Initial condition of stochastic self-assembly

Jason K. Davis and Suzanne S. Sindi

Phys. Rev. E **93**, 022109 — Published 5 February 2016

DOI: 10.1103/PhysRevE.93.022109

On the Initial Condition of Stochastic Self Assembly

Jason K. Davis* and Suzanne S. Sindi†

University of California, Merced

5200 N Lake Rd, Merced, CA 95343

(Dated: December 23, 2015)

Abstract

The formation of a stable protein aggregate is regarded as the rate limiting step in the establishment of prion diseases. In these systems, once aggregates reach a critical size the growth process accelerates and thus the waiting time until the appearance of the first critically-sized aggregate is a key determinant of disease onset. In addition to prion diseases, aggregation and nucleation is a central step of many physical, chemical and biological process. Previous studies have examined the first-arrival time at a critical nucleus size during homogeneous self-assembly under the assumption that at time $t = 0$ the system was in the all-monomer state. However, in order to compare to *in vivo* biological experiments where protein constituents inherited by a newly born cell likely contain intermediate aggregates, other possibilities must be considered. We consider one such possibility by conditioning the unique ergodic size distribution on sub-critical aggregate sizes – this "least-informed" distribution is then used as an initial condition. We make the claim that this initial condition carries fewer assumptions than an all-monomer one and verify that it can yield significantly different, averaged waiting times relative to the all-monomer condition under various models of assembly.

* jdavis8@ucmerced.edu

† ssindi@ucmerced.edu

I. INTRODUCTION

The self-assembly of particles into aggregates is fundamental to many physical, chemical, and biological processes [1–3]. In particular, protein aggregation is critical for prion diseases which encompass a number of fatal neurodegenerative diseases in mammals such as Creutzfeldt-Jacob and Kuru in humans, bovine spongiform encephalopathies (BSE) in cows, scrapie in sheep, and chronic wasting in elk [4, 5]. In these diseases a misfolded form of the protein appears and forms aggregates. These small aggregates of misfolded protein are thought to reproduce slowly until reaching a critical size (nucleation), at which point the aggregates are able more efficiently amplify [6–9]. As such, a key limiting step in the onset of prion diseases is the waiting time until the formation of this nucleus, which we will also call the assembly time. Beyond mammalian diseases, prions have been associated with a number of harmless heritable phenotypes in yeast [10, 11]. Both the harmless nature of the prions and the experimental tractability of yeast have made it an ideal model system to study the appearance of prion disease *in vivo*, and whose considerations will in turn affect our choice of *in silico* models.

Most prior models for computing the waiting times until nucleus formation have assumed that the initial condition of the stochastic process is the all monomer state. That is, no aggregates of any kind exist at time $t = 0$. However, for an *in vivo* assembly process, the system under study has presumably existed for some amount of time prior to the start of the experiment – time during which the assembly process has been active. In particular, since newly born yeast daughter cells inherit protein constituents from their mothers, it is probable that both monomers and small aggregates are transmitted. Therefore, there should be some uncertainty in the specific configuration of polymer sizes at time $t = 0$.

In this manuscript, we take a step towards more accurate modeling of nucleation in prion diseases by considering protein aggregation under more realistic initial conditions that incorporate this uncertainty in a "least-informed" manner. We first give an overview of common models used to study molecular assembly and then establish the least-informed distribution and its relationship to these models' ergodic distributions. We claim that this initial condition imposes fewer assumptions on the physical system, roughly agreeing with the all-monomer condition when assembly is rare but subsequently favoring larger polymer sizes as assembly becomes increasingly favored over disassembly. We demonstrate that the

mean assembly time is particularly sensitive to the initial condition for this latter case.

