

# Supporting Information

## Dimensional control over metal halide perovskite crystallization guided by active learning

Zhi Li,<sup>§</sup> Philip W. Nega,<sup>§</sup> Mansoor Ani Najeeb Nellikkal,<sup>‡</sup> Chaochao Dun,<sup>§</sup> Matthias Zeller,<sup>#</sup> Jeffrey J. Urban,<sup>§</sup> Wissam A. Saidi,<sup>◦</sup> Joshua Schrier,<sup>†</sup> Alexander J. Norquist,<sup>‡</sup> Emory M. Chan\*<sup>§</sup>

<sup>§</sup>Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA

<sup>‡</sup>Department of Chemistry, Haverford College, 370 Lancaster Avenue, Haverford, Pennsylvania 19041, USA

<sup>†</sup>Department of Chemistry, Fordham University, 441 E. Fordham Road, The Bronx, New York 10458, USA

<sup>#</sup>Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette IN 47907, USA

<sup>◦</sup>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

<b>1. Materials and methods .....</b>	3
<b>Chemicals .....</b>	3
<b>Equipment .....</b>	3
<i>Liquid handling robot</i> .....	3
<i>Powder X-ray Diffraction</i> .....	4
<i>UV-visible absorption spectra</i> .....	4
<i>Photoluminescence spectra</i> .....	4
<i>Thermogravimetric analysis</i> .....	4
<i>High-throughput powder X-ray Diffraction</i> .....	4
<i>High-throughput UV-Vis absorption spectra</i> .....	4
<i>High-throughput photoluminescence spectra</i> .....	4
<i>High-throughput microscopic images</i> .....	4
<b>Benchtop ASVC synthesis .....</b>	5
<b>Robotic Workflow.....</b>	5
<i>Solubility measurement</i> .....	5
<i>Overview of HT-synthesis and HT-characterization workflows</i> .....	5
<i>Step-by-step synthetic protocols</i> .....	6
<b>Product Scoring Rubric Based on Human Inspection.....</b>	8
<b>Software and Computation .....</b>	9
<i>Software and packages</i> .....	9
<i>Five-fold cross-validation</i> .....	9
<i>Machine learning metrics</i> .....	9
<i>Grid-point generation</i> .....	10
<i>Kennard-Stone algorithm</i> .....	10
<i>Selecting 48 primary reactions in 3D reaction-composition space</i> .....	10
<i>Generating the reaction pool in 6D reaction-composition space</i> .....	11
<i>Diverse-mini-batch active learning</i> .....	11
<i>Random-classification accuracy</i> .....	11
<i>Theoretical limit of CV accuracy</i> .....	11
<i>Selecting 24 reactions far from initial sampling + 5 AL runs</i> .....	12
<i>Synthetic dataset</i> .....	12
<i>Permutation-feature-importance analysis</i> .....	12

<i>Bootstrapping the “crystals only” dataset .....</i>	13
<i>One-tailed statistical hypothesis testing .....</i>	13
<i>Density functional theory (DFT) calculations .....</i>	14
<i>Detailed discussion of the mechanistic role of additives in controlling MHP dimensionality .....</i>	14
<b>2. Supplemental Figures and Tables .....</b>	<b>16</b>
<b>Single Crystal Structure Refinement Details.....</b>	<b>16</b>
Compound .....	17
<b>3. References .....</b>	<b>33</b>

## 1. Materials and methods

### Chemicals

Lead iodide ( $\text{PbI}_2$ ) (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.  $\gamma$ -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 K $\alpha$  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<sub>2</sub>, 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<sub>2</sub> depends on the concentration of coexisting morphI in the solution. To determine the solubility of PbI<sub>2</sub> for HT-ASVC reactions, we weighed fixed masses of PbI<sub>2</sub> 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<sub>2</sub> did not precipitate, which can disrupt the pipetting of liquid handler. The solubility was then calculated as total moles of PbI<sub>2</sub> divided by the total volume of the solutions. The solubility of PbI<sub>2</sub> 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<sub>2</sub>-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<sub>2</sub>-morphI mixture solution were 2.32 M (PbI<sub>2</sub>) 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 °C (measured by an IR camera at a set point temperature of 80 °C) 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 °C 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<sub>2</sub>, 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 °C 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<sub>2</sub>-morphI in DMF, is prepared by adding PbI<sub>2</sub>, 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 °C, with the actual temperature of the HHS reactor module measured and recorded in the software. The heating rate is approximately 7 °C/ min.
3. After the measured temperature of HHS reaches 80 °C, 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 °C (75 °C 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 °C 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 °C without disturbance for 16 h.

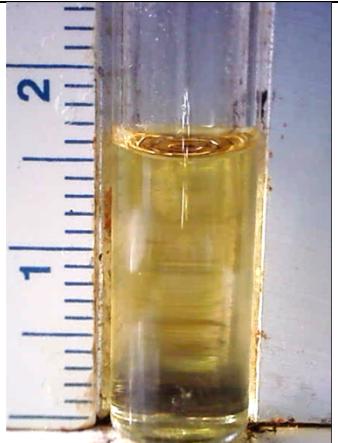
***Screening in the reaction-composition space of [Pb], [morph], [FAH], [H<sub>2</sub>O], V<sub>f</sub>, DMSO, and V<sub>f</sub>, GBL.***

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<sub>2</sub>-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	<b>No solid</b> was observed in the solutions.	NA	
3	<b>Red-colored crystals or powder</b> are the major solid products.	Large-red-flake crystals	

4	<b>Yellow-colored crystals or powder</b> are the major solid products.	Bulky-yellow crystals	
---	--	-----------------------	--

## Software and Computation

### *Software and packages*

Our custom-developed pipeline software: ESCALATE<sup>1</sup>, 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.

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN)$$

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<sup>2</sup> ( $C_{3D}$ ) 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<sub>2</sub>O], V<sub>f</sub>, DMSO, and V<sub>f</sub>, 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 5<sup>th</sup> 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  $i^{th}$  class in the dataset. For the dataset after the 5<sup>th</sup> 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	activation = ‘relu’ solver = ‘lbfgs’ alpha = 0.001,	0.27	0.24 ± 0.02
48		0.22	
57		0.24	
69		0.26	
83		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.

<b>Library</b>	<b>module</b>	<b>function</b>	<b>parameters</b>
Scikit-learn	dataset	<i>make_classification</i>	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<sub>2</sub>O]) 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 ( $\mu$ ) and standard deviation ( $\sigma$ ) 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	$\mu$	$\sigma$	$Z_c$	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 <sub>2</sub> O]	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	$\mu$	$\sigma$	$Z_c$	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 <sub>2</sub> O]	Negative	Positive or none	Left tail	-1.53	0.21	$-228 < -Z_0$	Reject null hypothesis

