; Liu, Z.; Thapa, J.; Zhao, Y.; Hartono, N. T. P.; Goyal, A.; Heumueller, T.; Batali, C.; Encinas, A.; Yoo, J. J.; Li, R.; Ren, Z.; Peters, I. M.; Brabec, C. J.; Bawendi, M. G.; Stevanovic, V.; Fisher, J.; Buonassisi, T. A Data Fusion Approach to Optimize Compositional Stability of Halide Perovskites. *Matter* **2021**, *4* (4), 1305–1322. https://doi.org/10.1016/j.matt.2021.01.008.

(39)   Dahl, J. C.; Wang, X.; Huang, X.; Chan, E. M.; Alivisatos, A. P. Elucidating the Weakly Reversible Cs–Pb–Br Perovskite Nanocrystal Reaction Network with High-Throughput Maps and Transformations. *J. Am. Chem. Soc.* **2020**, *142* (27), 11915–11926. https://doi.org/10.1021/jacs.0c04997.

(40)   Hino, H. Active Learning: Problem Settings and Recent Developments. *ArXiv201204225 Cs Stat* **2020**.

(41)   Ueno, T.; Ishibashi, H.; Hino, H.; Ono, K. Automated Stopping Criterion for Spectral Measurements with Active Learning. *Npj Comput. Mater.* **2021**, *7* (1), 1–9. https://doi.org/10.1038/s41524-021-00606-5.

(42)   Ramalingam, K.; Rajaraman, T. Metal – Organic Hybrids of Tin(IV) with Tuneable Band Gap: Synthesis, Spectral, Single Crystal X-Ray Structural, BVS and CSM Analysis of Morpholinium Hexahalostannate(IV). *J. Mol. Struct.* **2020**, *1218*, 128489. https://doi.org/10.1016/j.molstruc.2020.128489.

(43)   Xiong, K.; Liu, W.; Teat, S. J.; An, L.; Wang, H.; Emge, T. J.; Li, J. New Hybrid Lead Iodides: From One-Dimensional Chain to Two-Dimensional Layered Perovskite Structure. *J. Solid State Chem.* **2015**, *230*, 143–148. https://doi.org/10.1016/j.jssc.2015.07.004.

(44)   Owczarek, M.; Jakubas, R.; Pietraszko, A.; Medycki, W.; Baran, J. Investigation of Structure–Properties Relationship in a Novel Family of Halogenoantimonates(III) and Halogenobismuthates(III) with Morpholinium Cation: [NH2(C2H4)2O]MX4. Crystal Structure, Phase Transitions and Dynamics of Molecules. *Dalton Trans.* **2013**, *42* (42), 15069–15079. https://doi.org/10.1039/C3DT51726H.

(45)   Owczarek, M.; Szklarz, P.; Jakubas, R.; Miniewicz, A. [NH2(C2H4)2O]MX5: A New Family of Morpholinium Nonlinear Optical Materials among Halogenoantimonate(III) and Halogenobismuthate(III) Compounds. Structural Characterization, Dielectric and Piezoelectric Properties. *Dalton Trans.* **2012**, *41* (24), 7285–7294. https://doi.org/10.1039/C2DT30291H.

(46)   Kawasaki, T.; Takahashi, M.; Ohhara, T.; Tanaka, I.; Kusaka, K.; Hosoya, T.; Yamada, T.; Kurihara, K. Structure of Morpholinium Tribromoplumbate C4H8ONH2PbBr3 Studied Using Single-Crystal Neutron Diffraction. *J. Phys. Soc. Jpn.* **2012**, *81* (9), 094602. https://doi.org/10.1143/JPSJ.81.094602.

(47)   Li, H.-H.; Chen, Z.-R.; Cheng, L.-C.; Wang, Y.-J.; Feng, M.; Wang, M. Hybrid Polymeric Iodoplumbates Constructed from Morpholine and Its Derivatives: Structures and Properties. *Dalton Trans.* **2010**, *39* (45), 11000–11007. https://doi.org/10.1039/C0DT00622J.

(48)   Li, H.-H.; Chen, Z.-R.; Cheng, L.-C.; Liu, J.-B.; Chen, X.-B.; Li, J.-Q. A New Hybrid Optical Semiconductor Based on Polymeric Iodoplumbate Co-Templated by Both Organic Cation and Polyiodide Anion. *Cryst. Growth Des.* **2008**, *8* (12), 4355–4358. https://doi.org/10.1021/cg800959s.

(49)   Corradi, A. B.; Bruni, S.; Cariati, F.; Ferrari, A. M.; Saccani, A.; Sandrolini, F.; Sgarabotto, P. Organic/Inorganic Composite Materials: Synthesis and Properties of One-Dimensional Polymeric Haloplumbate(II) Systems. *Inorganica Chim. Acta* **1997**, *254* (1), 137–143. https://doi.org/10.1016/S0020-1693(96)05156-0.

(50)   Yue, H.-L.; Sung, H.-H.; Chen, F.-C. Seeded Space-Limited Crystallization of CH3NH3PbI3 Single-Crystal Plates for Perovskite Solar Cells. *Adv. Electron. Mater.* **2018**, *4*, 1700655. https://doi.org/10.1002/aelm.201700655@10.1002/(ISSN)1614-6840.hottopic-solarcells.

(51)   Shi, C.; Ye, L.; Gong, Z.-X.; Ma, J.-J.; Wang, Q.-W.; Jiang, J.-Y.; Hua, M.-M.; Wang, C.-F.; Yu, H.; Zhang, Y.; Ye, H.-Y. Two-Dimensional Organic–Inorganic Hybrid Rare-Earth Double Perovskite Ferroelectrics. *J. Am. Chem. Soc.* **2019**. https://doi.org/10.1021/jacs.9b11697.

(52)   Saidaminov, M. I.; Abdelhady, A. L.; Murali, B.; Alarousu, E.; Burlakov, V. M.; Peng, W.; Dursun, I.; Wang, L.; He, Y.; Maculan, G.; Goriely, A.; Wu, T.; Mohammed, O. F.; Bakr, O. M. High-Quality Bulk Hybrid Perovskite Single Crystals within Minutes by Inverse Temperature Crystallization. *Nat. Commun.* **2015**, *6*, 7586–7592. https://doi.org/10.1038/ncomms8586.

(53)   Shi, D.; Adinolfi, V.; Comin, R.; Yuan, M.; Alarousu, E.; Buin, A.; Chen, Y.; Hoogland, S.; Rothenberger, A.; Katsiev, K.; Losovyj, Y.; Zhang, X.; Dowben, P. A.; Mohammed, O. F.; Sargent, E. H.; Bakr, O. M. Low Trap-State Density and Long Carrier Diffusion in Organolead Trihalide Perovskite Single Crystals. *Science* **2015**, *347* (6221), 519–522. https://doi.org/10.1126/science.aaa2725.

(54)   Ferrando, A.; Martínez Pastor, J. P.; Suárez, I. Toward Metal Halide Perovskite Nonlinear Photonics. *J. Phys. Chem. Lett.* **2018**, *9* (18), 5612–5623. https://doi.org/10.1021/acs.jpclett.8b01967.

(55)   Park, H.; Ha, C.; Lee, J.-H. Advances in Piezoelectric Halide Perovskites for Energy Harvesting Applications. *J. Mater. Chem. A* **2020**, *8* (46), 24353–24367. https://doi.org/10.1039/D0TA08780G.

(56)   Newnham, R. *Properties of Materials: Anisotropy, Symmetry, Structure*; Oxford University Press: Oxford, New York, 2005.

(57)   Li, Z. multi-well crystallization block by ZhiLBNL, https://www.thingiverse.com/thing:4927375 (accessed 2021 -09 -20).

(58)   Kennard, R. W.; Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **1969**, *11* (1), 137–148. https://doi.org/10.1080/00401706.1969.10490666.

(59)   Xie, J. *Karoka/Kennard-Stone-Algorithm*; 2021.

(60)   Pendleton, I. M.; Cattabriga, G.; Li, Z.; Najeeb, M. A.; Friedler, S. A.; Norquist, A. J.; Chan, E. M.; Schrier, J. Experiment Specification, Capture and Laboratory Automation Technology (ESCALATE): A Software Pipeline for Automated Chemical Experimentation and Data Management. *MRS Commun* **2019**, *9*, 846–859. https://doi.org/10.1557/mrc.2019.72.

(61)   Yangui, A.; Garrot, D.; Lauret, J. S.; Lusson, A.; Bouchez, G.; Deleporte, E.; Pillet, S.; Bendeif, E. E.; Castro, M.; Triki, S.; Abid, Y.; Boukheddaden, K. Optical Investigation of Broadband White-Light Emission in Self-Assembled Organic–Inorganic Perovskite (C6H11NH3)2PbBr4. *J. Phys. Chem. C* **2015**, *119* (41), 23638–23647. https://doi.org/10.1021/acs.jpcc.5b06211.

(62)   Zhou, Y.; Game, O. S.; Pang, S.; Padture, N. P. Microstructures of Organometal Trihalide Perovskites for Solar Cells: Their Evolution from Solutions and Characterization. *J. Phys. Chem. Lett.* **2015**, *6* (23), 4827–4839. https://doi.org/10.1021/acs.jpclett.5b01843.

(63)   Gu, E.; Tang, X.; Langner, S.; Duchstein, P.; Zhao, Y.; Levchuk, I.; Kalancha, V.; Stubhan, T.; Hauch, J.; Egelhaaf, H. J.; Zahn, D.; Osvet, A.; Brabec, C. J. Robot-Based High-Throughput Screening of Antisolvents for Lead Halide Perovskites. *Joule* **2020**, *4* (8), 1806–1822. https://doi.org/10.1016/j.joule.2020.06.013.

(64) Islam, T.; Sarker, M. Z. I.; Uddin, A. H.; Yunus, K. B.; D.M, R. P.; Mia, M. A. R.; Ferdosh, S. Kamlet Taft Parameters: A Tool to Alternate the Usage of Hazardous Solvent in Pharmaceutical and Chemical Manufacturing/Synthesis - A Gateway towards Green Technology. *Anal. Chem. Lett.* **2020**, *10* (5), 550–561. https://doi.org/10.1080/22297928.2020.1860124.

(65) Nega, P. W.; Li, Z.; Ghosh, V.; Thapa, J.; Sun, S.; Hartono, N. T. P.; Nellikkal, M. A. N.; Norquist, A. J.; Buonassisi, T.; Chan, E. M.; Schrier, J. Using Automated Serendipity to Discover How Trace Water Promotes and Inhibits Lead Halide Perovskite Crystal Formation. *Appl. Phys. Lett.* **2021**, *119* (4), 041903. https://doi.org/10.1063/5.0059767.

(66) Kamlet-Taft solvent parameters http://www.stenutz.eu/chem/solv26.php (accessed 2021 -05 -17).

(67) Settles, B. Active Learning Literature Survey. *University of Wisconsin–Madison* **2010**, 67.

(68) Hinton, G.; Roweis, S. Stochastic Neighbor Embedding. 8.

(69) Zhdanov, F. Diverse Mini-Batch Active Learning. 9.

(70) MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. *Proc. Fifth Berkeley Symp. Math. Stat. Probab. Vol. 1 Stat.* **1967**, 281–297.

(71) Vlachos, A. A Stopping Criterion for Active Learning. *Comput. Speech Lang.* **2008**, *22* (3), 295–312. https://doi.org/10.1016/j.csl.2007.12.001.

(72) Zhu, J.; Wang, H.; Hovy, E.; Ma, M. Confidence-Based Stopping Criteria for Active Learning for Data Annotation. *ACM Trans. Speech Lang. Process.* **2010**, *6* (3), 1–24. https://doi.org/10.1145/1753783.1753784.

(73) Yang, Y.; Loog, M. A Variance Maximization Criterion for Active Learning. *Pattern Recognit.* **2018**, *78*, 358–370. https://doi.org/10.1016/j.patcog.2018.01.017.

(74) Cortes, C.; Jackel, L. D.; Chiang, W.-P. Limits on Learning Machine Accuracy Imposed by Data Quality. **1995**, 6.

(75) 4.2. Permutation feature importance — scikit-learn 0.24.2 documentation https://scikit-learn.org/stable/modules/permutation_importance.html#outline-of-the-permutation-importance-algorithm (accessed 2021 -06 -03).

(76) Efron, B.; Tibshirani, R. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Stat. Sci.* **1986**, *1* (1), 54–75. https://doi.org/10.1214/ss/1177013815.

