Assembly of heteropolymers via a network of reaction coordinates

Davide Chiuchiù [●],[1] James Ferrare [●],[1,2] and Simone Pigolotti[1,*]

[1]*Biological Complexity Unit, Okinawa Institute for Science and Technology, 1919-1 Tancha, Onna, Kunigami-gun, Okinawa 904-0412, Japan*
[2]*Tulane University, 6823 St. Charles Avenue, New Orleans, Lousiana 70118, USA*

In biochemistry, heteropolymers encoding biological information are assembled out of equilibrium by sequentially incorporating available monomers found in the environment. Current models of polymerization treat monomer incorporation as a sequence of discrete chemical reactions between intermediate metastable states. In this paper, we use ideas from reaction rate theory and describe nonequilibrium assembly of a heteropolymer via a continuous reaction coordinate. Our approach allows for estimating the copy error and incorporation speed from the Gibbs free energy landscape of the process. We apply our theory to several examples from a simple reaction characterized by a free energy barrier to more complex cases incorporating error correction mechanisms, such as kinetic proofreading.

I. INTRODUCTION

DNA, RNA, and proteins are the building blocks of all living systems. These heteropolymers are assembled to match a template; only a very small numbers of mismatches with the template are tolerable for maintaining biological information and for correct functioning of cells. However, the binding energies of different monomers usually differ by only a few $k_B T$'s, where k_B is the Boltzmann constant and T is the temperature. This means that, at physiological temperature, mismatches cannot be completely suppressed [1].

Our aim is to describe the chemical processes responsible for these errors. Specifically, we consider sequential assembly of heteropolymers where each incorporated monomer can be a right (r) or a wrong (w) match with a template. These two different outcomes can be represented as competing chemical reactions,

$$
\begin{array}{c}
hr \\
\nearrow^{+r} \\
\swarrow_{-r} \\
h \\
\searrow^{+w} \\
\nwarrow_{-w} \\
hw
\end{array}
\qquad (1)
$$
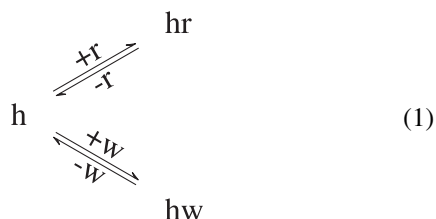
where h is the heteropolymer produced so far and hr/hw are the same heteropolymer with an addition of a r/w monomer at the tip, respectively. Each monomer incorporation is iteratively followed by a new one so that the whole polymerization process is described by the tree-shaped network of chemical reactions [2,3] in Fig. 1(a).

To achieve accurate and fast assembly, the reactions in Eq. (1) involve several intermediate steps, such as initial monomer discrimination [4], kinetic proofreading [4–6], and mismatch repair [7,8]. In general, each one of these error-correction mechanisms contributes simultaneously to polymerization accuracy, speed [9–17], and energetic cost [14,18–21].

Two approaches can provide insight into the error-correction mechanisms underlying heteropolymer assembly. The first approach is to measure their kinetic rates under different experimental conditions [9]. The second approach is to simulate heteropolymer assembly using molecular dynamics [22]. From the molecular dynamics, one can project the numerous degrees of freedom into a one-dimensional collective variable called the reaction coordinate [23]. The reaction coordinate simplifies a chemical process into a one-dimensional random motion [23–25]. The parameters of this random motion depend on the underlying reactants dynamics [26–28] and on the projection technique [23,28,29].

Although successful in describing protein folding [24,25,30] and in modeling reaction rates [19,30], approaches based on reaction coordinates found little use in studies of polymerization speed and accuracy. In principle, both reactions in Eq. (1) can be described by means of a reaction coordinate [Fig. 1(b)]. However, to study the complete polymerization process, we need to join the reaction coordinates characterizing each branch in Fig. 1(a). Mathematically, this amounts to impose appropriate boundary conditions at the nodes of the reaction network.

In this paper, we develop a model of heteropolymer assembly based on reaction coordinates and use it to compute the accuracy and speed of polymerization in different conditions. The paper is organized as follows. In Sec. II, we introduce our model. From the reaction coordinate, we derive effective incorporation and removal probabilities of right and wrong monomers. In Sec. III, we compute the accuracy and speed of a general heteropolymer assembly. In Sec. IV, we consider examples characterized by different Gibbs free energy landscapes. In Sec. V, we generalize our results to a case where the reaction leading to monomer incorporation is complemented by kinetic proofreading. Section VI is devoted to conclusions and perspectives.
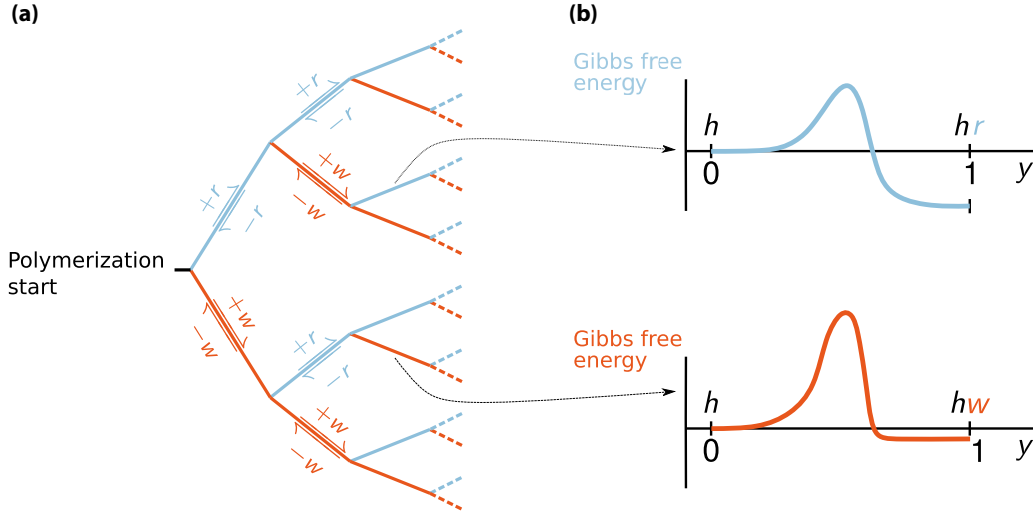
---

*simone.pigolotti@oist.jp

FIG. 1. Synthesis of heteropolymers. (a) Network of incorporation and removal reactions to synthesize a heteropolymer. Each edge in the network represents the chemical reaction $h \rightleftharpoons hx$ where h is the heteropolymer produced so far and hx is the same heteropolymer with the addition of monomer $x \in \{r, w\}$ at the tip. Three reactions compete at the nodes of the network: removal of the last incorporated monomer, incorporation of a r monomer, and incorporation of a w monomer. (b) Reaction coordinate description of the incorporation and removal reactions. The initial and final points of the free energy landscapes correspond to the reactants and products of the incorporation and removal reactions, respectively.

