Supporting Information

Robot-Accelerated Perovskite Investigation and Discovery

Zhi Li,[§] Mansoor Ani Najeeb,[‡] Liana Alves,[‡] Alyssa Z. Sherman,[‡] Venkateswaran Shekar,[∥] Peter Cruz Parrilla,[‡] Ian M. Pendleton,[‡] Wesley Wang,[§o] Philip W. Nega,[§] Matthias Zeller,[#] 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 Computer Science, Haverford College, 370 Lancaster Avenue, Haverford, Pennsylvania 19041, USA
[o]Department of Chemistry, University of California, Berkeley, 420 Latimer Hall, Berkeley, California 94720, USA
[#]Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette IN 47907, USA
[†]Department of Chemistry, Fordham University, 441 E. Fordham Road, The Bronx, New York, 10458, USA

*To whom the correspondence should be addressed: jschrier@fordham.edu, anorquis@haverford.edu, and EMChan@lbl.gov.

Table of Contents

1. Materials and methods

Chemicals

Lead iodide and formic acid were purchased from Sigma Aldrich Chemicals with 99% and ≥95% purity, respectively. Methylammonium iodide (98%), ethylammonium iodide (98%), n-butylammonium iodide (98%), formamidinium iodide (≥98%), cyclohexylmethylammonium iodide (98%) and all other ammonium iodides in this work were purchased from GreatCell Solar. γ-Butyrolactone (GBL) (≥98%) was purchased from Spectrum Chemical. Dimethyl sulfoxide (DMSO) (≥99.5%) and dimethylformamide (DMF) (anhydrous, 99.8%) were purchased from Sigma Aldrich Chemicals.

Equipment

Liquid handling robot

A Hamilton Microlab NIMBUS4 liquid handling robot was used in this study for the high-throughput synthesis of metal halide perovskites. 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. Inverse temperature crystallization reactions were performed on a Hamilton Heater and Shaker II (HHS2) module, which can be heated up to 105 °C (actual 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. A complete, step-by-step procedure for our robotic workflow is shown in the "Robotic Workflow" section.

Powder X-ray Diffraction

Powder X-ray diffraction 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-Vis absorption spectra

UV-visible absorption spectra were collected using an Agilent Cary-5000 UV-Vis-NIR spectrophotometer. Absorbance spectra of powders ground from large perovskite crystals were measured using an internal diffuse reflectance accessory.

Photoluminescence spectra

Photoluminescence (PL) spectra of powders ground from large perovskite crystals were measured using a Horiba Jobin Yvon Fluorolog-3 spectrofluorometer. PL spectra were collected from 365 to 650 nm with 1 nm wavelength steps and 0.01 s integration time per step.

Conductivity measurements

Conductivity measurements were performed on $AcetPbI_3$ single crystals using a customized probe station. Selected needle-shaped $AcetPbI_3$ single crystals with lengths of $0.3 - 0.6$ mm were placed on double-sided tape on clean glass slides. Care was taken not to break the crystals. Carbon paste was applied at opposite ends of the needle crystals as electrodes. In this configuration, the cross-sectional area of electron transport is approximately the cross-sectional area of the needle crystals, which is in the range of $0.3 - 0.4$ mm^2. Platinum probes were placed to form contacts with carbon electrodes, and a Keithley 2400 source meter was used to measure the *I-V* plots. Due to the high electrical resistance of the $AcetPbI_3$ crystals, the current across a single crystal is very small ($<1nA$), requiring that the internal resistance of the source meter be taken into account. The conductivity of $AcetPbI_3$ single crystals was estimated based on a shunt model containing $AcetPbI_3$ single crystals and internal resistance of the source meter (200 MΩ).

Solubility Measurements

The solubility of lead-ammonium compounds fundamentally depends on the solvent used (GBL, DMF or DMSO), the temperature, and the concentration of ammonium iodide in the solution. To determine the lead iodide solubility for the crystallization reactions, we weighed the lead iodide with different percentages of a selected ammonium iodide and stirred the suspensions in a heated oil bath at 75 °C and 450 rpm. The Solvent 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 as a homogeneous solution during the robotic run and that the lead content did not precipitate, which can disrupt the pipetting of liquid handler. The solubility was then calculated as total moles of PbI_2 or ammonium iodide divided by total volume of solvent being added. The solubilities of individual ammonium iodides were measured using the same procedure. Solubility data for lead iodides with individual ammonium iodide compounds are shown in Table S2.

Robotic Workflow

Workflow overview

Stock solutions of PbI_2-ammonium iodide mixture solution and ammonium iodide solution were prepared based on experimental data entry files generated by ESCALATE based on the compounds' solubility data. After stock solution preparation, all reaction components (i.e., the stock solutions, pure solvent, and formic acid) were placed in programmatically designated locations on NIMBUS operation deck. After addition of all reaction components, the resultant solutions were kept at an actual temperature of 95 ºC (measured by IR camera at a set point temperature of 105 ºC) for crystallization for 2.5 hours. An experimental flow chart is shown in Figure S4, describing the selection of compounds (e.g., PbI_2, ammonium iodides, solvents), benchtop solubility test of ammonium iodide and PbI_2-ammomnium iodide mixture, reaction components, stock solution preparation, and robotic reaction. Figure S4 also contains a photograph of NIMBUS operation deck, labeled with the location of each reaction component. During crystallization, the NIMBUS robot was kept undisturbed and fume hood ventilation was minimized by closing fume hood sash. Ambient environmental conditions such as fume hood temperature and humidity were recorded throughout the time of the reaction and was later used for data analysis and optimization. The 8x43 mm (diameter x height) glass scintillation vials were used as reaction vessels. 96, 1-ml glass vials were loaded on an SBS formatted aluminum microplate (Symyx Technologies, now Unchained Labs) on the HHS2 heater-shaker module. The vials were pre-heated to the set temperature before addition of the stock solutions. Formic acid addition was separated into two injections per well and followed by 15 and 20 minutes shaking after first and second addition, respectively to avoid premature precipitation of PbI_2, ammonium iodide or perovskite. The actual, step-by-step protocol of the robot is given below.

Step-by-step robotic procedure

1. Utilizing solubility data from Table S1, stock solutions are manually prepared with concentrations specified by ESCALATE and are named as Reagents 1-4 in order of their addition during the liquid handling process. Reagent 1 is pure solvent (GBL, DMF or DMSO), and Reagent 4 is formic acid. They are used without any additional treatment. Reagent 2, a stock solution of PbI_2 and ammonium iodide in the same solvent as Reagent 1, is prepared by adding PbI_2, ammonium iodide and then solvent to the stock solution container (in that order). Reagent 3 is the ammonium iodide solution, prepared by adding solvent to the ammonium iodide. Reagent 2 and Reagent 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 as a clear solution. Then, all reagent solutions are manually loaded into designated locations on liquid handler deck and kept at room temperature. Reagent 1 is stored in a reusable 50 mL polypropylene solvent container. 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 reaction microplate is pre-heated to a set point temperature at 105 ºC, with the actual temperature of the HHS2 reactor module measured and recorded in the software. The heating rate is approximately 7 ºC/ min.

3. After the measured temperature of the HHS2 reaches 80 ºC, the liquid handler dispenses Reagent 1 (0-500 µL), into each of the 96, 1-mL glass vials in the aluminum microplate, followed by Reagent 2 (0-500 µL) and Reagent 3 (0-500 µL). The temperature of the HHS2 will reach 105 ºC (95 ºC for solution) during dispensing of Reagents 1-3. The combined volume of Reagents 1-3 dispensed in each vial is maintained at 500 µL. The liquid dispensing process is programmed and automatically carried out without any delay time between each dispense.

4. After addition of Reagents 1-3, formic acid (Reagent 4) is dispensed into the 96 vials in two injections per vial, with 0-300 µL for each injection. The first addition of formic acid is followed by 15 minutes of shaking at 750 rpm, while the second addition of formic acid is followed by 20 minutes of shaking.

5. After the 20 minutes of shaking, the perovskite precursor solution is kept at 95 ºC (actual solution temperature) without shaking for 2.5 hours to allow for undisturbed crystal growth.

6. After 2.5 hours of crystal growth, the HHS2 temperature is maintained at 105 ºC while operators visually score reaction outcomes in solution. Operators remove 8 vials at a time from the reaction microplate and inspect each vial. Operators classify the reaction outcomes as: (1) clear solution without any solid; (2) fine powder; (3) small crystallites (crystal size < 0.1 mm); and (4) large (>0.1 mm) crystal(s) suitable for structure determination by single crystal x-ray diffraction (see the following section for detailed scoring criteria). To prevent the crystals from dissolving at room temperature, the total time of scoring and recording for each vial should take less than 4 seconds. After scoring all 96 reactions, only representative high scoring crystals with distinct sizes and shapes are isolated from selected vials via hot filtration in a Büchner funnel and washed with diethyl ether to remove residual reaction solution. The resulting crystals are ground into powders for powder XRD analysis. The powder XRD patterns can be visualized using the online scripts provided (see "Interactive Data Visualization and Analysis Interface" section).

Crystal Scoring Rubric

To score reaction outcomes, reaction vials were inspected at different angles by eye. We found such visual inspection to be more sensitive and flexible for scoring reaction outcomes compared to the side photographs by fixed camera settings, which were also taken for the record. A detailed crystal scoring rubric is shown below in Table S1. When multiple crystal types were observed in the same vial, we used the highest overall score for each reaction.

Table S1. Crystal scoring rubric. Scale bar: 1 mm. (see Figure S2 for more photos of different crystal classes)

Crystal Scores	Criteria	Photos
1	**No solid** observed in the solutions.	[Clear solution]
2	**Fine powder** observed with no visible crystal facets. Under illumination, reflection is diffuse rather than specular (as would be the case for the presence of smooth facets).	
3	**Small crystallites** observed with approximate size 0.1 mm or smaller. Distinct from Class 2 (fine powder) in exhibiting specular reflection at certain angles, indicating the presence of crystal facets.	
4	**Large crystals** (> 0.1 mm) with straight edges and large area of specular reflection from crystal facets.	
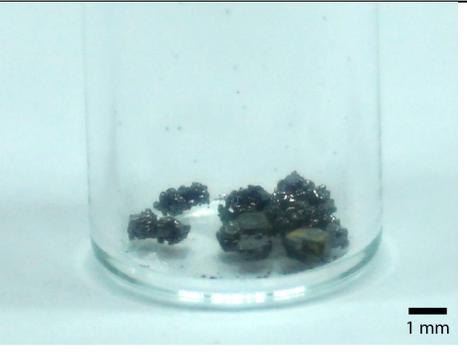
Software and Computation

Software and packages

Our custom-developed pipeline software: ESCALATE,[1] is used to specify experimental parameters in robot readable files, provide instructions for human operators, and capture experiment results and observations. The algorithms for data visualization and analysis were written in Python 3.6 in Jupyter notebooks using the following libraries: Numpy 1.14.6, Pandas 0.22.0, Scipy 1.0.1, Matplotlib 3.1.0 and Scikit-learn 0.19.2. We used the Quickhull algorithm implemented in Scipy to calculate the convex hull (as shown in Figure 6a and Figure S26 - S43).[2]

Features for machine learning

In machine learning, features are defined as measurable properties or characteristics of phenomena being observed. Selecting relevant, informative and independent features have positive impacts on machine learning performance. In the work, we select a total of 75 features to describe the robot-ready perovskite synthesis reactions. Those features include 8 reaction conditions: PbI_2 concentration, ammonium iodide concentration, formic acid concentration, 1st mixing time, 2nd mixing time, reaction temperature, reaction time and stirring rate. The other 67 features are computed property descriptors of organoammonium and are listed in Section 3 of this document ("Explanation of Features/Descriptors for Machine Learning Modeling").

Train/test split and Cross-validation

We used a "Stratified Shuffle Split" method from Sciki-learn to generate training/testing datasets for cross-validation 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.

Prediction metrics

Experimental results are interpreted as two-class classification problem: Class 4 formation was considered as a "positive" result while non-Class 4 formation was considered as "negative" result. Typical metrics for machine learning models are defined below. Here, TP is the total number of "True Positive" results (i.e., the predicted and actual scores are both Class 4). 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.

Accuracy = (TP + TN)/(TP + TN + FP + FN)
Precision = TP/(TP + FP)
Recall = TP/(TP + FN)
F1 = (2*Recall*Precision)/(Recall + Precision)
AUC is "Area Under Curve" for "Receiver Operating Characteristic" Curve.

Building machine learning models

We used modularized machine learning packages from Scikit learn Python library. Models used are Support Vector Machine (with RBF: Radial Basis Function and PUFK: Pearson VII Universal Function Kernel),[3] k-Nearest Neighbors, Gaussian Naïve Bayes, Quadratic Discriminant Analysis, Logistic Regression, Decision Tree, Random Forest, Neural Network. Model hyper-parameters were optimized using a grid search method with 5-fold cross validation for highest prediction accuracy, trained by random 80% of total datasets and then the machine learning models with optimized hyper-parameters were tested by the rest of 20% datasets which machine learning models have never been trained on. The test results/metrics (accuracy, precision, recall, etc.) are listed in Table S8.

2. Interactive Data Visualization and Analysis Interface
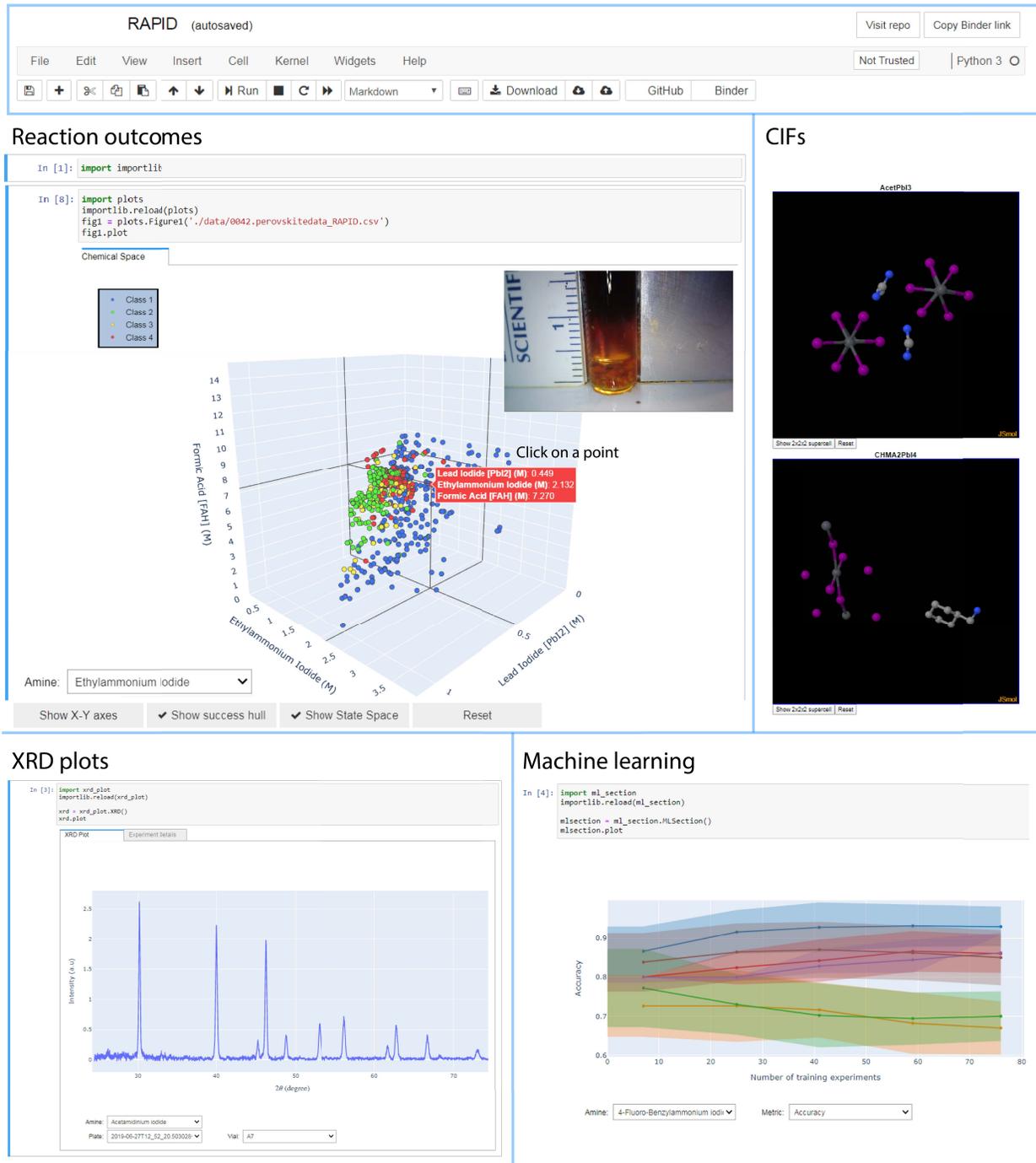
Figure S1. Screenshot of the data visualization and analysis interface.

The data visualization and analysis of all 45 chemical systems (8172 reactions), including 26 systems not shown in the manuscript, are available through our interactive data visualization and analysis interface (see Figure S1 for a screen shot of the webpage). The interface is an online script written in Python in a Jupyter notebook. The link to the Jupyter notebook is given below:

https://mybinder.org/v2/gh/darkreactions/rapid/master?filepath=RAPID.ipynb

The Jupyter notebook provides a graphical user interface that allows users without coding experience to navigate and view reaction outcomes, CIFs, powder-XRD patterns, and machine-learning plots. User instructions are provided on the web page and also below:

Instructions for visualizing data via MyBinder:

1. Use a web browser to load the URL above.
2. **The MyBinder page takes a few minutes (up to 20 min) to load** the data from a separate GitHub repository. If you see a spinning color wheel, the page is still loading. Please be patient!
3. When the interactive notebook is ready to use, you will see a menu bar at the top (see screenshot below). **In the "Cell" menu, select "Run All"** to display the data visualizations.
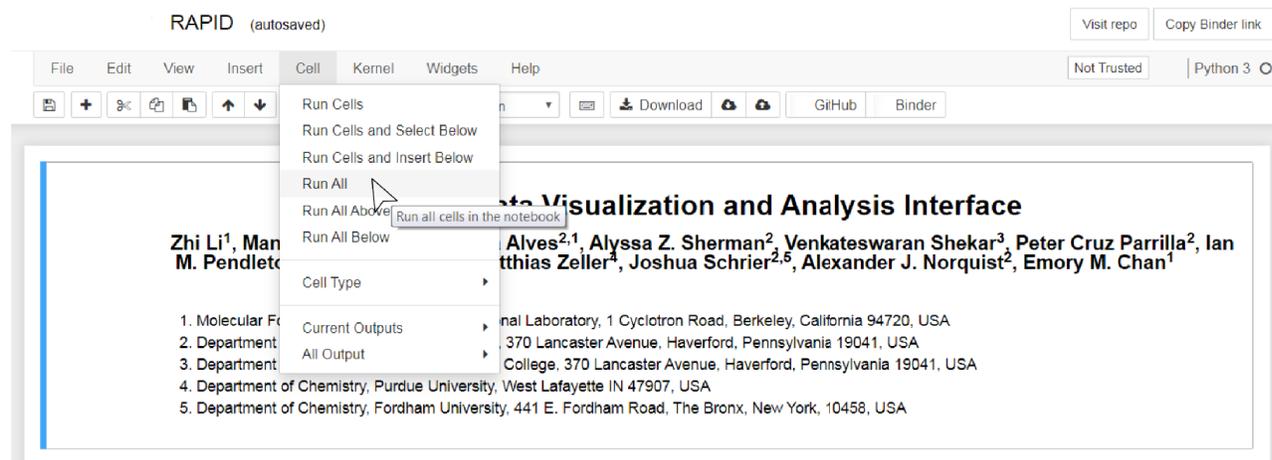
4. It may take a few minutes on average for the cells to process. If you do not see graphs that resemble Figure S1 after waiting a few minutes, refresh the page and try again starting with Step 2.
5. Once the visualizations are live, a user can interact with each plot to dive deep into the data, as described below for each data type:

- **Reaction outcomes**. After selecting a chemical system from the drop-down list, one can visualize all reaction outcomes of the system in a 3D plot similar to Figure 1. Users can drag to rotate the plot and scroll to zoom in and out. Click the data points in the plot to view the experimental conditions and images of crystals made with those reaction conditions. Convex hulls containing the experimental space and Class 4 space (see Figure 6a) can both be added on top of the plot by toggling on "Show State Space" and "Show success hull" tags (you made need to remove the convex hulls to click on certain data points).

- **CIFs**. Unit cells and $2\times2\times2$ supercells of AcetPbI$_3$ and (CHMA)$_2$PbI$_4$ single crystals can be visualized by clicking "Show Molecule" buttons. By scrolling and dragging, the crystals structures can be zoomed and rotated.

- **XRD plots**. Users can view the powder XRD patterns of reaction products from each chemical system by selecting corresponding drop-down option. The 2Θ angles and intensities are displayed when the cursor hovers over the curve. The related experimental conditions and crystal images are also attached.

- **Machine learning**. In this section, users can plot learning curves of 4 model metrics for all 45 chemical systems. The chemical systems and model metrics can be selected by clicking on the drop-down menu below the plot. Double clicking on one curve will remove all other curves from the plot.

6. **To access the source code for this notebook and the associated data** please visit: https://github.com/darkreactions/rapid

