Diffusion model

The diffusion model is used to determine the concentration profile of the mother liquor when perovskite precipitate is observed. This is calculated in four main steps. The first step is to create a mathematical model of one-dimensional diffusion. This is necessary to account for the species diffusion that results in precipitation, and to quantify species concentration at any given point in time and position. The second step is to calculate the initial input parameters for the model from the experimental data. The third step is to optimize the remaining free parameters to fit the diffusion model to the experimental data. With a satisfactory fit, the model reflects the experimental data. Finally, once all parameters have been fit, the model can then be queried at the time and position of crystal formation to determine the concentration conditions necessary for crystallization.

Mathematical Model of Antisolvent Diffusion in One Dimension

The following model approximates antisolvent concentration in a vial as function of position and time. This model calculates diffusion in one dimension using a finite volume approximation of Fick's second law. This model operates under several assumptions of antisolvent-solvent mixture: (1) the antisolvent is constantly evaporating and condensing from the vial; (2) the antisolvent and solvent are an ideal mixture, such that the non-ideal volume of mixing is zero; (3) any changes in volume result from the same volume of antisolvent diffusing into the vial; (4) no interaction between the molecules of the solvent and antisolvent impede or accelerate the rates of diffusion as described by Fick's first law[3].

The first step in creating this model is to discretize the space in the vial into a sequence of bins. Each bin, *i*, has a defined height, *h*, and starting species concentration, *C*. Diffusion from bin to bin is calculated based on the distances from the centers of each bin. For each bin, the change in concentration of antisolvent with respect to time is given by:

$$\frac{\delta C}{\delta t} = Flux_{in} + Flux_{out} \tag{1}$$

$$= D\frac{\delta C}{\delta h}\Big|_h - D\frac{\delta C}{\delta x}\Big|_{h+\Delta h} \tag{2}$$

$$\approx \frac{C_i^{t+1} - C_i^t}{\Delta t}$$

$$= \frac{1}{\Delta h}\left(D\frac{C_{i+1}^t - C_i^t}{\Delta h} - D\frac{C_i^t - C_{i-1}^t}{\Delta h}\right) \tag{3}$$

This discretized formula calculates the rate at which the concentration in each bin changes in a single time step, and can be numerically integrated to determine the concentration in each bin as a function of time.

This baseline, finite volume approximation for one dimensional diffusion is modified to reflect liquid volume build-up. The model assumes the solution behaves as an ideal mixture and the bin height increases linearly with species concentration. Bin height is written as a function of molar concentration, molar volume, and a height to volume scaling constant. The height to volume scaling factor is determined experimentally and converts from the recorded liquid height to volume. It is abbreviated as *R* and is in units of cm/microliter.

$$h_i^t = n_a\underline{V_a}R + n_s\underline{V_s}R + n_o\underline{V_o}R \tag{4}$$

This relationship connects the height of each bin to the moles of species present within it. **Figure S37.** provides a visual description of the relationship between concentration and bin height.

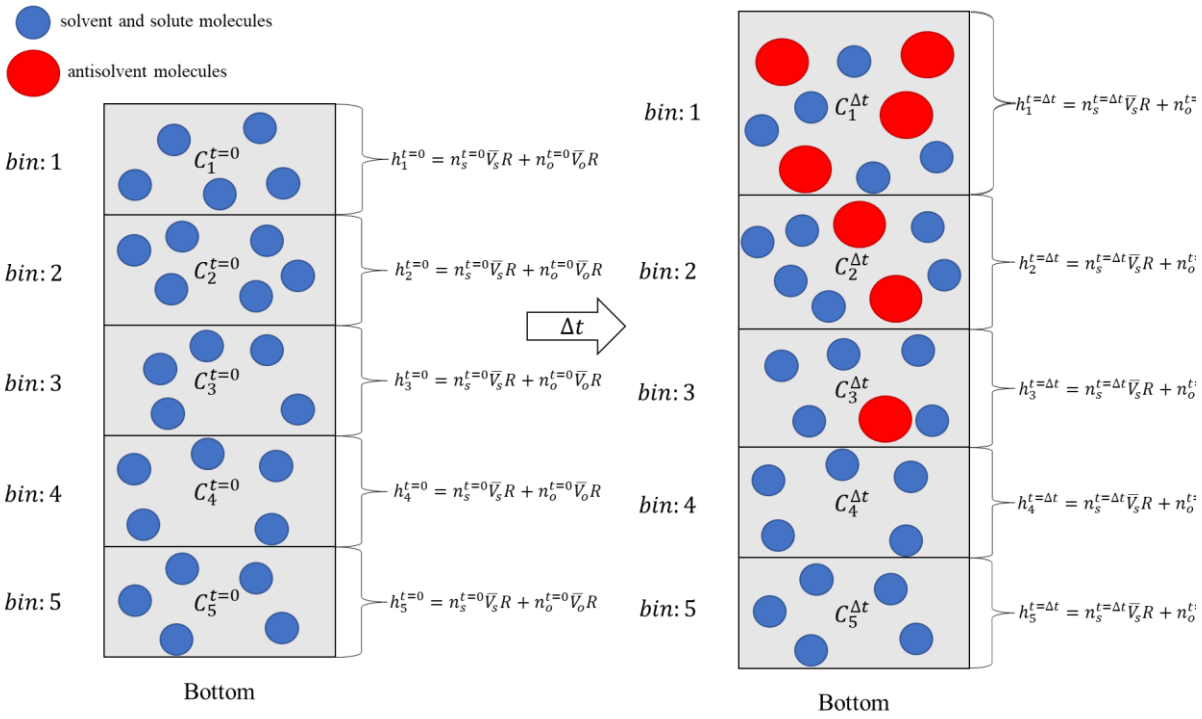

Figure S37. A diagram indicating how the influx of antisolvent molecules changes the height of the bins