II. MODELING MOLECULAR ASSEMBLY

Mathematical models of assembly have been studied for a century, and in particular two systems have received considerable attention: the Smoluchowski coagulation system [12] and the Becker-Döring equations [13]. The former system (with its generalization in [14]) models the coagulation (or coalescence) and fragmentation of polymers consisting of monomeric units, represented by the chemical equations

$$\underbrace{X_j + X_{i-j}}_{j=1,2,\ldots,i-1} \xrightarrow{k^+_{j,i-j}} X_i \xrightarrow{k^-_{k,i-k}} \underbrace{X_k + X_{i-k}}_{k=1,2,\ldots,i-1}, \tag{1}$$

where X_i denotes a polymer (or aggregate or cluster) of size i. We refer to this model more generally as the discrete coagulation-fragmentation model. The Becker-Döring model restricts the reactions to just monomer polymerization and depolymerization, represented by the chemical equations

$$X_1 + X_i \underset{k^-_{i+1}}{\overset{k^+_i}{\rightleftharpoons}} X_{i+1}. \tag{2}$$

Both models are depicted graphically in Figures 1 and 2. From these chemical equations, mathematical equations may be derived with additional assumptions. For example, use of the Law of Mass Action and the corresponding rate equations have been thoroughly studied for these models [13–17]. However, this approximation requires a large number of particles to be valid and fails to capture mesoscopic effects [18]. When these effects are non-negligible, one must instead assume a Markov property and obtain a continuous-time Markov chain. We may then study the first-passage time for the subset of states containing a nucleus (a polymer of size at least c, for some c). In particular, we will compute the mean first-passage (assembly) time denoted μ.

For any differential or stochastic system an initial condition (IC) must be specified. We note that in all of the cited works, the IC has exclusively been the all-monomer state, which contains only X_1 particles. Though convenient, it is not realistic for *in vivo* processes. Indeed, with our yeast prion model, cells transfer material from mother to daughter during budding, including partially formed aggregates [19, 20]. Furthermore, within a single cell, these processes of protein assembly and disassembly are always occurring, even before the

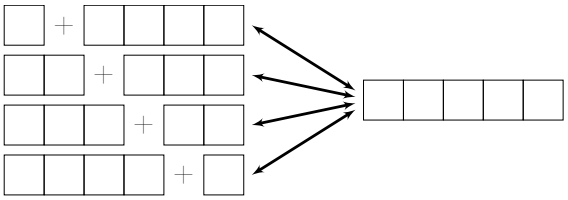

FIG. 1. A graphical depiction of Coagulation-Fragmentation type assembly. Under this assembly model polymers of any size may join and fragmentation may occur between any two adjacent, constituent monomers.

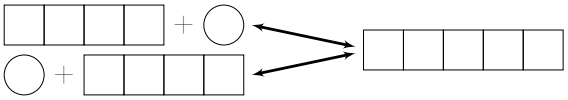

FIG. 2. A graphical depiction of Becker-Döring type assembly. Under this assembly model only monomer polymerization and depolymerization may occur.

"experimental clock" begins. Thus, in many experimental settings we believe it more accurate to think of the IC as being sampled from some distribution of states which reflects our uncertainty regarding the precise configuration of the polymers.

III. LEAST-INFORMED DISTRIBUTION FOR STOCHASTIC ASSEMBLY

Let us define the set of all mass-preserving polymer configurations $\Omega_m = \{\boldsymbol{n} : \sum_{i>0} i n_i = m\}$, where each $\boldsymbol{n} \in \Omega_m$ defines a system state with total mass m. A particular model of assembly will induce a directed graph on this set of states with edges weighted by the rates of transition between states. Thus, while the numbers and sizes of polymers may change over time, they do so in a way that preserves the total mass m of the system.