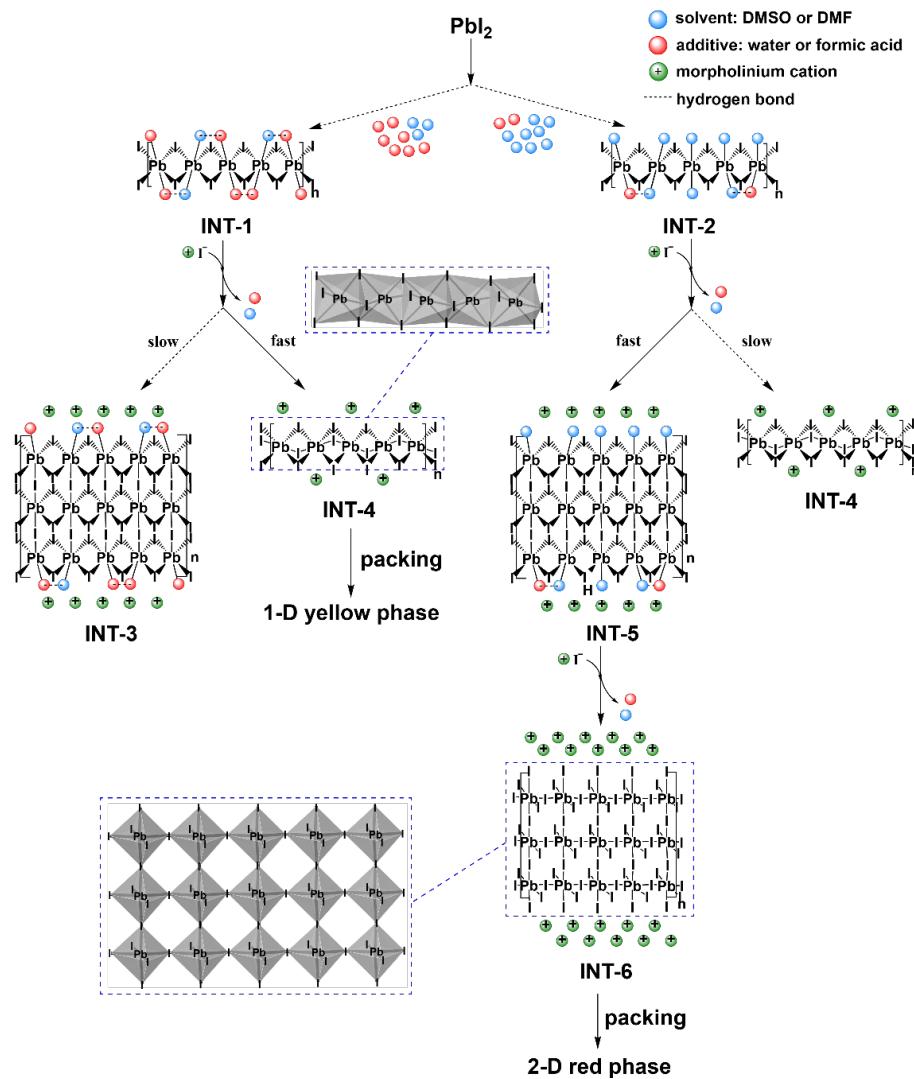
### Density functional theory (DFT) calculations

The first-principles DFT calculations were carried out within the Perdew-Burke-Ehrenzhof exchange-correlation functional revised for solids (PBEsol).<sup>3</sup> Because of the importance of van der Waals interactions in the hybrid perovskites<sup>4</sup>, we applied the Tkatchenko-Scheffler scheme to correct the PBEsol energies.<sup>5,6</sup> The electron-nucleus interactions were modeled using the projector augmented wave (PAW) pseudopotentials<sup>7,8</sup> 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.<sup>9</sup> 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.<sup>10</sup> Thermodynamic stability was assessed by accounting for vibrational energy as computed using phonopy.

### Detailed discussion of the mechanistic role of additives in controlling MHP dimensionality

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 (see Scheme S1 below). When morphI and PbI<sub>2</sub> are dissolved in the solution, solvent (i.e., DMF and DMSO) and additive molecules (i.e., formic acid and H<sub>2</sub>O) coordinate with PbI<sub>2</sub> to form 1D complexes (INT-1 and INT-2 in Scheme S1) as prenucleation intermediates.<sup>11–13</sup> When the additives' concentrations are low, the 1D PbI<sub>2</sub> 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<sub>2</sub> chains are substituted by water and formic acid to form intermediates similar to INT-1.<sup>12</sup> 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 stabilize 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.



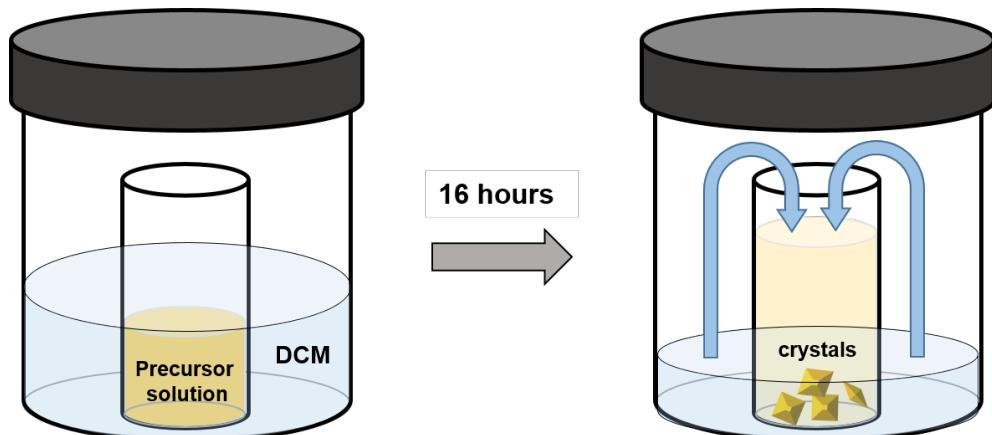
**Scheme S1.** Detailed reaction scheme for the formation of the 1D yellow phase and 2D red phase.

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- $\text{I}^-$  from morphI. Thus, **INT-1** is converted to **INT-4**—the 1D  $\text{PbI}_2$  chain that forms the framework of the yellow phase (see crystal structure in Figure 1b). Compared to **INT-1**  $\rightarrow$  **INT-4**, the  $\text{I}^-$  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 makes the reaction rate of **INT-2**  $\rightarrow$  **INT-4** slower than **INT-2**  $\rightarrow$  **INT-5**, which then becomes the dominant

pathway. With additional  $\text{I}^-$ , 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  $[\text{H}_2\text{O}]$  and  $[\text{FAH}]$ .

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.

## 2. Supplemental Figures and Tables



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

### Single Crystal Structure Refinement Details

Single crystals of morph $\text{PbI}_3$  and (morph) $_2\text{PbI}_4$  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 K $\alpha$  radiation ( $\lambda = 0.71073 \text{ \AA}$ ) at 100(2) K.

Data were collected, reflections were indexed and processed, and the files were scaled and corrected for absorption using APEX3.<sup>14</sup> Heavy atom positions were determined using SIR92.<sup>31</sup> 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  $I > 3\sigma(I)$ . Hydrogen atoms were placed in geometrically idealized positions. All calculations

were performed using Crystals v. 14.23c.<sup>32</sup> 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 charge from The Cambridge Crystallographic Data Centre via [www.ccdc.cam.ac.uk/data\\_request/cif](http://www.ccdc.cam.ac.uk/data_request/cif).

**Table S1.** Crystallographic data for [C<sub>4</sub>H<sub>10</sub>NO][PbI<sub>3</sub>] (yellow phase) and [C<sub>4</sub>H<sub>10</sub>NO]<sub>2</sub>[PbI<sub>4</sub>] (red phase).

Compound	[C <sub>4</sub> H <sub>10</sub> NO][PbI <sub>3</sub> ]	[C <sub>4</sub> H <sub>10</sub> NO] <sub>2</sub> [PbI <sub>4</sub> ]
Formula	C <sub>4</sub> H <sub>10</sub> I <sub>3</sub> NOPb	C <sub>8</sub> H <sub>20</sub> I <sub>4</sub> N <sub>2</sub> O <sub>2</sub> Pb
fw	676.04	891.08
Space-Group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> (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 / °	90	90
β / °	90	94.5715(13)
γ / °	90	90
V / Å <sup>3</sup>	1175.89(9)	1885.00(13)
Z	4	4
ρ <sub>calc</sub> / g cm <sup>-3</sup>	3.818	3.140
λ / Å	0.71073	0.71073
T / K	100(2)	100(2)
μ / mm <sup>-1</sup>	22.177	15.497
R <sub>1</sub> <sup>a</sup>	0.0235	0.0248
wR <sub>2</sub> <sup>b</sup>	0.0513	0.0583

<sup>a</sup> R<sub>1</sub> = Σ ||F<sub>o</sub>|| - ||F<sub>c</sub>|| / Σ |F<sub>o</sub>|. <sup>b</sup> wR<sub>2</sub> = [Σw(F<sub>o</sub><sup>2</sup> - F<sub>c</sub><sup>2</sup>)<sup>2</sup> / [Σw(F<sub>o</sub><sup>2</sup>)<sup>2</sup>]<sup>1/2</sup>].

**Table S2a.** Selected bond lengths (Å) in [C<sub>4</sub>H<sub>10</sub>NO][PbI<sub>3</sub>].

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 ( $^{\circ}$ ) in  $[C_4H_{10}NO][PbI_3]$ .