With dynamic bin heights, equation (2), is rewritten, where the constant distance between the center of two bins, Δh, is replaced by the dynamic distance between the bins, h_i.

$$\frac{C_i^{t+1}-C_i^t}{\Delta t} = \frac{1}{h_i}\left(D\frac{C_{i+1}^t-C_i^t}{\frac{h_{i+1}}{2}+\frac{h_i}{2}} - D\frac{C_i^t-C_{i-1}^t}{\frac{h_i}{2}+\frac{h_{i-1}}{2}}\right) \quad (5)$$

Equation (5) holds for the middle bins, where there is movement of each species into two adjacent bins. For the bottom bin, there is simply no diffusion into a lower bin.

$$\frac{C_i^{t+1}-C_i^t}{\Delta t} = \frac{1}{h_i}\left(D\frac{C_{i+1}^t-C_i^t}{\frac{h_{i+1}}{2}+\frac{h_i}{2}}\right) \quad (6)$$

The top bin has a few more factors affecting the change in concentration of the antisolvent. The bin is in contact with the air above the vial, from which antisolvent is condensing into, and evaporating from. We assume a constant influx rate, k_{in}, at which molecules enter the top vial; this is reasonable because the vapor pressure of antisolvent above the vial is constant throughout the experiment, so long as any liquid antisolvent remains. In principle, k_{in} could be related to the impingement rate on the surface implied by the kinetic theory of gases and the vapor pressure of the antisolvent, but we treat it as a fitting constant. The evaporation of antisolvent from the top layer is assumed to follow Raoult's Law, [4] such that the total rate is the product of an influx constant, k_{out} and the mole fraction of antisolvent, X_i. in that top layer. Putting this together, the expression for antisolvent diffusion in the top bin is given by

$$\frac{C_i^{t+1}-C_i^t}{\Delta t} = k_{in} + \frac{1}{h_i}\left(-D\frac{C_i^t-C_{i-1}^t}{\frac{h_i}{2}+\frac{h_{i-1}}{2}}\right) - k_{out}X_i \quad (7)$$

Calculating the molar fraction of antisolvent, X_i, at time *t* requires the calculation of moles of solvent and antisolvent in the bin. Calculating the moles of antisolvent is straightforward:

$$n_a = C_i^t R^{-1} h_i^t \tag{8}$$

The moles of all remaining species in the solution, n_δ, can be calculated with the same method:

$$n_\delta = \sum R^{-1} h_i^t (C_o^t) \tag{9}$$

Therefore, the mole fraction of antisolvent is:

$$X_i = \frac{C_i^t R h_i^t}{C_i^t R^{-1} h_i^t + \sum R^{-1} h_i^t (C_o^t)} \tag{10}$$

The changes of concentration of each bin per time step is used to determine the change in moles. To convert the diffusion to units of moles, the following expression is used, where n_{ai}^t is the number of mols of antisolvent "a" present in bin "i" at time t.

$$\frac{n_{ai}^{t+1} - n_{ai}^t}{\Delta t} = \frac{C_i^{t+1} - C_i^t}{\Delta t} R^{-1} h_i^t \tag{11}$$

By integrating over time, the model can determine the moles of each species in each bin. From the amount of mols in each bin, the height of each bin can be defined with equation (4), and the concentration of antisolvent, C_i^t, in the bin with the following:

$$C_i^t = \frac{n_{ai}^t}{R^{-1} h_i^t} \tag{12}$$

Governed by these formulas, this model can simulate the buildup and diffusion of antisolvent in the mother liquor and quantify the concentration of all species in the solution at varying heights over the duration of the simulation. While this explanation focused on the antisolvent, there are four other chemical species within the solvent that are described by the same formulas. However, there is one notable exception. For the remaining species, we assume that there is no condensation or evaporation. This means that for equation (7), k_{in} and k_{out} are both equal to zero.

Bin Spawning and Merging:

The height of each bin is determined by the moles of each species. To avoid negative moles and heights, the model dynamically merges bins together once they start to reach a minimum height threshold. This prevents bins from completely emptying and serves to ensure that mass is conserved. Once a bin has reached a minimum threshold it is merged with the smallest adjacent bin. The moles of all species in the two bins are added then added together, and the resulting height is recalculated using equation (4).

Bin spawning is used to address unbounded bin growth and provide greater spatial resolution at the top of the column. Unbounded growth occurs if the rate of antisolvent ingress into a bin far outstrips the rate of egress, the bin will grow to a critical point at which the antisolvent is not capable of diffusing out of the bin in the given timestep. Without any diffusion out, the bin grows unbounded until it is undefined. Even without unbounded growth, if the ingress rate of the antisolvent into a bin is consistently greater than the egress, the bin exhibits a slow creep in height. The larger bin height reduces spatial resolution at an area of the column in which the rate of change is greatest. Bin spawning solves both issues and is straightforward to implement. After a bin reaches a maximum threshold in size, it is split into two smaller bins of equal size, and the moles of each species are divided evenly amongst the two new bins. The height is then recalculated using equation (4).

Determining Model Starting Conditions:

The model presented generates a concentration profile over the length of the solution vial given starting values and initial parameters. As seen in equations (4-12), there are several input values necessary to run the finite volume approximation. The model requires the initial concentration values for all species in the mother solution.