We will require that this directed graph be strongly connected and aperiodic, implying the corresponding Markov chain will asymptotically reach a unique ergodic distribution over Ω_m [21]. It is this ergodic distribution that we claim is a more natural IC, provided that for some positive integer c of interest, we condition it on there being no polymers of size at least c. We have assumed that once a polymer with size at least c (nucleus) is formed, other more rapid processes act upon it [6–9] – thus our nucleation models are only valid when no such particles exist. To accommodate this, we simply zero the respective probabilities from the ergodic distribution and renormalize, reflecting our certainty that no nuclei are

present, but otherwise providing a prior of sorts on the distribution of the possible polymer configurations. We refer to this conditional distribution as "least informed" for this reason – we assume only that the system has existed long enough to deviate substantially from any particular initial state, but it has yet to assemble a particle of size at least c.

We evaluate the impact of the IC by studying the mean assembly time of the stochastic process μ and its dependency on the initial polymer configuration. We next provide novel, closed-form results on the ergodic distributions of constant-rate Coagulation-Fragmentation and Becker-Döring assembly models, use these results to derive the corresponding least-informed distributions, and then verify that significant differences in μ between the least-informed and all-monomer IC arise in particular parameter regimes.

A. Notation

We use standard notation for the ith canonical basis vector $\boldsymbol{e}_i$ and the Kronecker delta function δ_{ij}. We further denote the total mass of the system by m, the set of all mass-preserving configurations by Ω_m, the critical nucleus size by c, and we let $\mathcal{P} : [0, \infty) \times \Omega_m$ denote the solution to an appropriately-defined master equation, where $\mathcal{P}(t, \boldsymbol{n}) = \Pr[N_1(t) = n_1, N_2(t) = n_2, \dots]$ (such that $\boldsymbol{n} \in \Omega_m$ and the random variable $N_i(t)$ is understood to represent the number of X_i particles at time t). The use of the master equation is convenient since the initial condition is itself a distribution – the all-monomer IC corresponds to a point-mass, or degenerate distribution centered about a single state, while our least-informed distribution will be more generally some non-negative vector with entries summing to 1.

The mean assembly time of our stochastic process can be estimated using kinetic Monte Carlo simulations, but it is also computable via the solution of a high-dimensional, linear relationship stemming directly from the master equation (refer to Yvinec *et al.* [22] for details). This latter technique, which we employ for our figures, yields a vector of mean assembly times with entries corresponding to each discrete polymer configuration; the inner product of this vector with an initial condition will then yield the overall mean assembly time via the law of total expectation.

B. Coagulation-Fragmentation Model

A coagulation-fragmentation model allows polymers of any size to freely coagulate (or coalesce) as well as fragment into any size. When these rates are assumed to be constant (size-independent) the reactions are succinctly represented:

$$\underbrace{X_j + X_{i-j}}_{j=1,2,\ldots,i-1} \xrightarrow{\beta} X_i \xrightarrow{\gamma} \underbrace{X_k + X_{i-k}}_{k=1,2,\ldots,i-1}. \tag{3}$$

We assume polymers may coagulate on either end; however, we have double-counted pairs (X_j, X_{i-j}) and (X_{i-j}, X_j). Thus, the unordered pair (X_j, X_{i-j}) coagulates at rate 2β to form X_i. We define the operators $W_{i,j}^{\pm}$ such that $W_{i,j}^{\pm}\mathcal{P}(t, \boldsymbol{n}) = \mathcal{P}(t, \boldsymbol{n} \pm (\boldsymbol{e}_i + \boldsymbol{e}_j - \boldsymbol{e}_{i+j}))$ and write the master equation:

$$\begin{aligned}\frac{d\mathcal{P}}{dt} = &-\beta \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} n_i(n_j - \delta_{ij})\mathcal{P} - \gamma \sum_{i=1}^{\infty}(i-1)n_i\mathcal{P} \\ &+ \beta \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(n_i + 1 + \delta_{ij})(n_j + 1)W_{i,j}^{+}\mathcal{P} \\ &+ \gamma \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(n_{i+j} + 1)W_{i,j}^{-}\mathcal{P}.\end{aligned} \tag{4}$$

