ADVANCED MATERIALS

Supporting Information

for *Adv. Mater.*, DOI: 10.1002/adma.202001626

Artificial Chemist: An Autonomous Quantum Dot Synthesis Bot

*Robert W. Epps, Michael S. Bowen, Amanda A. Volk, Kameel Abdel-Latif, Suyong Han, Kristofer G. Reyes, Aram Amassian, and Milad Abolhasani**

Supporting Information

Artificial Chemist: An Autonomous Quantum Dot Synthesis Bot

*Robert W. Epps, Michael S. Bowen, Amanda A. Volk, Kameel Abdel-Latif, Suyong Han, Prof. Kristofer G. Reyes, Prof. Aram Amassian, and Prof. Milad Abolhasani**

SI. 1A: Chemical Inventory

Lead (II) oxide (*>99.9% trace metal basis*, 211907), tetraoctylammonium bromide (*98%*, 294136), zinc iodide (*>98%*, 223883), zinc chloride (*reagent grade, anhydrous*, *>98%*, 793523), oleic acid (*technical grade, 90%*, 364525), and γ-butyrolactone (*ReagentPlus, >99%*, B103608) were purchased from *Millipore Sigma*. Cesium hydroxide 50 wt% solution in water (*99.9% trace metal basis*, AC213601000), zinc bromide (*99.999% trace metal basis*, AC212770100), oleylamine (*>50%*, TCI America O0059), isopropyl alcohol (A416-4), toluene (*extra dry over molecular sieves, 99.85%*, AC364410025), methyl acetate (*extra pure 99%*, AC181380025), quinine sulfate dehydrate (AC418780050), fluorescein (*laser grade, 99%*, AC410620010), sodium hydroxide (*pellets*, S320), and rhodamine 6G (*99%*, AC419010250) were purchased from *Fisher Scientific*. Sulfuric acid (*60%*, 098361) was purchased from *Oakwood Chemical*. The toluene used in all precursors, starting perovskite QDs, and flow studies was dried further over fresh molecular sieves. All other chemicals were used as received.

SI.1B: Preparation of Starting $CsPbBr_3$

The green-emitting $CsPbBr_3$ QDs were prepared using the flow synthesis strategy described in Epps *et al.* and as adapted from Wei *et al.*. Briefly, a 0.04 M bromide solution was prepared by dissolving 1.3 mmol of tetraoctylammonium bromide in a solution of 26 mL of anhydrous toluene and 6.5 mL of oleic acid at room temperature. A 0.04 M cesium-lead-oleate solution was prepared by first forming a 0.2 M solution. 7 mL of oleic acid was combined with 1.3 mmol of lead (II) oxide and 0.12 mL of cesium hydroxide solution, followed by heating in an oil bath at 160 ^{o}C for 30 min then heating in an oven at 120 ^{o}C for 30 min. The final cesium-lead-oleate precursor was completed by diluting 6.5 mL of the stock solution with 26 mL of anhydrous toluene. SQDs were synthesized by continuously delivering the two bromide and cesium-lead-oleate precursors into a custom polyether ether

ketone (PEEK) cross junction at 1.2 mL/min each with an argon stream of 2 mL/min. The segmented gas-liquid flow continuously ran through a 2 m length of 0.04 in inner diameter (ID) fluorinated ethylene propylene (FEP) tubing followed by collection into a continuously stirred vial fed with an anti-solvent stream (γ-butyrolactone) flowing at 1.3 mL/min. QD collection was carried out for 10 min followed by 10 min of additional stirring. The mixture was then centrifuged at 6,500 rpm for 10 min and dispersed in anhydrous toluene at twice the fluid volume of the starting mixture. The dispersed nanocrystals were allowed to age for 12 h. The aged $CsPbBr_3$ QDs were then centrifuged at 4,000 rpm for 4 min. The final supernatant was collected and used for autonomous halide exchange studies.

SI.1C: Preparation of Halide Exchange Precursors

The 0.04 M halide salt precursors, zinc chloride ($ZnCl_2$) and zinc iodide (ZnI_2), were prepared by adding 0.8 mmol of zinc salt to 20 mL of anhydrous toluene and 0.4 mL of isopropyl alcohol. The precursors were then heated at 100 ^oC with ventilation for 5 h, sonicated for 2 min, and allowed to cool before use. The 0.025 M zinc bromide ($ZnBr_2$) precursor was prepared similarly with 0.5 mmol of zinc bromide, 20 mL of anhydrous toluene, and 0.25 mL of isopropyl alcohol. The oleic acid and oleylamine solutions were each prepared by combining with anhydrous toluene in a 50 vol% mixture.

SI.2A: Reactor Operation Procedure

The sampling process integrated into the modeling and condition selection algorithm begins operation as soon as flow conditions are made available. For all uninformed ensemble model training methods, the modeling algorithm first sends two random reaction conditions to the fluidic microprocessor, then operates at an offset with the reactor. In other words, while the flow synthesis reactor is running condition *i*, the algorithm is training and selecting input condition *i+1* with sample set *1…i-1*. For SNOBFIT, CMA-ES, and the pre-trained NN-EPLT,

model selection began with the first condition. Upon receiving a set of reaction conditions, the autonomous flow synthesis platform initiates flow of the precursors and carrier phase into the fluidic path. In order to minimize chemical consumption, the SQD pump does not begin until 100 s after the other pumps were activated. After 360 s (total) the algorithm begins spectral sampling. For UV-Vis absorption followed by photoluminescence sampling, 50 spectra were collected at intervals of 25 ms with integration times of 12 and 9 ms, respectively. This process is replicated five times in series for a single set of reaction conditions. An embedded spectra processing algorithm then isolates the reactive phase spectra from the segmented oil-liquid flow. E_P, E_{FWHM}, and PLQY values are extracted from the resulting spectra and the five replicates are then averaged together. The final outputs are returned to the modeling algorithm for continued condition selection.

SI.2B: Flow System Components

The flow system is driven by seven stainless steel syringes (*Harvard Apparatus*) each connected to a computer-controlled syringe pump (*Harvard Apparatus*, PhD Ultra and *Chemyx*, Fusion 4000). The channels connecting the syringe to the flow junctions are 50 cm each of 0.01 in ID fluorinated ethylene propylene (FEP) tubing, and the PFO and SQD delivery channels are both intersected by a selector valve (*VICI*, C25-3180EUHB), which provides the option to redirect flow to the respective refill vessel. The zinc halide and ligand mixtures are combined within two PEEK cross-junctions (*IDEX-H&*S, P-722) connected in series followed by a 10 cm, 0.01 in ID, FEP passive micromixer. This mixture is directed into a tee-junction (*IDEX-H&*S, P-715) to combine with the SQD channel which is followed by a second micromixer. The homogenous reactive solution is finally combined in a second tee-junction with PFO for isolated droplet formation. The reactor section is comprised of 150 cm of 0.03 in ID FEP followed by the reduced path length flow cell. Optical sampling is conducted using a miniaturized fiber-coupled spectrometer (*Ocean Optics*, HDX UV-Vis),

deuterium halogen lamp (*Ocean Optics*, DH-2000-BAL), and 365 nm high power light emitting diode (*Thorlabs*, M365LP1) all connected to the flow cell through fiber optic patch cords (*Ocean Optics,* QP600-1-SR).

SI.2C: Inline Spectra Sampling

UV-Vis absorption spectra for the transient, reactive phase slugs were automatically isolated from each set of 50 continuously collected samples by first base-line correcting the beer-lambert absorption at 1.6 eV. Spectra were then sorted in descending order of the measured absorption intensity at 3.4 eV (corresponding to the 365 nm emission source), and after discarding outlier spectra caused by the oil-reactive phase interface, the highest measured 20 samples were averaged together. A similar process was performed for photoluminescence (PL) spectra. Using the peak emission intensity, the spectra were sorted in descending order and the highest measured 20 samples were averaged. Photoluminescence quantum yield (PLQY) was calculated from the resulting spectra according to the equation

$$\Phi_S = \frac{\Phi_R I_{Int,S} A_{3.4eV,R} \eta_S}{I_{Int,R} A_{3.4eV,S} \eta_R}$$

where subscripts S and R correspond to the sample and the reference dye, Φ is the quantum yield, I_{Int} is the integral of the photoluminescence peak, $A_{3.4eV}$ is the measured absorption at 3.4 eV, and η is the refractive index.[1,2]

SI. 2D: Flow Cell Validation

The reduced path length flow cell was evaluated using a 1mm×10 mm quartz cuvette for calculation of the path length and a 10 mm×10 mm cuvette for validation of the PLQY measurements. As shown in Figure S1, the flow cell is able to effectively retain a linear relationship with the integral of PL as a function of absorption at the excitation wavelength for all absorption values below 0.1. Using the absorption curves of the same concentrations of

fluorescent dyes (quinine sulfate in 0.05 M sulfuric acid, fluorescein in 0.1 M sodium hydroxide, and rhodamine 6G in ethanol) between the 1 mm path length cuvette and the compressed tube flow cell, the path length of the compressed tubing section was calculated to be 206 ± 1 μm. Offline PLQY validation was conducted according to the equation[3]

$$\Phi_S = \frac{\Phi_R m_S \eta_S}{m_R \eta_R}$$

where *m* is the slope of the linear fit of the integral PL as a function of the absorption at 3.4 eV, with an intercept through the origin.

SI. 3A: Non-Dimensionalization of Input Parameters

Non-dimensional parameters X_{SQD}, X_{ZnX_2}, X_{ZnBr_2}, X_{OA}, and X_{OLA} were correlated to volumetric flow rate of the starting precursors through the relations

$$X_{SQD} = \frac{(Q_{SQD} - 120\mu L/min)}{84\mu L/\text{min}}$$

$$X_{ZnX2} = \frac{Q_{ZnX2}}{0.4 * (240\mu L/min - \text{Q}_{\text{SQD}})}$$

$$X_{ZnBr2} = \frac{Q_{ZnBr2}}{0.3 * (240\mu L/min - \text{Q}_{\text{SQD}})}$$

$$X_{OA} = \frac{Q_{OA}}{0.15 * (240\mu L/min - \text{Q}_{\text{SQD}})}$$

$$X_{OLA} = \frac{Q_{OLA}}{0.15 * (240\mu L/min - \text{Q}_{\text{SQD}})}$$

where Q is the input volumetric flow rate of the referenced precursor and all non-dimensional values (X) are constrained to values between 0 to 1. The scalar parameters shown in the denominator of these relationships were determined, through preliminary studies, to most effectively isolate a meaningful window of attainable reaction conditions. The flow rate of toluene, the reaction solvent, is then set according to the relationship

$$Q_{Toluene} = 240\mu L/min - Q_{SQD} + Q_{ZnX2} + Q_{ZnBr2} + Q_{OA} + Q_{OLA}.$$

thereby keeping the total flow rate of the reactive phase at 240 µL/min. The flow rate of the carrier phase, perfluorinated oil, was kept constant at 180 µL/min. For all experiments conducted in this study, the average set flow rates for Q_{SQD}, Q_{ZnX2}, Q_{ZnBr2}, Q_{OA}, Q_{OLA}, and $Q_{Toluene}$ are 148.5, 16.5, 13.8, 7.3, 7.1, and 46.7 µL/min respectively. The duration of unassisted experimentation was limited by the 50 mL toluene syringe, which on average emptied after 17.8 hr. However, the inclusion of additional syringe refill modules is a simple modification of the reaction system that would extend the unassisted run duration significantly.