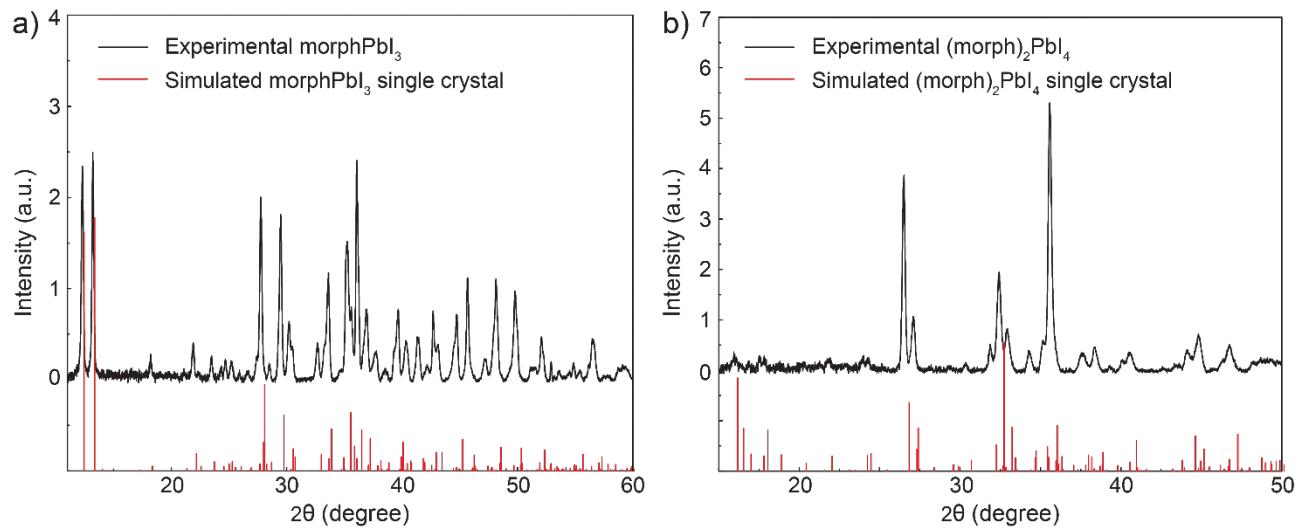
Bonding atoms	Angle ( $^{\circ}$ )
I2 – Pb1 – I1	88.806(10)
I2 – Pb1 – I3	113.808(10)
I1 – Pb1 – I3	112.907(10)
I2 – Pb1 – I1	89.054(11)
I1 – Pb1 – I1	167.705(9)
I3 – Pb1 – I1	78.974(10)
I2 – Pb1 – I2	168.201(9)
I1 – Pb1 – I2	88.048(10)
I3 – Pb1 – I2	77.851(9)
I1 – Pb1 – I2	91.609(11)
I2 – Pb1 – I3	82.648(10)
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 ( $\text{\AA}$ ) in  $[C_4H_{10}NO]_2[PbI_4]$ .

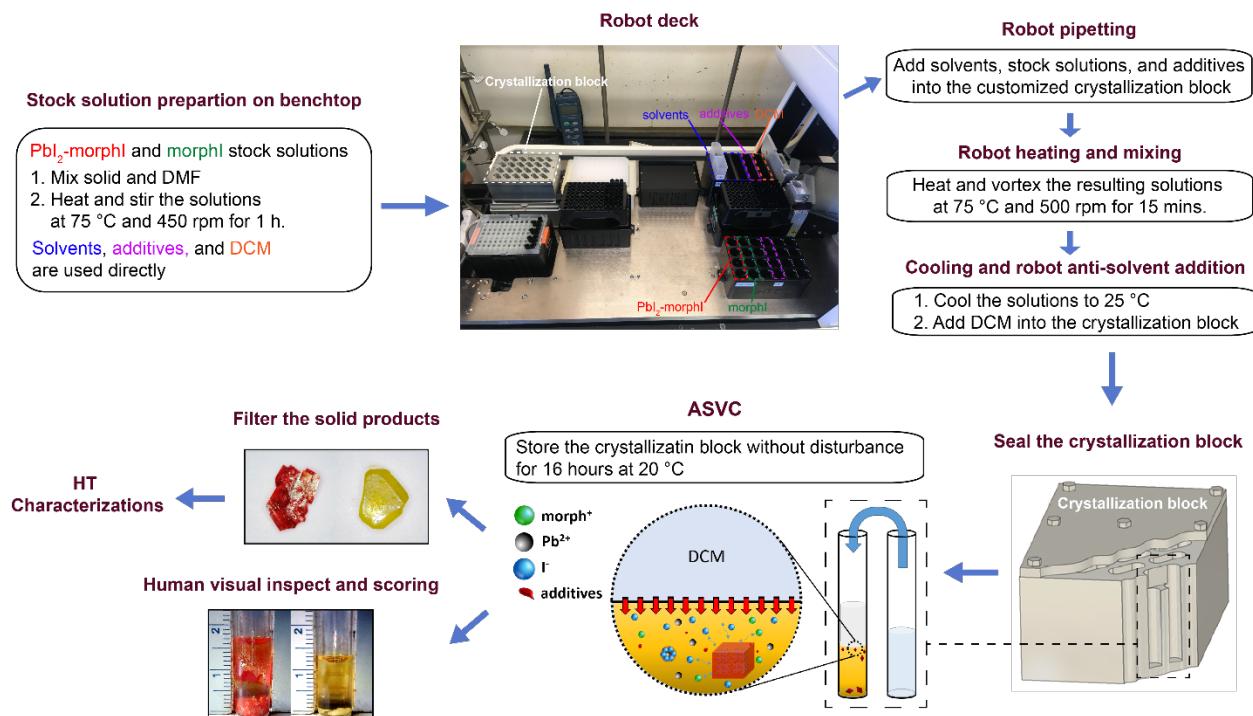
Bond	Length ( $\text{\AA}$ )
Pb1 – I1	3.1946(4)
Pb1 – I2	3.1828(3)
Pb1 – I2	3.2169(3)