An experiment begins with a 180 μL mixture of solvent, inorganic solute, and organic solute. The molarities of the organic and inorganic species are experimentally known. 20 microliters of FAH are added to this mixture. The moles of all the species given the starting 180 μL mixture can be calculated:

$$V_{total} * \frac{n_{inorganic}}{V_{total}} = n_{inorganic} \quad (13)$$

$$V_{total} * \frac{n_{organic}}{V_{total}} = n_{organic} \quad (14)$$

With the moles of both species of solutes, the moles of solvent can be calculated, given the molar mass and densities:

$$V_{solvent} = V_{total} - V_{organic} - V_{inorganic} \quad (15)$$

$$V_{species} = n_{species} * M_{species} * \rho_{species}^{-1} \quad (16)$$

$$V_{solvent} = V_{total} - \left(n_{inorganic} * M_{inorganic} * \rho_{inorganic}^{-1}\right) - \left(n_{organic} * M_{organic} * \rho_{organic}^{-1}\right) \quad (17)$$

$$n_{solvent} = V_{solvent} * \rho_{solvent} * M_{solvent}^{-1} \quad (18)$$

This accounts for the moles of all species in the starting mixture. The final molarities can be calculated after addition of 20 microliters of FAH:

$$[M]_{species} = \frac{n_{species}}{V_{Solvent} + V_{FAH} + V_{inorganic} + V_{organic}} \quad (19)$$

With these calculations, the model has all the starting concentrations for the species in the solution. The initial height of solution, the duration of the experiment, and the dimensions of the solvent column are all given by the experimental data. The diffusion coefficient for each solvent system is determined experimentally. The number of bins and the size of the time-steps are free for the user to vary. If increased spatial resolution is required, the number of bins can be increased. If increased temporal resolution is required, the size of the time steps can be decreased. Note, that the increased resolution comes at the expense of computation time. For our modeling, 20 bins were used, with a one second timestep.

The model requires a height to volume relationship, or *R* value. This allows for the conversion between solution height, and solution volume. This is fundamental to the model, and its use can be seen in equations (4) and (8-12).

Calculating the Height to Volume Relationship:

The height to volume relationship, *R*, is calculated experimentally and determines liquid volume given its height in the solution vial. This scalar constant is calculated by pipetting several aliquots of 100 microliters of solvent into an empty vial, measuring the observed height after each aliquot, and calculating the slope. This method is repeated for each of the solvent systems, and for water. After plotting each 100 microliters aliquot against its measured height, a linear fit is used to calculate the slope. A linear fit of the data, shown in **Figure S40,** indicates that each microliter of liquid increases the height by 0.002966 cm.

The accuracy of this calibration slope, *R*, is pivotal in accurately determining the buildup of antisolvent and the overall volume of the solution. To validate the calibration slope's accuracy, we examine its consistency against the recorded starting experimental values. As every experimental solution is initially 200 microliters and the starting height is reported, we see that the calibration slope is consistent with the recorded heights, with the average difference between the calculated height and the experimental height being 4.24%. The results of this are shown in **Figure S38.**

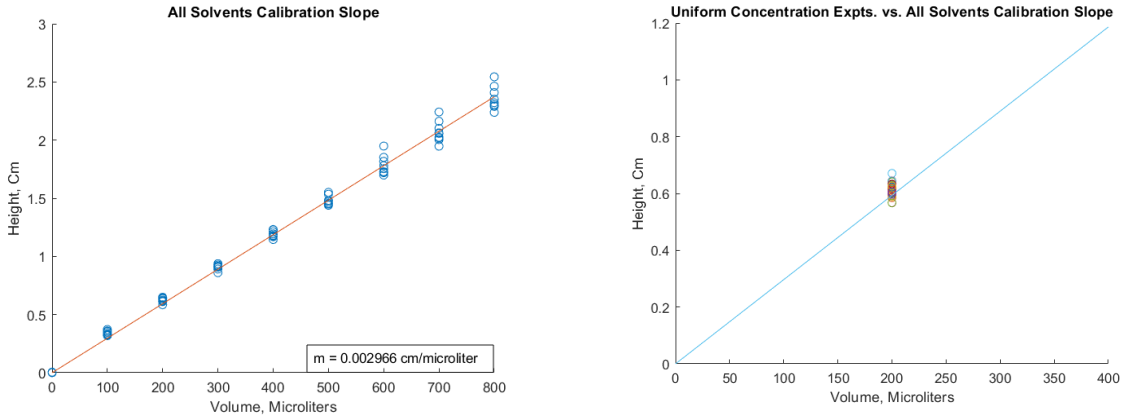

Figure S38. (a) The final *R* value, calculated after proper image calibration. (b) The predicted liquid height of each experiment given the starting volumes. Each colored point indicates an experiment's measured starting height. The distribution of starting heights around the calibration slope in conjunction with their low variance, demonstrates agreement between the reported heights, and the *R* value.

Determining the Diffusion Constant, and Propagation of Uncertainty:

The diffusion coefficient is determined experimentally through a laser refraction experiment. It is possible that inconsistencies in the reference positions of the beam gave rise to observed differences in diffusion rate between neat and mixed solvent systems. As the refraction study relies on the change in height of a diffracted laser beam over a fixed window, any two start and endpoints can be used to calculate the diffusion coefficient. This allows for additional small variations in the resulting diffusion coefficient. This change in height of the refracted beam can be seen in **Figure S39a**.