II. MODEL

We define our model of heteropolymer assembly with reaction coordinates through the following steps. We first introduce the reaction coordinate and the free energy landscape in each chemical reaction of the polymerization network. We then study the dynamics of the reaction coordinate dynamics and its boundary conditions at the nodes of the network. Finally, we compute the probabilities to incorporate and remove one monomer along each reaction coordinate.

A. Reaction coordinate and Gibbs free energy of the heteropolymer

We introduce the continuous reaction coordinate y along each edge of the polymerization network, Fig. 1(a). Without loss of generality, we choose the units of the reaction coordinate so that $y \in [0, 1]$, where $y = 0$ and $y = 1$ correspond to h and hx, respectively, i.e., to the states before and after monomer incorporation, see Fig. 1(b).

Each point along this reaction coordinate is characterized by a Gibbs free energy $G^{hx}(y)$ (from now on simply free energy). Such free energy depends on the previously incorporated sequence of monomers (h), on the candidate monomer to be incorporated (x), and on the stage of the incorporation process, i.e., the value of y. Implicitly, $G^{hx}(y)$ also depends on the reactant and product concentrations.

We introduce the free energy increments from the beginning of each incorporation reaction,

$$\Delta G^x(y) = G^{hx}(y) - G^{hx}(0), \qquad (2)$$

see Fig. 2. The free energy increments depend on the candidate monomer x but not on the whole history of incorporated monomers h. With this notation, the (absolute) binding free energy of monomer x is equal to $-\Delta G^x(1)$.

The free energy must be a continuous function of y and must also vary continuously when crossing the nodes of the network in Fig. 1. This means that we can decompose the free energy at an arbitrary stage of the polymerization process as

$$
\begin{aligned}
G^{hx}(y) &= G^{hx}(0) + \Delta G^x(y) \\
&= G^h(1) + \Delta G^x(y) \\
&= \left(\sum_{i \in h} \Delta G^i(1) \right) + \Delta G^x(y).
\end{aligned}
\qquad (3)
$$

B. Stochastic dynamics of the reaction coordinate and boundary conditions

Because of thermal fluctuations, the reaction coordinate y evolves according to a Langevin equation,

$$\frac{dy}{dt} = -\mu \frac{d}{dy} G^{hx}(y) + \sqrt{2D}\xi(t), \qquad (4)$$

where μ is a mobility, D is a diffusion coefficient, and $\xi(t)$ is white noise with $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$ where $\langle \ldots \rangle$ denotes an average over noise [31]. We assume that D satisfies the Einstein relation $D = k_B T \mu$ with temperature T and Boltzmann constant k_B. We also assume that D, μ, and T are constant. When the reaction coordinate reaches the boundaries, either $y = 0$ or $y = 1$, a new incorporation or removal reaction is commenced.

Equation (4) needs to be complemented by rules to specify which reaction initiates at the nodes of the reaction network. To this aim, we consider two intermediate values of the reaction coordinate: $y = \epsilon$ and $y = 1 - \epsilon$ with $\epsilon \ll 1$. Using these values, we coarse grain the evolution of the reaction
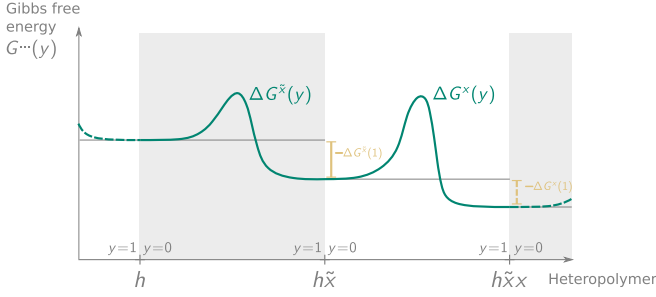
FIG. 2. Free energy $G^{\cdots}(y)$ of the heteropolymer depends on the reaction coordinate y and the sequence of incorporated monomers. The superscript of the free energy indicates the stage of the polymerization process (h, $h\tilde{x}$, or $h\tilde{x}x$ in this case). The functions $\Delta G^{\tilde{x}}(y)$ and $\Delta G^{x}(y)$ are the free energy increments along the reaction coordinate to incorporate $\tilde{x}$ and x, respectively. The total binding free energies for monomer $\tilde{x}$ and x are $-\Delta G^{\tilde{x}}(1)$ and $-\Delta G^{x}(1)$, respectively.

coordinate y as

$$(y=0) \xrightleftharpoons[\pi^x_{\varepsilon,0}]{\pi^x_{0,\varepsilon}} (y=\varepsilon) \xrightleftharpoons[\pi^x_{1-\varepsilon,\varepsilon}]{\pi^x_{\varepsilon,1-\varepsilon}} (y=1-\varepsilon) \xrightleftharpoons[\pi^x_{1,1-\varepsilon}]{\pi^x_{1-\varepsilon,1}} (y=1), \quad (5)$$

where the quantities $\pi^x_{\tilde{y},y}$ are the first-passage probabilities from y to $\tilde{y}$. For example, $\pi^x_{1-\epsilon,\epsilon}$ is the probability that the reaction coordinate reaches $y = 1 - \epsilon$ from $y = \epsilon$ without having reached $y = 0$ before.

The representation in Eq. (5) separates the dynamics in proximity of the nodes of Fig. 1(b) from the dynamics in the interval $[\epsilon, 1 - \epsilon]$. Thanks to this separation, we use detailed balance, probability conservation close to the nodes, and the continuity of $G^{hx}(y)$ to compute the first-passage probabilities $\pi^x_{\epsilon,0}$ and $\pi^x_{1-\epsilon,1}$ (see Appendix A). This procedure results in

$$\pi^x_{\epsilon,0} = \pi^x_{1-\epsilon,1} = \tfrac{1}{3} + O(\epsilon). \quad (6)$$

We compute the first-passage probabilities in the interior applying standard techniques [31,32] to the Fokker-Planck equation associated with Eq. (4). We obtain

$$\pi^x_{0,\epsilon} = \frac{\int_{\epsilon}^{1-\epsilon} \psi^x(y)dy}{\int_{0}^{1-\epsilon} \psi^x(y)dy},$$

$$\pi^x_{\epsilon,1-\epsilon} = \frac{\int_{1-\epsilon}^{1} \psi^x(y)dy}{\int_{\epsilon}^{1} \psi^x(y)dy},$$

$$\pi^x_{1-\epsilon,\epsilon} = \frac{\int_{0}^{\epsilon} \psi^x(y)dy}{\int_{0}^{1-\epsilon} \psi^x(y)dy},$$

$$\pi^x_{1,1-\epsilon} = \frac{\int_{\epsilon}^{1-\epsilon} \psi^x(y)dy}{\int_{\epsilon}^{1} \psi^x(y)dy}, \quad (7)$$

with

$$\psi^x(y) = \exp\left[\int_0^y \frac{\mu}{D} \frac{\partial \Delta G^x(z)}{\partial z} dz\right] = \exp\left[\frac{\Delta G^x(y)}{k_B T}\right], \quad (8)$$

where the last equality follows from the relation $D = k_B T \mu$.