**Table S3b.** Selected bond angles ( $^{\circ}$ ) in  $[C_4H_{10}NO]_2[PbI_4]$ .

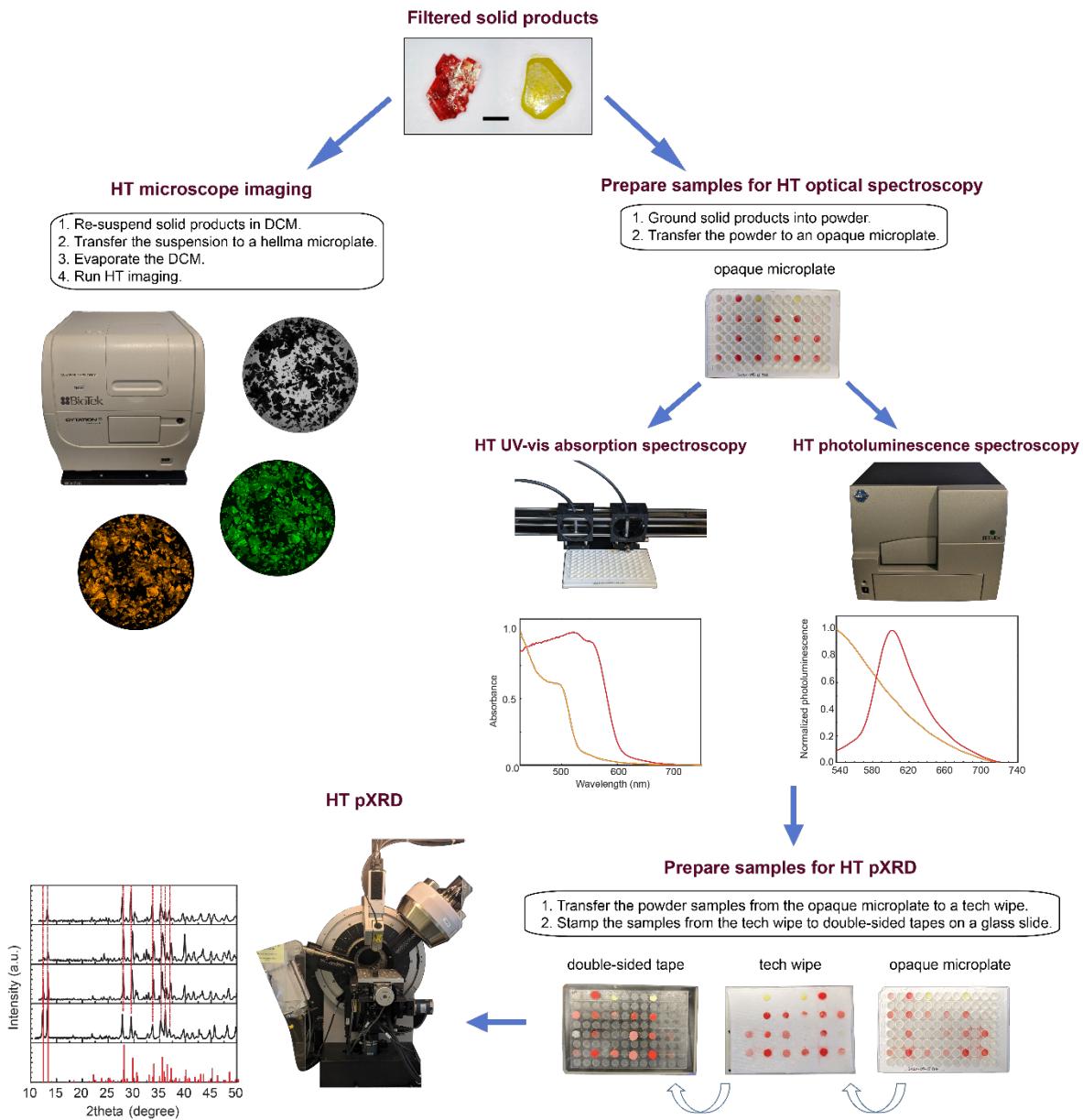
Bonding atoms	Angle ( $^{\circ}$ )
I1 – Pb1 – I1	179.154(12)
I1 – Pb1 – I2	86.583(8)
I1 – Pb1 – I2	87.353(8)
I1 – Pb1 – I2	93.191(8)
I1 – Pb1 – I2	92.812(8)
I2 – Pb1 – I2	85.472(3)
I2 – Pb1 – I2	174.363(12)
I2 – Pb1 – I2	100.147(12)



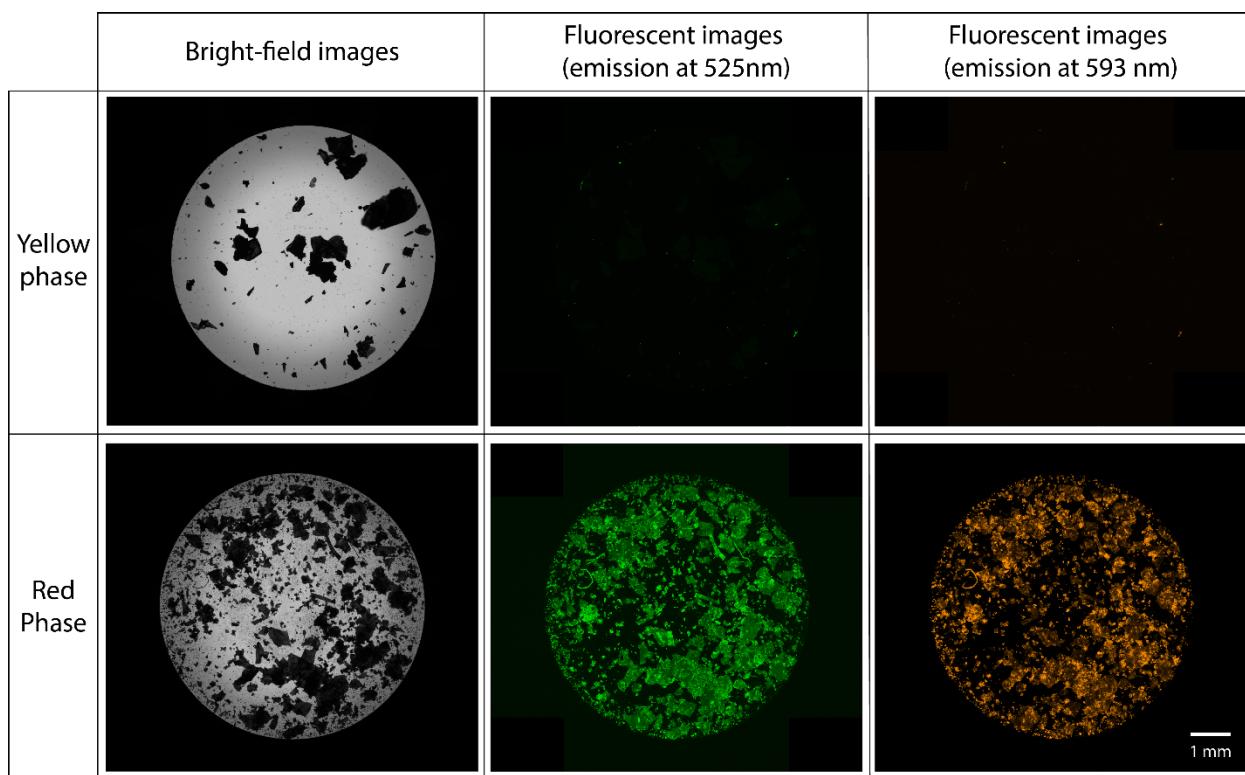
**Figure S2.** Powder X-ray diffraction (pXRD) patterns of the yellow phase (a) and red phase (b). Black lines are experimental pXRD patterns; red lines are simulated pXRD patterns based on structures derived from single crystal XRD analysis.



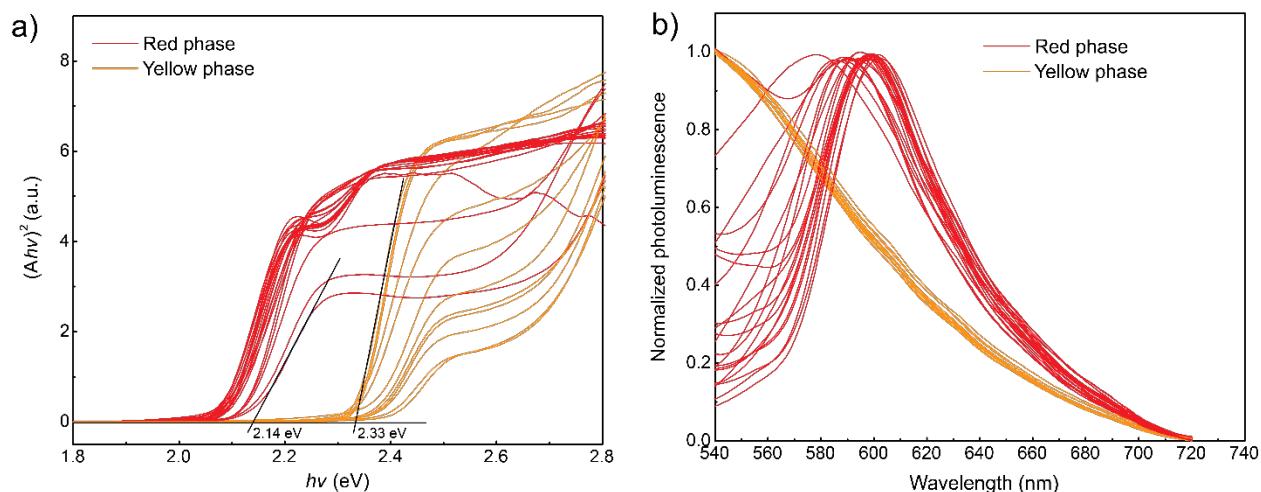
**Figure S3.** Robotic system and perovskite synthetic workflow based on high-throughput ASVC.