It may be verified by inspection that $\mathcal{P}(\boldsymbol{n}) \propto 1/\prod(\beta/\gamma)^{n_i} n_i!$ is a steady-state solution to (4). Since the ergodic distribution is unique, we conclude that Equation (4) asymptotically converges to this solution. This gives us the least-informed distribution:

$$\Pr\left[\boldsymbol{N} = \boldsymbol{n} \middle| N_i = 0, \forall\, i \geq c\right] = \frac{1}{\prod_{i<c}(\beta/\gamma)^{n_i} n_i!} \Bigg/ \sum_{\boldsymbol{n'}} \frac{1}{\prod_{i<c}(\beta/\gamma)^{n'_i} n'_i!}. \tag{5}$$

Note that by restricting our least-informed distribution to cases where $N_i = 0$ for all $i \geq c$ we have ensured that our distribution only includes states where all particles are smaller than the critical size c. We use this distribution as the IC for the master equation corresponding to the assembly process and compute the mean assembly time (μ) to any state containing a nucleus of size at least c. The non-dimensionalized mean assembly time $\beta\mu$ is shown for various parameter configurations in Figures 3 and 4. We note that when β/γ is small, fragmentation is favored relative to aggregation, and the mean assembly time for the all monomer IC and the least-informed IC agree. As β/γ increases, the mean assembly times separate.

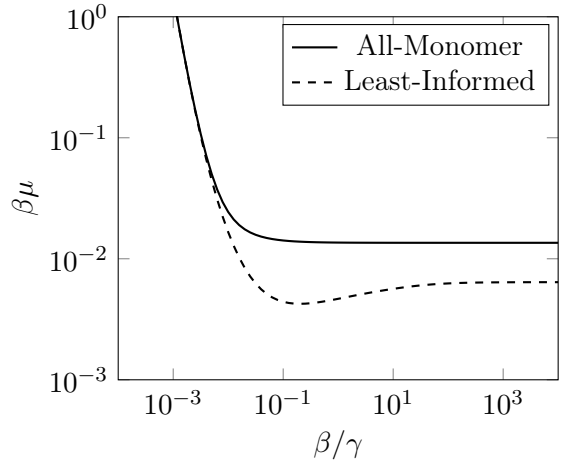

FIG. 3. Non-dimensionalized mean assembly times ($\beta\mu$) for Coagulation-Fragmentation with $c = 5$ and $m = 50$.

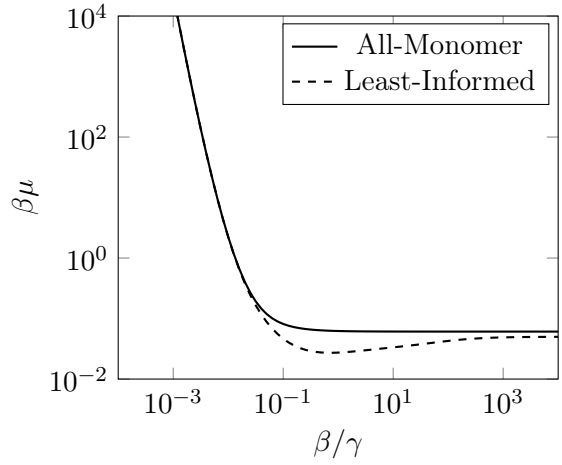

FIG. 4. Non-dimensionalized mean assembly times ($\beta\mu$) for Coagulation-Fragmentation with $c = 7$ and $m = 30$.

C. Becker-Döring Model

The Becker-Döring model with constant rate coefficients is described by the chemical equations

$$X_1 + X_i \underset{2\gamma}{\overset{2\beta}{\rightleftharpoons}} X_{i+1}. \tag{6}$$

They describe an assembly mechanism where monomers attach (polymerize) to a polymer-end with rate β and detach (depolymerize) with rate γ.