(77) Lehmann, E. L.; Romano, J. P. *Testing Statistical Hypotheses*, 3rd ed.; Springer Texts in Statistics; Springer-Verlag: New York, 2005. https://doi.org/10.1007/0-387-27605-X.

(78) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Phys. Rev. Lett.* **2008**, *100* (13), 136406. https://doi.org/10.1103/PhysRevLett.100.136406.

(79) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): Assessing the Accuracy of DFT Formation Energies. *Npj Comput. Mater.* **2015**, *1* (1), 1–15. https://doi.org/10.1038/npjcompumats.2015.10.

(80) Do, J.-L.; Friščić, T. Mechanochemistry: A Force of Synthesis. *ACS Cent. Sci.* **2017**, *3* (1), 13–19. https://doi.org/10.1021/acscentsci.6b00277.

(81) Guo, Y.; Shoyama, K.; Sato, W.; Matsuo, Y.; Inoue, K.; Harano, K.; Liu, C.; Tanaka, H.; Nakamura, E. Chemical Pathways Connecting Lead(II) Iodide and Perovskite via Polymeric Plumbate(II) Fiber. *J. Am. Chem. Soc.* **2015**, *137* (50), 15907–15914. https://doi.org/10.1021/jacs.5b10599.

(82) Zhang, K.; Wang, Z.; Wang, G.; Wang, J.; Li, Y.; Qian, W.; Zheng, S.; Xiao, S.; Yang, S. A Prenucleation Strategy for Ambient Fabrication of Perovskite Solar Cells with High Device Performance Uniformity. *Nat. Commun.* **2020**, *11* (1), 1006. https://doi.org/10.1038/s41467-020-14715-0.

(83) Xiao, S.; Zhang, K.; Zheng, S.; Yang, S. Good or Evil: What Is the Role of Water in Crystallization of Organometal Halide Perovskites? *Nanoscale Horiz.* **2020**, *5* (8), 1147–1154. https://doi.org/10.1039/D0NH00270D.

Supporting Information

Dimensional control over metal halide perovskite crystallization guided by active learning

Zhi Li,[§] Philip W. Nega,[§] Mansoor Ani Najeeb,[‡] Chaochao Dun,[§] Matthias Zeller,[#] Jeffrey J. Urban,[§] Wissam A. Saidi,[°] Joshua Schrier, [‡†] Alexander J. Norquist,[‡] Emory M. Chan*[§]

[§]Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA

[‡]Department of Chemistry, Haverford College, 370 Lancaster Avenue, Haverford, Pennsylvania 19041, USA

[†]Department of Chemistry, Fordham University, 441 E. Fordham Road, The Bronx, New York 10458, USA

[#]Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette IN 47907, USA

[°]Department of Mechanical Engineering and Materials Science, University of Pittsburgh, Pittsburgh, PA 15261, USA

*To whom correspondence should be addressed: EMChan@lbl.gov.

Table of Contents

1. Materials and methods

Chemicals

Lead iodide (PbI$_2$) (99%), formic acid (FAH) ($\geq$95 %), dimethylformamide (DMF) (99.8 %), dimethyl sulfoxide (DMSO) ($\geq$99.5 %), and dichloromethane (DCM) ($\geq$99.8 %) were purchased from Sigma Aldrich Chemicals. γ-Butyrolactone (GBL) ($\geq$98 %) was purchased from Spectrum Chemical. Morpholinium iodide (morphI) (98 %) was purchased from GreatCell Solar.

Equipment

Liquid handling robot

A Hamilton Microlab NIMBUS4 liquid handling robot was used in this study for the HT-ASVC synthesis of morpholinium lead iodide. The robot features four independent pipetting channels for transferring liquid. The pipettors aspirate reagent stock solutions stored in polypropylene containers organized in racks placed in programmatically defined positions on the robot deck. Stock solutions were used on the same day as they were prepared (within 8 hours) to avoid any possible solution degradation. New pipette tips were used for each stock solution. Solutions for HT-ASVC reactions were prepared on a Hamilton Heater and Shaker (HHS) module, which can be heated up to 105 °C (actual temperature solution temperature reaches 95 °C) and can vortex microplates up to 2000 rpm. Robotic protocols were programmed in the Hamilton Method Editor software; reaction time, shaking speed etc., were imported from Microsoft Excel.xls spread sheets generated by ESCALATE. Complete, step-by-step synthetic protocols are shown in the "Robotic Workflow" section.

Powder X-ray Diffraction

Powder X-ray diffraction (pXRD) measurements were performed on a Bruker AXS D8 Discover GADDS X-Ray Diffractometer, which is equipped with a Vantec-500 area detector and is operated at 35kV/40mA with a Co Kα radiation source with a wavelength of 1.79Å.

UV-visible absorption spectra

UV-visible (Vis) absorption spectra were collected using an Agilent Cary-5000 UV-Vis-NIR spectrophotometer. Absorbance spectra of powders ground from morpholinium lead iodide crystals were measured using an internal diffuse reflectance accessory.

Photoluminescence spectra

Photoluminescence (PL) spectra of powders ground from morpholinium lead iodide crystals were measured using a Horiba Jobin Yvon Fluorolog-3 spectrofluorometer. PL spectra were collected from 530 to 720 nm with 1 nm wavelength steps and 0.01 s integration time per step.

Thermogravimetric analysis

Thermogravimetric analysis were performed using TA Instruments Q5500 TGA-MS. The sample weight change was measured from room temperature to 450 °C with a ramp rate of 10 °C /min under nitrogen.

High-throughput powder X-ray Diffraction

High-throughput pXRD Measurements were performed on glass slides in a customized sample holder on the same Bruker X-Ray Diffractometer. The sample locations (e.g., A1) are programmatically defined in pXRD software. The diffractometer automatically performs pXRD measurements and switches samples until all samples are finished.

High-throughput UV-Vis absorption spectra

High-throughput UV-Vis-NIR absorption spectra were collected using a custom-built reflection-mode UV-Vis-NIR absorption spectrometer. The spectra were measured from 350 nm to 2500 nm and averaged over 100 acquisitions (each acquisition takes one second). A motorized XY stage enables automated measurement on samples in a 96-well microplate.

High-throughput photoluminescence spectra

High-throughput PL spectra were collected using a Biotek Synergy 4 UV-Vis Absorption/Fluorescence Microplate Reader. The PL spectra were measured from 540 nm to 720 nm with 450 nm excitation.

High-throughput microscopic images

High-throughput optical micrographs were acquired using a Biotek Cytation 5 Cell Imaging Multi-Mode Reader with a 4x objective lens. Bright-field and fluorescent (excitation at 469 nm; emission at 525 nm and 593 nm) images are measured for solid products in a 96-well Hellma quartz microplate.

Benchtop ASVC synthesis

A precursor solution was made by dissolving 1.6 mmol morphI, 1.6 mmol PbI_2, and 50 μL formic acid in 1 mL DMF in a 4 mL clear scintillation vial. The precursor solution was heated and stirred at 75 °C and 450 rpm for 1 hour. After dissolution, the 4 mL vial (uncapped) containing 1 mL precursor solution was cooled to room temperature and placed in a 20 mL scintillation vial. Then 5 mL DCM was added between the 4 mL vial and 20 mL vial. Care was taken not to add DCM in the precursor solution. Then the 20 mL vial was sealed by a PTFE-lined solid-top storage cap. The reaction was stored in the dark for 16 hours without disturbance. Large yellow-colored crystals formed after 16 hours of crystallization. The crystals were isolated via vacuum filtration. Then they were dried in a desiccator before characterization.

Robotic Workflow

Solubility measurement

The solubility of PbI_2 depends on the concentration of coexisting morphI in the solution. To determine the solubility of PbI_2 for HT-ASVC reactions, we weighed fixed masses of PbI_2 with different percentages of morphI and stirred the suspensions in a heated oil bath at 75 °C and 450 rpm. DMF was added gradually until the solid was completely dissolved. The solution was then brought to room temperature to ensure its stability. This step also ensured that the solution remained a homogeneous solution during the robotic run and that the PbI_2 did not precipitate, which can disrupt the pipetting of liquid handler. The solubility was then calculated as total moles of PbI_2 divided by the total volume of the solutions. The solubility of PbI_2 in DMF is 2.32 M (mixed with 2.91 M morphI). Using the same method, the solubility of morphI in DMF was calculated as 2.36 M.

Overview of HT-synthesis and HT-characterization workflows

Stock solutions of PbI_2-morphI mixture solution and morphI solution were prepared based on experimental data entry files generated by ESCALATE using the solubility data. The concentrations of PbI_2-morphI mixture solution were 2.32 M (PbI_2) and 2.91 M (morphI). The concentration of morphI-only solution was 2.36 M. After stock solution preparation, all reaction components (i.e., the stock solutions, pure solvents, and additives) were placed in programmatically designated locations on NIMBUS operation deck. During the addition of all reaction components, the reaction solutions were kept at an actual temperature of 75 ºC (measured by an IR camera at a set point temperature of 80 ºC) for dissolution. A synthetic flow chart is shown in Figure S3, describing stock solution preparation and robotic procedures. Figure S3 contains a photograph of NIMBUS operation deck, labeled with the locations of stock solutions. A customized robot-compatible crystallization block (circled in white dash square), containing 24 pairs of wells, was placed on the Hamilton Heater and Shaker (HHS). For one pair of wells in the block, perovskite stock solution occupied one well, and the other well was taken by anti-solvent (DCM). The 8x43 mm (diameter x height) glass scintillation vials were used as reaction and anti-solvent vessels. The vials were pre-heated to 55 ºC before addition of the stock solutions. Formic acid was added to each reaction vial, followed by 15 minutes shaking to avoid premature precipitation of PbI_2, morphI, or perovskite. After vortexing, the crystallization block was cooled to room temperature before DCM was added to the anti-solvent wells in the crystallization block. After DCM addition, we sealed the block with a metal cap and stored the block at 20 ºC without disturbance for 16 h. The actual, step-by-step synthetic protocol of the robot is given in the below section.

After crystallization was completed, we opened the crystallization block and took side photos of each reaction vial. Based on visual inspection, we recorded the morphologies and colors of solid products. Then we labeled the reaction outcomes by the colors of the major products. The examples of side photos, notes, and scores are shown in Table S1 in the "**Product Scoring Rubric Based on Human-Inspection**" section.

The solid products were filtered and dried. Then we took HT bright-field and fluorescent images of the products in a quartz-bottom 96-well microplate (Hellma Analytics) using a Biotek Cytation 5 Cell Imaging Multi-Mode Reader (Figure S5). After the microscopic imaging, solid products were ground to powder and placed in an opaque round-bottom 96-well microplate. Then HT UV-Vis absorption spectra were measured using a custom-built automated reflection-mode UV-Vis-NIR absorption spectrometer. Representative absorption spectra are shown in Figure S6a. HT PL spectra of the powders were collected in the same microplate using a Biotek Synergy 4 plate reader (Figure S6b). For HT pXRD, the powder samples need to be transferred to a flat surface. After HT PL measurement, the microplate was covered by ten layers of nonwoven wipes (55% cellulose and 45% polyester) and then a plastic plate. The microplate was flipped, and the powder samples were knocked out to the top layer of the nonwoven wipes. Then the microplate was lifted and removed. Care should be taken not to move the powder samples on the wipe. A glass slide, covered with double-sided tapes on one side, was stamped on the powder samples on the nonwoven wipe (with the taped side down). Then the wipe was peeled off, leaving the powder samples on the double-sided tapes on the glass slide. The powder samples on the glass slide kept the same corresponding locations as they have in the round-bottom microplate. Then HT pXRD was performed on the samples. The whole HT-characterization workflow is illustrated in the flow charts in Figure S4.

Step-by-step synthetic protocols
Primary screening in the reaction-composition space of [Pb], [morph], and [FAH]

1. Stock solutions are manually prepared with concentrations specified in the "overview" section and are named Reagents 1-5 in order of their addition during the liquid handling process. Reagents 1, 4, and 5 are pure DMF, formic acid, and DCM, respectively. They are used without any additional treatment. Reagent 2, a stock solution of PbI_2-morphI in DMF, is prepared by adding PbI_2, morphI and then DMF to the stock solution container (in that order). Reagent 3 is the morphI solution, prepared by adding DMF to morphI. Reagents 2 and 3 stock solutions are prepared in glass containers and stirred in a heated oil bath at 75 °C and 450 rpm for one hour to completely dissolve any solids. Reagents 2 and 3 are then cooled to room temperature, where they should remain a clear solution. Then, all reagent solutions are manually loaded into designated locations on liquid handler deck and kept at room temperature. Reagents 1 and 5 are stored in reusable 50 mL polypropylene solvent containers. Sets of four, 15 mL Falcon polypropylene centrifuge tubes are used as containers for Reagents 2-4.