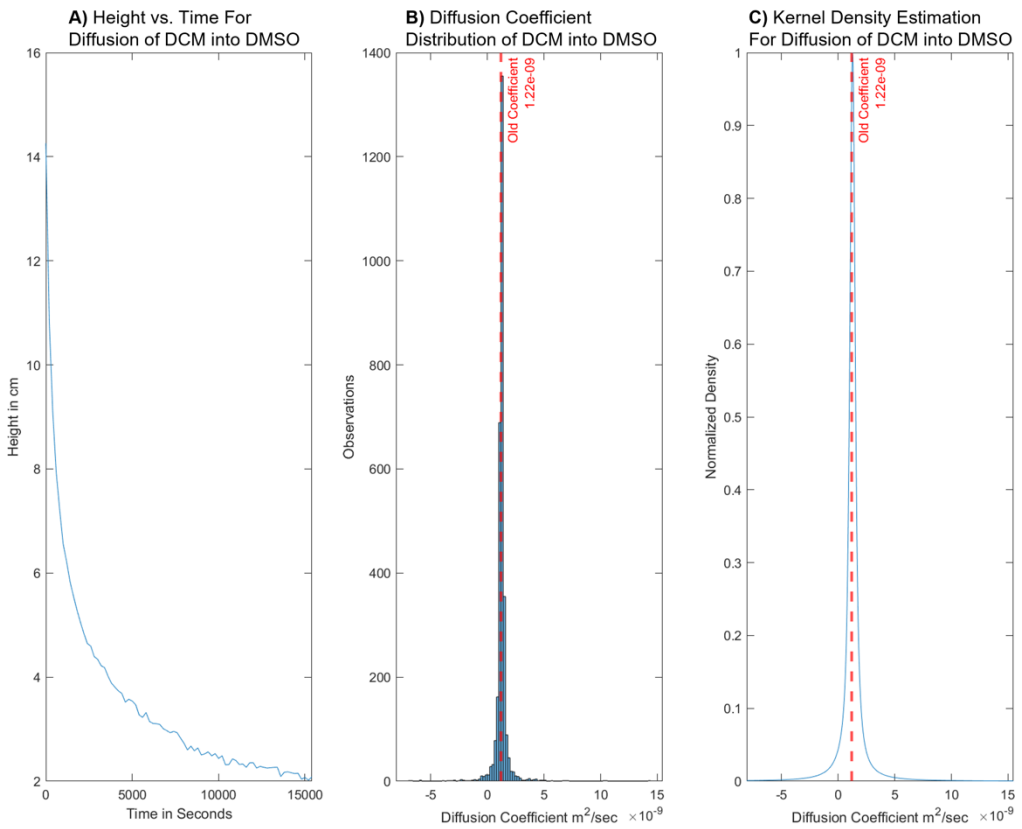

Figure S39. An analysis of uncertainty in the diffusion coefficient calculation for DCM in to DMSO.

To account for these two sources of error, every possible diffusion coefficient in the window of our laser diffraction experiment was calculated. The width of the laser beam in the resulting images of the experiment was used as a starting error, as the height of the refracted beam cannot be measured more accurately than its width. This error is propagated throughout the calculation of every diffusion coefficient to determine the time points that yield a diffusion coefficient with the lowest uncertainty.

The diffusion coefficient with lowest propagated error was used for each of the five solvent systems. To corroborate our findings, we performed further statistical analyses on all the possible diffusion coefficients. The first approach was to calculate all the possible diffusion coefficients given our diffraction experiment window. For this case, we did not propagate out the possible error. Using all possible diffusion coefficients, a histogram was created to get an estimation of the probability function. The center of the bin with the most occurrences is the most likely estimate of the diffusion coefficient. See **Figure S39b**.

Our second approach for determining the most accurate diffusion coefficient was a kernel density estimation, using the propagated uncertainty to determine the width of the gaussian kernel centered about each calculated diffusion coefficient. The maximum of the KDE was assumed to be the most likely diffusion coefficient. This can be seen in **Figure S39c**. All three approaches for calculating the final diffusion coefficient gave similar results. This can be seen in Table S12.

Table S12. Diffusion coefficient analysis data.

Solvent Type	Original Diff. Coeff. (m^2/sec)	Histogram Diff. Coeff. (m^2/sec)	Kernel Density Diff. Coeff. (m^2/sec)	Lowest Uncertainty Diff. Coeff. (m^2/sec)	Lowest Propagated Uncertainty Range (m^2/sec)	Average Propagated Uncertainty Range (m^2/sec)
GBL	5.88×10^{-10}	5.00×10^{-10}	5.83×10^{-10}	5.26×10^{-10}	4.56×10^{-10}	2.88×10^{-9}
GBL:DMF	1.33×10^{-10}	1.25×10^{-10}	1.13×10^{-10}	1.08×10^{-10}	1.04×10^{-10}	6.04×10^{-10}
DMF	8.42×10^{-10}	7.50×10^{-10}	5.87×10^{-10}	5.55×10^{-10}	5.66×10^{-10}	3.34×10^{-9}
DMF:DMSO	2.06×10^{-10}	1.75×10^{-10}	2.05×10^{-10}	2.07×10^{-10}	7.50×10^{-11}	3.12×10^{-10}
DMSO	1.33×10^{-9}	1.10×10^{-9}	1.30×10^{-10}	1.24×10^{-9}	2.79×10^{-10}	1.29×10^{-9}

The diffusion coefficients with the lowest uncertainty would result in a smaller window between the possible lower and upper bounds for the diffusion rates. Because of this, the diffusion coefficients with the lowest associated uncertainty were chosen as the possible interval for the diffusion coefficient.

