First assembly times and equilibration in stochastic coagulation-fragmentation

Maria R. D'Orsogna,[1,2] Qi Lei,[3] and Tom Chou[1,4]
[1]*Department of Biomathematics, UCLA, Los Angeles, California 90095-1766, USA*
[2]*Department of Mathematics, CSUN, Los Angeles, California 91330-8313, USA*
[3]*Institute for Computational and Engineering Sciences, University of Texas, Austin, Texas 78712-1229, USA*
[4]*Department of Mathematics, UCLA, Los Angeles, California 90095-1555, USA*

(Received 11 February 2015; accepted 15 June 2015; published online 6 July 2015)

We develop a fully stochastic theory for coagulation and fragmentation (CF) in a finite system with a maximum cluster size constraint. The process is modeled using a high-dimensional master equation for the probabilities of cluster configurations. For certain realizations of total mass and maximum cluster sizes, we find exact analytical results for the expected equilibrium cluster distributions. If coagulation is fast relative to fragmentation and if the total system mass is indivisible by the mass of the largest allowed cluster, we find a mean cluster-size distribution that is strikingly broader than that predicted by the corresponding mass-action equations. Combinations of total mass and maximum cluster size under which equilibration is accelerated, eluding late-stage coarsening, are also delineated. Finally, we compute the mean time it takes particles to first assemble into a maximum-sized cluster. Through careful state-space enumeration, the scaling of mean assembly times is derived for all combinations of total mass and maximum cluster size. We find that CF accelerates assembly relative to monomer kinetic only in special cases. All of our results hold in the infinite system limit and can be only derived from a high-dimensional discrete stochastic model, highlighting how classical mass-action models of self-assembly can fail. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4923002]

I. INTRODUCTION

Aggregation and self-assembly of particles into clusters are ubiquitous phenomena in chemistry, molecular biology, polymer, and material science, and have been extensively studied, both experimentally and theoretically.[1–5] Nearly, all theories developed thus far have employed mass-action, mean-field kinetics that describe the dynamics of the mean concentration of clusters of a given size.[6–10] Typically, only monomer attachment and detachment kinetics[2] or irreversible aggregation of clusters[1,11–13] is treated.

Mass-action models and their generalizations[14] have revealed rich behavior including metastable kinetics, gelation, the emergence of multiple time scales, and nontrivial convergence to equilibrium and coarsening. Most of these theories have focused primarily on infinite systems with unbounded cluster sizes.[4] However, in macromolecular self-assembly in cell biology,[15–17] aggregation is naturally constrained by a maximum cluster size. Moreover, kinetic theories designed to resolve discrepancies in classical nucleation theory often invoke multi-stage mechanisms across different cluster sizes and rely on understanding the formation times of finite-sized critical clusters.[14,18–22] Another desired result of self-assembly studies is an estimate of the time it takes for a full cluster to first appear, a quantity that can only be determined through a fully stochastic treatment. These estimates may be useful in determining how fast-growing protein aggregates, filaments, or viral capsids form.

Thus, to obtain a clear understanding of the overall assembly process, one must study discrete, finite-sized stochastic models of particle aggregation. While some stochastic models have been already introduced in the literature, they have typically focused only on coalescence, neglecting fragmentation.[11,12] In general, attachment-only processes are simpler to study due to their irreversible nature. Our analysis, to the best of our knowledge, is the first to consider stochastic properties of an equilibrium reversible coagulation and fragmentation (CF) process, which, in the singular limit of infinitesimally small fragmentation rates, approaches those of irreversible kinetics for finite times. Among our new results are also an analysis and enumeration of first assembly times in CF processes.

We first derive and analyze a fully stochastic model for self-assembly of discrete finite-sized systems that incorporates coagulation (the fusion of two clusters) and fragmentation (the breaking of a cluster into two smaller ones). By using numerical and analytical approaches, we find unexpected behavior in coarsening, equilibrium cluster size distributions, and first assembly times that cannot be predicted from classical mass-action models or from previous models involving only monomer attachment and detachment processes.[23,24] Our results are qualitatively universal in that they can occur at large system sizes and do not depend on microscopic CF rates.

II. STOCHASTIC CF MODEL

We begin by introducing $P(n_1, n_2, \ldots, n_N; t) \equiv P(\mathbf{n}; t)$, the probability of the system being in a state with exactly n_i clusters of size i, at time t.[23–26] We also impose a finite total mass M and a maximum cluster size N. These constraints can be

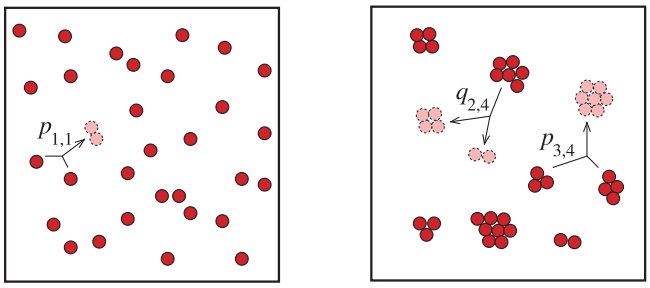

FIG. 1. Short-time and long-time snapshots of CF in a closed system with total mass $M = 30$ and maximum cluster size $N = 8$. Possible transitions are depicted in light red. CF rates are labeled by $p_{i,j}$ and $q_{i,j}$, respectively.

lifted by setting $M, N \to \infty$ in the desired order. As illustrated in Fig. 1, clusters of size i and j are allowed to coalesce with rate $p_{i,j}$ only if the merger yields a cluster of size $i + j \le N$. A cluster of size $i + j$ can also fragment into two smaller clusters of size i and j with rate $q_{i,j}$. The explicit master equation for the CF process depicted in Fig. 1 is

$$\begin{aligned}\dot{P}(\mathbf{n};t) = &-\Lambda(\mathbf{n})P(\mathbf{n};t)\\ &+\frac{1}{2}\sum_{i+j\le N} p_{i,j}(n_i+1)(n_j+1)W_i^+W_j^+W_{i+j}^-P(\mathbf{n};t)\\ &+\frac{1}{2}\sum_{i=1}^{[N/2]} p_{i,i}(n_i+1)W_i^+W_i^+W_{2i}^-P(\mathbf{n};t)\\ &+\frac{1}{2}\sum_{i+j\le N} q_{i,j}(n_{i+j}+1)W_i^-W_j^-W_{i+j}^+P(\mathbf{n};t)\\ &+\frac{1}{2}\sum_{k=1}^{[N/2]} q_{i,i}(n_{2i}+1)W_i^-W_i^-W_{2i}^+P(\mathbf{n};t),\end{aligned} \tag{1}$$

where $P(\mathbf{n};t) = 0$ if any $n_i < 0$, $[x]$ is the integer part of x, and $\Lambda(\mathbf{n})$ is the total exit rate from configuration $\mathbf{n}$ given by

$$\Lambda(\mathbf{n}) = \frac{1}{2}\sum_{i+j\le N}(p_{i,j}n_in_j + q_{i,j}n_{i+j}) - \frac{1}{2}\sum_{i=1}^{[N/2]}(p_{i,i}n_i - q_{i,i}n_{2i}). \tag{2}$$

Each term in $\Lambda(\mathbf{n})$ includes an intrinsic rate and a combinatoric factor for the number of clusters that can merge or fragment. Finally, the raising or lowering operators $W_j^{\pm}$ add or subtract one cluster of size j from state $\mathbf{n}$ so that $W_j^{\pm}P(\mathbf{n};t) = P(n_1,\ldots,n_j \pm 1,\ldots,n_N;t)$. By construction, the mass conservation constraint $\sum_{k=1}^{N} kn_k = M$ holds at all times. For simplicity, we will henceforth assume an all-monomer initial condition: $P(\mathbf{n};t=0) = \delta_{n_1,M}\delta_{n_2,0},\ldots,\delta_{n_N,0}$.

It is important to note that our "microscopic" coagulation and fragmentation rates, $p_{i,j}$ and $q_{i,j}$, are multiplied by the appropriate factors of cluster numbers n_i, n_j that determine the overall rate for two clusters to come together or for one to fall apart, respectively. Thus, any i,j-dependence in $p_{i,j}$, $q_{i,j}$ reflects microscopic details, such as size-dependent cluster shape and molecular binding energies. Our probability density and master equation describe the numbers of interacting clusters within an ensemble. This formulation is distinctly different from the typical master equation describing the stochastic evolution of the *size* of a single isolated cluster.[27]

In the appropriate state-space basis, the master equation can be written in linear form $\dot{\mathbf{P}} = \mathbf{AP}$, where $\mathbf{P}$ is a vector whose elements are the probabilities of all possible configurations and where $\mathbf{A}$ is the matrix of transition rates among them. The state-space is extremely high-dimensional and the transition matrix $\mathbf{A}$ is sparse. In fact, for $M \gg N!$, we can estimate state-space size to be $o(M^{N-1})$. In more general cases, the size of the state-space can be numerically computed by constructing an appropriate generating function that represents a constrained partition of an integer M into elements of maximum size N.[28]

In order to connect our master equation to classical, mass-action descriptions, we consider the mean number of clusters of size k, $\langle n_k(t)\rangle \equiv \sum_{\mathbf{n}} n_k P(\mathbf{n};t)$. Upon taking the appropriate sums in Eq. (1), we find, after some algebra, the first equation in a new moment hierarchy,

$$\begin{aligned}\langle \dot{n}_k\rangle = &\frac{1}{2}\sum_{i+j=k} p_{i,j}\langle n_in_j\rangle - \sum_{i=1}^{N-k} p_{i,k}\langle n_in_k\rangle + p_{k,k}\langle n_k\rangle^*\\ &+\sum_{i=1}^{N-k} q_{i,k}\langle n_{i+k}\rangle - \frac{1}{2}\sum_{i+j=k} q_{i,j}\langle n_k\rangle + q_{k,k}\langle n_{2k}\rangle^*\\ &-\frac{1}{2}p_{\frac{k}{2},\frac{k}{2}}\langle n_{\frac{k}{2}}\rangle - \frac{1}{2}q_{\frac{k}{2},\frac{k}{2}}\langle n_k\rangle.\end{aligned} \tag{3}$$

The summed terms are analogous to those arising in unbounded ($N \to \infty$) systems, while the four new unsummed terms arise as a consequence of the maximum cluster-size constraint. The last two terms arise only when k is even, while the terms with asterisks arise only if $2k \le N$. Mass conservation is preserved in Eqs. (3) which satisfies $\sum_{k=1}^{N} k\langle \dot{n}_k(t)\rangle = 0$ and the total mass is set by the initial conditions.