2. When the robot protocol is initiated, the crystallization block is pre-heated to a set point temperature at 80 ºC, with the actual temperature of the HHS reactor module measured and recorded in the software. The heating rate is approximately 7 ºC/ min.

3. After the measured temperature of HHS reaches 80 ºC, the liquid handler dispenses Reagent 1 (0-300 µL), into each of the 24, 1-mL glass vials (in "precursor solution" positions) in the crystallization block, followed by Reagent 2 (0-300 µL) and Reagent 3 (0-300 µL). The temperature of the HHS will reach 80 ºC (75 ºC for solution) during dispensing of Reagents 1-3. The liquid dispensing process is programmed and automatically carried out without any delay time between each dispense.

4. After addition of Reagents 1-3, 0-300 μL formic acid (Reagent 4) is dispensed into the 24 vials. The combined volume of Reagents 1-4 dispensed in each vial is maintained at 300 μL. The addition of formic acid is followed by 15 minutes of shaking at 500 rpm.
5. After the 15 minutes of shaking, heating is stopped, and the perovskite precursor solution is cooled with a fan to 25 ºC in ~50 minutes.
6. DCM (800 μL) is added into the 24 vials in "anti-solvent" positions in the crystallization block. After DCM addition is completed, the block is manually sealed with a metal cap and screws. The crystallization block is then stored at 20 ºC without disturbance for 16 h.

Screening in the reaction-composition space of [Pb], [morph], [FAH], [H$_2$O], V$_{f, DMSO}$, and V$_{f, GBL}$.
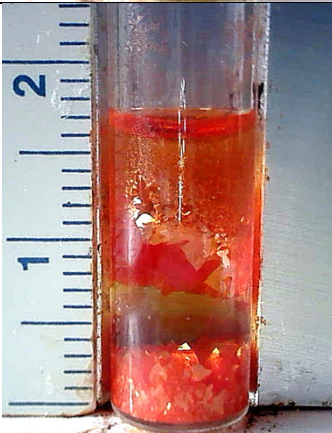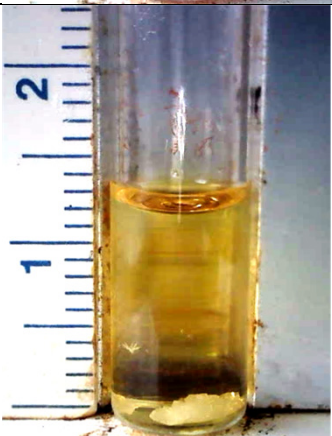
1. Stock solutions are manually prepared as Reagents 1-8 in order of their addition during the liquid handling process. Reagents 1, 2, 3, 6, 7, and 8 are pure DMF, DMSO, GBL, DI water, and DCM, respectively. They are used without any additional treatment. Reagents 4 and 5 are stock solutions of PbI$_2$-morphI and morphI in DMF, and they are prepared in the same way as in primary screening. All reagent solutions are manually loaded into designated locations on liquid handler deck and kept at room temperature. Reagents 1, 2, 3, 7, 8 are stored in reusable 50 mL polypropylene solvent containers. Reagents 4-6 are stored in 15 mL Falcon centrifuge tubes.
2. This step is the same as step 2 in primary screening.
3. This step is similar to step 3 in primary screening. The liquid handler dispenses Reagents 1-5 (0-300 μL) into the glass vials (in "precursor solution" positions) in the crystallization block.
4. After addition of Reagents 1-5, formic acid (Reagent 6) and water (Reagent 7) are dispensed into the 24 vials. The combined volume of Reagents 1-7 dispensed in each vial is maintained at 300 μL. The addition of water is followed by 15 minutes of shaking at 500 rpm.

   Step 5 and 6 are the same as primary screening.

Product Scoring Rubric Based on Human Inspection

To score the reaction outcomes of primary screening, reaction vials were inspected at different angles by eye. In some reactions, there was a mixture of both yellow- and red-colored solid products, so we labeled the reaction outcomes by the colors of the major products. Thus, there are three classes of reaction outcomes in the primary screening: (1) class 1, clear solution without any solid; (2) class 2, red phase; (3) class 3, yellow phase.

Crystal Scores	Criteria	Notes	Photos

1	**No solid** was observed in the solutions.	NA	
2	**Red-colored crystals or powder** are the major solid products.	Large-red-flake crystals	
3	**Yellow-colored crystals or powder** are the major solid products.	Bulky-yellow crystals	

Software and Computation

Software and packages

Our custom-developed pipeline software: ESCALATE[1], was used to specify experimental parameters in robot readable files, provide instructions for human operators, and capture experiment results and observations. All algorithms in this work were written in Python 3.6 in Jupyter notebooks using the following libraries: Numpy 1.18.0, Pandas 0.22.0, Scipy 1.3.0, Matplotlib 3.1.0, Scikit-learn 0.21.3, and modAL 0.3.5.

Five-fold cross-validation

We used a "Stratified Shuffle Split" method from Scikit-learn to generate training/testing datasets for cross-validation (CV) of machine learning models. In the case of 5-fold cross-validation, there are 5 different train/test splits on the dataset: in each split, 80 % of the data were used to train the machine learning model, while 20% of the data were reserved for testing. The testing sets were randomly drawn from whole datasets in a stratified style (i.e., testing sets have the same percentage of samples of each target class as the whole datasets). Before each drawing, the datasets were shuffled, so the testing datasets are not necessarily exclusive between splits.

Machine learning metrics

CV accuracies were calculated by averaging the prediction accuracies of five different train/test splits created by CV on the dataset. The prediction accuracy is defined in the equation below.

Accuracy = (TP + TN)/(TP + TN + FP + FN)

Here, TP is the total number of "True Positive" results (i.e., the predicted and actual scores are both "positive"). Likewise, "FP" is the number of "False Positive" results, "TN" is the number of "True Negative" results, and "FN" is the number of "False Negative" results.

Our experimental results were interpreted as either two-class or three-class classification problems. In the two-class case, "yellow phase" outcome was considered "positive" result while "red phase" outcome was considered "negative" result. The "clear solution" outcomes were excluded. The prediction accuracy was calculated using the above equation directly.

In the three-class case, we calculated the overall prediction accuracy by averaging the prediction accuracies from all possible one-vs-all classifications [e.g., yellow phase (positive) vs. non-yellow phase (negative)].

Grid-point generation

Grids were generated as cuboids/hypercuboids in reaction-composition space. The ranges of the cuboid/hypercuboid' sides are the allowed ranges of the concentration/volume fraction (Table S4 and S5). The interval number of a grid is n. The total numbers of grid points are n^3 and n^6 for the primary screening and 6D screening.

Kennard-Stone algorithm

Kennard-Stone (KS) algorithm was used to generate uniform sampling pools from candidate pools in reaction-composition space. Before performing KS sampling, the features of the candidate pools were scaled by standardization (subtracting the mean and dividing the result by the standard deviation). The sampling algorithm works as follows:

1. First, two points with the largest Euclidean distance (p_1 and p_2) are selected as the initial points of the sampling pool and removed from the candidate pool.
2. The separation distance between a candidate point and the sampling pool is defined as the distance from the candidate point to its closest sampling point (p_1 or p_2).
3. The candidate point with the largest separation distance (p_3) is selected and added to the sampling pool. This step is repeated until the sampling pool reaches the required size k.

Selecting 48 primary reactions in 3D reaction-composition space

1. First, we constructed a 3D convex hull[2] (C_{3D}) of the allowed reaction-composition space of [Pb], [morph], and [FAH]. Pure FAH, Reagent 2, and Reagent 3 were used as vertices to construct an initial convex hull, and then the concentration constraints (Table S4) are applied on the initial convex hull to generate C_{3D}.
2. Then, we generated a 3D grid (with interval number n) in the smallest cuboid containing C_{3D}. The number of grid points contained in C_{3D} is defined as N_C. We chose n that minimized N_C while kept $N_C \geq 48$. The calculated n and N_C are 9 and 51.
3. Last, using the Kennard-Stone algorithm, we selected 48 data points from the 51 grid points in C_{3D} as our primary sampling points.

Generating the reaction pool in 6D reaction-composition space

We constructed a 6D convex hull (C_{6D}) of the allowed reaction-composition space of [Pb], [morph], [FAH], [H_2O], $V_{f, DMSO}$, and $V_{f, GBL}$. The constraints of the experimental variables are listed in Table S5. Then we generated a 6D grid (interval number $n = 20$, see "*Grid-point generation*") in the smallest hypercuboid containing C_{6D}. The grid points contained in C_{6D} are defined as the reaction pool (the grid points outside C_{6D} are excluded since they are inaccessible to our experimental workflow). There are 469,326 points in the reaction pool.

Diverse-mini-batch active learning

To initiate AL, 24 reactions were selected from the reaction pool using the KS algorithm. The RF model was trained on the 24 initial reactions and then generated prediction uncertainties (U) for all the reactions in the reaction pool. The reactions with $U > 0.5$ were prefiltered, and the k-mean clustering algorithm was used to divide the prefiltered reactions into 24 clusters ($k = 24$). In k-mean clustering, the weight of a reaction was its U value. After clustering, the centroids of the clusters were determined. Then 24 different reactions closest to the centroids were selected for the next AL iteration. The AL process continued until the stopping criterion was met.

Random-classification accuracy

The random-classification accuracy was calculated as one minus the Gini impurity of the dataset (after the 5th AL run). The Gini impurity (g) is defined below.

$$g = \sum_{i=1}^{3} f_i(1 - f_i)$$

f_i is the number fraction of ith class in the dataset. For the dataset after the 5th AL run, f_1, f_2, and f_3 are 0.41, 0.33, and 0.26 respectively. Therefore, the Gini impurity and random-classification accuracy are 0.66 and 0.34.

Theoretical limit of CV accuracy

The theoretical limit of CV accuracy of any ML models on the tested dataset (initial sampling + 5 AL runs) was calculated as one minus intrinsic error of the dataset. The intrinsic error was estimated as one minus the average CV accuracy of multiple deep neural networks that overfit the dataset. The hidden layer numbers of the deep neural networks are 40, 48, 57, 69, 83, and 100. The CV accuracies of these models were calculated using the "leave-one-out" strategy.

Number of hidden layers	Hyper-parameters (fixed)	Intrinsic error	Average intrinsic error
40		0.27	
48		0.22	
57	activation = 'relu'	0.24	
69	solver = 'lbfgs'	0.26	0.24 ± 0.02
83	alpha = 0.001,	0.21	
100		0.26	

Selecting 24 reactions far from initial sampling + 5 AL runs

We selected 24 reactions located far from the 168 tested reactions (initial sampling + 5 AL runs) using the KS algorithm in the 6D reaction-composition space. In the KS sampling, the tested reactions were treated as the initial points of the sampling pool. The reaction, with the largest Euclidean distance to the sampling pool, was picked from the candidate pool and added to the sampling pool. The selection process continued until all 24 reactions were picked and added to the sampling pool. Then the newly added 24 reactions in the sampling pool were performed to test the RF model's predicting ability on unseen reactions.

Synthetic dataset

The synthetic dataset was generated using the "*make_classification*" function in the Scikit-learn library. All parameters of the synthetic dataset are shown in the table below. The synthetic dataset has six input features, three output classes, and the same weight distribution of classes as our experimental dataset, which is [0.43, 0.29, 0.28] for Class 1, 2, and 3, respectively. The dataset has 50,000 data points. The "class_sep" is set as 0.8 to create overlaps between the classes (the larger the value is, the less the overlap will be), and "flip_y" = 0.1 means that there are 10% of data points whose classes are assigned randomly.

The same AL experiment was performed on the synthetic dataset: 1) there were 48 data points selected by the KS algorithm to initiate the AL process; 2) AL queried 24 data points in each iteration.

Library	module	function	parameters
Scikit-learn	dataset	*make_classification*	n_samples = 50000, n_features = 6, n_redundant = 0, n_repeated = 0, n_informative = 6, n_classes = 3, n_clusters_per_class = 1, weights = [0.43, 0.29, 0.28], class_sep = 0.8, random_state = 1, flip_y = 0.1

Permutation-feature-importance analysis
The permutation-feature-importance is defined as the decrease in a ML model's prediction score when the value of a feature is randomly shuffled. Since the shuffling breaks relationship between the feature and the outcome, the decrease in the model score is indicative of how much the model depends on that feature. In this work, we calculated the reduction of the five-fold CV accuracy of the RF model on the "crystals only" dataset after each feature was randomly shuffled.