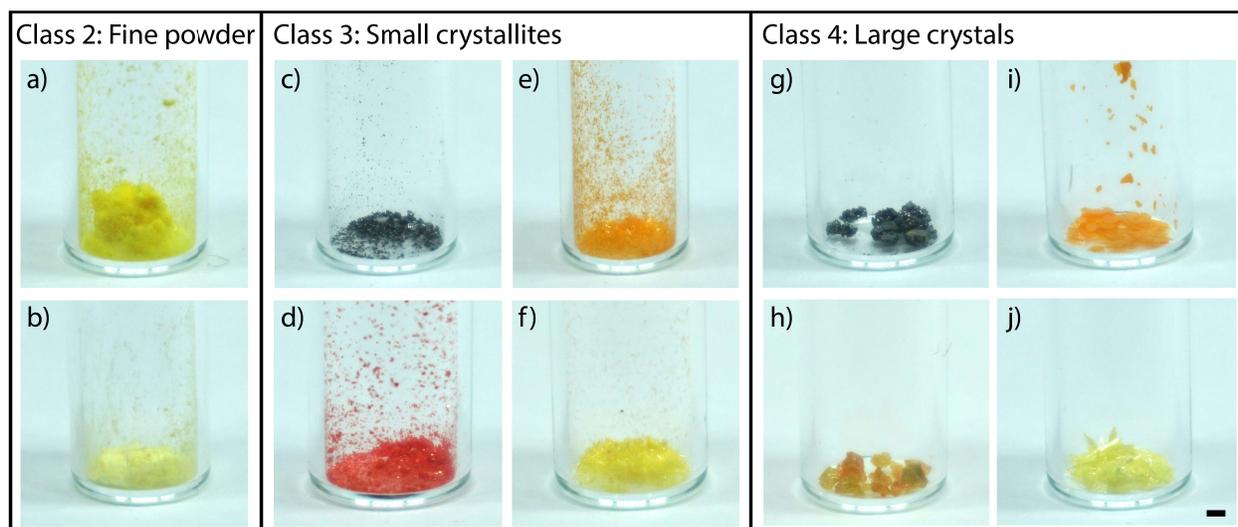# 3. Supplemental Figures and Tables

Figure S2. Representative photographs showing examples of crystals scored as Class 2 (fine powder): Piper / PbI_2 (a) and DiMA / PbI_2 (b). Reaction outcomes scored as Class 3 (small crystallites): MA / PbI_2 (c), Morph / PbI_2 (d), CHMA / PbI_2 (e) and 4FBenA / PbI_2 (f). Reaction outcomes scored as Class 4 (large crystals): MA / PbI_2 (g), EtA / PbI_2 (h), CHMA / PbI_2 (i) and Acet / PbI_2 (j). Scale bar: 1 mm.
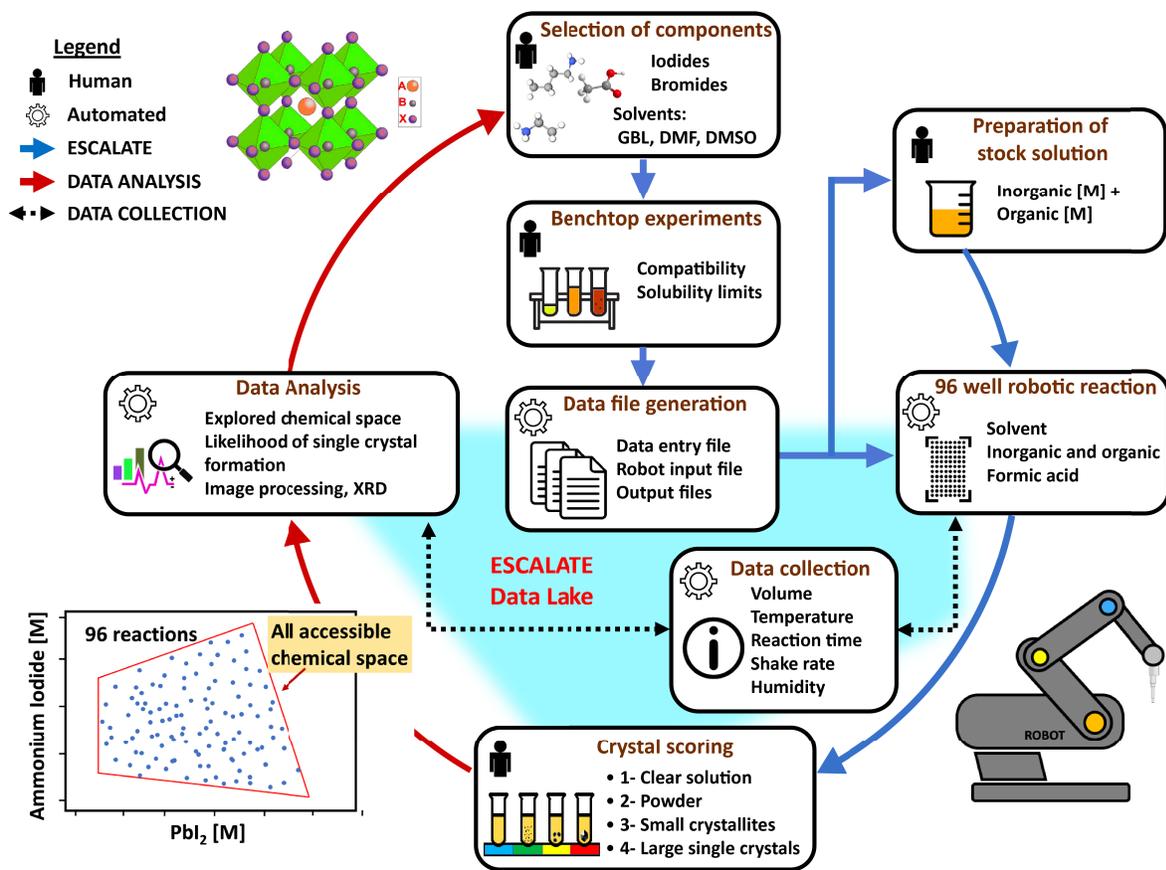
Figure S3. Life cycle of an experiment.
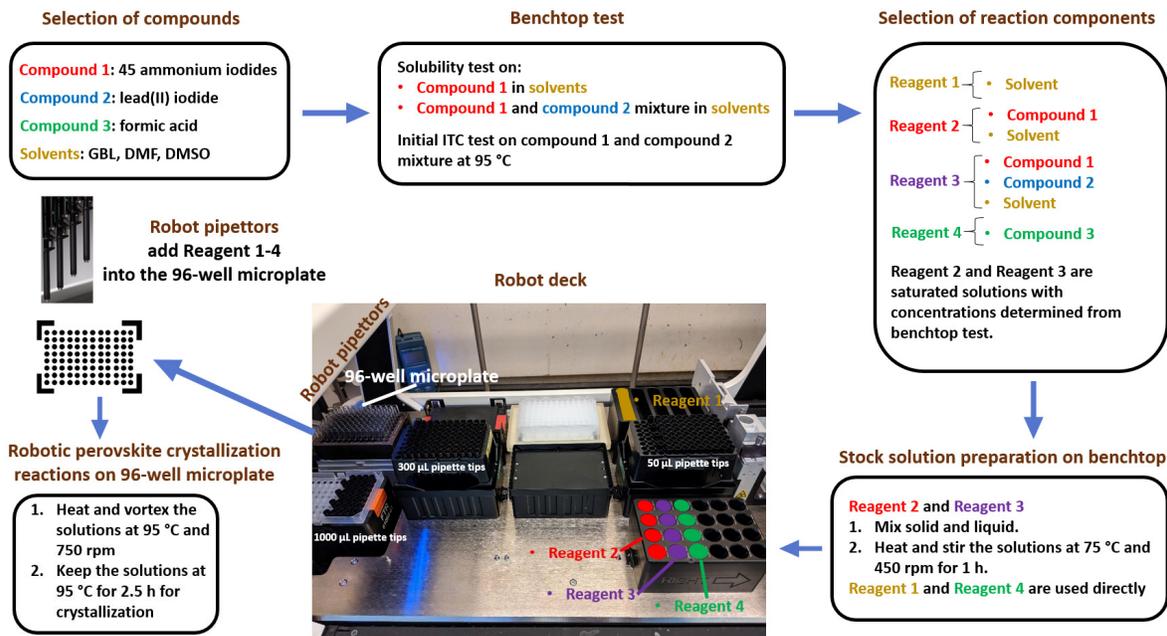
Figure S4. Robotic system and process flow for designing and performing automated reactions.
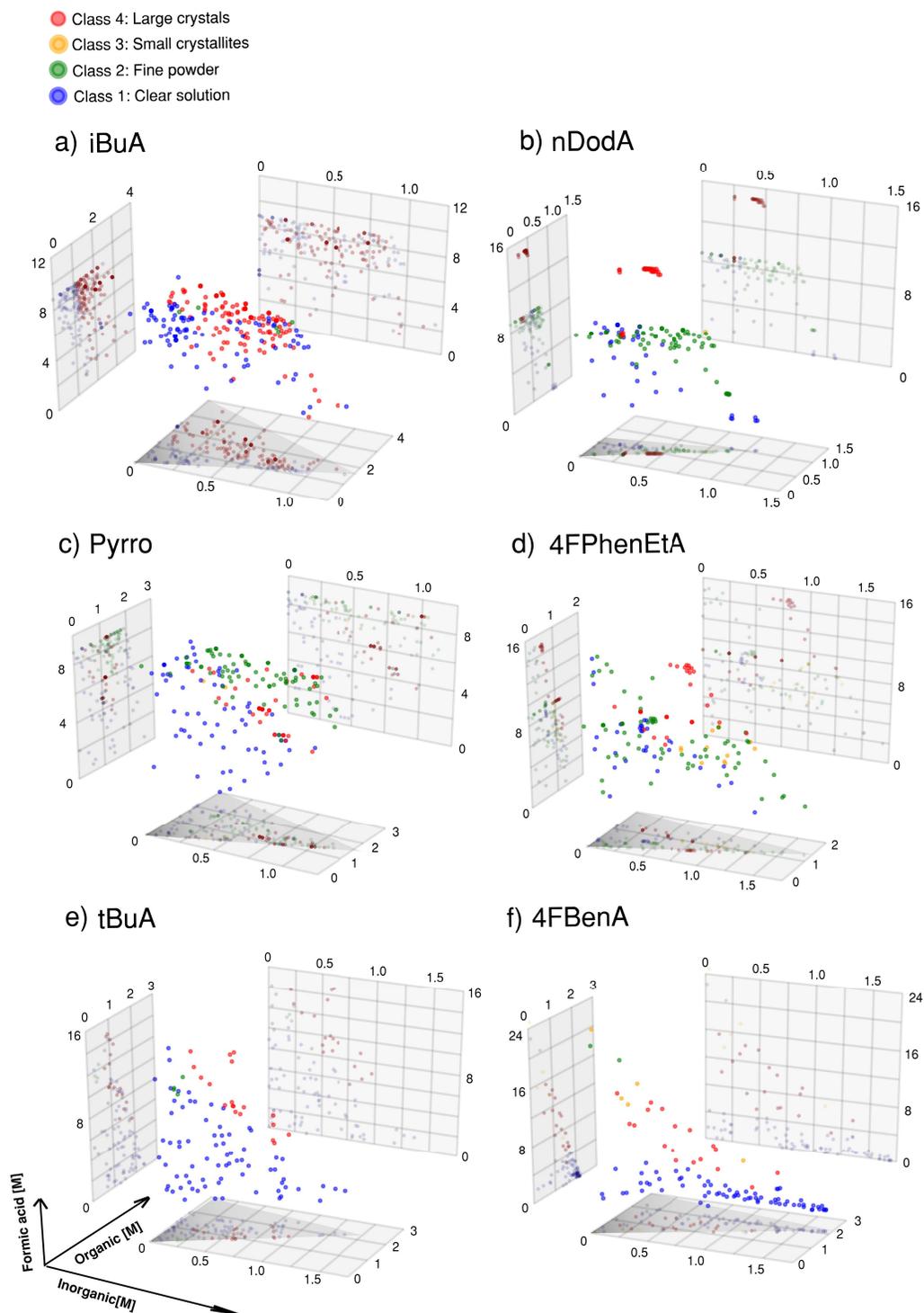
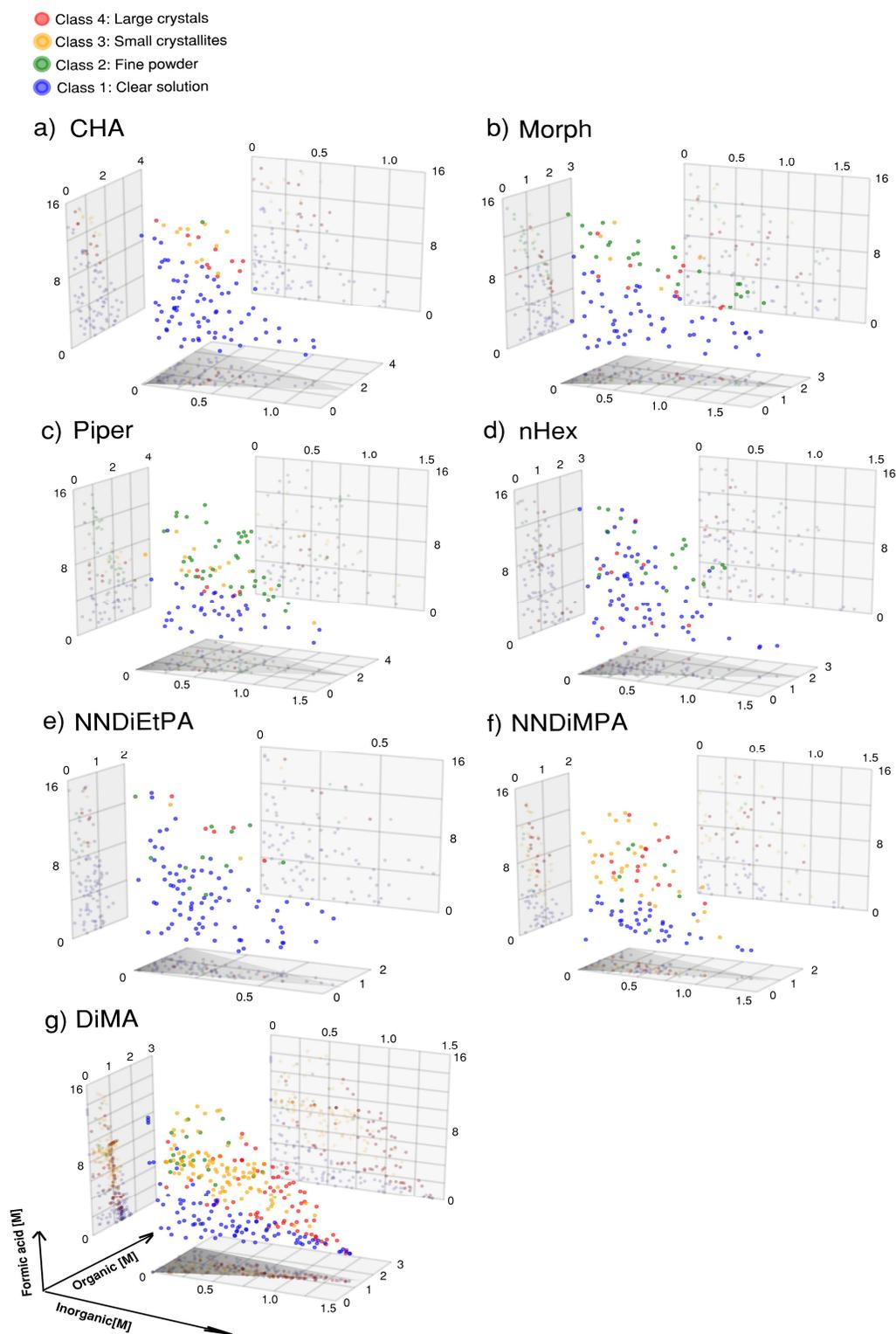**Figure S5.** Crystal quality of 912 lead iodide perovskite reactions as a function of the organic cation and the concentrations of the organic, inorganic (PbI$_2$), and formic acid precursors. Organic precursors are iso-butylammonium iodide (a), n-dodecylammonium iodide (b), pyrrolidinium iodide (c), 4-fluoro-phenethylammonium iodide (d), t-butylammonium iodide (e), 4-fluoro-benzylammonium iodide (f). Light gray boxes show two-dimensional projections of the data. Dark gray triangles illustrate the accessible experimental space in the organic-inorganic plane, as constrained by precursor solubilities.

Figure S6. Quality of 867 lead iodide perovskite reactions as a function of the organic cation and the concentrations of the organic, inorganic (PbI$_2$), and formic acid precursors. Organic precursors are cyclohexylammonium iodide (a), morpholinium iodide (b), piperidnium iodide (c), n-

hexylammonium iodide (d), N,N-diethylpropane-1,3-diammonium iodide (e), N,N-dimethylpropane-1,3-diammonium iodide (f) and dimethylammonium iodide (g). Light gray boxes show two-dimensional projections of the data. Dark gray triangles illustrate the accessible experimental space in the organic-inorganic plane, as constrained by precursor solubilities.

Table S2. Solubility data for the different ammonium iodides used in this work.

Ammonium Iodides (Abbreviations)	Maximum Concentration of PbI_2 in solvents (M)	Required Ammonium Concentration to reach maximum PbI_2 (M)	Required Ammonium Percent (for maximum PbI_2) (%)	Maximum Concentration Ammonium (M)	Solvents
Acetamidinium iodide (Acet)	1.26	0.96	76.2	3.82	GBL
Guanidinium iodide (G)	1.4	1.41	100.7	4.07	GBL
n-Butylammonium iodide (nBuA)	1.36	1.39	102.2	3.13	GBL
Ethylammonium Iodide (EtA)	1.1	2.2	200.0	3.99	GBL
Methylammonium iodide (MA)	0.95	0.67	70.5	2.14	GBL
Phenethylammonium iodide (PhenEtA)	1.66	0.83	50.0	2.04	GBL
Formamidinium Iodide (F)	1.47	1.1	74.8	3.64	GBL
Imidazolium Iodide (Im)	0.88	0.44	50.0	2.12	GBL
Benzylammonium Iodide (BenA)	1.47	1.49	101.4	2.74	GBL
neo-Pentylammonium iodide (neoPenA)	1.63	1.66	101.8	2.06	GBL
i-Propylammonium iodide (iPropylA)	1.73	2.14	123.7	2.85	GBL
iso-Butylammonium iodide (iBuA)	2.26	2.83	125.2	4.75	GBL
Dimethylammonium iodide (DiMA)	1.52	1.52	100.0	3.55	DMF
n-Dodecylammonium iodide (nDodA)	1.24	1.25	100.8	1.65	GBL

Pyrrolidinium Iodide (Pyrro)	1.55	1.94	125.2	2.9	DMSO
Ethane-1,2-diammonium iodide (EtDiA)	2.17	1.08	49.8	2.14	DMF
4-Methoxy-Phenylammonium iodide (4MeOPhenA)	1.56	1.56	100.0	3.11	GBL
N-propylammonium Iodide (nPropylA)	1.27	1.27	100.0	3.15	GBL
iso-Pentylammonium iodide (iPenA)	1.52	1.52	100.0	2.61	GBL
Propane-1,3-diammonium iodide (Pro13DiA)	1.34	1.01	75.4	1.9	DMF
4-Fluoro-Phenethylammonium iodide (4FPhenEtA)	1.85	1.39	75.1	2.12	GBL
4-Fluoro-Phenylammonium iodide (4FPhenA)	1.94	1.45	74.7	1.93	GBL
4-Trifluoromethyl-Benzylammonium iodide (4TFMBenA)	1.72	1.72	100.0	2.82	GBL
4-Trifluoromethyl-Phenylammonium iodide (4TFMPhenA)	1.67	2.08	124.6	2.81	GBL
tert-Octylammonium iodide (tOctylA)	1.4	1.4	100.0	2.23	GBL
Phenylammonium Iodide (PhenA)	2.55	2.55	100.0	2.77	GBL
t-Butylammonium Iodide (tBuA)	2.21	3.31	149.8	3.7	GBL
Cyclohexylammonium iodide (CHA)	2	2.5	125.0	3.42	DMSO
4-Fluoro-Benzylammonium iodide (4FBenA)	1.84	2.31	125.5	2.97	DMF
Pyridinium Iodide (Pyrid)	2.36	2.36	100.0	3.28	DMSO
Diethylammonium iodide (DiEtA)	1.66	1.82	109.6	3.11	GBL
Morpholinium Iodide (Morph)	1.5	1.89	126.0	2.26	DMF

n-Octylammonium Iodide (nOctylA)	1.37	1.74	127.0	2.53	DMSO
n-Hexylammonium iodide (nHexA)	1.88	1.85	98.4	2.87	GBL
Piperidinium Iodide (Piper)	1.62	1.99	122.8	4.54	DMSO
Cyclohexylmethylammonium iodide (CHMA)	1.04	1	96.2	2.07	GBL
Piperazine-1,4-diium iodide (PiperaDiium)	0.93	0.47	50.5	1.01	DMSO
Butane-1,4-diammonium Iodide (Bu14DiA)	1.07	1.39	129.9	1.53	DMF
1,4-Benzene diammonium iodide (Ben14DiA)	1.14	1.51	132.5	1.12	DMSO
N,N-Dimethylethane- 1,2-diammonium iodide (NNDiMEtA)	1.07	1.57	146.7	2.07	DMF
N,N-Diethylpropane-1,3-diammonium iodide (NNDiEtPA)	0.97	0.77	79.4	1.71	GBL
2-Pyrrolidin-1-ium-1-ylethylammonium iodide (PyrrolidinEtA)	0.99	1.19	120.2	1.99	DMF
N,N-dimethylpropane-1,3-diammonium iodide (NNDiMPA)	1.52	0.75	49.3	1.74	DMF
Di-isopropylammonium iodide (DiisoPropylA)	1.26	1.25	99.2	1.54	GBL
4-methoxy-phenethylammonium-iodide (4MeOPhenEtA)	1.13	1.36	120.4	1.9	GBL

Table S3. Experiment parameters in the chemical space explored

A-cations[a]	Organoammoniums: Acet, G, nBuA, EtA, MA, PhenEtA, F, Im, BenA, neoPenA, iPropylA, iBuA, DiMA, nDodA, Pyrro, EtDiA, 4MeOPhenA, nPropylA, iPenA, Pro13DiA, 4FPhenEtA, 4FPhenA, 4TFMBenA, 4TFMPhenA, tOctylA, PhenA, tBuA, CHA, 4FBenA, Pyrrid, DiEtA, Morph, nOctylA, nHexA, Piper, CHMA, PiperaDiium, Bu14DiA, Ben14DiA, NNDiMEtA, NNDiEtPA, PyrrolidinEtA, NNDiMPA, DiisoPropylA, 4MeOPhenEtA
Concentration of organoammonium iodide (organoammonium source)	Lower bound: 0; Upper bound: solubility of the organoammonium iodide
B-cation	Pb^{2+}
Concentration of PbI_2 (Pb^{2+} source)	Lower bound: 0; Upper bound: solubility of the PbI_2 under required amount of organoammonium iodide
X-halide	I^-
Formic acid concentration	Lower bound: 0; Upper bound: 16 M
Solvents[b]	GBL, DMF, DMSO
Crystallization temperature[c]	95 °C
Crystallization time	2.5 h

[a] See Table S2 and Figure 1 for full chemical names and structures.
[b] Solvents were selected based on the solubility of the organoammonium and lead iodide/organoammonium mixture. In general, we began with GBL. If the concentration of lead iodide in lead iodide/organoammonium mixture in GBL was <0.2 M, we proceeded to test other solvents. Reactions for each A-cation (organoammonium) were performed using only one solvent.
[c] All analysis and machine learning described in the main text were performed on reactions run at 95 °C. The publicly available dataset does include reactions performed at other temperatures (67, 70, and 80 °C) for single organoammonium (i.e., EtA).

Explanation of the variables in the chemical space

Variables that have been varied:
- **A-cation type**: we used 45 commercially available organoammonium as A-cation. (see Figure 1 for organoammonium structures)
- **Solvent type**: solvent type is not an independent variable but determined by the organoammonium type based on the solubility requirement. Each organoammonium species was used with only one solvent. We tested GBL, DMSO, and DMF as these are the most common solvents for perovskites single crystal synthesis. In general, we began with GBL, as the weak solvent interactions with the lead iodide species in solution facilitate ITC behavior at lower temperatures that are within the range of what we can achieve on our system. If the concentration of lead iodide in lead iodide/organoammonium mixture in GBL was <0.2 M, then we proceeded to investigate other solvents.

- **Organoammonium iodide concentration**: organoammonium iodide concentration was varied from 0 to the solubility limit of organoammonium iodide in the selected solvent.
- **PbI$_2$ concentration**: pure PbI$_2$ is not very soluble by itself in the solvents used here so the organoammonium iodide was mixed with PbI$_2$ to ensure optimal solubility of PbI$_2$. Therefore, PbI$_2$ concentration was varied from 0 to the solubility limit of PbI$_2$ under the required amount of organoammonium iodide.
- **Formic acid concentration**: the formic acid concentration was set to the range of 0 to 16 M.

