

Kinetics of self-assembly via facilitated diffusion: Formation of the transcription complex

ZiYa Kalay

Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Yoshida Ushinomiya-cho, 606-8501, Kyoto, Japan

(Received 27 April 2015; revised manuscript received 14 July 2015; published 29 October 2015)

We present an analytically solvable model for self-assembly of a molecular complex on a filament. The process is driven by a seed molecule that undergoes facilitated diffusion, which is a search strategy that combines diffusion in three dimensions and one dimension. Our study is motivated by single-molecule-level observations revealing the dynamics of transcription factors that bind to the deoxyribonucleic acid at early stages of transcription. We calculate the probability that a complex made up of a given number of molecules is completely formed, as well as the distribution of completion times, upon the binding of a seed molecule at a target site on the filament (without explicitly modeling the three-dimensional diffusion that precedes binding). We compare two different mechanisms of assembly where molecules bind in sequential and random order. Our results indicate that while the probability of completion is greater for random binding, the completion time scales exponentially with the size of the complex; in contrast, it scales as a power law or slower for sequential binding, asymptotically. Furthermore, we provide model predictions for the dissociation and residence times of the seed molecule, which are observables accessible in single-molecule tracking experiments.

DOI: [10.1103/PhysRevE.92.042716](https://doi.org/10.1103/PhysRevE.92.042716)

PACS number(s): 87.10.Mn, 05.20.Dd, 82.20.Fd

I. INTRODUCTION

Many biochemical processes involve formation of mesoscopic molecular structures that perform complex tasks. One well-known example is the transcription complex which plays the key role in accessing the information coded in the deoxyribonucleic acid (DNA) [1]. During the process of transcription, a complex consisting of ribonucleic acid (RNA) polymerase and transcription factors is assembled on the DNA in order to *read* the genetic information and produce RNA molecules. Thanks to powerful methods of molecular biology, it has been possible to study the number and types of molecules involved in the formation of the transcription complex; nevertheless, kinetics of the formation of transcription complex is much less known and is now an active area of biophysics [2–4]. A key experimental finding [5] regarding the kinetics of transcription factors is that at least some of the molecules that bind to sites on DNA undergo *facilitated diffusion* [6,7], during which molecules diffusing in three dimensions (3D) can temporarily bind to the DNA, diffusing along the filament and densely exploring it, which is thought to be an efficient search strategy [5]. Other aspects of facilitated diffusion, such as how it would affect the noise in transcriptional regulation, have also been explored in previous theoretical work [8,9].

We consider the kinetics of a self-assembly process in which a *seed* molecule diffuses in 3D and gets temporarily attached to a filament that carries a target site (see Fig. 1 for an illustration of the process and of facilitated diffusion). In this work, we do not explicitly model diffusion in 3D, which has been studied earlier [6], and focus on the dynamics of a nucleation process initiated by the seed, as described below. While it is associated with the filament, the seed molecule undergoes 1D diffusion and when it occupies the target site, it becomes bound at a constant rate, triggering the subsequent binding of other molecules that bind and unbind at constant rates. When w molecules are assembled, the process is complete. We consider the case where the seed molecule is initially bound at the target site and focus on the kinetics of the rest of the process. Behavior of the seed molecule is motivated by the observation

of facilitated diffusion of transcription factors in bacterial cells, as mentioned above. We envisage that the seed is an essential molecule for transcription initiation which possesses binding sites for other molecules or induces the binding of additional transcription factors, such as the RNA polymerase [1–4]. In eucaryotic cells, RNA polymerase II (RNAP2) transcribes the majority of genes. Although the binding order of molecules that form the transcription initiation complex is not clear, experiments suggest that a number of transcription factors need to bind both before and after the binding of RNAP2 [4]. Therefore, if one were to think of RNAP2 as the seed molecule, then the model presented here would be applicable to the latter part of the assembly process, starting with the

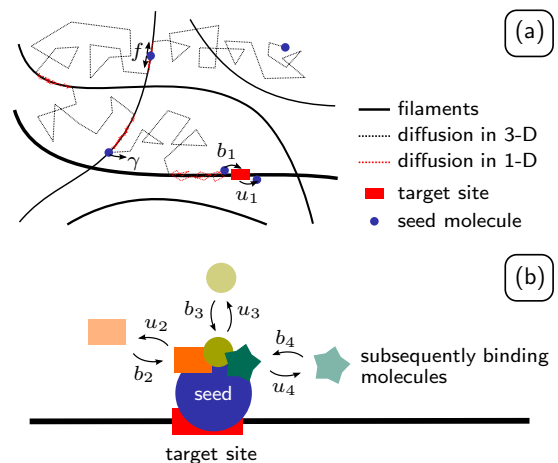


FIG. 1. (Color online) Illustration of the process of self-assembly via facilitated diffusion. In (a), facilitated diffusion of the seed molecule is shown, where it performs Brownian motion in 3D (not explicitly modeled in this work) and temporarily binds to filaments, diffusing in 1D (at a rate $\propto f$). When the seed arrives at the target site, it binds at rate b_1 , and, if it is already bound, becomes unbound at rate u_1 . In (b), assembly process is shown for $w = 4$. After the seed is bound, additional molecules are recruited via reversible binding.

binding of RNAP2. Alternatively, the seed can represent a molecule that binds during the initial stages of assembly. A candidate for such a molecule is TFIID, which significantly changes the local conformation of the DNA, paving the way for subsequent molecules to bind [1].

Our main results consist of an exact expression for the probability that the assembly completely forms upon the binding of a seed molecule and the distribution of the completion time. In addition to these quantities that describe the kinetics of assembly, and to link model predictions with quantities that can be accessed in single-molecule observations, distributions of the time at which the seed dissociates from the filament, as well as the residence time in an interval containing the target site are also presented.

In what follows, we first describe a mathematical model corresponding to the process described above and illustrated in Fig. 1. We then present results of analytic calculations for the quantities mentioned above. Last, we discuss the applicability of the model as well as previous findings on the same problem and state our conclusions. We also provide a comparison of analytical results with simulations of the process, presented in Appendix E.

II. AN ANALYTICALLY SOLVABLE MODEL OF ASSEMBLY

A. Model description and assumptions

We consider the formation of a molecular complex consisting of w elements that are assembled sequentially or in random order. Formation of the complex is initiated by a *seed molecule*, which we will just refer to as the *seed*. The seed is envisaged to diffuse in 3D (not explicitly modeled) and temporarily attaches to filaments along which it undergoes 1D diffusion. When the seed occupies the target site on the filament, it can become bound at rate b_1 .

After the seed is bound, a total number of $w - 1$ molecules start to assemble, which are, so to speak, recruited by the seed. If the seed is the only molecule in the complex, then it becomes unbound at rate u_1 , returning to diffusion along the filament. When two or more molecules (including the seed) are bound at the same time, the seed cannot become unbound. Note that this assumption may not be applicable in general, and the model considered in this work is appropriate for the case where the binding of subsequent molecules stabilizes the complex. While the seed is unbound and diffusing on a filament, it dissociates at a constant rate γ . When w molecules are assembled, the process is complete.

In this work, we do not explicitly model the motion of the seed in 3D and focus on the dynamics when the seed molecule is initially bound at the target site. This approach allows us to study the kinetics of the assembly process in the presence of a low concentration of seed molecules, which is often a good assumption in cell biology [10], accounting for the effect of facilitated diffusion. For a treatment of the problem of searching for a target site in a filament via 3D diffusion interrupted by periods of 1D exploration, we refer the readers to existing literature [7–9, 11, 12].

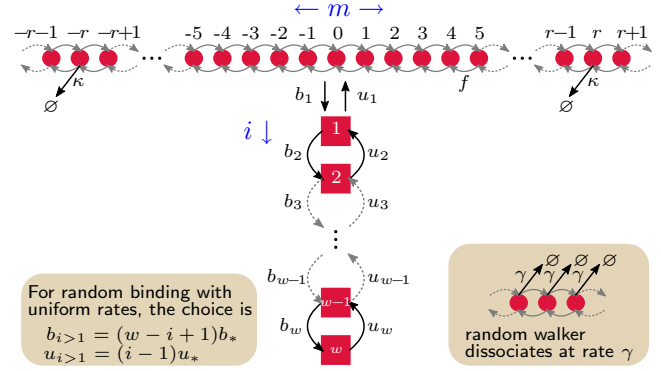


FIG. 2. (Color online) Illustration of the assembly model as a Markov chain. Circles and squares correspond to different states of the system, and arrows show state transitions with corresponding rates. Circles denote different positions on the filament, indexed by m , modeled as a lattice, where the seed molecule performs a random walk between adjacent sites, with a hopping rate f . As shown in the lower right, the random walker disappears from the system at a constant rate γ only while it is diffusing along the filament (states indicated by circles). Squares correspond to the bound states indexed by i . For sequential and random binding models, the rates $b_{i>1}$ and $u_{i>1}$ are interpreted in a different fashion (see text for explanation).

While a cell can contain a large number of transcription factors as well as corresponding target sites, we consider a regime where the concentration of seed molecules is low such that the competition for the target site is negligible, and the dynamics of the system is well characterized by a single seed and a target site. Nevertheless, we do provide a generalization for multiple molecules under the assumption of negligible competition (see Sec. III D).

B. Mathematical details and predicted quantities

Based on the biologically inspired model illustrated in Fig. 1, we consider a model for diffusion and assembly that can be analytically solved by standard tools of statistical mechanics [13]. See Fig. 2 for an illustration of the corresponding mathematical model.

Diffusion along the filament is modeled as a continuous time random walk in a 1D lattice, where the random walker, that is, the seed, hops between adjacent lattice sites at a (symmetric) rate f (see Fig. 2). For a DNA filament, it is natural to think that the lattice sites correspond to base pairs that are separated by ≈ 0.34 nm [1]. The lattice site with index $m = 0$ is where the target is located. When the seed is occupying site $m = 0$, it can become bound at rate b_1 , upon which the system would transition to the first bound state $i = 1$. Being in the first bound state, the system can either go back to the state where the seed is diffusing, at rate u_1 , or transition to the next bound state, $i = 2$, if an auxiliary molecule becomes bound. Note that while the meaning of the index m is straightforward (the position of the seed), the physical picture ascribed to the i^{th} bound state depends on the details of how molecules are assembled. We consider two different models of assembly: sequential and random.

In sequential assembly, $w - 1$ auxiliary molecules can reversibly bind in order, once the seed becomes bound.

When the i^{th} auxiliary molecule binds or unbinds, the system transitions to the $i + 1^{\text{st}}$ or i^{th} bound state, respectively. The binding and unbinding rates for the i^{th} auxiliary molecule is equal to b_{i+1} and u_{i+1} , respectively.

If auxiliary molecules assemble in random order, equally likely in any of the $(w - 1)!$ possible ways, then the model illustrated in Fig. 2 can still be used, provided that the binding and unbinding rates of different auxiliary molecules are sufficiently similar. We suppose that each molecule binds and unbinds at the uniform rates b_* and u_* , respectively. Provided that this is the kinetics at the level of individual molecules, the transition rates for the whole system would become $b_{i>1} = (w - i + 1)b_*$ and $u_{i>1} = (i - 1)u_*$, which was also employed in a previous study [14]. This is a consequence of treating the reactions involving auxiliary molecules as Poisson processes with exponential waiting times. When j of the auxiliary molecules are bound, such that the system is in state $i = j + 1$, binding of any of the remaining $w - 1 - j$ molecules would take the system to the $j + 2^{\text{nd}}$ state, or unbinding of any of the j molecules would take the system to the j^{th} state. The former takes place at rate $(w - j)b_*$, whereas the latter at ju_* , since the minimum of a set of independent exponential random variables is also distributed exponentially, with a parameter that equals the sum of all individual random variables' parameters.