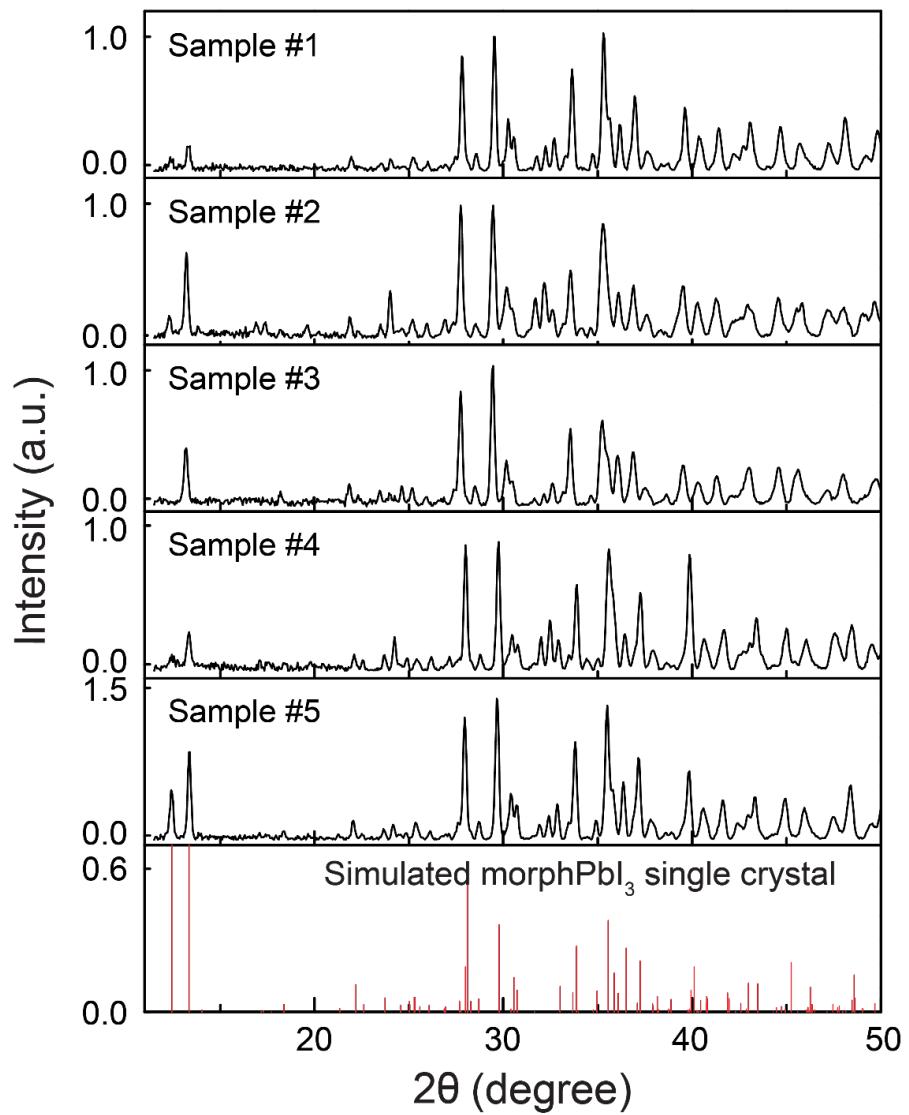
**Figure S4.** High-throughput tools and characterization workflow for metal halide perovskites.



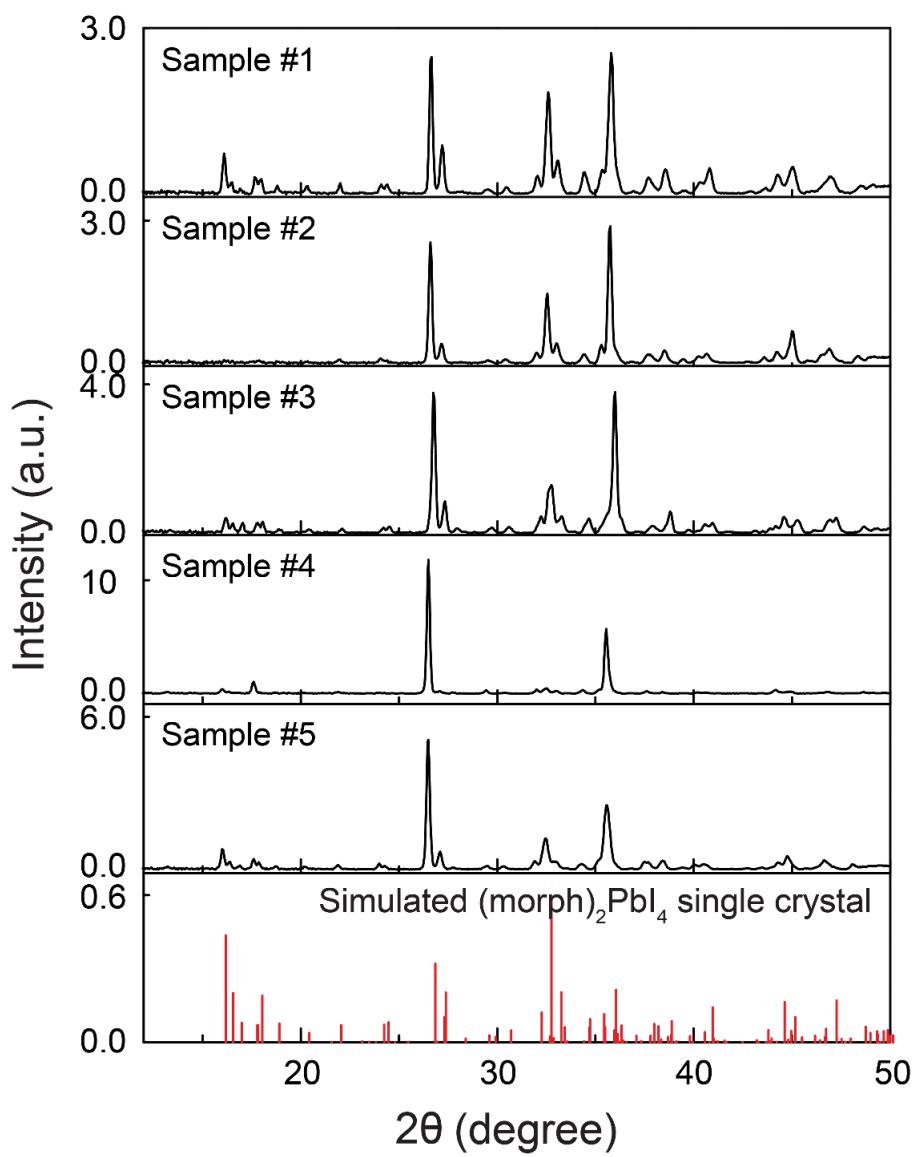
**Figure S5.** Representative high-throughput microscopic images [bright-field and fluorescent ( $\lambda_{\text{ex}} = 469$  nm)] for the yellow and red phases.



**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] <sup>a</sup>	Lower bound: 0; Upper bound: 2.91 M (concentration of morphI in Reagent 2)
[Pb] <sup>a</sup>	Lower bound: 0.1 M; Upper bound: 2.32 M (concentration of PbI <sub>2</sub> in Reagent 2)
[FAH] <sup>a</sup>	Lower bound: 0; Upper bound: 16 M
Solvents <sup>b</sup>	DMF
Crystallization temperature <sup>b</sup>	20 °C
Crystallization time <sup>b</sup>	16 h

<sup>a</sup> Variables that have been varied.<sup>b</sup> Variables that have been kept constant.**Table S5.** Experiment parameters and their constraints in 6D screening.