SI.3B: Neural Network-Based Modeling

The ensemble model consists of 500 classical cascade-forward neural networks (illustrated in Figure S7) of randomized architecture with model and data weighting assigned through *Adaboost.RT*. For every network, the number of layers used is randomly selected between either 1 or 2 and the number of nodes in each is randomly assigned to an integer value between 5 and 25. Models are trained for five inputs and three outputs using the Levenberg-Marquardt method with *Adaboost* data weights to bias training. Exchanges using $ZnCl_2$ and ZnI_2 were treated as separate systems, and their respective models were trained independently. Measurements with which any of the three output metrics are unable to be derived are discarded from model training. Before insertion into the boost function as well as the subsequent decision policies, the three output parameters are converted into a scalar quality metric (Z) through an objective function, shown in the equation

$$Z(E_P, \Phi, E_{FHWM}) = A_{PE}\frac{|E_P - E_{P,Set\ Point}|}{B_{PE}} + A_{PLQY}(1-\Phi) + A_{FWHM}\frac{E_{FWHM}}{B_{FWHM}}$$

Scaling constants B_{PE} and B_{FWHM} are 1.2 eV and 0.4 eV, and output parameter weight fractions A_{PE}, A_{PLQY}, and A_{FWHM} are 0.85, 0.1, and 0.05 respectively. Preliminary model selection and objective function performance studies are further shown in Figure S8 and S9.

SI. 3C: Decision-Making Policies

For a given set of synthesis conditions X, we let $f^*(X)$ be an unknown, ground truth experimental response. In Bayesian Optimization, we attempt to identify the inputs X^* that maximize / minimize the ground-truth response $f^*(X)$ through a strategic selection of sequential experiments that result in noisy observations $\hat{f}(X)$ which we will assume to be normally distributed, additive perturbations from the ground truth:

$$\hat{f}(X) = f^*(X) + W$$

where $W \sim N(0, \sigma^2)$ is a normally distributed noise with variance σ_W^2. After n such experiments, we assume we may obtain a time-n estimate $f_n(X)$ of the ground truth response, as well as a quantification $\sigma_n(X)$ of the uncertainty of this estimate. Specifically, we assume that

$$f^*(X) \sim N\big(f_n(X), \sigma_n^2(X)\big) \tag{1}$$

That is, we shall assume that the unknown ground truth response at any input X is normally distribution with mean estimate $f_n(X)$ and variance $\sigma_n^2(X)$. The functions $f_n(X)$ and $\sigma_n^2(X)$ can be estimated, for example, using an ensemble of candidate models.

Given time-n beliefs on the ground truth response, the *Artificial Chemist* selects the next set of synthesis conditions X_{n+1} to test using a decision-making policy. For example, the *Artificial Chemist* can select an experiment at random in order to explore the space of synthesis conditions. This policy is called **Pure Exploration** and can be conducted independent of time-n beliefs. In contrast and in a slightly more informed strategy, if the *Artificial Chemist* has high confidence in the estimate f_n, it could try the predicted optimal conditions according to this estimate. This policy is called the **Pure Exploitation (EPLT)** policy:

$$X_{n+1}^{EPLT} = argmax\big(f_n(X)\big)$$

If, on the other hand, a large amount of uncertainty prevents the *Artificial Chemist* from trusting the estimate, it could instead consider the experimental conditions whose estimated response has the largest amount of uncertainty. This is called the **Maximum Variance (MV)** policy:

$$X_{n+1}^{MV} = argmax\left(\sigma_n^2(X)\right)$$

These simple policies often perform poorly when attempting to find the optimal set of conditions within a limited number of experiments because they do not adequately balance between resolving uncertainties in the estimate and looking for optimal values of the response. In other words, given a small experimental budget, the *Artificial Chemist* would like to develop a reasonable estimate of the ground truth only as much as that estimate helps it to determine the optimal values of the ground truth. It should be noted that we do not wish to expend experimental runs learning the function globally. The problem of achieving the trade-off of learning the ground truth to high degree of fidelity *vs.* focusing on finding its extrema is called the exploration versus exploitation dilemma. Below, we outline two policies that better strike this balance.

The **Upper Confidence Bound (UCB)** policy attempts to maximize

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

where k is a parameter in this case defined as $k = \frac{2}{\sqrt{2}}$. It includes both an exploitation term f_n and an exploration term σ_n. The quantity $f_n(X) + k\sigma_n(X)$ can be thought of as an *optimistic estimate* of the ground truth response function value $f^*(X)$. Specifically, Dasgupta[4] showed that under the normality assumptions in (1), the probability that the ground truth response $f^*(X)$ deviates from the mean estimate $f_n(X)$ by an amount at most $k\sigma_n(X)$ is given by

$$\mathbb{P}\left[|f^*(X) - f_n(X)| \geq k\sigma_n(X)\right] \leq \frac{1}{3k^2}$$

Due to the symmetry of the normal distribution, this means that with confidence level $1 - \frac{1}{6k^2}$, the ground truth response $f^*(X)$ is at most $f_n(X) + k\sigma_n(X)$. It is important to note here that this derivation of the UCB policy differs slightly from the typical treatment, which assumes a more Frequentist interpretation. In the standard treatment, sample mean and variances are calculated empirically from the data, and an appeal to Hoeffding's inequality rather than that of Dasgupta is used to determine confidence levels on the similarly obtained optimistic estimates of the ground truth.

The **Expected Improvement (EI)** policy attempts to maximize a measure of improvement. After n experiments, we let y_n^* be the maximal response value observed. For an experiment X, the *improvement* is defined as

$$[f^*(X) - y_n^*]^+ = \begin{cases} f^*(X) - y_n^* & \text{if } f^*(X) > y_n^* \\ 0 & \text{otherwise} \end{cases}$$

In other words, the improvement is a measure of how much more the experimental response will be if the *Artificial Chemist* was to run experiment X compared to the best response it had seen so far. If running experiment X would not result in surpassing the previous best value, then the improvement value is simply 0. Of course, this cannot be evaluated because the *Artificial Chemist* does not know $f^*(X)$. However, the *Artificial Chemist* can use its time-n beliefs to calculate the *expected improvement*:

$$X_{n+1}^{EI} = argmax(\mathbb{E}_n[f^*(X) - y_n^*]^+)$$

where the $\mathbb{E}_n$ is the expected value integration operator under time-n beliefs on f^*. Under normality assumptions (1), this expectation can be calculated according to the following expression taken from[5]

$$\mathbb{E}_n[f^*(X) - y_n^*]^+ = [\Delta_n(X)]^+ + \sigma_n(X)\varphi\left(\frac{\Delta_n(X)}{\sigma_n(X)}\right) - |\Delta_n(X)|\phi\left(\frac{\Delta_n(X)}{\sigma_n(X)}\right)$$

where $\Delta_n(X) = f_n(X) - y_n^*$ and φ, ϕ are the standard normal probability density and cumulative density functions, respectively. The above derivation is made assuming the ground

truth response can be observed without noise. To include noise, several heuristics and augmentations to the above are typically employed. For example, instead of using the previously observed best value, the *Artificial Chemist* could instead use the maximum of the current estimate $y_n^* = \max(f_n(X))$.

SI. 4: Offline Analyses Sample Preparation

Samples were prepared for offline analyses by first collecting the synthesized product in flow and combining with different methyl acetate, depending on the target composition (for target emissions of 2.8, 2.6, 2.4, 2.2, and 2.0 eV, ratios of crude solution to methyl acetate of 5:4, 5:4, 5:4, 1:1, and 2:3 were used respectively). The solution was mixed for 5 min then centrifuged at 6,400 rpm for 20 min. TEM samples were drop-cast onto 200 mesh copper grids with carbon coating (*Ted Pella*). TEM images, shown in full in Figure S10, were collected using the FEI Talos F200X operated at 200 kV. X-ray diffraction (XRD) and grazing incidence X-ray refraction (GIXRD) (see Figure S11) samples were drop-cast onto fused silica substrates under inert conditions. XRD and GIXRD were performed with a Rigaku SmartLab X-ray Diffractometer (Cu K_α source, 1.54 Å, 44mA, 40kV) in Bragg-Bretano and parallel beam configurations, respectively.

SI References

[1] S. Kedenburg, M. Vieweg, T. Gissibl, and H. Giessen. “Linear refractive index and absorption measurements of nonlinear optical liquids in the visible and near-infrared spectral region”, Opt. Mat. Express 2, 1588-1611 (2012).

[2] G. M. Hale and M. R. Querry. Optical constants of water in the 200-nm to 200-μm wavelength region, Appl. Opt. 12, 555-563 (1973).

[3] HORIBA UK Limited, “A Guide to Recording Fluorescence Quantum Yields”, *Technical manual*

[4] Anirban Dasgupta, “Best Constants in Chebyshev Inequalities with Various Applications”, *Metrika*, 51, 185, 2000.

[5] Peter I. Frazier, “A Tutorial on Bayesian Optimization”, 2018.