Defining the operators $W_i^{\pm}\mathcal{P}(t, \boldsymbol{n}) = \mathcal{P}(t, \boldsymbol{n} \pm (\boldsymbol{e}_1 + \boldsymbol{e}_i - \boldsymbol{e}_{i+1}))$, the corresponding master

equation is

$$\frac{d\mathcal{P}}{dt} = -\beta n_1(n_1-1)\mathcal{P} - 2\beta n_1 \sum_{i=2}^{\infty} n_i \mathcal{P} - 2\gamma \sum_{i=2}^{\infty} n_i \mathcal{P} \\ + \beta(n_1+2)(n_1+1)W_1^+\mathcal{P} + 2\beta(n_1+1)\sum_{i=2}^{\infty}(n_i+1)W_i^+\mathcal{P} + 2\gamma\sum_{i=1}^{\infty}(n_{i+1}+1)W_i^-\mathcal{P}. \tag{7}$$

One may verify that $\mathcal{P}(\boldsymbol{n}) \propto 2^{n_1}/\prod(2\beta/\gamma)^{n_i}n_i!$ is a steady-state solution to (7). Then, as before, we obtain the least-informed distribution

$$\Pr\left[\boldsymbol{N} = \boldsymbol{n} \middle| N_i = 0, \forall\, i \geq c\right] = \frac{2^{n_1}}{\prod_{i<c}(2\beta/\gamma)^{n_i}n_i!} \Bigg/ \sum_{\boldsymbol{n}'} \frac{2^{n_1'}}{\prod_{i<c}(2\beta/\gamma)^{n_i'}n_i'!}. \tag{8}$$

We plot comparisons of the mean first-passage time for this distribution with the all-monomer IC across various parameters in Figures 5 and 6. As with the Coagulation-Fragmentation model, we observe that when β/γ is small, the mean assembly time for the all monomer IC and the least-informed IC agree. As β/γ increases, the mean assembly times separate; however, unlike the Coagulation-Fragmentation model, in the Becker-Döring assembly model the IC with the fastest mean assembly depends on the critical size c and mass of the system m.

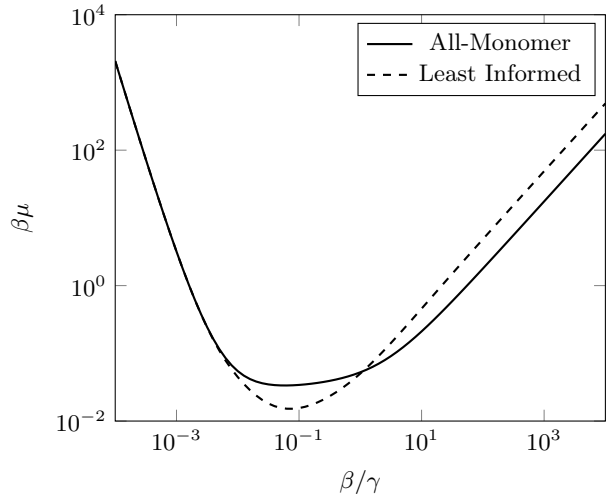

FIG. 5. Non-dimensionalized mean assembly times ($\beta\mu$) for Becker-Döring with $c = 5$ and $m = 50$.

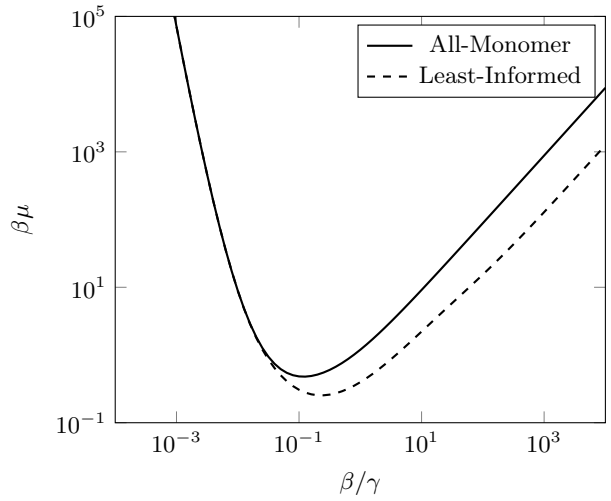

FIG. 6. Non-dimensionalized mean assembly times ($\beta\mu$) for Becker-Döring with $c = 7$ and $m = 30$.