In addition to studying the kinetics of the assembly, we are also interested in making predictions for observables relevant in single-molecule tracking experiments. Supposing that the seed is labeled (via fluorescent dyes, quantum dots, etc.) and its position can be tracked, one may be able to observe the time it takes for the seed to dissociate from the filament, or the time it takes for it to exit from an interval of length $2r$ centered at the target site, given the seed was observed to be bound at $t = 0$. To be able to calculate the statistics of these two times from the model, we introduce, for solely mathematical convenience, two *leaky* sites at $m = -r$ and r , where the seed disappears from the system at rate κ . In the next section, we consider the limits $\kappa \rightarrow 0$ and $\kappa \rightarrow \infty$ depending on which calculation is of concern.

Based on the model illustrated in Fig. 2, we write a set of master equations for the probability of finding the seed (unbound) at lattice site m , denoted by $P_m(t)$, and the probability of finding the system in the i^{th} bound state, denoted by $Q_i(t)$, at time t . Note that $Q_w(t)$ is the probability of having a fully assembled complex at time t , from which we will derive the probability of completion as well as the first completion time, in the next section. The master equation as well as its analytical solution is given in Appendix A. In the next section, we present results derived from the probabilities P_m and Q_i , assuming that they are known, and always refer to Appendices for calculation details.

III. RESULTS

In this section we present results for the kinetics of the assembly process obtained by solving the model illustrated in Fig. 2. Results are displayed in a way that highlights the difference between the kinetics for sequential and random binding models. For convenience, and to be able to treat the case of random binding as described in the previous section, we

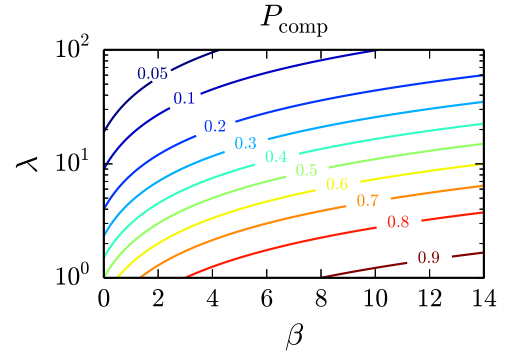


FIG. 3. (Color online) Contour plot of the probability of completion as a function of λ and β .

consider uniform rates b_* and u_* as the binding and unbinding rates of each auxiliary molecule, regardless of order.

All rates and times are measured in units of f and $1/f$, respectively, where f is the hop rate between adjacent sites in the lattice, proportional to the 1D diffusion coefficient.

A. Probability of completion

A key quantity that characterizes the efficiency of the assembly process is the probability that the molecular complex completely forms before the seed dissociates from the filament, given the seed was initially bound [that is, $Q_1(0) = 1$, $Q_{i \neq 1}(0) = 0$, and $P_m(0) = 0$]. We refer to this quantity as the probability of completion and denote it by P_{comp} . We note that P_{comp} is the probability of arriving at the bound state w at any time as $t \rightarrow \infty$, which is also the fraction of system trajectories that end at w . Therefore, we have $P_{\text{comp}} = \lim_{t \rightarrow \infty} Q_w(t)$, under the condition $u_w = 0$, ensuring that trajectories that reach the last bound state are terminated. Performing the calculation, we obtain (see Appendix B 2)

$$P_{\text{comp}} = \frac{1}{1 + \frac{\lambda(w)}{1 + \beta}}, \quad (1)$$

where $\beta = b_1/\sqrt{\gamma(\gamma + 4f)}$ and $\lambda(w)$ is a constant formed by the combination of all the rates b_i and u_i except b_1 , explicitly given in (B11). If the complex is made up of just a pair of molecules, then $\lambda(w)$ has a particularly simple form, given by $\lambda(2) = u_1$. Note that β quantifies the ratio of the affinity to the binding site (i.e., nucleation rate given the particle occupies the binding site) to the rate at which the seed is carried away from the binding site, either via dissociation or diffusion. When $f/\gamma \ll 1$, meaning dissociation is much more rapid than diffusion along the filament, we have $\beta \propto b_1/\gamma$, and when $f/\gamma \gg 1$, meaning diffusion is much faster than dissociation, $\beta \propto b_1/\sqrt{f\gamma}$.

Since λ and β do not share any parameters, they constitute a good pair of knobs that can be used to investigate the behavior of P_{comp} . In Fig. 3, a contour plot of P_{comp} is displayed as a function of λ and β . We note that it becomes less and less probable for the complex to be completed as λ increases or β decreases. This is in line with what one may expect intuitively, since larger λ values correspond to relatively larger

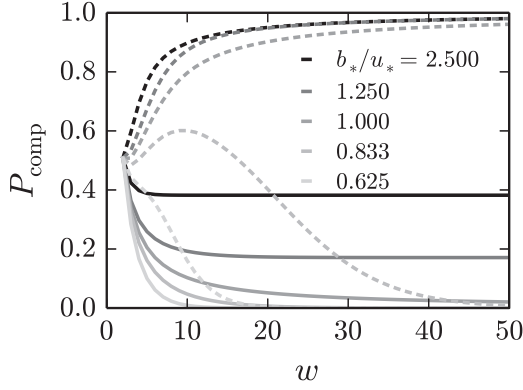


FIG. 4. P_{comp} as a function of the total number of molecules in the complex for different values of the ratio b_*/u_* (obtained by varying u_*). Solid and dashed curves correspond to sequential and random binding, respectively. Parameter values are $\gamma = 0.1$, $b_1 = 2$, $b_* = 0.25$, and $u_1 = 1$, measured in units of f , the hopping rate along the filament.

unbinding rates $u_{i>1}$, and smaller β values imply that the seed is diffusing fast and/or it dissociates from the filament quickly.

As λ and β are combinations of many parameters, it is informative to explore the behavior of P_{comp} as a function of parameters whose physical meaning is more direct. In this respect, next we display how P_{comp} changes with the total number of molecules in the complex and the ratio of the binding and unbinding rates b_*/u_* .

In Fig. 4, P_{comp} is plotted as a function of w , for five different values of b_*/u_* , indicated by curves with different shades of gray. In this figure and throughout the article, dashed and solid curves correspond to random and sequential binding models, respectively, unless otherwise noted. We see that P_{comp} has greater values for random binding compared to sequential binding for the same set of parameter values. In random binding, P_{comp} can be nonmonotonic in the number of bound states, depending on the value of the ratio b_*/u_* .

The behavior of P_{comp} as a function of w can be intuitively understood, to a certain extent, by ignoring the diffusion states and associating $m = 0$ with dissociation (see Fig. 2). Then the problem can be viewed as (biased) random walk in a 1D lattice with $w + 1$ sites, $i = 0, 1, \dots, w$, where arriving at the top (site 0) means dissociation and arriving at the bottom (site w) means completion, and the random walker starts at site $i = 1$. We take $b_1 = b_*$ for simplicity. In the next three paragraphs, we provide an intuitive explanation for the behavior observed in Fig. 4.

In sequential binding, the rate of acquiring and losing an auxiliary molecule does not depend on the number of already bound molecules. When $b_*/u_* > 1$, the random walk is biased downward, leading to the completion of the process with a probability that increases with b_*/u_* (see Fig. 4). On the contrary, for $b_*/u_* < 1$, the bias is against any motion towards completion. Therefore, the only way for the random walker to end up at the bottom first is via a highly improbable sequence of downward steps, whose probability is expected to dramatically diminish with increasing w , leading P_{comp} to zero as a function of w .

In random binding, when n molecules are bound, the rate at which the complex grows is given by $(w - n)b_*$, and

the rate at which it shrinks is nu_* . Therefore, there is a state with $n_* = b_*w/(u_* + b_*)$ bound molecules, which is more stable than others in the sense that the growth and shrinking rates are approximately balanced. Note that we have $b_*/u_* \approx n_*/(w - n_*)$, implying that for $b_*/u_* > 1$ we expect to have, *on average*, more bound molecules than missing ones. On the other hand, for $b_*/u_* < 1$, we expect to have more missing molecules. When $b_*/u_* > 1$, the stable state is in the lower half of the simplified lattice, and fluctuations are more likely to drive the system to the bottom, i.e., completion. In the other case, $b_*/u_* < 1$, the stable state is in the upper half, and fluctuations are more likely to take the system to the top, i.e., dissociation. This intuitive picture is in line with the results shown in Fig. 4. Note that the curve that corresponds to $b_*/u_* = 0.833$ indicates that the stable state in this case moves from the lower half to the upper half at around $w \approx 15$.

In the random binding model, when $b_*/u_* < 1$, we observe a transient increase in P_{comp} , although it eventually decays to zero with increasing w . To understand this, we note that the completion probability can be expressed by $(1 - p)q$, where p is the probability of dissociation during the first transition, and q is the probability of completion starting from site 2 (as the random walker of the simplified problem would end up in state 2 if it did not dissociate after the first step). Note that $p = 1/[1 + (w - 1)b_*/u_*]$, the probability of taking the first step upward, decreases with w (the number of auxiliary molecules). Therefore, one would expect an initial increase in P_{comp} due to the increasing factor $(1 - p)$. Nevertheless, for $b_*/u_* < 1$, this increase is balanced by a decrease in q , as completion before dissociation gets more and more improbable as the number of steps required to arrive at w increases, as discussed in the previous paragraph.

On the whole, we find that there is a qualitative difference in the behavior of P_{comp} for random and sequential binding and that there can be an optimal value of w that maximizes P_{comp} in random binding, depending on the relative strength of binding and unbinding rates of the auxiliary molecules.

Next we consider how the probability of completion depends on the ratio b_*/u_* for different values of the total number of molecules. Figure 5(a) shows P_{comp} as a function of b_*/u_* for $3 \leq w \leq 7$. We note that P_{comp} monotonically increases with b_*/u_* for both random and sequential binding models. In sequential binding, P_{comp} does not depend on w for $b_*/u_* \gg 1$. In contrast, the values of P_{comp} at $b_*/u_* \gg 1$ increase with w by as much as $\approx 25\%$ in the random binding model. In Fig. 5(b), the ratio $P_{\text{comp}}^{\text{seq}}/P_{\text{comp}}^{\text{ran}}$, where the superscripts denote the binding order of auxiliary molecules, is plotted as a function of b_*/u_* . We observe that P_{comp} is greater for the random binding model, especially for $b_*/u_* > 1$, and that there is a large drop in the ratio around $b_*/u_* = 1$.

Last, we present results quantifying the effect of facilitated diffusion on how likely the process is going to be completed. Given the seed is initially bound with no other bound molecules, setting $b_1 = 0$ would prevent the possibility of rebinding since the seed cannot become bound again if it ever transits to the diffusive state. Therefore, the ratio $\rho = P_{\text{comp}}/P_{\text{comp}}(b_1 = 0)$ would give us the relative enhancement of completion probability due to facilitated diffusion. Note that

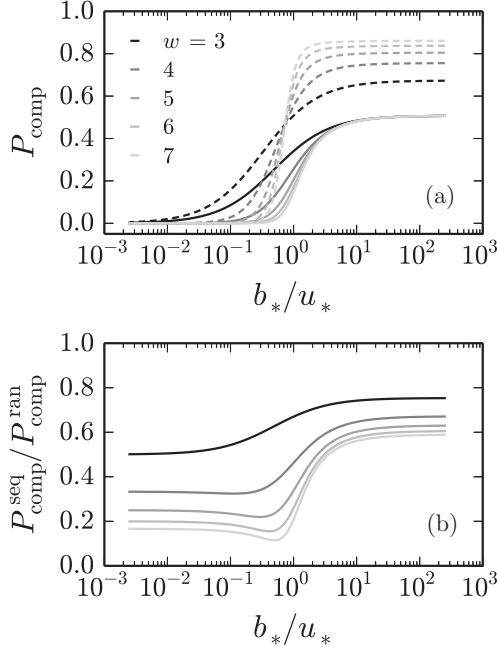


FIG. 5. Probability of completion as a function of the ratio b_*/u_* . In (a), P_{comp} is plotted for different values of the total number of molecules in the complex. Solid and dashed curves correspond to sequential and random binding, respectively. In (b), the ratio of P_{comp} for sequential binding to that for random binding is displayed. Parameter values for both graphs are $\gamma = 0.1$, $b_1 = 1$, $b_* = 0.25$, and $u_1 = 1$, measured in units of f , the hopping rate along the filament.

setting $f = 0$ maximizes the enhancement due to rebinding, as in this case a seed that becomes unbound does not leave the binding site ($m = 0$) via 1D diffusion, such that rebinding occurs at the maximal rate b_1 .