Variables that have been kept constant:
- **B-cation type**: we used Pb^{2+} as B-cation.
- **X-halide type**: we used I- as X-halide.
- **Crystallization time**: the crystallization time is set to be 2.5 h.
- **Reaction temperature**: majority of our reactions are performed at 95 °C (for all 45 chemical systems), although we did explore other temperatures (67 °C, 70 °C, 80 °C) for single chemical system (i.e., EtA) to demonstrate our experimental capability of varying the reaction temperature in ITC. Practically, we can vary temperature from 95 °C to room temperature. In this work, we performed all analysis and machine learning on reactions at 95 °C as the reaction temperature is kept at 95 °C for nearly all chemical systems. The complete data set, which includes non-95 °C reactions, is publicly available at https://github.com/darkreactions/rapid.[4]
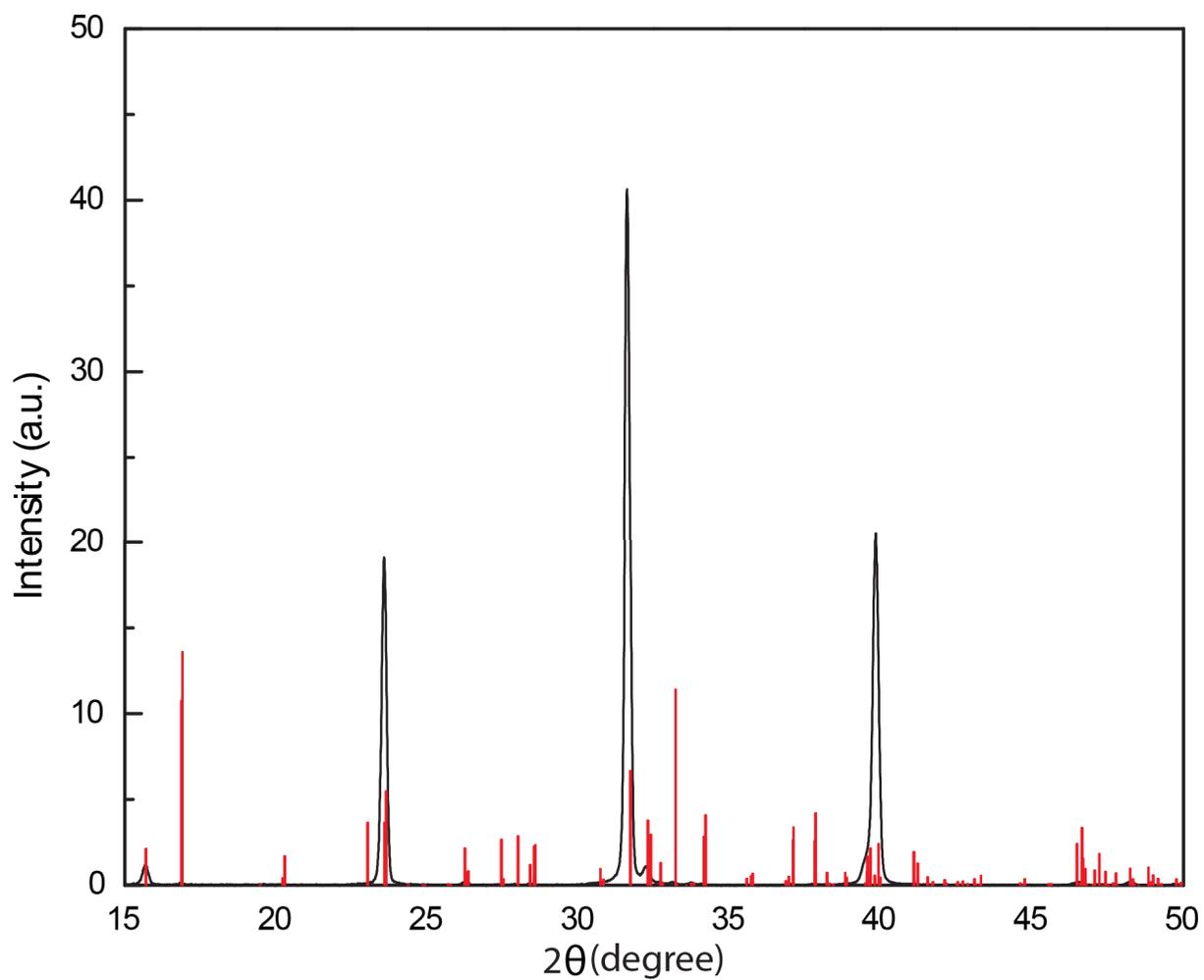
Figure S7. Experimental powder XRD pattern (black curve) of crystals grown from the iBuA / PbI$_2$ system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 1876240).
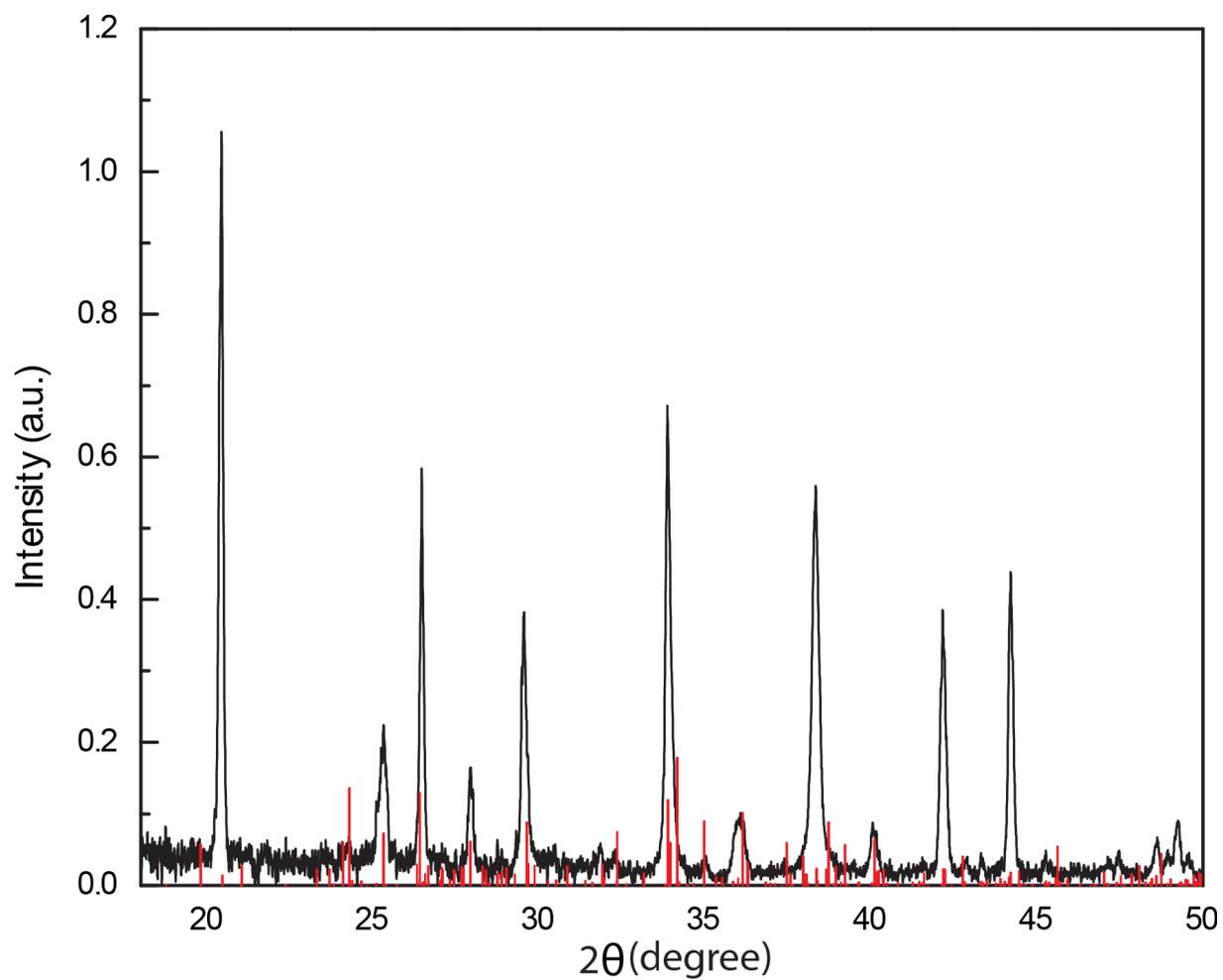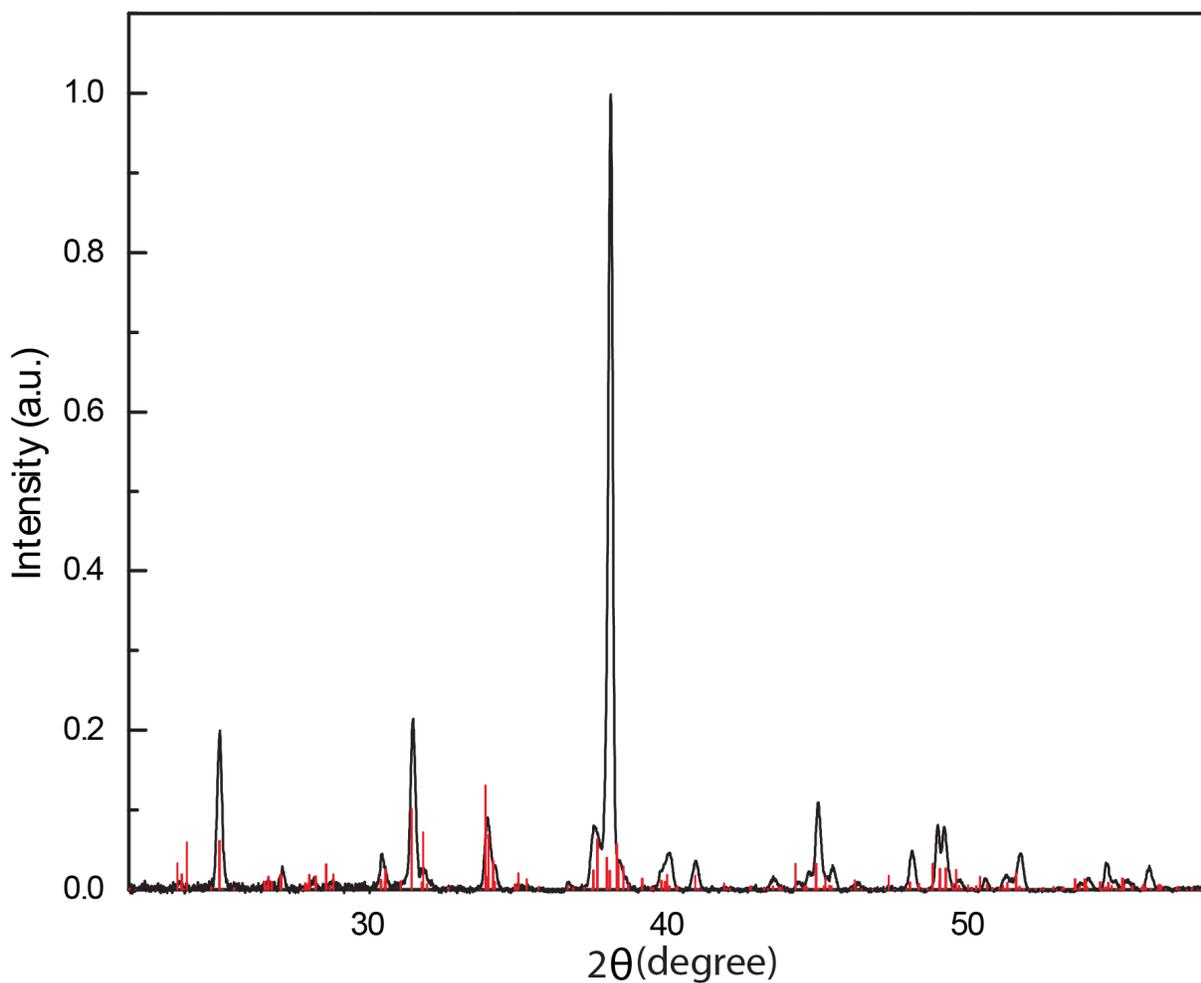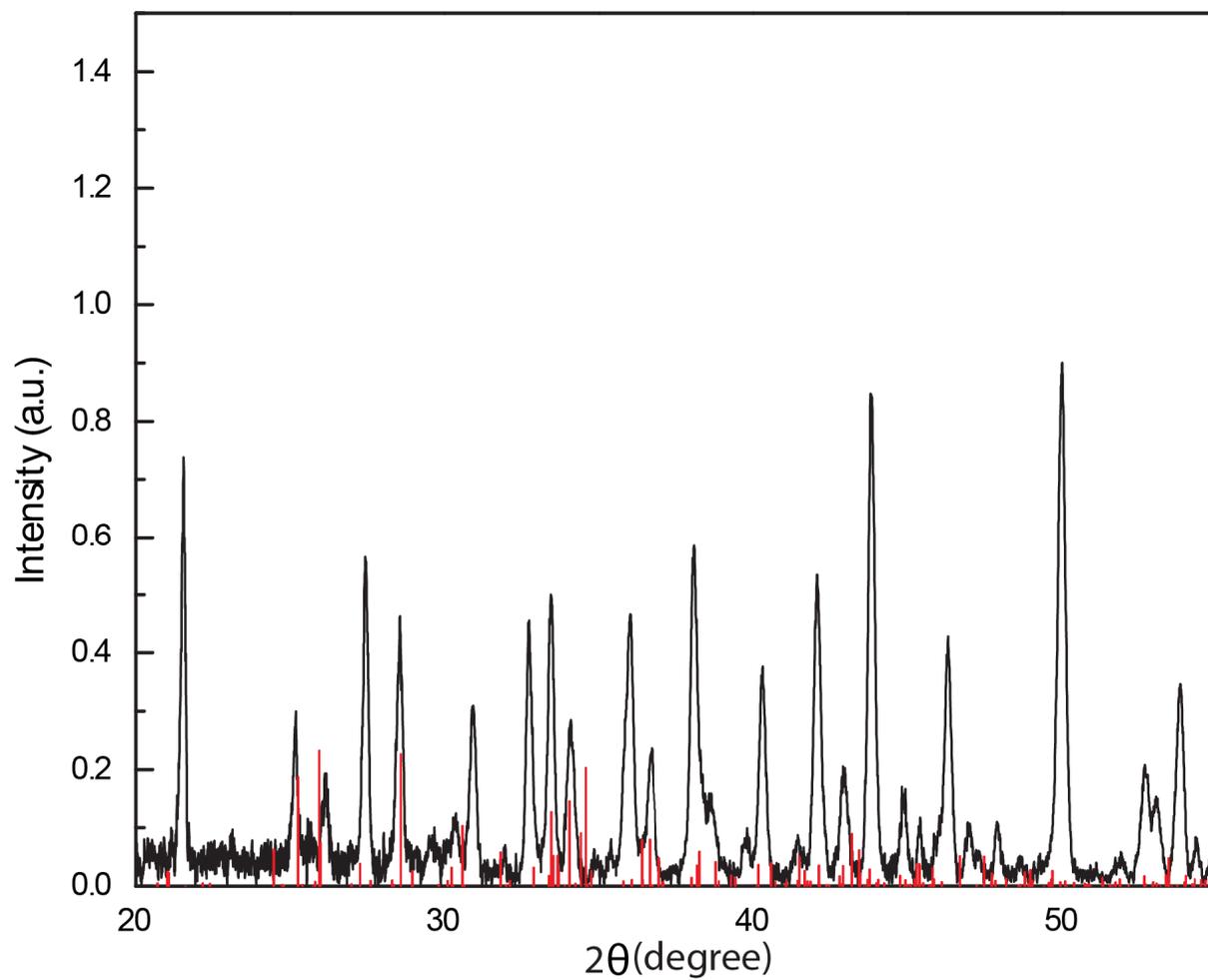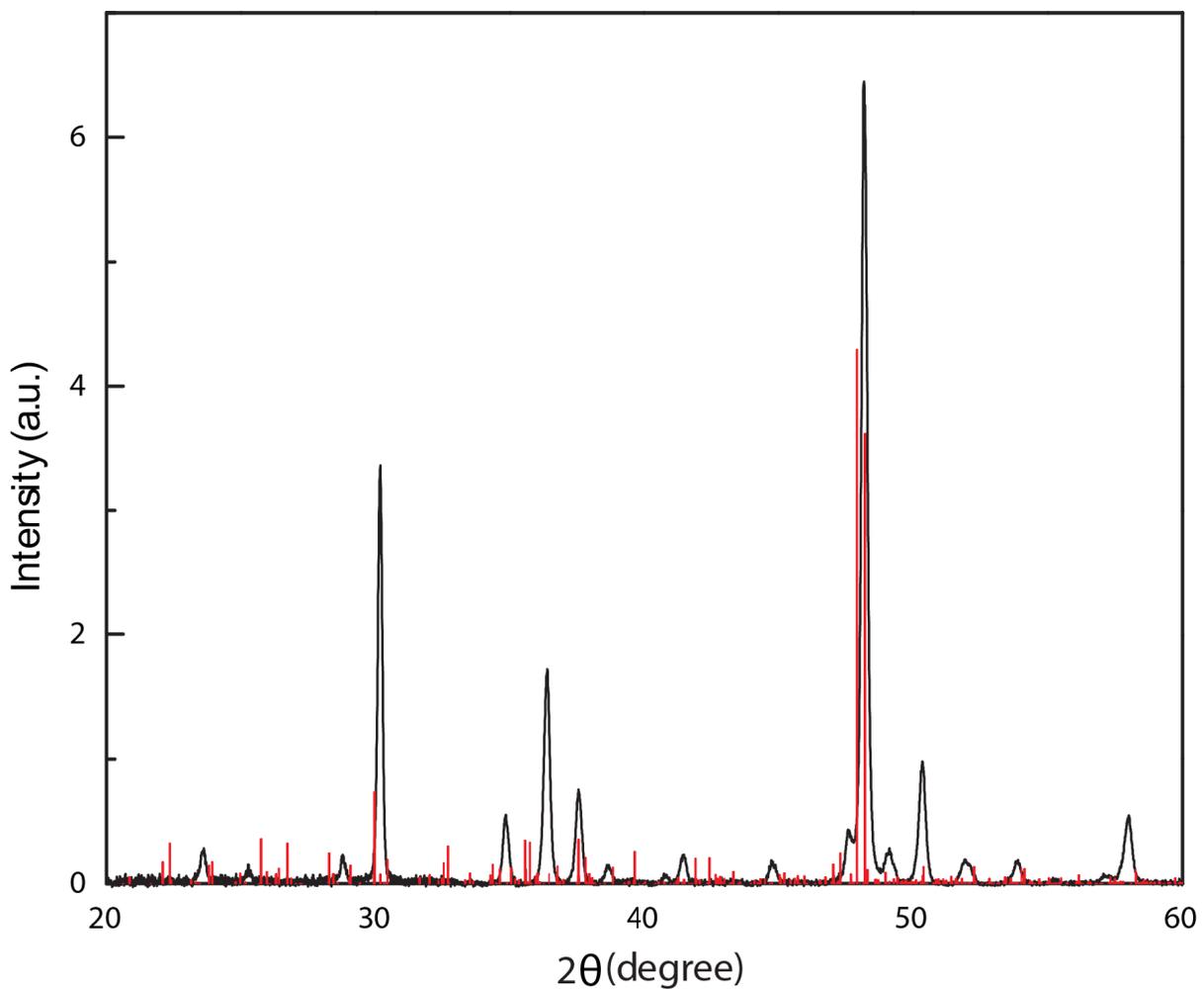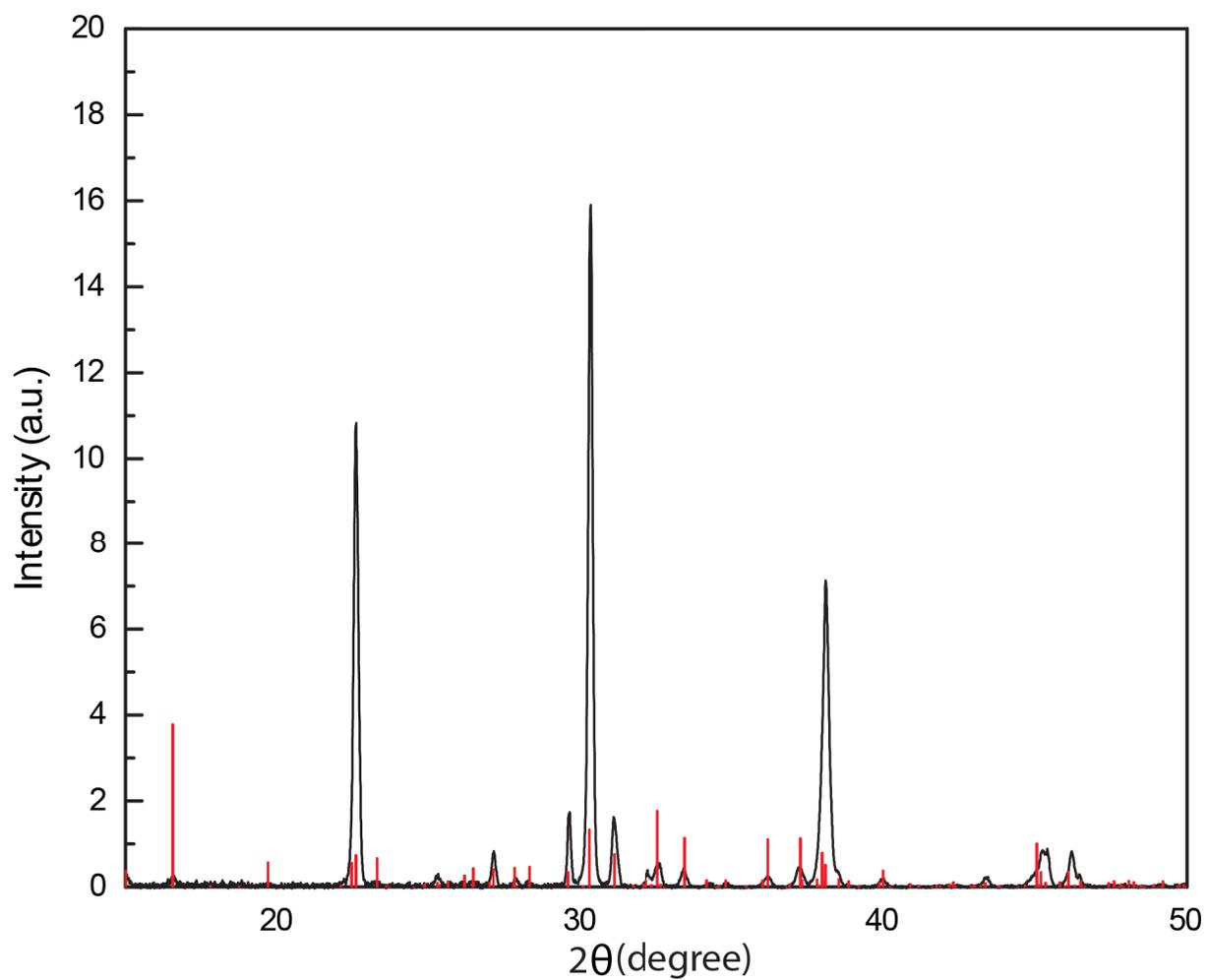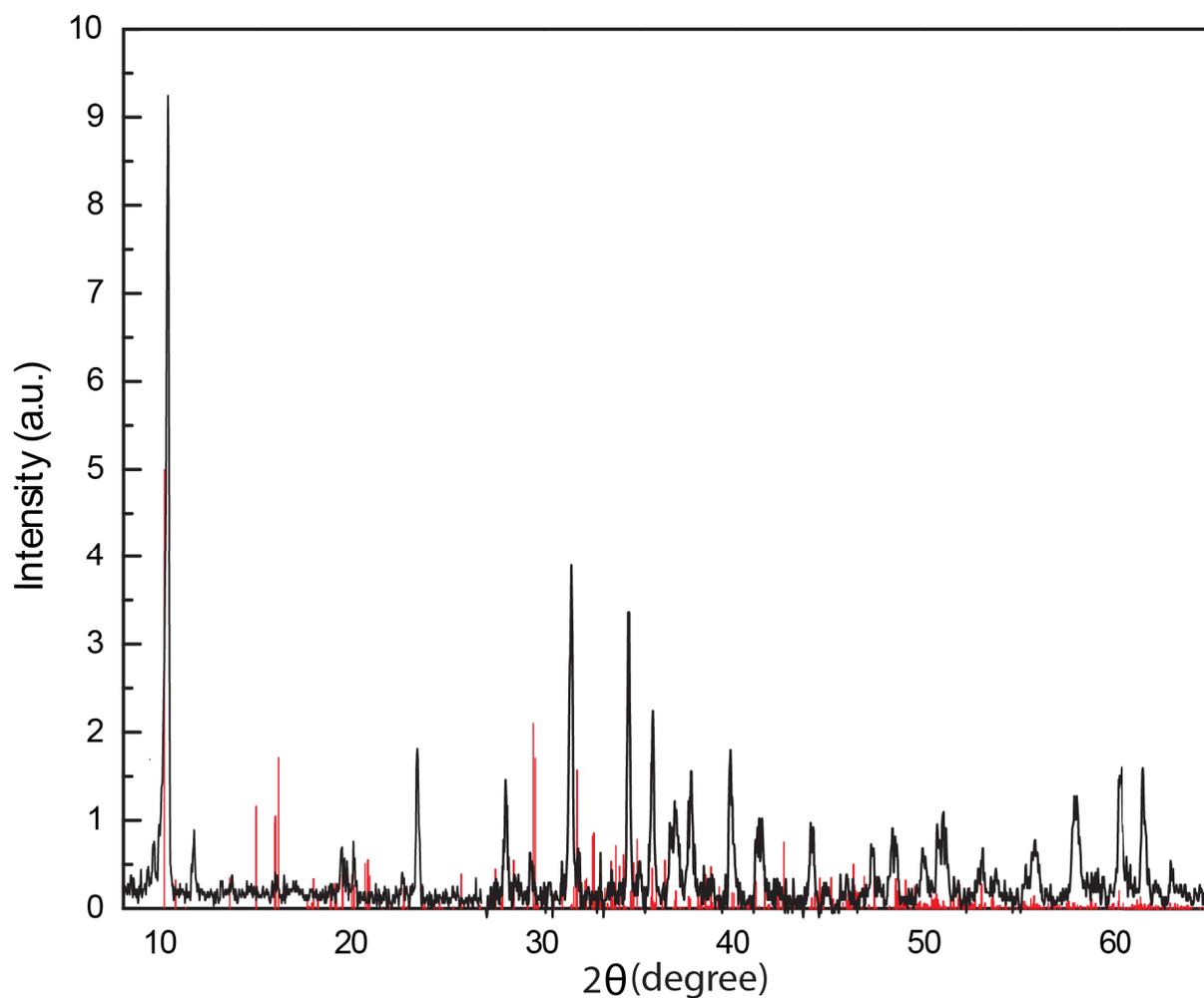
Figure S8. Experimental powder XRD pattern (black curve) of crystals grown from the nDodA / PbI$_2$ system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 692951).
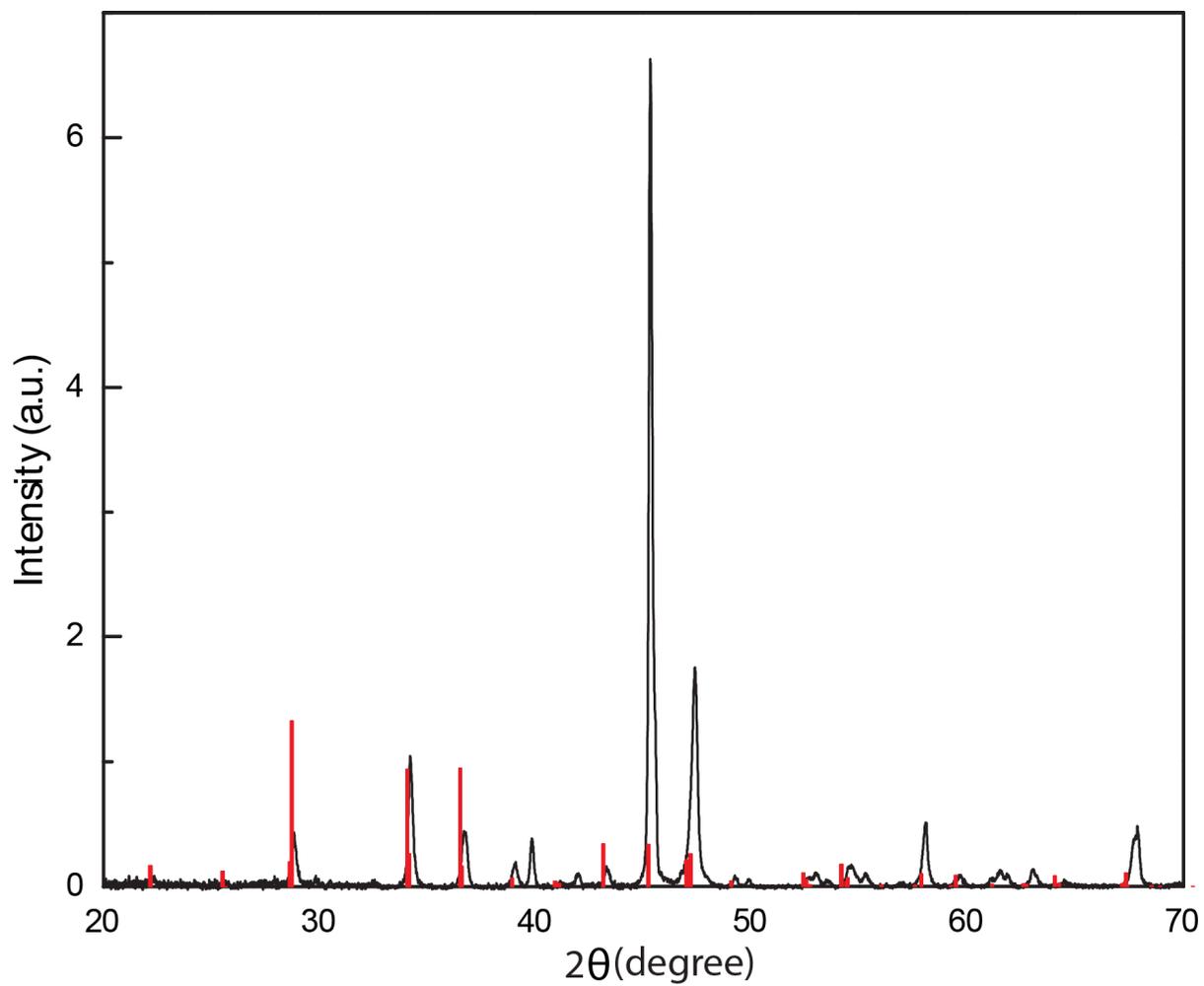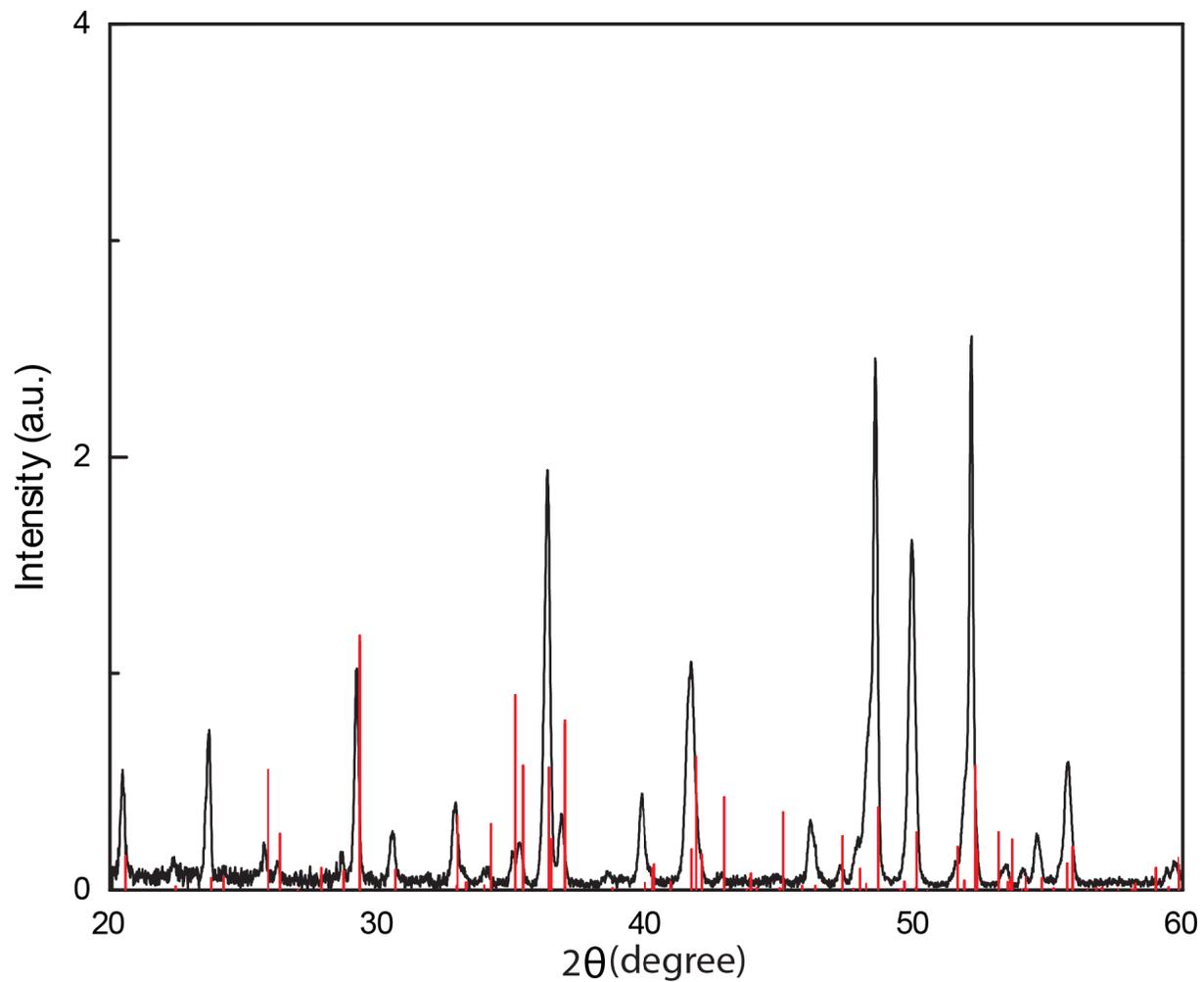
Figure S9. Experimental powder XRD pattern (black curve) of crystals grown from the Pyrro / PbI$_2$ system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 1886890).
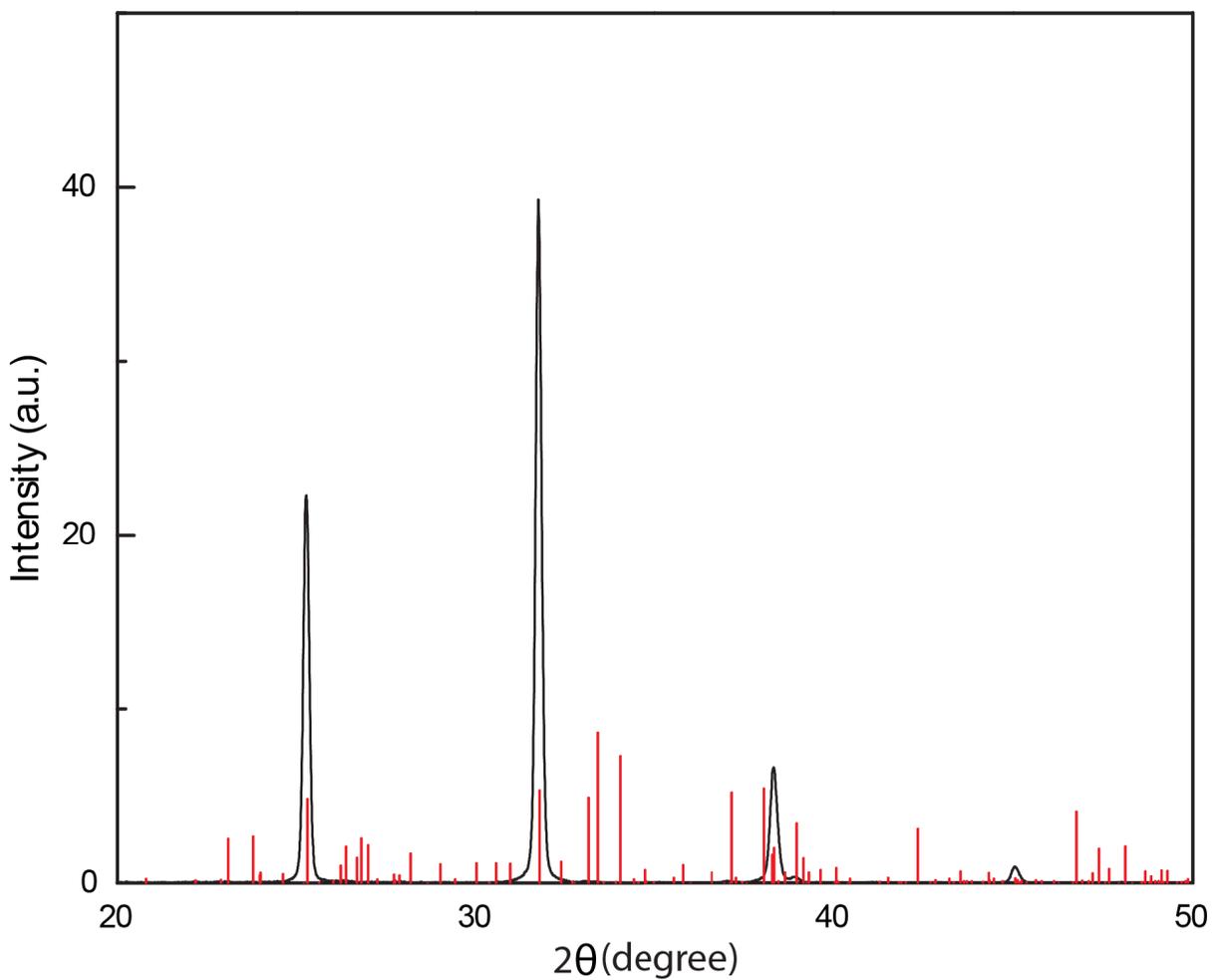
Figure S10. Experimental powder XRD pattern (black curve) of crystals grown from the 4FPhenEtA / PbI_2 system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 1893385).