Bootstrapping the "crystals only" dataset
We used bootstrapping to create 1000 samples from the "crystals only" dataset. Each sample was randomly selected and had the same size as the "crystals only" dataset. In each sampling, the same reaction could be drawn repeatedly.

One-tailed statistical hypothesis testing
To validate the effects of the important features (i.e., [Pb], [morph], [FAH], [H_2O]) on yellow/red phase formation, we performed one-tailed statistical hypothesis tests with a confidence level of 0.99 (i.e., significance level = 0.01). A critical value of $Z_0 = 2.8$ (the value to separate the "rejection" and "fail to reject" regions for null hypothesis) was calculated from a normal probability distribution using the confidence level after Bonferroni correction [we are testing four hypotheses at the same time, so the confidence level used for calculating Z_0 should be 1-(0.01)/4]. The actual Z value for each important feature (Z_c) was calculated using the mean (μ) and standard deviation (σ) of the 1000 slopes (from bootstrapping) with the below equation.

$$Z_c = \frac{(\mu - 0)}{(\sigma/\sqrt{1000})}$$

For yellow phase formation,

Feature	Hypothesized effect on yellow phase formation	Null hypothesis	Type of test	μ	σ	Z_c	Conclusion
[morph]	Negative	Positive or none	Left tail	-2.33	0.24	$-310 < -Z_0$	Reject null hypothesis
[Pb]	Positive	Negative or none	Right tail	2.37	0.16	$471 > Z_0$	Reject null hypothesis
[FAH]	Positive	Negative or none	Right tail	1.09	0.24	$139 > Z_0$	Reject null hypothesis
[H_2O]	Positive	Negative or none	Right tail	1.51	0.22	$224 > Z_0$	Reject null hypothesis

For red phase formation,

Feature	Hypothesize effect on red phase formation	Null hypothesis	Type of test	μ	σ	Z_c	Conclusion
[morph]	Positive	Negative or none	Right tail	2.33	0.24	$299 > Z_0$	Reject null hypothesis
[Pb]	Negative	Positive or none	Left tail	-2.36	0.15	$-498 < -Z_0$	Reject null hypothesis
[FAH]	Negative	Positive or none	Left tail	-1.09	0.24	$-141 < -Z_0$	Reject null hypothesis
[H_2O]	Negative	Positive or none	Left tail	-1.53	0.21	$-228 < -Z_0$	Reject null hypothesis

Density functional theory (DFT) calculations

The first-principles DFT calculations were carried out within the Perdew-Burke-Ehrenzhof exchange-correlation functional revised for solids (PBEsol).[3] Because of the importance of van der Waals interactions in the hybrid perovskites[4], we applied the Tkatchenko-Scheffler scheme to correct the PBEsol energies.[5,6] The electron-nucleus interactions were modeled using the projector augmented wave (PAW) pseudopotentials[7,8] as implemented in the Vienna Ab initio Simulation Package (VASP) package. Due to the importance of computational setup, planewave cutoff, kgrid, and energy tolerances were properly chosen, as tested before.[9] In our calculations, we relaxed the atomic positions and the lattice while constraining the system to maintain the initial lattice symmetry to avoid complications from shallow minima that exist in the potential energy surface.[10] Thermodynamic stability was assessed by accounting for vibrational energy as computed using phonopy.
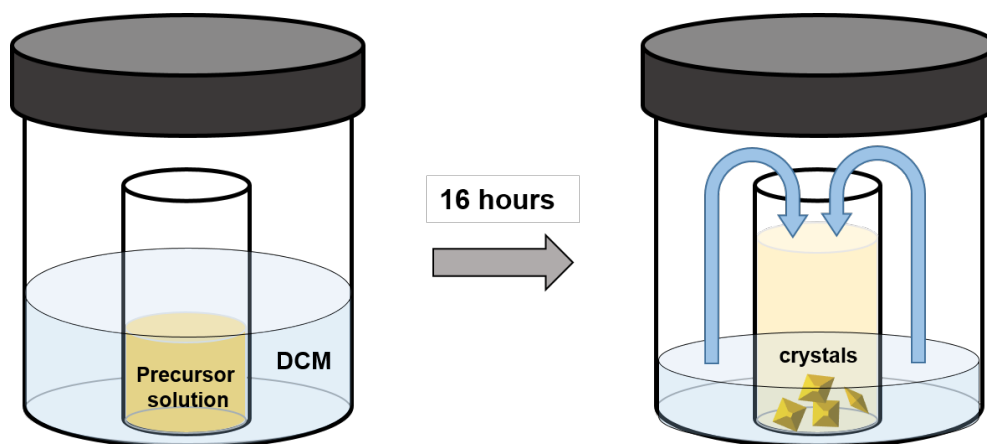
2. Supplemental Figures and Tables

Figure S1. Illustration of benchtop ASVC of morpholinium lead iodide.

Single Crystal Structure Refinement Details

Single crystals of morphPbI$_3$ and (morph)$_2$PbI$_4$ were transferred to the goniometer head of a Bruker Quest diffractometer with a fixed chi angle, a sealed tube fine focus X-ray tube, single crystal curved graphite incident beam monochromator, a Photon100 area detector and an Oxford Cryosystems low-temperature device. Examination and data collection were performed with Mo Kα radiation ($\lambda = 0.71073$ Å) at 100(2) K.

Data were collected, reflections were indexed and processed, and the files were scaled and corrected for absorption using APEX3.[11] Heavy atom positions were determined using SIR92.[31] All other non-hydrogen sites were located from Fourier difference maps. All non-hydrogen sites were refined using anisotropic thermal parameters with full matrix least squares procedures on F_o^2 with $I > 3\ (I)$. Hydrogen atoms were placed in geometrically idealized positions. All calculations were performed using Crystals *v.* 14.23*c*.[32] Complete crystallographic data, in CIF format, has been deposited with the Cambridge Crystallographic Data Centre. CCDC 210021 and 2110022 contain the supplementary crystallographic data for this paper. These data can be obtained free of

Table S1. Crystallographic data for $[C_4H_{10}NO][PbI_3]$ (yellow phase) and $[C_4H_{10}NO]_2[PbI_4]$ (red phase).

Compound	$[C_4H_{10}NO][PbI_3]$	$[C_4H_{10}NO]_2[PbI_4]$
Formula	$C_4H_{10}I_3NOPb$	$C_8H_{20}I_4N_2O_2Pb$
fw	676.04	891.08
Space-Group	$P2_12_12_1$ (no. 19)	$C2/c$ (no. 15)
a / Å	8.1767(3)	9.3920(4)
b / Å	8.7039(4)	8.6729(3)
c / Å	16.5225(7)	23.2152(9)
α / º	90	90
β / º	90	94.5715(13)
γ / º	90	90
V / Å^3	1175.89(9)	1885.00(13)
Z	4	4
ρ_{calc} / g cm^{-3}	3.818	3.140
λ / Å	0.71073	0.71073
T / K	100(2)	100(2)
μ / mm^{-1}	22.177	15.497
R_1,[a]	0.0235	0.0248
wR_2[b]	0.0513	0.0583

[a] $R_1 = \Sigma \left\| F_o \right| - \left| F_c \right\| / \Sigma \left| F_o \right|$. [b] $wR_2 = [\Sigma w(F_o{}^2 - F_c{}^2)^2 / [\Sigma w(F_o{}^2)^2]^{1/2}$.

Table S2a. Selected bond lengths (Å) in $[C_4H_{10}NO][PbI_3]$.

Bond	Length (Å)
Pb1 – I1	3.3096(4)
Pb1 – I1	3.1525(4)
Pb1 – I2	3.2443(4)
Pb1 – I3	3.4061(4)
Pb1 – I3	3.0831(4)

Table S2b. Selected bond angles (º) in $[C_4H_{10}NO][PbI_3]$.

Bonding atoms	Angle (º)
I2 – Pb1 – I1	88.806(10)

I2 – Pb1 – I3	113.808(10)
I1 – Pb1 – I3	112.907(10)
I2 – Pb1 – I1	89.054(11)
I1 – Pb1 – I1	167.705(9)
I3 – Pb1 – I1	78.974(10)
I2 – Pb1 – I2	168.201(9)
I1 – Pb1 – I2	88.048(10)
I3 – Pb1 – I2	77.851(9)
I1 – Pb1 – I2	91.609(11)
I2 – Pb1 – I3	82.648(10)
I1 – Pb1 – I3	81.472(10)
I3 – Pb1 – I3	157.398(11)
I1 – Pb1 – I3	86.246(11)
I2 – Pb1 – I3	85.640(10)

Table S3a. Selected bond lengths (Å) in $[C_4H_{10}NO]_2[PbI_4]$.

Bond	Length (Å)
Pb1 – I1	3.1946(4)
Pb1 – I2	3.1828(3)
Pb1 – I2	3.2169(3)

Table S3b. Selected bond angles (º) in $[C_4H_{10}NO]_2[PbI_4]$.