C. Effective probabilities of monomer incorporation and rejection

From the probabilities $\pi^x_{\tilde{y},y}$, we now compute the effective probabilities $p^x_{\rightarrow}$ and $p^x_{\leftarrow}$ to incorporate and reject monomer x along each edge of the reaction network in Fig. 1. To this end, we assume that the coarse grained dynamics in Eq. (5) is at steady state. We then use adiabatic elimination [33] to obtain (see Appendix B)

$$p^x_{\rightarrow} = \frac{\epsilon}{3} \frac{1}{\int_0^1 \exp\left[\frac{\Delta G^x(y)}{k_B T}\right] dy} + O(\epsilon), \quad (9a)$$

$$p^x_{\leftarrow} = \frac{\epsilon}{3} \frac{\exp\left[\frac{\Delta G^x(1)}{k_B T}\right]}{\int_0^1 \exp\left[\frac{\Delta G^x(y)}{k_B T}\right] dy} + O(\epsilon). \quad (9b)$$

Equations (9) relate the free energy landscapes $G^x(y)$ and the incorporation and removal probabilities of the polymerization process. From Eq. (9), the probabilities are consistent with the detailed balance condition,

$$\frac{p^x_{\rightarrow}}{p^x_{\leftarrow}} = \exp\left[-\frac{\Delta G^x(1)}{k_B T}\right], \quad (10)$$

which connects the ratios of forward and backward probabilities to the binding free energy $-\Delta G^x(1)$, see Fig. 2.

III. RESULTS

We now address the accuracy and speed of a polymerization process in the reaction coordinate framework. We consider a copy polymer made up of a number N^r of right monomers and N^w of wrong monomers with $N = N^r + N^w$. For large N, we define the error rate,

$$\eta = \lim_{N\to\infty} \frac{N^w}{N}. \quad (11)$$

To compute η from the incorporation and removal probabilities $p^x_{\rightarrow}$ and $p^x_{\leftarrow}$, we first recast Eq. (11) into the implicit equation,

$$\frac{\eta}{1-\eta} = \lim_{N\to\infty} \frac{N^w}{N^r} = \lim_{n\to\infty} \frac{n^w_{\rightarrow} - n^w_{\leftarrow}}{n^r_{\rightarrow} - n^r_{\leftarrow}}, \quad (12)$$

where we have introduced the numbers $n^r_{\rightarrow}$, $n^r_{\leftarrow}$, $n^w_{\rightarrow}$, and $n^w_{\leftarrow}$ of r and w incorporation and removal reactions which have occurred in the process and n is the total number of observed chemical reactions. For large n, we have

$$n^r_{\rightarrow} \sim nc p^r_{\rightarrow}, \quad (13a)$$

$$n^r_{\leftarrow} \sim nc(1-\eta)p^r_{\leftarrow}, \quad (13b)$$

$$n^w_{\rightarrow} \sim nc p^w_{\rightarrow}, \quad (13c)$$

$$n^w_{\leftarrow} \sim nc\eta p^w_{\leftarrow}, \quad (13d)$$

where $c = [p^r_{\rightarrow} + (1-\eta)p^r_{\leftarrow} + p^w_{\rightarrow} + \eta p^w_{\leftarrow}]^{-1}$ is a normalization constant so that $n = n^r_{\rightarrow} + n^r_{\leftarrow} + n^w_{\rightarrow} + n^w_{\leftarrow}$. Substituting Eqs. (13) into Eq. (12) gives

$$\frac{\eta}{1-\eta} = \frac{p^w_{\rightarrow} - \eta p^w_{\leftarrow}}{p^r_{\rightarrow} - (1-\eta)p^r_{\leftarrow}}. \quad (14)$$

Equation (14) is a general "self-consistency" relation for the error rate that holds also for discrete models of polymerization [2,3,14]. In our case, we substitute Eqs. (9) in Eq. (14)

and take the limit $\epsilon \to 0$, obtaining

$$\frac{\eta}{1-\eta} = \left(\frac{\eta - \exp\left[-\frac{\Delta G^w(1)}{k_B T}\right]}{1 - \eta - \exp\left[-\frac{\Delta G^r(1)}{k_B T}\right]} \right)$$
$$\times \left(\exp\left[\frac{\Delta G^w(1) - \Delta G^r(1)}{k_B T}\right] \frac{\int_0^1 \exp\left[\frac{\Delta G^r(y)}{k_B T}\right] dy}{\int_0^1 \exp\left[\frac{\Delta G^w(y)}{k_B T}\right] dy} \right).$$

(15)

Solving Eq. (15) for η yields an explicit expression of the error rate from the energy potentials.

Equation (15) identifies different regimes of error correction. To identify a first regime, we observe that

$$\eta = \frac{1}{1 + \exp\left[\frac{\Delta G^w(1) - \Delta G^r(1)}{k_B T}\right]} \quad \text{if}$$
$$\exp\left[\frac{\Delta G^w(1) - \Delta G^r(1)}{k_B T}\right] \frac{\int_0^1 \exp\left[\frac{\Delta G^r(y)}{k_B T}\right] dy}{\int_0^1 \exp\left[\frac{\Delta G^w(y)}{k_B T}\right] dy} = 1,$$

(16)

In the regime where Eq. (16) holds, the error depends only on the binding free energy difference $\Delta G^w(1) - \Delta G^r(1)$. This regime is called the energetic discrimination regime in the literature [3,16]. Systems near equilibrium operate in this regime because the Boltzmann factors of the binding free energies determine, via detailed balance, the probabilities to incorporate different monomers.

To identify a second error-correction regime in Eq. (15), we consider the case where $\Delta G^r(y)$ and $\Delta G^w(y)$ are characterized by energy barriers with heights δ^r and δ^w, respectively (see Fig. 1(b) and Kramers [19]). When such barriers are large, we can approximate the integrals in Eq. (15) by using the Laplace method [34],

$$\frac{\int_0^1 \exp\left[\frac{\Delta G^r(y)}{k_B T}\right] dy}{\int_0^1 \exp\left[\frac{\Delta G^w(y)}{k_B T}\right] dy} \approx \exp\left[\frac{\delta^r - \delta^w}{k_B T}\right] \sqrt{\frac{\Sigma^w}{\Sigma^r}}, \quad (17)$$

where Σ^r and Σ^w are the curvatures of $\Delta G^r(y)$ and $\Delta G^w(y)$ at their maxima, respectively. Equation (17) implies that activation barriers suppress the polymerization error via the second term in parentheses in Eq. (15). The regime where this suppression occurs is the kinetic discrimination regime [3,16]. The first factor on the right-hand side of Eq. (17) represents the contribution of a difference in activation energy barrier $\delta^r - \delta^w$ to accuracy. This effect is also present in models based on discrete-step reactions [2,3,16,18]. The factor $\sqrt{\Sigma^w/\Sigma^r}$ is a correction to activation energies based on the width of the activation barriers. This factor permits kinetic discrimination at equal barrier heights, provided that the barrier for right monomers is significantly more narrow than for wrong monomers.

