DOI: 10.1021/cg900830y

CRYSTAL GROWTH & DESIGN Article

2010, Vol. 10
2515–2521

A Stochastic Model for Nucleation Kinetics Determination in Droplet-Based Microfluidic Systems

Limay Goh, Kejia Chen, Venkateswarlu Bhamidi, Guangwen He, Nicholas C. S. Kee, Paul J. A. Kenis, Charles F. Zukoski, III, and Richard D. Braatz*

Department of Chemical & Biomolecular Engineering, University of Illinois at Urbana-Champaign, 600 South Matthews Avenue, Urbana, Illinois 61801

Received July 18, 2009; Revised Manuscript Received April 9, 2010

ABSTRACT: The measured induction times in droplet-based microfluidic systems are stochastic and are not described by the deterministic population balances or moment equations commonly used to model the crystallization of amino acids, proteins, and active pharmaceutical ingredients. A stochastic model in the form of a Master equation is formulated for crystal nucleation in droplet-based microfluidic systems for any form of nucleation rate expression under conditions of time-varying supersaturation. An analytical solution is provided to describe the (1) time evolution of the probability of crystal nucleation, (2) the average number of crystals that will form at time t for a large number of droplets, (3) the induction time distribution, and (4) the mean, most likely, and median induction times. These expressions are used to develop methods for determining nucleation kinetics. Nucleation kinetics are determined from induction times measured for paracetamol and lysozyme at high supersaturation in an evaporation-based high-throughput crystallization platform, which give low prediction errors when the nucleation kinetics were used to predict induction times for other experimental conditions. The proposed stochastic model is relevant to homogeneous and heterogeneous crystal nucleation in a wide range of droplet-based and microfluidic crystallization platforms.

I. Introduction

Numerous studies have been carried out to obtain a better fundamental understanding of nucleation mechanisms.[1−8] More recently, high-throughput microfluidic crystallization systems have been used to crystallize a variety of organic compounds including amino acids, proteins, and active pharmaceutical ingredients.[9−15] The goals of such studies include the identification of conditions to produce protein crystals for X-ray and neutron crystallography, the search for polymorphs for pharmaceutical compounds, and exploration of the behavior of nucleation and crystallization kinetics at high supersaturation.[2,16−20] Such applications have the potential to impact structure−function analysis, pharmaceutical design, bioseparations, controlled drug delivery, treatment of protein condensation diseases, and study of human degenerative conditions.[16,21,22] Estimates of the nucleation kinetics can be used to revise crystallization conditions to produce higher quality protein crystals, for the design of kinetic processes such as separations, and for investigations into the fundamentals of nucleation. Microfluidic crystallization systems enable the measurement of a large number of induction times for various experimental conditions using only micrograms of starting material, and many researchers have looked into the estimation of nucleation kinetics from such data.

Techniques for the estimation of nucleation kinetics in macroscale crystallizers (e.g., see the reviews[23,24]) are poorly applicable at the microscale. Because of the small volumes, fluctuations in measured induction times at the microscale can be large.[25,26] Several methods have been used to estimate nucleation kinetics in droplets. The deterministic population balance models and moment equations used by various researchers[3,20] to estimate nucleation kinetics do not capture this stochastic behavior, which has been characterized in a series of high-throughput constant-supersaturation experiments.[25−29] Nucleation kinetics have been fit to average induction times, to the proportion of drops not containing crystals as a function of time, and to the average number of crystals per drop at long times.[25,27,28,30,31] The Poisson distribution provided a good fit to distributions in the number of lysozyme and lactose crystals nucleated at constant supersaturation in droplets using temperature jump techniques.[25,28] Izmailov et al.[26] developed the induction time statistics for the classical nucleation rate expression by assuming that the induction times could be fit by a gamma distribution. Knezic et al.[27] employed electrodynamic levitation and a two-step model to analyze cluster size fluctuations and cluster rearrangements to study nucleation kinetics at high supersaturation.

The aforementioned papers do not fully characterize the distribution of crystal nuclei in droplets in cases where supersaturation varies with time. Time-varying supersaturation occurs, for example, in an evaporation-based microwell crystallization platform (see Figure 1) designed to force a phase change in every experiment.[13,32] An alternative microfluidic crystallization system utilizes microchannels to form droplets of varying concentrations.[10] In either system, the supersaturation in each droplet can be monotonically increased in a known way as the evaporation proceeds. In Section IIa of this paper, a stochastic model in the form of a Master equation is formulated for any form of nucleation rate expression under conditions of time-varying supersaturation. An analytical solution is given for the Master equation to describe the (1) time evolution of the probability of crystal nucleation, (2) the average number of crystals that will form at time t for a large number of droplets, (3) the induction time statistics, and (4) the mean, most likely, and median induction times. Section IIb describes how to apply these expressions to estimate the nucleation kinetics from experimental data collected from droplet-based microfluidic systems. Section III estimates kinetic parameters in the classical nucleation expression by fitting mean induction times and induction time distributions for the nucleation

*Corresponding author: Phone: 217-333-5073. Fax: 217-333-5052. E-mail: braatz@illinois.edu.

Published on Web 05/07/2010

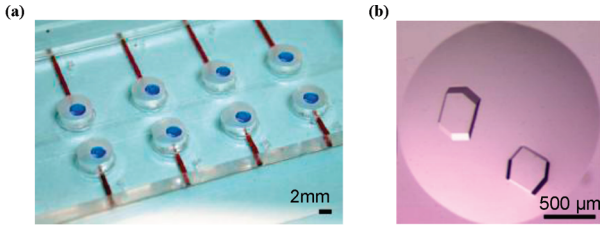

Figure 1. (a) Evaporation-based microwell crystallization platform with droplets adhering to a glass slip, and (b) crystals formed within an evaporating hanging drop as observed from above. In (a), dark red and blue inks are used to visualize the evaporation channel and the droplet, respectively.

of lysozyme and paracetamol in aqueous droplets in a high-throughput evaporation-based crystallization platform.[13] Section IV summarizes this study.