Figure S11. Experimental powder XRD pattern (black curve) of crystals grown from the tBuA / PbI$_2$ system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 612444).
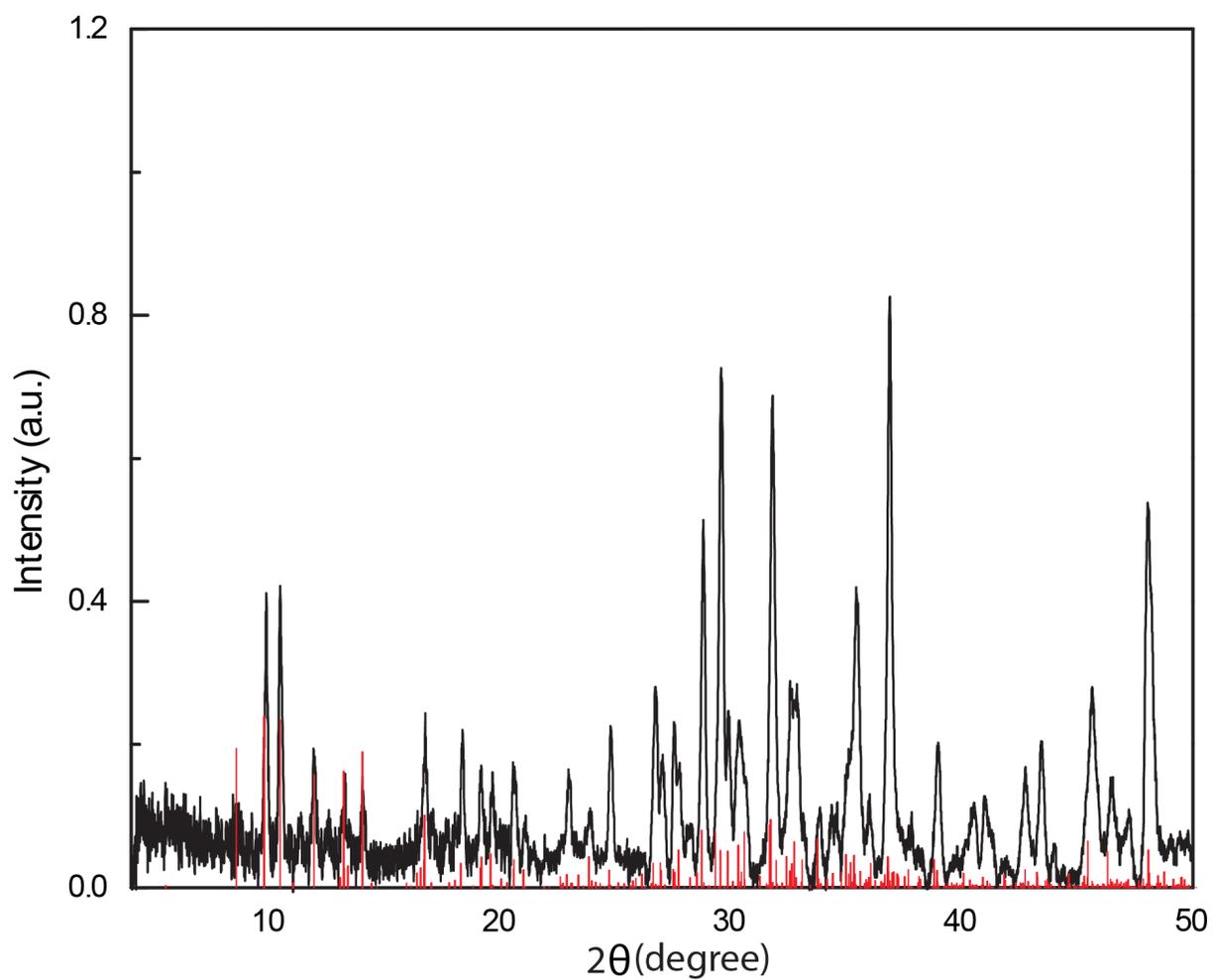
Figure S12. Experimental powder XRD pattern (black curve) of crystals grown from the 4FBenA / PbI$_2$ system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 1819854).
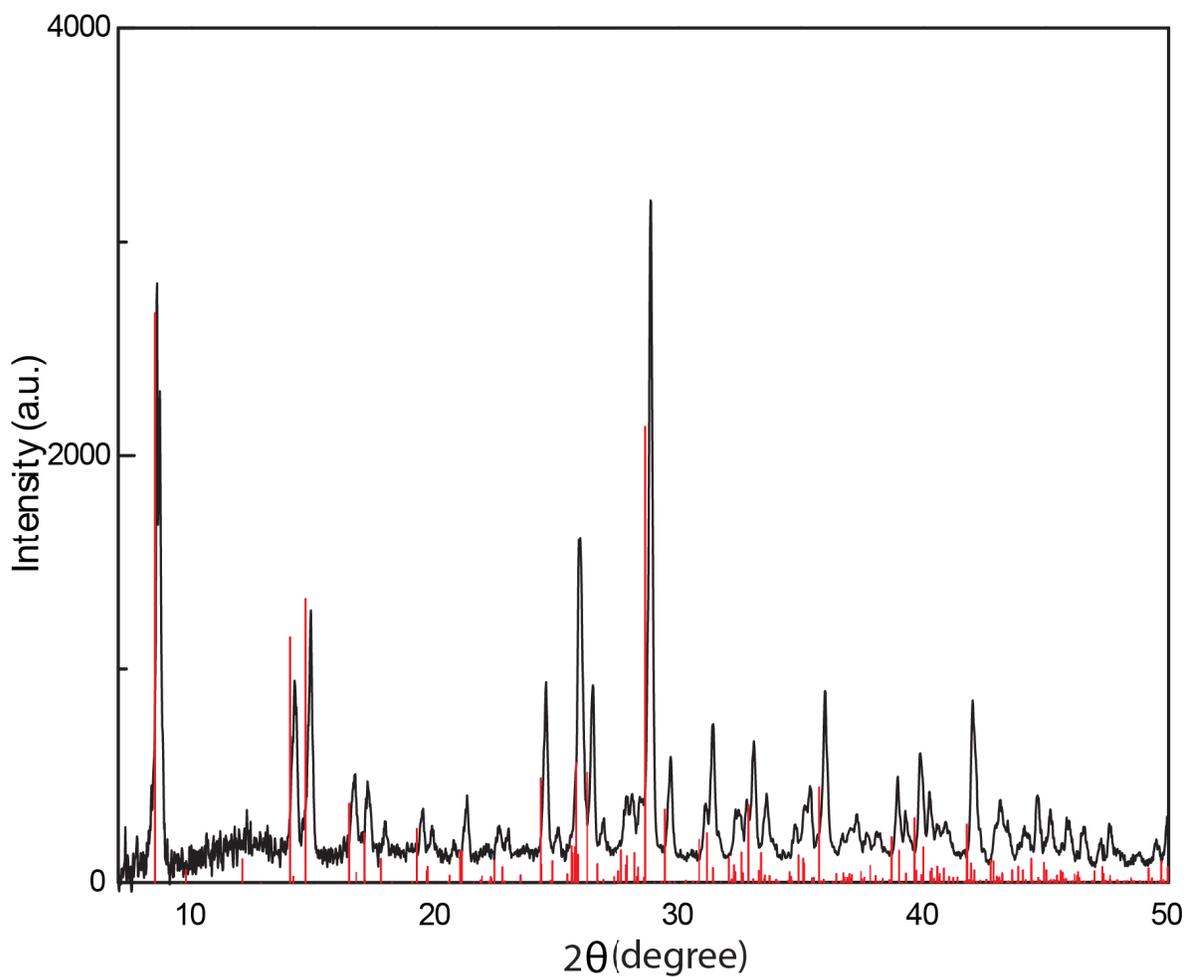
Figure S13. Experimental powder XRD pattern (black curve) of crystals grown from the CHA / PbI$_2$ system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 1409220).
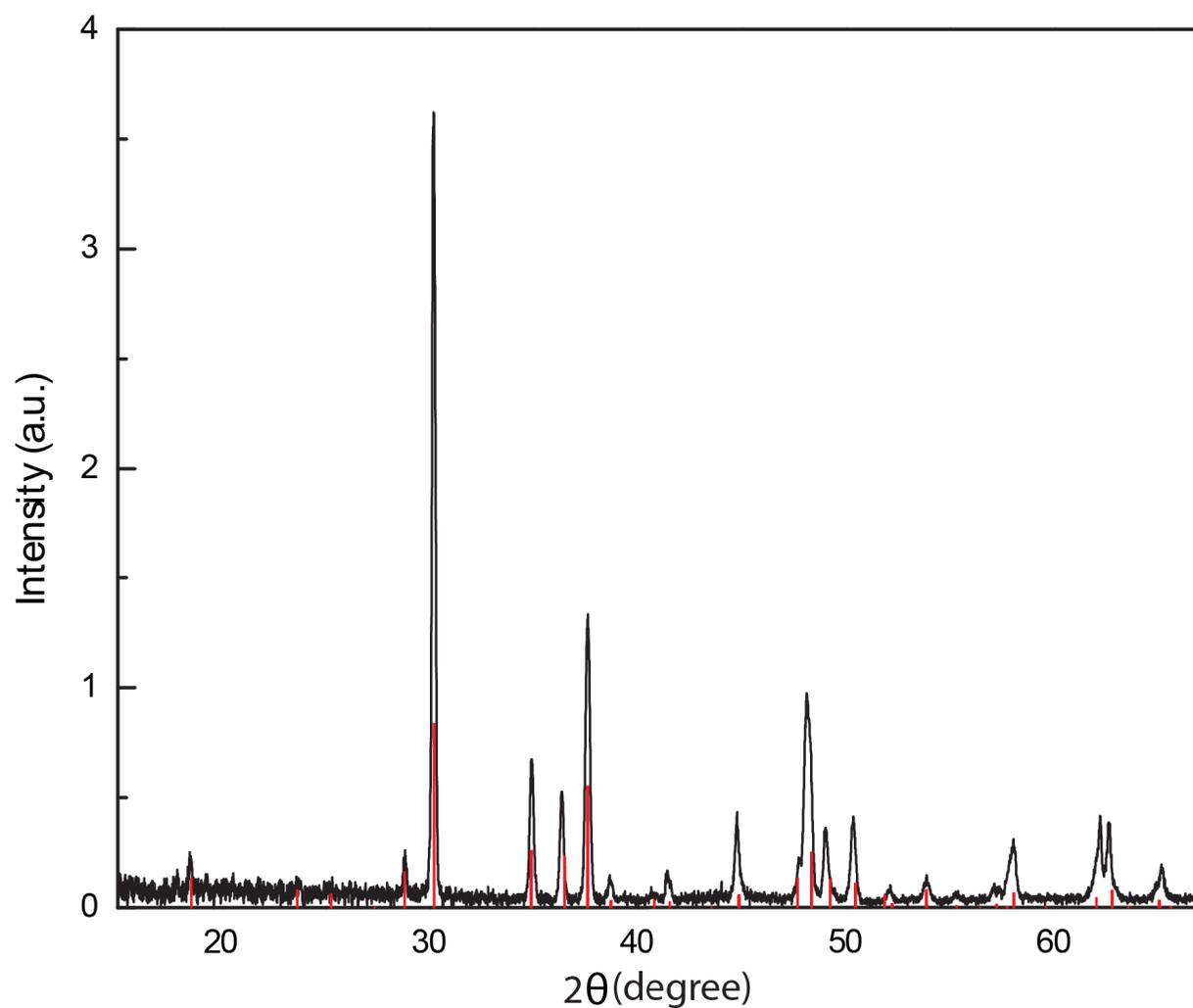
Figure S14. Experimental powder XRD pattern (black curve) of crystals grown from the Morph / PbI$_2$ system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 616872).

Figure S15. Experimental powder XRD pattern (black curve) of crystals grown from the Piper / PbI$_2$ system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 1135285).
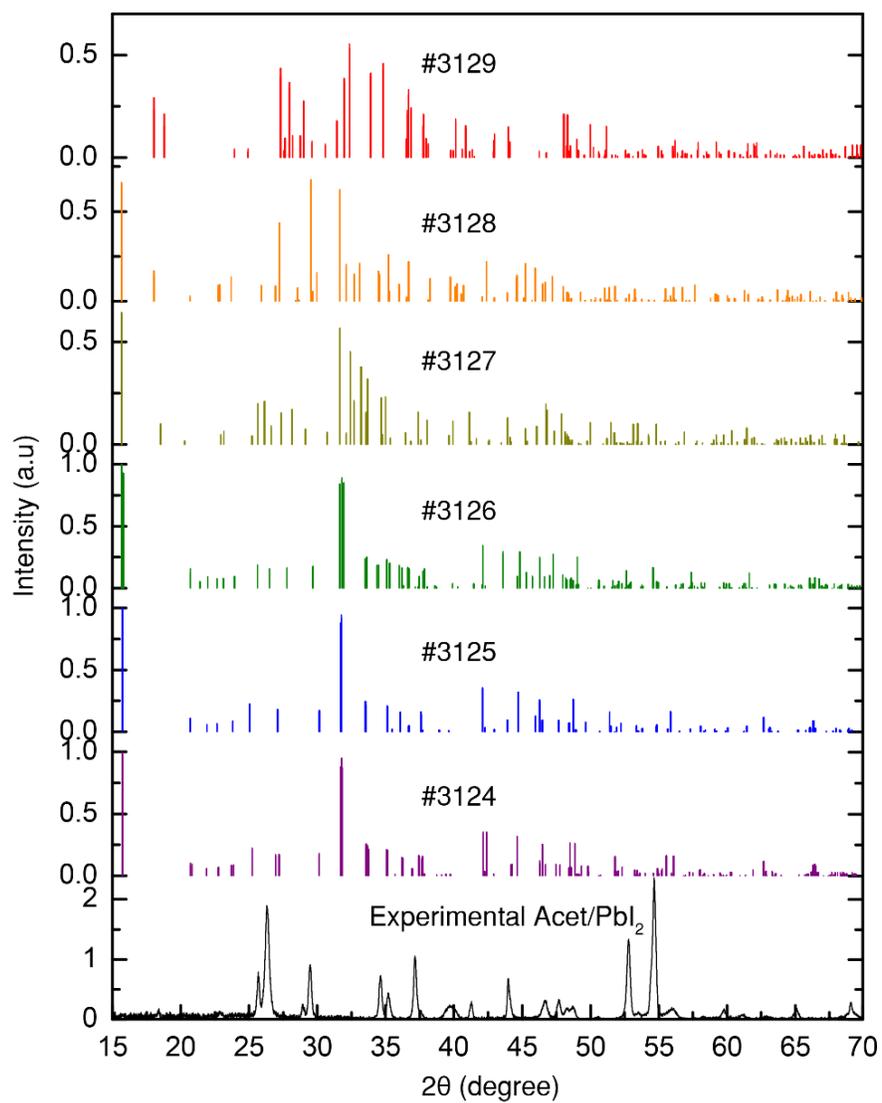
Figure S16. Experimental powder XRD pattern (black curve) of crystals grown from the nHex / PbI$_2$ system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 665695).

Figure S17. Experimental powder XRD pattern (black curve) of crystals grown from the NNDiEtPA / PbI$_2$ system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 995700).

Figure S18. Experimental powder XRD pattern (black curve) of crystals grown from the NNDiMPA / PbI$_2$ system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 995699).

Figure S19. Experimental powder XRD pattern (black curve) of crystals grown from the DiMA / PbI$_2$ system and simulated powder XRD pattern (red sticks) of reference single crystal (CCDC: 1408123).

Figure S20. Experimental powder XRD pattern (black curve) from Acet / PbI$_2$ system and simulated powder pattern (colored sticks) from calculated single crystal structures of AcetPbI$_3$ from Khazana database (#3124-3129 are the database IDs).[5]

Single Crystal Structure Refinement Details

The structure for $CH_2C(NH_2)_2PbI_3$ (CCDC Deposition Number 1938171) was refined as a 4-component inversion twin by merohedry and inversion. The structure emulates space group $P6_3/mcm$. The higher symmetry is broken by a slight shift of one of the $(PbI_3)_n$ chains, creating both an offset of Pb and I atoms, and a slight rotation of the other $(PbI_3)_n$ chains. These distortions caused a rotational offset of the iodine atoms. The structure was refined as being twinned by merohedry (180 degree around the reciprocal c-axis) and inversion twinning, yielding twin rates of 0.292(15), 0.238(15) (rotation around (0 0 1)), 0.258(16) (inversion) and 0.212(15) (rotation around (0 0 1) and inversion).

Due to the pseudosymmetry, parameters of the acetamidinium cations are highly correlated. To overcome correlation effects, all C-N bond distances were restrained to be similar in length (SADI restraint of Shelxl). The ADPs of carbon atoms C1 and C3, and of C2 and C4 were each constrained to be identical. An overall rigid bond restraint was applied for C and N atoms (RIGU 0.001 of Shelxl).

Other relevant crystallographic data are listed in Table 1 and Table S4-6.

Table S4. Selected bond lengths (Å) in $CH_2C(NH_2)_2PbI_3$.

Bond	Length (Å)
Pb1 – I1	3.236(2)
Pb1 – I2	3.233(2)
Pb2 – I1	3.193(2)
Pb2 – I2	3.235(2)
Pb3 – I3	3.218(2)
Pb3 – I4	3.193(3)
Pb4 – I3	3.268(2)
Pb4 – I4	3.189(3)

Table S5. Selected bond angles (º) in $CH_2C(NH_2)_2PbI_3$.

Bonding atoms	Angle (º)
I1 – Pb1 – I1	84.16(6)
I2 – Pb1 – I1	177.75(5)
I2 – Pb1 – I1	94.39(5)
I2 – Pb1 – I1	97.41(4)
I2 – Pb1 – I1	97.42(4)
I2 – Pb1 – I2	84.08(6)
I2 – Pb1 – I2	84.09(6)
I1 – Pb2 – I1	85.55(6)
I1 – Pb2 – I2	177.51(6)
I1 – Pb2 – I2	177.52(6)
I1 – Pb2 – I2	93.70(4)
I1 – Pb2 – I2	96.76(5)
I2 – Pb2 – I2	84.03(7)
I2 – Pb2 – I2	84.04(7)
I3 – Pb3 – I3	84.46(7)
I4 – Pb3 – I3	179.73(8)
I4 – Pb3 – I3	179.73(9)
I4 – Pb3 – I3	95.33(4)
I4 – Pb3 – I3	95.34(4)
I4 – Pb3 – I4	84.87(8)
I3 – Pb4 – I3	82.88(6)
I3 – Pb4 – I3	82.88(6)
I4 – Pb4 – I3	178.57(8)
I4 – Pb4 – I3	96.05(4)
I4 – Pb4 – I4	85.00(8)

Table S6. Selected bond lengths (Å) in $(C_7H_{16}N)_2PbI_4$.

Bond	Length (Å)
Pb1 – I1	3.2129(5)
Pb1 – I2	3.1680(4)
Pb1 – I2	3.1729(4)
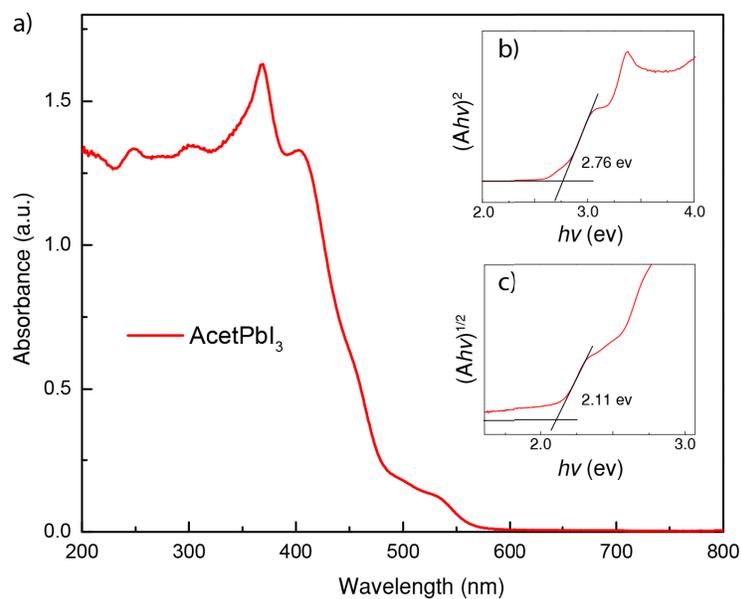
Figure S21. (a) UV−visible absorption spectrum of ground, large crystals of AcetPbI$_3$ measured in diffusive reflectance mode. Tauc plots of (a) showing direct band gap (b) and indirect band gap (c) determination.
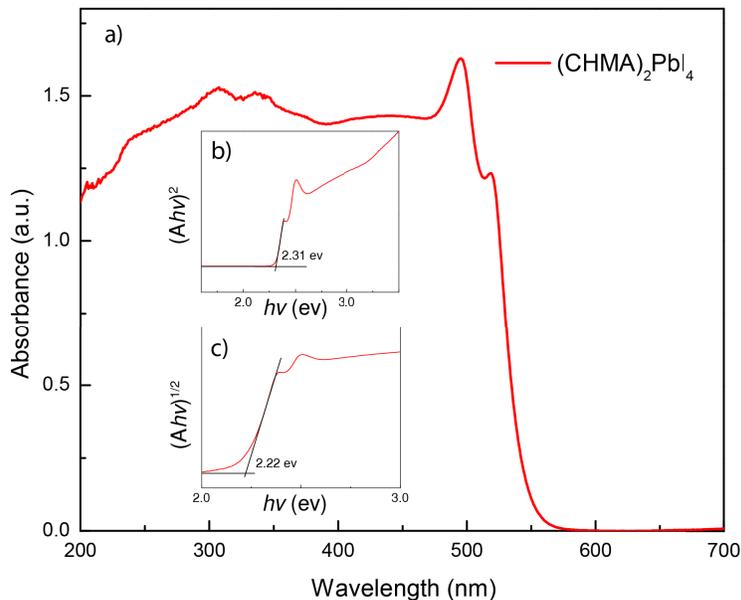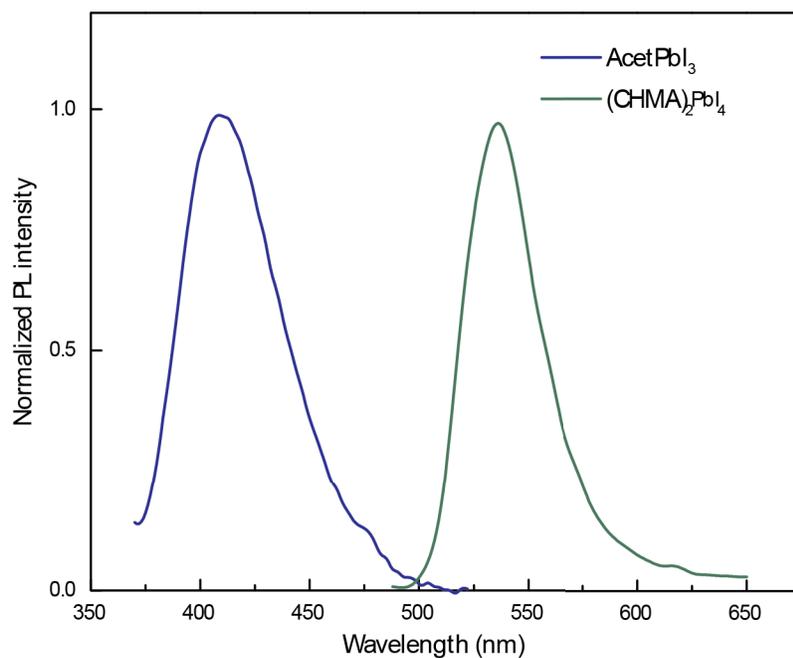

Figure S22. (a) UV−visible absorption spectrum of ground, large crystals of (CHMA)$_2$PbI$_4$ measured by diffusive reflectance method. Tauc plots of (a) showing direct band gap (b) and indirect band gap (c) determination.