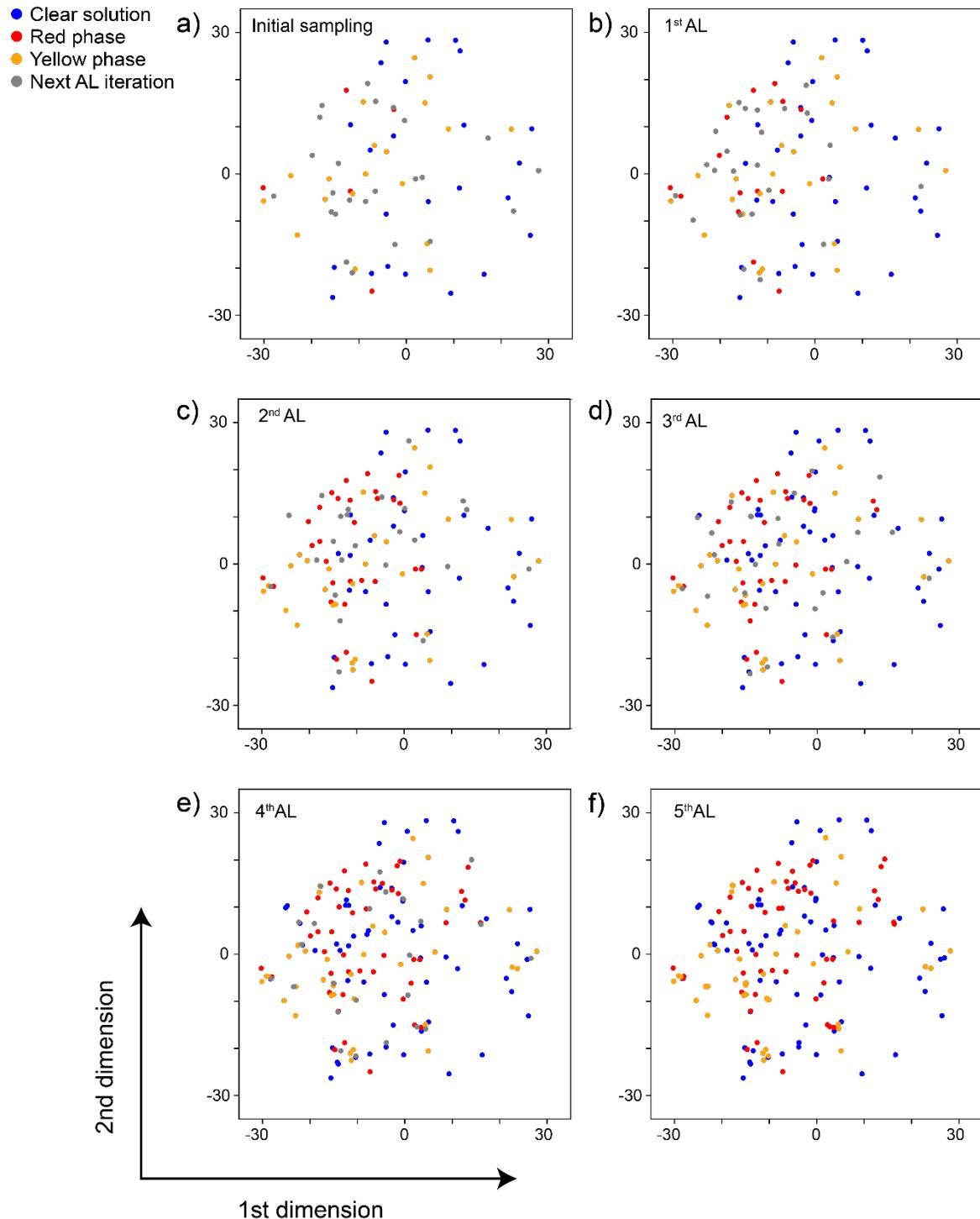
Experiment variables	Constraints
[morph] <sup>a</sup>	Lower bound: 0; Upper bound: 2.91 M (concentration of the morphI in Reagent 2)
[Pb] <sup>a</sup>	Lower bound: 0.1 M; Upper bound: 2.32 M (concentration of PbI <sub>2</sub> in Reagent 2)
[FAH] <sup>a</sup>	Lower bound: 0; Upper bound: 16 M
[H <sub>2</sub> O] <sup>a</sup>	Lower bound: 0; Upper bound: 13.5 M
V <sub>f</sub> (DMSO) <sup>a</sup>	Lower bound: 0; Upper bound: 1
V <sub>f</sub> (GBL) <sup>a</sup>	Lower bound: 0; Upper bound: 1
Solvents <sup>a</sup>	DMF, DMSO, GBL
Crystallization temperature <sup>b</sup>	20 °C
Crystallization time <sup>b</sup>	16 h

<sup>a</sup> Variables that have been varied.<sup>b</sup> 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

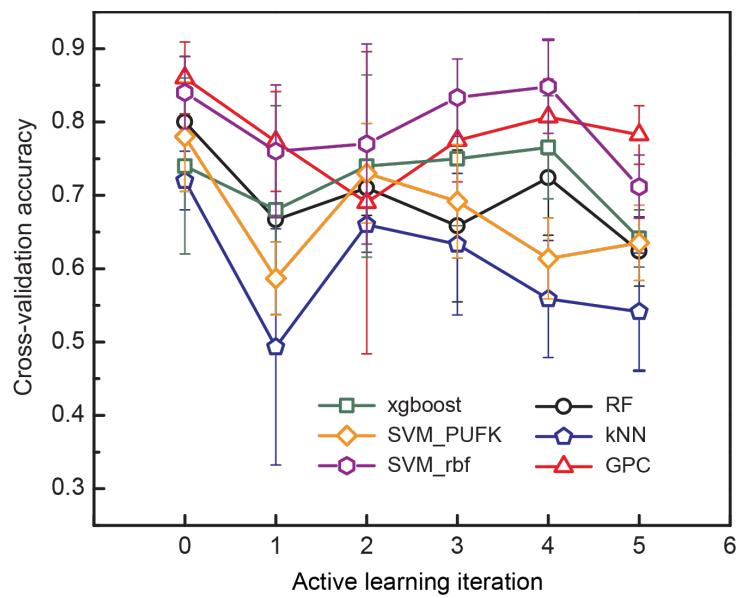
\* see explanations of the hyper-parameters in Scikit-learn documents at [https://scikit-learn.org/stable/supervised\\_learning.html](https://scikit-learn.org/stable/supervised_learning.html)



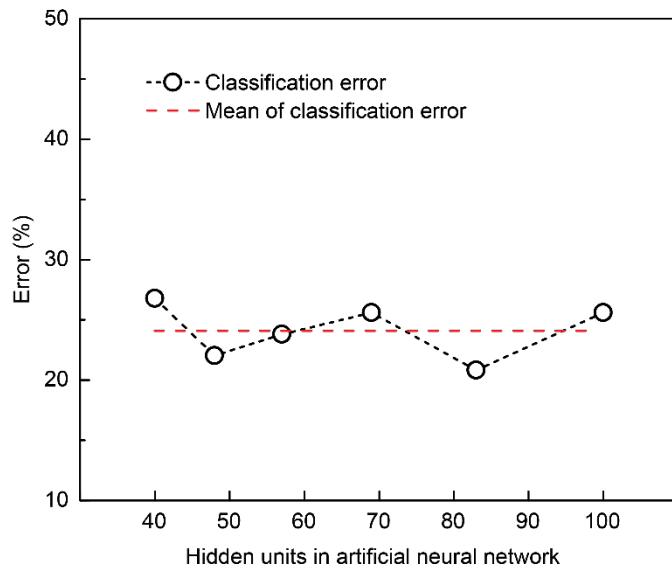
**Table S7.** Machine learning models tested after each iteration of AL (their CV accuracies are shown in Figure S10)

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)