Bonding atoms	Angle (º)
I1 – Pb1 – I1	179.154(12)
I1 – Pb1 – I2	86.583(8)
I1 – Pb1 – I2	87.353(8)
I1 – Pb1 – I2	93.191(8)
I1 – Pb1 – I2	92.812(8)
I2 – Pb1 – I2	85.472(3)
I2 – Pb1 – I2	174.363(12)
I2 – Pb1 – I2	100.147(12)
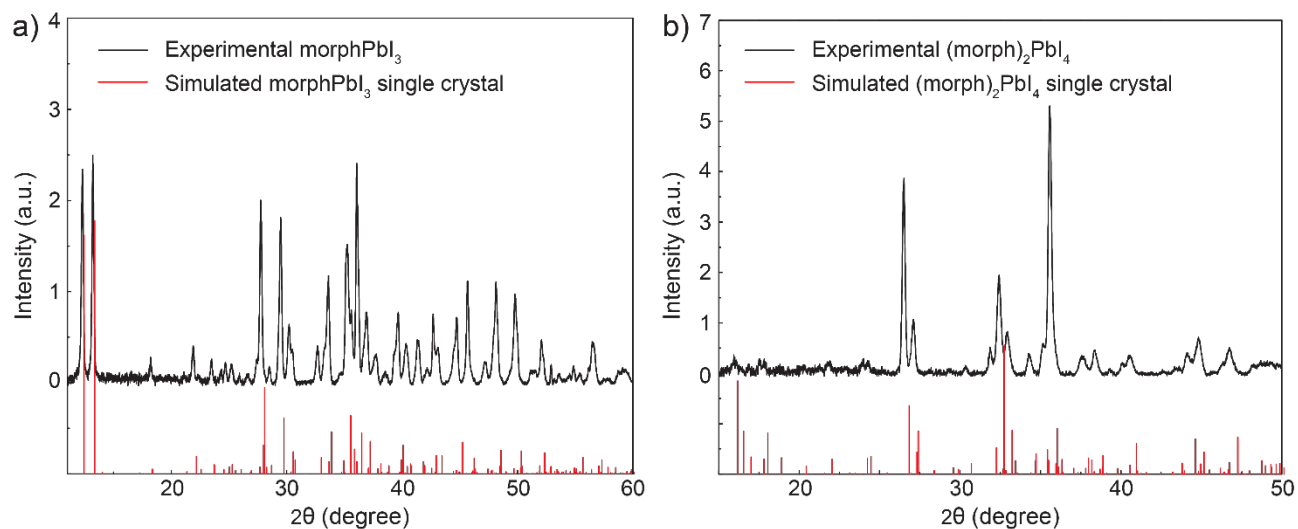
Figure S2. Powder X-ray diffraction (pXRD) patterns of the yellow phase (a) and red phase (b). Black lines are experimental pXRD patterns; red lines are simulated pXRD patterns based on structures derived from single crystal XRD analysis.
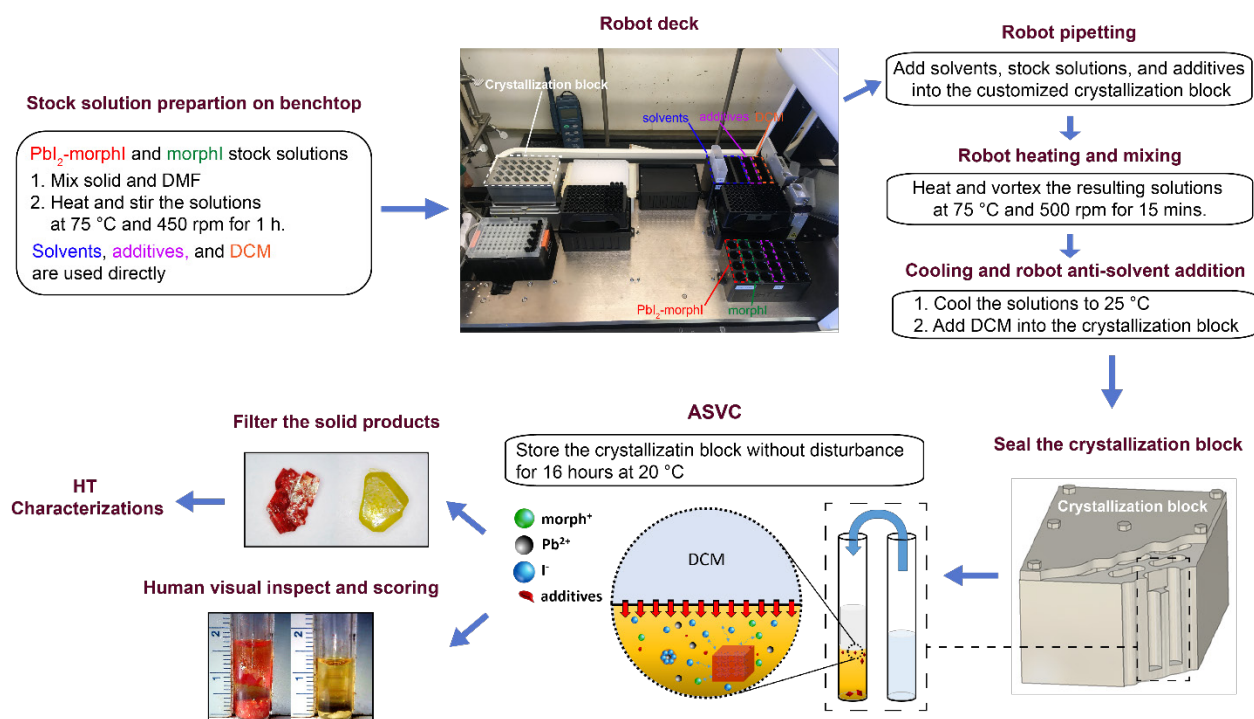
Figure S3. Robotic system and perovskite synthetic workflow based on high-throughput ASVC.
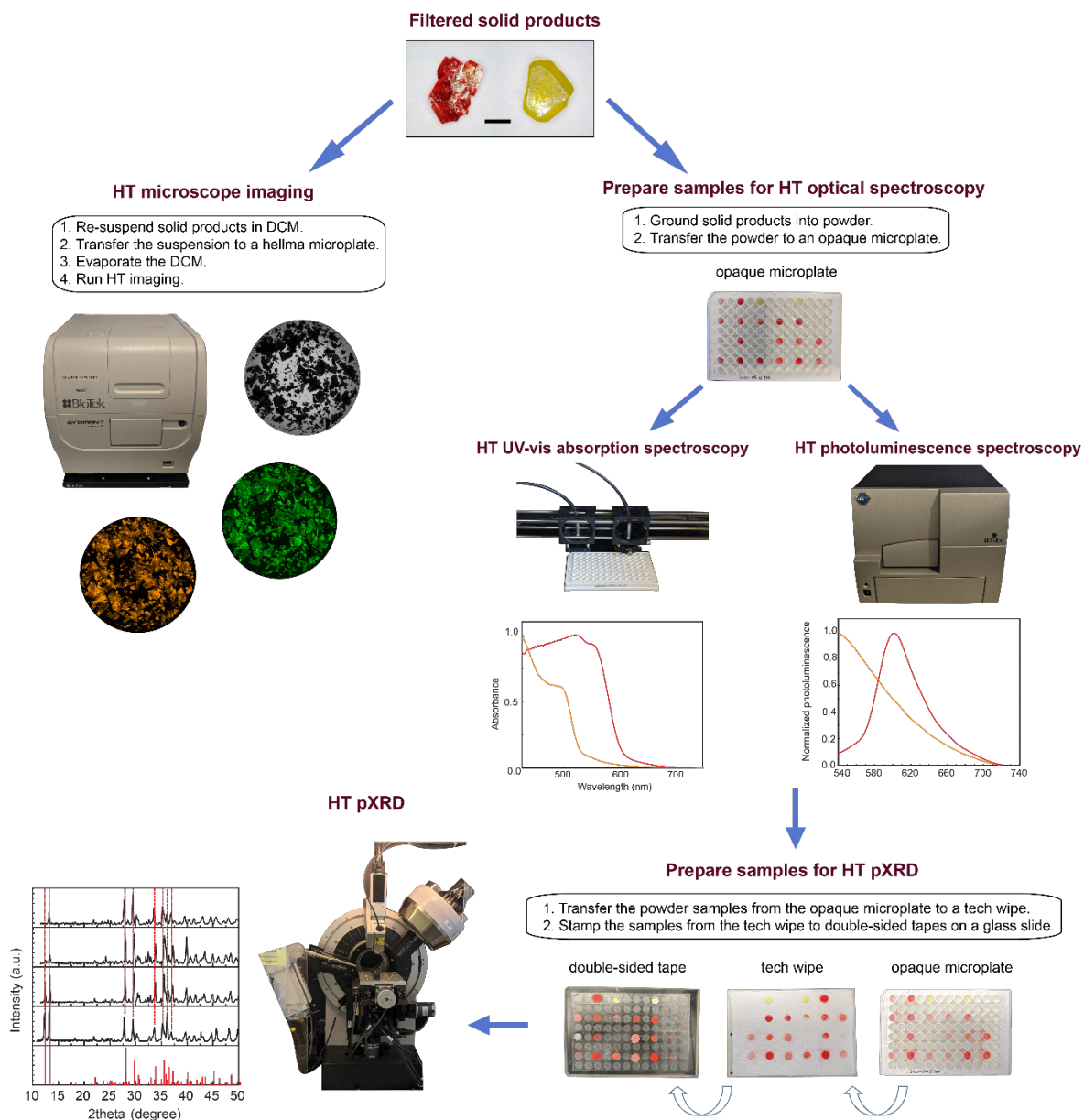
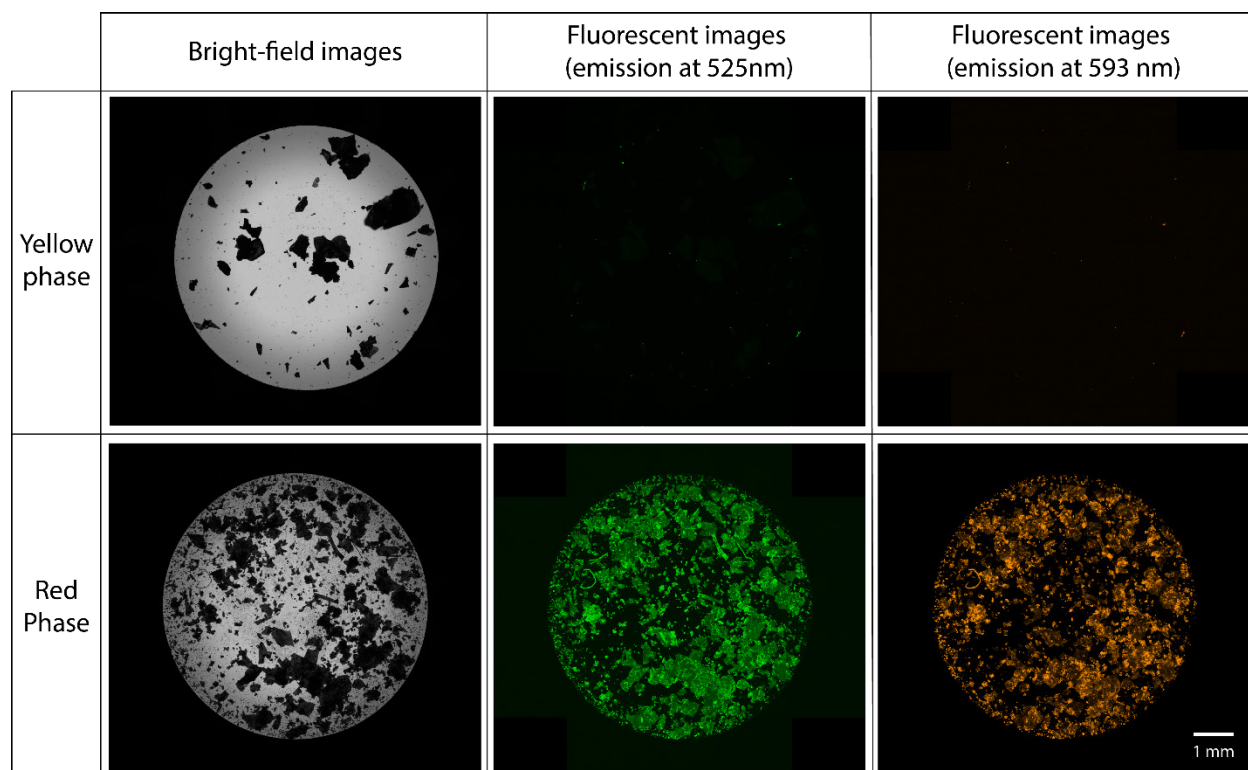**Figure S4**. High-throughput tools and characterization workflow for metal halide perovskites.

	Bright-field images	Fluorescent images (emission at 525nm)	Fluorescent images (emission at 593 nm)
Yellow phase			
Red Phase			