Upon setting the right hand side of Eq. (3) to zero and judiciously matching individual terms, we can find the conditions for detailed balance at equilibrium to hold. Note that two conditions arise: one for the merging or fragmentation of clusters of different sizes, the other for clusters of the same size. The respective conditions for detailed balance are

$$\begin{aligned}p_{i,j}n_in_jP^{\mathrm{eq}}(\mathbf{n}) &= q_{i,j}(n_{i+j}+1)W_i^-W_j^-W_{i+j}^+P^{\mathrm{eq}}(\mathbf{n}),\ i \ne j,\\ p_{i,i}\binom{n_i}{2}P^{\mathrm{eq}}(\mathbf{n}) &= q_{i,i}(n_{2i}+1)W_i^-W_i^-W_{2i}^+P^{\mathrm{eq}}(\mathbf{n}),\end{aligned} \tag{4}$$

where $P^{\mathrm{eq}}(\mathbf{n})$ represents the probability $P(\mathbf{n};t)$ at equilibrium. Detailed balance can also be applied at the level of the expected cluster size distribution $\langle n_i\rangle$ by taking averages over Eq. (4). We find

$$\begin{aligned}p_{i,j}\langle n_in_j\rangle &= q_{i,j}\langle n_{i+j}\rangle, \quad \text{for } i \ne j,\\ p_{i,i}\langle n_i(n_i-1)\rangle &= 2q_{i,i}\langle n_{2i}\rangle.\end{aligned} \tag{5}$$

The traditional mass-action equations that exploit a mean-field assumption can be easily derived in the thermodynamic limit from Eq. (3) by assuming large system sizes, $\langle n_k\rangle \gg 1$, and by using a mean-field approximation that neglects correlations, $\langle n_in_j\rangle \to \langle n_i\rangle\langle n_j\rangle \equiv c_ic_j$. Under these approximations,

Eq. (3) becomes

$$\dot{c}_k = \frac{1}{2}\sum_{i+j=k} p_{i,j}c_i c_j - \sum_{i=1}^{N-k} p_{i,k}c_i c_k + \sum_{i=1}^{N-k} q_{i,k}c_{i+k} - \frac{1}{2}\sum_{i+j=k} q_{i,j}c_k, \tag{6}$$

where $\langle n_k(t)\rangle$ is replaced by $c_k(t)$, the corresponding mass-action concentration of clusters of size k. These mean-field "mass-action" equations are widely used.[4]

Equation (6) generalizes the Becker-Döring equations for self-assembly via monomer attachment and detachment to those that involve coagulation and fragmentation of larger clusters. When only monomer attachment and detachment take place, the $M \to \infty$ limit renders the process mean-field-like since unlimited monomer numbers $n_1 \to \infty$ extinguish correlations and $\langle n_1 n_j\rangle = \langle n_1\rangle\langle n_j\rangle$. Besides being combinatorically more complex, at finite times, the mean-field limit of the CF process will not be exact in the mean cluster numbers even in the $M \to \infty$ limit. However, at equilibrium or asymptotically long times, the mean-field Becker-Döring equations do provide exact results for the mean cluster numbers, as we explicitly shown in the Appendix.

The symmetric rate matrices $p_{i,j}$ and $q_{i,j}$ are determined by the microscopic structure and free energies of cluster aggregation and fragmentation in classical nucleation models.[3,14,19] For processes involving linear polymerization,[16,20] uniform rates $p_{i,j} = p$, $q_{i,j} = q$ provide a good approximation. Other rate structures appropriate for self-assembly of globular structures can be easily motivated[29,30] and will be discussed below.

Here, we point out that the qualitative results we will describe arise not from specific details of the rate structure $p_{i,j}$, $q_{i,j}$ but from the constraints on the structure of our stochastic theory and on the state-space trajectories. Thus, without loss of universality in our results, we may consider uniform rates $p_{i,j} = p$, $q_{i,j} = q$ and rescale time in units of p^{-1}. Since the strong binding limit arises in many biomolecular processes such as viral capsid elongation[15] and actin filament assembly,[16] we further assume $q/p = \varepsilon \ll 1$.

A. Mean cluster numbers

Let us first consider a small system with $N = 8$ and $\varepsilon = 10^{-6}$. In Fig. 2, we plot $\langle n_k(t)\rangle$ and $c_k(t)$, the mean cluster numbers found through kinetic Monte Carlo (KMC) simulations[31,32] and through the numerical solution of the mean-field approximation, respectively. Fig. 2(a) shows that for $M = 24$, $\langle n_k(t)\rangle$ and $c_k(t)$ are similar at short times and evolve into an "intermediate" cluster distribution that persists up to $t \sim \varepsilon^{-1}$ before significant fragmentation and equilibration occur during late-stage "coarsening." The intermediate cluster numbers, $\langle n_k^*\rangle$ and c_k^*, can be accurately approximated by setting $\varepsilon = 0$ in the appropriate model and solving for the nontrivial steady-state. At times much greater than ε^{-1}, equilibrium values $c_k^{\rm eq}$ and $\langle n_k^{\rm eq}\rangle$ remain modestly different. Since fragmentation is much slower than coagulation, we expect that at equilibrium, to order ε, particles will be aggregated into states with the fewest total number of clusters. Given that

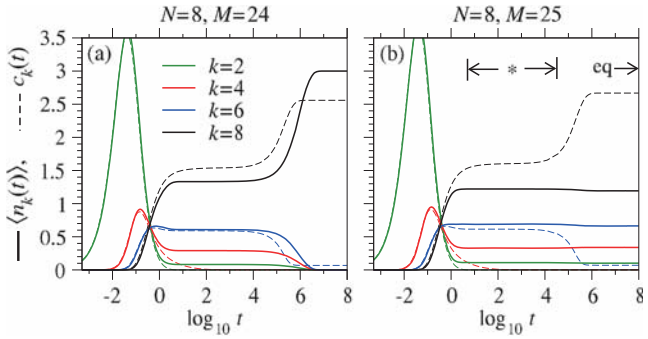

FIG. 2. Cluster numbers for the case $N = 8$ and $\varepsilon = 10^{-6}$. Solid curves represent $\langle n_k(t)\rangle$ derived from kinetic Monte Carlo (KMC) simulations, while dashed lines are the corresponding mean-field values $c_k(t)$. (a) Since total mass $M = 24$ is divisible by $N = 8$, the single dominant equilibrium (eq) configuration is that of three complete clusters, so that $\langle n_k^{\rm eq}\rangle = 3\delta_{k,8} + O(\varepsilon)$. The cluster numbers in the long-lived metastable regime (*) are composed of many other configurations. (b) When the initial monomer number is increased by just one to $M = 25$, the continuum mean-field description breaks down further and $c_k^{\rm eq}$, and $\langle n_k^{\rm eq}\rangle$ differ dramatically. At intermediate times, the metastable cluster numbers $\langle n_k^*\rangle$ nearly coincide with those at equilibrium, $\langle n_k^*\rangle \approx \langle n_k^{\rm eq}\rangle$, leading to an apparent accelerated equilibration without late-stage coarsening.

$M = 24$ divides $N = 8$, the unique fewest-cluster configuration is $(0, 0, 0, 0, 0, 0, 0, 3)$, containing three clusters of size $N = 8$. This state occurs with probability $1 - O(\varepsilon)$, while probabilities of configurations with $m > 3$ clusters scale as ε^{m-3}. From Fig. 2(a), it is clear that the stochastic approach yields $\langle n_k^{\rm eq}\rangle = 3\delta_{k,8} + O(\varepsilon)$.

Before equilibrating, the system can occupy states that are "trapped." For example, the dynamics of state $(0, 0, 1, 0, 0, 0, 3, 0)$ is very slow since the only two possible coagulation events that can occur—a trimer merging with a heptamer or two heptamers merging with each other—would violate the maximum cluster size $N = 8$ constraint. It is only through slow fragmentation ($\varepsilon \ll 1$) that the system can exit these trapped configurations and evolve towards equilibrium. As a result, there are many more configurations that contribute to the metastable regime than to the equilibrium state. Since these trapped configurations contain incomplete clusters of size $k < N$, we find a nontrivial distribution in the metastable values of $\langle n_k(t)\rangle$, where smaller clusters may be present in appreciable numbers. As we will see, these traps will also profoundly influence assembly times.

When $M = 25$, the expected cluster numbers at long times are dramatically different, as shown in Fig. 2(b). Indeed, the KMC simulations show very little difference between metastable and equilibration mean cluster concentrations. Since $M = 25$ is not divisible by $N = 8$, the most probable configurations have a total of four clusters instead of the three that arose when $M = 24$. However, several equilibrium four-cluster states exist. In addition to $(1, 0, 0, 0, 0, 0, 0, 3)$, ten others with comparable weights arise,

$$\begin{aligned} &(0,1,0,0,0,0,1,2),\ (0,0,1,0,0,1,0,2),\\ &(0,0,0,1,1,0,0,2),\ (0,1,0,0,0,0,1,2),\\ &(0,0,1,0,0,0,2,1),\ (0,0,0,1,0,1,1,1),\\ &(0,0,0,0,2,0,1,1),\ (0,0,0,0,1,2,0,1),\\ &(0,0,0,1,0,0,3,0),\ (0,0,0,0,0,3,1,0). \end{aligned} \tag{7}$$

This entropic explosion of minimal cluster states leads to a broadening of $\langle n_k^{\rm eq}\rangle$. Moreover, these configurations happen to be the same trapped states that contribute at long times to the metastable distribution. Thus, for $M = 25$, fragmentation simply introduces more pathways among the same states, slightly re-adjusting relative weights at equilibrium. It does not thin the number of configurations to a final, unique configuration (such as $(0, 0, 0, 0, 0, 0, 0, 3)$ in the $M = 24$ case). As a result, metastable and equilibration values are not too dissimilar, leading to an apparent disappearance of coarsening and acceleration towards the equilibrium distribution $\langle n_k^{\rm eq}\rangle$. An accelerated equilibration and a dispersed cluster distribution $\langle n_k^{\rm eq}\rangle$ are general features that occur in constant-rate ($p_{i,j} = p, q_{i,j} = q$) CF whenever the total mass takes on values $M = \sigma N + 1$, where σ is an integer.