The relative enhancement ratio ρ is explicitly given by

$$\rho = \frac{1 + \lambda}{1 + \frac{\lambda}{1 + \beta}}. \quad (2)$$

Expanding (2) around $\beta = 0$, we get

$$\rho = 1 + \frac{\lambda}{1 + \lambda} \beta + O(\beta^2),$$

which clearly shows that there is no enhancement when the affinity to binding site is zero, $\beta = 0$, or when dissociation or diffusion rates diverge, that is, $\gamma \rightarrow \infty$ or $f \rightarrow \infty$, implying $\beta \rightarrow 0$. Note that the enhancement initially increases linearly with the ratio β [defined below (1)].

To demonstrate the behavior of the enhancement factor, a contour plot of ρ is displayed in Fig. 6 as a function of λ and β , plotted in the same domain as P_{comp} in Fig. 3. We see that facilitated diffusion enhances the probability of completion as λ and β increase. Greater β values correspond to a situation where the seed spends greater amount of time bound to the filament, in the vicinity of a lattice site, and hence explains the increase in ρ . While the parameter λ depends on many model parameters, it is roughly proportional to u_*/b_* . Therefore, greater λ values correspond to relatively larger unbinding rates, which hinders the completion of the process both with and

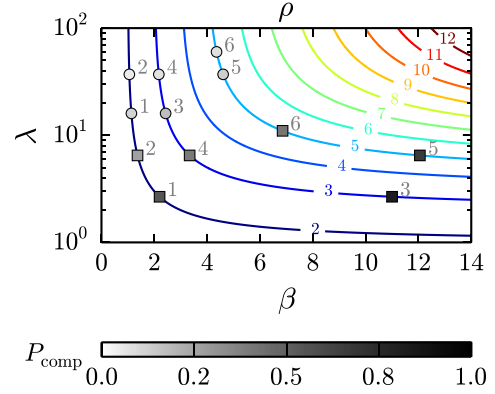


FIG. 6. (Color online) Contour plot of the enhancement ratio due to facilitated diffusion. Sequential and random binding models are indicated by circles and squares, respectively, and symbols with the same number have identical binding and unbinding rates per molecule: $u_* = 0.25$ for 1 and 3; 0.4 for 2, 4 and 5; 0.5 for 6. Points on the same curve correspond to the same enhancement factor for different values of u_* , where other parameters are fixed as $w = 5$, $\gamma = 0.1$, $b_1 = 2$, $b_* = 0.25$, and $u_1 = 1$, where rates are measured in units of f . Data points are color coded in grayscale according to their P_{comp} values and the corresponding color bar is indicated below the graph.

without diffusion. Higher ρ values for increasing λ suggest that facilitated diffusion can counter this hindrance effect to a limited extent.

In Fig. 6, we also include a number of data points to compare the effect of facilitated diffusion in random and sequential binding. Sequential and random binding models are indicated with circles and squares, respectively. Data points with the same index correspond to the same value of b_* and u_* , and hence the same per molecule binding-unbinding rates for auxiliary molecules, while the rest of the parameters are determined by imposing the condition that ρ is fixed. Each data point is colored in grayscale according to its P_{comp} value, and the color bar is shown at the bottom of the graph. Comparing the relative positions of the data points with the same index and considering the corresponding P_{comp} values, we conclude that with the same kinetic constants per reaction (b_* and u_*), random binding has larger probability of completion; however, achieving the same enhancement factor with random binding requires stronger association with the filament as compared to sequential binding (compare symbols located on the same line).

B. Completion time

In the previous section we demonstrated the behavior of the completion probability as a function of a subset of the model parameters and found that the order in which auxiliary molecules bind has a significant effect on the chance of completion, random binding being more efficient. Next we present model predictions regarding the time dependence of the process, providing the complementary information to answer the question: Given the process completes, how long does it take?

We define T_{comp} as the completion time, the time at which all w molecules are assembled for the first time, given the

seed is the only molecule initially bound. Let $f_{\text{comp}}(t)$ be the probability density function of the random variable T_{comp} , conditioned on the completion of the process. In terms of $Q_i(t)$, the conditional distribution $f_{\text{comp}}(t)$ can be expressed as

$$\begin{aligned} f_{\text{comp}}(t) &= \frac{1}{P_{\text{comp}}} \left\{ -\frac{d}{dt} [1 - Q_w(t; u_w = 0)] \right\}, \\ &= \frac{1}{P_{\text{comp}}} \left. \frac{dQ_w}{dt} \right|_{u_w=0}, \end{aligned}$$

where the quantity in square brackets in the first line is the survival probability, that is, the probability of not having visited the w^{th} bound state until time t , and we need the normalization constant P_{comp} , as f_{comp} is conditioned on the completion of the process. In Appendix B 2, we show that the Laplace transform of $f_{\text{comp}}(t)$ is given by

$$\begin{aligned} \tilde{f}_{\text{comp}}(\epsilon) &= \left(\frac{(-1)^{w-1}}{P_{\text{comp}}} \left[\prod_{i=2}^w K_i(w) \right] \right. \\ &\quad \times \left. \left\{ \frac{\epsilon}{\epsilon + b_2 + u_1 \alpha(\epsilon) / [\alpha(\epsilon) + b_1] + u_2 K_2(w)} \right\} \right) \Big|_{u_w=0}, \end{aligned} \quad (3)$$

where $\alpha(\epsilon) = \sqrt{(\gamma + \epsilon)(\gamma + 4f + \epsilon)}$, $K_i(w)$ is a continued fraction involving the rate constants [see (A8)], and the Laplace transform is defined as

$$\tilde{f}(\epsilon) = \int_0^\infty dt e^{-\epsilon t} f(t). \quad (4)$$

Throughout the text, we will use tildes to distinguish Laplace-transformed quantities.

In the following, we first discuss how the mean and variance of T_{comp} depend on the binding order and model parameters and then demonstrate the behavior of the full distribution $f_{\text{comp}}(t)$.

1. Mean and variance of the completion time

The m^{th} moment of T_{comp} can be calculated from the Laplace transform of $f_{\text{comp}}(t)$ as

$$\langle (T_{\text{comp}})^m \rangle = (-1)^m \lim_{\epsilon \rightarrow 0} \frac{d^m \tilde{f}_{\text{comp}}}{d\epsilon^m}, \quad (5)$$

which follows from (4). We define $\mu_{\text{comp}} = \langle T_{\text{comp}} \rangle$ and $\text{CV}(T_{\text{comp}}) = [\langle (T_{\text{comp}})^2 \rangle - \mu_{\text{comp}}^2] / \mu_{\text{comp}}^2$ as the mean and coefficient of variance of the completion time.

Figure 7 shows μ_{comp} as a function of w for a set of b_*/u_* values. Plots in Figs. 7(a) and 7(b) correspond to identical parameter values, but the axes are scaled differently. Looking at Fig. 7(a), where the y axis is scaled logarithmically, we first note that μ_{comp} increases exponentially with w for the case of random binding (dashed curves). The exponent increases as b_*/u_* decreases to approach the value 1 from above, which is simple to grasp intuitively, as higher relative unbinding rates would lead to longer completion times. When the ratio b_*/u_* is below 1, results show the opposite trend, where the exponent decreases with decreasing b_*/u_* . This reflects the conditional nature of the completion time. In the previous section, we showed that P_{comp} approaches 0 for $b_*/u_* < 1$ as

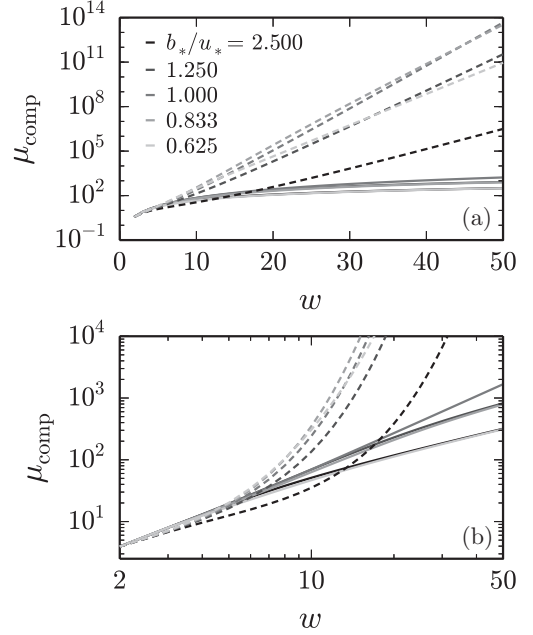


FIG. 7. Mean completion time as a function of the total number of molecules in the complex. In all plots, solid and dashed curves correspond to sequential and random binding, respectively. In (a), μ_{comp} is plotted on semilogarithmic axes (y) and indicates that the mean completion time asymptotically grows exponentially for random binding. In (b), μ_{comp} is plotted on log-log axes and suggests that the mean completion time asymptotically grows as a power law, or slower, for sequential binding. Parameter values for all graphs are $\gamma = 0.1$, $b_1 = 2$, $b_* = 0.25$, and $u_1 = 1$, measured in units of f , the hopping rate along the filament.

w increases, meaning that the fraction of trajectories that lead to the completion of the process becomes negligible. Results in Fig. 7(a) suggest that while it is quite unlikely for the process to complete when $b_*/u_* < 1$, the trajectories that do lead to completion take shorter and shorter times as u_* increases (this point will be discussed further below). Note that μ_{comp} clearly increases slower than exponentially for sequential binding (solid curves).

In Fig. 7(b), the same data is plotted on logarithmic axes. The most prominent feature here is that μ_{comp} for sequential binding (solid curves) increases as a power law for $b_*/u_* = 1$ and slower than a power law for $b_*/u_* \neq 1$. The rate of increase as a function of b_*/u_* follows a trend akin to that for random binding in Fig. 7(a).

Results above show that the behavior of the average completion time qualitatively differs for random and sequential binding models as the number of molecules in the complex increases. Since completion of the assembly is a stochastic process with many steps, one may expect significant variance in the values of T_{comp} such that a typical value may lie far away from μ_{comp} . To investigate this, we calculate the coefficient of variance (CV) of T_{comp} , defined just below (5). Figure 8 shows $\text{CV}(T_{\text{comp}})$ as a function of w for the same set of b_*/u_* values in Fig. 7. As seen in Fig. 8(a), we find that random and sequential binding models display approximately the same level of variability in T_{comp} for the first few w values. However, for $w \gg 2$, the coefficient of variance can be significantly

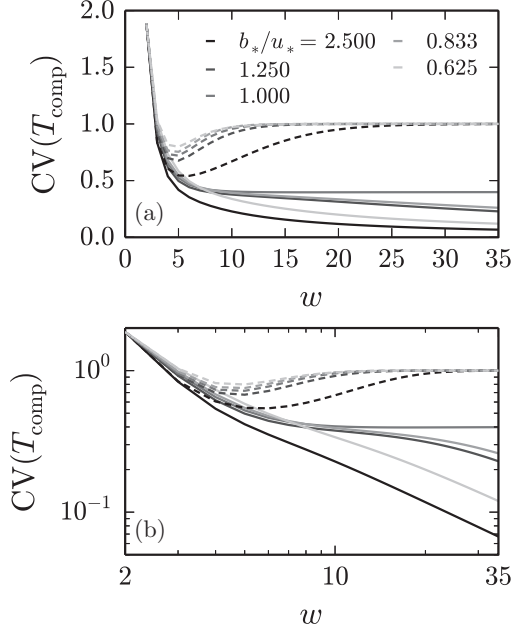


FIG. 8. Coefficient of variance of the completion time for sequential and random binding models. Solid and dashed lines correspond to sequential and random binding, respectively, for various values of the ratio b_*/u_* . Parameter values are $\gamma = 0.1$, $b_1 = 1$, $b_* = 0.25$, and $u_1 = 1$, measured in units of f , the hopping rate along the filament.

smaller for sequential binding compared with random binding. In random binding (dashed curves), CV transiently dips below 1 for a range of w values depending on b_*/u_* , but eventually approaches the value 1. In sequential binding, however, the asymptotic behavior of CV depends on the value of b_*/u_* , as shown in the log-log plot in Fig. 8(b).