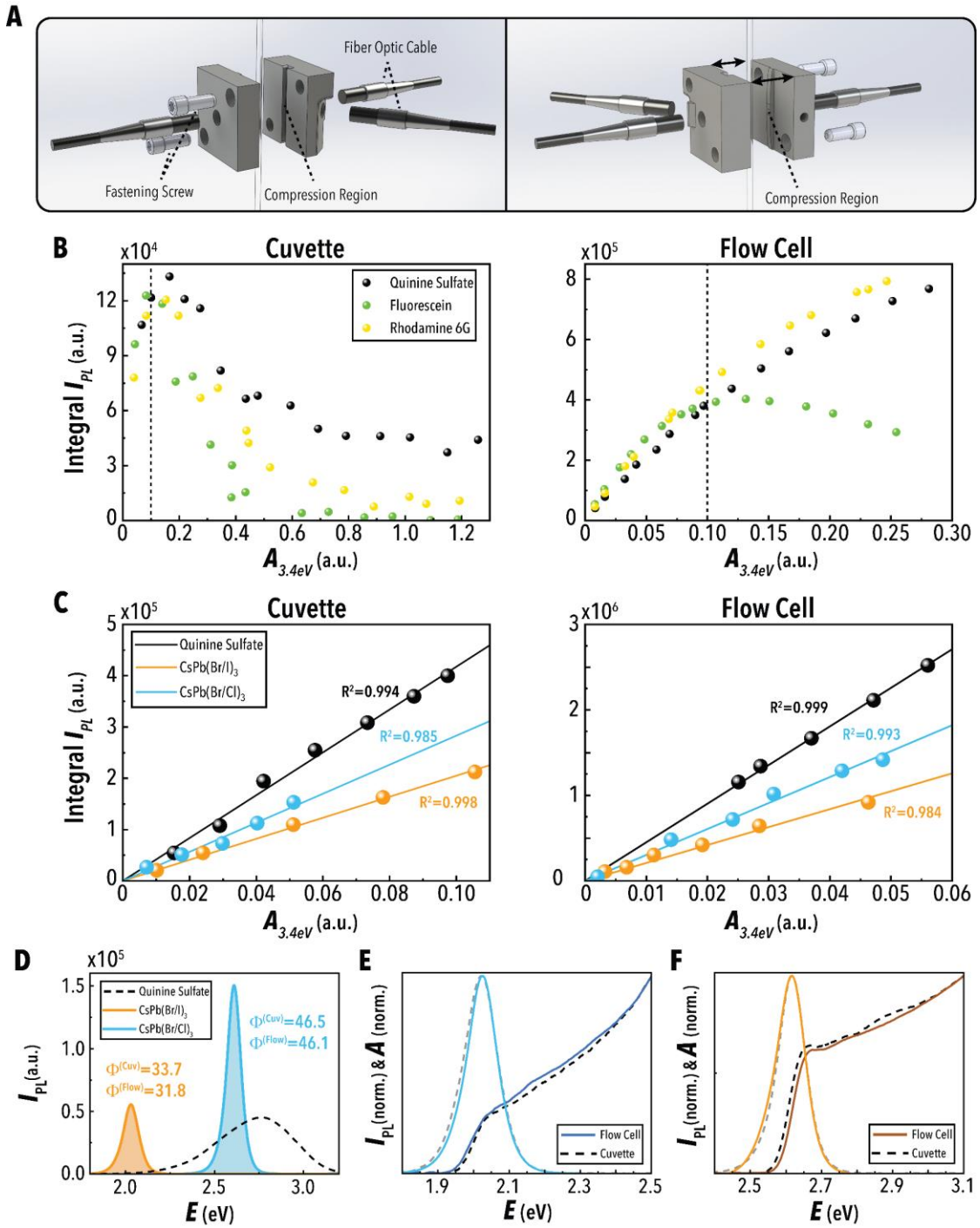

Figure S1. (A) Illustrations of compressed tube flow cell through an exploded view. (B) Integral photoluminescence as a function of absorption at 3.4 eV for three fluorescent dyes taken at equivalent concentrations between a 1x10 mm cuvette and the custom flow cell and (C) for quinine sulfate and two mixed halide perovskite solutions between a 10x10 mm cuvette and the custom flow cell with (D) corresponding photoluminescence spectra and quantum yield measurements with normalized absorption and photoluminescence spectra for (E) $CsPb(Br/Cl)_3$ and (F) $CsPb(Br/I)_3$ perovskite solutions with measurements from both the flow cell and cuvette.

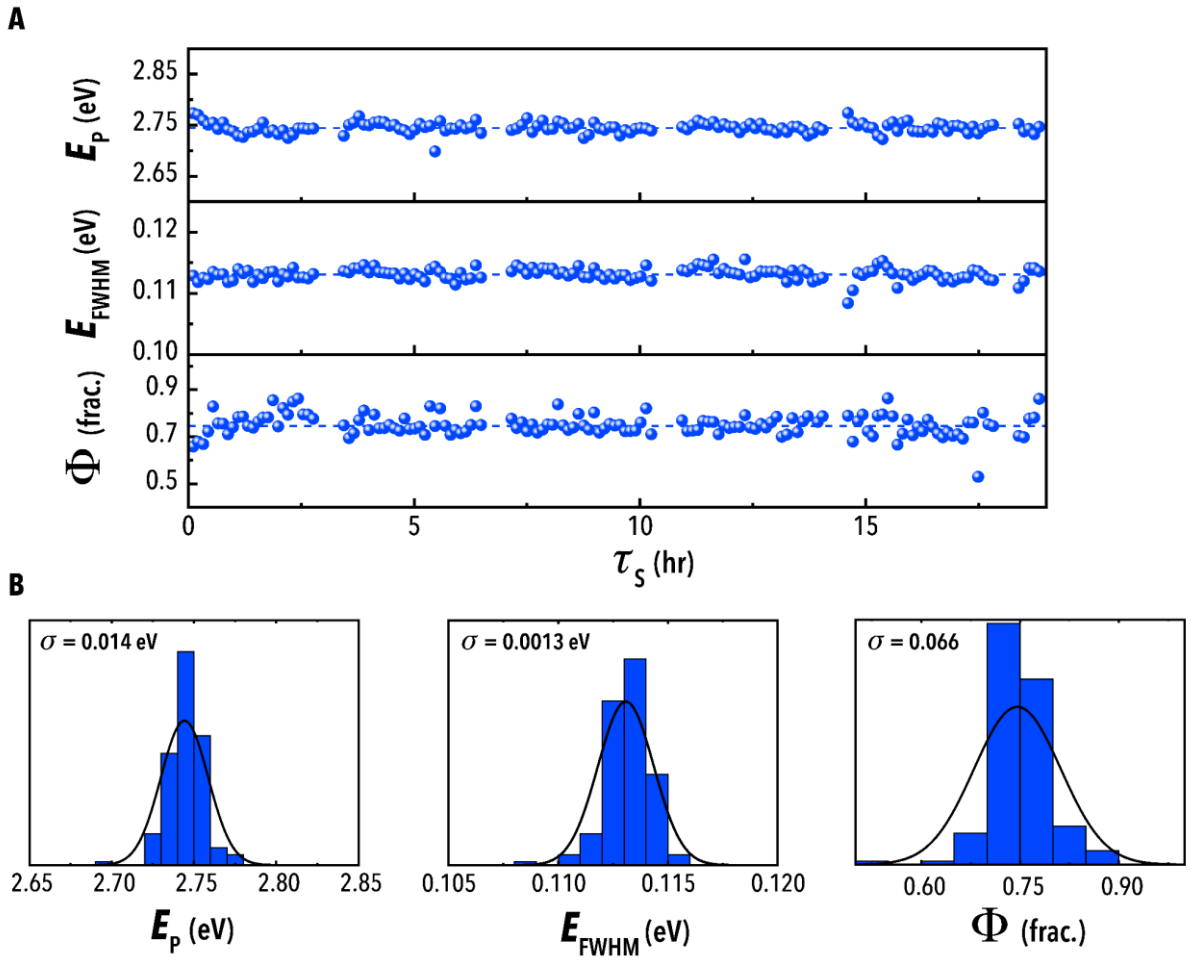

Figure S2. (A) Peak emission energy (E_p), emission full-width at half-maximum (E_{FWHM}), and photoluminescence quantum yield (ϕ) for continuous sampling in the *Artificial Chemist* of constant flow conditions ($X_{SQD}, X_{ZnX_2}, X_{ZnBr_2}, X_{OA}, X_{OLA} = 0.5$) and the average measured value as indicated by the dashed line with (B) histograms of corresponding data.

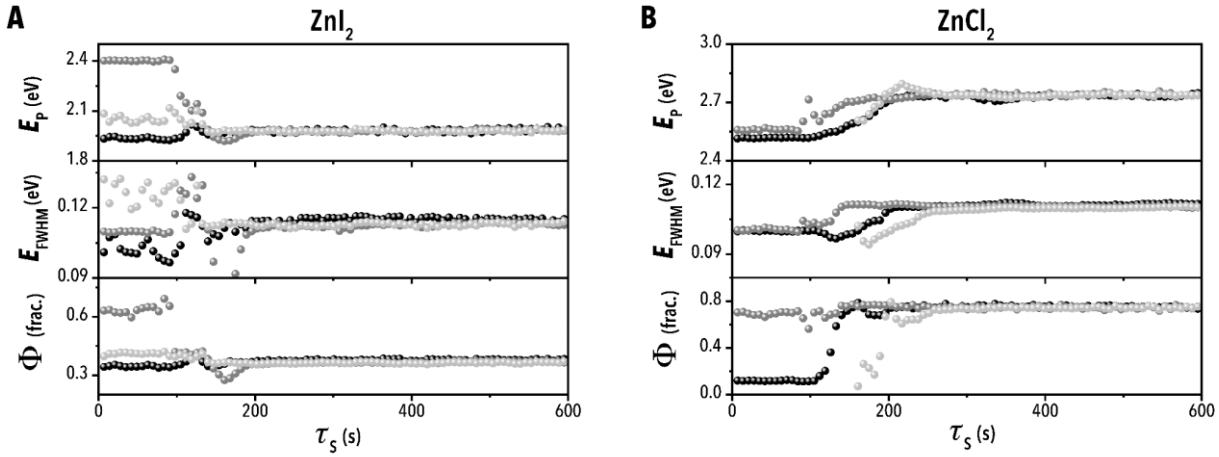

Figure S3. Peak emission energy (E_p), emission full-width at half-maximum (E_{FWHM}), and photoluminescence quantum yield (ϕ) sampled continuously for entirety of flow stabilization in the Artificial chemist when moving from three randomly selected experimental conditions to constant flow conditions, X_{SQD}, X_{ZnX_2}, X_{ZnBr_2}, X_{OA}, $X_{OLA} = 0.5$, for (A) ZnI_2 and (B) $ZnCl_2$ as the exchanging halide sources. All flow rates except for the starting quantum dot stream were set at time zero. The starting quantum dot solution was stopped at time zero then set after 100 s.