Model Fitting and Parameter Optimization:

Recall from the explanation of the diffusion model, that k_{in} is the rate of antisolvent condensation into the solvent solution. The k_{out} value, along with the antisolvent molar fraction, determines the rate of evaporation. These two values result in the build-up of antisolvent in the solution column. The antisolvent buildup of the modeled system and the resulting concentration profiles can be changed by tuning these two parameters. To find the optimal k_{in} and k_{out} values for each experiment, the modeled height increase of the solution must match the experimental increase. An example of this data can be seen in **Figure S40.**

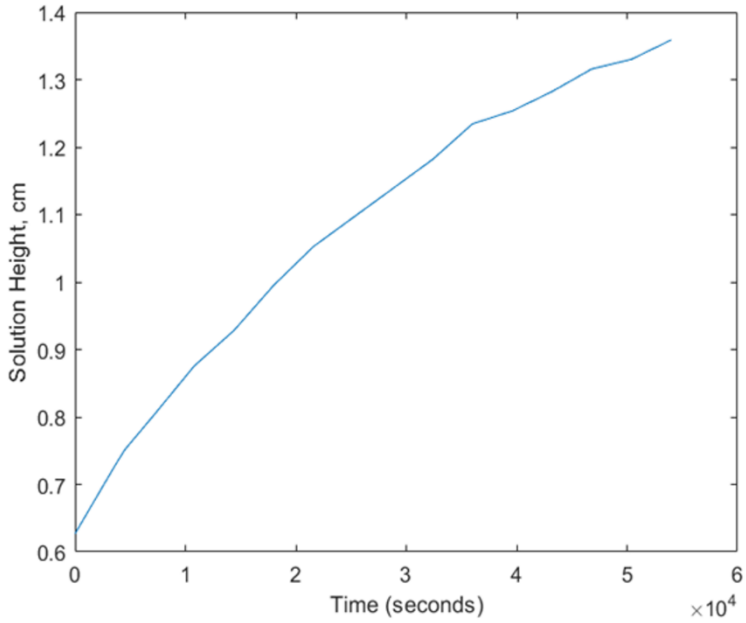

Figure S40. Solution height data for the PbI_2 / aep / DMF:DMSO reaction.

The k_{in} and k_{out} constants are fit to agree with the experimental solution heights. Sum of square error (SSE) between the model height and the experimental heights at every time point defines the objective function to minimize. Before choosing the optimization algorithm, we first assessed the numerical stability of the objective function to different candidate values of the parameters . For all 47 experiments, there is a well-defined local minimum when the parameters are approximately equal, and the SSE becomes undefined when they have very different magnitudes. A representative example of this topology can be seen in PbI_2 / aep / DMF:DMSO shown in **Figures S41**. This sort of behavior is physically reasonable, as condensation and evaporation rates should be similar in magnitude. Reasonably, one would not expect to see incredibly high rates of evaporation, with low condensation, or vice versa. The undefined values are also unsurprising. Remember from equation (4) that our bin heights increase with antisolvent concentration. If the antisolvent condensation rate far outstrips the outward rate of diffusion and the evaporation rate, the height of the bin will grow rapidly. Eventually it will reach a critical point at which the antisolvent cannot diffuse out of the bin given the time-step. The bin height will grow unbounded and eventually lead to undefined values. This behavior explains why the SSE is undefined when $k_{out}$is less than k_{in}.

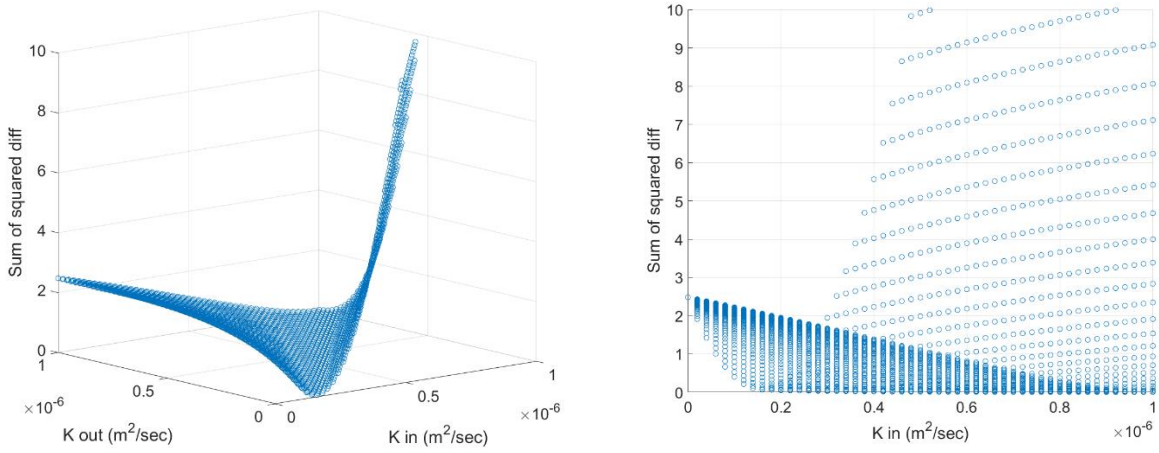

Figure S41. (a) Mesh Plot of the SSE for k_{in} and k_{out} for the PbI_2 / aep / DMF:DMSO reaction. Note the valley of minimums along the line k_{in}= k_{out} and (b) different view of the mesh plot focusing on k_{in} vs SSE.