IV. DISCUSSION

We see in Figures 3-6 that when γ is large relative to β, there is virtually no difference between the standard all-monomer IC and our least-informed distribution. This is because for both models we studied, as β/γ approaches 0, the all-monomer state becomes increasingly probable: our IC is effectively no different than the all-monomer IC. However, as $\beta/\gamma \to \infty$, the least-informed distributions begin favoring configurations that minimize the total number of polymers, assigning higher probability mass to configurations with larger polymers. This in turn has a potentially substantial effect on the mean assembly time, which is demonstrated in Figures 7 and 8.

This dependence of the initial condition on the dynamics of the specific assembly process is a powerful feature of our least-informed framework. Unilaterally applying an all-monomer IC for *in vivo* processes requires the *a priori* assumption that assembly is slow. Our initial condition replicates that assumption when it is indeed rare, but also provides a physically meaningful result when it is not.

In particular, for models where assembly is restricted to particular sizes (e.g. the Becker-Döring where aggregation proceeds only through monomer addition), this can have a dramatic effect on the overall time to nucleation. When assembly is favored, the monomer population may deplete before a critically-sized polymer has formed, leading to a "trap" of sorts where no further assembly is possible until some fragmentation has occurred [22]. This

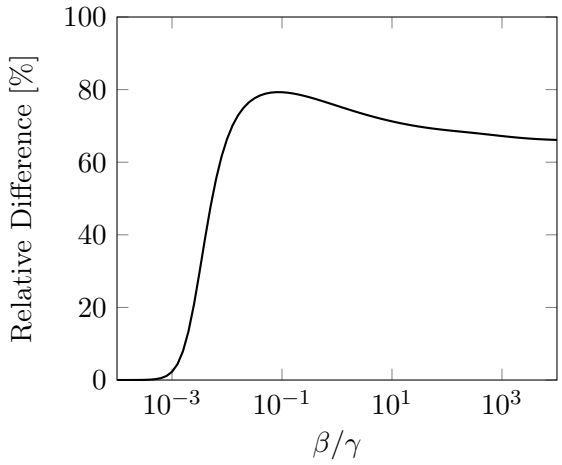

FIG. 7. $(1-\mu_{\mathrm{LI}}/\mu_{\mathrm{AM}})\times 100\%$, the relative difference of mean assembly times under the different IC (least-informed, LI, and all-monomer, AM). Computed with $c = 5$ and $m = 100$ with Coagulation-Fragmentation assembly.

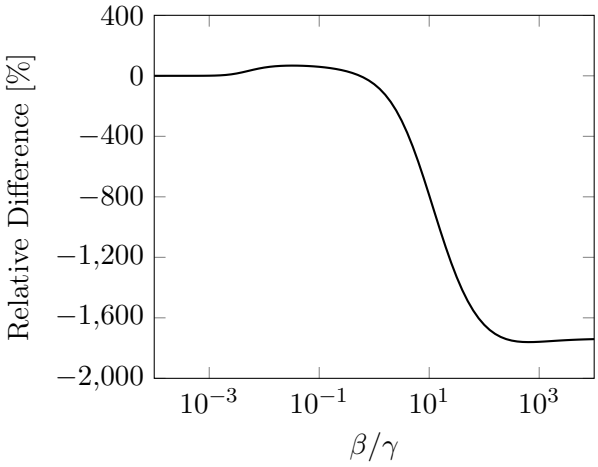

FIG. 8. Same as Figure 7, but with Becker-Döring assembly.

counter-intuitive behavior is well-modeled by our least-informed IC, leading to a substantially longer mean first-passage time since the all-monomer IC allows for rapidly-nucleating trajectories (see Figures 6 and 8).