As M is further increased, the number of equilibrium states decreases, generally increasing the difference between metastable and equilibrium cluster distributions. This occurs until $M = 32$, when total mass is again divisible by the maximum cluster size $N = 8$. Cyclic behaviors persist for general values of M, N, even in the case of $M, N \to \infty$, as long as their ratio stays finite.

Fig. 3(a) shows $\langle n_k^{\rm eq}\rangle$ for $N = 8$ across various values of M. Note that whenever $M = \sigma N$ is an integer multiple of N, the asymptotic solution is simply $\langle n_k^{\rm eq}\rangle = \sigma\delta_{k,N} + O(\varepsilon)$. The inclusion of only one extra monomer, so that $M = \sigma N + 1$, dramatically changes the cluster size distribution, broadening it to include smaller cluster sizes. A similar behavior was also reported for a monomer attachment and detachment self-assembly process.[32] This incommensurability effect is a consequence of the discrete nature of the finite-sized problem and does not arise in the corresponding mean-field solution. For comparison, $c_k^{\rm eq}$ is a monotonic function of M and is plotted in Fig. 3(b). Note that as N increases, the discrepancy between $c_k^{\rm eq}$ and $\langle n_k^{\rm eq}\rangle$ also increases, *even when there is perfect divisibility*. This new observation reveals how exact mean values of cluster numbers can differ qualitatively from those derived from mass-action models.

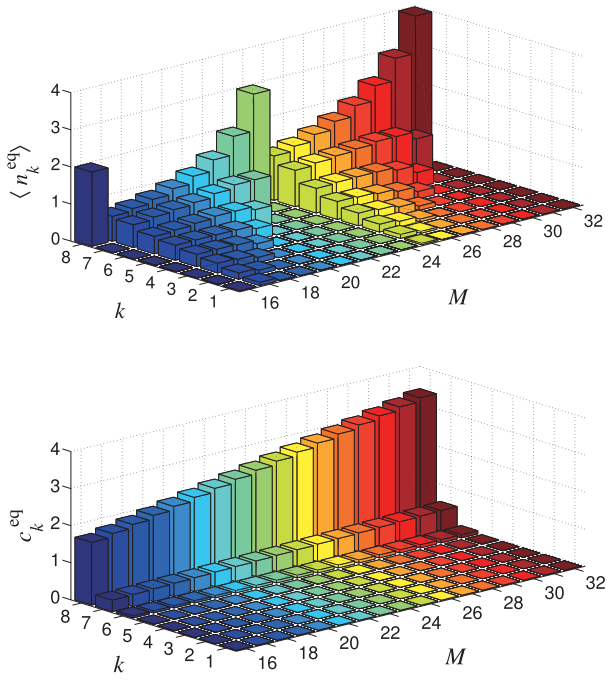

FIG. 3. Top: Equilibrium cluster sizes $\langle n_k^{\rm eq}\rangle$ of the fully discrete coagulation-fragmentation process as a function of $1 \le k \le 8$ and mass $16 \le M \le 32$. Note the incommensurability-induced broadening upon adding only one monomer to increase the mass from $M = \sigma N$ to $M = \sigma N + 1$. Bottom: Corresponding equilibrium mass-action cluster sizes $c_k^{\rm eq}$. Note the monotonic increase of $c_k^{\rm eq}$ as a function of M and the discrepancies with the stochastic results in the left panel. Due to slow convergence of the solution to Eq. (6), the discrepancy between $\langle n_k^{\rm eq}\rangle$ and $c_k^{\rm eq}$ is appreciable even when M is divisible by N. In Fig. 6 of the Appendix, we calculate and plot $\langle n_k^{\rm eq}\rangle$ and $c_k^{\rm eq}$ for $N = 20$ and $41 \le M \le 100$ to illustrate the effect for higher values of M, N.

The equilibrium values $\langle n_k^{\rm eq}\rangle$ can be evaluated analytically by considering detail-balance between the fewest-cluster "ground states" and the first "excited states" which have $O(\varepsilon)$ probability. Given M, N, we can write $M = \sigma N + j$, where $\sigma = [M/N]$ is an integer, and the remainder $0 \le j \le N - 1$. For large remainders $N/2 \le j \le N - 1$, we find that $\langle n_{k<j}(t \to \infty)\rangle = 0$ and

$$\langle n_{k\ge j}^{\rm eq}\rangle = \frac{(N-j)!\prod_{\ell=0}^{k-j-1}(\sigma+\ell)}{(k-j)!\prod_{\ell=2}^{N-j}(\sigma+\ell)}. \tag{8}$$

This formula *holds for all values of M and N*. For $1 \le j \le N/2 - 1$, an outline of the calculation for $\langle n_k^{\rm eq}\rangle$ shows that it depends on specific choices of M, N, as shown in the Appendix for $N = 5$.

B. Size-dependent $p_{i,j}$ and $q_{i,j}$

Many linear polymerization processes can be described using uniform attachment and detachment rates $p_{i,j} = p$ and $q_{i,j} = q$, as discussed above. However, some applications may require coagulation and fragmentation rates to carry an explicit dependence on cluster sizes. For instance, there may be cases where attachment rates depend on cluster surface areas. If we assume cluster volume to be proportional to the number of constituent monomers, the surface area scales as $i^{2/3}$, motivating the form $p_{i,j} \propto (ij)^{2/3}$. More generally, in studies of reversible polymerization and/or gelation events, attachment rates are often modeled via $p_{i,j} \propto (ij)^{\alpha}$, with $1/3 < \alpha \le 1$. These forms allow for faster attachment when cluster sizes are large. Some authors have also considered the alternate case $p_{i,j} = A + Bij + C(i + j)$, where A, B, C are variable prefactors. Detachment rates are often kept constant, although in certain cases, they may also depend on i, j.[12]

While we do not expect general trends to differ from the uniform case, for completeness, we have investigated several cases where $p_{i,j}, q_{i,j}$ explicitly depend on i, j. In Fig. 4, we plot the time evolution of $\langle n_k(t)\rangle$ for $M = 24$, $N = 8$ and $M = 25$, $N = 8$ under the assumption $p_{i,j} = (ij)^{2/3}$ and $q_{i,j} = 10^{-6}$. These curves are to be compared with those in Fig. 2 where $p_{i,j} = 1$ and $q_{i,j} = 10^{-6}$.

At equilibrium, we distinguish between the two cases $M = 24$ and $M = 25$. When $M = 24$, there is only one dominant equilibrium configuration, namely, $(0, 0, 0, 0, 0, 0, 0, 3)$, regardless of the form of coagulation and fragmentation rates, as long as $q_{i,j}/p_{i,j} \sim O(\varepsilon)$. The system will overwhelmingly populate the dominant configuration, leading to $\langle n_k^{\rm eq}\rangle = 3\delta_{k,8} + O(\varepsilon)$. In general, when $M = \sigma N$, there will always be one

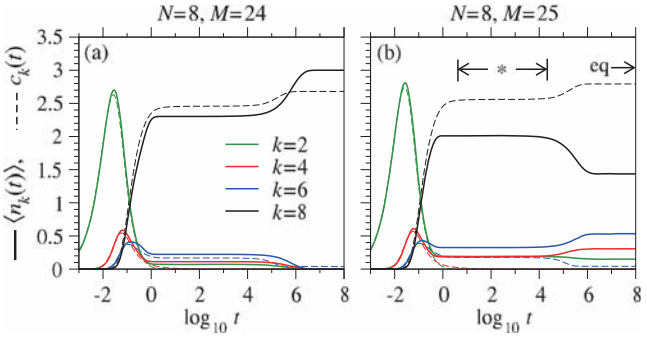

FIG. 4. Cluster numbers for $N = 8$. (a) $M = 24$ and (b) $M = 25$. Solid curves represent $\langle n_k(t)\rangle$ derived from KMC simulations, while the dashed lines are the corresponding mean-field values. Here values of $p_{i,j} = (ij)^{2/3}$ and $q_{i,j} = 10^{-6}$. The metastable (*) and fully equilibrated (eq) regimes are indicated.

dominant equilibrium configuration, so $\langle n_k^{\rm eq}\rangle = \sigma\delta_{k,N} + O(\varepsilon)$ will be insensitive to the detailed form of $q_{i,j} \ll p_{i,j}$. Indeed, for $M = 24$ as shown in Figs. 4(a) and 2(a), the equilibrium distributions $\langle n_k^{\rm eq}\rangle$ corresponding to $p_{i,j} = (ij)^{2/3}$ and $p_{i,j} = 1$, respectively, are indistinguishable. In the case of $M = 25$, when more states are present at equilibrium, their relative weights will depend on the specific choices of $p_{i,j}$ and $q_{i,j}$ in a nontrivial way. Hence, as shown in Figs. 4(b) and 2(b), the corresponding $\langle n_k^{\rm eq}\rangle$ depend on the specific choices for $p_{i,j}, q_{i,j}$. In particular, $\langle n_8^{\rm eq}\rangle$ is larger for $p_{i,j} = (ij)^{2/3}$ (Fig. 4(b)) than for $p_{i,j} = 1$ (Fig. 2(b)). Conversely, all other $\langle n_{k<8}^{\rm eq}\rangle$ are smaller for $p_{i,j} = (ij)^{2/3}$ than for $p_{i,j} = 1$. This behavior arises because $p_{i,j} = (ij)^{2/3}$ will favor assembly of larger clusters compared to $p_{i,j} = 1$.

A similar trend is observed for the metastable distribution $\langle n_k^*\rangle$. For $p_{i,j} = (ij)^{2/3}$, $\langle n_8^*(t)\rangle$ is larger and $\langle n_{k<8}^*(t)\rangle$ are smaller, compared to their values when $p_{i,j} = 1$. Since there are many configurations contributing to the metastable regime, even when $M = \sigma N$, changes in $\langle n_k^*\rangle$ as a function of $p_{i,j}$ are observed for both $M = 24$ and $M = 25$. Other choices of $p_{i,j}, q_{i,j}$ lead to similar behaviors and are not shown here. For nonuniform values of $p_{i,j}$ and $q_{i,j}$ and for small values of M, N, one can evaluate the equilibrium values analytically by using detailed balance and simple combinatoric arguments, as shown in Sec. II A for uniform p, q.