Finding Seed Boundaries for k_{in} and k_{out} :

Knowing that the area of interest is located near the line k_{in} equals k_{out} narrows the search area. Further constraints arise from the nature of evaporation and condensation behavior. First, both constants must be positive. Second, the kinetic theory of gases provides an upper bound on k_{in}. Taken from the perspective of the kinetic theory of gases, the maximum k_{in} flux rate is J, given by the product of the number density of gas molecules, the average speed of those gas molecules, and their rate of collision with an area *a*, given by:

$$J = n_g * C * \frac{a}{4} \tag{20}$$

This is an upper bound for the rate of condensation, as molecules cannot enter faster than they collide with the surface. Each of these terms follows from the kinetic theory of gases. The number density is

$$n_g = \frac{mols}{cm^3} \tag{21}$$

Using the ideal gas law, we can then write:

$$n_g = \frac{P}{RT} \tag{22}$$

The system is sealed, so the pressure is the the vapor pressure of DCM and remaining experimental conditions can be used to calculate:

$$\frac{0.02326\ mols}{liter} = \frac{0.5655\ atm}{0.0821\ \frac{l*atm}{mol*k} * 296.15\ k} \tag{23}$$

The average speed of the molecules (derived via the Maxwell-Boltzmann treatment of the kinetic theory of gases) is:

$$C_{avg} = \sqrt{\frac{8RT}{\pi M}} \tag{24}$$

Inputting in the calculated values:

$$271.719\ m/s = \frac{8*8.314\ J*mol^{-1}*K^{-1}}{\pi*.08493\ kg/mol} \tag{25}$$

Assuming the value for *a* is 0.02, or the relative chance a molecule striking the liquid will enter the liquid state, our values from equations (23) and (25) are used:

$$0.03159\ mols/cm^2sec = \ 2.326 * 10^{-5}\frac{mols}{cm^3} * 27{,}171.9\frac{cm}{sec} * \frac{0.02}{4} \quad (26)$$

One can calculate the surface area of the interface:

$$0.0385\ cm^2 = \ \pi(.035\ cm)^2 \quad (27)$$

This allows the calculation of the max k_{in}, given the area:

$$1.22 * 10^{-4}\frac{mol}{sec} = \ 0.03159\frac{mols}{cm^2sec} * \ 0.0385\ cm^2 \quad (28)$$

Because the minima of the objective function are near k_{in}= k_{out}, the upper bound for k_{out} to be the same as k_{in} can be set. With some experimentation, we set the range of the k_{in} and k_{out} values from 1.0E-10 mol/sec to 1 mol/sec. This range was chosen to be certain that we have covered the physically plausible range of evaporation and condensation values. 1.0E-10 mol/sec was chosen as the lower bound, as lower rates gave consistently higher values for the SSE. The Nelder-Mead method [5] was used for the optimization, as it is robust against undefined values, and computationally inexpensive. To avoid getting trapped in local minima, the algorithm is started at 10 different seed locations spanning orders of magnitudes ranging from 1.0E-10 mol/sec to 1 mol/sec. The results of this process are shown in **Figure S42.** for the PbI_2 / aep / DMF:DMSO reaction.