We estimate the average polymerization speed using a similar argument to that leading to Eq. (15). For large number N of incorporated monomers, the average speed is equal to N divided by the total time $\mathcal{T}$ needed to assemble the polymer,

$$v = \lim_{N \to \infty} \frac{N}{\mathcal{T}} = \lim_{N \to \infty} \frac{(n^r_\to - n^r_\leftarrow) + (n^w_\to - n^w_\leftarrow)}{\mathcal{T}}, \quad (18)$$

where we expressed N in terms of the number of incorporation and removal reactions. For large N, we can approximate the polymerization time as

$$\mathcal{T} \sim n\langle\tau\rangle, \quad (19)$$

where $\langle\tau\rangle$ is the average time it takes to either incorporate or remove a monomer. Substituting Eqs. (13) and (19) into Eq. (18) gives the estimate for the polymerization speed,

$$v = \frac{c[p^r_\to - (1-\eta)p^r_\leftarrow + p^w_\to - \eta p^w_\leftarrow]}{\langle\tau\rangle}. \quad (20)$$

The numerator of Eq. (18) is the probability of an incorporation minus the probability of a removal, whereas the denominator provides the timescale of these events. In practice, calculating $\langle\tau\rangle$ is not straightforward since one has to take into account contributions from incorporation attempts that are not finalized. In Appendix C, we provide a more formal derivation of Eq. (20) together with an explicit expression for $\langle\tau\rangle$.

IV. EXAMPLES

To address the validity and practical implications of Eqs. (15) and (20), we consider two examples of potentials $\Delta G^r(y)$ and $\Delta G^w(y)$. In both cases, we work in dimensionless units by fixing $D = 1$, $k_B T = 1$, and $\mu = 1$.

A. Linear potential

As a first example, we consider linear free energy landscapes,

$$\Delta G^r(y) = -m_r y, \quad (21a)$$
$$\Delta G^w(y) = -m_w y. \quad (21b)$$

Despite their simplicity, the potentials in Eq. (21) are useful to understand the physics of the process. Upon increasing the slopes m_r and m_w, polymerization becomes increasingly irreversible. Substituting the potentials Eq. (21) into the expression for the error Eq. (15) and performing the integrals, we obtain

$$\frac{\eta}{1-\eta} = \frac{m_w(1 - e^{-m_r})[1 - e^{-m_w}\eta]}{m_r(1 - e^{-m_w})[1 + e^{-m_r}(1-\eta)]}, \quad (22)$$

which implies

$$\eta = \frac{m_w}{m_r + m_w} \quad \text{for } m_r, m_w \gg 1. \quad (23)$$

The exact solution of Eq. (22) shows that the error is approximately a function of m_w/m_r when m_r, m_w are large, as predicted by Eq. (23), Fig. 3(a). We compared the predictions from Eqs. (22) and (20) with numerical simulations of the incorporation process from Eq. (4). Our theory yields reliable predictions for a broad range of parameters, Figs. 3(c) and 3(d).
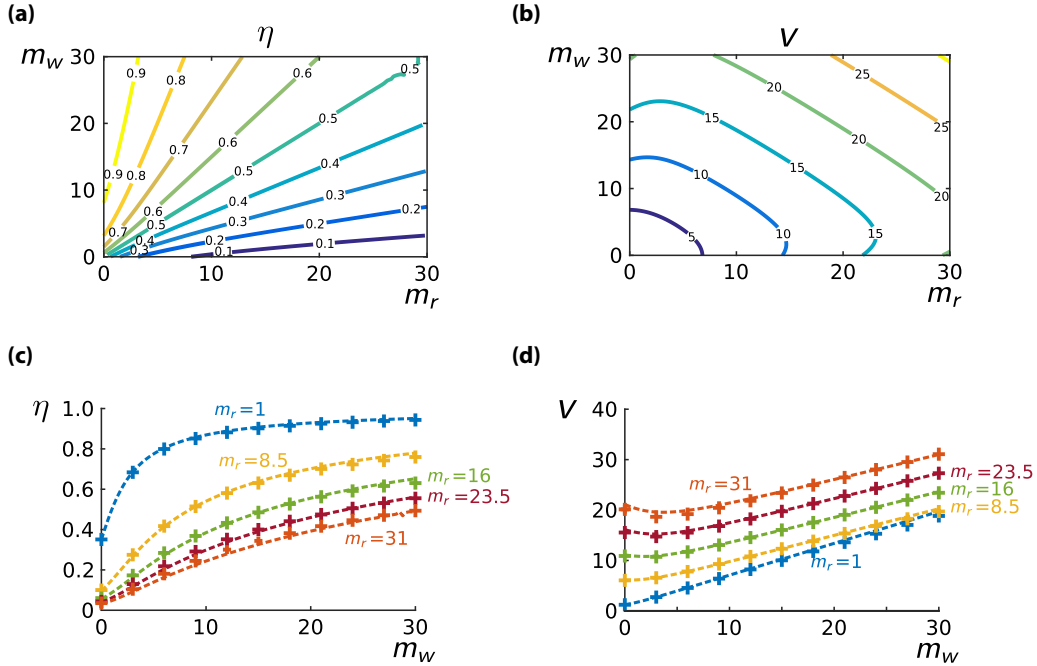
FIG. 3. Error rate η and velocity v for linear free energy landscapes. (Top) Surface plots of η and v from Eqs. (15) and (20) when $G^r(y) = -m_r y$ and $G^w(y) = -m_w y$ as a function of the irreversibility parameters m_r and m_w. (Bottom) Error η and speed v as a function of m_w for different values of m_r. Crosses represent the average η and v values measured from 700 numerical simulations of the stochastic incorporation process. The Langevin dynamics of Eq. (4) was integrated with the Euler-Maruyama scheme [35].

B. Potential with an activation barrier

As a second example, we consider the potential,

$$
\begin{aligned}
\Delta G^x(y) &= a_x \left\{ \exp\left[-\frac{(y - \frac{1}{2})^2}{2c_x^2} \right] - \exp\left(-\frac{1}{8c_x^2} \right) \right\} \\
&+ \frac{b_x}{2} \left(\frac{2c_x + \frac{1}{2} - y}{\sqrt{(2c_x + \frac{1}{2} - y)^2 + c_x^2}} - \frac{2c_x + \frac{1}{2}}{\sqrt{(2c_x + \frac{1}{2})^2 + c_x^2}} \right),
\end{aligned}
\tag{24}
$$

where a_x, b_x, and c_x are monomer-dependent parameters that control the shape of the free energy potentials. Key features of the potential of Eq. (24) are the binding energy $-\Delta G^x(1)$, the height of the activation barrier δ^x, and its width $\sigma^x = 4c_x$, Fig. 4.