When $b_*/u_* = 1$, random and sequential binding models both give rise to a CV that attains a constant value as w increases, that is, the standard deviation of T_{comp} increases at the same rate as its mean. This implies that the distribution of completion times remain well dispersed no matter how large the number of auxiliary molecules gets.

In the case of sequential binding, for $b_*/u_* \neq 1$, we observe that the CV decays to zero with w . While it is not straightforward to provide a simple explanation for this behavior for all values of b_*/u_* , we can gain some insight into it by considering the extreme cases $b_*/u_* \gg 1$ and $b_*/u_* \ll 1$, as discussed below.

When $b_*/u_* \gg 1$, almost all transitions are towards completion, meaning that we have $\text{mean}(T_{\text{comp}}) \approx wb_*^{-1}$, that is, the number of transitions to completion multiplied by the average duration of a transition. Since the variance of the duration between transitions (towards completion) is given by b_*^{-2} , variance of the completion time would be $\text{var}(T_{\text{comp}}) \approx wb_*^{-2}$. Therefore, we expect the CV to decay as w^{-1} in the limit $b_*/u_* \gg 1$, which is consistent with the power-law behavior observed in Fig. 8(b).

When $b_*/u_* \ll 1$, we expect the process to reach completion very rarely, as the complex is much more likely to shrink than grow at any state. Nevertheless, when the complex does form, the maximally likely system trajectory would

consist of w consecutive transitions towards completion, since any transition towards the unbound state would introduce an additional multiplicative factor of b_*/u_* in the likelihood of a trajectory that reaches completion. Therefore, we expect CV to decay in the same way as it does for $b_*/u_* \gg 1$, as explained in the paragraph above; the only difference is that completion only rarely occurs for $b_*/u_* \ll 1$, while it is the typical outcome for $b_*/u_* \gg 1$.

We note that a previous study by D’Orsogna and Chou [14] employed a similar model, although in a different context (ligand-receptor binding) and without spatial extent, and found that random binding results in faster mean completion times compared with sequential binding, except when $b_*/u_* \approx 1$, in the absence of any cooperativity. Our findings are consistent with this result up to a certain value of w , say, w_c . As seen in Fig. 7(b), the crossover value w_c above which random binding leads to longer completion time increases with b_*/u_* . We also would like to point out that the presence of diffusive states and the possibility of dissociation significantly affects the behavior of P_{comp} .

Overall, we find that sequential binding model results in more precise timings compared to the random binding model. In addition, as clearly seen in Fig. 7, sequential binding is orders of magnitude faster than random binding for $w \gtrsim 10$. Note that the predictions of an assembly model is only feasible if the completion time is less than the longest time scale in a cell, i.e., duration of the cell cycle.

2. Distribution of the completion time

Here we demonstrate the full distribution of T_{comp} , which is obtained by taking the inverse Laplace transform of the expression in (3) (see Appendices B 2 and D).

Figures 9(a)–9(c) show $f_{\text{comp}}(t)$ for sequential and random binding models, indicated by the superscripts “seq” and “ran.” All graphs display $f_{\text{comp}}(t)$ as a function of t , for $w = 4, 8, 16$ and $b_*/u_* = 2.5, 1, 0.625$, where different b_*/u_* values are color coded. Note that time axes are logarithmically scaled and cover a broad range containing four to six decades. Looking at Fig. 7(a), we see that $f_{\text{comp}}^{\text{seq}}(t)$ has a bell-shaped form, whose peak increases with w as one would expect. We observe that the distribution is most widely spread around its peak when $b_*/u_* = 1$. This is most consistent with the CV shown in Fig. 8, for $w \gg 2$, where CV for sequential binding attains its maximum at $b_*/u_* = 1$ and decreases for other values, implying a narrower distribution. In Fig. 8(b), identical parameter values are used to plot $f_{\text{comp}}^{\text{ran}}(t)$. This time, we note that the distribution gets spread over a wide range of t much quicker with increasing w . To better visualize this case, we display the same data on logarithmically scaled axes in (c). We immediately note that the curves for $w = 16$ are now clearly visible and that the distribution becomes significantly uniform over a broad range of t values as w increases [for instance, $f_{\text{comp}}^{\text{ran}}(t)$ for $w = 16$ attains the value $\approx 10^{-4}$ over a range $\approx 10^4$, implying that almost all of the probability is contained in the plateau]. This behavior is also in agreement with the results obtained in Fig. 8, where CV for random binding approaches 1 with increasing w , indicating that the distribution remains well spread over a range of t that grows with μ_{comp} .

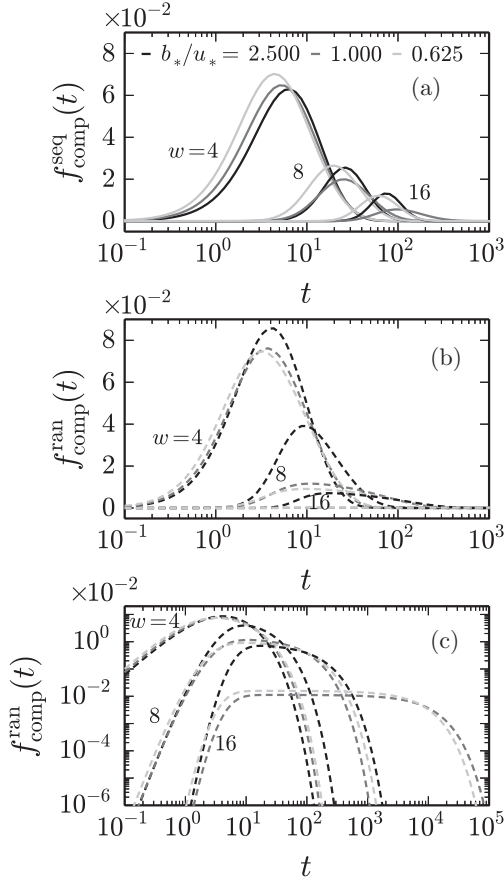


FIG. 9. Distribution of the completion time for different values of the total number of molecules and the ratio b_*/u_* . Parameter values are $\gamma = 0.1$, $b_1 = 2$, $b_* = 0.25$, and $u_1 = 1$, measured in units of f , the hopping rate along the filament.

C. Residence time in an interval

Using techniques of single molecule microscopy, it is possible to directly observe trajectories of individual molecules, especially in one and two dimensions. Analyzing trajectories that exhibit binding and unbinding events, one can extract useful information about reaction kinetics at the molecular level [15,16]. Nevertheless, this almost always requires fitting a model to the data.

If a labeling method can be developed to directly observe the completion time of a molecular complex, then model predictions presented so far can be used to fit experimental data and extract kinetic parameters. Nevertheless, this would be a challenging task. More often than not, there is uncertainty in the number of molecules in the complex as well as in determining which molecule(s) would best characterize the completion of the complex.

In this section, we present model predictions for the amount of time the seed spends in an interval of length $2r$ centered around the target site, given it was initially bound and did not dissociate until the time of measurement. We refer to this time as the *residence time*, which is also the first-passage time of the seed at a distance r from the target site, given dissociation is prevented by setting $\gamma = 0$. The choice $\gamma = 0$ is justified, as a particle cannot be tracked anymore after it dissociates.

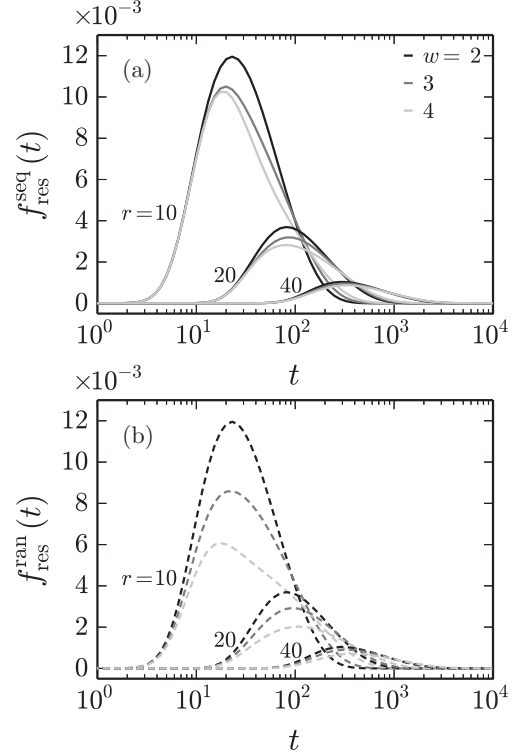


FIG. 10. Distribution of the residence time for sequential and random binding models. Each graph shows groups of three curves, characterized by the same w value, corresponding to three different values of the interval length $2r$, measured in units of lattice spacing. Panels (a) and (b) show $f_{\text{res}}(t)$ as a function of t for sequential and random binding, respectively. Parameter values are $\gamma = 0.1$, $b_1 = 2$, $b_* = 0.25$, and $u_1 = 1$, measured in units of f , the hopping rate along the filament.

Therefore, all measurements of T_{res} thus defined corresponds to a model where $\gamma = 0$. Although the information contained in the residence time is more indirect compared with the information that would be contained in the completion time, the residence time is probably easier to measure in practice.

To calculate the distribution of the residence time, denoted by $f_{\text{res}}(t)$, we consider the limit $\kappa \rightarrow \infty$ in the model illustrated in Fig. 2, which amounts to placing perfectly absorbing boundaries at $m = -r$ and r . Supposing that the seed molecule does not dissociate, we can calculate the first-passage time at site $-r$ or r from the knowledge of $P_m(t)$ and $Q_i(t)$. Details of the calculation are presented in Appendix C. Formulas for the mean and variance of $f_{\text{res}}(t)$ for the first few w are also given in Appendix C.

Figure 10 shows $f_{\text{res}}(t)$ as a function of t for sequential (a) and random (b) binding models for the first few values of w . We observe that the profile of $f_{\text{res}}(t)$ contains useful information about the binding model at smaller values of r . As r gets larger, distributions with different parameter values start to look similar. In the presence of measurement errors, this would make it difficult to infer the molecular details about the assembly process via measurement of T_{res} .

One possible way of directly observing the residence time could be achieved by employing nanomaterials such as DNA

origami frames. Two-dimensional frames made up of DNA that contain a stretched filament can be observed with atomic force microscopy as well as light microscopy [17]. This allows making measurements in a virtually 2D space such that the seed molecule would not go out of focus while it is bound to the filament.

D. When multiple seed molecules are present

We expect the medium to contain multiple seed molecules undergoing facilitated diffusion such that the overall rate of completion depends on how frequently a *new* seed molecule binds to the target site. By a *new* molecule, we mean any molecule except the one that has not dissociated from the filament after becoming unbound at the target site (following an incomplete assembly). When the concentration of the seed molecules is low, competition among different molecules for the target site is approximately negligible. In this case, arrival of new molecules at the first bound state can be approximated by a Poisson process, where the time until arrival is distributed exponentially. In other words, the system attempts to complete the assembly process at constant rate. We would like to remark that this approximation was considered in similar contexts in Refs. [12] and [9], where the low concentration assumption is also discussed in the light of biologically relevant values of molecular concentrations.