III. FIRST ASSEMBLY TIMES

In this section, we present a complete derivation of the properties of the expected first assembly times as a function of all possible initial configurations. The time to assemble the first maximal cluster is fundamentally a stochastic quantity that cannot be studied using mean-field mass-action approximations.[33] The analysis of assembly times to form the first cluster of size N requires a careful enumeration of all possible paths in state-space, which becomes increasingly complex as M, N increase.

We first derive the corresponding backward Kolmogorov equation (BKE) for the probability distribution function $P(\mathbf{n}; t|\mathbf{m}; 0)$ of being in state $\mathbf{n} = \{n_1, n_2, \ldots, n_N\}$ at time t given that the system started in the initial configuration $\mathbf{m} = \{m_1, m_2, \ldots, m_N\}$ at $t = 0$. The distribution of first assembly times can be evaluated via BKE describing the evolution of $P(\mathbf{n}; t|\mathbf{m}; 0) \equiv P(\mathbf{m}; t)$ in terms of the initial configuration $\mathbf{m}$,

$$\begin{aligned}\dot P(\mathbf{m};t) = &-\Lambda(\mathbf{m})P(\mathbf{m};t)\\ &+\frac{1}{2}\sum_{i+j\le N} p_{i,j}m_i m_j W_i^- W_j^- W_{i+j}^+ P(\mathbf{m};t)\\ &-\frac{1}{2}\sum_{i=1}^{[N/2]} p_{i,i}m_i W_i^- W_i^- W_{2i}^+ P(\mathbf{m};t)\\ &+\frac{1}{2}\sum_{i+j\le N} q_{i,j}m_{i+j} W_i^+ W_j^+ W_{i+j}^- P(\mathbf{m};t)\\ &+\frac{1}{2}\sum_{k=1}^{[N/2]} q_{i,i}m_{2i} W_i^+ W_i^+ W_{2i}^- P(\mathbf{m};t),\end{aligned} \tag{9}$$

where the operators $W_i^\pm$ now act on the indices m_i describing the *initial* configuration. The third term on the right-hand-side of Eq. (9) arises from the $-m_i$ part of the mass action term $m_i(m_i - 1)$ when $j = i$.

In vector representation, it can be shown that the BKE can be rewritten as $\dot{\mathbf{P}} = \mathbf{A}^\dagger\mathbf{P}$, where $\mathbf{A}^\dagger$ is the transpose of the transition matrix $\mathbf{A}$, as can be verified by comparing Eqs. (1) and (9). From the BKE in Eq. (9), we can determine the evolution of the survival probability, defined as $S(\mathbf{m}; t) \equiv \sum_{\{\mathbf{n}\}, n_N=0} P(\mathbf{n}; t|\mathbf{m}; 0)$, where the sum is restricted to final configurations where $n_N = 0$. $S(\mathbf{m}; t)$ thus describes the "survival" probability that no maximum cluster has yet formed up to time t, given that the system started in some configuration $\mathbf{m}$ at $t = 0$.

Upon performing the sum defining $S(\mathbf{m}; t)$ over Eq. (9), we find that $S(\mathbf{m}; t)$ also obeys the BKE in Eq. (9) with $P(\mathbf{m}; t)$ replaced by $S(\mathbf{m}; t)$, with the "boundary condition" $S(m_1, m_2, \ldots, m_N \ge 1; t) = 0$ and initial condition $S(m_1, m_2, \ldots, m_N = 0; 0) = 1$. Since the survival probabilities for a given initial condition $\mathbf{m}$ are coupled to those at other initial conditions $\mathbf{m}'$, we define the vector $\mathbf{S}(t)$ such that each component corresponds to the survival probability for a specific initial configuration. The backward equation can then be written in the form $\dot{\mathbf{S}} = \mathbf{A}^\dagger\mathbf{S}$, where the subspace of $\mathbf{A}^\dagger$ is restricted to states where no complete clusters exist. The first assembly time distribution, given any initial configuration, can be defined as

$$\mathbf{G} \equiv -\frac{\partial \mathbf{S}}{\partial t}, \tag{10}$$

where each element of $\mathbf{G}$ corresponds to the first assembly time distribution $G(\mathbf{m}; t)$ for a specific initial condition $\mathbf{m}$. The resulting *mean* assembly time $T(\mathbf{m})$ is finally given by[24]

$$T(\mathbf{m}) = \int_0^\infty t\, G(\mathbf{m}; t)\mathrm{d}t. \tag{11}$$

Upon using the definition of $G(\mathbf{m}; t)$ and integration by parts in the integral in Eq. (11), we can write

$$T(\mathbf{m}) = \int_0^\infty S(\mathbf{m}; t)\mathrm{d}t = \tilde S(\mathbf{m}; s = 0), \tag{12}$$

where $\tilde S(\mathbf{m}; s)$ is the Laplace transform of $S(\mathbf{m}; t)$. Finally, by solving the matrix equation for $\mathbf{S}$, we find

$$T(\mathbf{m}) = \int_0^\infty S(\mathbf{m}; t)\mathrm{d}t = \tilde S(\mathbf{m}; s = 0) = -[(\mathbf{A}^\dagger)^{-1}\mathbf{1}]_{\mathbf{m}}, \tag{13}$$

where the subscript $\mathbf{m}$ indicates the vector element corresponding to the configuration of the initial condition and $\mathbf{1}$ is the unit vector corresponding to the initial condition $S(\mathbf{m}; t = 0) = 1$, for all $\mathbf{m}$, with $m_N = 0$.

A simple evaluation of $T(\mathbf{m})$ using the above procedures is shown in the Appendix for $M = 5, N = 4$ and using $p_{i,j} = p, q_{i,j} = q$, with $q/p \equiv \varepsilon \ll 1$. As can be seen, regardless of the initial condition $\mathbf{m}$, the first assembly time is infinitely large as $\varepsilon \to 0$. This is due to the presence of traps, states into which the system will evolve via coagulation before a first full cluster is assembled and which can be exited only via fragmentation. Since fragmentation events are characterized by time scales of $O(1/\varepsilon)$, the presence of traps will necessarily lead to diverging mean first assembly times in the $\varepsilon \to 0$ limit.

A. Trap formation and first assembly times

Following the previous discussion, we can determine whether the mean first assembly time will diverge in the $\varepsilon \to 0$ limit for given M and N simply by determining whether traps arise. To do this, we consider the following partitioning of M:

$$M = \nu(N-1) + j. \tag{14}$$

Here, for notational simplicity, we have defined the integer $\nu \equiv [M/(N-1)]$ and j is the remainder. Note that for the purposes of this discussion, the definitions of ν and j are slightly different from those in Eq. (8) of Sec. II A and in the Appendix.

If $j \neq 1$, our trapped state is $(0, \ldots, 1, \ldots, \nu, 0)$, where the "1" is in the jth position. The cluster it represents cannot be merged with any other cluster of size $N-1$ because the maximal cluster size condition would be violated. Hence, fragmentation events of order $O(1/\varepsilon)$ are necessary to leave this trap. If $j = 1$, state $(1, \ldots, \nu, 0)$ can turn into state $(0, \ldots, \nu-1, 1)$, through a coagulation event and a trap is not generated. We thus consider the next partition with $j = 1$ and one less cluster of size $N-1$: $M = (\nu-1)(N-1) + N$. Here, the remaining N monomers can be split into two clusters of size $N-k$ and k, which can merge for all $k \geq 2$, leading to a trap-free state. For instance, state $(0, \ldots, 1, 0, \ldots, 1, 0, \ldots, \nu-1, 0)$, where the single digits refer to sites k and $N-k$, will turn into state $(0, \ldots, \nu-1, 1)$, precluding the formation of a possible trap. The next partition is $M = (\nu-2)(N-1) + 2N - 1$. In this case, there are partial clusters of size $2k-1$ and $(N-k)$, that for $k \geq 2$ cannot merge to a form complete cluster, since merging them would lead to a cluster of at least size $N+1$, exceeding the maximum size limit. Hence, we have a trapped state at $(0, \ldots, 1, \ldots, 2, \ldots, \nu-2, 0)$ where the single digit occupies position $2k-1$ and there are two clusters of size $N-k$. This trapped state exists insofar as the latter are clusters of different sizes and $N-k > 2k-1$, which for $k = 2$ is equivalent to $N > 3k-1 = 5$. Also, $\nu \geq 2$ must be satisfied for the decomposition to hold. We have thus shown that kinetic traps arise for all M, N such that $M = \nu(N-1) + j$, with $j \neq 1$ and for $M = \nu(N-1) + 1$ if $N > 5$ and $\nu \geq 2$. Note that the case $N > 5$, $\nu = 1$ and $j = 1$ is equivalent to $M = N$, in which case no traps arise.

TABLE I. Mean first assembly times of a maximum-sized cluster under CF and monomer-only kinetics. We enumerate all qualitatively different possibilities in the $\varepsilon \ll 1$ limit. Here $\nu \equiv [M/(N-1)]$ is the integer part of $M/(N-1)$.

N, M values	T, monomer	T, CF
$N = 2$	$O(1)$	$O(1)$
$N = 3, M$ odd	$O(1)$	$O(1)$
$N = 3, M$ even	$O(1/\varepsilon)$ traps	$O(1/\varepsilon)$ traps
$N = 4, M \neq 3\nu + 1$	$O(1/\varepsilon)$ traps	$O(1/\varepsilon)$ traps
$N = 4, M = 3\nu + 1$	$O(1/\varepsilon)$ traps	$O(1)$
$N \geq 5, M \neq N$	$O(1/\varepsilon)$ traps	$O(1/\varepsilon)$ traps
$N \geq 5, M = N$	$O(1/\varepsilon)$ traps	$O(1)$

We now analyze the remaining cases individually. The equivalent M for $N = 5$ and $j = 1$ is $M = 4\nu + 1$. If $\nu \geq 2$, then we can immediately identify a trap at $(0, 0, 3, \nu-2, 0)$. The choice $\nu = 1$ reduces to the specific case $M = N = 5$. Here, we can enumerate all final states and not find any traps. They are $(0, 0, 0, 0, 1)$, $(1, 0, 0, 1, 0)$, and $(0, 1, 1, 0, 0)$. Hence, for $M = N = 5$, there will be no traps. The choice $N = 4$, $M = \nu(N-1) + 1 = 3\nu + 1$ also leads to a simple enumeration process where states with the fewest number of incomplete clusters are $(1, 0, \nu, 0)$, $(0, 2, \nu-1, 0)$ which both lead to the complete state $(0, 0, \nu-1, 1)$ through a coagulation event, so that no traps arise in this case as well. The choice $N = 3$, $M = \nu(N-1) + 1 = 2\nu + 1$ implies that the sole possible states with the fewest number of incomplete clusters are of the form $(1, \nu, 0)$ which can turn into state $(0, \nu-1, 1)$ via coagulation, leading also to a trap free scenario. Finally, for $N = 2$, $M = \nu + 1$ and no traps arise as well. A table summarizing results for all possible combinations of M, N is shown in Table I.