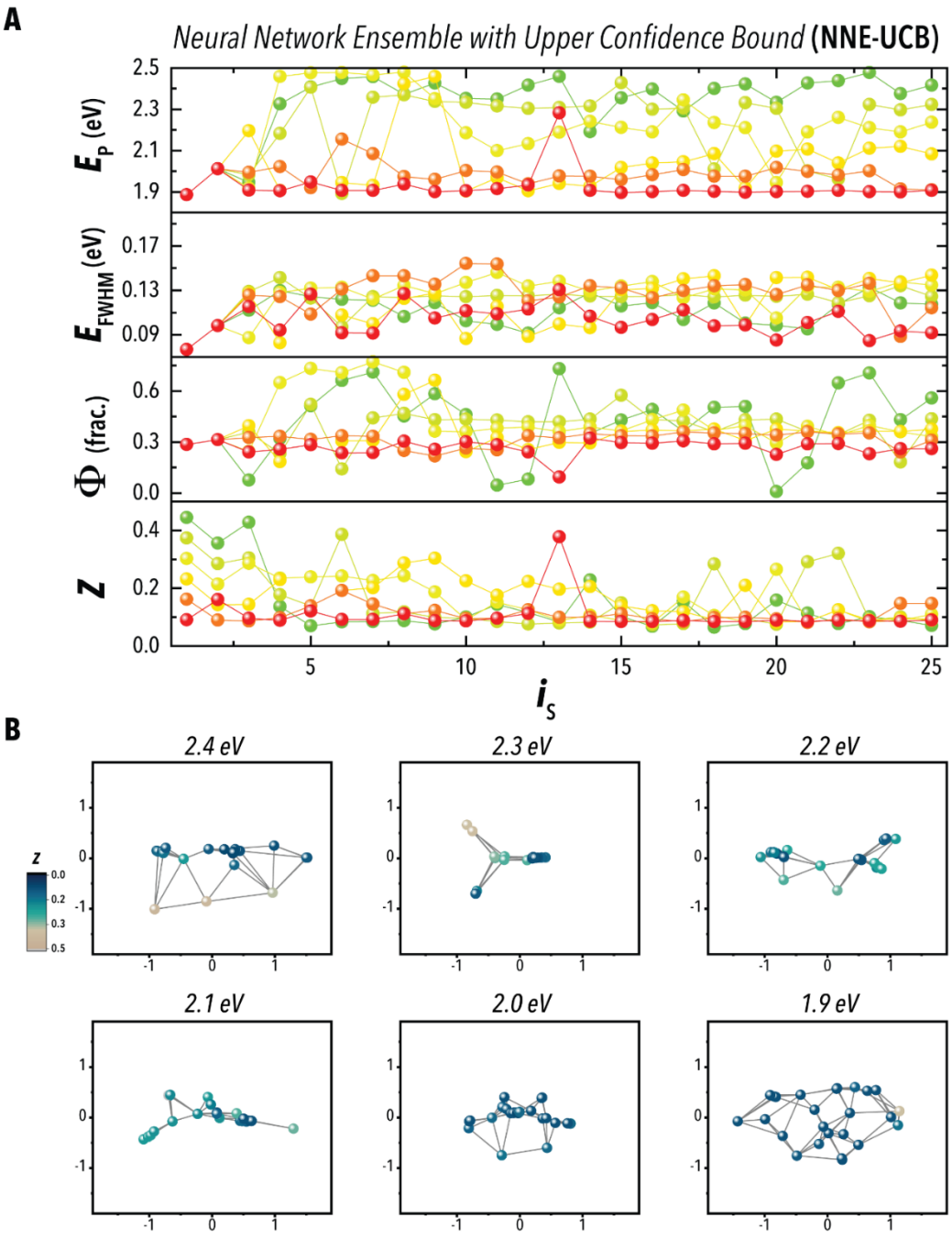

Figure S4A. (A) Measured peak emission energy (E_p), emission full-width at half-maximum (E_{FWHM}), photoluminescence quantum yield (ϕ), and objective function (Z) values for 25 samples selected using the neural network ensemble with the upper confidence bound decision policy for six different target emissions (● 2.4 eV, ● 2.3 eV, ● 2.2 eV, ● 2.1 eV, ● 2.0 eV, and ● 1.9 eV) with ZnI_2 as the exchanging halide source and no prior knowledge. (B) Isomap representations of the corresponding input conditions selected by the method with measured Z. Isomaps were formed in Euclidian space on the non-dimensionalized input variables with the four nearest neighbors.

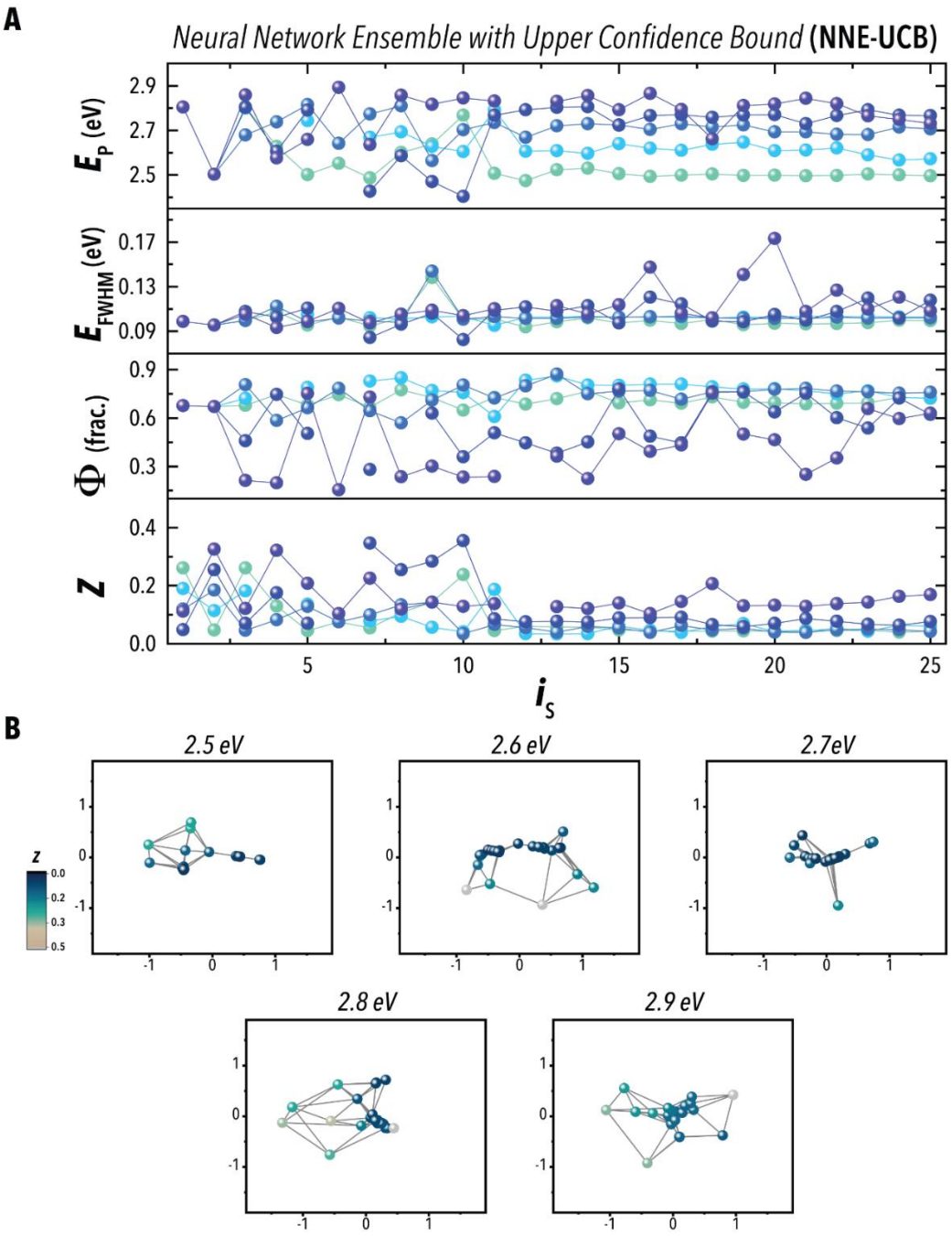

Figure S4B. (A) Measured peak emission energy, emission full-width at half-maximum, photoluminescence quantum yield, and objective function values for 25 samples selected using the neural network ensemble with the upper confidence bound decision policy for five different target emissions (● 2.9 eV, ● 2.8 eV, ● 2.7 eV, ● 2.6 eV, and ● 2.5 eV) with $ZnCl_2$ as the exchanging halide source and no prior knowledge. (B) Isomap representations of the corresponding input conditions selected by the method with measured *Z*. Isomaps were formed in Euclidian space on the non-dimensionalized input variables with the four nearest neighbors.

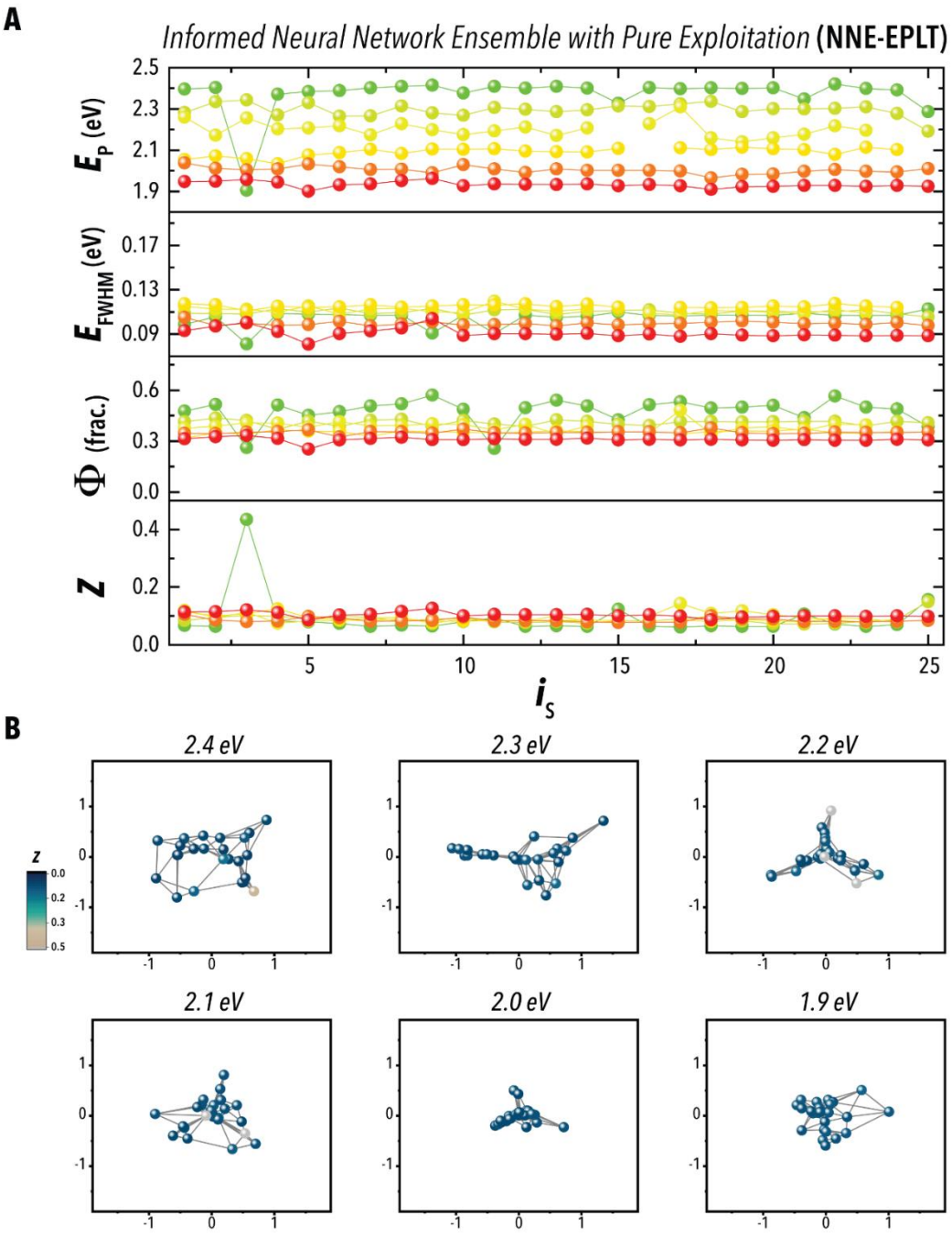

Figure S4C. (A) Measured peak emission energy, emission full-width at half-maximum, photoluminescence quantum yield, and objective function values for 25 samples selected using the informed (*i.e.*, with prior knowledge) neural network ensemble with the pure exploitation decision policy for six different target emissions (● 2.4 eV, ● 2.3 eV, ● 2.2 eV, ● 2.1 eV, ● 2.0 eV, and ● 1.9 eV) with ZnI_2 as the exchanging halide source and prior knowledge transfer. (B) Isomap representations of the corresponding input conditions selected by the method with measured *Z*. Isomaps were formed in Euclidian space on the non-dimensionalized input variables with the four nearest neighbors.