Let T_{dis} denote the time at which a seed that started in the first bound state dissociates, regardless of whether the process completes. Distribution of T_{dis} , denoted by f_{dis} , is given in Appendix B 1. Next we define $f_{\text{arr}}(t)$ and $f'_{\text{dis}}(t)$ to be distributions of the first arrival time, T_{arr} , of a new seed at the first bound state and the first dissociation time, T'_{dis} , of a seed that was initially bound and assuming that the process is not allowed to complete, that is, $b_w \rightarrow 0$. Note that the time T'_{dis} is introduced for convenience, and its role in calculating the completion time will become clear in what follows.

Suppose that initially there are no molecules in the vicinity of the target site. After a time T_{arr} , we expect a molecule to become bound, which would lead to completion with probability P_{comp} after a time T_{comp} , or to dissociation without completion with probability $1 - P_{\text{comp}}$, after time T'_{dis} . If we call this an *attempt*, then we can formulate the distribution of the first completion time in terms of the number of attempts that lead to the completion of the process. Note that this requires $\langle T'_{\text{dis}} \rangle \ll \langle T_{\text{arr}} \rangle$, implying that binding of a new molecule while another has not yet dissociated from the filament is unlikely.

Let $g_{\text{comp}}(t)$ be the distribution of the completion time when the assumption above holds. We can express $g_{\text{comp}}(t)$ as

$$\begin{aligned} g_{\text{comp}}(t) = & P_{\text{comp}}[f_{\text{arr}} * f_{\text{comp}}] \\ & + (1 - P_{\text{comp}})P_{\text{comp}}[f_{\text{arr}} * f'_{\text{dis}} * f_{\text{arr}} * f_{\text{comp}}] \\ & + (1 - P_{\text{comp}})(1 - P_{\text{comp}})P_{\text{comp}} \\ & \times [f_{\text{arr}} * f'_{\text{dis}} * f_{\text{arr}} * f'_{\text{dis}} * f_{\text{arr}} * f_{\text{comp}}] \\ & + \dots, \end{aligned} \quad (6)$$

where $*$ denotes convolution, that is, $f * g = \int_0^t ds f(t-s)g(s)$. In (6), the first, second, and third terms correspond to the probability that the process is completed after the first, second, and third attempts, multiplied by the distribution of

the time each route takes. Taking the Laplace transform of (6), thereby converting convolutions into products, we obtain

$$\begin{aligned} \tilde{g}_{\text{comp}}(\epsilon) = & \frac{P_{\text{comp}} \tilde{f}_{\text{comp}}(\epsilon)}{(1 - P_{\text{comp}}) \tilde{f}'_{\text{dis}}(\epsilon)} \\ & \times \sum_{n=1}^{\infty} [(1 - P_{\text{comp}}) \tilde{f}_{\text{arr}}(\epsilon) \tilde{f}'_{\text{dis}}(\epsilon)]^n \\ = & \frac{P_{\text{comp}} \tilde{f}_{\text{arr}}(\epsilon) \tilde{f}_{\text{comp}}(\epsilon)}{1 - (1 - P_{\text{comp}}) \tilde{f}_{\text{arr}}(\epsilon) \tilde{f}'_{\text{dis}}(\epsilon)}, \end{aligned} \quad (7)$$

where we assumed the convergence of the geometric sum, which certainly holds as $\epsilon \rightarrow 0$, since $0 \leq P_{\text{comp}} \leq 1$ and then $\tilde{f}_{\text{arr}}(\epsilon)$ and $\tilde{f}'_{\text{dis}}(\epsilon)$ are Laplace transforms of normalized probability distributions.

Since we assume that the arrival of a new seed is a Poisson process, we have $f_{\text{arr}}(t) = \alpha e^{-\alpha t}$, where α is a function of the 3D diffusion coefficient as well as the unspecific binding-unbinding rates between the seed and the filaments. The distribution $f'_{\text{dis}}(t)$ can be calculated as described in Appendix B 1.

We can then calculate the mean completion time, starting from the state where no seed is bound, as

$$\langle T_{\text{comp}} \rangle = - \lim_{\epsilon \rightarrow 0} \frac{d \tilde{g}_{\text{comp}}}{d \epsilon},$$

which follows from (5). Substituting (7) into the equation above and performing the limit, we get

$$\langle T_{\text{comp}} \rangle = \mu_{\text{arr}} + \mu_{\text{comp}} + \left(\frac{1 - P_{\text{comp}}}{P_{\text{comp}}} \right) (\mu_{\text{arr}} + \mu_{\text{dis}'}),$$

where $\mu_{\text{arr}} = \langle T_{\text{arr}} \rangle = \alpha^{-1}$ and $\mu_{\text{dis}'} = \langle T_{\text{dis}'} \rangle$, where the angular brackets imply expected values, as in (5). If the complex disappears from the system, then we can estimate the rate of formation as

$$k_{\text{comp}} = \frac{1}{\langle T_{\text{comp}} \rangle}.$$

We note that the theory developed by Berg *et al.* [6], or possibly an extension of the widely applicable mean first-passage time calculations given by Bénichou *et al.* [18], can be used to approximate the association rate α as a function of parameters characterizing the whole system, including the 3D diffusion coefficient and the ratio of the average inter-filament distance to filament radius.

IV. DISCUSSION AND CONCLUSIONS

In this work, we presented a model and its analytic solution for the formation of a complex of molecules on a filament, applicable to the assembly of the transcription complex. The process is driven by seed molecules that undergo facilitated diffusion, which consists of 3D diffusion interrupted by episodes where a molecule associates with a filament and undergoes 1D diffusion in search of a target site. In this work, we did not explicitly model the 3D diffusion of a seed molecule, which was studied earlier [6]. Once the seed molecule becomes bound to the target site, a number of auxiliary molecules can reversibly bind, forcing the seed to stay bound, until an assembly of a given size forms. We believe

that including spatial degrees of freedom and accounting for the effect of facilitated diffusion is the major contribution of this study to the existing body of work on the kinetics of aggregation.

Two quantities were used to characterize the process: (1) the probability that the assembly completely forms and (2) the time it takes for the process to complete, conditioned on completion; once a seed molecule becomes bound. Mathematical expressions for these quantities are given in (1) and (3), respectively. In Appendix E, we also provide a comparison of analytical expressions with simulations, verifying the validity of the results.

Similarly to previous studies (see, for instance, Refs. [14] and [19], where the latter has an experimental component), we found that the order in which auxiliary molecules bind would matter and compared two different models where auxiliary molecules bind in a strictly sequential order and in completely random order.

The findings indicate that the probability of completion is greater for random binding than it is for sequential binding (see Figs. 4 and 5). Interestingly, in random binding, when the unbinding rates of auxiliary molecules are relatively larger than their binding rates, there can be an optimal size for the complex for which the chance of completion is maximal (see Fig. 4). While the probability of completion is larger for the random binding model, it can take a long time to reach completion, especially when the complex contains much more than a few molecules. Calculating the completion time distribution for the two models, we found that the mean completion time grows exponentially with the size of the complex for the random binding model, while it grows as a power law or slower for the sequential binding model (see Fig. 7). In addition, completion time is much more broadly distributed for random binding compared to that in sequential binding (see Figs. 8 and 9).

Therefore, there is a trade-off between the probability of completion and the completion time, and an optimal strategy may consist of a hybrid model, where the first few molecules bind in random order to stabilize the forming complex, and the rest of the molecules bind sequentially to reduce the completion time, ensuring that it does not scale exponentially with the number of molecules in the complex.

Facilitated diffusion enhances the probability of completion by increasing the chances for the seed molecule to quickly rebound to the target site even if it becomes unbound before the process completes (this so-called *rebinding* effect was also discussed by other authors, for instance, in the context of enzymatic reactions [20] and in gene expression [11]). We quantified the amount of enhancement by deriving a formula for it as a function of all model parameters [see (2)]. Inside the parameter range considered here, facilitated diffusion is found to enhance the completion probability more strongly for the sequential binding model compared with the random binding model (see Fig. 6).

Note that the current model can be generalized to also let auxiliary molecules undergo facilitated diffusion. One possible way of achieving this is generalizing the previously studied *island growth model* [21,22], where monomers adsorb to a surface and undergo diffusion limited aggregation to form immobile islands to allow for dissociation and readsorption of monomers.

Our results are relevant for the case where the concentration of seed molecules is sufficiently low so the competition for the target site is negligible (also see Refs. [12] and [9]). Under this assumption, we also provided an approximate analytic expression for the Laplace transform of the completion time distribution for multiple seed molecules, from which we obtained the average completion time.

While the model we consider here is appealing for artificial systems where, for instance, stretched-out DNA filaments are placed in nanoengineered structures [17,23], applicability of the model to the formation of the transcription complex *in vivo* depends on the validity of two key assumptions. First, the DNA is assumed to be a 1D filament in the vicinity of a target site (a regulatory sequence of a gene). If the dissociation rate γ is much larger than the hop rate f , then this assumption is more likely to hold, since the seed would not be able to explore a large section of the filament at a time. We should note that DNA can be packed inside cells in a highly organized manner; it could be condensed by proteins in bacteria and is organized in chromosomes in eukaryotic cells [1]. In eukaryotic cells, before transcription of a gene begins, the structure of the DNA around the regulatory region of the gene loosens up such that transcription factors can directly bind to the base pairs. How well this loosened-up section of the chromatin can be approximated by a filament should eventually be verified by experiments *in vivo*. Second, we assume that the hopping rate along the DNA is uniform. Nevertheless, recent studies showed that DNA-binding molecules can act like “road blocks” that can hinder the 1D diffusion of transcription factors [24]. In the presence of molecules that act as road blocks as well as nonuniformity in the rate of diffusion along different sections of the DNA due to other reasons, the model considered here may underestimate, for instance, the variance and higher moments of the completion time distribution. To improve on this point, one can use a more involved model of diffusion in 1D, allowing for random transition rates, permeable barriers [25–28], and results derived for transport in random environments [29,30].

In this work, we did not investigate the presence of cooperativity in binding rates and assumed that the motion of molecules can be described by Brownian diffusion (continuous-time random walk with exponential waiting times). Inclusion of cooperative binding-unbinding of auxiliary molecules and considering anomalous diffusion can affect the stability of the complex being formed [14] and lead to correlated bursts in gene expression [11].

A key factor that determines the efficiency of search for target sites via facilitated diffusion is the relative affinity of the seed to unspecific sites on the DNA compared with that to the target site. It is worthwhile to note that this point becomes even more significant when one accounts for the degradation of DNA-binding molecules [31]. In our model, the seed hops between all adjacent sites, including the target site, at the rate f , and in all of the plots we choose $b_1 = 2f$, implying that binding to the target site happens at only twice the rate at which the seed hops between any two sites. Therefore, results illustrated here correspond to the case where the target site is not strongly “sticky.” Note that this is consistent with experimental findings showing that the probability of binding at the first encounter is not necessarily close to 1 [5].

In Sec. III C we presented the model prediction for the time it takes for a seed to escape from an interval around the target site, provided that it does not dissociate from the filament until it is observed to escape. Single molecule observations can be performed by tagging multiple molecules and can be sophisticated enough to provide direct information about molecular interactions (for a review on multiple experimental methods, see, for instance, Ref. [3]). Nevertheless, in the most basic setting, assuming that the seed molecule can be tracked, the dissociation time and the residence time for the seed can be directly accessed, providing evidence for performing model selection. However, we remark that the applicability of an analysis using the residence time defined above would often be limited to observations in the vicinity of the binding site and may require high position precision. This is because the *in vivo* unspecific binding strength of molecules to the DNA filament is not necessarily high enough for the molecule to cover a large distance on the DNA without being dissociated. To provide a quick estimate, we expect the diffusive displacement of a molecule to go as $d = \sqrt{\langle m^2 \rangle} \sim \sqrt{f\tau}$, where τ is the typical time a molecule spends attached, which can be estimated as $\tau \sim 1/\gamma$. Therefore, the distance a molecule covers typically goes as $d \sim \sqrt{f/\gamma}$. Nevertheless, several single molecule observations of RNA polymerase *in vitro* suggest that the dissociation times can be long to allow significant motion of the particle along the DNA (see, for instance, Refs. [32] and [23]). Finally, as also noted in Sec. III C, we remark that the discriminative power of this analysis diminishes as the length of the interval increases (see Fig. 10).