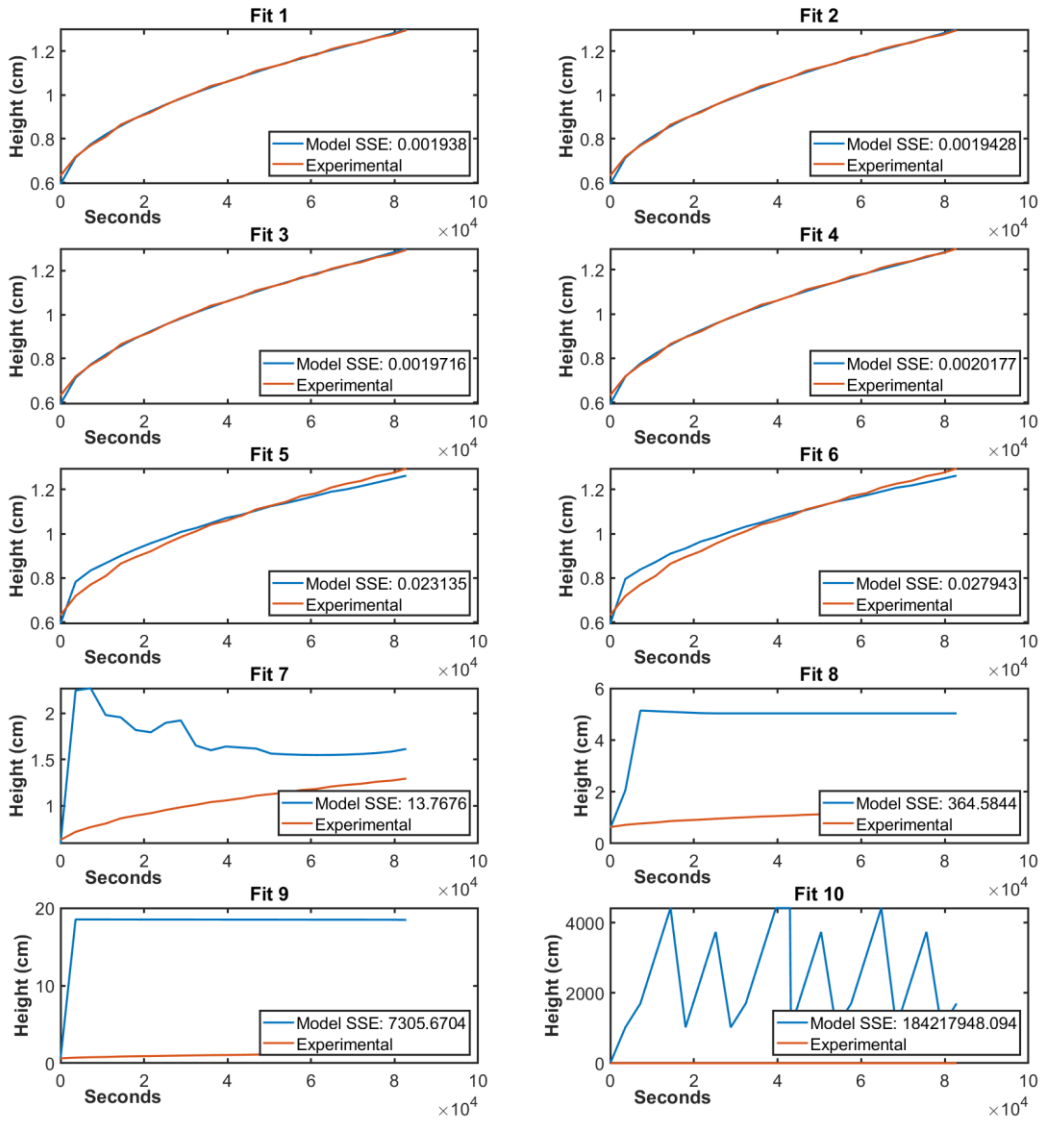

Figure S42. The ten best model fits for the PbI_2 / aep / DMF:DMSO reaction.

The highest quality fits have similar k_{in} and k_{out} values. This means that the random seeding is effective, as despite starting from different magnitudes, the minimum returns to the same area of the parameter space, with most of the experiments converging to a k_{in} and k_{out} value in the 1E-7 mol/sec range. Furthermore, experiments with slightly different k_{in} and k_{out} values produce similar concentration profiles. When querying the best four fits at a random time point, the concentration profiles of high-quality fits are nearly identical. This leads us to conclude that parameter choices that are equally good matches for experimental data, give effectively the same concentration profiles. An example of this can be seen for the PbI_2 / aep / DMF:DMSO reaction in **Figure S43.** The best four fits with low SSE are nearly indistinguishable from one another.

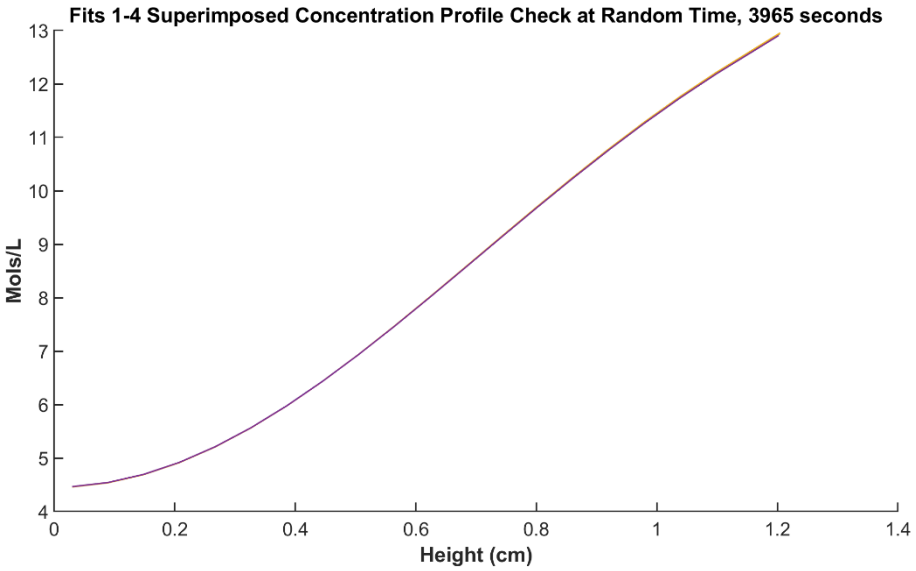

Figure S43. Superimposed concentration profiles of the four best fit models for experiment the PbI_2 / aep / DMF:DMSO reaction.

This optimization process is repeated three times to account for the lower, upper, and mean values of the diffusion coefficient. This results in three sets of values for the best fitting k_{in} and k_{out} values, and a resulting window of possible concentrations. However, in general the quality of fit is insensitive to these differences, as shown for the example of PbI_2 / aep / DMF:DMSO reaction in **Figure S44**.