II. Theoretical Development

A. Statistics of Nucleation in Droplets with Time-Varying Supersaturation. Assume that the time for a nucleus to grow large enough to be observable is negligible and that the solution is spatially uniform (these assumptions are made in nearly all published studies on crystallization in microfluidic devices). Define $\kappa(t) > 0$ as the nucleation rate in a whole droplet (in #/s); that is, $\kappa(t)\,dt$ is the probability that a critical nucleus will form during an infinitesimal time interval dt. The time evolution of the probability $P_n(t)$ that a droplet contains n crystals is described by the Master equation:[33]

$$\frac{dP_0(t)}{dt} = -\kappa(t)P_0(t), \quad P_0(0) = 1 \tag{1a}$$

$$\frac{dP_n(t)}{dt} = \kappa(t)(P_{n-1}(t) - P_n(t)), \quad P_n(0) = 0, n = 1, 2, \ldots \tag{1b}$$

where the differential equations for $n \geq 1$ assume that earlier nuclei do not grow fast enough to significantly deplete solute from the solution. This latter assumption was explicitly made by Dombrowski et al.[28] and implicitly made by other researchers (e.g., ref 25), and is reasonable provided that the crystals observed in each droplet have approximately the same size. The differential equations (1) describe a nonstationary Poisson process[34] and can be solved recursively or by defining a probability-generating function (see Supporting Information) to give

$$P_0(t) = e^{-\int_0^t \kappa(s)\,ds} \tag{2a}$$

$$P_n(t) = \frac{1}{n!}\left[\int_0^t \kappa(s)\,ds\right]^n e^{-\int_0^t \kappa(s)\,ds}, n = 1, 2, \ldots \tag{2b}$$

Figure 2 shows the time evolution of the probabilities for droplets in which the overall nucleation rate κ is constant:

$$P_n(t) = \frac{1}{n!}(\kappa t)^n e^{-\kappa t}, n = 0, 1, 2, \ldots \tag{3}$$

which has been used to fit experimental data for nucleation in droplets of constant supersaturation.[25,35,36] The shapes of the curves in Figure 2b are the same as those observed experimentally for nucleation of lysozyme crystals in droplets under constant supersaturation (see Figure 4 of ref 25).

Consider a large number of droplets in a high-throughput microfluidic device in which each droplet moves from undersaturated to saturated to supersaturated conditions. The mean number of crystals at time t (averaged over a sufficiently large number of droplets) is[37]

$$\begin{aligned} E\{N(t)\} = \bar{N}(t) &= \sum_{n=0}^{\infty} nP_n(t) \\ &= \sum_{n=0}^{\infty} \frac{n}{n!}\left[\int_{t_{sat}}^{t} \kappa(s)\,ds\right]^n e^{-\int_{t_{sat}}^{t} \kappa(s)\,ds} \\ &= \int_{t_{sat}}^{t} \kappa(s)\,ds \end{aligned} \tag{4}$$

so that the time when the mean number of crystals is equal to 1 is

$$\int_{t_{sat}}^{t_{n,\mathrm{mean}}} \kappa(s)\,ds = 1 \tag{5}$$

where the subscript n indicates that this induction time is an average over the number of crystals. The *most likely* time of having exactly 1 crystal in the system is given by

$$\begin{aligned} \left.\frac{dP_1}{dt}\right|_{t=t_{ml}} &= \kappa(t_{n,\mathrm{ml}})e^{-\int_{t_{sat}}^{t_{n,\mathrm{ml}}} \kappa(s)\,ds}\left[1 - \int_{t_{sat}}^{t_{n,\mathrm{ml}}} \kappa(s)\,ds\right] \\ &= 0 \Rightarrow \int_{t_{sat}}^{t_{n,\mathrm{ml}}} \kappa(s)\,ds = 1 \end{aligned} \tag{6}$$

Hence, the most likely time $t_{n,\mathrm{ml}}$ for there to be exactly one crystal in the system is equal to the time for the mean number of crystals to be equal to 1. At this time, the probability distributions are $P_n = 1/(n!e)$ and the variance in the number of crystals n is

$$\mathrm{Var}[n] = \sum_{n=0}^{\infty}(n - \bar{N})^2 P_n = \frac{1}{e}\sum_{n=0}^{\infty}\frac{(n-1)^2}{n!} = 1 \tag{7}$$

The large value for the variance indicates a high probability in a particular experiment that there is either no nucleus or multiple nuclei at $t = t_{n,\mathrm{mean}} = t_{n,\mathrm{ml}}$. This is also seen by the low probability for having exactly 1 crystal in the system at the mean induction time, $P_1 = 1/e$. In general, it can be shown that the variance in the number of crystals n is

$$\mathrm{Var}[n(t)] = \int_{t_{sat}}^{t} \kappa(s)\,ds \tag{8}$$

which is monotonically increasing for experiments with positive supersaturation (i.e., $\kappa(t) \geq 0$). These results hold for any time evolution of the overall nucleation rate $\kappa(t)$ and hence any time evolution of the supersaturation.