We study this model for different cases, corresponding to different parameter choices. In the first case, we fix $-\Delta G^r(1) = -\Delta G^w(1)$ upon choosing $b_r = b_w = b$ and $c_r = c_w = 1/20$. This enforces a kinetic discrimination regime [3] where the binding energy $-\Delta G^r(1)$ quantifies the degree of irreversibility. For highly irreversible processes, the error η should mainly depend on the activation energy difference $\delta^r - \delta^r$, see Eq. (17). We also expect that the reaction speed should increase for more irreversible processes. Equations (15) and (20) confirm such a qualitative picture, see Figs. 5(a) and 5(b). Also in this case, numerical simulations are in excellent quantitative agreement with our theory, Figs. 5(c) and 5(d).

As a second case, we fix $a_r = a_w = 5$ and $b_r = b_w = 1$. In this way, we have that $-\Delta G^r(1) \approx -\Delta G^w(1)$ and $\delta^r \approx \delta^w$.

Energetics alone would not permit monomer discrimination in this case [3]. However, Eq. (17) predicts that the difference in the barrier widths σ_r and σ_w should allow to discriminate r and w monomers [see Fig. 6(a)]. We confirmed the existence of such a kinetic discrimination regime with numerical simulations, Fig. 6(c).

V. KINETIC PROOFREADING

In this section, we sketch a generalization of our framework to include kinetic proofreading [5,6]. We assume that the
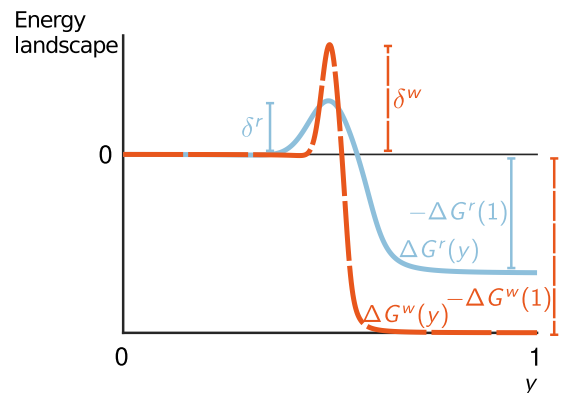

FIG. 4. Free energy potentials with a barrier for r and w monomers from Eq. (24) for r and w monomers. We chose the parameters so that the free energy landscapes for r and w monomers have different binding energies [$-\Delta G^r(1)$ and $-\Delta G^w(1)$], different barrier heights (δ^r and δ^w), and different barrier widths (σ^r and σ^w).
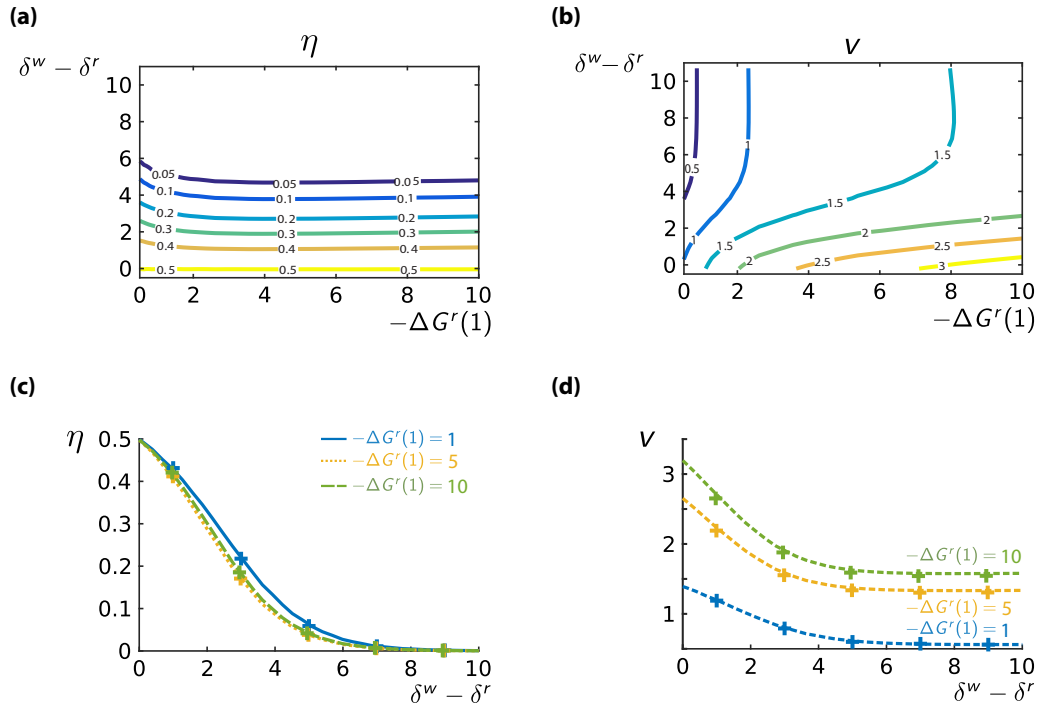
FIG. 5. Equations (15) and (20) predict η and v in a kinetic discrimination regime. (Top) Contour plots of η and v from Eqs. (15) and (20) as a function of the activation energy difference $\delta^w - \delta^r$ and the binding energy $-\Delta G^r(1)$. In this example, we chose $a_r = 1$, $b_r = b_w$, and $c_r = c_w = 1/20$ to ensure a kinetic discrimination regime where $\Delta G^r(1) = \Delta G^w(1)$. Large values of $-\Delta G^r(1)$ correspond to highly irreversible processes. (Bottom) Plots of η and v versus the activation energy difference $\delta^w - \delta^r$ at fixed values of $\Delta G^r(1)$. Crosses corresponds to the average values of η and v measured from 300 stochastic simulations of the incorporation process with Eq. (4). The Langevin dynamics was simulated with a weak 2.0 Runge-Kutta stochastic scheme [35].
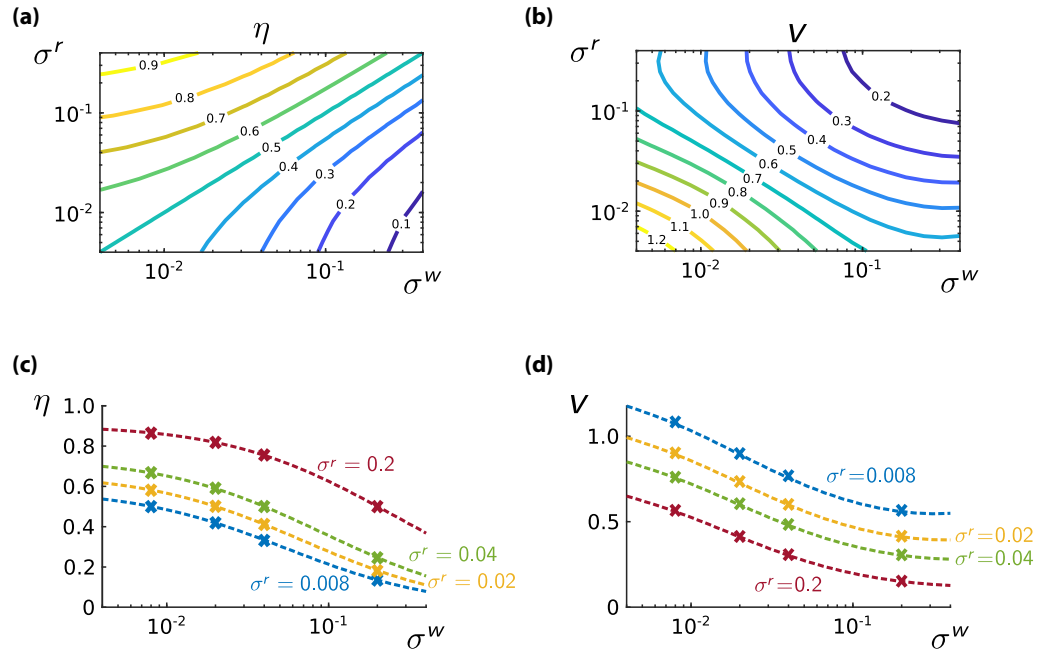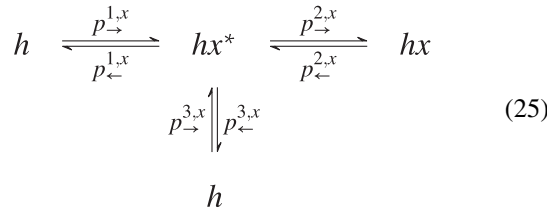