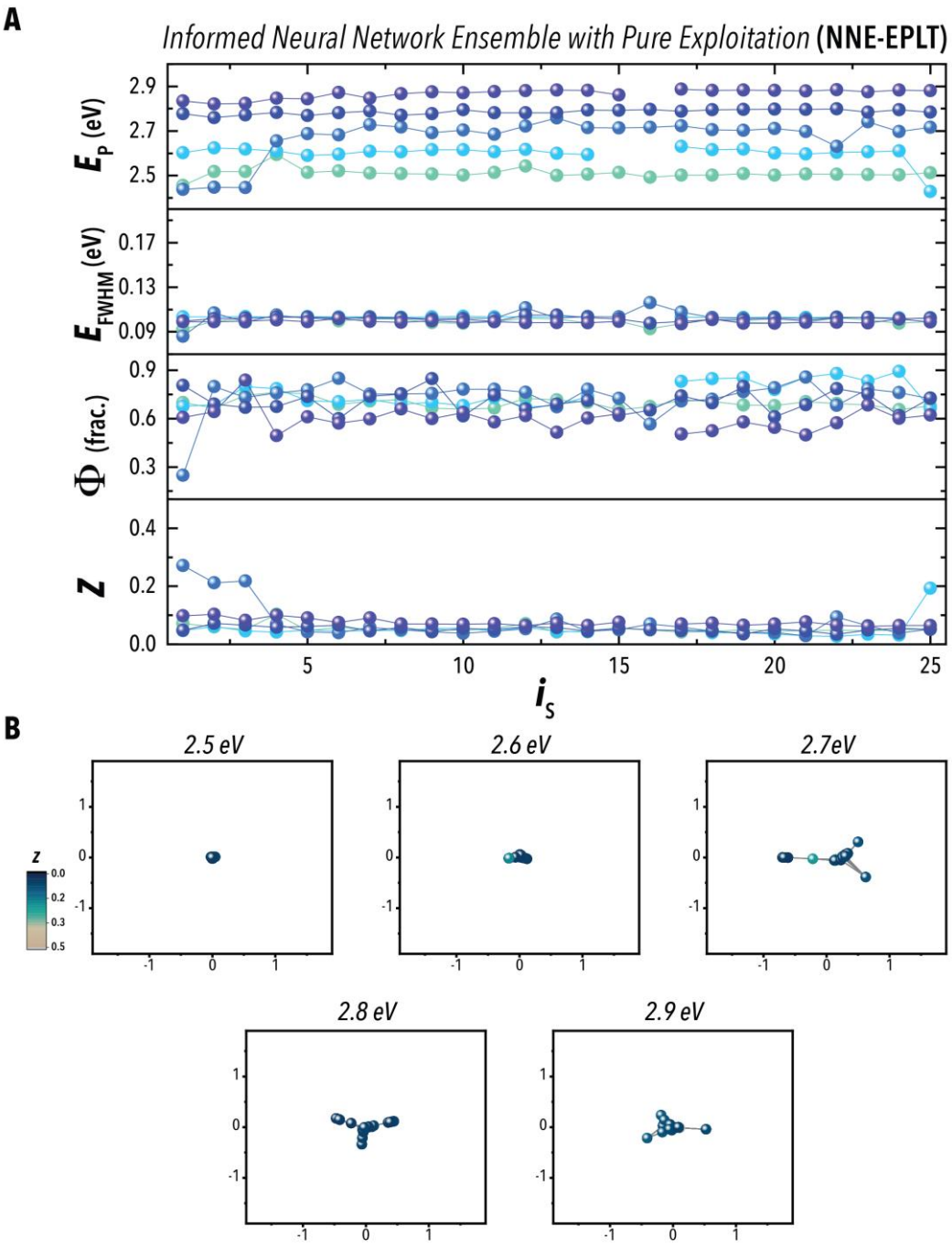

Figure S4D. (A) Measured peak emission energy, emission full-width at half-maximum, photoluminescence quantum yield, and objective function values for 25 samples selected using the informed (*i.e.*, with prior knowledge) neural network ensemble with the pure exploitation decision policy for five different target emissions (● 2.9 eV, ● 2.8 eV, ● 2.7 eV, ● 2.6 eV, and ● 2.5 eV) with $ZnCl_2$ as the exchanging halide source and prior knowledge transfer. (B) Isomap representations of the corresponding input conditions selected by the method with measured *Z*. Isomaps were formed in Euclidian space on the non-dimensionalized input variables with the four nearest neighbors.

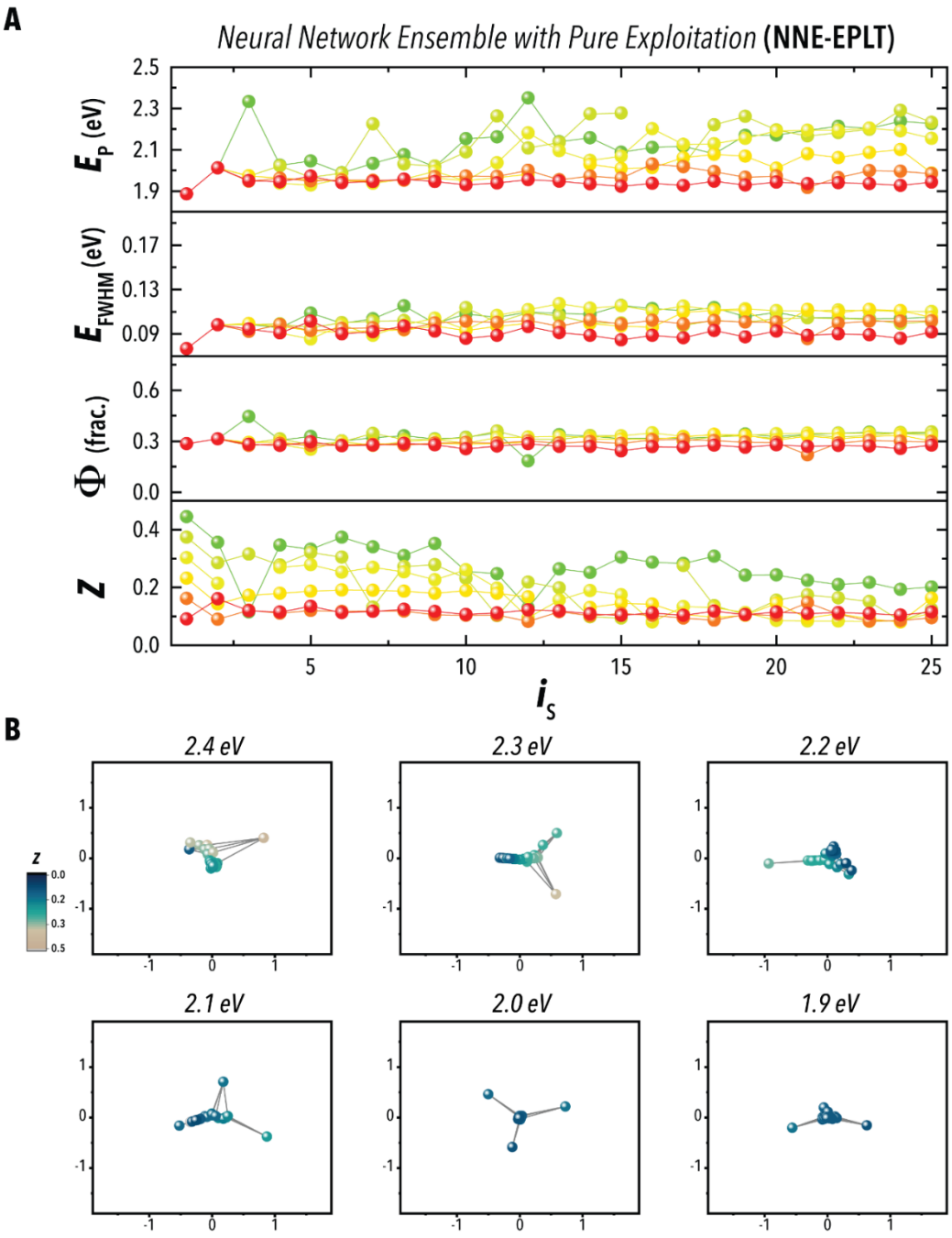

Figure S4E. (A) Measured peak emission energy, emission full-width at half-maximum, photoluminescence quantum yield, and objective function values for 25 samples selected using the neural network ensemble with the pure exploitation decision policy for six different target emissions (● 2.4 eV, ● 2.3 eV, ● 2.2 eV, ● 2.1 eV, ● 2.0 eV, and ● 1.9 eV) with ZnI_2 as the exchanging halide source and no prior knowledge. (B) Isomap representations of the corresponding input conditions selected by the method with measured *Z*. Isomaps were formed in Euclidian space on the non-dimensionalized input variables with the four nearest neighbors.

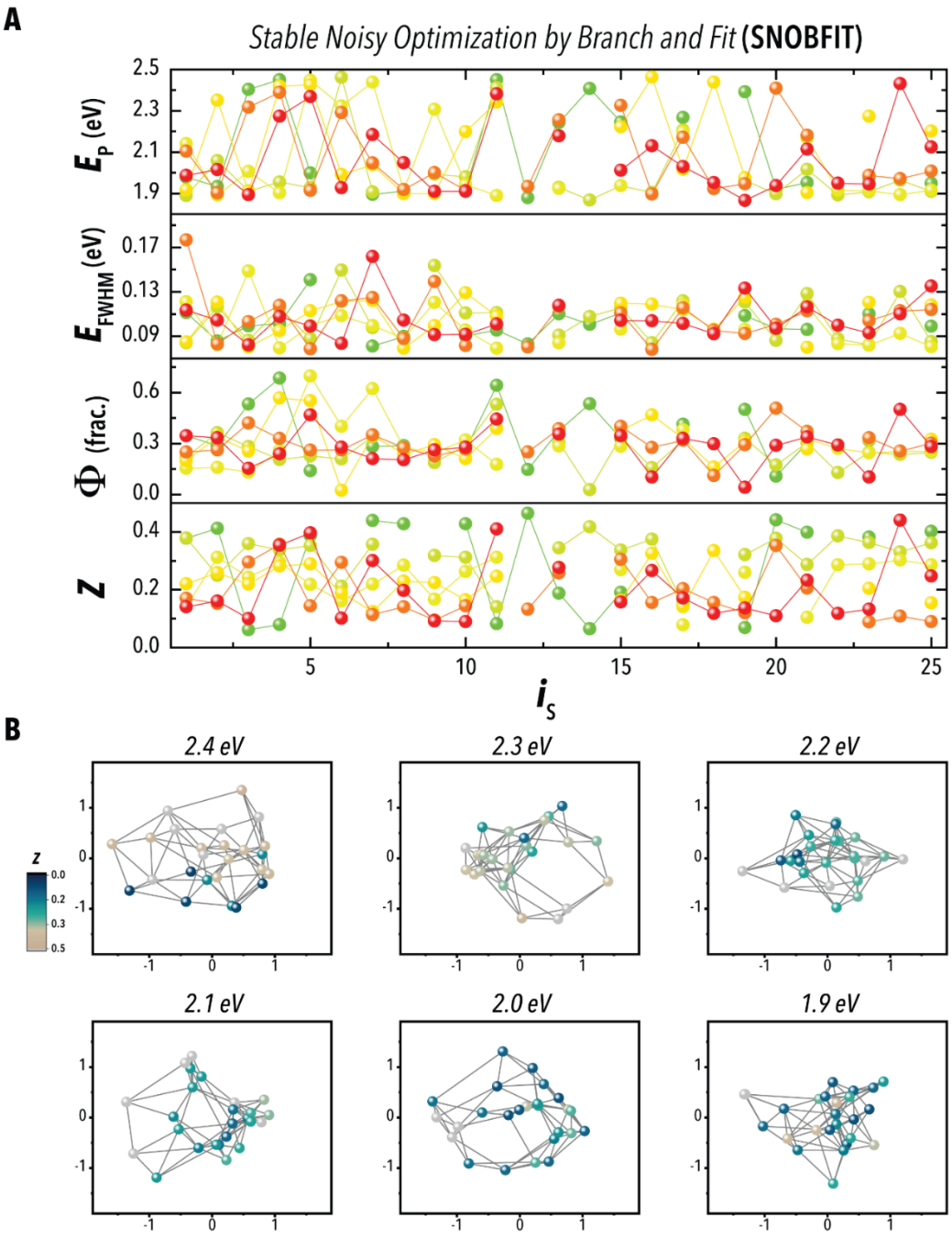

Figure S4F. (A) Measured peak emission energy, emission full-width at half-maximum, photoluminescence quantum yield, and objective function values for 25 samples selected using Stable Noisy Optimization by Branch and Fit for six different target emissions (● 2.4 eV, ● 2.3 eV, ● 2.2 eV, ● 2.1 eV, ● 2.0 eV, and ● 1.9 eV) with ZnI_2 as the exchanging halide source and no prior knowledge. (B) Isomap representations of the corresponding input conditions selected by the method with measured *Z*. Isomaps were formed in Euclidian space on the non-dimensionalized input variables with the four nearest neighbors.