The cumulative distribution function (CDF) for the time T_n when at least n crystals have nucleated is

$$\begin{aligned} P(T_n \leq t) &= F(t) \\ &= 1 - \sum_{m=0}^{n-1}\frac{1}{m!}\left[\int_{t_{sat}}^{t} \kappa(s)\,ds\right]^m e^{-\int_{t_{sat}}^{t} \kappa(s)\,ds} \end{aligned} \tag{9}$$

The corresponding probability distribution function (PDF) is[38,39]

$$f(t) = \frac{dF(t)}{dt} = \frac{\kappa(t)\left[\int_{t_{sat}}^{t} \kappa(s)\,ds\right]^{n-1}}{(n-1)!} e^{-\int_{t_{sat}}^{t} \kappa(s)\,ds} \tag{10}$$

whose integral gives the probability of T_n lying within a particular interval, $\int_{t_1}^{t_2} f(t)\,dt = P(t_1 \leq T_n \leq t_2)$. These induction time

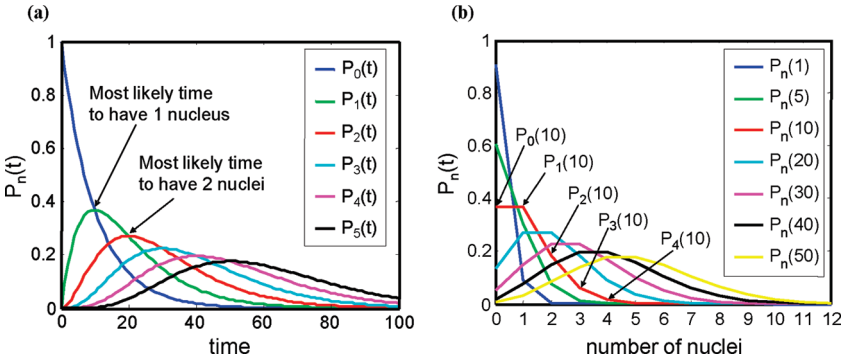

Figure 2. (a) Time evolution of probabilities $P_n(t)$ for $n = 0, 1, 2, ..., 7$ for $\kappa = 0.1$, and (b) the corresponding $P_n(t)$ vs n for different times. The lines in (b) are drawn to guide the eye.

distributions can be used to compute a variety of induction times that can be compared to experiments. For the induction time defined by the appearance of at least 1 nucleus ($n = 1$):

$$P(T_1 \le t) = F(t) = 1 - e^{-\int_{t_{sat}}^{t} \kappa(s)\,ds}$$

$$f(t) = \frac{dF(t)}{dt} = -\frac{d}{dt} e^{-\int_{t_{sat}}^{t} \kappa(s)\,ds} = \kappa(t) e^{-\int_{t_{sat}}^{t} \kappa(s)\,ds} \quad (11)$$

For time-varying nucleation rate $\kappa(t)$, the mean time for the appearance of at least n crystals is

$$t_{mean} = \int_0^\infty t \frac{\kappa(t)\left[\int_{t_{sat}}^{t} \kappa(s)\,ds\right]^{n-1}}{(n-1)!} e^{-\int_{t_{sat}}^{t} \kappa(s)\,ds} dt \quad (12)$$

with the *mean induction time* for the nucleation of at least 1 crystal being

$$t_{mean} = \int_0^\infty t\kappa(t) e^{-\int_{t_{sat}}^{t} \kappa(s)\,ds}\,dt \quad (13)$$

This analytical expression corresponds to the induction time that is normally experimentally reported in the literature.[40] The variance of the distribution of times about the mean time for the appearance of at least n crystals is

$$\mathrm{Var}(t_{ind}) = \int_0^\infty t^2 \frac{\kappa(t)\left[\int_{t_{sat}}^{t} \kappa(s)\,ds\right]^{n-1}}{(n-1)!} e^{-\int_{t_{sat}}^{t} \kappa(s)\,ds}\,dt - t_{mean}^2 \quad (14)$$

which for $n = 1$ is

$$\mathrm{Var}(t_{ind}) = \int_0^\infty t^2 \kappa(t) e^{-\int_{t_{sat}}^{t} \kappa(s)\,ds}\,dt - t_{mean}^2 \quad (15)$$

The *maximum likelihood* (most likely) *induction time* t_{ml} must occur at a time t that satisfies $df/dt = 0$ for f in eq 11, which occurs if and only if

$$\frac{d\kappa}{dt}(t_{ml}) - \kappa^2(t_{ml}) = 0 \quad (16)$$

That is, t_{ml} can be computed from all (typically 1 or 2) of the finite positive roots of the above expression:

$$t_{ml} = \arg\max_{t_k} \left\{ f(t_k) \middle| \frac{d\kappa}{dt}(t_k) - \kappa^2(t_k) = 0 \right\} \quad (17)$$

The *median induction time* t_{median} for the nucleation of at least 1 crystal satisfies

$$P(T_1 \le t_{median}) = 0.5 \Leftrightarrow e^{-\int_{t_{sat}}^{t_{median}} \kappa(s)\,ds} = 0.5 \Leftrightarrow \int_{t_{sat}}^{t_{median}} \kappa(s)\,ds = \ln 2 \quad (18)$$

which is smaller than the induction times (5)−(6) but has a similar value when the overall nucleation rate $\kappa(t)$ increases rapidly right before a crystal nucleates, which typically occurs in an evaporation-based crystallization platform with constant or increasing evaporation rate.[20]

B. Determination of Nucleation Kinetics. The overall nucleation rate $\kappa(t)$ could be experimentally estimated from the ratio of the proportion of droplets not containing nuclei to its derivative as a function of time in eq 1a, but taking derivatives of data is inaccurate. Least-squares estimation is an alternative approach that first parametrizes $\kappa(t)$ in terms of nucleation parameters and then fits these parameters to experimental data.[41] For example, the overall nucleation rate is given by $\kappa(t) = J(S(t))V(t)$ for homogeneous nucleation in droplets, where $S(t) = (C(t) - C_{sat})/C_{sat}$ is the relative supersaturation determined from a material balance on the solute,[42] $C(t)$ is the concentration (g/L) of solute to be crystallized at time t, C_{sat} (g/L) is the solubility, $J(S)$ is the number of crystals formed per time per unit volume (#/s-L), and $V(t)$ is the volume (L) of the droplet determined from the evaporation rate and a material balance on the solvent.[13] The nucleation parameters A and B in the classical homogeneous nucleation expression[43−46]

$$J(S) = AC \exp(-B/(\ln(S+1))^2) \quad (19)$$

can be determined by numerical solution of

$$\min_{A,B} \sum_{i=1}^{n_{exp}} \sum_k \left(\ln P_{0,i}(t_k) + \int_0^{t_k} J(S_i(s)) V_i(s)\,ds \right)^2 \quad (20)$$

which minimizes deviations from eq 2a, summed over all experimental conditions i and times t_k in which the proportion of droplets without crystals, $P_{0,i}(t_k)$, are measured. As eq 20 was derived based only on eq 1a, its application does not require the assumption that the growth of any crystals does not significantly deplete solute from the solution.