Figure S23. Photoluminescence spectra of ground, large crystals of AcetPbI$_3$ (blue curve, λ_{ex} = 350 nm) and (CHMA)$_2$PbI$_4$ (green curve, λ_{ex} = 400 nm).
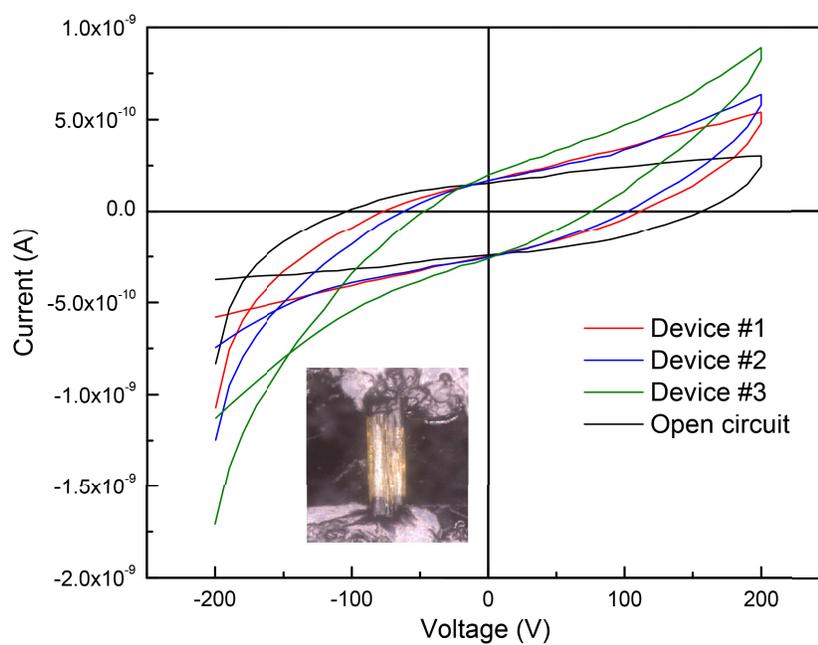
Figure S24. *I-V* plots for 3 different needle-shaped single crystals of AcetPbI$_3$ at room temperature. Inset is the photo of one device.
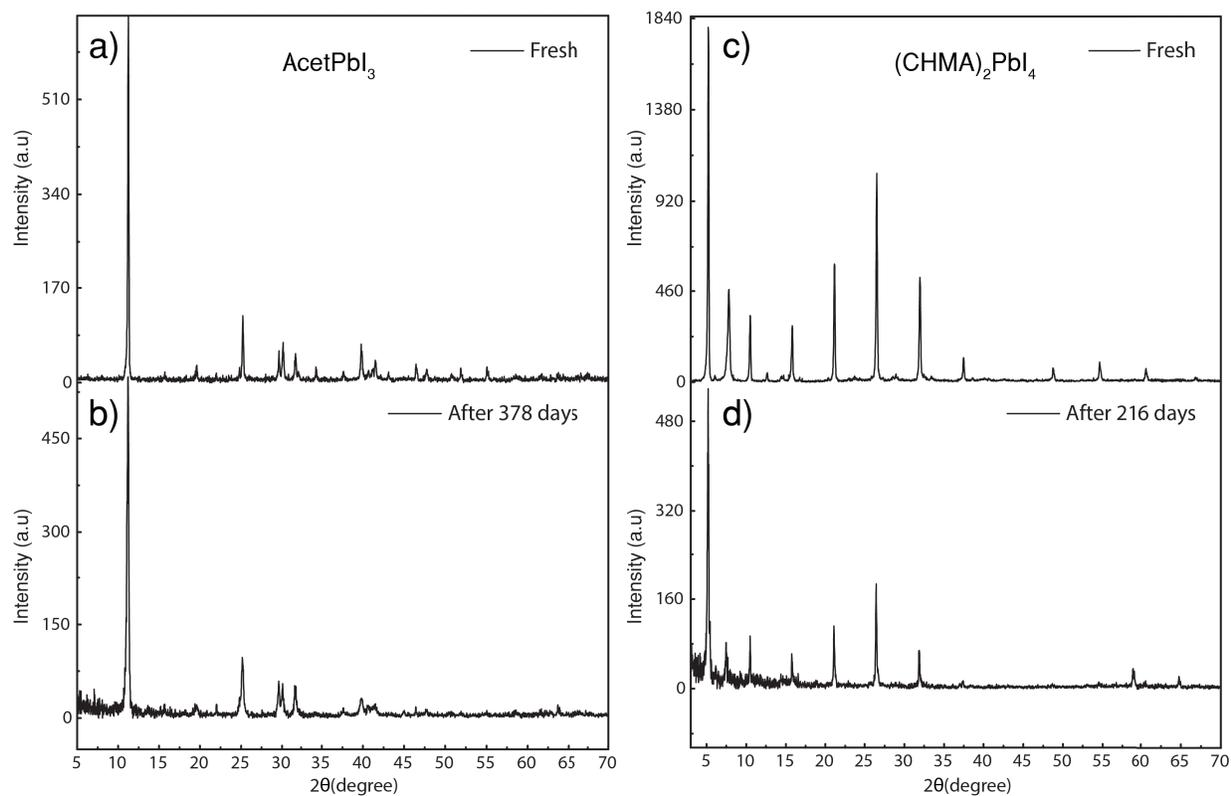
Figure S25. Powder XRD patterns of the AcetPbI$_3$ and (CHMA)$_2$PbI$_4$ before (a, c) and after (b, d) storage for >7 months at room temperature under air at 23-33% average humidity.
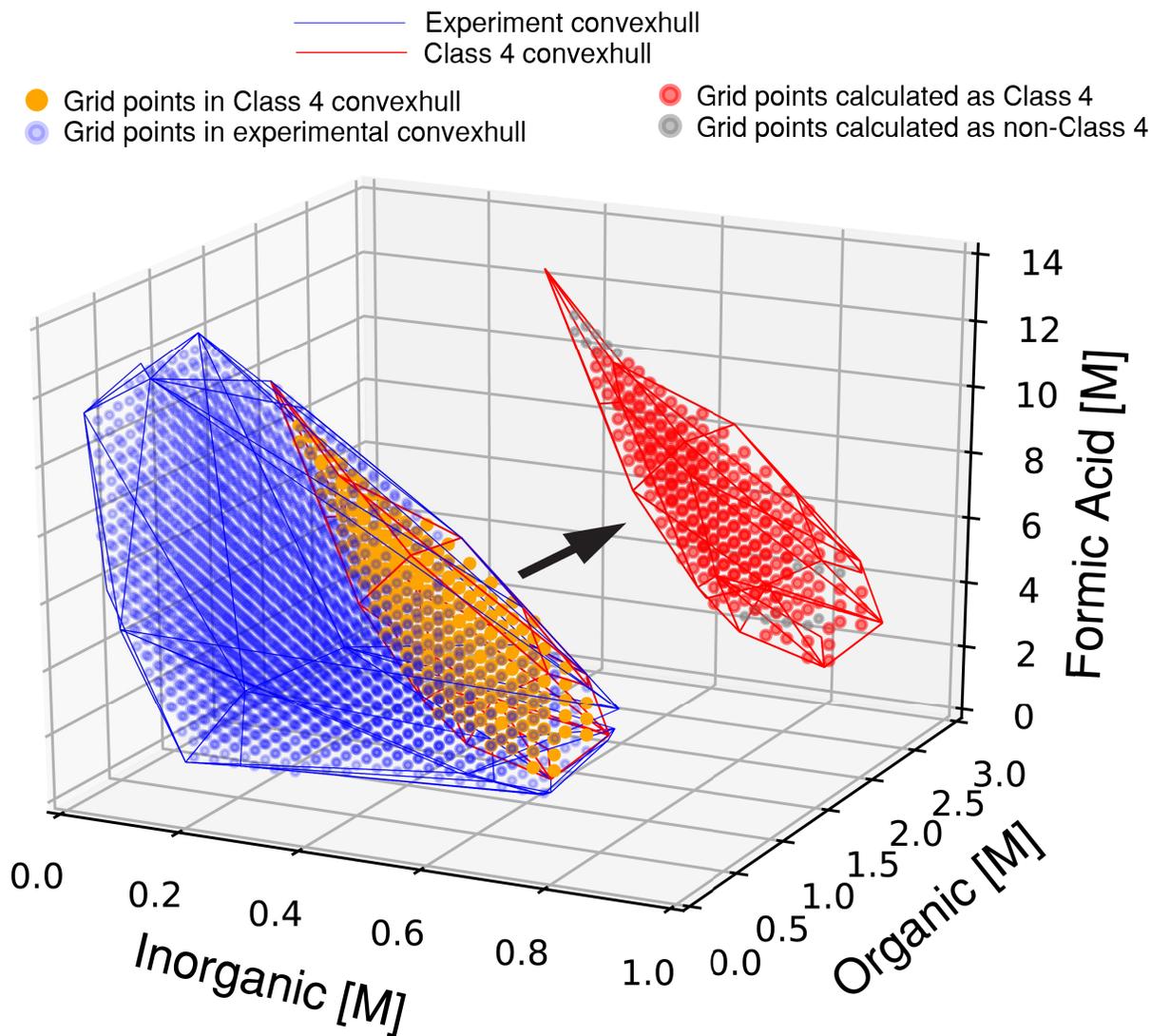
Figure S26. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the MA / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
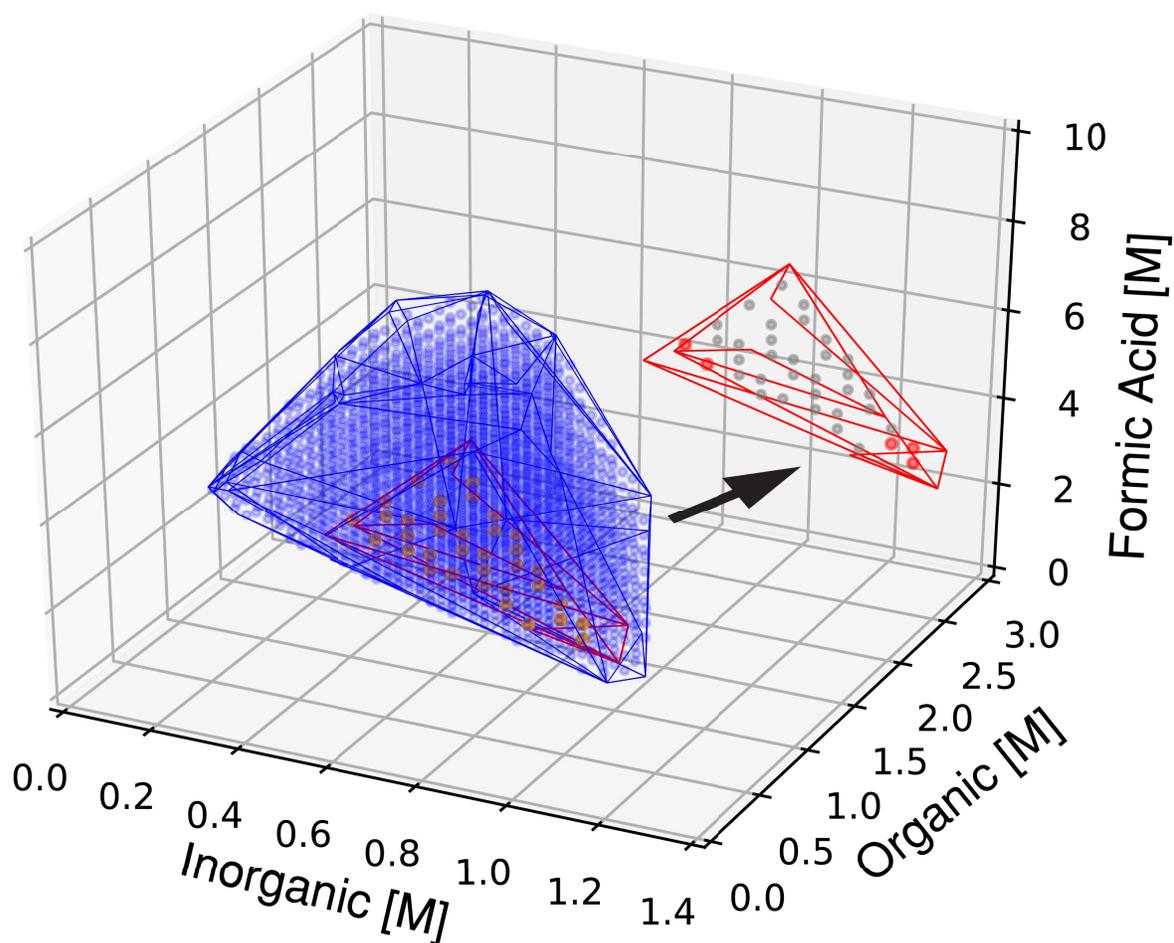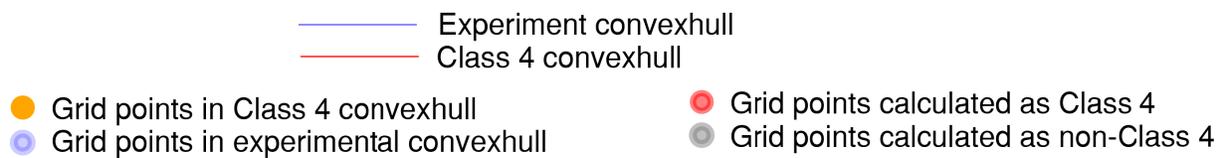
Figure S27. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the nBuA / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the *k*-NN algorithm.
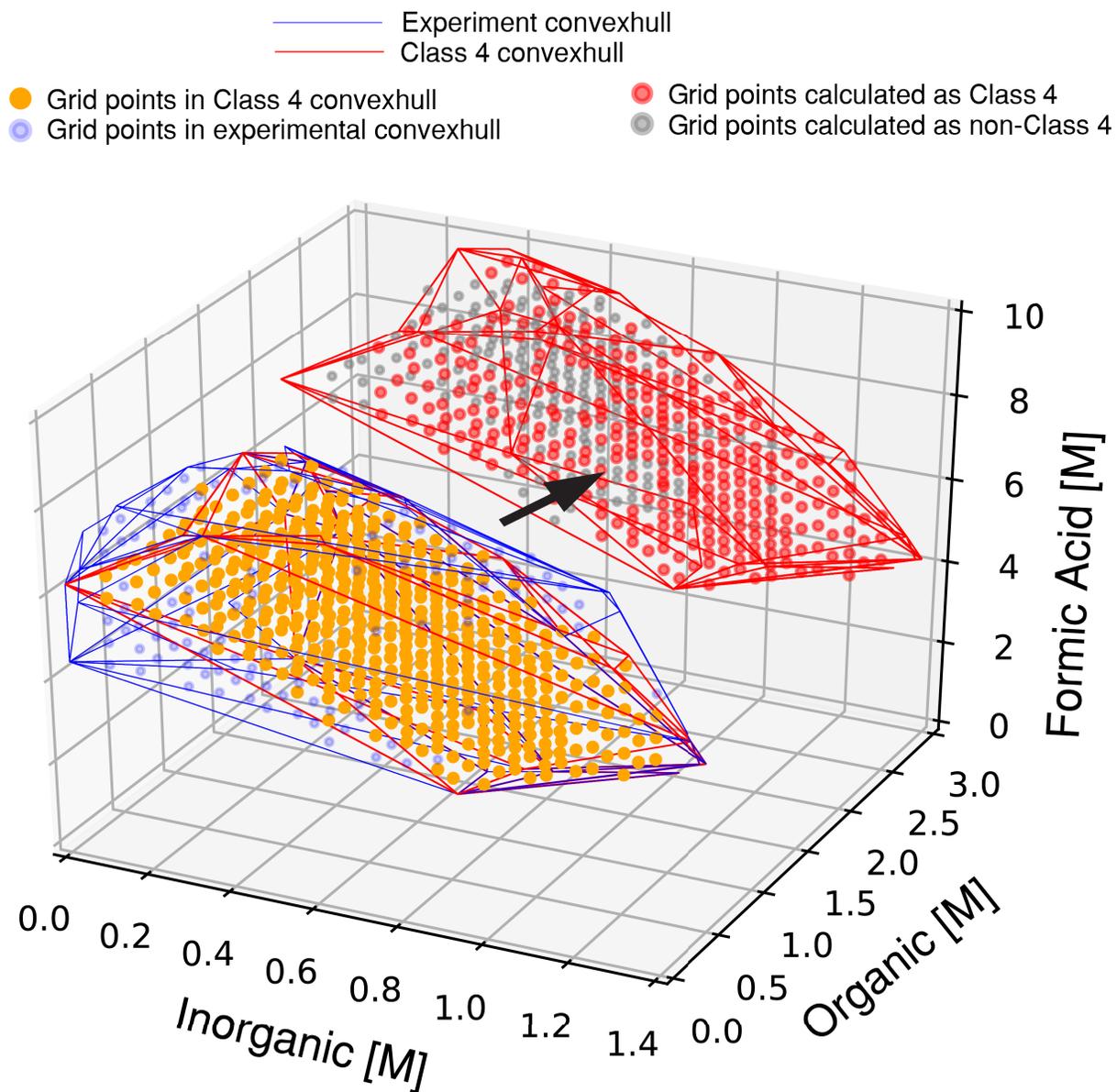
Figure S28. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the FA / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
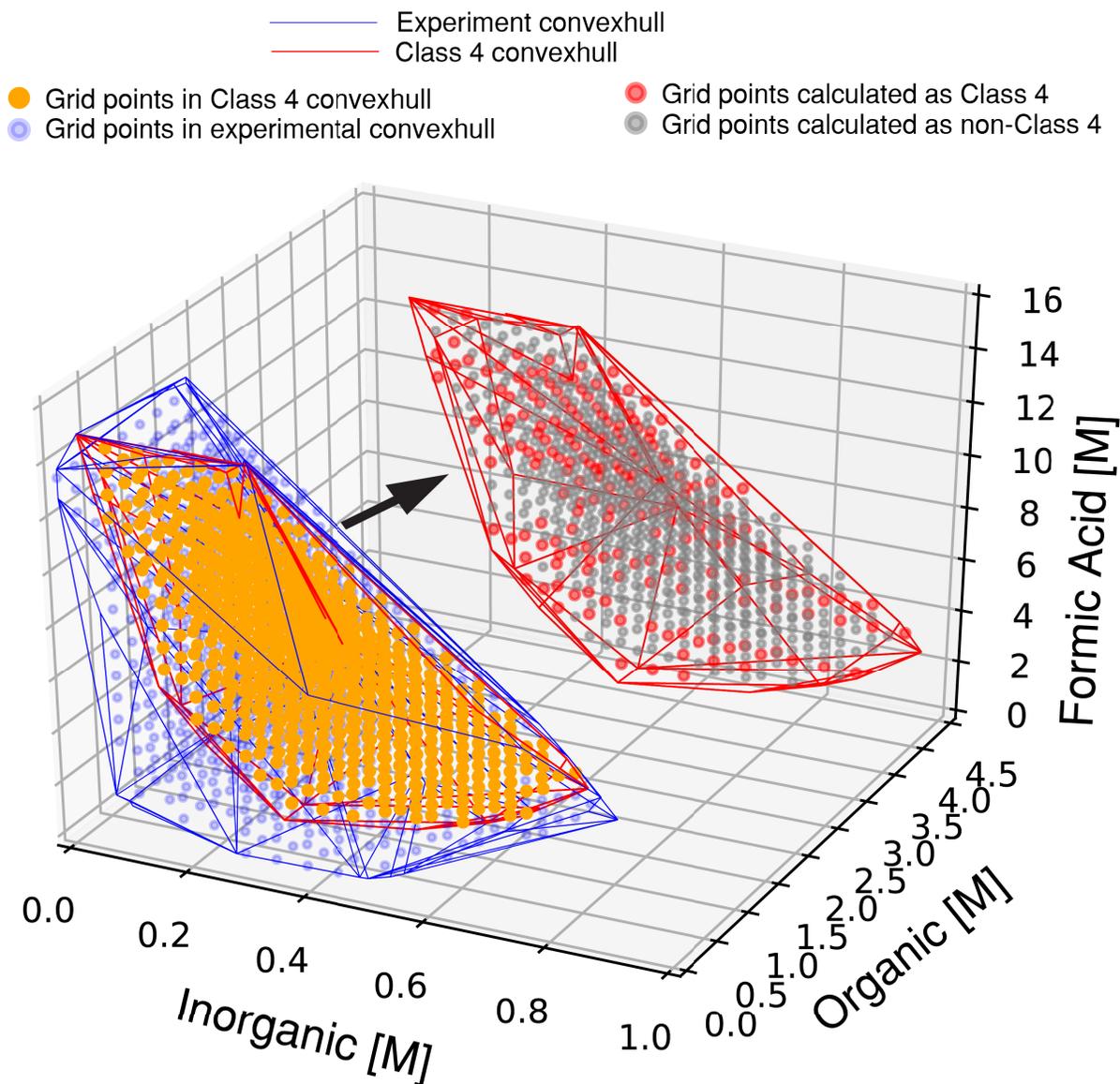
Figure S29. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the Acet / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the *k*-NN algorithm.
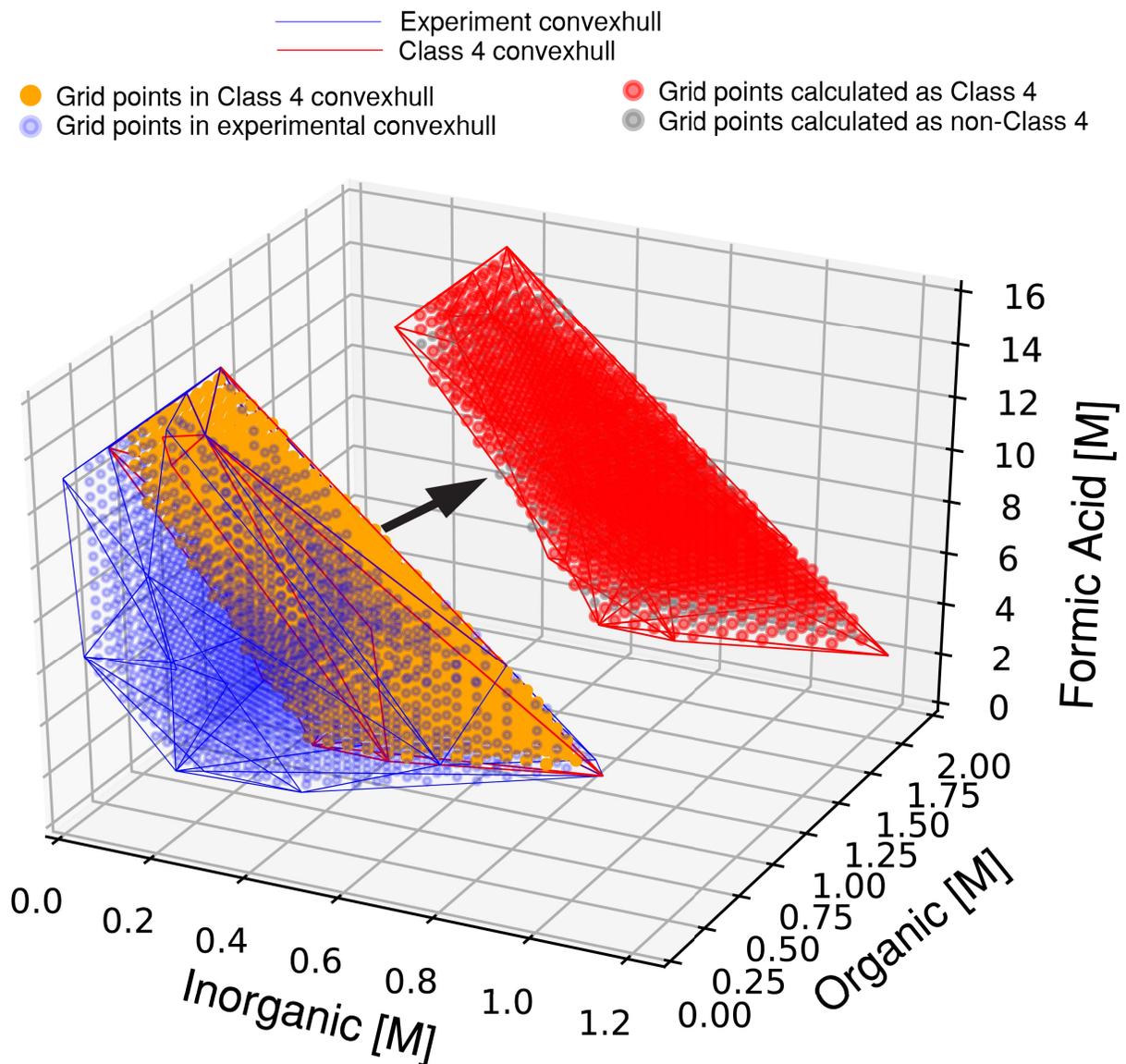
Figure S30. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the CHMA / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
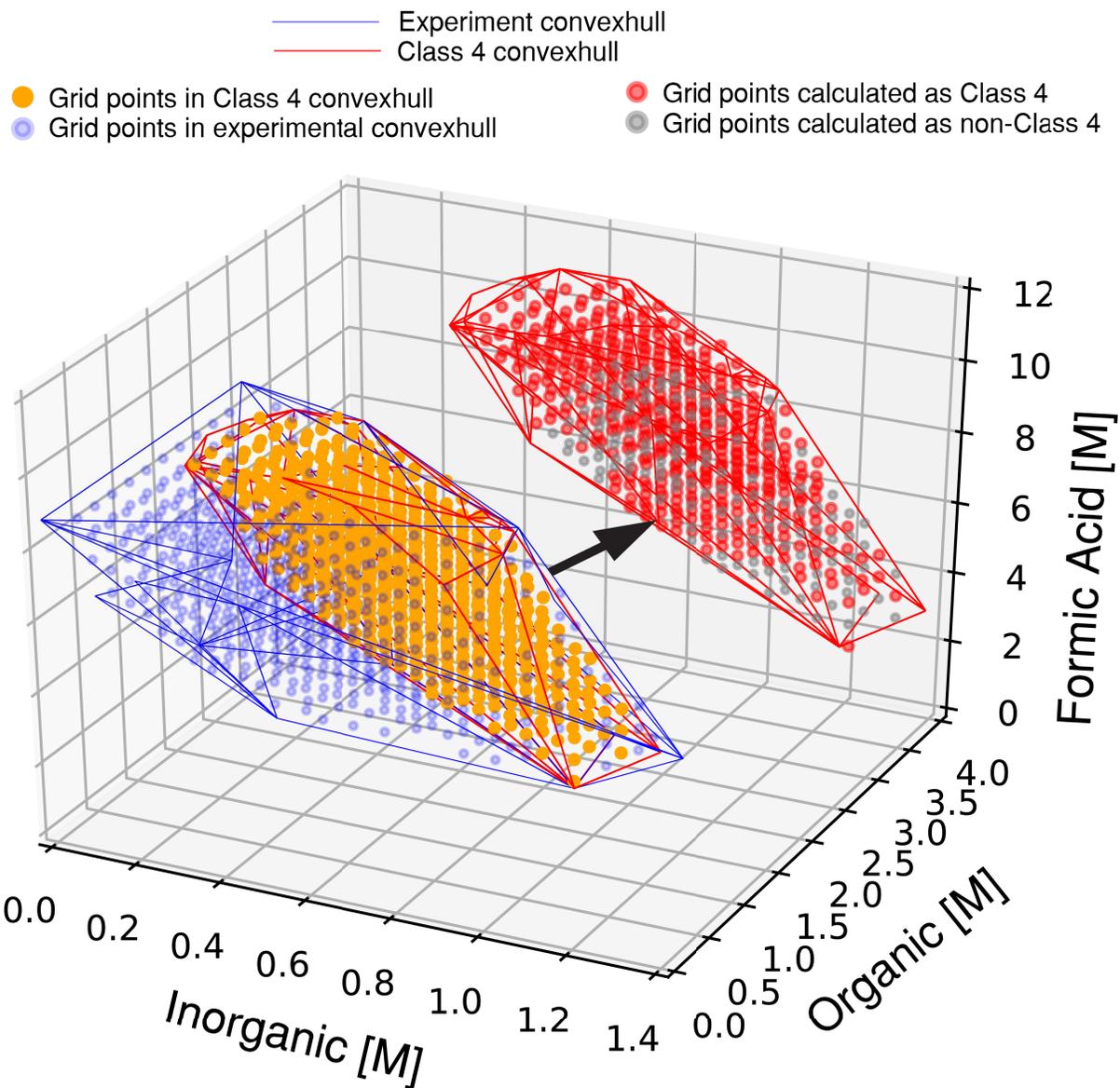
Figure S31. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the iBuA / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
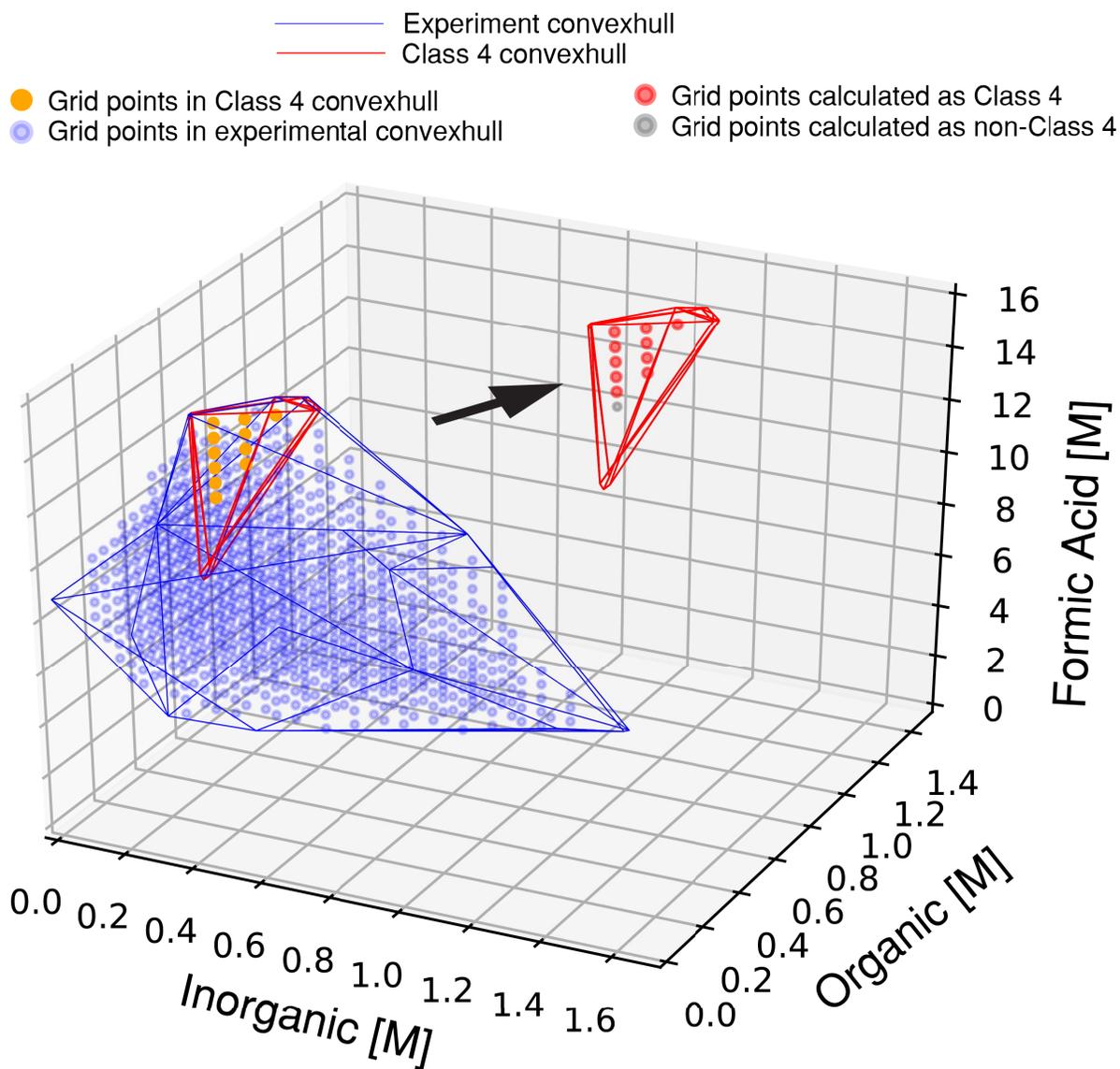
Figure S32. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the nDodA / PbI₂ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.