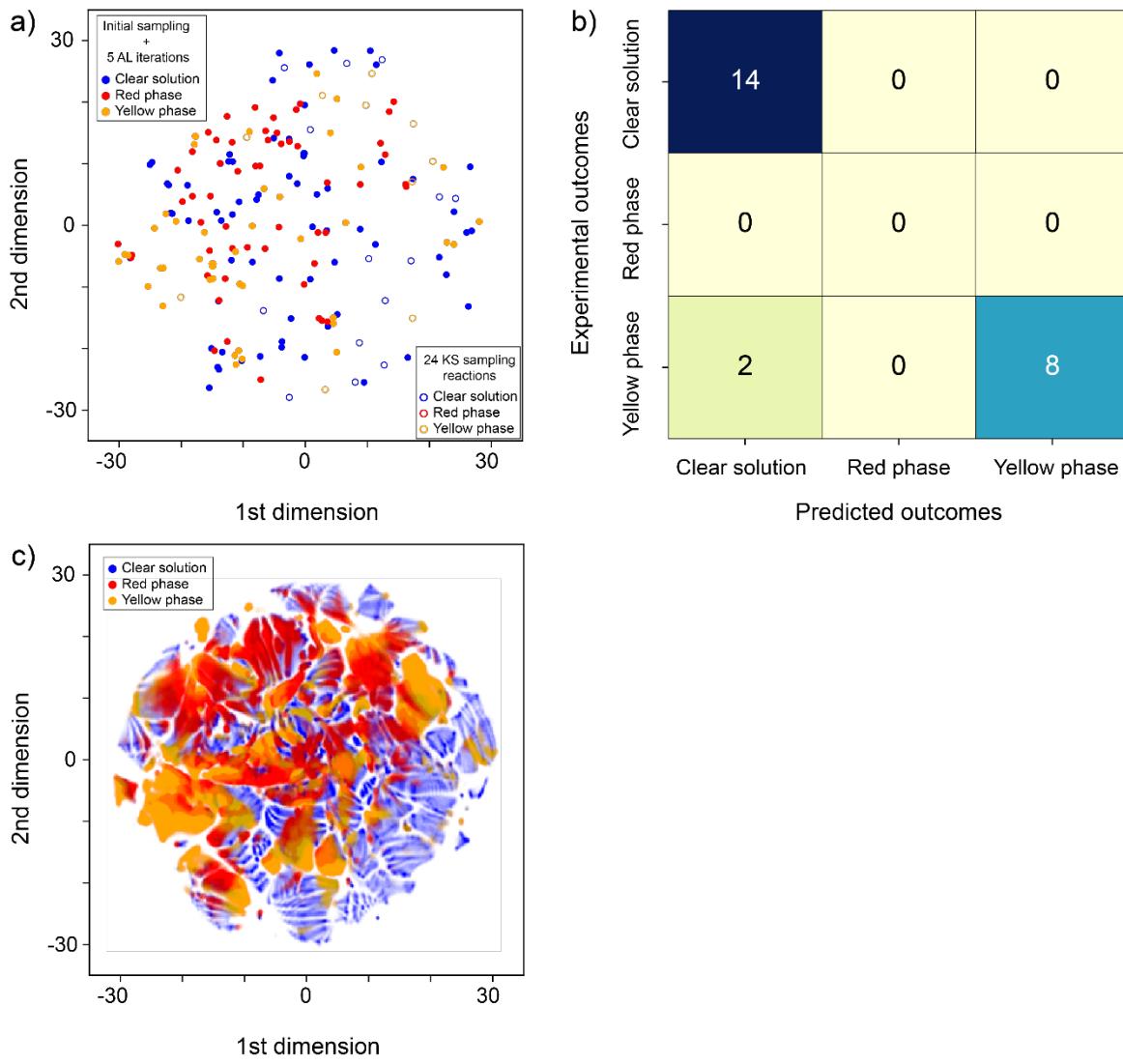
\* see explanations of the hyper-parameters in Scikit-learn and XGBoost documents [ref]



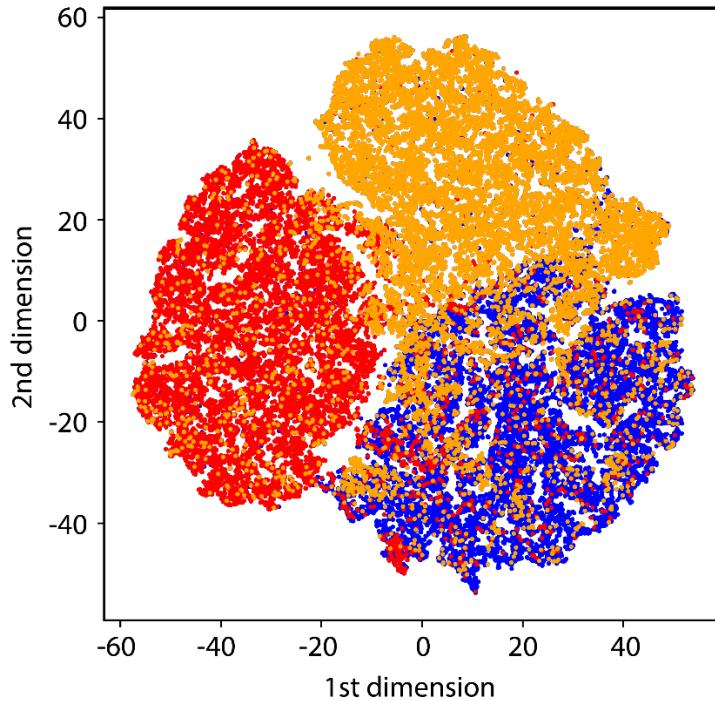
**Figure S10.** Cross-validation accuracies of different machine learning models on the dataset collected after each iteration of AL.



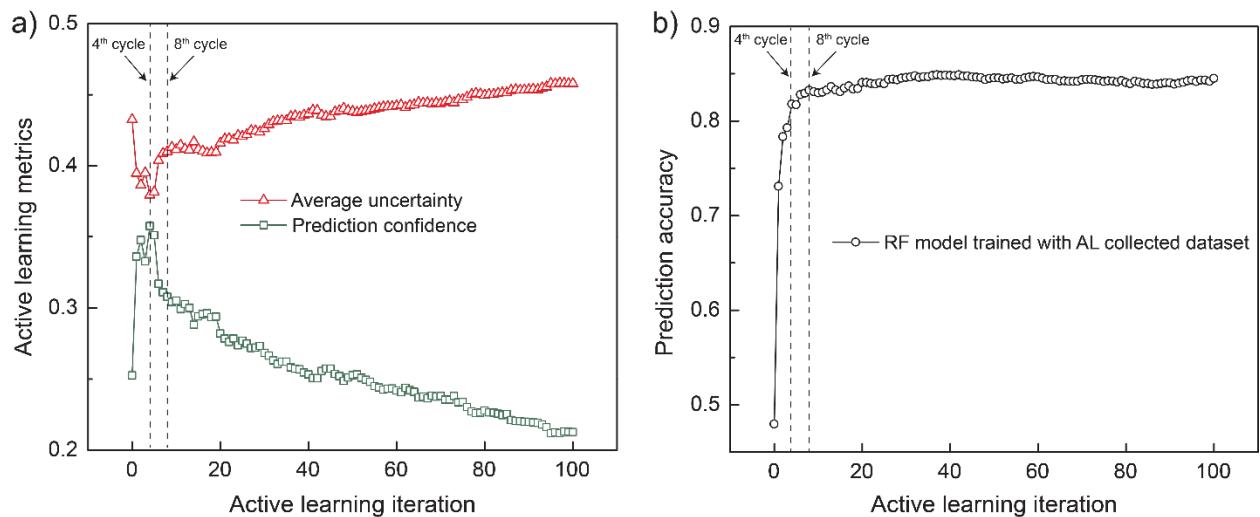
**Figure S11.** Classification errors (one minus cross-validation accuracies) of artificial neural networks with different numbers of hidden units. The red dashed line represents the mean value of the classification errors.