An alternative formulation fits a mean or most likely induction time $t_{ind} \in \{t_{n,ml}, t_{n,mean}, t_{mean}, t_{ml}, t_{median}\}$:

$$\psi = \min_{A,B} \sum_{i=1}^{n_{exp}} (t_{ind,i} - t_{ind,measured,i})^2 \quad (21)$$

Table 1. Induction Times $\{t_{n,\mathrm{ml}}, t_{n,\mathrm{mean}}, t_{\mathrm{mean}}, t_{\mathrm{ml}}, t_{\mathrm{median}}\}$ from eqs *5*, *6*, *13*, *17*, and *18* for Paracetamol in Water for 11 Experimental Conditions with Different Evaporation Rates and Initial Solute Concentrations[a]

experimental condition #	evaporation rate (μg/h)	C_0 (g/kg water)	t_{mean} (h)	$t_{n,\mathrm{mean}}, t_{n,\mathrm{ml}}$ (h)	t_{median} (h)	t_{ml} (h)[b]	$t_{\mathrm{ind,measured}}$ (h)	t_{lower} (h)[c]	t_{upper} (h)
1	34.7	11.9	13.2	13.5	13.3	13.4	13.8	12.2	14.1
2	30.2	11.9	15.1	15.4	15.2	15.4	15.3	14.0	16.1
3	30.2	9.90	15.5	15.8	15.7	15.7	15.8	14.5	16.4
4	26.1	11.9	17.4	17.8	17.5	17.7	17.8	16.1	18.5
5	26.1	9.90	17.9	18.2	18.0	18.2	18.3	16.8	18.9
6	22.2	11.9	20.4	20.7	20.5	20.8	20.3	18.8	21.6
7	22.2	9.90	21.0	21.3	21.1	21.3	21.3	19.6	22.1
8	22.2	7.94	21.5	21.9	21.7	21.8	22.0	20.4	22.5
9	18.7	11.9	24.0	24.5	24.1	24.4	23.3	22.2	25.5
10	18.7	9.90	24.8	25.2	24.9	25.2	24.8	23.2	26.1
11	18.7	7.94	25.5	25.9	25.6	25.8	25.7	24.2	26.6

[a] Each measured induction time is the time in which one or more crystals was observed averaged over multiple droplets and t_{lower}, t_{upper} are 90% confidence limits. The nucleation parameters A and B were fit in eq 21 for the odd-numbered experimental conditions based on the mean induction time t_{mean}. Even-numbered experiments are used to assess accuracy of the model predictions. [b] An example calculation of the maximum likelihood induction time t_{ml} in Supporting Information. [c] t_{lower} is determined from $\int_0^{t_{\mathrm{lower}}} f(t)\,\mathrm{d}t = 0.05$ and t_{upper} from $\int_0^{t_{\mathrm{upper}}} f(t)\,\mathrm{d}t = 0.95$ where $f(t)$ is the induction time distribution (11).

where t_{ind} is given by eqs 5, 6, 13, 17, or 18. Unbiased estimates of the nucleation parameters can be determined using $t_{\mathrm{ind}} \in \{t_{n,\mathrm{ml}}, t_{n,\mathrm{mean}}\}$ by comparing each value to the time when the average number of crystals in a large number of droplets is equal to 1 (for an example of such data, see refs 25−35.). A weakness of using $t_{\mathrm{ind}} \in \{t_{n,\mathrm{ml}}, t_{n,\mathrm{mean}}\}$ is their dependence on the assumption that the early nuclei do not significantly deplete solute from the solution before additional nuclei form. This assumption was not needed to derive the expressions for $t_{\mathrm{ind}} \in \{t_{\mathrm{mean}}, t_{\mathrm{ml}}, t_{\mathrm{median}}\}$. Fitting the mean induction time t_{mean} averages experimental errors over all of the data points, whereas fitting the median induction time t_{median} is highly insensitive to outliers in the data.[47] An experimental measurement of t_{ml} is the time at the peak of a histogram of measured induction times, which is not as precise to estimate as t_{mean} or t_{median}.

If sufficient induction times are collected to construct a cumulative distribution, these data can be directly fit to estimate the nucleation kinetics by minimizing the sum of squared errors of the cumulative distribution functions:

$$\psi_F = \min_{A,B} \sum_{i=1}^{n_{\mathrm{exp}}} \sum_{k} \left(F_i(t_k) - F_{\mathrm{measured},i}(t_k)\right)^2 \qquad (22)$$

with $F(t)$ given in eq 11.

The above approaches for the determination of nucleation kinetics also apply to heterogeneous nucleation, with suitable modifications to the overall nucleation rate $\kappa(t)$. For example, consider droplets attached to a solid surface but otherwise surrounded by humid air. If the crystals nucleate at the solid−liquid interface, then this heterogeneous surface nucleation is described by inserting $\kappa(t) = J_{\mathrm{het}}(S(t))A(t)$ into the above expressions (1)−(18), where $A(t)$ is the contact area between the droplet and the solid surface. If the nucleation occurs at the liquid−air interface, then $A(t)$ would be defined as the contact area between the liquid and humid air. If nucleation occurs at the interface between aqueous droplets surrounded by an oil, then $A(t)$ would be the contact area between the two liquids.