FIG. 6. Different barrier widths allow for kinetic discrimination in the absence of binding and activation energy differences. (Top) Surface plots of η and v from Eqs. (15) and (20) as a function of the barrier widths σ_r and σ_w. To ensure that $G^r(y)$ and $G^w(y)$ have approximately the same binding and activation energies we fixed $a_r = a_w = 5$, $b_r = b_w = 1$, and $c_r, c_w \leqslant 0.05$. (Bottom) Plots of η and v versus σ_w for selected values of σ^r. Crosses corresponds to the average η and v values measured from 400 simulations of the incorporation process with Eq. (4). The Langevin dynamics was simulated with the weak 2.0 Runge-Kutta stochastic scheme [35].

reaction $h \rightleftharpoons hx$ can be decomposed into three subreactions,

$$
h \quad \underset{p_{\leftarrow}^{1,x}}{\overset{p_{\rightarrow}^{1,x}}{\rightleftharpoons}} \quad hx^* \quad \underset{p_{\leftarrow}^{2,x}}{\overset{p_{\rightarrow}^{2,x}}{\rightleftharpoons}} \quad hx
$$

$$
p_{\rightarrow}^{3,x} \, \Big\| \, p_{\leftarrow}^{3,x} \tag{25}
$$

$$
h
$$

where each subreaction occurs with probabilities $p_{\rightarrow}^{i,x}$'s and $p_{\leftarrow}^{i,x}$'s, and hx^* is an intermediate metastable complex. The extra pathway $hx^* \rightleftharpoons h$ represents kinetic proofreading. Such a reaction can improve accuracy when driven towards the reactants h so that wrong monomers undergo an additional checkpoint [3,5].

Every subreaction in Eq. (25) is described by its own reaction coordinate y which evolves according to a Langevin equation,

$$
\frac{dy}{dt} = -\mu \frac{d}{dy} G^{i,x}(y) + \sqrt{2D}\xi(t), \tag{26}
$$

where $G^{i,hx}(y)$ is the free energy landscapes along the ith subreaction. Also, in this case, we take $y \in [0, 1]$ for all subreactions with $y = 1$ always in the direction of incorporation of monomer x. Similar to Eq. (3), we decompose the free energies for the subreactions as

$$
G^{1,h\tilde{x}x}(y) = \Delta G^{1,x}(y) + G^{2,h\tilde{x}}(1), \tag{27a}
$$

$$
G^{2,h\tilde{x}x}(y) = \Delta G^{2,x}(y) + G^{1,h\tilde{x}x}(1), \tag{27b}
$$

$$
G^{3,h\tilde{x}x}(y) = \begin{cases} \Delta G^{3,x}(y) + G^{2,h\tilde{x}}(1) & \text{for } hx^* \to h, \\ \Delta G^{3,x}(y) + G^{1,h\tilde{x}x}(1) & \text{for } hx^* \leftarrow h, \end{cases} \tag{27c}
$$

where we specified that monomer $\tilde{x}$ was incorporated before attempting to incorporate monomer x. Here, $G^{3,h\tilde{x}x}(y)$ depend on the direction of the subreaction because the heteropolymer total energy now depends also on the sequence of subreactions.

We now compute the probabilities $p_{\rightarrow}^{i,x}$'s and $p_{\leftarrow}^{i,x}$ with $i \in \{1, 2, 3\}$ from Eq. (26) with the same procedure which leads to Eq. (9). This yields

$$
p_{\rightarrow}^{i,x} = \epsilon \pi_{\epsilon,0}^{i,x} \frac{1}{\int_0^1 \exp\left[\frac{\Delta G^{i,x}(y)}{k_B T}\right] dy} + O(\epsilon), \tag{28a}
$$

$$
p_{\leftarrow}^{i,x} = \epsilon \pi_{1-\epsilon,1}^{i,x} \frac{\exp\left[\frac{\Delta G^{i,x}(1)}{k_B T}\right]}{\int_0^1 \exp\left[\frac{\Delta G^{i,x}(y)}{k_B T}\right] dy} + O(\epsilon), \tag{28b}
$$

with

$$
\pi_{\epsilon,0}^{1,x} = \pi_{1-\epsilon,1}^{2,x} = \pi_{\epsilon,0}^{3,x} = \tfrac{1}{5}, \tag{29a}
$$

$$
\pi_{1-\epsilon,1}^{1,x} = \pi_{\epsilon,0}^{2,x} = \pi_{\epsilon,0}^{3,x} = \tfrac{1}{3}. \tag{29b}
$$

Equations (29) state that the all subreactions from reactants h and hx^*, respectively, can start with equal probabilities.

To obtain an equation for η, we need to compute the effective incorporation and removal probabilities $p_{\rightarrow}^x$ and $p_{\leftarrow}^x$ in Eq. (14) from Eqs. (28) and (29). To this end, we assume

that the reactions in Eq. (25) are at steady state. We then use adiabatic elimination [33] to obtain (see Appendix D)

$$
p_{\rightarrow}^x = \frac{p_{\rightarrow}^{2,x}(p_{\rightarrow}^{1,x} + p_{\rightarrow}^{3,x})}{p_{\leftarrow}^{1,x} + p_{\leftarrow}^{3,x} + p_{\rightarrow}^{2,x}}, \tag{30a}
$$

$$
p_{\leftarrow}^x = \frac{p_{\leftarrow}^{2,x}(p_{\leftarrow}^{1,x} + p_{\leftarrow}^{3,x})}{p_{\leftarrow}^{1,x} + p_{\leftarrow}^{3,x} + p_{\rightarrow}^{2,x}}. \tag{30b}
$$

Substituting Eqs. (28) and (30) in Eq. (14) finally provides an expression for η in terms of the free energy landscapes $G^{i,x}(y)$.