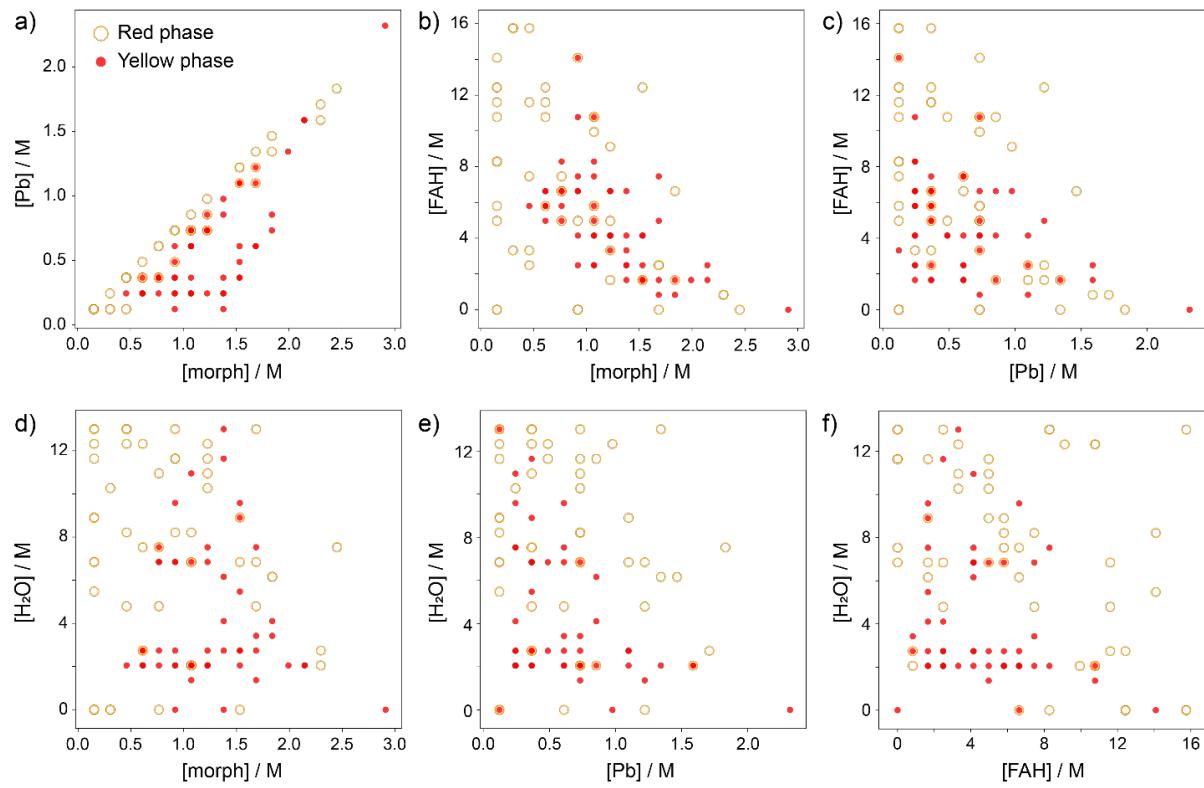
**Figure S12.** a) Projections of the reaction-composition space using t-SNE for initial sampling + five AL iterations and 24 unseen reactions (labeled by experimental outcomes) selected by the KS algorithm. b) Confusion matrix of experimental outcomes vs. predicted outcomes. c) Projections of the reaction pool labeled with predicted outcomes.



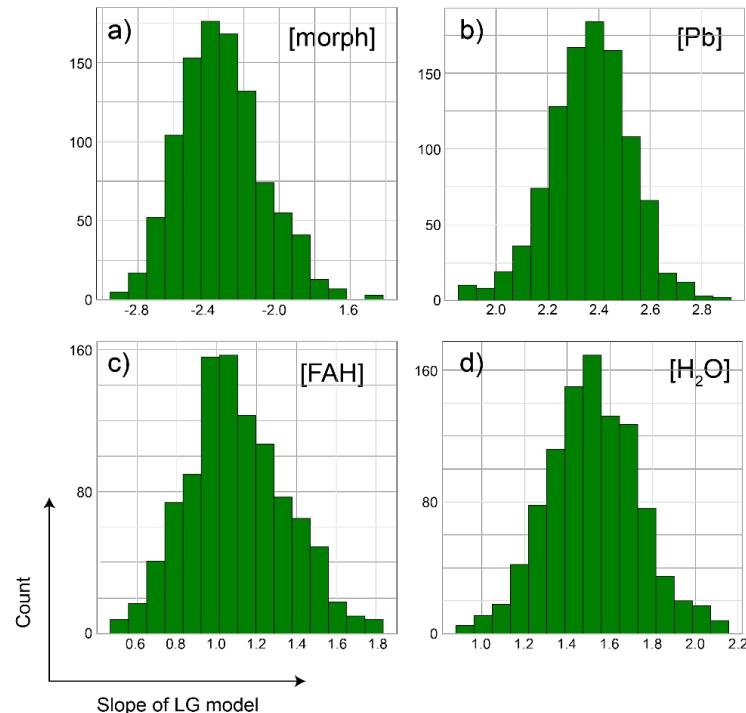
**Figure S13.** Projections of the 6D synthetic dataset (50,000 data points) using t-SNE.



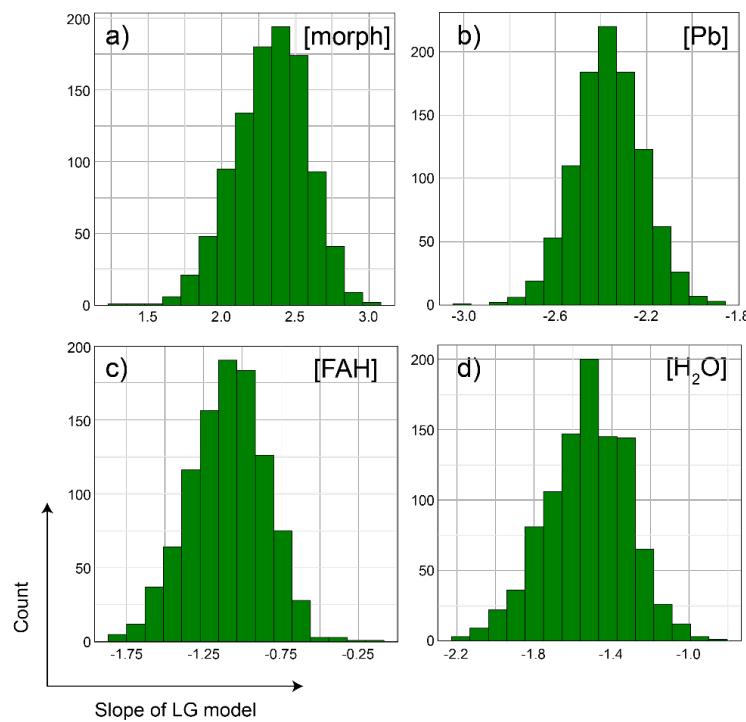
**Figure S14.** (a) Prediction confidence and average uncertainty of the RF model on the synthetic dataset in each iteration of AL. (b) Prediction accuracy (on the whole synthetic dataset) of the RF model in each iteration of AL.



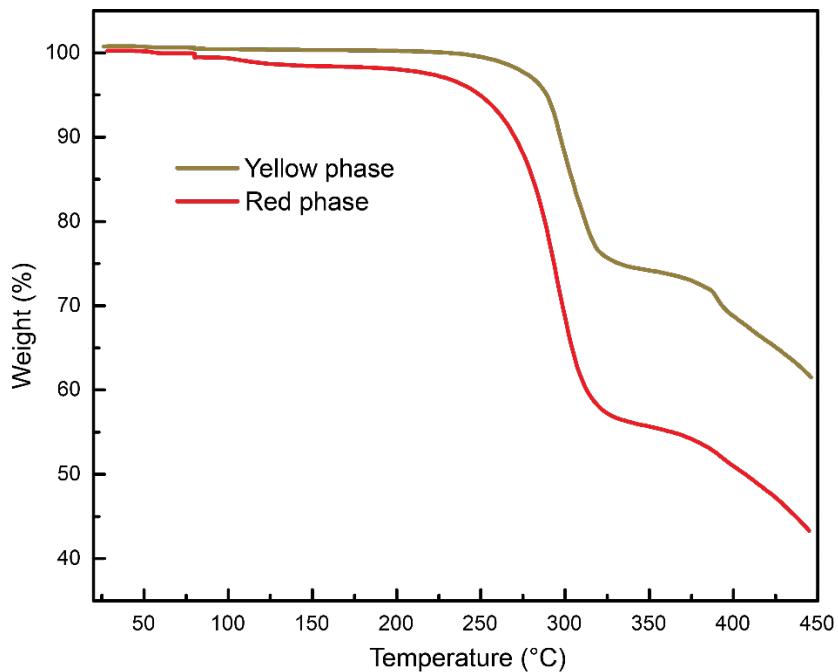
**Figure S15.** 2D projection plots of the “crystals only” dataset using axis from [Pb], [morph], [FAH], and [H<sub>2</sub>O].



**Figure S16a.** Distributions of logistic regression (LR) slopes for a) [morph], b) [Pb], c) [FAH], and d) [H<sub>2</sub>O] when treating “yellow phase” class as “1”.



**Figure S16b.** Distributions of logistic regression (LR) 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.jpcllett.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) 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>.
- (12) 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>.

- (13) 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/DONH00270D>.
- (14) Bruker (2018). Apex3 V2018.1-0, SAINT V8.38A, Bruker AXS Inc.: Madison (WI), USA, 2018.