The next section applies two of the above approaches to estimate nucleation kinetics from several sets of measured induction times, both to validate the statistical model for nucleation in droplets and for illustration purposes.

III. Results and Discussion

This section applies the above analyses to experimental data for the nucleation of organic crystals in aqueous droplets in an evaporation-based microfluidic platform (for details on the experimental system and procedures, see ref 13). Computational fluid dynamics indicated that natural convection was sufficient to ensure that the solutions were spatially uniform.[48] It was observed during the experiments that the crystals did not stick to the glass slips,[32] so the applications assume that the nucleation was homogeneous.

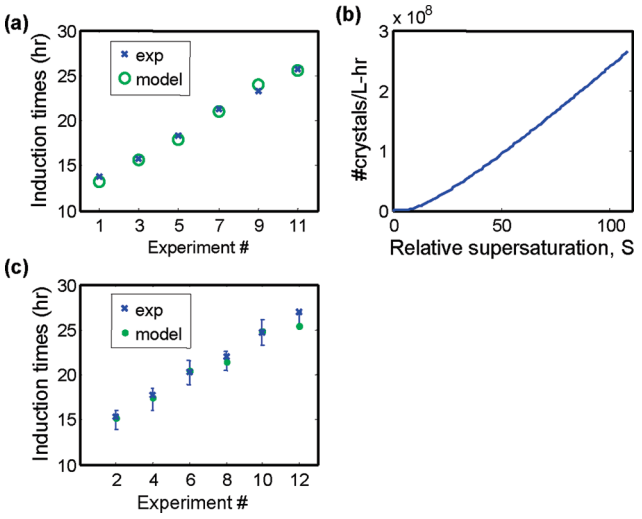

Figure 3. (a) Measured and model (t_{mean}) mean induction times for paracetamol in water for six experimental conditions with kinetic parameters $A = 3.52 \times 10^5\ \mathrm{g^{-1}\,h^{-1}}$ and $B = 14.3$, (b) corresponding nucleation rate expression 19, and (c) comparison of model and experimental mean induction times for five additional experimental conditions, showing 90% prediction intervals. [The nucleation kinetics were the same regardless of whether the time to grow to a visible size was taken into consideration (using growth kinetics obtained from Finnie et al.,[52] $G = 0.0183(C(t)/C_{\mathrm{sat}} - 1)^2$ m/h), supporting the assumption of negligible growth time.]

A. Prediction of Mean Induction Time for Different Experimental Conditions. The mean time in which paracetamol crystals were first observed in an aqueous droplet was recorded for each of 11 experimental conditions with varying evaporation rates and initial solution concentrations (see Table 1). The nucleation parameters A and B for the classical nucleation expression 19 were obtained by numerical solution of eq 21[49] to fit the mean induction times to the model values t_{mean} in eq 13[50] for six experimental conditions.[51] The measured induction times are well fit by the classical nucleation model (Figure 3a). The nucleation rate expression with the

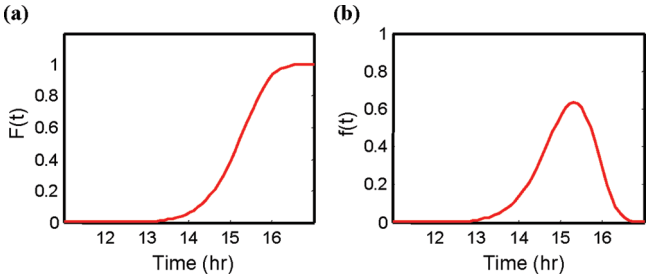

Figure 4. (a) Model cumulative distribution function and (b) probability distribution function 11 for Experimental Condition #2 for paracetamol in water.

Table 2. Experimental Conditions for Measuring Induction Times for Lysozyme in NaCl Aqueous Solution with Initial NaCl (precipitant) Concentration of 0.36 M[a]

experimental condition #	evaporation rate (L/h)	C_0 (g/L solution)	number of measured induction times
1	1.961×10^{-7}	18	30
2	2.056×10^{-7}	18	15

[a] The solubility of $C_{sat}(t) = 1.5694[C_{NaCl}(t)]^{-2.94}$ g/L was incorporated into the time-varying supersaturation to correctly capture the change in solubility. The nucleation parameters A and B were fit in eq 22 for Experimental Condition #1. Induction times for Experimental Condition #2 were used to assess accuracy of model predictions.

best-fit nucleation parameters predicted the induction times for five other experimental conditions with an average deviation of <1/2 h (see Figure 3c). All of the predicted induction times are within 90% confidence intervals computed from the PDF (11) for the induction time (Table 1). The induction times $\{t_{n,ml}, t_{n,mean}, t_{mean}, t_{ml}, t_{median}\}$ obtained using the same nucleation kinetics are very close to each other, for the nucleation of paracetamol in water for the 11 experimental conditions in Table 1. The corresponding CDF and PDF for the induction times are shown in Figure 4 for Experimental Condition #2. The induction time distribution is very different from that obtained for nucleation in droplets operating at constant supersaturation, which is an exponential distribution peaked at zero time, as seen by inserting a constant κ in eq 10. In the microfluidic system in Figure 1, the supersaturation and the nucleation rate $\kappa(t)$ start at zero and slowly increase as water evaporates from the drop, reaching large values only after ~14 h, so that the value of the induction time distribution is negligible for the first ~14 h.