Figure S33. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the Pyrro / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
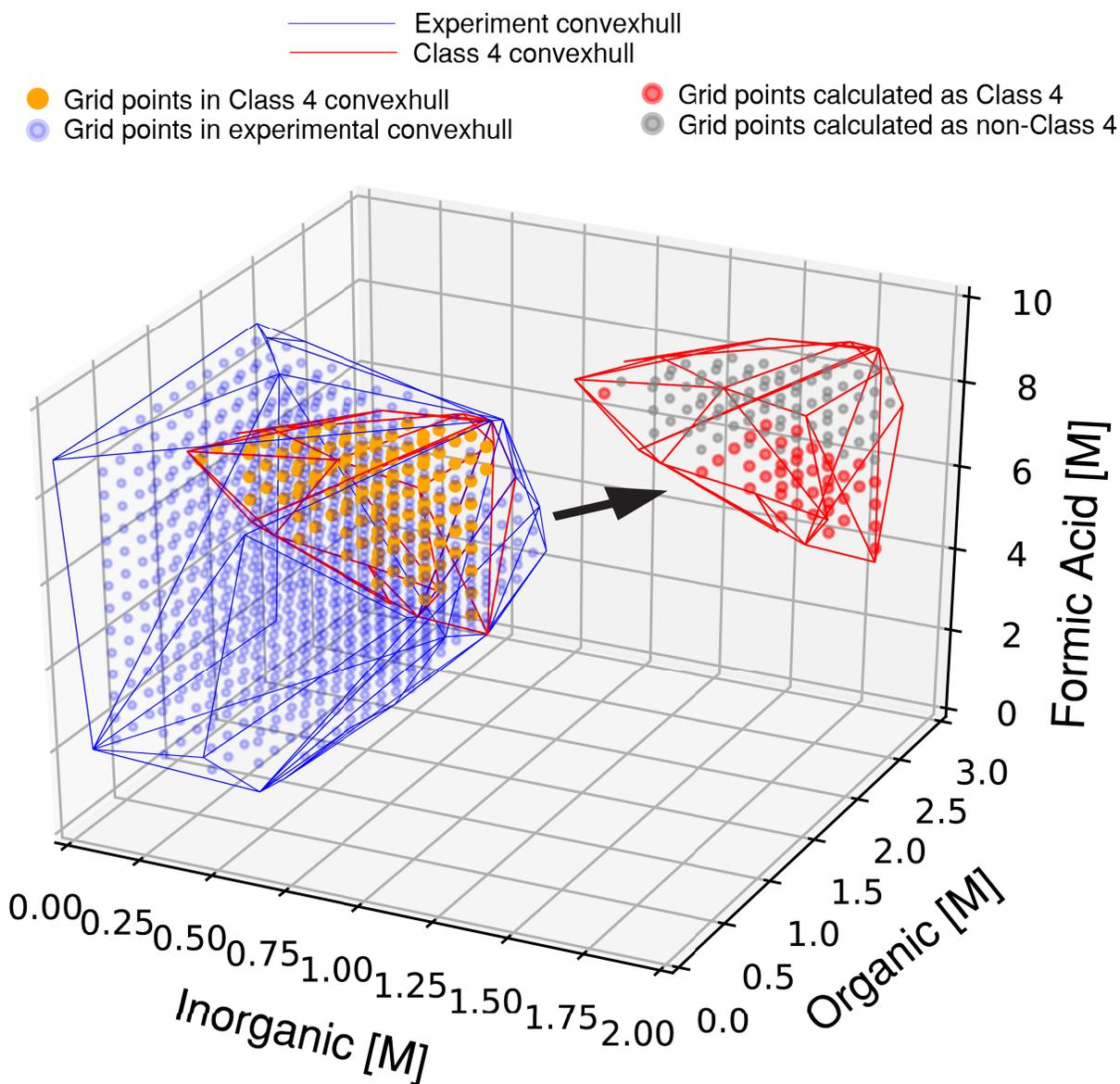
Figure S34. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the 4FPhenEtA / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
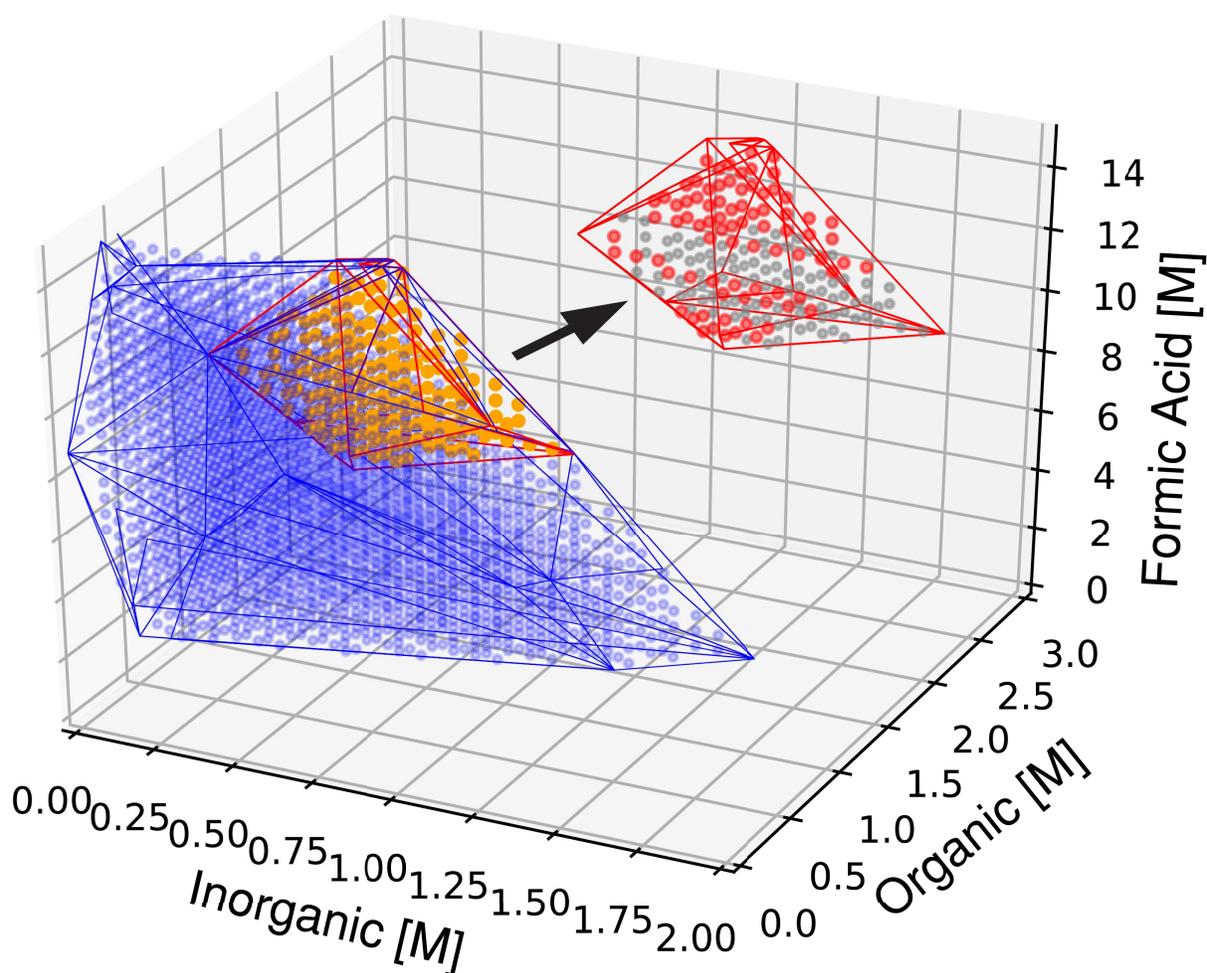
Figure S35. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the tBuA / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
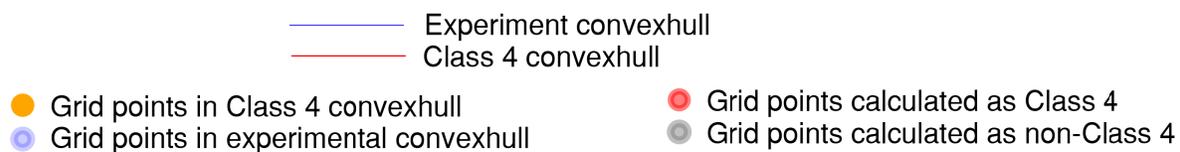
Figure S36. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the 4FBenA / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
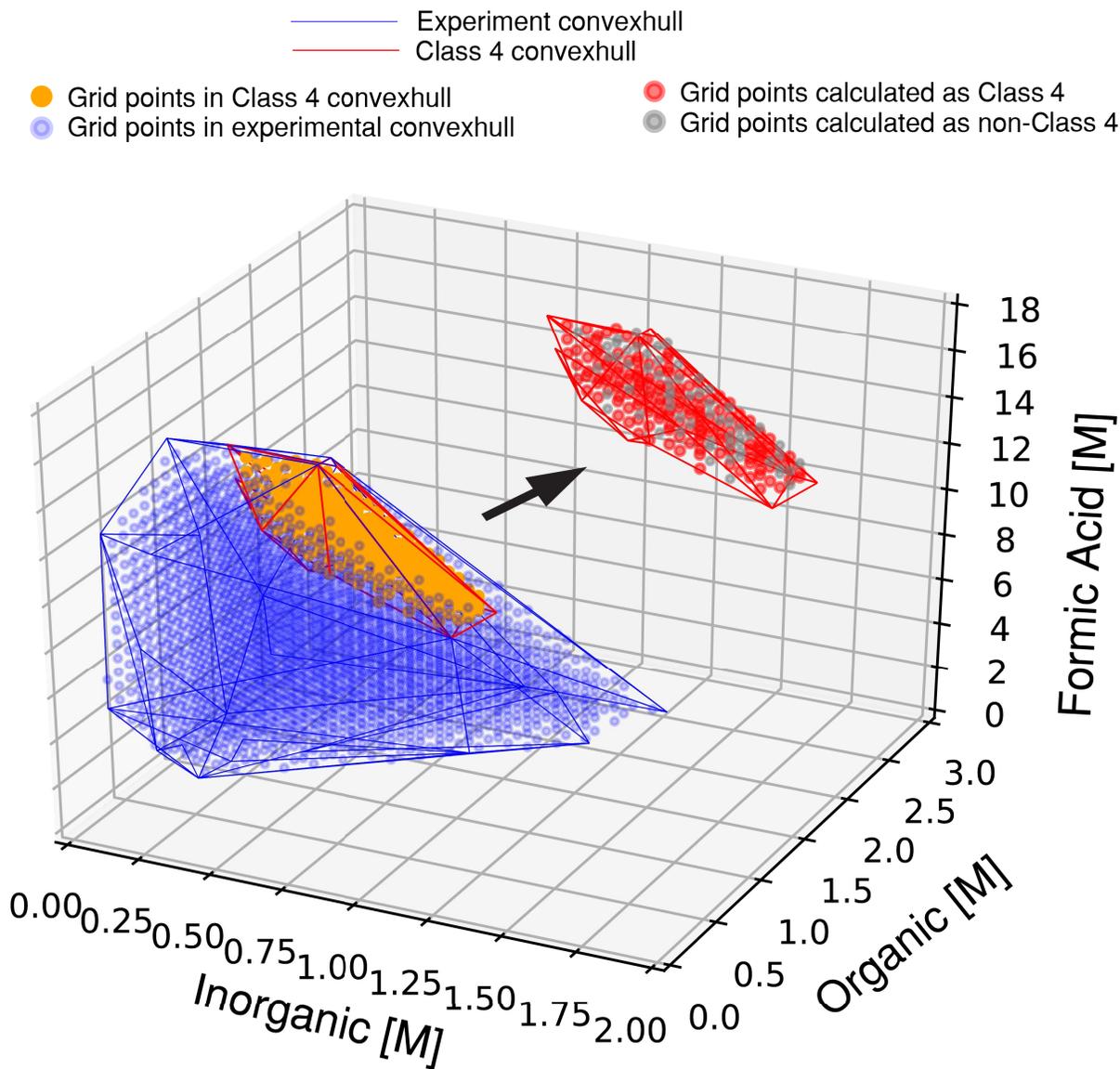
Figure S37. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the CHA / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
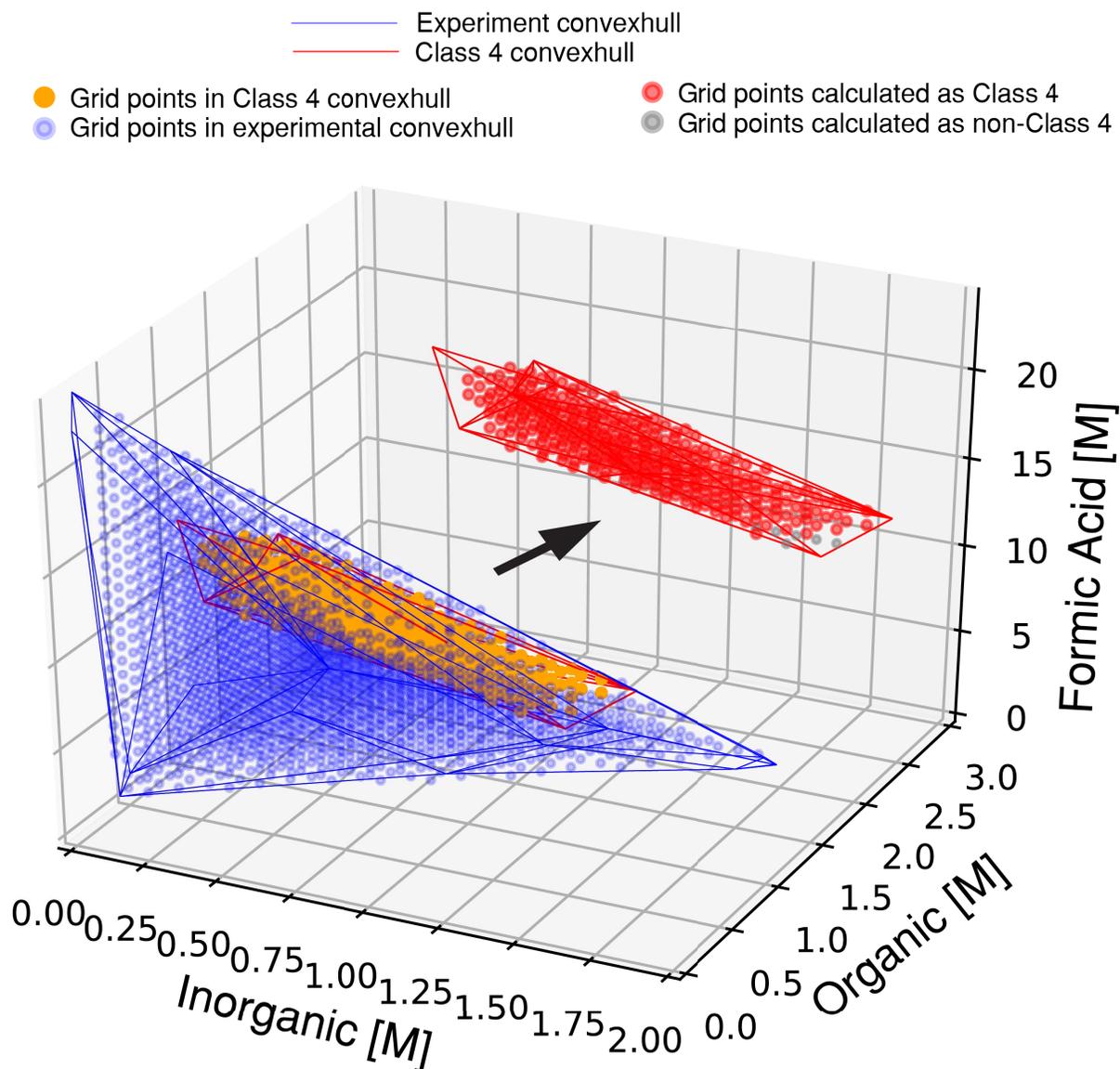
Figure S38. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the Morph / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
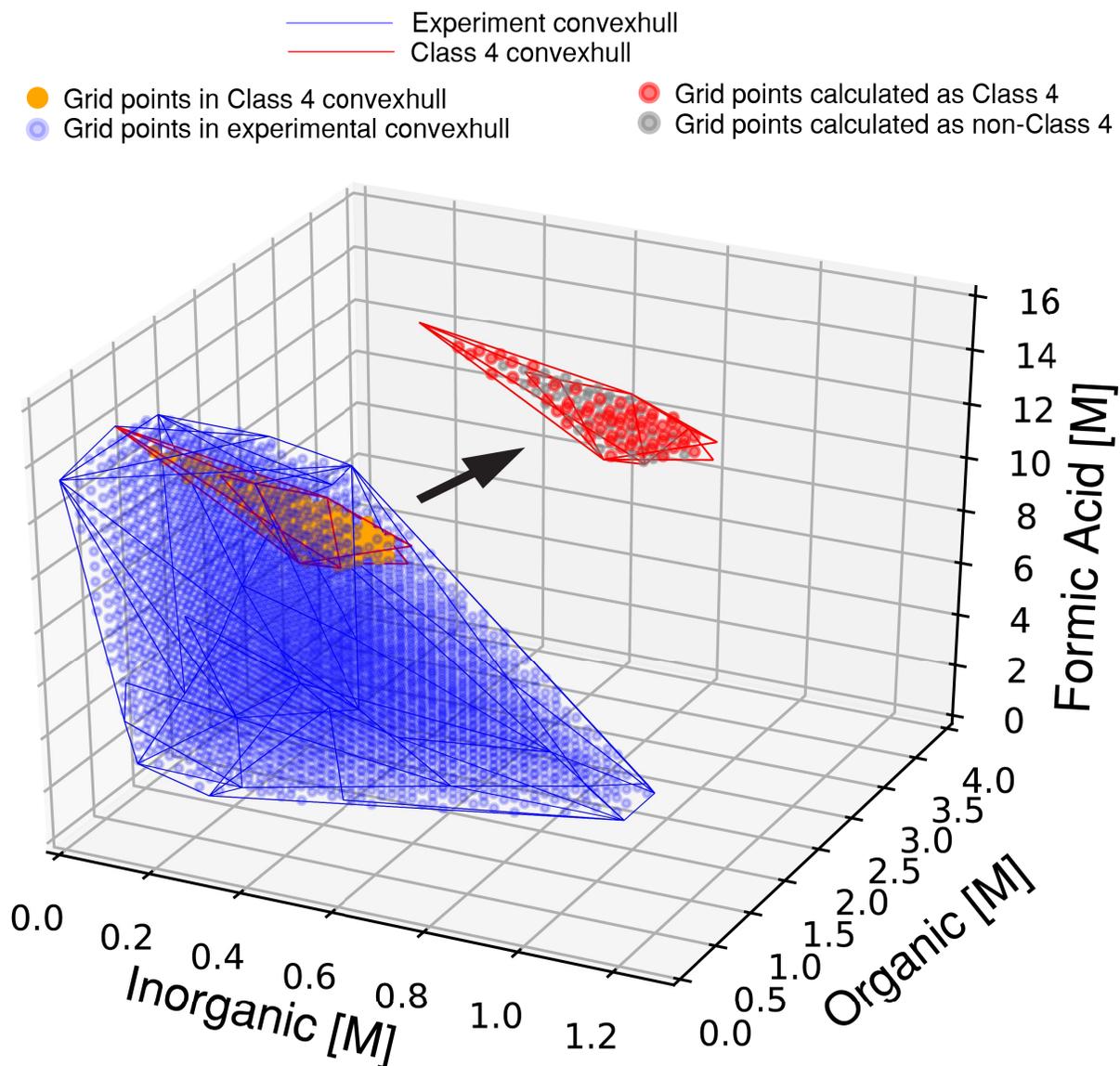
Figure S39. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the Piper / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the *k*-NN algorithm.
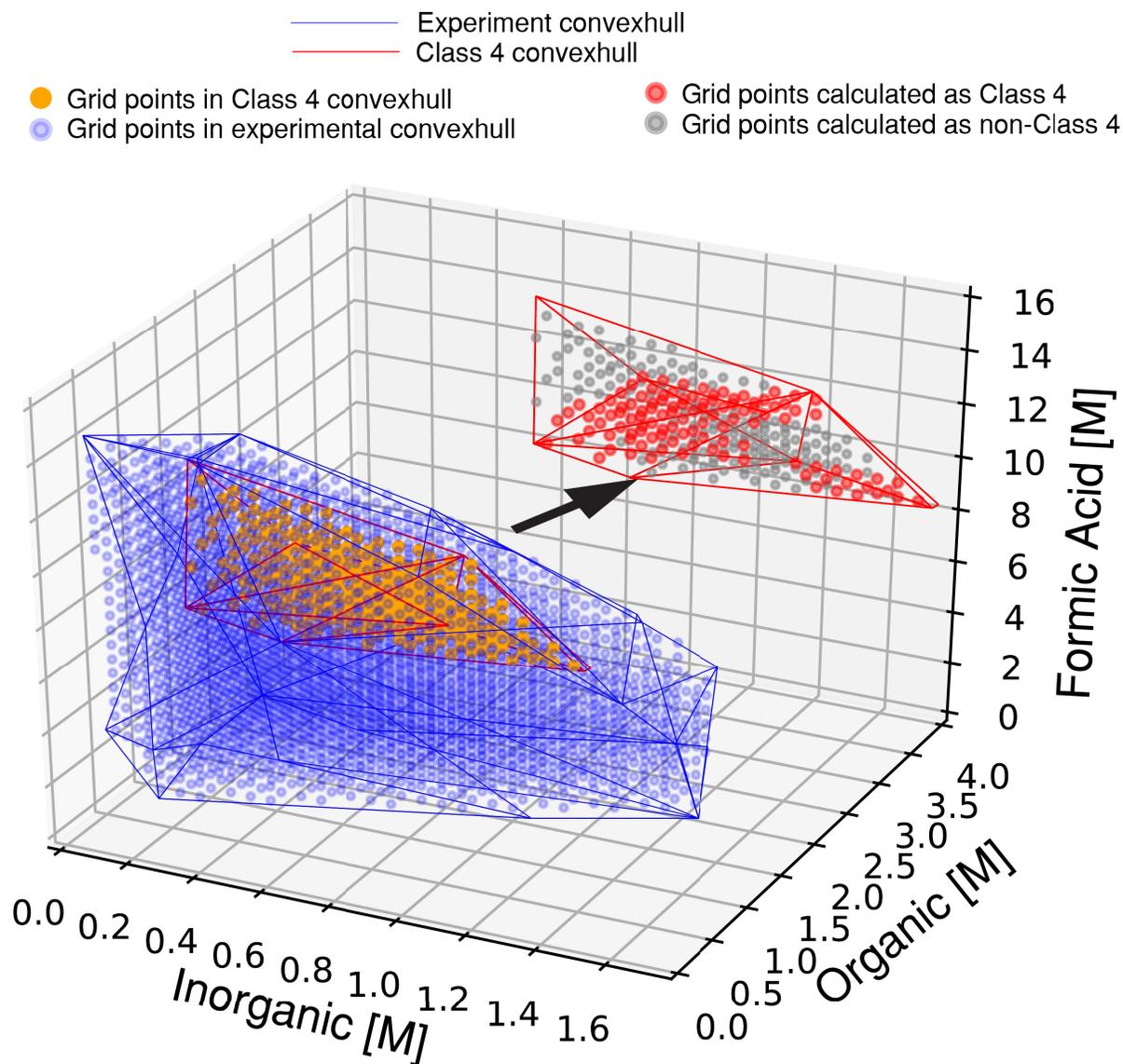
Figure S40. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the nHex / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
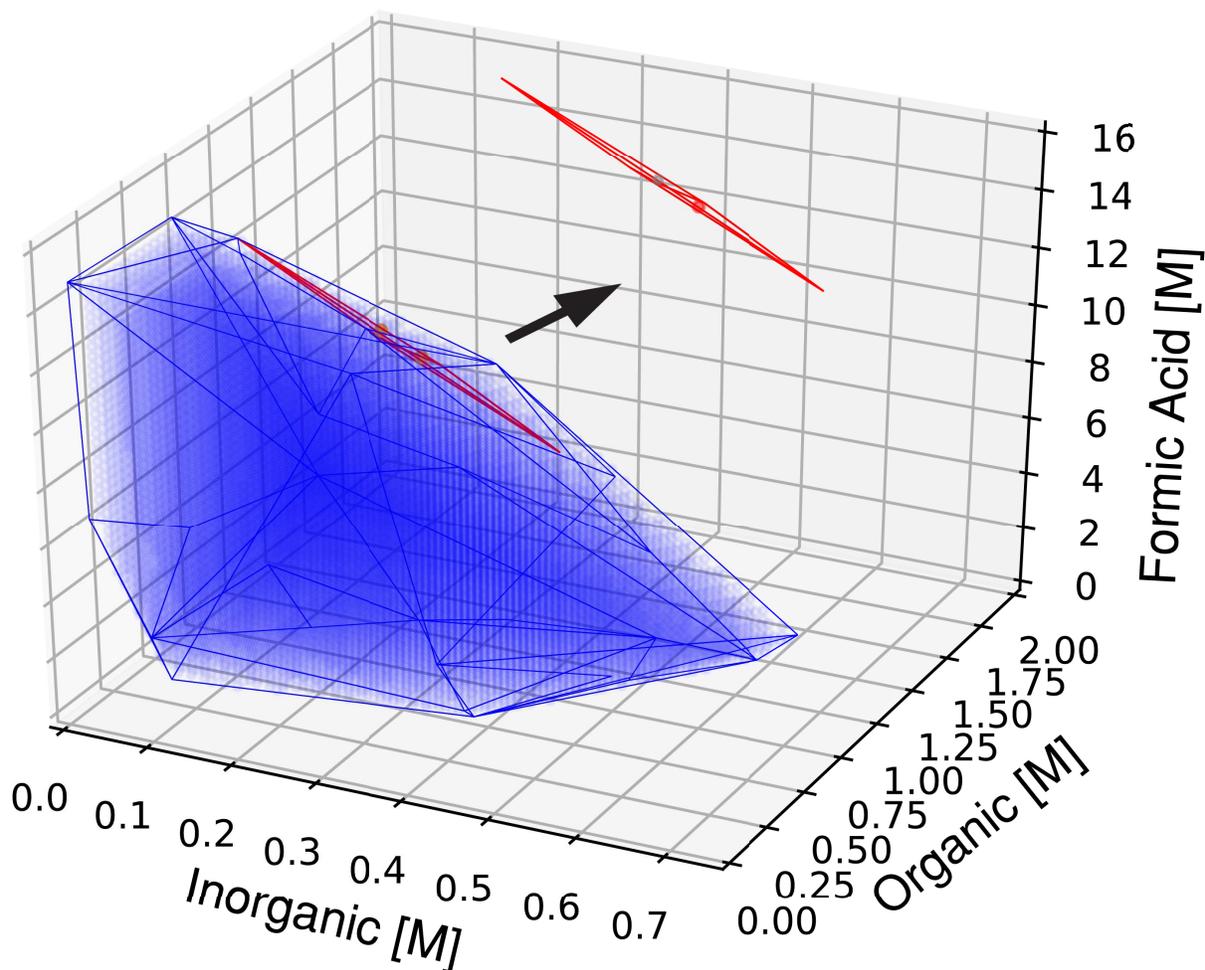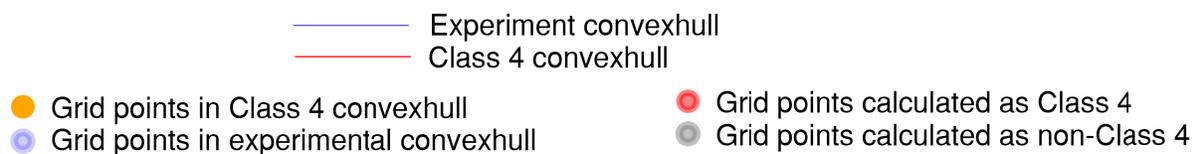
Figure S41. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the NNDiEtPA / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
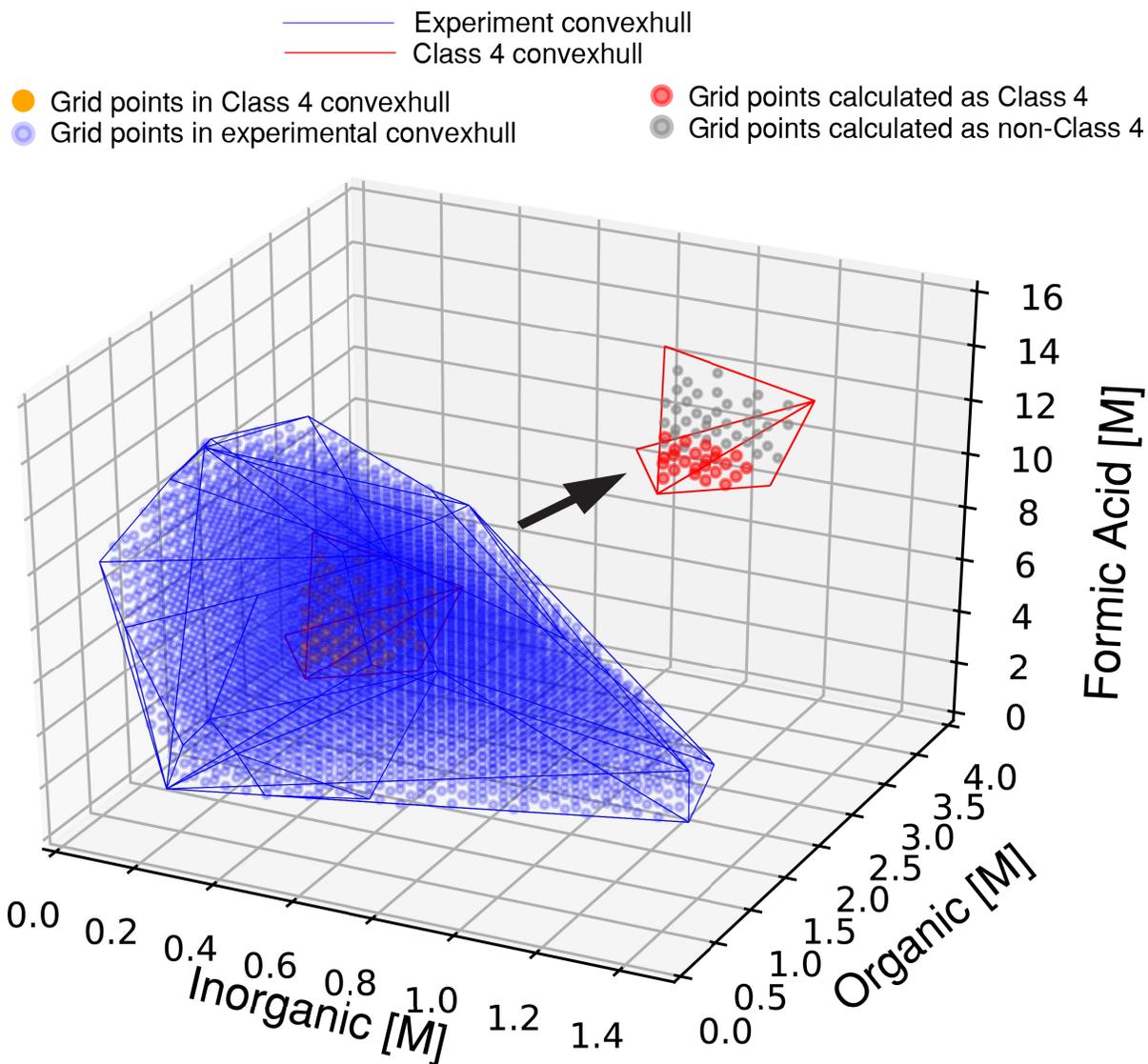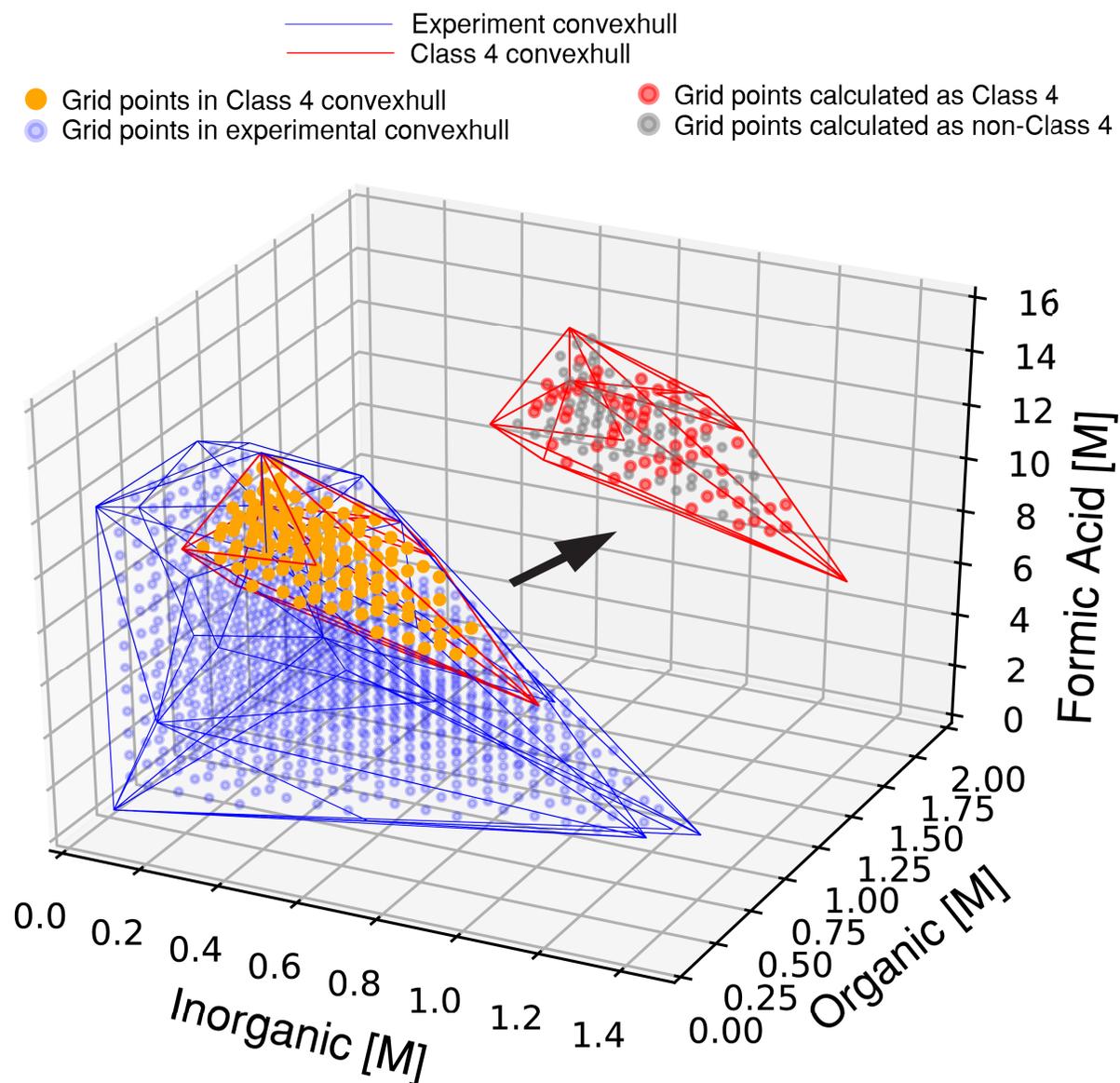
Figure S42. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the NNDiMPA / PbI$_2$ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the k-NN algorithm.
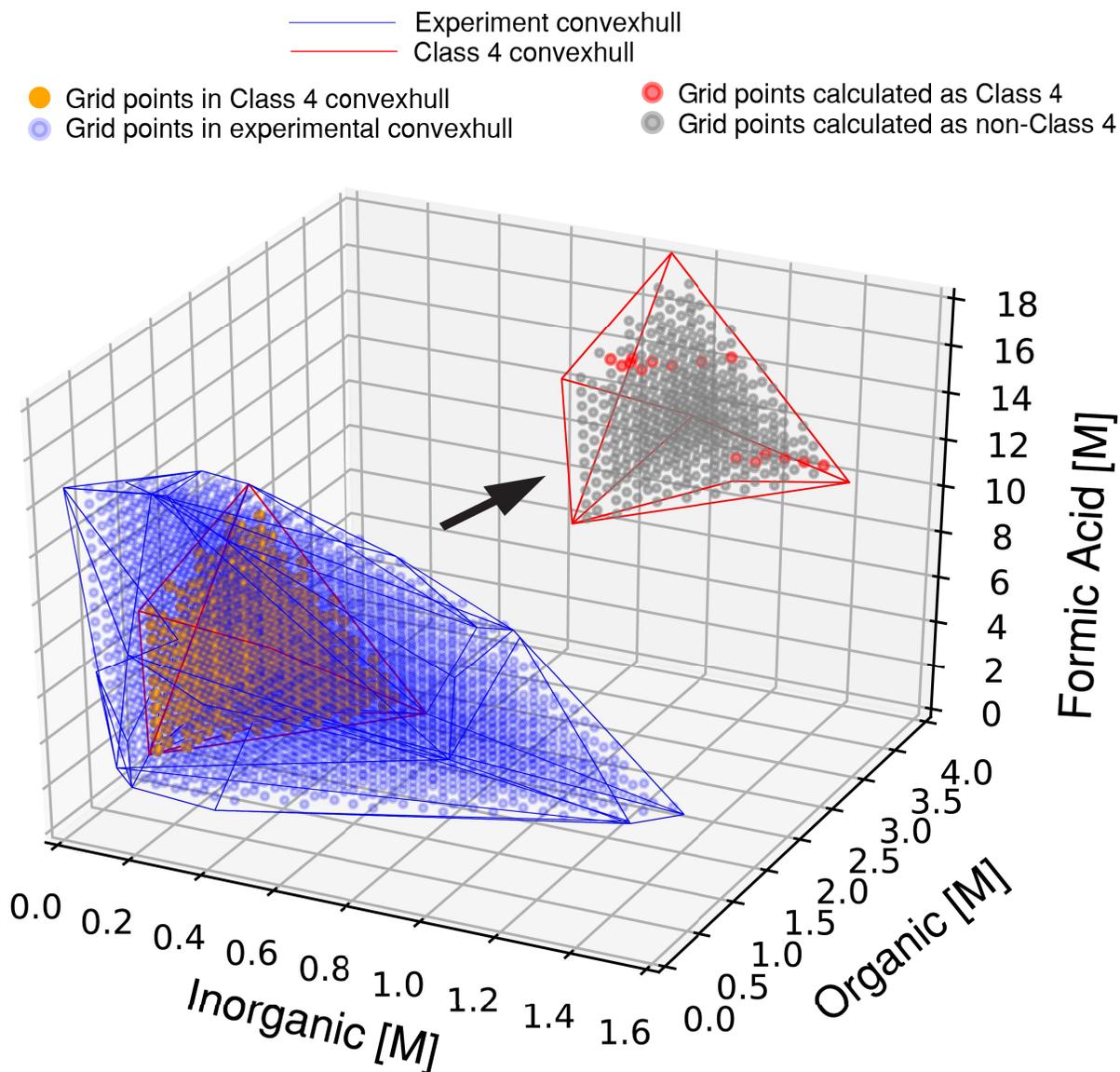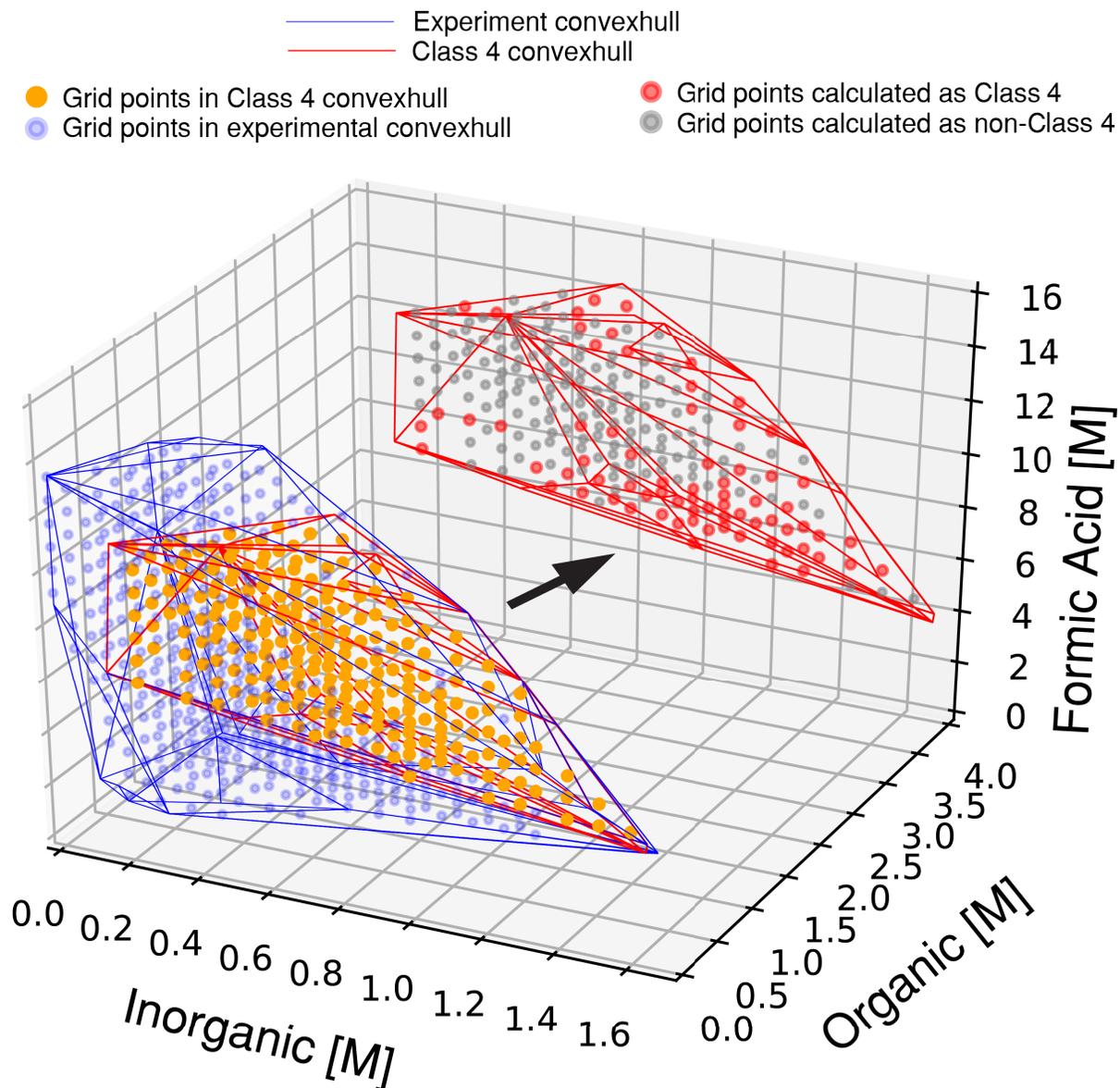
Figure S43. Convex hulls, and the grid points contained within them, for the explored experimental space (blue lines/circles) and Class 4 outcomes (red lines and orange circles), for the DiMA / PbI₂ system. Insets: grid points within the Class 4 convex hull, with colors indicating Class 4 (red circles) and non-Class 4 (grey circles) outcomes as determined by the *k*-NN algorithm.