B. Comparison between CF and monomer-only kinetics

In Ref. 24, we studied maximal size first assembly in the case of monomer activity, where clusters made of multiple particles could not merge nor could a cluster of size k fragment into any subunits other than a monomer and a cluster of size $k-1$. Evidently, the introduction of coagulation and fragmentation events allows for the presence of more paths in state-space. Here, we wish to investigate how the richer state-space connections in CF affect mean first assembly times. As illustration, consider state-space for the two small-system cases $N = 4, M = 7$ and $N = 5, M = 6$, shown in Fig. 5. When self-assembly is driven only by monomer attachment and detachment, the mean first assembly times in the $\varepsilon \to 0$ limit will be either $T \sim O(1)$ if no kinetic traps arise or $T \sim O(1/\varepsilon)$ otherwise. When coagulation and fragmentation are allowed, more paths between states are allowed as well (red, dashed lines), and one might expect that traps can be bypassed through these new transitions, leading to mean assembly times $T \sim O(1)$. This is the case for $N = 4, M = 7$ shown in Fig. 5(a). Here, when only monomer kinetics are allowed, the system can be trapped in $(1, 3, 0, 0)$ or $(0, 2, 1, 0)$, requiring slow dissociation in order to leave these states. Coagulation

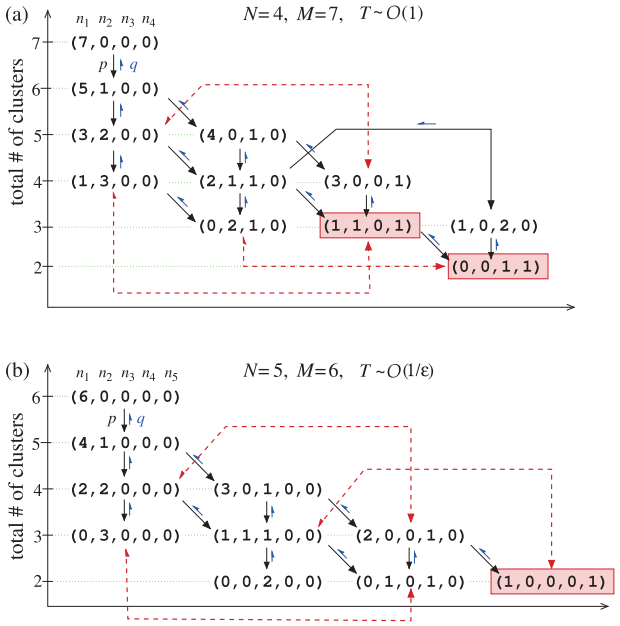

FIG. 5. (a) State-space for $M=7$, $N=4$. For $q_{i,j}/p_{k,\ell}\equiv\varepsilon\ll 1$ the equilibrium configuration is dominated by the lowest cluster number state $(0,0,1,1)$. Short arrows indicate the transitions when only monomer kinetics are allowed while red dashed lines represent the extra transitions introduced by CF which, in this case, short-circuits all kinetic traps by accelerating transitions from $(1,3,0,0)$ and $(0,2,1,0)$ to $(0,0,1,1)$, giving rise to $T\sim O(1)$. (b) For $N=5, M=6$, the only configurations that contribute to equilibrium are $(0,0,2,0,0)$,$(0,1,0,1,0)$, and $(1,0,0,0,1)$. Here, the new transitions opened by coagulation do not relieve the trap at $(0,0,2,0,0)$, and the overall mean first assembly time scales as $T\sim O(1/\varepsilon)$.

allows direct access to the maximal cluster states ($n_4=1$), leading to $T\sim O(1)$. However, in the case $N=5, M=6$, as shown in Fig. 5(b), probability can remain trapped in states such as $(0,0,2,0,0)$ and $(0,1,0,1,0)$ before reaching the only "final" state $(1,0,0,0,1)$ and $T\sim O(1/\varepsilon)$.

The complete results obtained from the enumeration process in Sec. III A and those derived from monomer activity listed in Ref. 24 are summarized in Table I and provide scaling of the mean first assembly time for all integer combinations of M and N in the slow detachment limit $q_{i,j}/p_{k,\ell}\sim O(\varepsilon)$.

Note that in some cases where monomer kinetics lead to diverging first assembly times in the $\varepsilon\to 0$ limit due to kinetic traps, CF can short-circuit these traps, dramatically accelerating the first assembly process, resulting in a finite mean first assembly time. In particular, when $M=N$, monomer attachment and detachment leads to a diverging mean assembly time, whereas CF allows to recover a finite first assembly time. As an example, we discuss the simple cases of $M=5, N=4,5$ in the Appendix, both for CF and for monomer dynamics only.

Finally, note that the inclusion of coagulation and fragmentation, which opens up many more paths among configurations, relieves kinetic traps under only two special cases: $N=M\geq 5$ and $N=4$, $M=3\nu+1$. Provided $q_{i,j}\ll p_{i,j}$, the qualitative features of the equilibrium configurations and mean assembly times that arise from structural properties such as combinatorics and kinetic traps remain unchanged from those found for constant $q\ll p$.

IV. CONCLUSIONS

Summarizing, we have derived the full master equation for a general CF process and investigated the process through KMC simulations and application detailed balance on a high-dimensional state-space. Our analyses reveal striking differences in mean cluster distributions between the full stochastic model and the associated mass-action approximation. These differences include an effectively accelerated equilibration process arising when CF is allowed; a dramatic divergence between mean-field and exact cluster numbers; and a diverging mean first assembly time that can be reduced by CF only for the enumerated cases shown in Table I. Our scaling results for the MFPT are universal in that they do not depend on the specific forms of $p_{i,j}$, $q_{i,j}$ provided the system remains in the slow detachment regime $q_{i,j}/p_{k,\ell}=\varepsilon\to 0$.

Since our discrete, finite-size stochastic model captures many features expected in cell and molecular processes, the results illustrated in this work may influence many processes including telomere clustering in the yeast nuclei,[17,34,35] filament[16] and viral capsid assembly,[36] amyloid polymerization,[18,20,37] and claritin coating of vesicles.[38,39] Moreover, the mechanisms we describe may contribute to observed effects in self-assembly, such as sample volume-dependent lag times for the formation of critical nuclei.[22]

ACKNOWLEDGMENTS

This work was supported by NSF Grant No. DMS-1021818 and ARO Grant Nos. MURI W1911NF-11-10332 and W911NF-14-1-0472.

APPENDIX: CALCULATIONAL DETAILS

In this appendix, we present details of our calculations of equilibration cluster numbers $\langle n_k^{\rm eq}\rangle$ and mean first passage times $T(\mathbf{m})$. We also motivate how our results hold for large values of M and N, and discuss properties of metastable states.

1. Equilibrium analytic results for $N=5$

Under the assumption of slow fragmentation, it is possible to derive algorithms to find asymptotic values for $\langle n_k^{\rm eq}\rangle$ for general M,N. In particular, we need to first identify the states with the lowest number of clusters—the ones that carry $O(1)$ probability—and the ones with the second lowest number of clusters—that carry $O(\varepsilon)$ probability. Once these states are found, we can derive the probabilities for the most probable lowest cluster number states by using the conditions of detailed balance between these two sets of states. We have written a Matlab code (available upon request) that will generate values of $\langle n_k^{\rm eq}\rangle$ for *general* values of M,N. The results can be written in a compact way as shown in Eq. (5) of the main text by using the notation $M=\sigma N+j$. For illustration, we consider the case $N=5$ and find the following for $\langle n_k^{\rm eq}\rangle$ for various values of j: For $j=0$,

$$\langle n_5\rangle=\sigma,\ \langle n_{k\neq 5}\rangle=0. \tag{A1}$$

For $j=4$,

$$\langle n_4\rangle=1,\ \langle n_5\rangle=\sigma\ \langle n_{k\neq 4,5}\rangle=0. \tag{A2}$$

For $j = 3$,

$$\begin{aligned}\langle n_1\rangle &= 0,\\ \langle n_2\rangle &= 0,\\ \langle n_3\rangle &= \frac{2}{\sigma+2},\\ \langle n_4\rangle &= \frac{2\sigma}{\sigma+2},\\ \langle n_5\rangle &= \frac{\sigma(\sigma+1)}{\sigma+2}.\end{aligned} \tag{A3}$$

For $j = 2$,

$$\begin{aligned}\langle n_1\rangle &= 0,\\ \langle n_2\rangle &= \frac{42}{5\sigma^2+25\sigma+42},\\ \langle n_3\rangle &= \frac{30\sigma}{5\sigma^2+25\sigma+42},\\ \langle n_4\rangle &= \frac{15\sigma(\sigma+1)}{5\sigma^2+25\sigma+42},\\ \langle n_5\rangle &= \frac{\sigma(5\sigma^2+15\sigma+22)}{5\sigma^2+25\sigma+42}.\end{aligned} \tag{A4}$$

For $j = 1$,

$$\begin{aligned}\langle n_1\rangle &= \frac{12(11\sigma+17)}{3\sigma^4+32\sigma^3+159\sigma^2+322\sigma+204},\\ \langle n_2\rangle &= \frac{12\sigma(9\sigma+11)}{3\sigma^4+32\sigma^3+159\sigma^2+322\sigma+204},\\ \langle n_3\rangle &= \frac{12\sigma(3\sigma^2+8\sigma+13)}{3\sigma^4+32\sigma^3+159\sigma^2+322\sigma+204},\\ \langle n_4\rangle &= \frac{4\sigma(3\sigma^3+14\sigma^2+30\sigma+13)}{3\sigma^4+32\sigma^3+159\sigma^2+322\sigma+204},\\ \langle n_5\rangle &= \frac{\sigma(3\sigma^4+23\sigma^3+99\sigma^2+157\sigma+54)}{3\sigma^4+32\sigma^3+159\sigma^2+322\sigma+204}.\end{aligned} \tag{A5}$$

Corresponding formulas apply for higher values of N, which increase in complexity with increasing values of j. For $N/2 \le j \le N-1$, we can introduce the compact notation shown in Eq. (4) of the main text. As can be seen, the expressions for $j = 2,3,4$ above match those in the compact formulation of Eq. (8).