Figure S5. Representative high-throughput microscopic images [bright-field and fluorescent (λ_{ex} = 469 nm)] for the yellow and red phases.
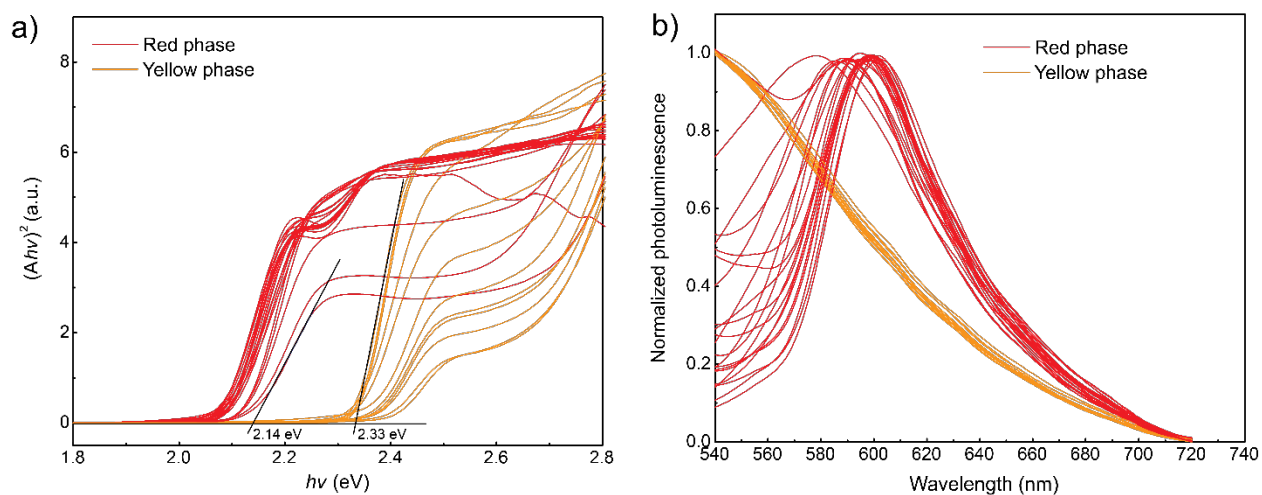
Figure S6. High-throughput a) UV-vis absorption spectra and b) PL spectra of primary screening reaction products.
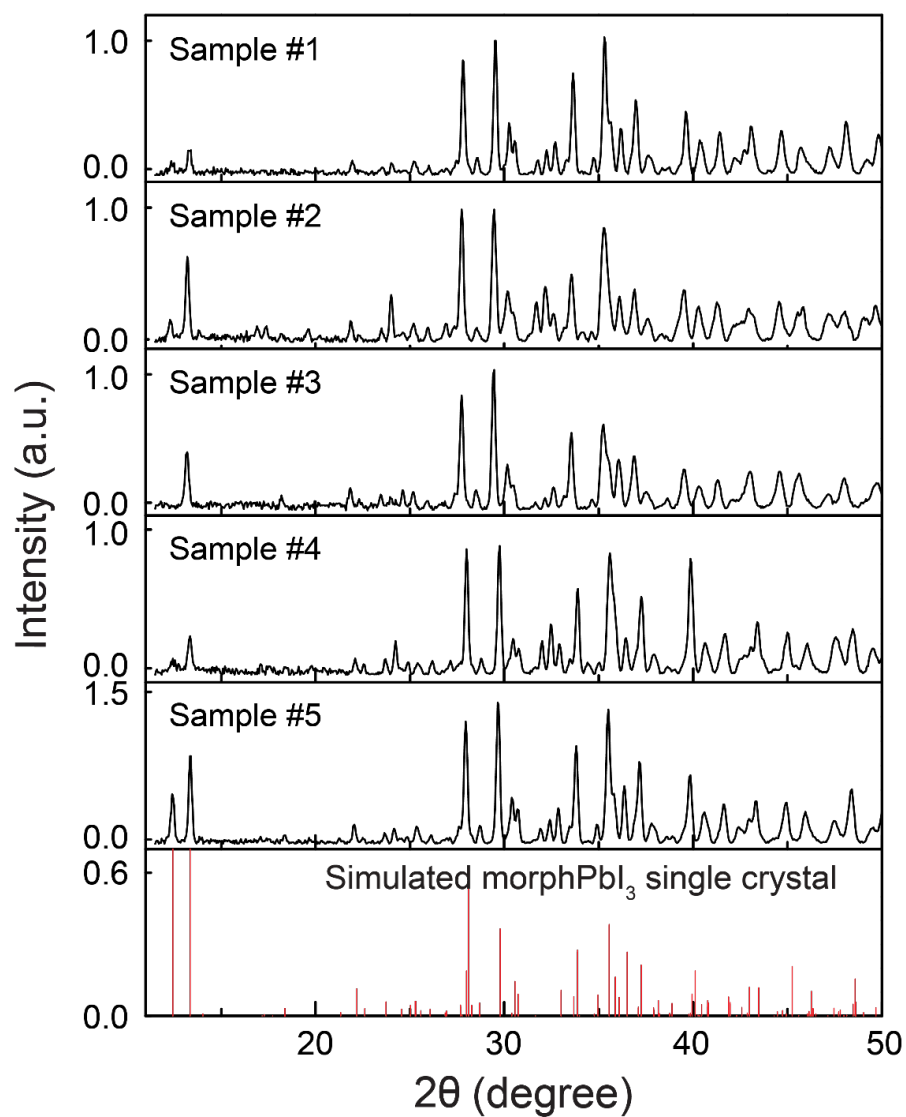
Figure S7. pXRD patterns of representative yellow phase samples from primary screening.
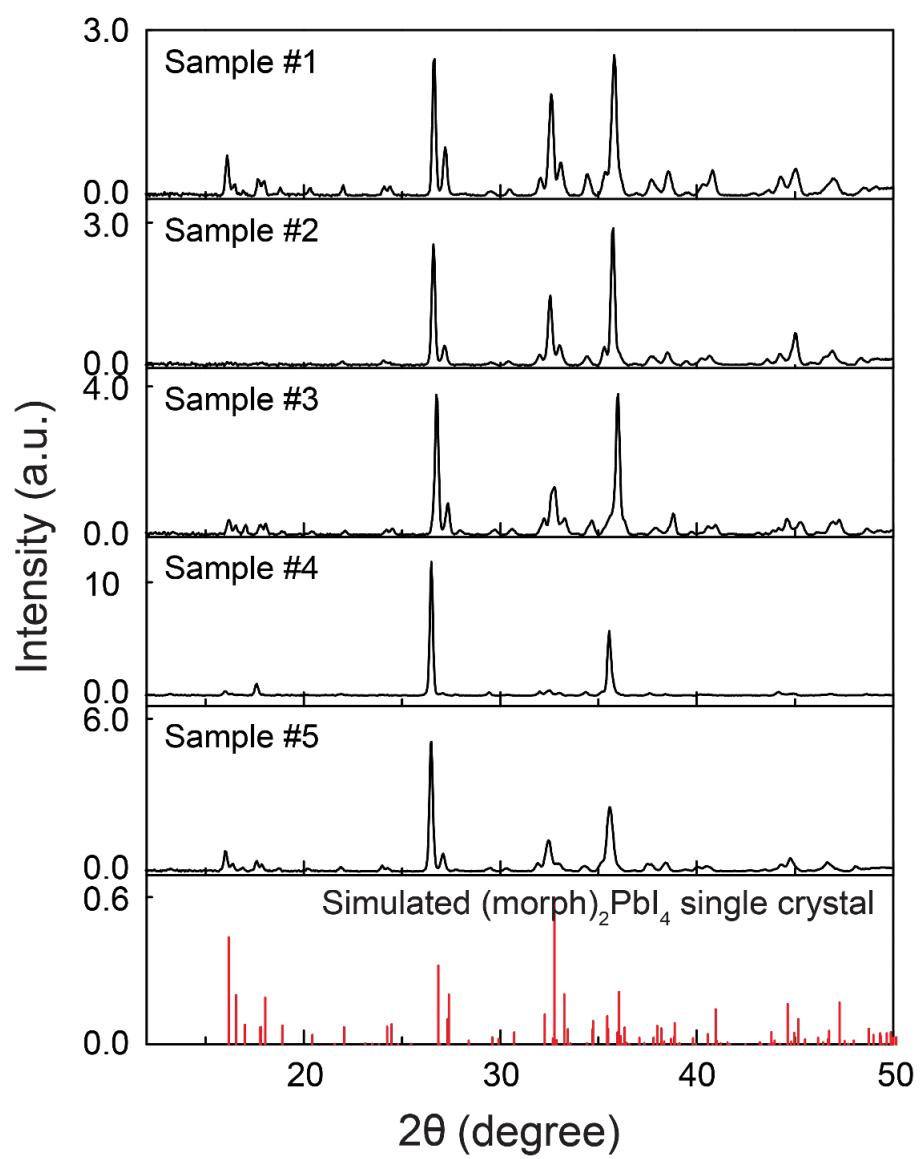
Figure S8. pXRD patterns of representative red phase samples from primary screening.

Table S4. Experiment variables and their constraints in primary screening.

Experiment variables	Constraints
[morph][a]	Lower bound: 0; Upper bound: 2.91 M (concentration of morphI in Reagent 2)
[Pb][a]	Lower bound: 0.1 M; Upper bound: 2.32 M (concentration of PbI_2 in Reagent 2)
[FAH][a]	Lower bound: 0; Upper bound: 16 M
Solvents[b]	DMF
Crystallization temperature[b]	20 °C
Crystallization time[b]	16 h

[a] Variables that have been varied.
[b] Variables that have been kept constant.

Table S5. Experiment parameters and their constraints in 6D screening.

Experiment variables	Constraints
[morph][a]	Lower bound: 0; Upper bound: 2.91 M (concentration of the morphI in Reagent 2)
[Pb][a]	Lower bound: 0.1 M; Upper bound: 2.32 M (concentration of PbI_2 in Reagent 2)
[FAH][a]	Lower bound: 0; Upper bound: 16 M
[H_2O][a]	Lower bound: 0; Upper bound: 13.5 M
V_f(DMSO)[a]	Lower bound: 0; Upper bound: 1
V_f(GBL)[a]	Lower bound: 0; Upper bound: 1
Solvents[a]	DMF, DMSO, GBL
Crystallization temperature[b]	20 °C
Crystallization time[b]	16 h

[a] Variables that have been varied.
[b] Variables that have been kept constant.

Table S6. Tested machine learning models and their CV accuracies for initial sampling in 6D composition space screening.

Model	Hyper-parameters sets* (tested)	Best Hyper-parameters (from grid-search)	CV accuracy
Random Forest	bootstrap = [True, False], max_features = ['auto', 'sqrt'], n_estimators = [100, 300, 500, 800, 1200], max_depth = [2, 5, 8, 15, 25, 30], min_samples_split = [2, 5, 10, 15, 100], min_samples_leaf = [1, 2, 4, 8], criterion = 'entropy', class_weight = 'balanced', random_state = 42	bootstrap = True max_features = 'auto' n_estimators = 800 max_depth = 5 min_samples_split = 5 min_samples_leaf = 1 criterion = 'entropy', class_weight = 'balanced' random_state = 42	0.80 ± 0.09
Support Vector Machine (Pearson VII Universal Function Kernel)	C = [0.001, 0.01, 0.1, 1, 10, 100, 1000], decision_function_shape = 'ovr', probability = True, class_weight = 'balanced'	C = 1, decision_function_shape = 'ovr', probability = True, class_weight = 'balanced'	0.78 ± 0.07

* see explanations of the hyper-parameters in Scikit-learn documents at https://scikit-learn.org/stable/supervised_learning.html
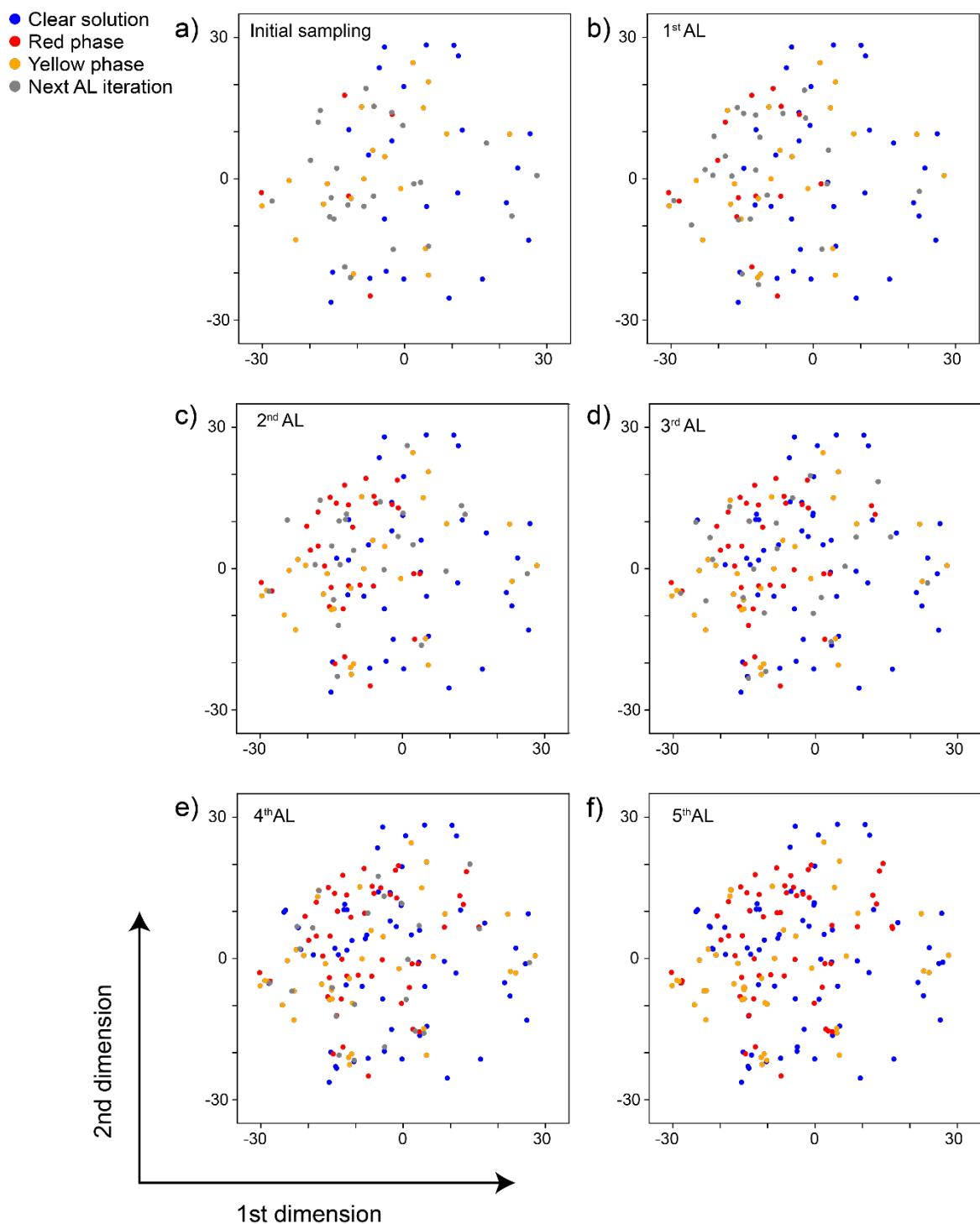
Figure S9. Projections of the reaction-composition space using the t-distributed stochastic neighborhood embedding (t-SNE) algorithm for initial sampling (a) and 1st to 5th AL iterations (b-f). The gray points represent the next AL iteration.