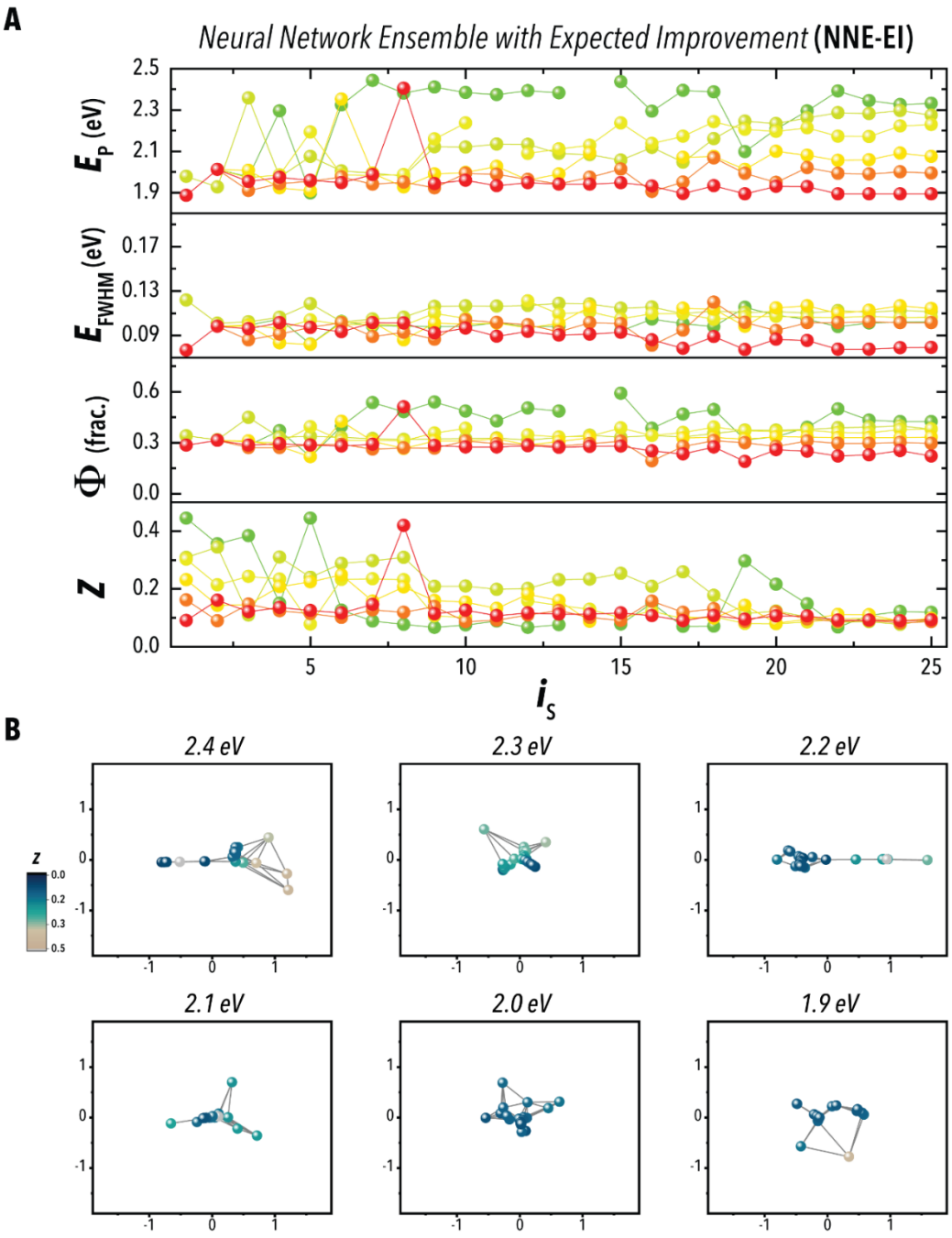

Figure S4G. (A) Measured peak emission energy, emission full-width at half-maximum, photoluminescence quantum yield, and objective function values for 25 samples selected using the neural network ensemble with the expected improvement decision policy for six different target emissions (● 2.4 eV, ● 2.3 eV, ● 2.2 eV, ● 2.1 eV, ● 2.0 eV, and ● 1.9 eV) with ZnI_2 as the exchanging halide source and prior knowledge transfer. (B) Isomap representations of the corresponding input conditions selected by the method with measured *Z*. Isomaps were formed in Euclidian space on the non-dimensionalized input variables with the four nearest neighbors.

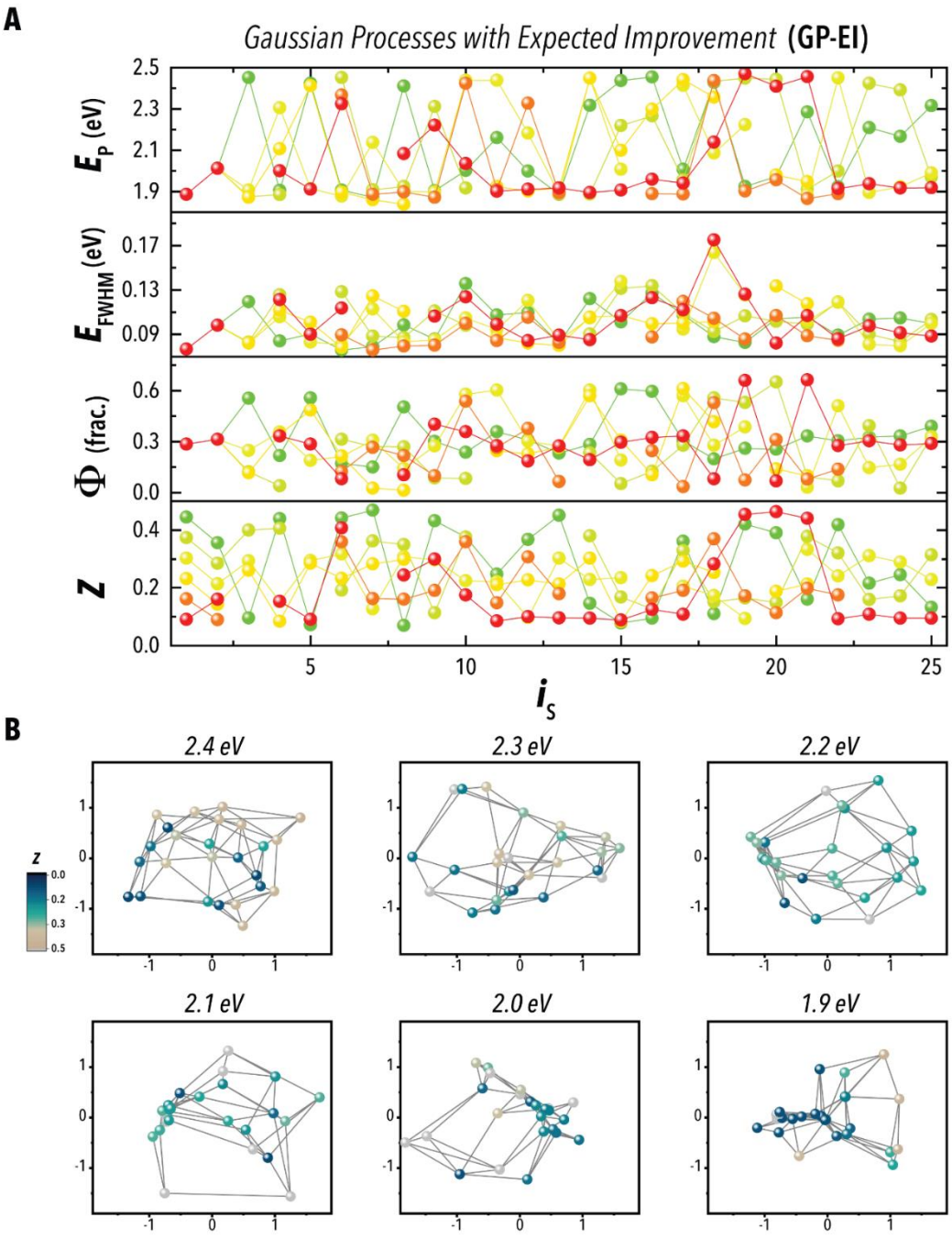

Figure S4H. (A) Measured peak emission energy, emission full-width at half-maximum, photoluminescence quantum yield, and objective function values for 25 samples selected using Gaussian process regressions with the expected improvement decision policy for six different target emissions (● 2.4 eV, ● 2.3 eV, ● 2.2 eV, ● 2.1 eV, ● 2.0 eV, and ● 1.9 eV) with ZnI_2 as the exchanging halide source and prior knowledge transfer. (B) Isomap representations of the corresponding input conditions selected by the method with measured *Z*. Isomaps were formed in Euclidian space on the non-dimensionalized input variables with the four nearest neighbors.