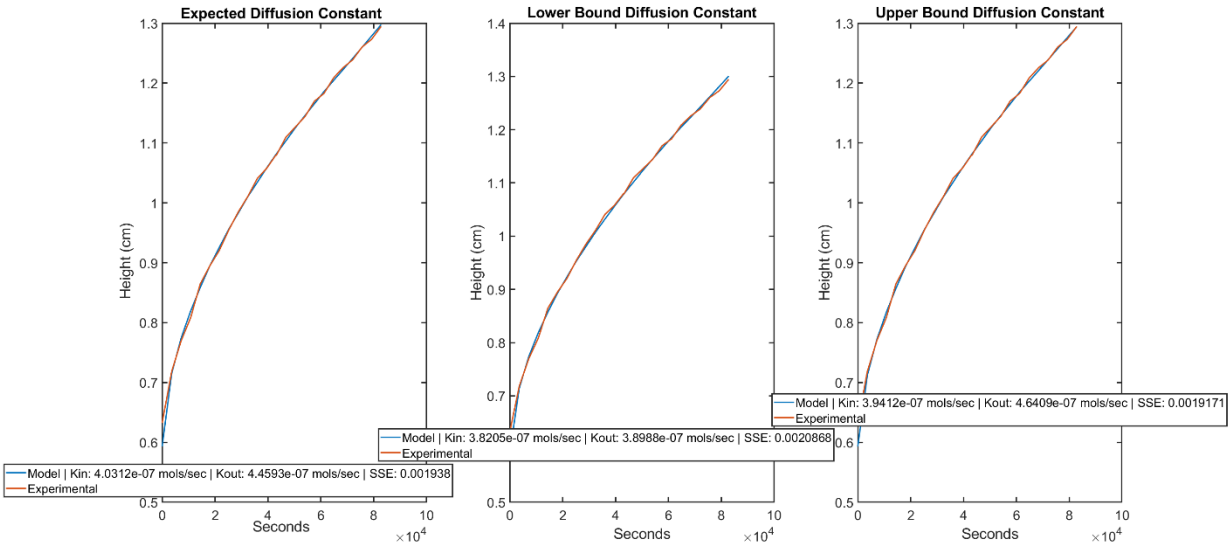

Figure S44. Solution height build up under the expected, lower, and upper bound diffusion coefficients for the PbI_2 / aep / DMF:DMSO reaction.

Experiment Truncation

Because the model assumes a constant condensation rate of antisolvent into the solution, the experiments must be stopped before the antisolvent supply runs out. Additionally, it is useful to have a uniform cut off time to the model to save on computation time, particularly during the parameter estimation process. Because the antisolvent builds up at different rates for every experiment and crystallization occurs at various times, a cutoff was implemented based upon total solution height (0.72 cm) rather than elapsed time. This value was chosen because it stopped all models from running before their antisolvent supply ran out and did not prematurely stop before crystallization. This criterion merely enforces the validity of underlying assumptions, and does not affect the final concentration estimate.

Results

Having determined an experimental cutoff point for the model and chosen an optimization algorithm to determine the values of our free parameters, calculate the concentration profile of the mother liquor at crystallization were calculated. Once the model has simulated the diffusion over the length of the experiment, we simply query the concentrations at the time and position of crystallization. The results of our model are shown in Figure **S45**.

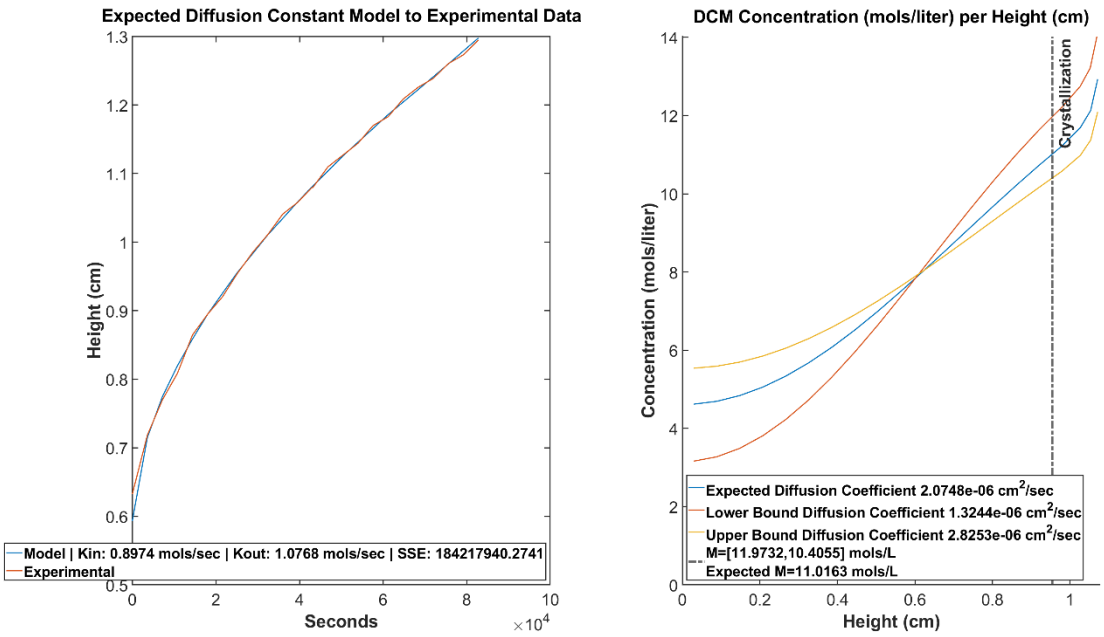

Figure S45. Best fit model and concentration profile at crystallization in the PbI_2 / aep / DMF:DMSO reaction.

The dotted blue line indicates the height of crystallization. The intersection between the concentration curve and the dotted line shows the antisolvent concentration at the time and location of crystallization. The blue line gives the expected concentration, with the yellow and red lines indicating the upper and lower bounds. Other chemical species concentrations are reported in an output file but are not visualized above.

Practical Limitations of the Model

While we have established that the finite volume approximation holds for realistic rates of diffusion, condensation, and evaporation processes, very unphysical guesses can result in numerical errors or large increases in computer time as bins are merged and spawned. Such unrealistic parameters might occur during an unsupervised numerical optimization. To combat this, we imposed limitations on the maximum number of total bins (to prevent runaway spawning) and the bin merging criteria (to prevent negative concentrations) in the code. This has no physical consequence, as these extreme parameter choices give a poor objective function and never get selected. It is merely serves the practical purpose of keeping each simulation to a limited computational budget.