Table S7. Machine learning models tested after each iteration of AL (their CV accuracies are shown in Figure S10)

Model	Hyper-parameters*
Random Forest	bootstrap = True max_features = 'auto' n_estimators = 800 max_depth = 5 min_samples_split = 5 min_samples_leaf = 1 criterion = 'entropy', class_weight = 'balanced' random_state = 42
Support Vector Machine (Pearson VII Universal Function Kernel)	C = 1, decision_function_shape = 'ovr', probability = True, class_weight = 'balanced'
Support Vector Machine (Radial Basis Function Kernel)	C = 1, gamma = 0.001 decision_function_shape = 'ovr', probability = True, class_weight = 'balanced'
XGBoost	booster='gbtree', n_estimators=100
K-nearest neighbor	n_neighbors = 1, p = 2, weights = 'distance'
Gaussian Process	kernel=1**2 * RBF(length_scale=1)

* see explanations of the hyper-parameters in Scikit-learn and XGBoost documents [ref]
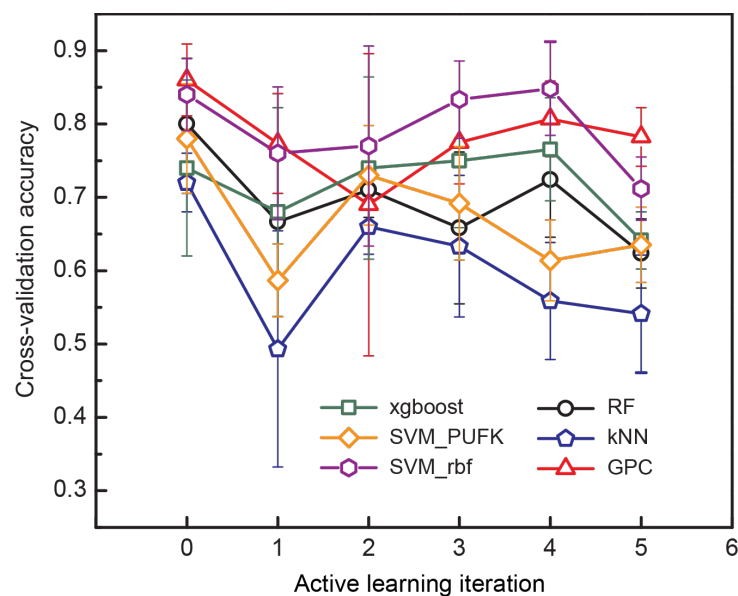
Figure S10. Cross-validation accuracies of different machine learning models on the dataset collected after each iteration of AL.
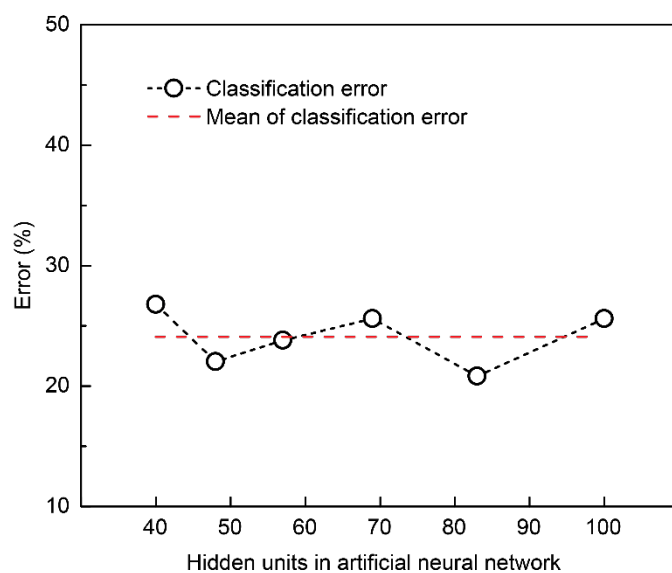

Figure S11. Classification errors (one minus cross-validation accuracies) of artificial neural networks with different numbers of hidden units. The red dashed line represents the mean value of the classification errors.
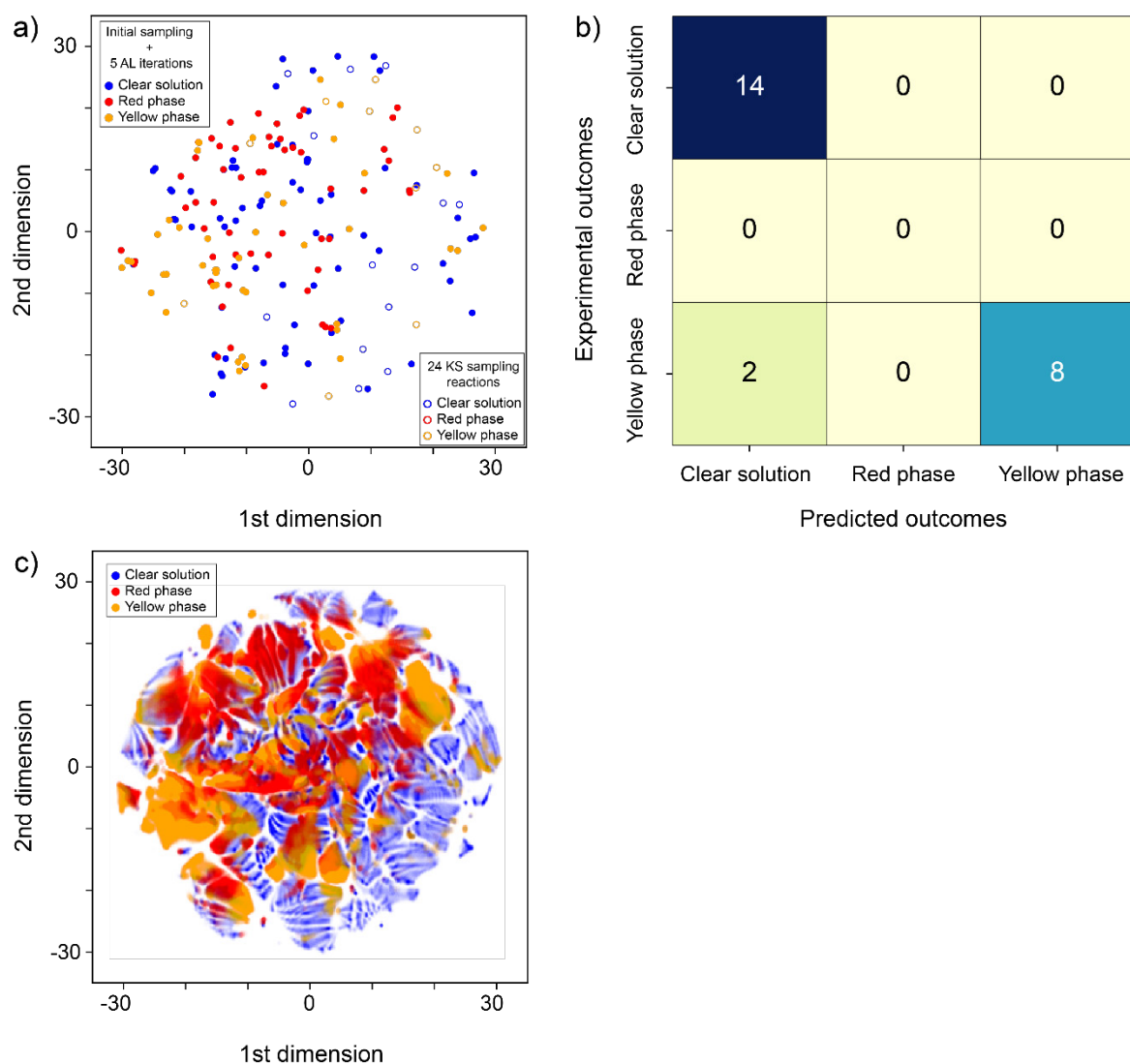
Figure S12. a) Projections of the reaction-composition space using t-SNE for initial sampling + five AL iterations and 24 unseen reactions (labeled by experimental outcomes) selected by the KS algorithm. b) Confusion matrix of experimental outcomes vs. predicted outcomes. c) Projections of the reaction pool labeled with predicted outcomes.
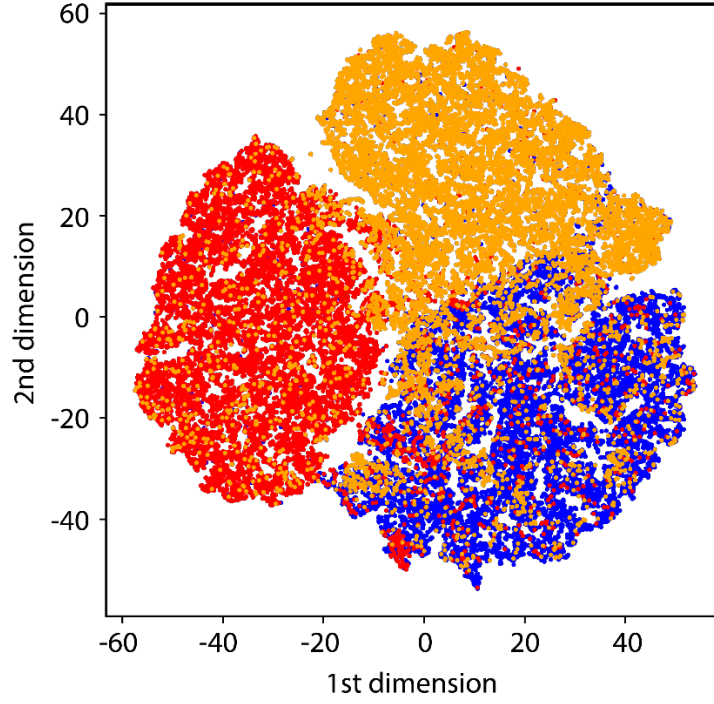
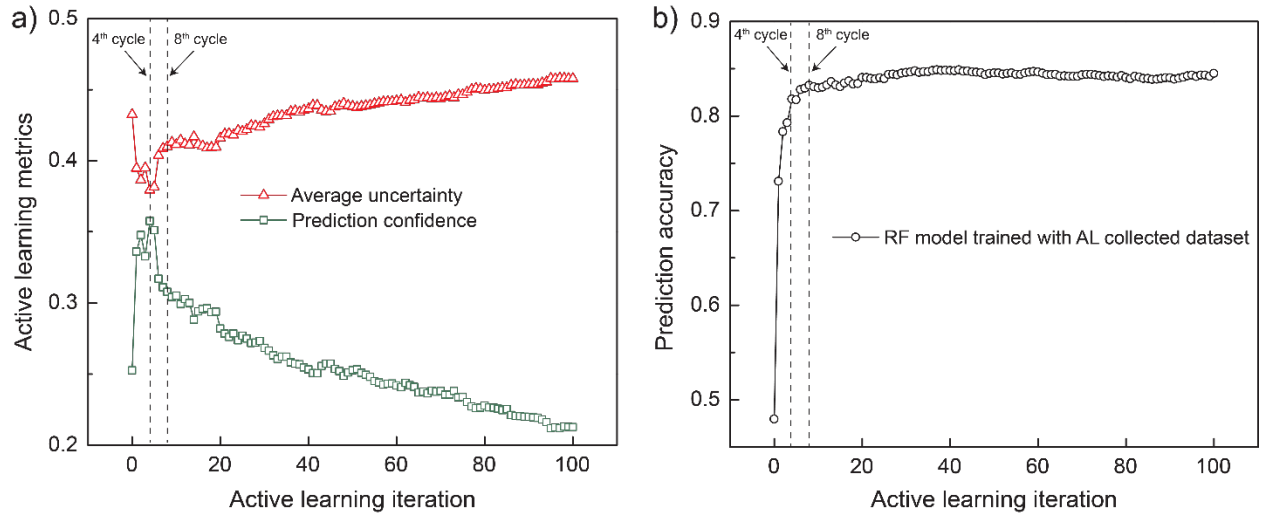**Figure S13.** Projections of the 6D synthetic dataset (50,000 data points) using t-SNE.