2. Large values of *M,N*

In the main text, we use small values of (M,N) to illustrate the enumeration methods employed for pedagogical reasons, since larger values lead to an increasingly large state space and make graphical representations impossible. This can be already deduced from Fig. 5 where values of $M = 7$ and $N = 4$ or $M = 6$ and $N = 5$ yield a complex arrangement of states linked by a tangle of arrows. While biophysical systems of interest usually involve small to medium values of (M,N), our results are valid for *all* (M,N), including Eq. (8). Similarly, emulsification of cluster size distribution occurs periodically for all values of (M,N), with the distribution being broadest when $M = \sigma N + 1$ and the most localized when $M = \sigma N$, where σ is an integer number. This occurs regardless of how large M and N are as long as their ratio remains finite. As discussed in the main text, upon using traditional mass-action equations, emulsification does not occur for *any* combination of (M,N). Emulsification is a result of the incommensurability between (M,N) and arises only when considering the full stochastic master equation. Qualitatively, this behavior has also been observed in cases where the kinetics are restricted to monomer activity.[32] The discrepancies between stochastic and mass-action results are shown in Fig. 3 of the main text for $N = 8$. We present a similar comparison here, in Fig. 6 for the larger case of $N = 20$, to re-iterate that emulsification is not an artifact of small numbers, but arises from the constraints on state-space.

If one assumes N is finite and takes $M \to \infty$, then there is no mass constraint even at very long times and the commensurability effect due to cluster size and mass discreteness should vanish. In this limit, the mean-field Becker-Döring equations provide exact results for the mean cluster numbers. This can be shown explicitly for the simple case of uniform rates $q_{i,j} = q$, $p_{i,j} = p$ for $i \neq j$ and $p_{i,i} = 2p$. Note that for $p > q$ in the $M \to \infty$ limit, all clusters will grow indefinitely and equilibrium will never truly be reached. We thus consider a finite but very large M before applying the detailed balance Eqs. (4) to show consistency of the mean-field approximation. To begin, we assume that in the long time or equilibrium limit, the clusters are independent and that $P^{\rm eq}(\mathbf{n}) = P(n_2),\ldots,P(n_N)$. We do not include $P(n_1)$ in this *ansatz* since now $n_1 = M \to \infty$. Inserting this form into Eqs. (4), we find

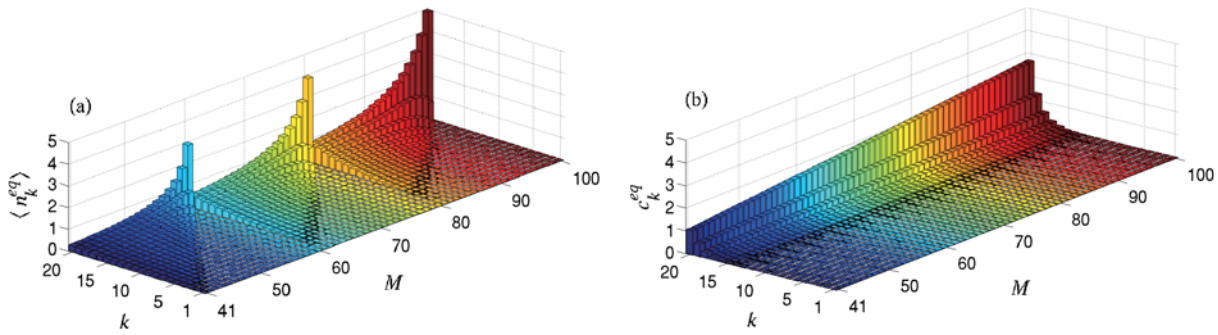

FIG. 6. (a) Equilibrium cluster sizes $\langle n_k^{\rm eq}\rangle$ as a function of $1 \le k \le 20$ and mass M for $N = 20$ and $\varepsilon \to 0$. Note the incommensurability-induced broadening upon adding only one monomer to increase the mass from $M = \sigma N$ to $M = \sigma N + 1$. (b) Corresponding equilibrium mass-action cluster sizes $c_k^{\rm eq}$. Note the monotonic increase of $c_k^{\rm eq}$ as a function of M and the large discrepancies with the stochastic results in the top panel. Note the large discrepancies even for cases in which M is divisible by N.

$$pn_i n_j P(n_i)P(n_j)P(n_{i+j}) = q(n_{i+j}+1)P(n_i-1)P(n_j-1)P(n_{i+j}+1) \quad \text{(A6)}$$

for $i \neq j$ and $i, j \neq 1$. For $i = j \neq 1$, we find

$$pn_i(n_i-1)P(n_i)P(n_{2i}) = q(n_{2i}+1)P(n_i-2)P(n_{2i}+1). \quad \text{(A7)}$$

Both relationships can be consistently solved by assuming the form

$$P(n_j) = \frac{A_j^{n_j}}{n_j!}e^{-A_j}, \quad \text{(A8)}$$

where the exponential term is a normalizing factor: since $M \to \infty$, an infinite number of clusters of size j are possible. This assumption leads to

$$pA_jA_i = qA_{i+j} \quad \text{(A9)}$$

for both $i \neq j$ and $i = j$. The detailed balance conditions coupling free monomers to each other and to clusters of size n_j for $j \neq 1$ stemming from Eqs. (4) can also be written. Using Eqs. (A8) leads to the recursion relation,

$$\begin{aligned} pM(M-1) &= qA_2, \\ p(M+1)(M+2) &= qA_2, \\ pMA_j &= qA_{j+1}, \\ p(M+1)A_j &= qA_{j+1}, \end{aligned} \quad \text{(A10)}$$

which, in the $M \to \infty$ limit, gives approximately

$$\begin{aligned} pMA_j &= qA_{i+j}, \\ pM^2 &= qA_2. \end{aligned} \quad \text{(A11)}$$

Eqs. (A9) and (A11) yield the solution

$$A_2 = \frac{q}{p}\left(\frac{pM}{q}\right)^2, \qquad A_k = \frac{q}{p}\left(\frac{pM}{q}\right)^k. \quad \text{(A12)}$$

Expressing our results in terms of $q/p \equiv \varepsilon$, we find

$$P^{\text{eq}}(n_k) = \exp\left[-\varepsilon\left(\frac{M}{\varepsilon}\right)^k\right]\frac{\varepsilon^{n_k}\left(\frac{M}{\varepsilon}\right)^{kn_k}}{n_k!}, \quad \text{(A13)}$$

and finally

$$P^{\text{eq}}(\mathbf{n}) = \exp\left[-\varepsilon\left(\frac{M}{\varepsilon}\right)^{\frac{N^2+N-2}{2}}\right]\frac{\varepsilon^{\sum_{k=2}^N n_k}\left(\frac{M}{\varepsilon}\right)^{\sum_{k=2}^N kn_k}}{\prod_{k=2}^N n_k!}.$$

These equations lead to the following estimate for the mean cluster size:

$$\langle n_k\rangle \equiv \sum_{n_k=0}^{\infty} n_k P^{\text{eq}}(n_k) = \varepsilon\left(\frac{M}{\varepsilon}\right)^k, \quad \text{(A14)}$$

which can straightforwardly be shown to satisfy Eqs. (6) with $\dot{c}_k = 0$.

3. Metastable states

Since the metastable configurations are attained before fragmentation and equilibration set in, we can evaluate the mean cluster concentrations at intermediate times by considering the irreversible coagulation process where $\varepsilon = 0$. We have derived a numerical recursion relation and implemented a Matlab code to find these values. In Fig. 7, we juxtapose the intermediate, metastable, and the final equilibrium mean cluster sizes.

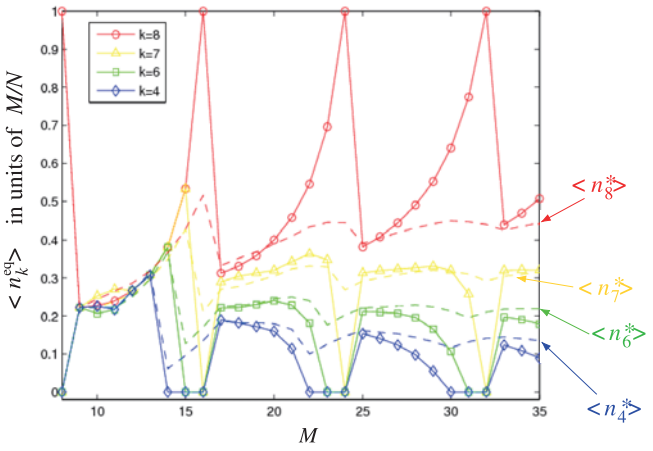

FIG. 7. A comparison between expected metastable $\langle n_k^*\rangle$ (dashed curves) and equilibrium configurations $\langle n_k^{\text{eq}}\rangle$ (solid curves with symbols) for a coagulation-fragmentation process with $N = 8$, plotted in units of M/N as functions of M. Note that equilibrium and metastable values are very similar for $M = N\sigma + 1$ with σ an integer. They diverge substantially when $M = N\sigma$ is a multiple of N.

Both the metastable and equilibrium regimes evolve with a periodicity of N. In the metastable regime, the values of $\langle n_k^*\rangle$, to first order, are given by the contribution of all states that are reached with highest probability and only due to coagulation events that may include encountering trapped states as discussed above. On longer time scales, fragmentation events may relieve traps and thin the number of possible states where mass is distributed. This effect is not felt for $M = \sigma N + 1$, where the maximal cluster size constraint does not allow the final number of states to be thinned by fragmentation. In this case, fragmentation simply creates more transitions among existing states and changes to the probability weights of these states are minor. Indeed, the smallest discrepancies between the final equilibration values of $\langle n_k^{\text{eq}}\rangle$ and the metastable values $\langle n_k^*\rangle$ are observed for $M = N\sigma + 1$. As the remainder j in the decomposition $M = N\sigma + j$ increases, fragmentation effects may cause the thinning of the final set of states, so that discrepancies between metastable and equilibrium values of the mean cluster concentrations are observed.