VI. CONCLUSIONS

In this paper, we described the assembly of heteropolymers by means of continuous reaction coordinates. In the simplest cases, our results are consistent with those derived for reactions occurring in discrete steps [2,3,10,12–14,16,17]. Moreover, our formalism reveals discrimination mechanisms that are not easily described with discrete reactions. One example is the possibility to discriminate according to barrier widths as described by Eq. (17) and confirmed in simulations, Fig. 6(c).

For simplicity, in this paper, we developed our formalism by means of a reaction coordinate characterized by a Markovian dynamic. In general, only specific projection techniques yield reaction coordinates with negligible non-Markovian contributions [23,28,29,36], and the resulting Langevin equation might not be in the form of Eq. (4). Our framework can be adapted to such situations as well as to non-Markovian reaction coordinates, describing, for example, enzymes undergoing slow conformational changes.

The framework described here is microscopically reversible. This allows to characterize nonequilibrium work and heat exchanges during the polymerization process from the diffusive dynamics of the reaction coordinate, similar to recent studies of the ATP synthase [37,38] and small-scale technological devices [39,40]. This analysis would permit characterizing thermodynamic limits of information processing of these processes [14,41–43].

APPENDIX A: FIRST-PASSAGE TIME PROBABILITIES AT THE NODES

Because of detailed balance, the probabilities $\pi_{\epsilon,0}^x$ and $\pi_{1-\epsilon,1}^x$ are related to the free energy difference when passing from one edge of the reaction network to another, i.e.,

$$
\pi_{\epsilon,0}^x \propto \exp\left[-\frac{G^{h\tilde{x}x}(\epsilon)}{k_B T}\right], \quad \pi_{1-\epsilon,1}^{\tilde{x}} \propto \exp\left[-\frac{G^{h\tilde{x}}(1-\epsilon)}{k_B T}\right], \tag{A1}
$$

where we specified that the monomer $\tilde{x} \in \{r, w\}$ was incorporated before monomer $x \in \{r, w\}$. After the incorporation of $\tilde{x}$, the enzyme can catalyze three reactions: Removal of $\tilde{x}$ or incorporation of either r or w. The probabilities of these three

events must be normalized

$$\pi_{1-\epsilon,1}^{\tilde{x}} + \pi_{\epsilon,0}^r + \pi_{\epsilon,0}^w = 1. \tag{A2}$$

Combining Eqs. (A1) and (A2) gives

$$\pi_{\epsilon,0}^x = \frac{\exp\left[-\frac{G^{h\tilde{x}x}(\epsilon) - G^{h\tilde{x}}(1-\epsilon)}{k_B T}\right]}{1 + \exp\left[-\frac{G^{h\tilde{x}r}(\epsilon) - G^{h\tilde{x}}(1-\epsilon)}{k_B T}\right] + \exp\left[-\frac{G^{h\tilde{x}w}(\epsilon) - G^{h\tilde{x}}(1-\epsilon)}{k_B T}\right]}, \tag{A3a}$$

$$\pi_{1-\epsilon,1}^{\tilde{x}} = \frac{1}{1 + \exp\left[-\frac{G^{h\tilde{x}r}(\epsilon) - G^{h\tilde{x}}(1-epsilon)}{k_B T}\right] + \exp\left[-\frac{G^{h\tilde{x}w}(\epsilon) - G^{h\tilde{x}}(1-\epsilon)}{k_B T}\right]}. \tag{A3b}$$

Substituting Eq. (3) into Eq. (A3), taking the limit of small ϵ, using the continuity of $G^{h\tilde{x}x}(y)$, and then renaming $\tilde{x}$ with x finally gives Eq. (6).

APPENDIX B: EFFECTIVE INCORPORATION AND REMOVAL PROBABILITIES

The dynamics of the reaction coordinate y in the coarse grained description of Eq. (5) obeys a Markov chain,

$$P_0^x(\nu + 1) = \pi_{0,\epsilon}^x P_\epsilon^x(\nu) + \left[1 - \pi_{0,\epsilon}^x\right] P_0^x(\nu) \\ + \text{external fluxes}, \tag{B1a}$$

$$P_\epsilon^x(\nu + 1) = \pi_{\epsilon,0}^x P_0^x(\nu) + \pi_{\epsilon,1-\epsilon}^x P_{1-\epsilon}^x(\nu) \\ + \left[1 - \left(\pi_{0,\epsilon}^x + \pi_{1-\epsilon,\epsilon}^x\right)\right] P_\epsilon^x(\nu), \tag{B1b}$$

$$P_{1-\epsilon}^x(\nu + 1) = \pi_{1-\epsilon,\epsilon}^x P_\epsilon^x(\nu) + \pi_{1-\epsilon,1}^x P_1^x(\nu) \\ + \left[1 - \left(\pi_{\epsilon,1-\epsilon}^x + \pi_{1,1-\epsilon}^x\right)\right] P_{1-\epsilon}^x(\nu), \tag{B1c}$$

$$P_1^x(\nu + 1) = \pi_{1,1-\epsilon}^x P_{1-\epsilon}^x(\nu) + \left[1 - \pi_{1-\epsilon,1}^x\right] P_1^x(\nu) \\ + \text{external fluxes}, \tag{B1d}$$

where the first-passage probabilities appear as transition probabilities and the quantities $P_0^x(\nu)$, $P_\epsilon(\nu)$, $P_{1-\epsilon}(\nu)$, and $P_1(\nu)$ are the probabilities that the reaction coordinate reaches the points $y = 0$, $y = \epsilon$, $y = 1 - \epsilon$, and $y = 1$ after ν consecutive transitions, respectively. The *external fluxes* in Eqs. (B1a) and (B1d) are the probability fluxes from the remaining reactions which originate from the nodes $y = 0$ and $y = 1$ in the network of Fig. 1(a).