Recent experimental studies on the kinetics of formation of the transcription initiation complex aim to test competing hypotheses of sequential and random binding. In this respect, we believe that model predictions, in conjunction with cutting-edge experimental methods, would be useful for revealing the dynamics of such mesoscopic systems.

ACKNOWLEDGMENTS

This research was supported in part by Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Young Scientists (KAKENHI Grant No. 26730150) and the World Premier International Research Center (WPI) Initiative of the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan.

APPENDIX A: MASTER EQUATION AND ITS SOLUTION

In this section we present the mathematical formulation of the model illustrated in Fig. 2 in terms of a set of master equations, and its solution.

We model the seed molecule as a random walker hopping between nearest-neighboring sites in a 1D lattice (see Fig. 2). When the random walker occupies the site $m = 0$, it can transition to a bound state at rate b_1 . We consider a total number of w bound states, each of which is accessible in a sequential manner, e.g., a stack of bound states. Among the bound states, the rate of transition from state i to $i + 1$ is denoted by b_{i+1} , and the rate of transition from i to $i - 1$ is denoted by u_i . When the random walker occupies the first bound state, it becomes unbound at rate u_1 and returns to the lattice. Note that b_i and

u_i characterize the rate of going deeper and shallower in the stack of bound states, respectively.

Let $P_m(t)$ denote the probability of finding the random walker unbound at site m and $Q_i(t)$ be the probability that it is in the i^{th} bound state. The master equations that govern these probabilities are given by

$$\begin{aligned} \frac{dP_m}{dt} = & f(P_{m-1} + P_{m+1} - 2P_m) - \gamma P_m \\ & + [\delta_{m,0}(-b_1 P_0 + u_1 Q_1) - \kappa(\delta_{m,-r} + \delta_{m,r})P_m], \end{aligned} \quad (\text{A1})$$

where κ is a parameter that adjusts the strength of an absorbing boundary at $-r$ and r , and

$$\begin{aligned} \frac{dQ_1}{dt} = & b_1 P_0 - (u_1 + b_2)Q_1 + u_2 Q_2, \\ \frac{dQ_2}{dt} = & b_2 Q_1 - (u_2 + b_3)Q_2 + u_3 Q_3, \\ & \vdots \\ \frac{dQ_{w-1}}{dt} = & b_{w-1} Q_{w-2} - (u_{w-1} + b_w)Q_{w-1} + u_w Q_w, \\ \frac{dQ_w}{dt} = & b_w Q_{w-1} - u_w Q_w. \end{aligned} \quad (\text{A2})$$

We start with the solution for $P_m(t)$. Let us denote by $\varphi_{m-n}(t)$ the solution of (A1) with the initial condition $P_m(0) = \delta_{m,n}$, when the terms inside the square brackets are set to zero, which corresponds to random walk in an infinite lattice where the random walker disappears at rate γ . Using Laplace and discrete Fourier transforms to convert differential equations to algebraic equations, $\varphi_{m-n}(t)$ can be obtained as

$$\varphi_{m-n}(t) = e^{-(2f+\gamma)t} I_{m-n}(2ft), \quad (\text{A3})$$

where $I_m(t)$ denotes the modified Bessel function of the first kind [33]. The Laplace transform of (A3) is given by (see Ref. [34], p. 75)

$$\mathcal{L}\{e^{-at} I_\nu(bt)\} = \frac{[(\epsilon + a + b)^{1/2} - (\epsilon + a - b)^{1/2}]^{2\nu}}{(2b)^\nu [(\epsilon + a)^2 - b^2]^{1/2}}, \quad (\text{A4})$$

where, $\text{Re}[\nu] > -1$, $\text{Re}[\epsilon] > \max\{\text{Re}[b - a], -\text{Re}[b + a]\}$ and \mathcal{L} denotes the Laplace transform defined in (4) with ϵ as the Laplace variable. We use tildes ($\tilde{\cdot}$) to denote Laplace transformed variables as in the main text. To obtain the full solution of (A1), we note that it is a first-order linear differential equation, which allows us to express its solution as

$$P_m = \sum_n P_n(0) \varphi_{m-n} + \int_0^t ds \sum_n \varphi_{m-n}(t-s) [\cdots](n,s), \quad (\text{A5})$$

where $[\cdots](n,s)$ corresponds to the expression in square brackets on the right-hand side of (A1) as a function of n and s . Note that we need to express Q_1 in terms of P_0 in order to obtain a closed equation for P_m 's. To achieve this, we formally solve the system of equations given in (A2) for an initial condition where the first bound state is occupied with probability $Q_1(0)$ and all other bound states are initially unoccupied, that is, $Q_i(0) = 0$ for $i > 1$. Taking the Laplace

transform of the system in (A2), and solving recursively, we find

$$\begin{aligned}\tilde{Q}_w &= \tilde{Q}_{w-1} \frac{b_w}{\epsilon + u_w}, \\ \tilde{Q}_{w-1} &= \tilde{Q}_{w-2} \frac{b_{w-1}}{\epsilon + u_{w-1} + b_w - \frac{u_w b_w}{\epsilon + u_w}}, \\ \tilde{Q}_{w-2} &= \tilde{Q}_{w-3} \frac{b_{w-2}}{\epsilon + u_{w-2} + b_{w-1} - \frac{u_{w-1} b_{w-1}}{\epsilon + u_{w-1} + b_w - \frac{u_w b_w}{\epsilon + u_w}}}, \\ &\vdots\end{aligned}$$

Note that we readily have

$$\tilde{Q}_2 = \tilde{Q}_1 \frac{b_2}{\epsilon + u_2}, \quad (\text{A6})$$

which can be substituted in the Laplace transform of the first equation in (A2) to obtain an equation that only involves \tilde{Q}_1 and \tilde{P}_0 , whose solution is

$$\tilde{Q}_1 = \frac{Q_1(0) + b_1 \tilde{P}_0}{\epsilon + u_1 + b_2 + K_2(w)u_2}, \quad (\text{A7})$$

where $K_j(w)$ is defined as

$$\begin{aligned}
K_i(w) &= \frac{1}{u_i} \prod_{j=i}^w \frac{(-u_j b_j)}{(\epsilon + u_j + b_{j+1})} \\
&= - \frac{b_i}{\epsilon + u_i + b_{i+1} - \frac{u_{i+1} b_{i+1}}{\epsilon + u_{i+1} + b_{i+2} - \frac{u_{i+2} b_{i+2}}{\epsilon + u_{i+2} + b_{i+3} - \dots - \frac{u_w b_w}{\epsilon + u_w}}}},
\end{aligned} \tag{A8}$$

where the big-K notation is one of the convenient ways of denoting parts of continued fractions, defined as [35]

$$\mathbf{K}_i^j \frac{a_i}{b_i} = \frac{a_i}{b_i + \frac{a_{i+1}}{b_{i+1} + \frac{a_{i+2}}{b_{i+2} + \cdots + \frac{a_j}{b_j}}}}.$$

Note that in the limit $\epsilon \rightarrow 0$, we have

$$\lim_{\epsilon \rightarrow 0} K_i(w) = -\frac{b_i}{u_i}. \quad (\text{A9})$$

Combining all these, we observe that the solution for \tilde{Q}_i can be compactly written as

$$\tilde{Q}_i = (-1)^{i-1} \tilde{Q}_1 \prod_{j=2}^i K_j(w), \quad (\text{A10})$$

for $i \geq 2$. For $w \geq 2$, the total probability of being bound is given by

$$\sum_{i=1}^w \tilde{Q}_i = (1 - K_2\{1 - K_3[\cdots - K_{w-1}(1 - K_w)\cdots]\})\tilde{Q}, \quad (\text{A11})$$

which is obtained by summing up (A10) and rearranging terms in the summation [the argument w of $K_i(w)$ is omitted for brevity in notation]. When there is only a single bound state, that is, $w = 1$, the probability of being bound is simply equal to Q_1 .

Taking the Laplace transform of (A5), we can obtain an algebraic equation for P_m in terms of the probabilities P_{-r} , P_0 , and P_r , which reads

$$\tilde{P}_m = \Sigma'_m - b'_m - \kappa \tilde{\varphi}_{m+r} \tilde{P}_{-r} - \kappa \tilde{\varphi}_{m-r} \tilde{P}_r, \quad (\text{A12})$$

where we defined

$$\begin{aligned}\Sigma'_m &= \Sigma_m + \frac{u_1 Q_1(0)}{\epsilon + u_1 + b_2 + K_2(w)u_2} \tilde{\varphi}_m, \\ \Sigma_m &= \sum_n P_n(0) \tilde{\varphi}_{m-n}, \\ b'_m &= \left[1 - \frac{u_1}{\epsilon + u_1 + b_2 + K_2(w)u_2} \right] b_1 \tilde{\varphi}_m.\end{aligned}$$

Substituting $m = -r, 0$, and r in (A5), we then obtain the following system of linear equations:

$$\begin{bmatrix} 1 + \kappa \tilde{\varphi}_0 & b'_{-r} & \kappa \tilde{\varphi}_{-2r} \\ \kappa \tilde{\varphi}_r & 1 + b'_0 & \kappa \tilde{\varphi}_{-r} \\ \kappa \tilde{\varphi}_{2r} & b'_r & 1 + \kappa \tilde{\varphi}_0 \end{bmatrix} \begin{bmatrix} \tilde{P}_{-r} \\ \tilde{P}_0 \\ \tilde{P}_r \end{bmatrix} = \begin{bmatrix} \Sigma'_{-r} \\ \Sigma'_0 \\ \Sigma'_r \end{bmatrix}, \quad (\text{A13})$$

which can be solved for \tilde{P}_{-r} , \tilde{P}_0 , and \tilde{P}_r to complete the whole solution.

APPENDIX B: PROBABILITY OF COMPLETION, FIRST COMPLETION, AND DISSOCIATION TIME DISTRIBUTIONS

In this section, we outline the calculation of the probability of completion before dissociation from the filament, the first completion time given completion precedes dissociation, and the dissociation time, all for a random walker that is initially occupying the first bound state. The appropriate limit in (A1) is $\kappa \rightarrow 0$. The full solution for \tilde{P}_m is obtained by solving the system of equations in (A13) and substituting the resulting

expressions in (A12). The explicit solution is given by

$$P_m(\epsilon) = \Sigma_m + \frac{\tilde{\varphi}_m [u_1 Q_1(0) - b_1 \Sigma_0(\epsilon + b_2 + K_2(w)u_2)]}{u_1 + (1 + b_1 \tilde{\varphi}_0)[\epsilon + b_2 + K_2(w)u_2]}. \quad (\text{B1})$$

When there is only a single bound state, (B1) reduces to

$$P_m(\epsilon) = \Sigma_m + \frac{\tilde{\varphi}_m [u_1 Q_1(0) - \epsilon b_1 \Sigma_0]}{u_1 + \epsilon(1 + b_1 \tilde{\varphi}_0)}.$$

Note that the expressions above hold for all initial conditions, described by Σ_m and $Q_1(0)$. We consider the case where the random walker is initially occupying the first bound state, that is, $\Sigma_m = 0$ and $Q_1(0) = 1$.