The largest discrepancies are seen when $M = \sigma N$ is an exact multiple of N. Here, fragmentation effects lead to a single final configuration to which all mass will aggregate into $(0,0,\ldots,\sigma)$. In this case, $\langle n_N^{\text{eq}}\rangle = \sigma + O(\varepsilon)$ and $\langle n_{k\neq N}^{\text{eq}}\rangle = O(\varepsilon)$. On the other hand, metastable configurations will still include a large set of other trapped states, leading to a broader population and to $\langle n_k^*\rangle = O(1)$ for all values of k.

4. First assembly times and traps for $M = 5$, $N = 4$

In the previous subsections and in the main text, we considered the non-dimensional fragmentation rate $q/p \equiv \varepsilon$ and the $\varepsilon \to 0$ limit. Here, for completeness, we will also look at behaviors for larger values of q/p. We thus replace ε

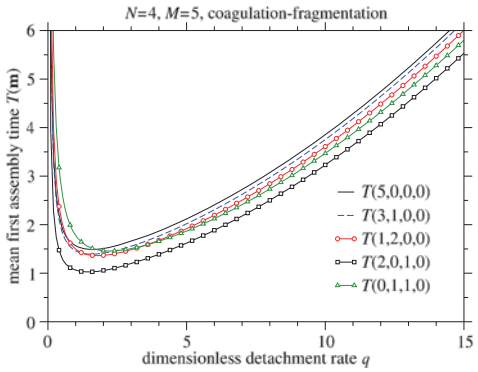

FIG. 8. First assembly times under CF for given initial configurations for $M = 5$, $N = 4$. Note the minimum in T as a function of the fragmentation rate q and the scaling $T \sim 1/q$ as $q \to 0$.

with the non-dimensional rate q/p and use the notation $q/p \equiv q$ for simplicity. The limit $\varepsilon \to 0$ is now recast as $q \to 0$.

As an example of first assembly time events, we consider the simple case $M = 5$, $N = 4$ where possible $\mathbf{m}$ states with $n_N = 0$ are $(5,0,0,0)$, $(3,1,0,0)$, $(1,2,0,0)$, $(2,0,1,0)$, $(0,1,1,0)$, and the corresponding matrix $\mathbf{A}^\dagger$ is

$$\mathbf{A}^\dagger = \begin{pmatrix} -10 & q & 0 & 0 & 0 \\ 10 & -q-6 & 2q & q & 0 \\ 0 & 3 & -3-2q & 0 & q \\ 0 & 3 & 0 & -q-3 & q \\ 0 & 0 & 2 & 1 & -2q \end{pmatrix}.$$

Upon taking the inverse of its transpose and summing the entries in each row according to Eq. (13), we find the mean first assembly times $T(\mathbf{m})$ as a function of the fragmentation rate q and plot them in Fig. 8,

$$\begin{aligned} T(5,0,0,0) &= \frac{270+624q+359q^2+54q^3+4q^4}{540q+300q^2}, \\ T(3,1,0,0) &= \frac{270+570q+329q^2+54q^3+4q^4}{540q+300q^2}, \\ T(1,2,0,0) &= \frac{90+135q+75q^2+13q^3+q^4}{135q+75q^2}, \\ T(2,0,1,0) &= \frac{90+210q+120q^2+23q^3+2q^4}{270q+150q^2}, \\ T(0,1,1,0) &= \frac{540+630q+270q^2+49q^3+4q^4}{540q+300q^2}. \end{aligned} \tag{A15}$$

From the above expressions for $T(\mathbf{m})$, it is clear that the mean first assembly time will diverge as $1/q$ for $q \to 0$. We can indeed predict the behavior of the mean first assembly time as $q \to 0$ for all choices of M, N based on simple considerations. In the case $M = 5, N = 4$, the divergence is due to the presence of a "trap" as described in the main text: a special configuration into which the system evolves before reaching any of the $n_N \geq 1$ "absorbing" states and from which it can emerge only via fragmentation. If they exist, once the system has reached any of these traps, the only way to leave them and proceed to an absorbing state is by fragmentation. The latter event, leaving a trap via fragmentation, is associated to a typical time scale of $O(1/q)$.

In the case $M = 5, N = 4$, there is only one trap $(0,1,1,0)$ because binding of the lone dimer to the lone trimer would lead to a pentamer that exceeds the maximum cluster size $N = 4$. In order to reach the sole $n_N = 1$ state possible, $(1,0,0,1)$, fragmentation from either the dimer or the trimer of state $(0,1,1,0)$ is necessary. In the first case, the intermediate state $(2,0,1,0)$ is reached before subsequent attachment of one of the two monomers to the trimer to reach $(1,0,0,1)$. In the second case, the intermediate state is $(1,2,0,0)$, from which two dimers can merge, also leading to state $(1,0,0,1)$. Thus, paths out of the trapped state $(0,1,1,0)$ will take $T \sim O(1/q)$. There are of course other trajectories to reach the absorbing state $(1,0,0,1)$ that do not pass through $(0,1,1,0)$. However, these are of order $O(1)$. Since the mean assembly time is an average over all paths, in the limit $q \to 0$, the mean will be dominated by trajectories that cross traps, yielding $\langle T \rangle \sim 1/q$. We can thus conclude that any time the chosen values of M, N lead to the creation of a trap, the mean first assembly time will scale as $O(1/q)$ for $q \to 0$, while values of M, N that do not lead to the creation of traps will result in mean first assembly times of $O(1)$.

5. First assembly times and traps for $M = 5$, $N = 5$

We now consider the case $M = 5, N = 5$ where, according to the considerations laid out in the main text, no traps should emerge and the mean first assembly time should be finite as $q \to 0$. State space is given by configurations $(5,0,0,0,0)$, $(3,1,0,0,0)$, $(1,2,0,0,0)$, $(2,0,1,0,0)$, $(0,1,1,0,0)$, and $(1,0,0,1,0)$. In this basis, the matrix $\mathbf{A}^\dagger$ is given by

$$\begin{pmatrix} -10 & q & 0 & 0 & 0 & 0 \\ 10 & -q-6 & 2q & q & 0 & 0 \\ 0 & 3 & -2q-3 & 0 & q & q \\ 0 & 3 & 0 & -q-3 & q & q \\ 0 & 0 & 2 & 1 & -2q-1 & 0 \\ 0 & 0 & 1 & 2 & 0 & -2q-1 \end{pmatrix}.$$

We can now write the inverse of its transpose and sum over the entries on each row to find the mean first assembly times $T(\mathbf{m})$ as a function of the fragmentation rate q so that

$$\begin{aligned} T(5,0,0,0,0) &= \frac{864+846q+299q^2+51q^3+4q^4}{540+270q}, \\ T(3,1,0,0,0) &= \frac{810+819q+299q^2+51q^3+4q^4}{540+270q}, \\ T(1,2,0,0,0) &= \frac{360+210q+43q^2+4q^3}{270}, \\ T(2,0,1,0,0) &= \frac{720+750q+293q^2+51q^3+4q^4}{540+270q}, \\ T(0,1,1,0,0) &= \frac{540+630q+270q^2+49q^3+4q^4}{540+270q}, \\ T(1,0,0,1,0) &= \frac{540+630q+270q^2+49q^3+4q^4}{540+270q}. \end{aligned} \tag{A16}$$

As predicted, these mean first assembly times do not diverge as $q \to 0$ as explicitly shown in Fig. 9. Also note that the mean

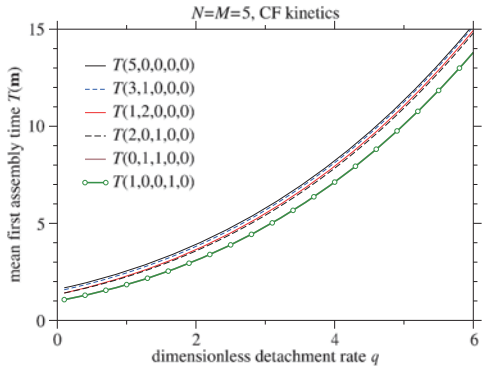

FIG. 9. First assembly times for given initial configurations in the case $M = 5$, $N = 5$ under CF dynamics. Note the monotonic behavior as a function of q. In this case, there are no traps when $q \to 0$. Different colors correspond to different initial conditions.

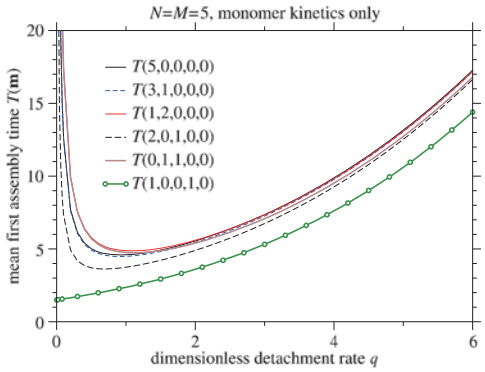

FIG. 10. First assembly times for given initial configurations in the case $M = 5$, $N = 5$ when only monomer attachment and detachment kinetics are allowed. For most initial conditions, there is non-monotonic behavior as the $(0, 1, 1, 0, 0)$ trap is encountered as $q \to 0$. The only initial condition that avoids the $(1, 0, 0, 1, 0)$ trap state, leading to a finite mean first assembly time as shown by the green circle line and a monotonic dependence of $T(1, 0, 0, 1, 0)$ on q.