Figure S14. (a) Prediction confidence and average uncertainty of the RF model on the synthetic dataset in each iteration of AL. (b) Prediction accuracy (on the whole synthetic dataset) of the RF model in each iteration of AL.
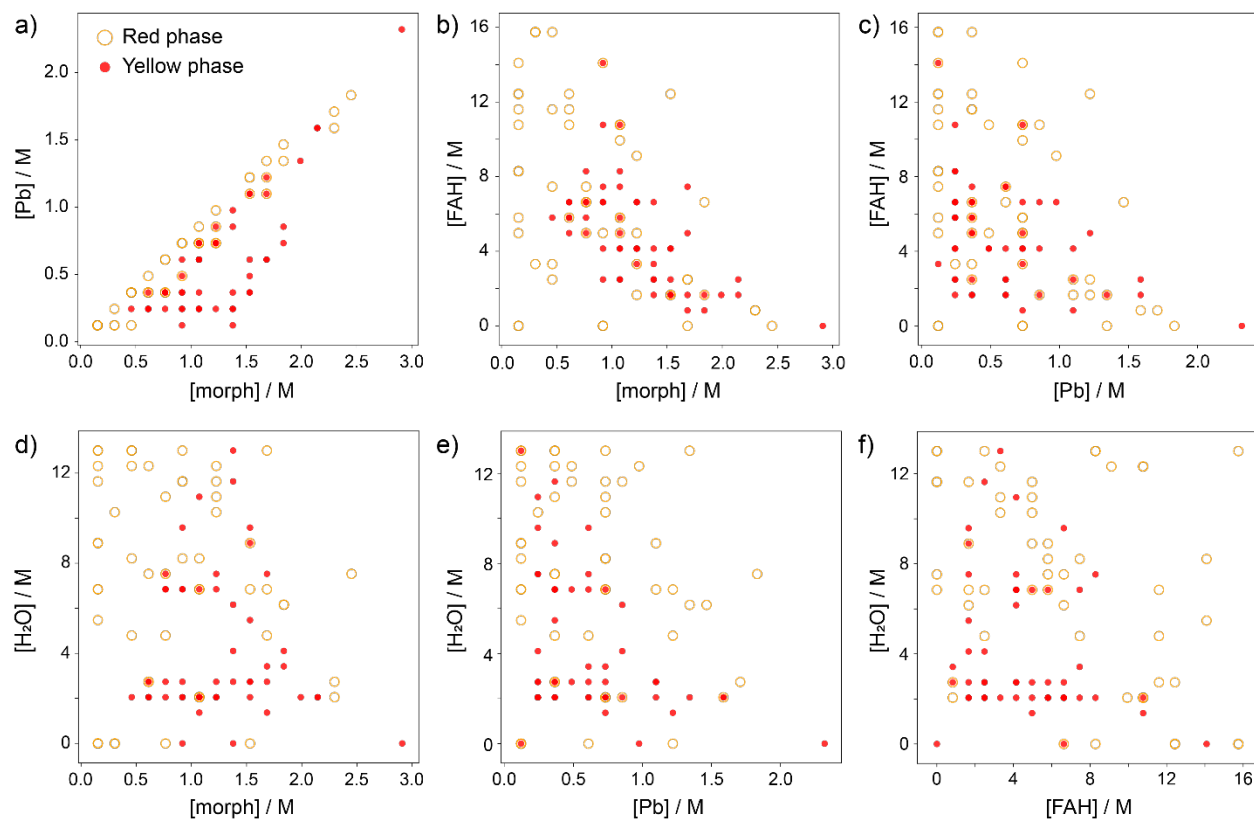
Figure S15. 2D projection plots of the "crystals only" dataset using axis from [Pb], [morph], [FAH], and [H$_2$O].
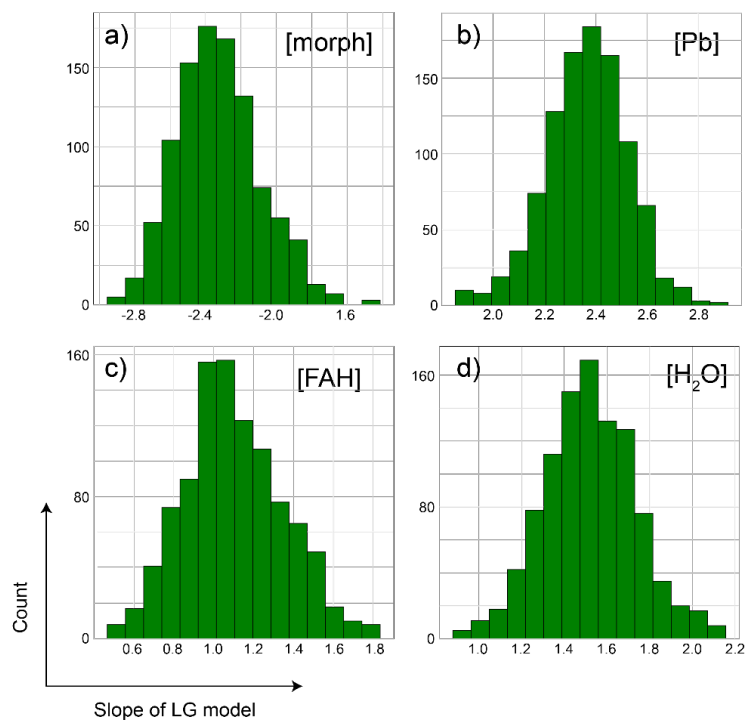
Figure S16a. Distributions of logistic regression (LG) slopes for a) [morph], b) [Pb], c) [FAH], and d) [H_2O] when treating "yellow phase" class as "1".
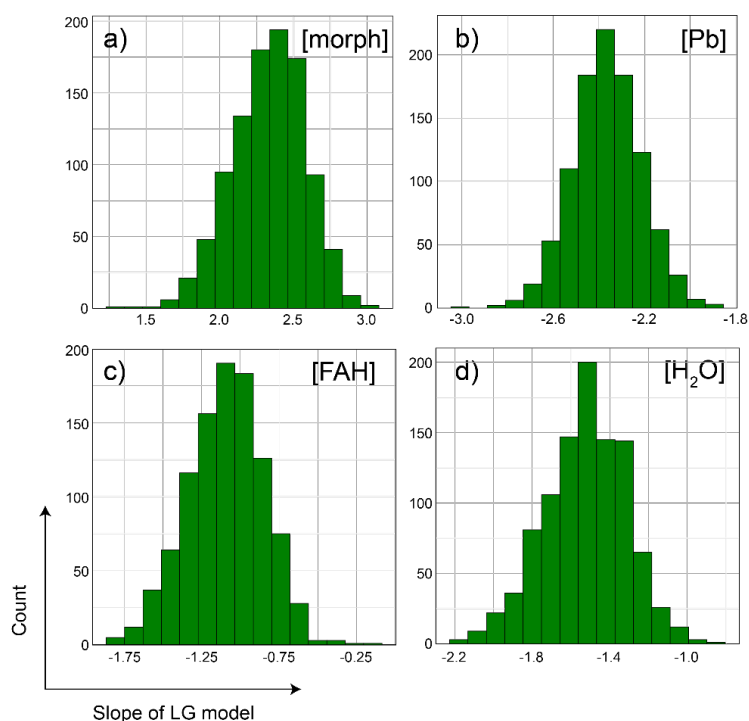

Figure S16b. Distributions of logistic regression (LG) slopes for a) [morph], b) [Pb], c) [FAH], and d) [H_2O] when treating "red phase" class as "1".
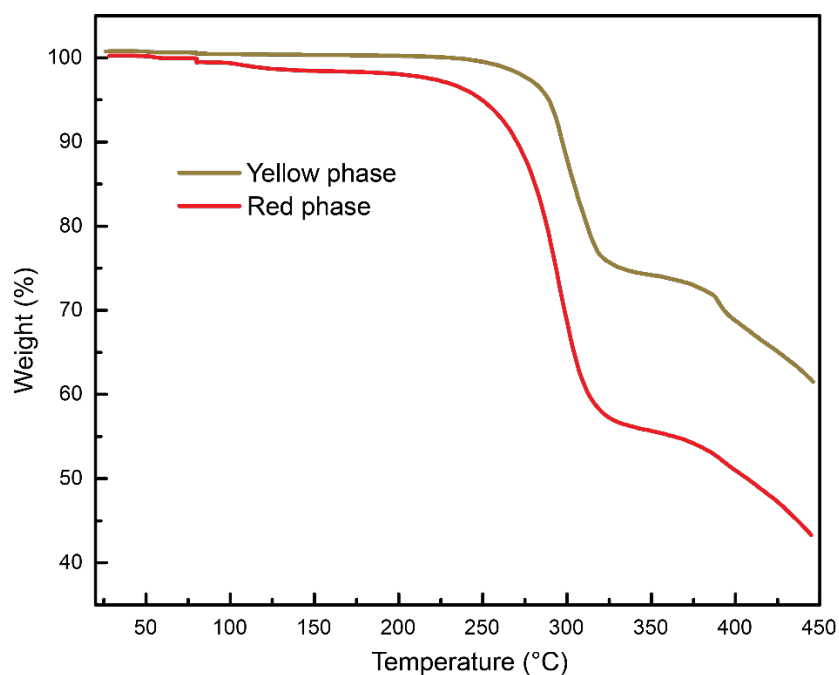
Figure S17. Thermogravimetric analysis of the yellow phase and red phase.

3. References

(1)   Pendleton, I. M.; Cattabriga, G.; Li, Z.; Najeeb, M. A.; Friedler, S. A.; Norquist, A. J.; Chan, E. M.; Schrier, J. Experiment Specification, Capture and Laboratory Automation Technology (ESCALATE): A Software Pipeline for Automated Chemical Experimentation and Data Management. *MRS Commun* **2019**, *9*, 846–859. https://doi.org/10.1557/mrc.2019.72.

(2)   Schrier, J. Solution Mixing Calculations as a Geometry, Linear Algebra, and Convex Analysis Problem. *J. Chem. Educ.* **2021**, *98* (5), 1659–1666. https://doi.org/10.1021/acs.jchemed.0c01456.

(3)   Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Phys. Rev. Lett.* **2008**, *100* (13), 136406. https://doi.org/10.1103/PhysRevLett.100.136406.

(4)   Saidi, W. A.; Choi, J. J. Nature of the Cubic to Tetragonal Phase Transition in Methylammonium Lead Iodide Perovskite. *J. Chem. Phys.* **2016**, *145* (14), 144702. https://doi.org/10.1063/1.4964094.

(5)   Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102* (7), 073005. https://doi.org/10.1103/PhysRevLett.102.073005.

(6)   Al-Saidi, W. A.; Voora, V. K.; Jordan, K. D. An Assessment of the VdW-TS Method for Extended Systems. *J. Chem. Theory Comput.* **2012**, *8* (4), 1503–1513. https://doi.org/10.1021/ct200618b.

(7)   Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50* (24), 17953–17979. https://doi.org/10.1103/PhysRevB.50.17953.

(8)   Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1999**, *59* (3), 1758–1775. https://doi.org/10.1103/PhysRevB.59.1758.

(9)   Shan, W.; Saidi, W. A. Segregation of Native Defects to the Grain Boundaries in Methylammonium Lead Iodide Perovskite. *J. Phys. Chem. Lett.* **2017**, *8* (23), 5935–5942. https://doi.org/10.1021/acs.jpclett.7b02727.

(10) Saidi, W. A.; Shadid, W.; Castelli, I. E. Machine-Learning Structural and Electronic Properties of Metal Halide Perovskites Using a Hierarchical Convolutional Neural Network. *Npj Comput. Mater.* **2020**, *6* (1), 1–7. https://doi.org/10.1038/s41524-020-0307-8.

(11) Bruker (2018). Apex3 V2018.1-0, SAINT V8.38A, Bruker AXS Inc.: Madison (WI), USA, 2018.

(12) SHELXTL Suite of Programs, Version 6.14, 2000-2003, Bruker Advanced X-Ray Solutions, Bruker AXS Inc., Madison, Wisconsin: USA.

(13) Sheldrick, G. M. A Short History of SHELX. *Acta Crystallogr. A* **2008**, *64* (1), 112–122. https://doi.org/10.1107/S0108767307043930.

(14) Sheldrick, G.M. University of Göttingen, Germany, 2018.

(15) Sheldrick, G. M. Crystal Structure Refinement with SHELXL. *Acta Crystallogr. Sect. C Struct. Chem.* **2015**, *71* (1), 3–8. https://doi.org/10.1107/S2053229614024218.

(16) Hübschle, C. B.; Sheldrick, G. M.; Dittrich, B. ShelXle: A Qt Graphical User Interface for SHELXL. *J. Appl. Crystallogr.* **2011**, *44* (Pt 6), 1281–1284. https://doi.org/10.1107/S0021889811043202.