Table S7. Convex hull volume and likelihood of large single crystal formation, P_4 of explored chemical space for the 19 perovskite syntheses producing Class 4 crystals

Ammoniums	Explored experiment space ($[M]^3$)	Likelihood of large single crystal formation, P_4	Ammoniums	Explored experiment space ($[M]^3$)	Likelihood of large single crystal formation, P_4
MA	5.02	0.10	tBuA	11.57	0.06
EtA	8.97	0.04	4FBenA	14.03	0.15
nBuA	3.96	0.002	CHA	12.56	0.02
FA	3.64	0.53	Morph	14.69	0.03
Acet	13.96	0.15	Piper	18.34	0.01
CHMA	6.05	0.46	nHex	15.05	0.004
iBuA	10.97	0.42	NNDiEtPA	4.97	0.0001
nDodA	1.76	0.01	NNDiMPA	8.01	0.05
Pyrro	9.43	0.04	DiMA	10.00	0.11
4FPhenEtA	9.31	0.03			

Table S8. Prediction metrics of machine learning models

Model	Hyper-parameters	Accuracy	Precision	Recall	F1	AUC
Gaussian Naïve Bayes		0.338	0.26	0.91	0.41	0.63
Quadratic Discriminant Analysis		0.680	0.40	0.51	0.45	0.64
Logistic regression	C = 1000	0.772	0.62	0.24	0.35	0.79
K-nearest neighbor	'n_neighbors': 1, 'weights': 'uniform'	0.845	0.71	0.66	0.68	0.78
SVM (RBF kernel)	kernel = 'rbf', C = 1, gamma = 10	0.841	0.69	0.66	0.68	0.87
SVM (linear kernel)	kernel = 'linear', C = 10	0.680	0.43	0.87	0.58	0.78
SVM (PUFK)	C = 10	0.869	0.78	0.67	0.72	0.91
Decision tree	'max_depth': 9, 'min_samples_leaf': 10	0.817	0.68	0.50	0.57	0.87
Random forest	'max_depth': 9, 'min_samples_leaf': 1, criterion='entropy', n_estimators = 100	0.828	0.82	0.41	0.54	0.91
Neural network	'activation': 'logistic', 'alpha': 0.001, 'hidden_layer_sizes': 1, 'solver': 'sgd'	0.748	0.00	0.00	0.00	0.62

Explanation of Learning Curves

In Figure S44 - S49, Model prediction metrics for Class 4 large crystal formation of six perovskite systems are plotted against various number of training experiments. First, the perovskite experiments were split into training sets and testing sets using 5-fold cross validation as described in "Software and Computation" section. Next, subsets containing a certain number of experiments were randomly drawn from training sets to be used as training experiments. Following this procedure, we have performed five drawings to simulate an unbiased learning process. Last, machine learning models were trained by training experiments and tested by testing sets. For a certain number of training experiments, there are a total of 25 (5 x 5) train-test combinations. In the plots, the solid points are the prediction metrics averaged over 25 different train-test combinations and the shaded bands are corresponding standard deviation.

Figure S44. Accuracy (a), precision (b), recall (c) and f1 (d) for predicting MA / PbI$_2$ Class 4 crystal formation for six different models: PUFK-SVM model; a control model assuming no Class 4 crystals or all Class 4 crystals (dashed line); 1-NN models (k = 1) based on the concentrations of all three precursors (organic, inorganic, formic acid); and 1-NN models based on only one precursor.