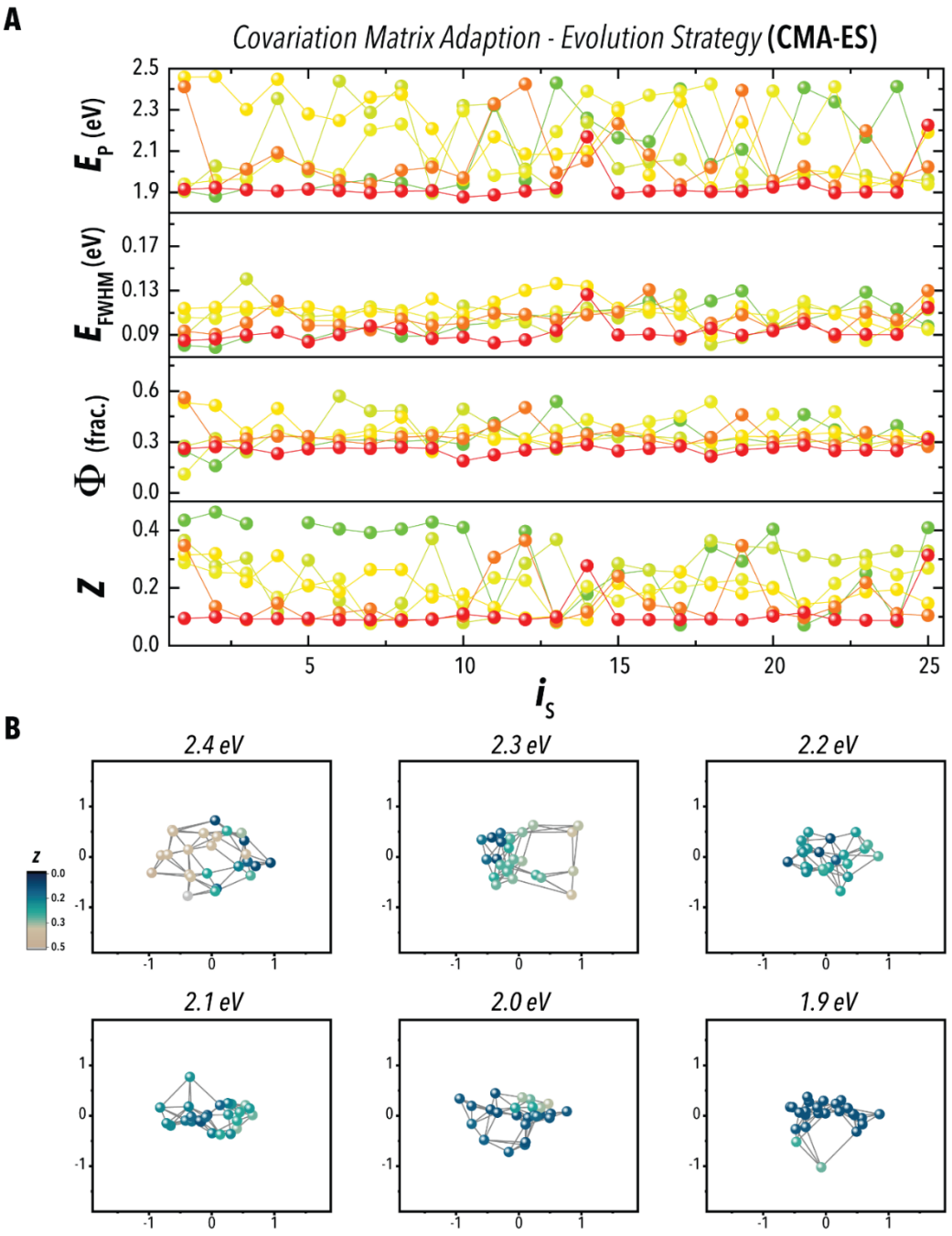

Figure S4I. (A) Measured peak emission energy, emission full-width at half-maximum, photoluminescence quantum yield, and objective function values for 25 samples selected using Covariation Matrix Adaption – Evolution Strategy for six different target emissions (● 2.4 eV, ● 2.3 eV, ● 2.2 eV, ● 2.1 eV, ● 2.0 eV, and ● 1.9 eV) with ZnI_2 as the exchanging halide source and prior knowledge transfer. (B) Isomap representations of the corresponding input conditions selected by the method with measured *Z*. Isomaps were formed in Euclidian space on the non-dimensionalized input variables with the four nearest neighbors. A population size of 10 was used.

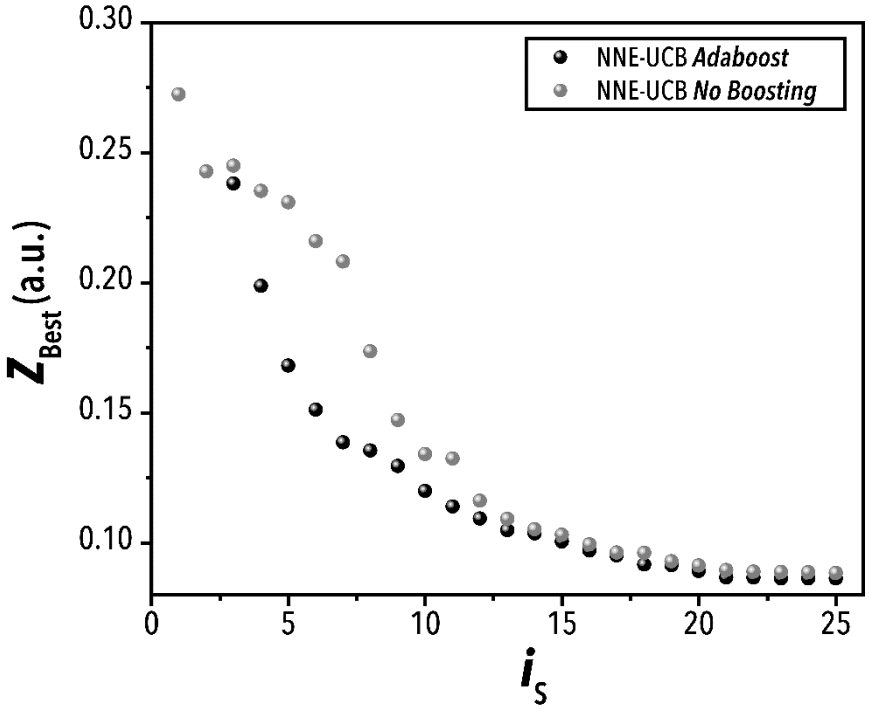

Figure S5. Mean of best three measured *Z* values as a function of sample number averaged over six different target emissions selected by the uninformed neural network ensemble with the upper confidence bound decision policy with and without the use of *Adaboost*.

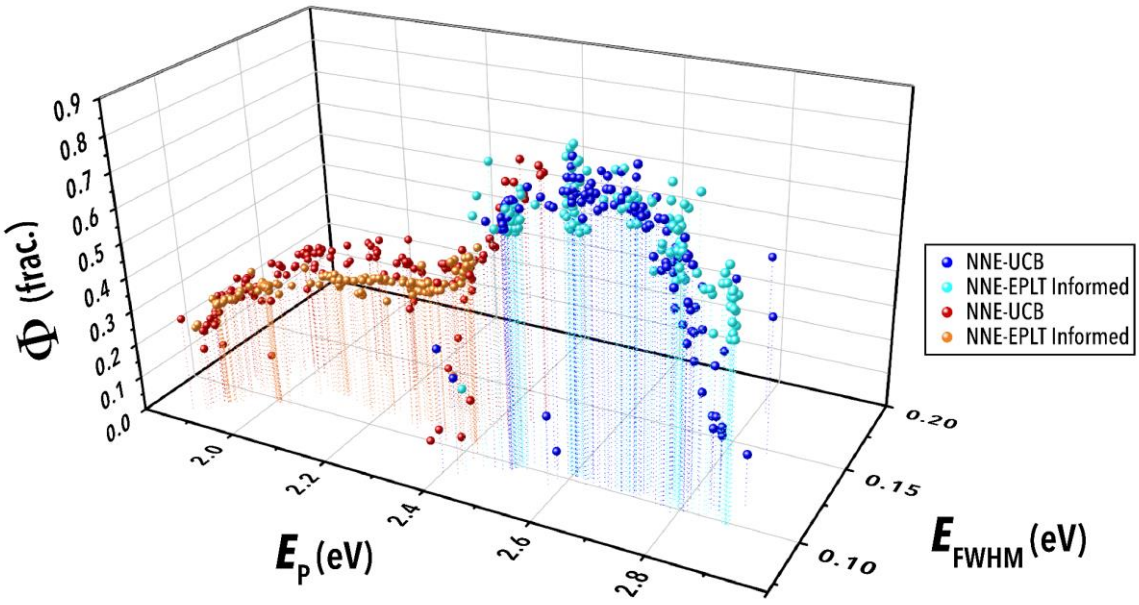

Figure S6. Measured peak emission energy, emission full-width at half-maximum, and photoluminescence quantum yield values from the highest performing experiment selection methods tested, the uninformed neural network ensemble with the upper confidence bound policy (● ZnI_2 and ● $ZnCl_2$) and the informed neural network ensemble with the pure exploitation policy (● ZnI_2 and ● $ZnCl_2$).

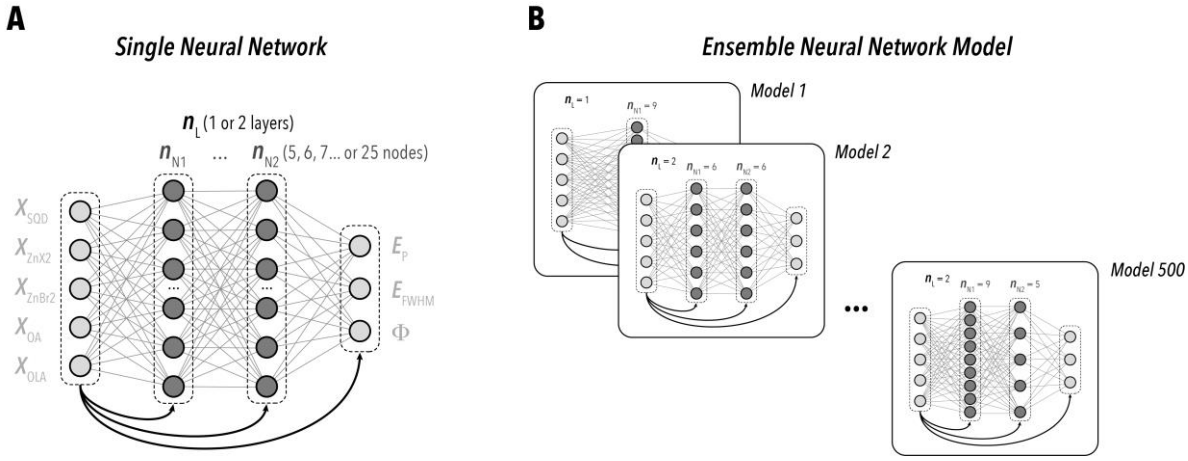

Figure S7. Illustrated diagrams of (A) a single neural network architecture with a randomize number of layers and nodes and (B) the ensemble neural network model built from the collection of 500 neural networks of randomized architecture.

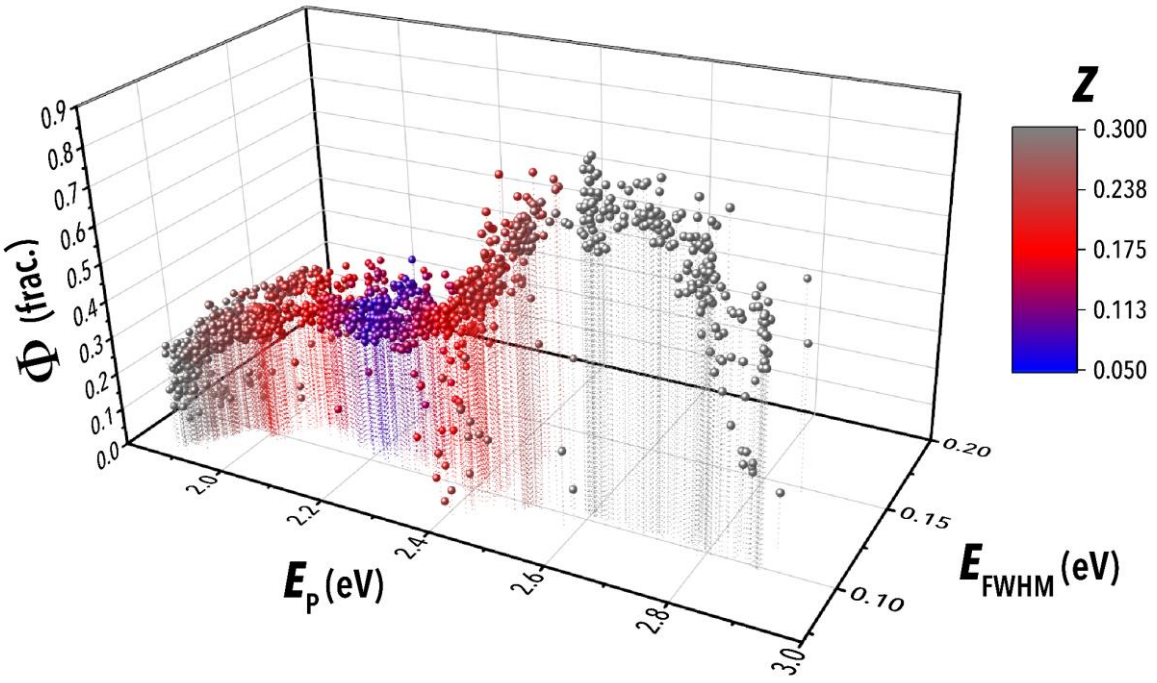

Figure S8. Measured peak emission energy, emission full-width at half-maximum, and photoluminescence quantum yield values from all experiments conducted with the objective function value, *Z*, colormap for a target emission of 2.2 eV. The objective function effectively isolates the target emission while simultaneously search for the optimal full-width at half-maximum and photoluminescence quantum yield.