We argue that for *in vivo* experiments, sampling from the least-informed distribution is more in line with our physical knowledge and allows one to make fewer *a priori* assumptions about the parameter regime a system may be in. Conversely, for *in vitro* experiments (where, for example, protein is heated and denatured in a controlled fashion), the all-monomer IC may still be most appropriate.

We lastly note that our least-informed distribution is but one of many plausible distributions to sample the IC from. In the case of our yeast system, for example, one might consider sampling polymer configurations from some appropriately defined branching process [23, 24] to model the mother-daughter budding dynamics. The point remains, however, that only in very special circumstances should the IC be treated without uncertainty – doing so glosses over important biological or physical details of the system under study.

ACKNOWLEDGMENTS

The authors would like to acknowledge funding from NSF grant #1344279.

[1] K. Thorkelsson, P. Bai, and T. Xu, Nano Today **10**, 48 (2015).

[2] G. M. Whitesides and M. Boncheva, Proceedings of the National Academy of Sciences **99**, 4769 (2002).

[3] G. M. Whitesides and B. Grzybowski, Science **295**, 2418 (2002).

[4] D. M. Fowler and J. W. Kelly, Cell **137**, 20 (2009).

[5] T. P. Knowles, M. Vendruscolo, and C. M. Dobson, Nature Reviews Molecular Cell Biology **15**, 384 (2014).

[6] T. P. Knowles, C. A. Waudby, G. L. Devlin, S. I. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson, Science **326**, 1533 (2009).

[7] J. Masel, V. A. Jansen, and M. A. Nowak, Biophysical chemistry **77**, 139 (1999).

[8] J. T. Jarrett and P. T. Lansbury, Cell **73**, 1055 (1993).

[9] Y. Ohhashi, K. Ito, B. H. Toyama, J. S. Weissman, and M. Tanaka, Nature chemical biology **6**, 225 (2010).

[10] M. F. Tuite and T. R. Serio, Nature Reviews Molecular Cell Biology **11**, 823 (2010).

[11] S. S. Sindi and T. R. Serio, Current opinion in microbiology **12**, 623 (2009).

[12] M. v. Smoluchowski, Zeitschrift fur Physik **17**, 557 (1916).

[13] R. Becker and W. Döring, Annalen der Physik **416**, 719 (1935).

[14] P. Blatz and A. Tobolsky, The journal of physical chemistry **49**, 77 (1945).

[15] J. Ball, J. Carr, and O. Penrose, Communications in mathematical physics **104**, 657 (1986).

[16] J. Ball and J. Carr, Proceedings of the Royal Society of Edinburgh: Section A Mathematics **108**, 109 (1988).

[17] J. Ball and J. Carr, Journal of statistical physics **61**, 203 (1990).

[18] M. D'Orsogna, G. Lakatos, and T. Chou, The Journal of chemical physics **136**, 084110 (2012).

[19] M. F. Tuite and B. S. Cox, Nature Reviews Molecular Cell Biology **4**, 878 (2003).

[20] A. Derdowski, S. S. Sindi, C. L. Klaips, S. DiSalvo, and T. R. Serio, Science **330**, 680 (2010).

[21] J. L. Doob, *Stochastic processes*, Vol. 101 (New York Wiley, 1953).

[22] R. Yvinec, M. R. D'Orsogna, and T. Chou, The Journal of chemical physics **137**, 244107 (2012).

[23] P. Whittle, in *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 61 (Cambridge Univ Press, 1965) pp. 475–495.

[24] P. Olofsson, S. S. Sindi, *et al.*, Journal of Applied Probability **51**, 453 (2014).