first assembly time is the same for the two different initial conditions $(0, 1, 1, 0, 0)$ and $(1, 0, 0, 1, 0)$. This is because, as one can explicitly verify from the last two column and last two row entries on the matrix $\mathbf{A}^\dagger$ above, all pathways in and out of the above states collectively carry the same weight. Interestingly, if we revert to the monomer activity case only and disallow transitions from state $(0, 1, 1, 0, 0)$ to state $(0, 0, 0, 0, 1)$ and between states $(1, 2, 0, 0, 0)$ and $(1, 0, 0, 1, 0)$, we recover a non-monotonic behavior, indicative of kinetic trapping, in this case due to the configuration $(0, 1, 1, 0, 0)$. Using the same basis as above, the matrix $\mathbf{A}_m^\dagger$ in the case of monomer activity only is given by

$$\mathbf{A}_m^\dagger = \begin{pmatrix} -10 & q & 0 & 0 & 0 & 0 \\ 10 & -q-6 & 2q & q & 0 & 0 \\ 0 & 3 & -2q-2 & 0 & q & 0 \\ 0 & 3 & 0 & -q-3 & q & q \\ 0 & 0 & 2 & 1 & -2q & 0 \\ 0 & 0 & 0 & 2 & 0 & -q-1 \end{pmatrix},$$

from which the mean first assembly times are derived as

$$\begin{aligned} T(5,0,0,0,0) &= \frac{60 + 211q + 217q^2 + 68q^3 + 12q^4 + q^5}{60q(1+q)}, \\ T(3,1,0,0,0) &= \frac{60 + 205q + 211q^2 + 68q^3 + 12q^4 + q}{60q(1+q)}, \\ T(1,2,0,0,0) &= \frac{90 + 225q + 210q^2 + 68q^3 + 12q^4 + q^5}{60q(1+q)}, \\ T(2,0,1,0,0) &= \frac{30 + 135q + 55q^2 + 11q^3 + q^4}{60q}, \\ T(0,1,1,0,0) &= \frac{90 + 225q + 200q^2 + 67q^3 + 12q^4 + q^5}{60q(1+q)}, \\ T(1,0,0,1,0) &= \frac{90 + 45q + 10q^2 + q^3}{60}. \end{aligned} \tag{A17}$$

Note that while they were identical under CF, $T(0, 1, 1, 0, 0)$ and $T(1, 0, 0, 1, 0)$ are now different. This is because compared to CF, monomer activity restricts the number of paths in and out of these states making them no longer equivalent, as can also be seen by the last two columns of the matrix $\mathbf{A}_m^\dagger$. We plot the mean first assembly times for monomer activity in Fig. 10. Upon comparing the latter and Fig. 9, we note the non-monotonic behavior in Fig. 10 as $q \to 0$ due to the presence of the kinetic trap $(0, 1, 1, 0, 0)$. The trap is bypassed in CF where a dimer and a trimer can coalesce, whereas under monomer activity only this is not possible. The only initial condition that is able to escape the trap is $(1, 0, 0, 1, 0)$, for which indeed, the mean first assembly time is finite as $q \to 0$. We also note that mean first assembly times are faster under CF than under monomer activity, as CF events allow for more pathways to emerge towards a completed cluster.

Under CF, as we increase q for small values of q, the mean first assembly times shown in Fig. 8 and in most curves of Fig. 10 decrease. In general, a larger q leads to a more rapid dissociation which leads to expect longer assembly times. On the other hand, due to the multiple pathways to cluster completion, increasing q actually allows for more mixing among them, so that at times, fragmentation allows the system to return to more favorable paths, leading to shorter first assembly times. This is the effect at play for $q = 0$ when, as we have seen, due to the presence of traps, the first assembly time diverges and where, upon raising the fragmentation rate q to a non-zero value, the first assembly time becomes finite. Fragmentation thus allows the system to visit paths that lead to absorbed states, which would otherwise not be accessible. Increasing q for larger values of q however does not significantly change the availability of paths towards cluster completion so that $T(\mathbf{m})$ will eventually increase with q. Taken together, these two trends give rise to a fragmentation rate q^* where the first assembly time is shortest. These trends persist for all cases where the choice of M, N leads to traps. A minimum in mean first assembly times is not observed for trap-free choices of M, N, where instead, $T(\mathbf{m})$ is monotonically increasing with q, as seen in Fig. 9 when $M = N = 5$ and in Fig. 10 for the $(1, 0, 0, 1, 0)$ initial condition.

[1]M. von Smoluchowski, "Drei vorträge über diffusion Brownsche molekularbewegung und koagulation von kolloidteichen," Phys. Z. **17**, 557–585 (1916).

[2]R. Becker and W. Döring, "Kinetische behandlung der keimbildung in übersättigten dämpfen," Ann. Phys. **24**, 719–752 (1935).
[3]K. Binder and D. Stauffer, "Statistical theory of nucleation, condensation and coagulation," Adv. Phys. **25**, 343–396 (1976).
[4]P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, 2010).
[5]J. Grant, R. L. Jack, and S. Whitelam, "Analyzing mechanisms and microscopic reversibility of self-assembly," J. Chem. Phys. **135**, 214505 (2011).
[6]O. Penrose, "The Becker-Döring equations at large times and their connection with the LSW theory of coarsening," J. Stat. Phys. **89**, 305–320 (1997).
[7]J. A. D. Wattis and J. R. King, "Asymptotic solutions of the Becker-Döring equations," J. Phys. A: Math. Gen. **31**, 7169–7189 (1998).
[8]J. A. D. Wattis, "An introduction to mathematical models of coagulation-fragmentation processes: A discrete deterministic mean-field approach," Physica D **222**, 1–20 (2006).
[9]P. Smereka, "Long time behavior of a modified Becker-Döring system," J. Stat. Phys. **132**, 519–533 (2008).
[10]T. Chou and M. R. D'Orsogna, "Coarsening and accelerated equilibration in mass-conserving heterogeneous nucleation," Phys. Rev. E **84**, 011608 (2011).
[11]A. H. Marcus, "Stochastic coalescence," Technometrics **10**, 133–143 (1968).
[12]D. J. Aldous, "Deterministic and stochastic models for coalescence (aggregation and coagulation): A review of the mean-field theory for probabilists," Bernoulli **5**, 3–48 (1999).
[13]M. Mobilia, P. L. Krapivsky, and S. Redner, "Kinetic anomalies in addition-aggregation processes," J. Phys. A: Math. Gen. **36**, 4533–4542 (2003).
[14]J. F. Lutsko and M. A. Duran-Olivencia, "Classical nucleation theory from a dynamical approach to nucleation," J. Chem. Phys. **138**, 244908 (2013).
[15]P. van der Schoot and R. Zandi, "Kinetic theory of virus capsid assembly," Phys. Biol. **4**, 296–304 (2007).
[16]L. Edelstein-Keshet and G. B. Ermentrout, "Models for the length distributions of actin filaments: I. Simple polymerization and fragmentation," Bull. Math. Biol. **60**, 449–475 (1998).
[17]N. Hozé and D. Holcman, "Modeling capsid kinetics assembly from the steady state distribution of multi-sizes aggregates," Phys. Lett. A **378**, 531–534 (2014).
[18]C.-C. Lee, A. Nayak, A. Sethuraman, G. Belfort, and G. J. McRae, "A three-stage kinetic model of amyloid fibrillation," Biophys. J. **92**, 3448–3458 (2007).
[19]P. G. Vekilov, "A two-step mechanism of nucleation of crystals in solution," Nanoscale **2**, 2346–2357 (2010).
[20]M.-T. Alvarez-Martinez, P. Fontes *et al.*, "Dynamics of polymerization shed light on the mechanisms that lead to multiple amyloid structures of the prion protein," Biochim. Biophys. Acta **10**, 1305–1317 (2011).
[21]P. A. Korevaar, C. Grenier *et al.*, "Model-driven optimization of multi-component self-assembly process," Proc. Natl. Acad. Sci. U. S. A. **110**, 17205–17210 (2013).
[22]J. Szavits-Nossan, K. Eden *et al.*, "Inherent variability in the kinetics of autocatalytic protein self-assembly," Phys. Rev. Lett. **113**, 098101 (2014).
[23]J. S. Bhatt and I. J. Ford, "Kinetics of heterogeneous nucleation for low mean cluster populations," J. Chem. Phys. **118**, 3166–3176 (2003).
[24]R. Yvinec, M. R. D'Orsogna, and T. Chou, "First passage times in homogeneous nucleation and self-assembly," J. Chem. Phys. **137**, 244107 (2012).
[25]S. Gueron, "The steady-state distribution of coagulation–fragmentation processes," J. Math. Biol. **37**, 1–27 (1998).
[26]M. R. D'Orsogna, B. Zhao, B. Berenji, and T. Chou, "Combinatoric and mean-field analysis of heterogeneous self-assembly," J. Chem. Phys. **139**, 121918 (2013).
[27]D. Kashchiev, *Nucleation: Basic Theory with Applications* (Butterworth Heinemann, Oxford, 2000).
[28]W. Watkins, "Generating functions," Coll. Math. J. **18**, 195–211 (1987).
[29]P. G. J. van Dongen and M. H. Ernst, "Kinetics of reversible polymerization," J. Stat. Phys. **37**, 301–324 (1984).
[30]M. Thorn, H. P. Breuer, F. Petruccione, and J. Honerkamp, "A master equation investigation of coagulation reactions: Sol-gel transition," Macromol. Theory Simul. **3**, 585–599 (1994).
[31]A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, "A new algorithm for Monte-Carlo simulation of Ising spin systems," J. Comput. Phys. **17**, 10–18 (1975).
[32]M. R. D'Orsogna, G. Lakatos, and T. Chou, "Stochastic self-assembly of incommensurate clusters," J. Chem. Phys. **136**, 084110 (2012).
[33]X. Li and A. Kolomeisky, "Mechanisms and topology determination of complex chemical biological network systems from first-passage theoretical approach," J. Chem. Phys. **139**, 114106 (2013).
[34]N. Hozé and D. Holcman, "Coagulation-fragmentation for a finite number of particles and application to telomere clustering in the yeast nucleus," Phys. Lett. A **376**, 845–849 (2012).
[35]N. Hozé *et al.*, "Spatial telomere organization and clustering in yeast *Saccharomyces cerevisiae* nucleus is generated by a random dynamics of aggregation-dissociation," Mol. Biol. Cell **24**, 1791–1800 (2013).
[36]A. Yu. Morozov, R. F. Bruinsma, and J. Rudnick, "Assembly of viruses and the pseudo-law of mass action," J. Chem. Phys. **131**, 155101 (2009).
[37]T. Pöschel, N. V. Brilliantov, and C. Frömmel, "Kinetics of prion growth," Biophys. J. **85**, 3460–3474 (2003).
[38]B. Shraiman, "On the role of assembly kinetics in determining the structure of clathrin cages," Biophys. J. **72**, 953–957 (1997).
[39]L. Foret and P. Sens, "Kinetic regulation of coated vesicle secretion," Proc. Natl. Acad. Sci. U. S. A. **105**, 14763–14768 (2008).