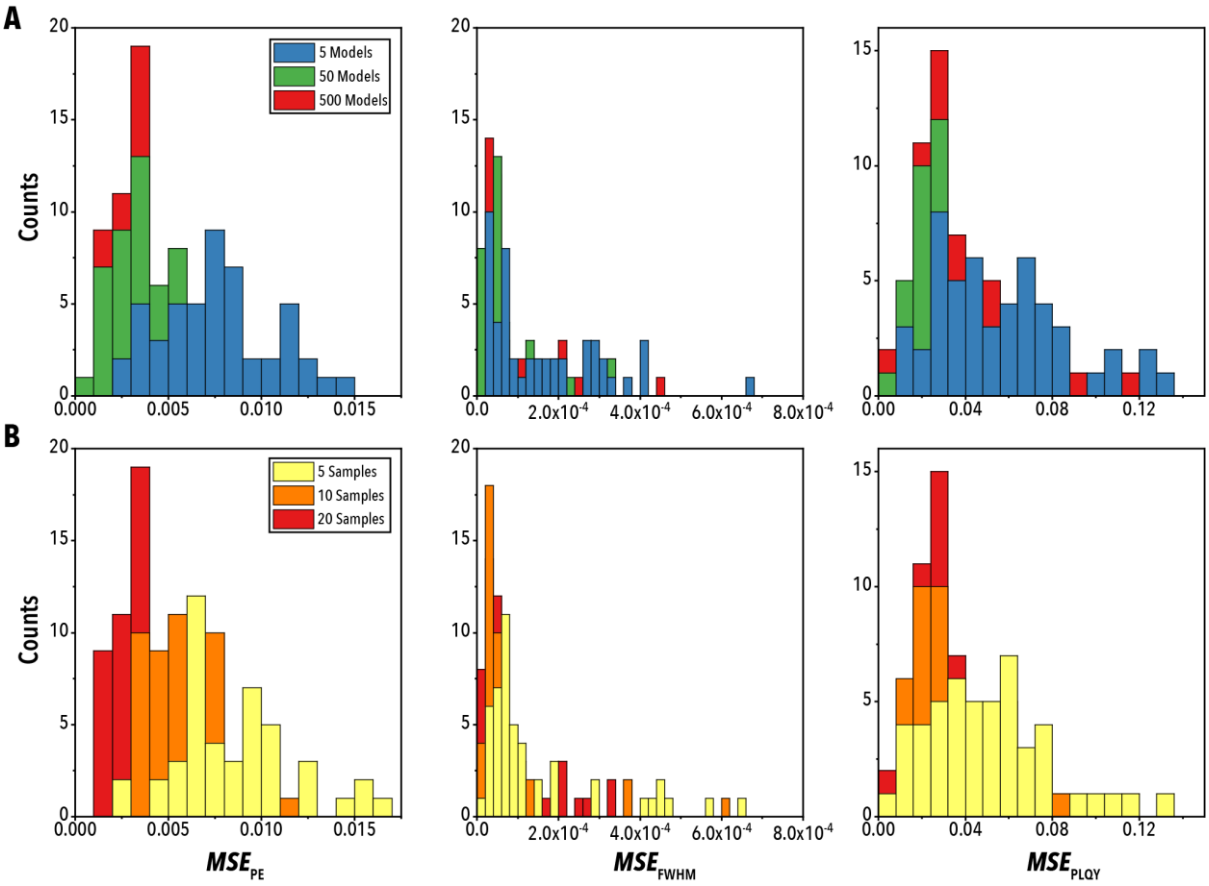

Figure S9. Mean squared error (MSE) of simulated neural network model predictions using 200 randomly selected and sampled experimental conditions when compared to the measured values for peak emission energy (PE), emission full-width at half-maximum (FWHM), and photoluminescence quantum yield (PLQY) (A) as a function of ensemble size for a model trained on 20 experimental conditions and (B) as a function of training set size for an ensemble of 500 models. Histograms were collected by calculating the MSE of 20 randomly selected measured values relative to the model predictions for 50 replicates of each set of model parameters.

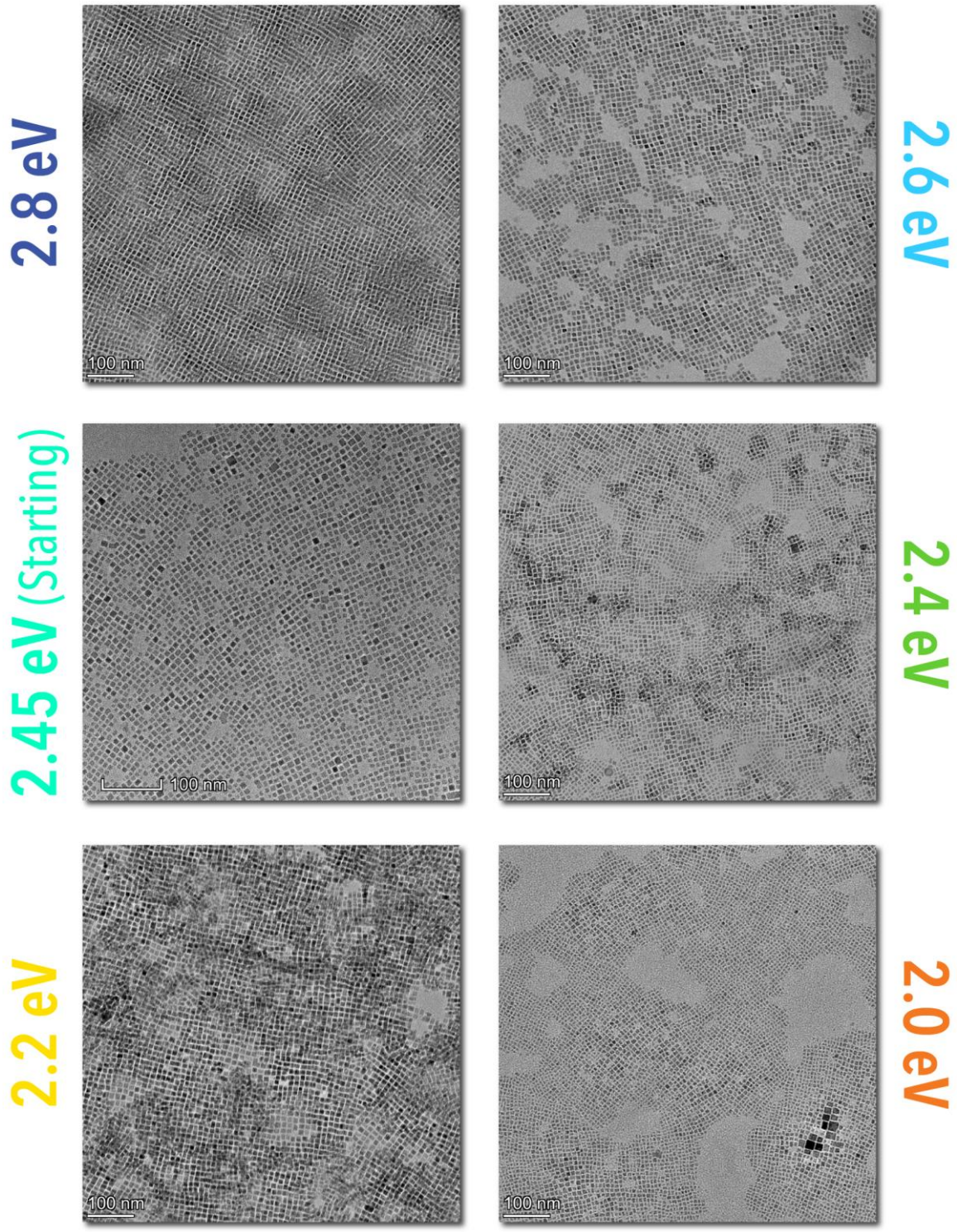

Figure S10. Transmission electron microscopy (TEM) images of the starting $CsPbBr_3$ quantum dots and the products of five selected target emissions synthesized by the *Artificial Chemist*.

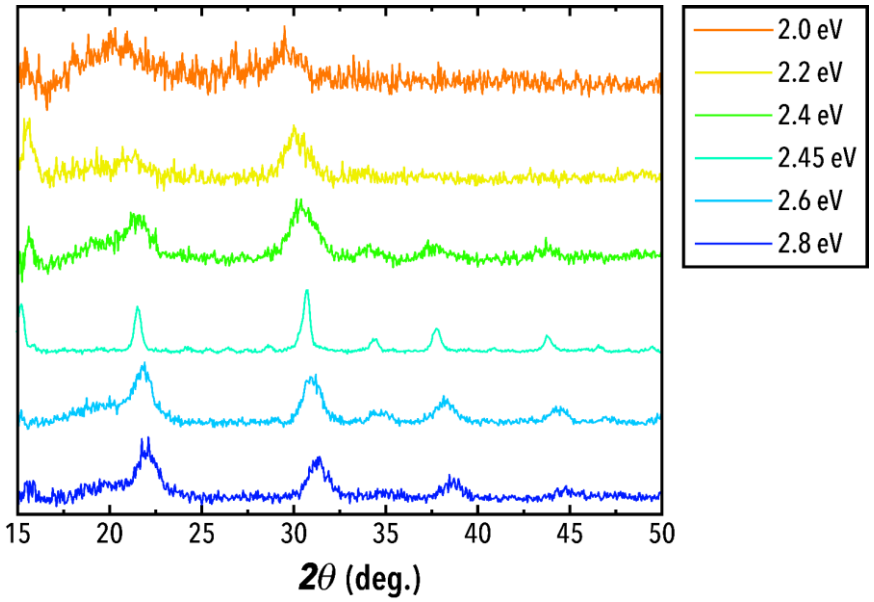

Figure S11. Grazing incidence x-ray diffraction patterns for the starting $CsPbBr_3$ quantum dots and the products of five selected target emissions synthesized by the *Artificial Chemist*.