B. Determine Nucleation Kinetics by Fitting the Cumulative Distribution. The cumulative distribution of induction times were collected for the evaporation of lysozyme in NaCl aqueous solution in many droplets at two experimental conditions (see Table 2). The kinetic parameters A and B in the classical nucleation expression 19 were fit to the cumulative distribution of induction times for Experimental Condition #1 by numerical solution of eq 22. Both the mean and variation of the measured induction times are closely described by the nonhomogeneous Poisson model (11) with fitting of only two kinetic parameters (see Figure 5ab). The classical nucleation model (19) with the two best-fit kinetic parameters accurately predict the cumulative distribution of induction times for Experimental Condition #2 (Figure 5ab), providing some confidence in the statistical assumptions underlying the model (11). The distribution of induction times at one experimental condition provided enough information to estimate A and B accurately enough to predict the

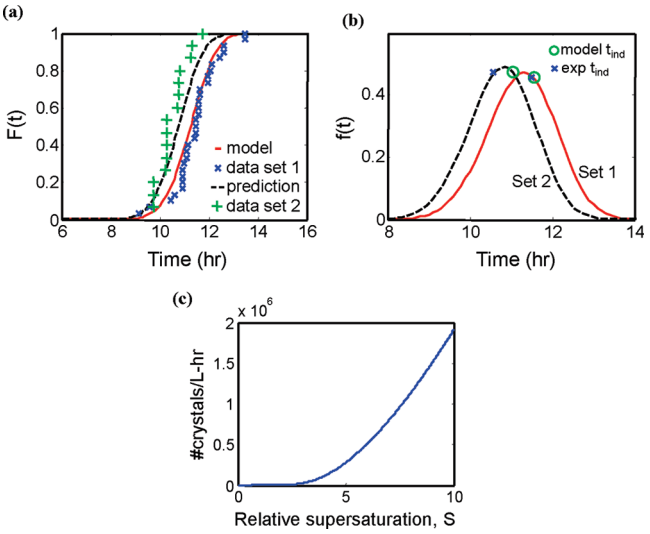

Figure 5. Lysozyme−NaCl−water system: (a) experimental and model cumulative distributions of induction times at two experimental conditions, with model (19) using parameters $\ln(A, g^{-1} h^{-1}) = 12.5$ and $B = 9.7$ fit to induction times for Experimental Condition #1, (b) probability distribution functions with model mean induction time (o) and experimental mean induction time ($\times$), and (c) corresponding nucleation rate expression 19. The 95% confidence intervals for $\ln(A, g^{-1} h^{-1})$ and B are [11.9, 13.1] and [7.4, 12.0] using t-statistics and [11.3, 13.7] and [6.8, 12.7] using F-statistics (see Supporting Information for calculation and discussion of the confidence intervals). The time for a nucleus to grow large enough to be visible in these experiments is very short compared to the induction time.[20]

cumulative distribution function at another experimental condition.

The goodness of fit of the cumulative distribution function $F(t)$ in 11 to the experimentally measured induction times was also evaluated using the Kolmogorov−Smirnov statistic[53]

$$D_n = \sup_t |F_N(t) - F(t)| \quad (23)$$

where the null hypothesis is that $F(t)$ in eq 23 is the distribution 11. The statistical analysis, with details in Supporting Information, accepted the distribution function 11 as appropriate for describing the empirical distribution function at an alpha level of 0.1.

IV. Conclusion

In this study, a stochastic model was presented that characterizes the statistics of crystal nucleation in microfluidic systems with time-varying supersaturation and for any form of nucleation rate expression. The analytical solution for this model is applicable to microfluidic crystallization platforms in which supersaturation changes with time due to changes in solute/precipitant concentration, temperature, pH, etc., and can be used to identify nucleation kinetics by fitting the measured average crystal number vs time (eq 4), the proportion of droplets that do not contain crystals vs time (eq 20), induction times (eq 21), or induction time distributions (eq 22). The kinetic parameters for the nucleation of paracetamol in aqueous droplets obtained by fitting the mean induction times (13) to six experimental conditions gave predicted induction times with an average deviation <1/2 h for six other experimental conditions. The nucleation kinetic parameters in lysozyme−NaCl−water droplets determined by fitting the cumulative induction time distribution (11) at one experimental

condition accurately predicted the cumulative induction time distribution at another experimental condition. The Komogorov−Smirnov goodness-of-fit test indicated that the measured induction time distribution at each experimental condition is consistent with the theoretical distribution (11). The stochastic model is widely applicable to crystal nucleation in a wide range of droplet and microfluidics-based devices including levitated droplet systems,[27] continuous-flow plug-based crystallization,[8] and patterned substrate-based systems.[29] As discussed in Section II, the model should also be applicable to heterogeneous nucleation at solid−liquid, liquid−vapor, and liquid−liquid interfaces.

Acknowledgment. Financial support is acknowledged from the National Science Foundation (Grant No. 0426328), the National Institutes of Health (R21 EB004513), 3M, and the Singapore Agency for Science, Technology and Research.

Supporting Information Available: (a) The derivation of the analytical solution of the Master equation, (b) an example calculation of the maximum-likelihood induction time, (c) a description of the quantification of uncertainty in kinetic parameters, and (d) a description of Kolmogorov−Smirnov statistics. This information is available free of charge via the Internet at http://pubs.acs.org/.