Figure S45. Accuracy (a), precision (b), recall (c) and f1 (d) for predicting EtA / PbI$_2$ Class 4 crystal formation for six different models: PUFK-SVM model; a control model assuming no Class 4 crystals or all Class 4 crystals (dashed line); 1-NN models ($k = 1$) based on the concentrations of all three precursors (organic, inorganic, formic acid); and 1-NN models based on only one precursor.
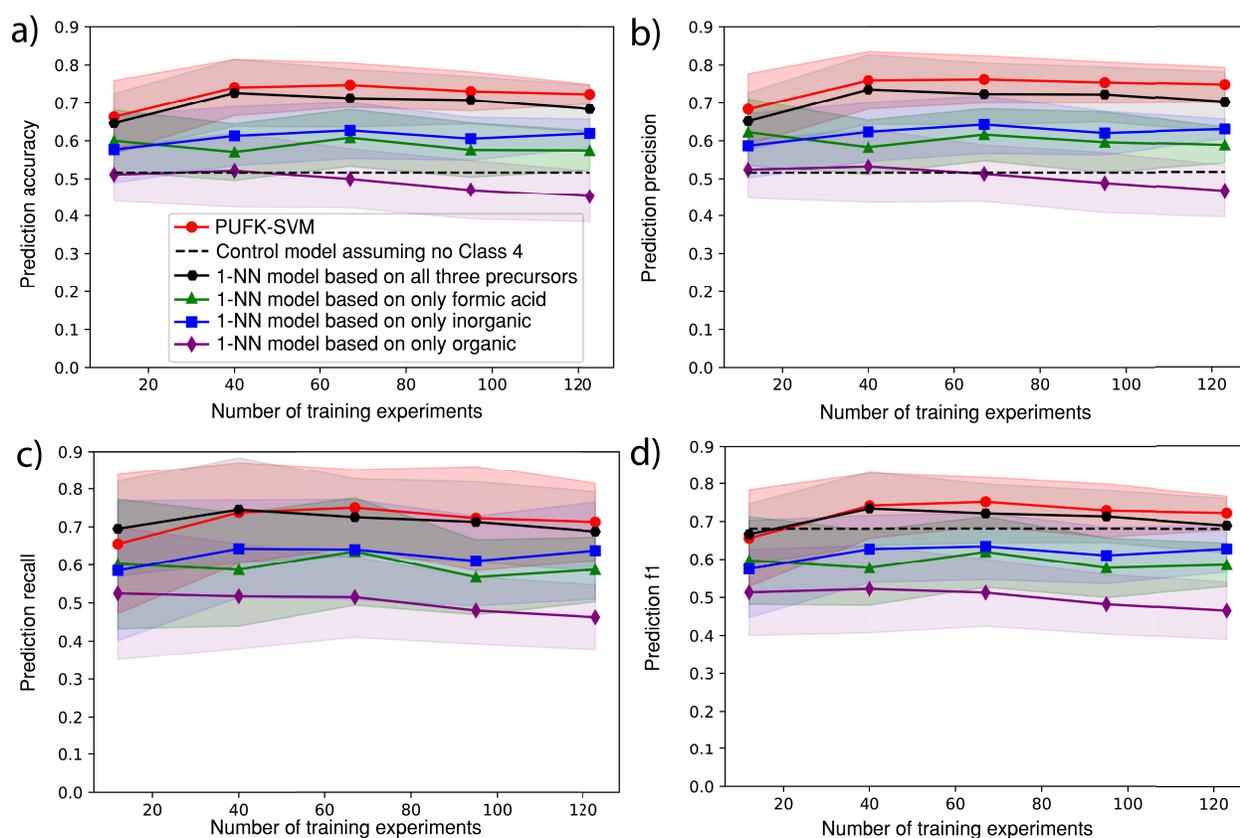
Figure S46. Accuracy (a), precision (b), recall (c) and f1 (d) for predicting nBuA / PbI$_2$ Class 4 crystal formation for six different models: PUFK-SVM model; a control model assuming no Class 4 crystals or all Class 4 crystals (dashed line); 1-NN models ($k = 1$) based on the concentrations of all three precursors (organic, inorganic, formic acid); and 1-NN models based on only one precursor.
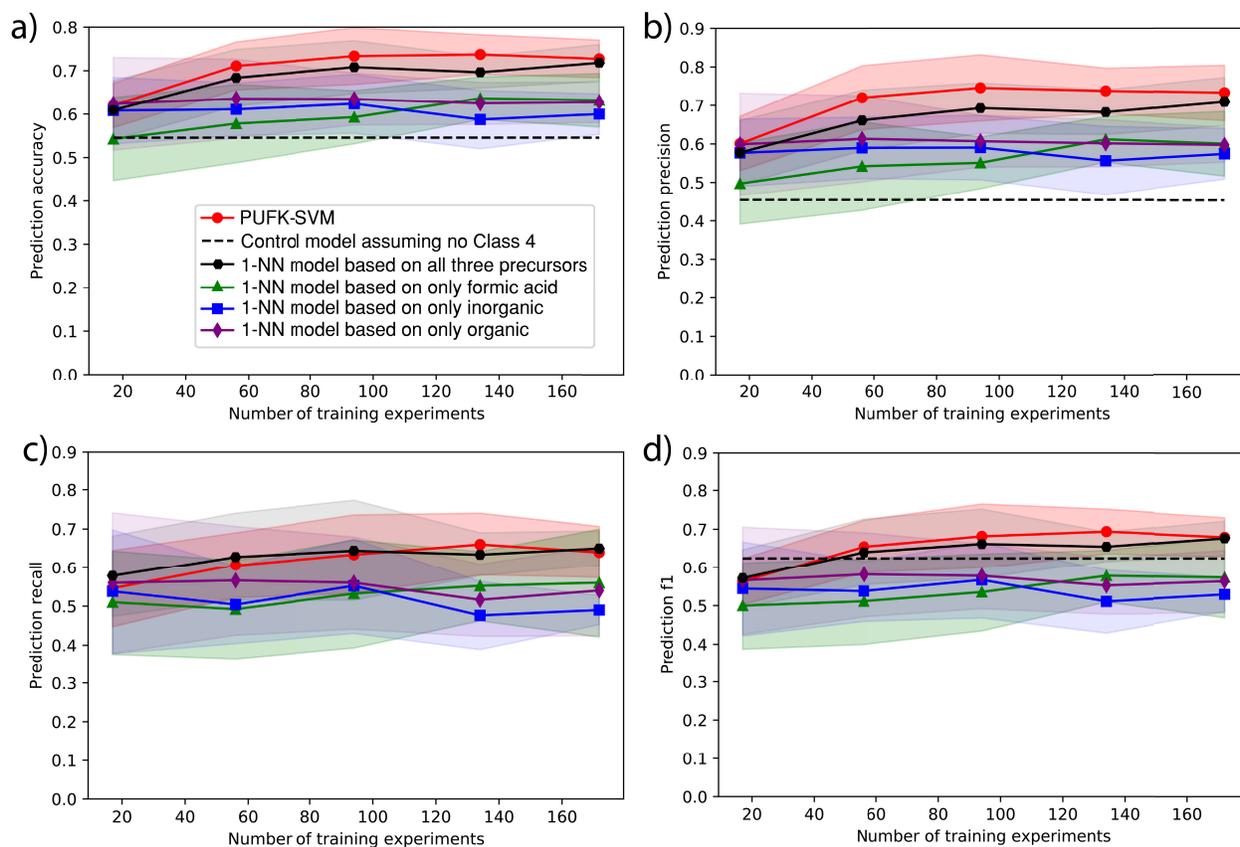
Figure S47. Accuracy (a), precision (b), recall (c) and f1 (d) for predicting FA / PbI$_2$ Class 4 crystal formation for six different models: PUFK-SVM model; a control model assuming no Class 4 crystals or all Class 4 crystals (dashed line); 1-NN models ($k = 1$) based on the concentrations of all three precursors (organic, inorganic, formic acid); and 1-NN models based on only one precursor.
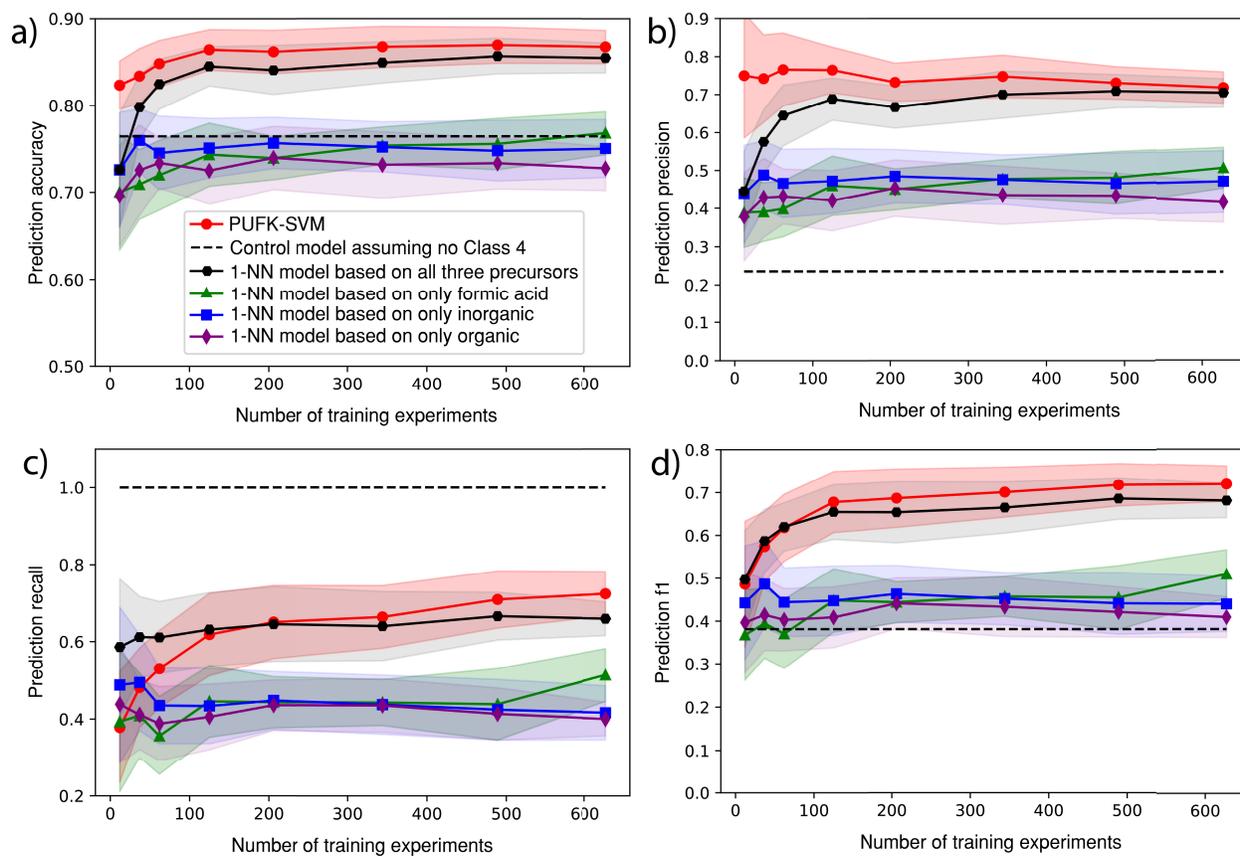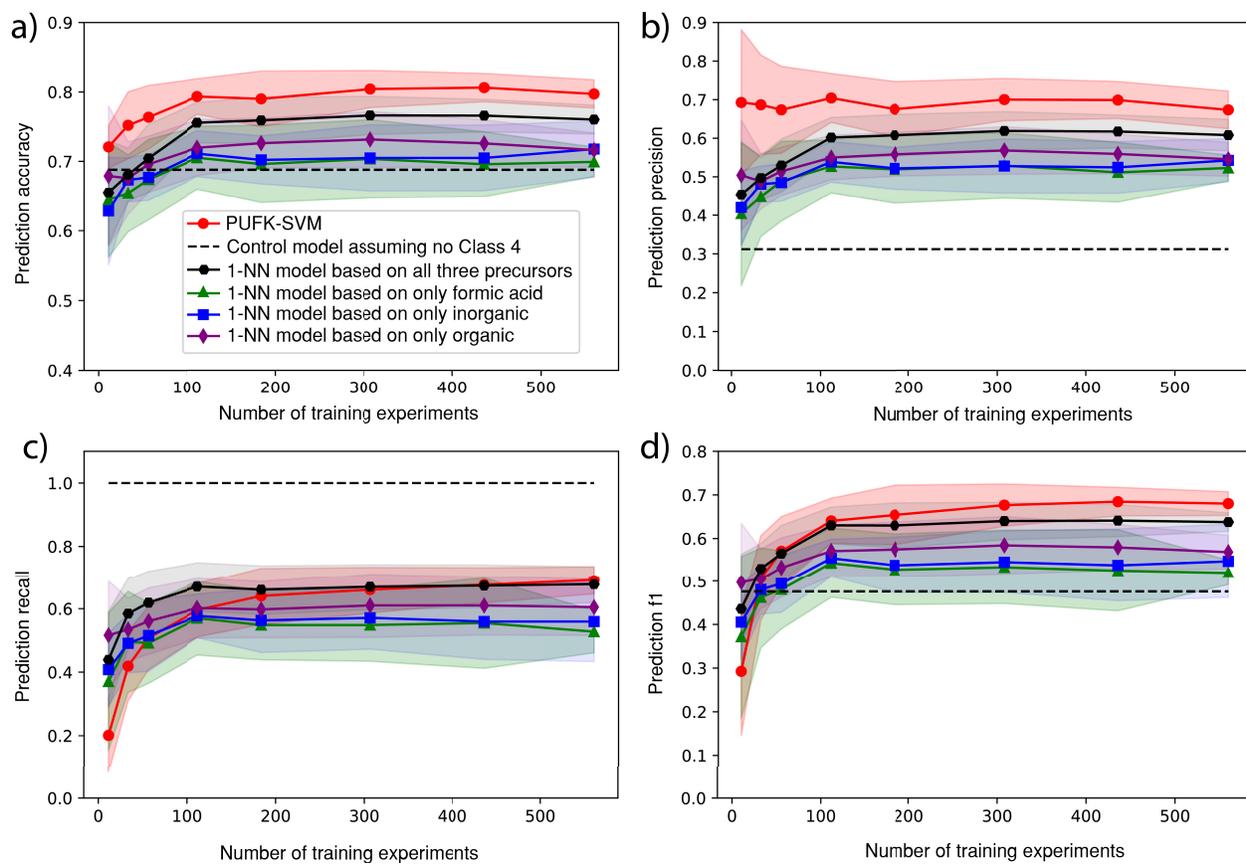
Figure S48. Accuracy (a), precision (b), recall (c) and f1 (d) for predicting Acet / PbI$_2$ Class 4 crystal formation for six different models: PUFK-SVM model; a control model assuming no Class 4 crystals or all Class 4 crystals (dashed line); 1-NN models (k = 1) based on the concentrations of all three precursors (organic, inorganic, formic acid); and 1-NN models based on only one precursor.
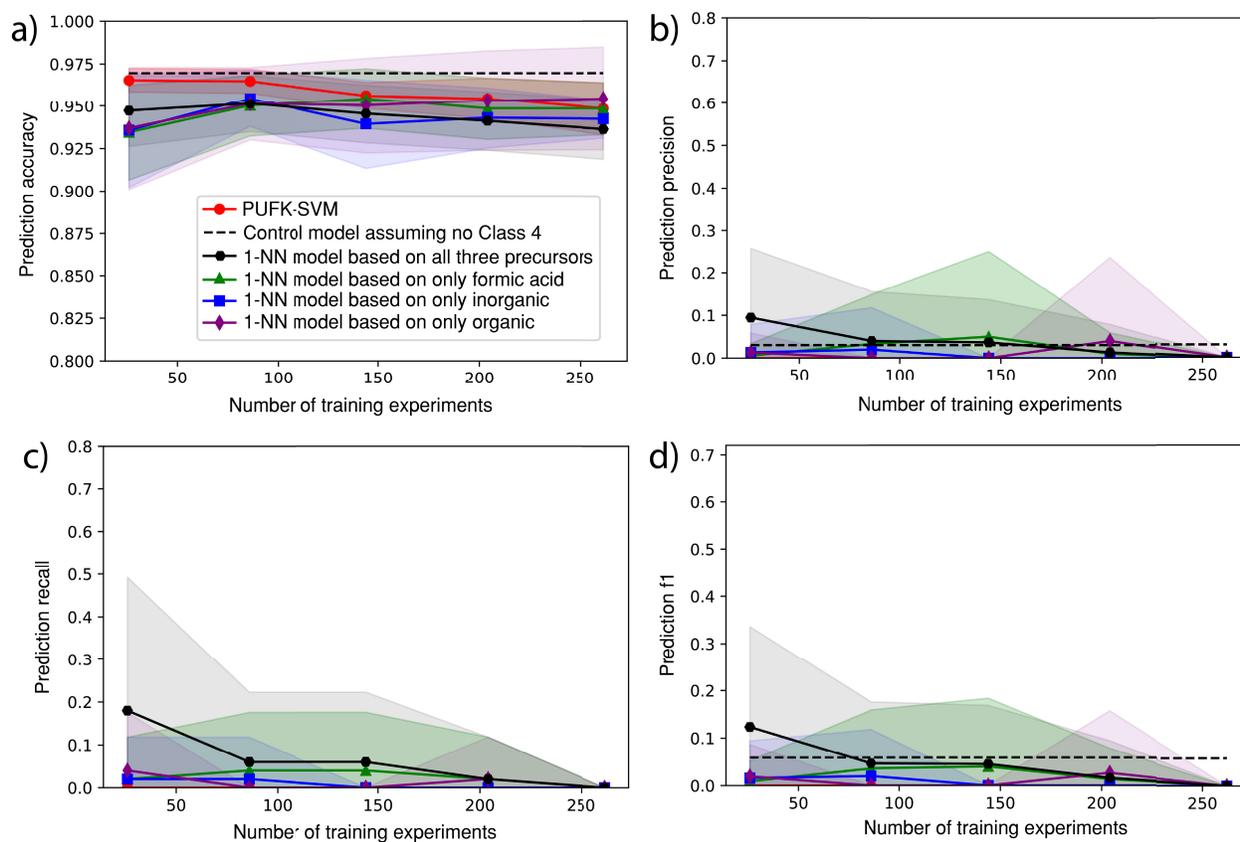
Figure S49. Accuracy (a), precision (b), recall (c) and f1 (d) for predicting CHMA / PbI$_2$ Class 4 crystal formation for six different models: PUFK-SVM model; a control model assuming no Class 4 crystals or all Class 4 crystals (dashed line); 1-NN models (k = 1) based on the concentrations of all three precursors (organic, inorganic, formic acid); and 1-NN models based on only one precursor.
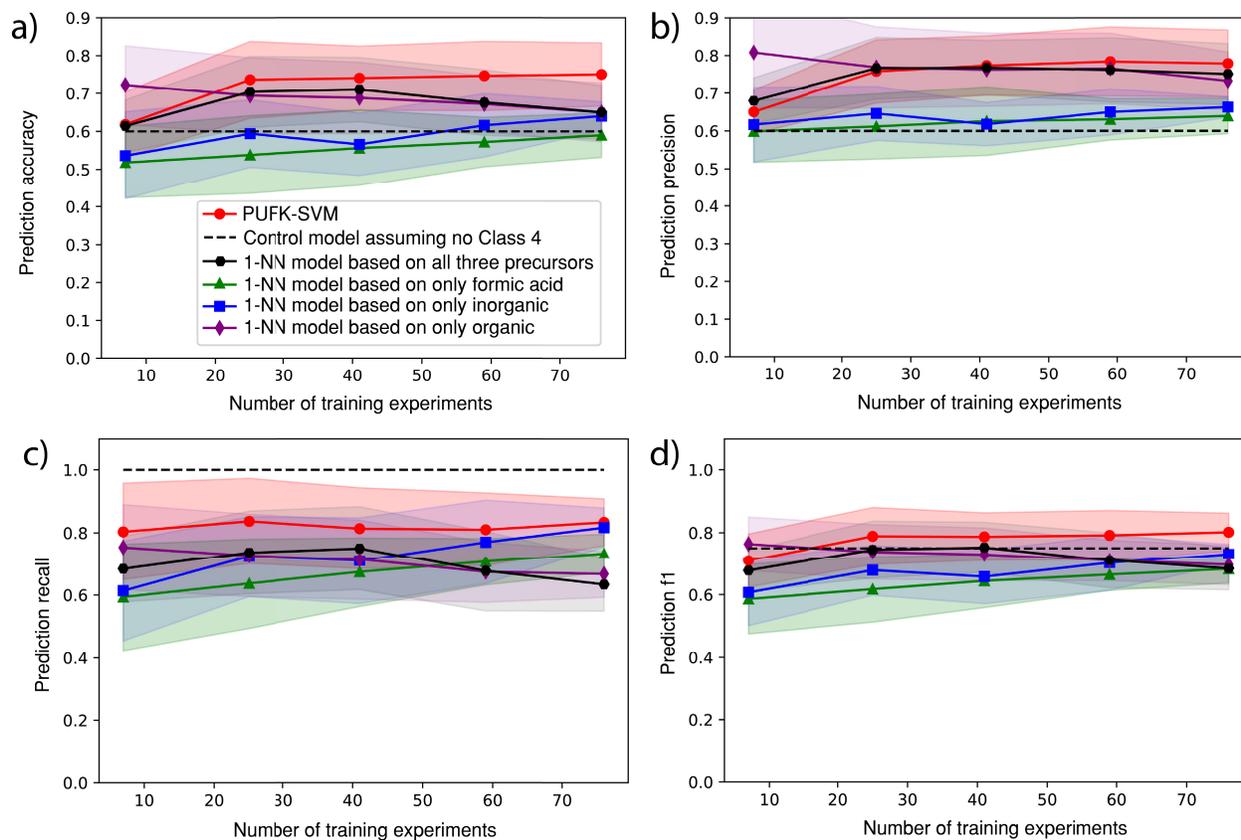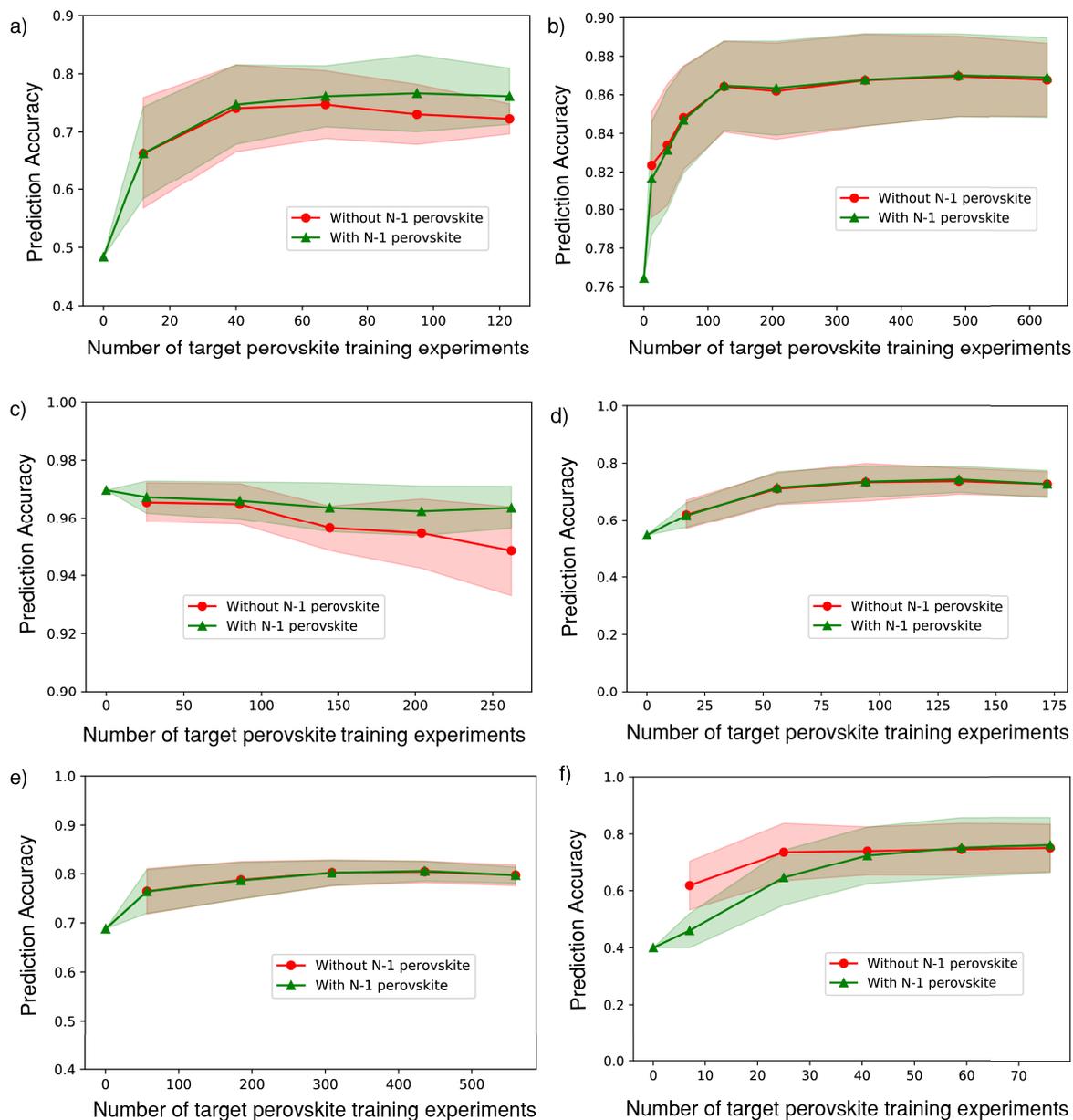
Figure S50. Testing machine learning models for generality across perovskites. PUFK-SVM prediction accuracy vs number of target system training experiments for Class 4 crystal formation. Two types of learning curves were generated for the target systems of MA / PbI$_2$ (a), EtA / PbI$_2$ (b), nBuA / PbI$_2$ (c), FA / PbI$_2$ (d), Acet / PbI$_2$ (e) and CHMA / PbI$_2$ (f). Red curves are trained with *only* experimental data from one perovskite (the "targeted" perovskite). These experiments were split into training sets and testing sets using 5-fold cross validation described in "Software and Computation" section. Shaded bands represent the standard deviation across different training sets. Training experiments were drawn from training sets using the method described in the "***Explanation of learning curves***" section. Green curves were trained with experimental data from the four other non-target perovskites; these data are incorporated into training experiments from the beginning of the learning curves (when there is no target training data) to test the hypothesis that these machine learning models can be generalized across perovskites. The insignificant difference between the red and green curves disproves this hypothesis.
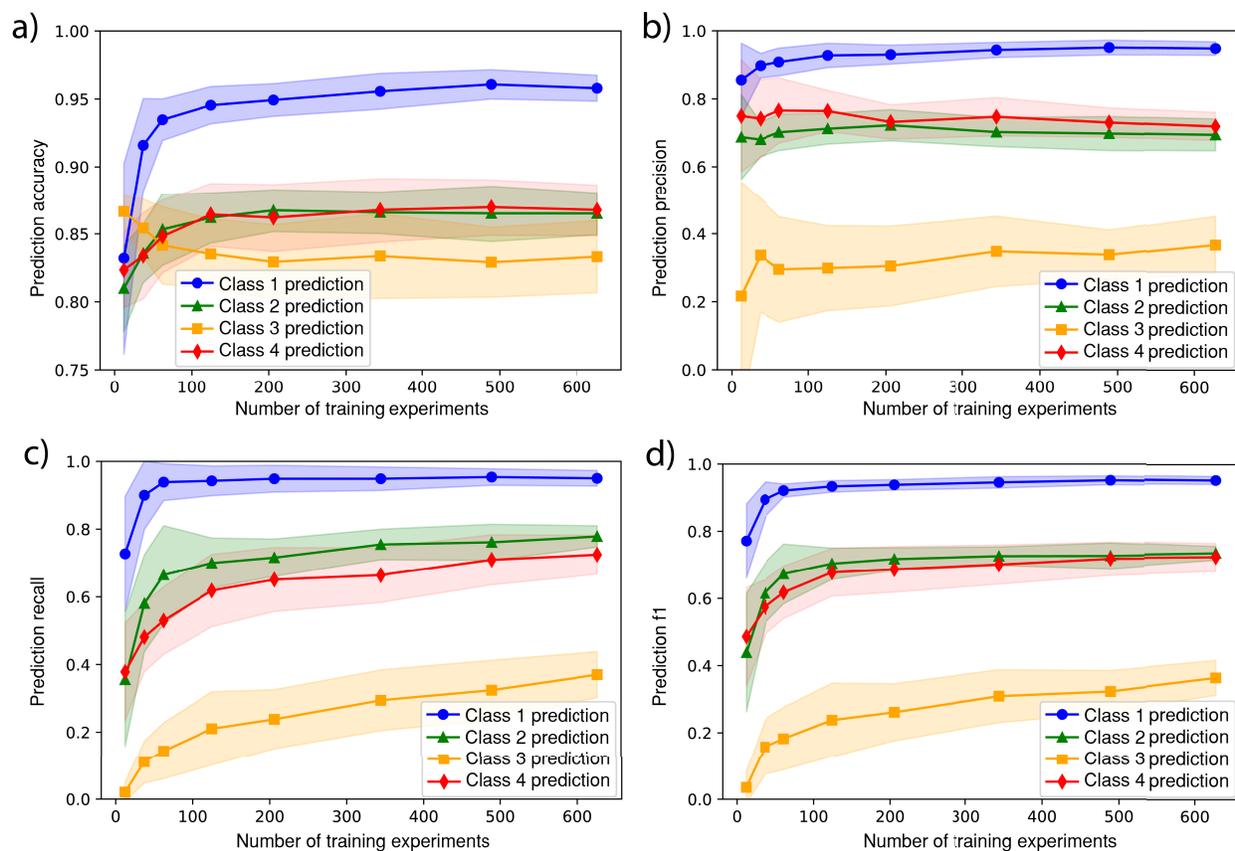
Figure S51. **Predicting individual Classes (1-4) using binary classification machine learning models.** Accuracy (a), precision (b), recall (c) and f1 (d) for predicting EtA / PbI_2 Class 1-4 reaction outcomes for PUFK-SVM model. The PUFK-SVM model is trained for binary classification using Class X reaction outcomes as positive results and non-Class X outcomes as negative results, for X = 1, 2, 3, 4 (e.g., the blue learning curves represent models that consider Class 1 outcomes as positive results and non-Class 1 outcomes as negative results).

4. Features/Descriptors Used for Machine Learning Modeling

Experimental Features

_rxn_M_inorganic
This feature specifies the molarity of the inorganic component (PbI_2) in the reaction solution (unit: molar/liter). (See explanation of experiment parameters in Table S3).

_rxn_M_organic
This feature specifies the molarity of the organic component (organoammonium iodide) in the reaction solution (unit: molar/liter). (See explanation of experiment parameters in Table S3).

_rxn_M_acid
This feature specifies the molarity of the formic acid in the reaction solution (unit: molar/liter). (See explanation of experiment parameters in Table S3)

_rxn_temperatureC
This feature specifies the temperature of the reaction solution at which the crystals growth occurred. (See explanation of experiment parameters in Table S3)

_rxn_mixingtime1S
This feature specifies the duration of the first mixing time (in seconds) after the addition of solvent, organic component, inorganic component, and the first addition of formic acid. (See Robotic Workflow section in the SI for more information)

_rxn_mixingtime2S
This feature specifies the duration of the second mixing time (in seconds) after the addition of the second portion of formic acid. (See Robotic Workflow section in the SI for more information)

_rxn_reactiontimeS
This feature specifies the duration of time (in seconds) that the reaction was are heated undisturbed for to allow for crystal growth. (See Robotic Workflow section in the SI for more information)

_rxn_stirrateRPM
This feature specifies the rate at which the reaction microplate was shaken during the two mixing time (mixingtime1S, mixingtime2S). (See Robotic Workflow section in the SI for more information)

Computed Features

The following descriptors were computed using ChemAxon's Calculator (cxcalc) in the JChem 5.2.0 engine. The suffix corresponds to the name of the descriptor used by cxcalc. See the link below for more information:
https://docs.chemaxon.com/display/docs/cxcalc+calculator+functions

Table S9. Explanation of _feat_desciptors.

Descriptor name	Description
_feat_Mass	Molecular mass of _raw_SMILES (possibly a salt)
_feat_Mass_Nosalt	Molecular mass of _raw_SMILES.nosalt (amine only as neutral species without counterions)
_feat_AtomCount_C (atomcount -z 6)	Number of carbon atoms. (integer)
_feat_AtomCount_N (atomcount -z 7)	Number of nitrogen atoms. (integer)
_feat_AvgPol	Average molecular polarizability (at _rxn_pH)
_feat_MolPol	Molecular polarizability (at _rxn_pH)
_feat_Refractivity	Computed refractivity
_feat_isoelectric (isoelectricpoint)	Isoelectric point of the molecule
_feat_apKa1 (pka –a 2)	First acidic pKa value. Subsequent columns are the subsequent entries in the returned list.
_feat_apKa2	Second acidic pKa value…
_feat_bpKa1 (pka –b 4)	First basic pKa value. Subsequent columns are the subsequent entries in the returned list.
_feat_bpKa2	Second basic pKa value…
_feat_bpKa3	Third basic pKa value …
_feat_bpKa4	Fourth basic pKa value …
_feat_AliphaticRingCount	Number of aliphatic rings (integer)
_feat_AromaticRingCount	Number of aromatic rings (integer)
_feat_Aliphatic AtomCount	Number of aliphatic atoms in the molecule (integer)
_feat_AromaticAtomCount	Number of aromatic atoms in the molecule (integer)
_feat_BondCount	Number of bonds in the molecule (integer)
_feat_CarboaliphaticRingCount	Number of aliphatic rings comprised solely of carbon atoms (integer)

_feat_CarboaromaticRingCount	Number of aromatic rings comprised solely of carbon atoms (integer)
_feat_CarboRingCount	Number of rings comprised solely of carbon atoms (integer)
_feat_ChainAtomCount	Number atoms that are part of chain (not part of a ring) (integer)
_feat_ChiralCenterCount	Number of tetrahedral stereogenic centers (integer)
_feat_RingAtomCount	Number of atoms that are part of a ring (not part of a chain) (integer)
_feat_SmallestRingSize	Number of members in the smallest ring (integer)
_feat_LargestRingSize	Number of members in the largest ring (integer)
_feat_fsp3	Fraction of sp3 carbons (Fsp3 value)
_feat_HeteroaliphaticRingCount	Number of heteroaliphatic rings (integer)
_feat_HeteroaromaticRingCount	Number of heteroaromatic rings (integer)
_feat_RotatableBondCount	Number of rotatable bonds (integer)
_feat_BalabanIndex	Balaban molecular graph index
_feat_CyclomaticNumber	Cyclomatic number of molecular graph
_feat_HyperWienerIndex	Hyper Wiener Index of molecular graph
_feat_WienerIndex	Wiener Index of molecular graph
_feat_WienerPolarity	Wiener Polarity of molecular graph
_feat_MinimalProjectionArea	Minimal projection area
_feat_MinimalProjectionRadius	Minimal projection radius
_feat_MinimalProjectionRadius	Minimal projection radius
_feat_MaximalProjectionRadius	Maximal projection radius
_feat_LengthPerpendicularToTheMinArea (minimalprojectionsize)	Length perpendicular to the minimal projection area
_feat_LengthPerpendicularToTheMaxArea (maximalprojectionsize)	Length perpendicular to the maximum projection area
_feat_VanderWaalsVolume (volume)	van der Waals volume of the molecule
_feat_VanderWaalsSurfaceArea (vdwsa)	van der Waals surface area of the molecule
_feat_ASA (asa -H _rxn_pH)	Water accessible surface area of the molecule, computed at _rxn_pH
_feat_ASA+ (molecularsurfacearea -t ASA+ -H _rxn_pH)	Water accessible surface area of all atoms with positive partial charge,computed at _rxn_pH

_feat_ASA- (molecularsurfacearea -t ASA- -H _rxn_pH)	Water accessible surface area of all atoms with negative partial charge,computed at _rxn_pH
_feat_ASA_H (molecularsurfacearea -t ASA_H -H _rxn_pH)	Water accessible surface area of all hydrophobic atoms with positive partial charge,computed at _rxn_pH
_feat_ASA_P (molecularsurfacearea -t ASA+P -H _rxn_pH)	Water accessible surface area of all polar atoms with positive partial charge,computed at _rxn_pH
_feat_PolarSurfaceArea (polarsurfacearea -H _rxn_pH)	2D Topological polar surface area, computed at_rxn_pH
_feat_acceptorcount (acceptorcount -H _rxn_pH)	Hydrogen bond acceptor atom count in molecule, computed at _rxn_pH
_feat_Accsitecount (acceptorsitecount - H _rxn_pH)	Hydrogen bond acceptor multiplicity in molecule, computed at _rxn_pH
_feat_donorcount (donorcount -H _rxn_pH)	Hydrogen bond donor atom count in molecule, computed at _rxn_pH
_feat_donsitecount	Hydrogen bond donor multiplicity in molecule, computed at _rxn_pH
_feat_sol (solubility -H _rxn_pH)	Aqueous solubility (logS) computed at _rxn_pH

The descriptors describing the number of functional-group fragments (_feat_fr_) were computed. Only functional groups related to amines are considered here. These are determined from the molecular graph, using the SMARTS patterns described in RDKit 2018.03.4. These are all integer values. See the link below for more information:

http://www.rdkit.org/Python_Docs/rdkit.Chem.Fragments-module.html

Table S10. Explanation of _feat_fr desciptors.

Descriptor name	Description
_feat_fr_NH2	Number of primary amines
_feat_fr_NH1	Number of secondary amines
_feat_fr_NH0	Number of tertiary amines
_feat_fr_quatN	Number of quaternary amines
_feat_fr_ArN	Number of N functional groups attached to aromatics
_feat_fr_Ar_NH	Number of aromatic amines
_feat_fr_Imine	Number of imines
_feat_fr_amidine	Number of amidine groups

_feat_fr_dihydropyridine	Number of dihydropyridines
_feat_fr_guanido	Number of guanidine groups
_feat_fr_piperdine	Number of piperidine rings
_feat_fr_piperzine	Number of piperzine rings
_feat_fr_pyridine	Number of pyridine rings

5. References

1   I. M. Pendleton, G. Cattabriga, Z. Li, M. A. Najeeb, S. A. Friedler, A. J. Norquist, E. M. Chan and J. Schrier, *MRS Commun.* **2019**, *9*, 846–859.

2   Scipy.spatial.ConvexHull — SciPy v1.3.0 Reference Guide, https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.ConvexHull.html, (accessed 26 June 2019).

3   G. Zhang and H. Ge, *Comput. Biol. Chem* **2013**, *46*, 16–22.

4   Shekar, V. et al. RAPID Github Repository https://github.com/darkreactions/rapid (accessed Apr 23, 2020).

5   C. Kim, T. D. Huan, S. Krishnan and R. Ramprasad, *Sci. Data* **2017**, *4*, 170057.