1. Dissociation time distribution

We denote the distribution of the dissociation time by $f_{\text{dis}}(t)$. The molecule is allowed to visit any state any number of times; therefore, we do not have the restriction $u_w = 0$ that we used in the calculation for the completion time. In this respect, $f_{\text{dis}}(t)$ is not a conditional distribution, unlike $f_{\text{comp}}(t)$. The distribution of the first dissociation time can formally be written as

$$f_{\text{dis}}(t) = -\frac{d}{dt} \left[\sum_m P_m(t) + \sum_i Q_i(t) \right], \quad (\text{B2})$$

where the quantity inside the parentheses is the survival probability, the probability that the random walker is still diffusing in the lattice, or the system is in any of the bound states. In the Laplace domain, this expression becomes

$$\tilde{f}_{\text{dis}}(\epsilon) = 1 - \epsilon \left(\sum_m \tilde{P}_m + \sum_i \tilde{Q}_i \right). \quad (\text{B3})$$

The probability of being at any lattice site can be obtained by substituting $\tilde{\varphi}_m(\epsilon)$ and summing over m

$$\sum_m \tilde{P}_m(\epsilon) = \frac{u_1(\alpha + \gamma + 4f + \epsilon)}{(\alpha + \gamma + \epsilon)\{(\alpha + b_1)[\epsilon + b_2 + u_2 K_2(w)] + u_1 \alpha\}}. \quad (\text{B4})$$

Combining this with (A11), the explicit form of the dissociation time is obtained as

$$\tilde{f}_{\text{dis}}(\epsilon) = 1 - \frac{\epsilon \{u_1[4f + \gamma + \epsilon + \alpha(\epsilon)] + \Lambda_w[(\gamma + \epsilon)b_1 + (\gamma + \epsilon + b_1)\alpha(\epsilon) + \alpha(\epsilon)^2]\}}{[\gamma + \epsilon + \alpha(\epsilon)]\{b_1(\epsilon + b_2) + (\epsilon + b_2 + u_1)\alpha(\epsilon) + K_2(w)u_2[b_1 + \alpha(\epsilon)]\}}, \quad (\text{B5})$$

where Λ_w is equal to the quantity in the square brackets on the right-hand side of (A11).

In order to obtain the distribution of dissociation time given the process never reaches completion before dissociation, denoted by $f'_{\text{dis}}(t)$ in Sec. III D, one can eliminate the constant b_w and the variable Q_w in the set of Eqs. (A2) and drop the last equation, thereby removing the possibility of completion. Then the solution of this modified master equation for the dissociation time, performed in the same way as above, would provide $f'_{\text{dis}}(t)$.

2. Probability of completion and the completion time distribution

When the system transitions into the final bound state w , we call the process complete. We can calculate the probability of completion and the first completion time by setting u_w , the rate of leaving the final bound state, to zero and finding out the probability Q_w . Note that this is equivalent to setting $u_w = 0$ in the master equation and solving for Q_w . With this choice, system trajectories that arrive at the final bound state cannot leave, and the distribution of first completion times can be

written as

$$f_{\text{comp}}(t) = \frac{1}{P_{\text{comp}}} \left\{ -\frac{d}{dt} [1 - Q_w(t; u_w = 0)] \right\}, \quad (\text{B6})$$

$$= \frac{1}{P_{\text{comp}}} \left. \frac{dQ_w}{dt} \right|_{u_w=0}, \quad (\text{B7})$$

where P_{comp} is the probability of completion, acting as a normalization constant. Note that the quantity in square brackets in (B6) is the survival probability, defined as the probability of not having visited the w^{th} bound state up to time t . In this respect, (B6) is analogous to (B2), except that f_{comp} is a conditional distribution and needs to be normalized with the probability of completion, P_{comp} . In the Laplace domain, we have

$$\tilde{f}_{\text{comp}}(\epsilon) = \frac{\epsilon \tilde{Q}_w(\epsilon; u_w = 0)}{P_{\text{comp}}}, \quad (\text{B8})$$

since $Q_w(0) = 0$. Note that this is equivalent to solving the master equations (A1) and (A2) by setting $u_w = 0$ first and then calculating Q_w , since the solutions (A10) and (B1) are valid for $u_w = 0$.

P_{comp} is the probability of occupancy of state w in the long-time limit. Therefore, we have $P_{\text{comp}} = Q_w(t \rightarrow \infty; u_w = 0)$. To find P_{comp} we first calculate $\tilde{Q}_w(\epsilon; u_w = 0)$ from (A10). Then, using the limit theorem for the Laplace transform, $\lim_{t \rightarrow \infty} f(t) = \lim_{\epsilon \rightarrow 0} \epsilon \tilde{f}(\epsilon)$, we obtain the explicit result

$$P_{\text{comp}} = \frac{1}{1 + \lambda(w)/(1 + \beta)}, \quad (\text{B9})$$

where

$$\beta = b_1 / \sqrt{\gamma(\gamma + 4f)}, \quad (\text{B10})$$

$$\lambda(w) = [\pi_b(2; w)]^{-1} u_1 (\pi_b(3; w) + u_2 \{\pi_b(4; w) + \dots + u_{w-2} [\pi_b(w; w) + u_{w-1}] \dots \}), \quad (\text{B11})$$

$$\pi_b(i; w) = \prod_{k=i}^w b_k. \quad (\text{B12})$$

Note that $w \geq 2$ and $\lambda(2) = u_1$.

The Laplace transform of the first completion time valid for $w \geq 2$ is obtained from (B8) and is explicitly given by

$$\begin{aligned} \tilde{f}_{\text{comp}}(\epsilon) &= \frac{(-1)^{w-1}}{P_{\text{comp}}} \left(\prod_{i=2}^w K_i(w) \right) \\ &\times \left\{ \frac{\epsilon}{\epsilon + b_2 + u_1 \alpha(\epsilon) / [\alpha(\epsilon) + b_1] + u_2 K_2(w)} \right\} \Big|_{u_w=0}, \end{aligned} \quad (\text{B13})$$

where $\alpha(\epsilon) = \sqrt{(\gamma + \epsilon)(\gamma + 4f + \epsilon)}$ and $K_i(w)$ is defined in (A8).

APPENDIX C: RESIDENCE TIME DISTRIBUTION

In this section, we are concerned with the case where the random walker is initially occupying the first bound state and can be absorbed if it travels far enough from the binding site. We consider the residence time of the random walker in a symmetric interval centered around the target site, given the walker does not dissociate from the lattice before exiting this interval. To calculate the residence time (equivalently, the first exit or escape time), we consider (A1) in the limit $\kappa \rightarrow \infty$ and $\gamma \rightarrow 0$, which amounts to placing perfectly absorbing boundaries at $m = -r$ and r and eliminating the possibility of dissociation from the lattice.

After solving for \tilde{P}_m using the same method presented in the previous sections, the Laplace transform of the residence time distribution, which we denote by \tilde{f}_{res} , is then found by using (B3). The probability of finding the molecule in the unbound state is obtained by summing \tilde{P}_m over all lattice sites, resulting in

$$\sum_m P_m = \frac{u_1 [4f + \epsilon + \delta(\epsilon)] \{-(2f)^r + [2f + \epsilon + \delta(\epsilon)]^r\}^2}{[\epsilon + \delta(\epsilon)] (u_1 [(2f)^{2r} + \beta(\epsilon)] \delta(\epsilon) - [\epsilon + b_2 + u_2 K_2(w)] \{(2f)^{2r} [b_1 - \delta(\epsilon)] - \beta(\epsilon) [(b_1 + \delta(\epsilon))]\}},$$

where $\beta(\epsilon) = [\sqrt{\epsilon(4f + \epsilon)} + 2f + \epsilon]^{2r}$ and $\delta(\epsilon) = \sqrt{\epsilon(4f + \epsilon)}$. The probability of being in any bound state is given by the sum in (A11). Substituting these two sums in (B3) and performing the algebra, we obtain

$$\tilde{f}_{\text{res}}(\epsilon) = 1 - \frac{\epsilon (u_1 [4f + \epsilon + \delta(\epsilon)] \{(2f)^r - [2f + \epsilon + \delta(\epsilon)]^r\}^2 + \Lambda_w [\epsilon + \delta(\epsilon)] \{(2f)^{2r} [-b_1 + \delta(\epsilon)] + \beta(\epsilon) [b_1 + \delta(\epsilon)]\})}{[\epsilon + \delta(\epsilon)] (u_1 [(2f)^{2r} + \beta(\epsilon)] \delta(\epsilon) + [\epsilon + b_2 + u_2 K_2(w)] \{(2f)^{2r} [-b_1 + \delta(\epsilon)] + \beta(\epsilon) [b_1 + \delta(\epsilon)]\})}, \quad (\text{C1})$$

where Λ_w is equal to the quantity in the square brackets on the right-hand side of (A11). We denote the residency time as T_{res} . The moments of T_{res} can be calculated from

$$\langle T_{\text{res}}^n \rangle = (-1)^n \lim_{\epsilon \rightarrow 0} \frac{d^n \tilde{f}_{\text{res}}}{d\epsilon^n}. \quad (\text{C2})$$

The mean and variance of T_{res} , normalized by f^{-1} and f^{-2} , respectively (f being the hopping rate along the lattice, proportional to the 1D diffusion coefficient along the DNA), for the first few values of w are given by

$w = 1$:

$$\mu_{\text{res}} = \frac{1}{2} \left(r^2 + \frac{2 + r b_1}{u_1} \right), \quad \sigma_{\text{res}}^2 = \frac{3(2 + r b_1)^2 + 2(r + 2r^3) b_1 u_1 + r^2 (1 + 2r^2) u_1^2}{12 u_1^2},$$

TABLE I. Comparison of numerically exact and simulated values of P_{comp} . Values of u_* and w are indicated in the first two columns, and all other parameter values are fixed at $f = 1$, $b_1 = 2$, $u_1 = 1$, $b_* = 0.25$, $\gamma = 0.1$. In computing simulation results, an ensemble of $N = 10^5$ independent simulation runs were obtained, and this process was repeated 20 times to compute error due to finite sample size. $P_{\text{comp}}^{\text{seq}}$ (sim) and $P_{\text{comp}}^{\text{ran}}$ (sim) correspond to simulation results, given in the format $x [y, z]$, where x is the average, and $[y, z]$ is the 95% confidence interval computed via bootstrapping.

u_*	w	$P_{\text{comp}}^{\text{seq}}$ [Eq. (1)]	$P_{\text{comp}}^{\text{seq}}$ (sim)	$P_{\text{comp}}^{\text{ran}}$ [Eq. (1)]	$P_{\text{comp}}^{\text{ran}}$ (sim)
0.10	2	0.5076	0.5078 [0.5071, 0.5086]	0.5076	0.5084 [0.5076, 0.5092]
0.40	3	0.2839	0.2840 [0.2832, 0.2847]	0.4423	0.4420 [0.4413, 0.4429]
0.08	4	0.4202	0.4206 [0.4199, 0.4213]	0.7101	0.7105 [0.7099, 0.7111]
0.19	5	0.2708	0.2699 [0.2694, 0.2705]	0.6863	0.6859 [0.6852, 0.6866]
0.13	6	0.3397	0.3398 [0.3394, 0.3402]	0.8007	0.8009 [0.8005, 0.8014]
0.31	7	0.0858	0.0858 [0.0855, 0.0861]	0.5531	0.5527 [0.5520, 0.5534]
0.15	8	0.2979	0.2983 [0.2976, 0.2990]	0.8571	0.8570 [0.8566, 0.8575]

$w = 2$:

$$\mu_{\text{res}} = \frac{1}{2} \left[r^2 + \frac{(2 + rb_1)(1 + \frac{b_2}{u_2})}{u_1} \right],$$

$$\sigma_{\text{res}}^2 = (12u_1^2u_2^2)^{-1} (3(2 + rb_1)^2b_2^2 + (3(2 + rb_1)^2 + 2(r + 2r^3)b_1u_1 + r^2(1 + 2r^2)u_2^2)u_2^2 + 2b_2\{3(2 + rb_1)^2u_2 + u_1[12 + rb_1(6 + u_2 + 2r^2u_2)]\}),$$

$w = 3$:

$$\mu_{\text{res}} = \frac{1}{2} \left\{ r^2 + \frac{(2 + rb_1)[1 + \frac{b_2(b_3 + u_3)}{u_2u_3}]}{u_1} \right\},$$

$$\sigma_{\text{res}}^2 = (12u_1^2u_2^2u_3^2)^{-1} [(3(2 + rb_1)^2 + 2(r + 2r^3)b_1u_1 + r^2(1 + 2r^2)u_1^2)u_2^2u_3^2 + 3(2 + rb_1)^2b_2^2(b_3 + u_3)^2 + 2b_2(6(2 + rb_1)b_3u_1(b_3 + u_2) + b_3\{3(2 + rb_1)^2u_2 + u_1[24 + rb_1(12 + u_2 + 2r^2u_2)]\})u_3 + \{3(2 + rb_1)^2u_2 + u_1[12 + rb_1(6 + u_2 + 2r^2u_2)]\}u_3^2],$$

where r is considered dimensionless, an integer corresponding to the number of lattice sites from the target site.