References

(1) Blow, D. M.; Chayen, N. E.; Lloyd, L. F.; Saridakis, E. *Protein Sci.* **1994**, *3*, 1638.
(2) ten Wolde, P. R.; Frenkel, D. *Science* **1997**, *277*, 1975.
(3) Larson, M. A.; Saikumar, M. V.; Glatz, C. E. *J. Cryst. Growth* **1998**, *187*, 277.
(4) Aizenberg, J.; Black, A. J.; Whitesides, G. M. *Nature* **1999**, *398*, 495.
(5) Baird, J. K. *J. Cryst. Growth* **1999**, *204*, 553.
(6) Cacciuto, A.; Auer, S.; Frenkel, D. *Nature* **2004**, *428*, 404.
(7) Li, L.; Nachtergaele, S.; Seddon, A. M.; Tereshko, V.; Ponomarenko, N.; Ismagilov, R. F. *J. Am. Chem. Soc.* **2008**, *130*, 14324.
(8) Kreutz, J. E.; Li, L.; Roach, L. S.; Hatakeyama, T.; Ismagilov, R. F. *J. Am. Chem. Soc.* **2009**, *131*, 6042.
(9) Sanjoh, A.; Tsukihara, T. *J. Cryst. Growth* **1999**, *196*, 691.
(10) Zheng, B.; Roach, L. S.; Ismagilov, R. F. *J. Am. Chem. Soc.* **2003**, *125*, 11170.
(11) Lee, A. Y.; Lee, I. S.; Dettet, S. S.; Boerner, J.; Myerson, A. S. *J. Am. Chem. Soc.* **2005**, *127*, 14982.
(12) Squires, T. M.; Quake, S. R. *Rev. Mod. Phys.* **2005**, *77*, 977.
(13) Talreja, S.; Kim, D. Y.; Mirarefi, A. Y.; Zukoski, C. F.; Kenis, P. J. A. *J. Appl. Crystallogr.* **2005**, *38*, 988.
(14) Hansen, C. L.; Sommer, M. O. A.; Quake, S. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14431.
(15) Hansen, C. L.; Classen, S.; Berger, J.; Quake, S. R. *J. Am. Chem. Soc.* **2006**, *128*, 3142.
(16) Vekilov, P. G. *Cryst. Growth Des.* **2004**, *4*, 671.
(17) Li, L.; Mustafi, D.; Fu, Q.; Tereshko, V.; Chen, D. L. L.; Tice, J. D.; Ismagilov, R. F. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19243.
(18) Lau, B. T. C.; Baitz, C. A.; Dong, X. P.; Hansen, C. L. *J. Am. Chem. Soc.* **2007**, *129*, 454.
(19) Shim, J. U.; Cristobal, G.; Link, D. R.; Thorsen, T.; Jia, Y. W.; Piattelli, K.; Fraden, S. *J. Am. Chem. Soc.* **2007**, *129*, 8825.
(20) Talreja, S.; Kenis, P. J. A.; Zukoski, C. F. *Langmuir* **2007**, *23*, 4516.
(21) Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdos, J.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M. *Nature* **2002**, *416*, 507.
(22) Pan, W.; Kolomeisky, A. B.; Vekilov, P. G. *J. Chem. Phys.* **2005**, *122*, 174905.
(23) Rawlings, J. B.; Miller, S. M.; Witkowski, W. R. *Ind. Eng. Chem. Res.* **1993**, *32*, 1275.
(24) Fujiwara, M.; Nagy, Z. K.; Chew, J. W.; Braatz, R. D. *J. Process Control* **2005**, *15*, 493.
(25) Galkin, O.; Vekilov, P. G. *J. Phys. Chem. B* **1999**, *103*, 10965.
(26) Izmailov, A. F.; Myerson, A. S.; Arnold, S. *J. Cryst. Growth* **1999**, *196*, 234.
(27) Knezic, D.; Zaccaro, J.; Myerson, A. S. *J. Phys. Chem. B* **2004**, *108*, 10672.
(28) Dombrowski, R. D.; Litster, J. D.; Wagner, N. J.; He, Y. *Chem. Eng. Sci.* **2007**, *62*, 4802.
(29) Singh, A.; Lee, I. S.; Myerson, A. S. *Cryst. Growth Des.* **2009**, *9*, 1182.
(30) Weidinger, I.; Klein, J.; Stockel, P.; Baumgartel, H.; Leisner, T. *J. Phys. Chem. B* **2003**, *107*, 3636.
(31) Selimovic, S.; Jia, Y. W.; Fraden, S. *Cryst. Growth Des.* **2009**, *9*, 1806.
(32) He, G. Ph.D. Thesis, University of Illinois at Urbana-Champaign and National University of Singapore, **2007**.
(33) Fichthorn, K. A.; Weinberg, W. H. *J. Chem. Phys.* **1991**, *95*, 1090.
(34) Cinlar, E. *Introduction to Stochastic Processes*; Prentice-Hall: Englewood Cliffs, NJ, 1971.
(35) Galkin, O.; Vekilov, P. G. *J. Am. Chem. Soc.* **2000**, *122*, 156–163.
(36) Koop, T.; Luo, B. P.; Biermann, U. M.; Crutzen, P. J.; Peter, T. *J. Phys. Chem. A* **1997**, *101*, 1117.
(37) The lower limit of the integral has been changed to t_{sat} since nucleation can only occur once the supersaturation becomes positive (that is, $\kappa = 0$ from $t = 0$ to $t = t_{sat}$).
(38) To verify that $f(t)$ in (10) has properties of a probability distribution function, let $a = \int_{t_{sat}}^{t} \kappa(s)\,\mathrm{d}(s)$ (or $\mathrm{d}a = \kappa(s)\,\mathrm{d}s$). For $\kappa > 0$, $a = 0$ for $t = t_{sat}$, $a \rightarrow \infty$ as $t \rightarrow \infty$, and
$$\int_0^\infty f(s)\,\mathrm{d}s = \frac{1}{(n-1)!}\int_0^\infty a^{n-1}e^{-a}\,\mathrm{d}a = 1$$
where $\int_0^\infty a^{n-1}\,e^{-a}\,\mathrm{d}a$ is the gamma function.
(39) For constant κ, $f(t) = \kappa^n t^{n-1}e^{-\kappa t}/(n-1)!$ is the Erlang probability distribution, which is the Gamma distribution for integer n (Devore, J. L. *Probability & Statistics for Engineering and the Sciences*, Brooks/Cole Publishing Company: Belmont, CA, 1982). The expected value of a is
$$a_{\mathrm{mean}} = \int_0^\infty a f(a)\,\mathrm{d}a = n$$
and its variance is
$$\mathrm{Var}(a) = \int_0^\infty (a-n)^2 f(a)\,\mathrm{d}a = \int_0^\infty a^2 f(a)\,\mathrm{d}a = n$$
which agrees with Cinlar,[34] in which a nonstationary Poisson process was transformed into a stationary Poisson process with constant rate $\kappa = 1$. The corresponding Erlang distribution has a mean of $t_{mean} = n/\kappa$.
(40) Penkova, A.; Dimitrov, I.; Nanev, C. *Ann. N.Y. Acad. Sci.* **2004**, *1027*, 56.
(41) Beck, J. V.; Arnold, K. J. *Parameter Estimation in Engineering and Science*; Wiley: New York, 1977.
(42) This manuscript follows the common practice of writing nucleation kinetics in terms of a relative supersaturation defined in terms of concentrations. Strictly speaking, the driving force for nucleation should be written in terms of the chemical potential difference or a ratio of activities of the substance in the solid and liquid states; for example, see Mullin, J. W.; Sohnel, O. *Chem. Eng. Sci.* **1977**, *32*, 683, which can be important at high solute concentrations. The stochastic model and associated analysis are valid for nucleation kinetics written in terms of chemical potential, activities, or absolute supersaturation, with obvious modifications .
(43) Nielsen, A. E. *Kinetics of Precipitation*; Pergamon Press: Oxford, 1964.
(44) Walton, A. G. In *Nucleation*; Zettlemoyer, A. C., Ed.; Marcel Dekker: New York, 1969, p 238.
(45) The classical nucleation expression with constants A and B is presented as an example; the model and analysis also apply to A and B varying during the experiment by including those dependencies. For example, consider the expressions $A = 2v(kt\sigma)^{1/2}/h$ and $B = -16\pi\sigma^3 v^2 \Delta G_a/k^4T^4$ where v is the molecular volume, k is the Boltzmann constant, T is absolute temperature, σ is the surface energy per unit area of the nuclei, h is Planck's constant, and ΔG_a is the diffusion activation energy (Walton, A. G. In *Nucleation*; Zettlemoyer, A. C., Ed.; Marcel Dekker: New York, 1969; p 238). The dependency of the temperature on time would be included in A and B for a cooling crystallization (such as in Edd, J. F.; Humphry, K. J.; Irimia, D.; Weitz, D. A.; Toner, M. *Lab on a Chip* **2009**, *9*, 1859 or Stan, C. A.; Schneider, G. F.; Shevkoplyas, S. S.; Hashimoto, M.; Ibanescu, M.; Wiley, B. J.; Whitesides, G. M. *Lab on a Chip* **2009**, *9*, 2293). For variation in the surface energy and/or diffusion activation energy during an experiment (such as in Bhamidi, V.; Varanasi, S.; Schall, C. A. *Cryst. Growth Des.* **2002**, *2*, 395), the variation on temperature and/or solute concentration would be included explicitly, which would be written as a

function of time for each drop as computed from mass and/or energy conservation equations.
(46) The model and analysis apply to other nucleation expressions as well, and to systems in which the overall nucleation rate decreases with time due to partial rehydration of the droplet (such as in Talreja, S.; Perry, S. L.; Guha, S.; Bhamidi, V.; Zukoski, C. F.; Kenis, P. J. A. *J. Phys. Chem. B* **2010**, *114*, 4432) or the nucleation rate decreases with increased solute concentration. The latter could occur, for example, due to the formation of a glassy state. See Roos, Y; Karel, M. *J. Food Sci.* **1992**, *57*, 775. Debenedetti, P. G.; Stillinger, F. H. *Nature* **2001**, *410* (6825), 259. Burnett, D. J.; Thielmann, F.; Booth, J. *Int. J. Pharm.* **2004**, *287*, 123. Zobrist, B.; Marcolli, C.; Pedernera, D. A.; Koop, T. *Atmos. Chem. Phys.* **2008**, *8*, 5221.
(47) Martin, D. R.; Maronna, R. A.; Yohai, V. J. *Robust Statistics: Theory and Methods*; Wiley: New York, 2006.
(48) Goh, L. M. *M.S. Thesis, University of Illinois at Urbana-Champaign*, **2007**.
(49) Because of the low number of parameters, Eq 21 was solved by gridding over the model parameters A and B.
(50) The mean induction time (13) was numerically computed using the Matlab *quad* problem to find the value of inner integral and *trapz* to find the value of the outer integral, as well as by solving an equivalent system of two ordinary diffusion equations (ODEs) using *ode45*; one ODE for the inner integral and the second ODE for the outer integral. The same numerical method used for finding the mean can be used to determine the variance (15).
(51) For experiments at constant temperature such as for the specific experiments described in this paper, the value of $A = 2v(kt\sigma)^{1/2}/h$ has only a weak dependence on the solute properties and assuming a constant value for A is a very accurate approximation. The value of $B = -16\pi\sigma^3 v^2 \Delta G_a / k^4 T^4$ depends more strongly on solute properties, but was kept constant due to a lack of data available on the values of the surface energy σ and the diffusion activation energy ΔG_a as a function of the solute concentration.
(52) Finnie, S. D.; Ristic, R. I.; Sherwood, J. N.; Zikic, A. M. *J. Cryst. Growth* **1999**, *207*, 308.
(53) Hogg, R. V.; Tanis, E. A. *Probability and Statistical Inference*; Prentice-Hall: Upper Saddle River, NJ, 2001.