APPENDIX D: NUMERICAL INVERSE LAPLACE TRANSFORM

Based on previous experience, we pick the Gaver-Stehfest method [36,37] for the numerical inversion of Laplace transforms, which only requires the evaluation of the transformed function at real values of the Laplace variable and is suitable for bounded functions such as first-passage time distributions. We employ the algorithm described by Abate and Whitt [38] to approximate the inverse Laplace transform of $\tilde{f}(\epsilon)$ as

$$f(t) = \frac{\ln 2}{t} \sum_{k=1}^{2M} w_k \tilde{f}\left(k \frac{\ln 2}{t}\right),$$

$$w_k = (-1)^{M+k} \sum_{j=\lfloor (k+1)/2 \rfloor}^{\min(k, M)} j^{M+1} \times \frac{2j!}{(j!)^2(M-j)!(2j-k)!(k-j)!},$$

where M is a positive integer and $\lfloor (k+1)/2 \rfloor$ means the largest integer less than or equal to $(k+1)/2$. The number M is chosen

based on the available numerical precision, and, according to the estimate provided in Ref. [38], the result has around 2.2M digits of precision.

APPENDIX E: COMPARISON WITH SIMULATIONS

Assembly formation process is simulated according to the model description given in Sec. II using the Gillespie algorithm [39] that proceeds by determining the time of the next transition in the system, as well as which transition is going to take place. The system state is then updated, and the process is repeated until the random walker: (1) arrives at the bound state w , (2) decays (at rate γ while at an unbound state), or (3) reaches one of the sites $-r$ or r for the computation of T_{comp} , T_{dis} , and T_{res} , respectively. P_{comp} is obtained by computing the ratio of the number of times the assembly forms to the total number of independent simulation runs. In each simulation run, the system starts at the same state where the random walker is occupying the first bound state with certainty, in accordance with the model description in Sec. II.

In Table I and Fig. 11, we display the comparison between analytical and simulation results for P_{comp} and the cumulative distributions of T_{comp} , T_{dis} , and T_{res} for several randomly

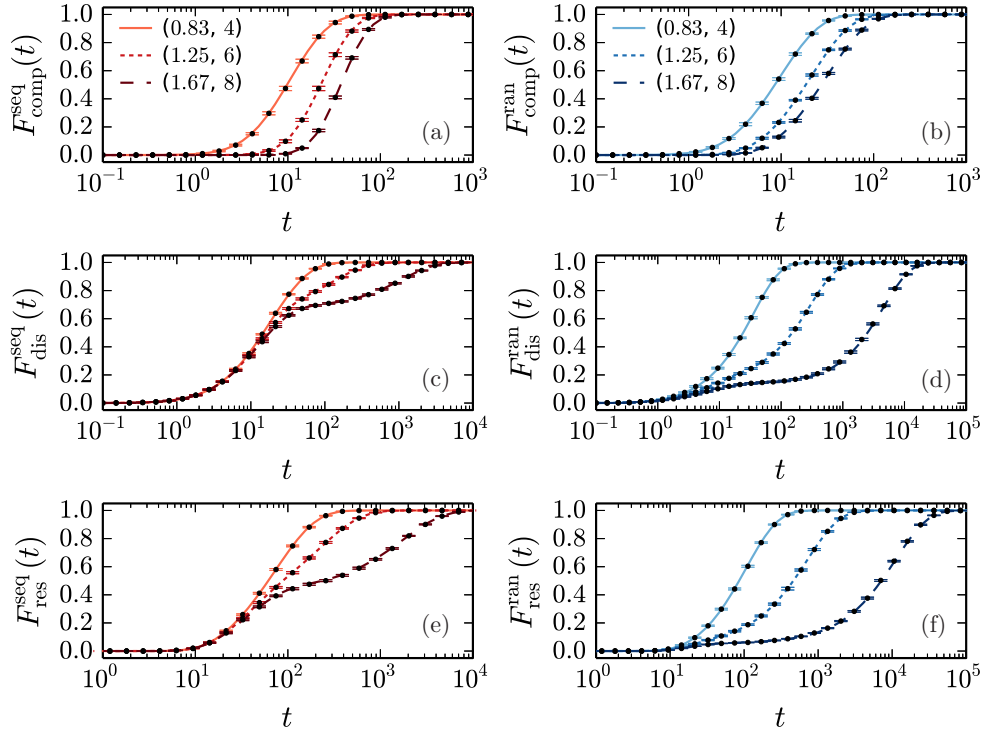


FIG. 11. (Color online) Comparison of numerically exact and simulated cumulative probability distributions of T_{comp} , T_{dis} , and T_{res} , defined in (E1). Curves correspond to the numerical inverse Laplace transform of (E1) [with the use of Eqs. (B13), (B5) and (C1)], and black filled circles along with error bars represent simulation results. Legends of the topmost graphs apply to all graphs in the same column, and values in parenthesis correspond to $(b_*/u_*, w)$. For the calculation of F_{res} , we set $r = 10$. Rest of the parameter values are the same as those in the caption of Table I. In computing simulation results, an ensemble of $N = 10^3$ independent simulation runs were obtained, and this process was repeated 20 times to compute error bars. Error bars correspond to 95% confidence intervals computed via bootstrapping.

selected parameter values. The cumulative probability distribution of T_x , which is straightforward to calculate from simulation data without any binning, is defined as

$$F_x^y(t) = \int_0^t ds f_x^y(s) = \mathcal{L}^{-1} \left\{ \frac{\tilde{f}_x^y(\epsilon)}{\epsilon} \right\}, \quad (\text{E1})$$

where y is either “seq” or “ran” for sequential and random binding models, respectively, and the inverse Laplace transform, denoted by \mathcal{L}^{-1} , is performed as in Appendix D. Parameter values are given in captions. As both sets of comparisons show, numerically exact and simulation results that are obtained independently are in excellent agreement.

-
- [1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 5th ed. (Garland Science, New York, 2007).
 - [2] T. J. Stasevich and J. G. McNally, *Chromosoma* **120**, 533 (2011).
 - [3] F. Mueller, T. J. Stasevich, D. Mazza, and J. G. McNally, *Crit. Rev. Biochem. Mol. Biol.* **48**, 492 (2013).
 - [4] A. Coulon, C. C. Chow, R. H. Singer, and D. R. Larson, *Nat. Rev. Genet.* **14**, 572 (2013).
 - [5] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf, *Science* **336**, 1595 (2012).
 - [6] O. G. Berg, R. B. Winter, and P. H. von Hippel, *Biochemistry (Mosc.)* **20**, 6929 (1981).
 - [7] P. H. von Hippel and O. G. Berg, *J. Biol. Chem.* **264**, 675 (1989).
 - [8] G. Tkačik and W. Bialek, *Phys. Rev. E* **79**, 051901 (2009).
 - [9] J. Pajmians and P. R. ten Wolde, *Phys. Rev. E* **90**, 032708 (2014).
 - [10] R. Phillips, J. Kondev, and J. Theriot, *Physical Biology of the Cell*, 1st ed. (Garland Science, New York, 2008).
 - [11] B. Meyer, O. Bénichou, Y. Kafri, and R. Voituriez, *Biophys. J.* **102**, 2186 (2012).
 - [12] K. Kaizu, W. de Ronde, J. Pajmians, K. Takahashi, F. Tostevin, and P. R. ten Wolde, *Biophys. J.* **106**, 976 (2014).
 - [13] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, 2010).
 - [14] M. R. D’Orsogna and T. Chou, *Phys. Rev. Lett.* **95**, 170603 (2005).
 - [15] I. Koyama-Honda, K. Ritchie, T. Fujiwara, R. Iino, H. Murakoshi, R. S. Kasai, and A. Kusumi, *Biophys. J.* **88**, 2126 (2005).
 - [16] R. S. Kasai, K. G. N. Suzuki, E. R. Prossnitz, I. Koyama-Honda, C. Nakada, T. K. Fujiwara, and A. Kusumi, *J. Cell. Biol.* **192**, 463 (2011).
 - [17] S. Yamamoto, D. De, K. Hidaka, K. K. Kim, M. Endo, and H. Sugiyama, *Nano Lett.* **14**, 2286 (2014).

- [18] O. Bénichou, C. Chevalier, J. Klafter, B. Meyer, and R. Voituriez, *Nat. Chem.* **2**, 472 (2010).
- [19] M. S. Luijsterburg, G. v. Bornstaedt, A. M. Gourdin, A. Z. Politi, M. J. Mon, D. O. Warmerdam, J. Goedhart, W. Vermeulen, R. v. Driel, and T. Hfer, *J. Cell Biol.* **189**, 445 (2010).
- [20] K. Takahashi, S. Tanase-Nicola, and P. R. ten Wolde, *Proc. Natl. Acad. Sci. USA* **107**, 2473 (2010).
- [21] N. V. Brilliantov and P. L. Krapivsky, *J. Phys. A* **24**, 4789 (1991).
- [22] H. Kallabis, P. L. Krapivsky, and D. E. Wolf, *Eur. Phys. J. B* **5**, 801 (1998).
- [23] M. Endo, K. Tatsumi, K. Terushima, Y. Katsuda, K. Hidaka, Y. Harada, and H. Sugiyama, *Angew. Chem. Int. Ed.* **51**, 8778 (2012).
- [24] G.-W. Li, O. G. Berg, and J. Elf, *Nat. Phys.* **5**, 294 (2009).
- [25] J. G. Powles, M. J. D. Mallett, G. Rickayzen, and W. A. B. Evans, *Proc. R. Soc. Lond. A* **436**, 391 (1992).
- [26] O. K. Dudko, A. M. Berezhkovskii, and G. H. Weiss, *J. Chem. Phys.* **121**, 11283 (2004).
- [27] Z. Kalay, P. E. Parris, and V. M. Kenkre, *J. Phys.: Condens. Matter* **20**, 245105 (2008).
- [28] Z. Kalay, *J. Phys. A* **45**, 215001 (2012).
- [29] V. M. Kenkre and P. Reineker, *Exciton Dynamics in Molecular Crystals and Aggregates*, Springer Tracts in Modern Physics Vol. 94, edited by G. Hoehler (Springer, Berlin, 1982).
- [30] B. D. Hughes, *Random Walks and Random Environments: Volume 2: Random Environments* (Oxford University Press, Oxford, 1996).
- [31] A. Burger, A. M. Walczak, and P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **107**, 4016 (2010).
- [32] J. H. Kim and R. G. Larson, *Nucl. Acids Res.* **35**, 3848 (2007).
- [33] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1965).
- [34] G. E. Roberts and H. Kaufman, *Table of Laplace Transforms* (Saunders, Philadelphia, PA, 1966).
- [35] Wikipedia, Kettenbruch, *Wikipedia, Die Freie Enzyklopädie* (2015) [accessed 10 July 2015].
- [36] D. P. Gaver, *Oper. Res.* **14**, 444 (1966).
- [37] H. Stehfest, *Commun. ACM* **13**, 47 (1970).
- [38] J. Abate and W. Whitt, *INFORMS J. Comput.* **18**, 408 (2006).
- [39] D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).