To simplify Eq. (B1), we perform adiabatic elimination [33] of the intermediate states $y = \epsilon$ and $y = 1 - \epsilon$: we impose the steady state regime $P_\epsilon^x(\nu + 1) = P_\epsilon^x(\nu)$ and $P_\epsilon^x(\nu + 1) = P_\epsilon^x(\nu)$ in Eqs. (B1b) and (B1c), respectively, we solve Eqs. (B1b) and (B1c) for $P_\epsilon^x(\nu)$ and $P_{1-\epsilon}^x(\nu)$, and we finally substitute the result back into Eqs. (B1a) and (B1d). This yields the effective Markov chain,

$$P_0^x(\nu + 1) = p_\leftarrow^x P_1^x(\nu) + [1 - p_\rightarrow^x] P_0^x(\nu) \\ + \text{external fluxes}, \tag{B2a}$$

$$P_1^x(\nu + 1) = p_\rightarrow^x P_0^x(\nu) + [1 - p_\leftarrow^x] P_1^x(\nu) \\ + \text{external fluxes}, \tag{B2b}$$

where we have defined the effective probabilities $p_\rightarrow^x$ and $p_\leftarrow^x$ to incorporate a monomer ($h \rightarrow hx$) and remove a monomer ($h \leftarrow hx$), respectively, as

$$p_\rightarrow^x = \frac{\pi_{1,1-\epsilon}^x \pi_{1-\epsilon,\epsilon}^x \pi_{\epsilon,0}^x}{\pi_{1,1-\epsilon}^x \pi_{1-\epsilon,\epsilon}^x + \pi_{0,\epsilon}^x \pi_{\epsilon,1-\epsilon}^x + \pi_{0,\epsilon}^x \pi_{1,1-\epsilon}^x}, \tag{B3a}$$

$$p_\leftarrow^x = \frac{\pi_{0,\epsilon}^x \pi_{\epsilon,1-\epsilon}^x \pi_{1-\epsilon,1}^x}{\pi_{1,1-\epsilon}^x \pi_{1-\epsilon,\epsilon}^x + \pi_{0,\epsilon}^x \pi_{\epsilon,1-\epsilon}^x + \pi_{0,\epsilon}^x \pi_{1,1-\epsilon}^x}. \tag{B3b}$$

Substituting Eqs. (6)–(8) into Eq. (B3) and then expanding for small ϵ finally gives Eq. (9).

APPENDIX C: DERIVATION OF THE POLYMERIZATION SPEED VIA REACTION COORDINATES

To derive the polymerization speed, we consider a mean field formulation of the polymerization process in Fig. 1(a) where the enzyme can remove any monomer in the copy heteropolymer. Removal of r and w monomers occurs with probabilities $1 - \eta$ and η, respectively. This assumption simplifies the reaction tree of Fig. 1(a) into the closed network of Fig. 7(a) where the incorporation and removal probabilities $p_\rightarrow^x$ and $p_\leftarrow^x$ are defined as in Eq. (9).

We now introduce the reaction coordinate in this mean field description, Fig. 7(b). For later convenience, we also consider the values of the reaction coordinate $y = 0$, $y = \epsilon$, $y = 1 - \epsilon$, and $y = 1$ together with the probabilities $\pi_{\tilde{y},y}^x$'s defined in Eqs. (6) and (7).

Using the scheme in Fig. 7(b), we define the probability $P_{0,1}(\zeta)$ that $y = 0$ or $y = 1$ after ζ consecutive transitions and the probabilities $P_\epsilon^r(\zeta)$, $P_{1-\epsilon}^r(\zeta)$, $P_\epsilon^w(\zeta)$, $P_{1-\epsilon}^w(\zeta)$ that $y = \epsilon$ or $y = 1 - \epsilon$ for the r and w monomers after ζ consecutive transitions. These probabilities evolve according to the Markov chain,

$$\vec{P}(\zeta + 1) = A\vec{P}(\zeta), \tag{C1}$$

where

$$\vec{P}(\zeta) = \left[P_{0,1}(\zeta), P_\epsilon^r(\zeta), P_{1-\epsilon}^r(\zeta), P_\epsilon^w(\zeta), P_{1-\epsilon}^w(\zeta)\right]^T, \tag{C2}$$

and

$$A = \begin{bmatrix} 0 & \pi_{0,\epsilon}^r & \pi_{1,1-\epsilon}^r & \pi_{0,\epsilon}^w & \pi_{1,1-\epsilon}^w \\ c'\pi_{\epsilon,0}^r & 0 & \pi_{\epsilon,1-\epsilon}^r & 0 & 0 \\ c'(1-\eta)\pi_{1-\epsilon,1}^r & \pi_{1-\epsilon,\epsilon}^r & 0 & 0 & 0 \\ c'\pi_{\epsilon,0}^w & 0 & 0 & 0 & \pi_{\epsilon,1-\epsilon}^w \\ c'\eta\pi_{1-\epsilon,1}^w & 0 & 0 & \pi_{1-\epsilon,\epsilon}^w & 0 \end{bmatrix}, \tag{C3}$$
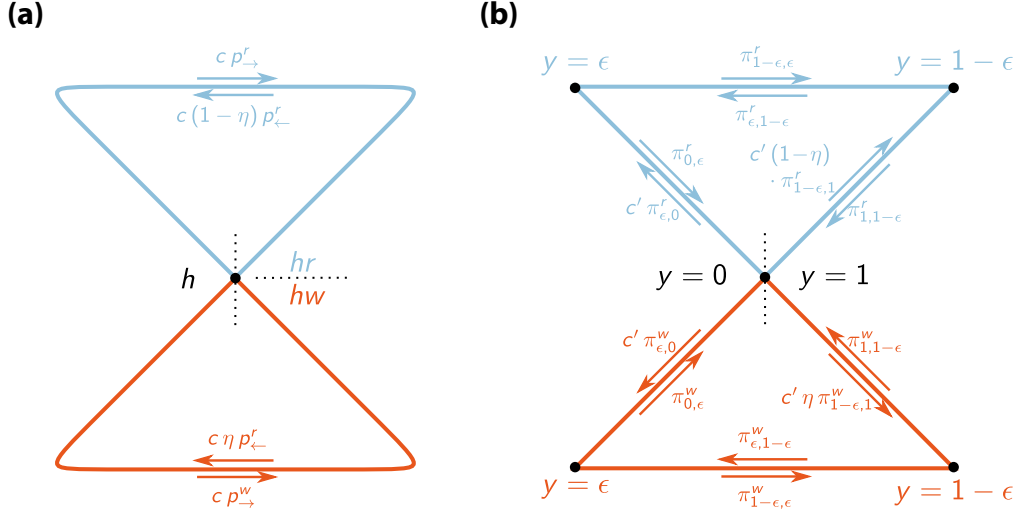
FIG. 7. Mean field representation of the polymerization process. (a) Mean field version of the heteropolymer assembly in Fig. 1(a) where the enzyme can remove any monomer in h. Removal involves r and w monomers $(1-\eta)$ and η times, respectively. The probabilities $p^r_\rightarrow$, $p^r_\leftarrow$ are defined as in Eq. (9). The constant $c = [p^r_\rightarrow + (1-\eta)p^r_\leftarrow + p^w_\rightarrow + \eta p^w_\leftarrow]^{-1}$ normalizes the probabilities and is defined as in Eq. (13). (b) The same as in panel (a), but we have now explicitly introduced the intermediate reaction coordinate values $y = \epsilon$ and $y = 1-\epsilon$, as well as the transition probabilities $\pi^x_{y,\tilde{y}}$ defined as in Eqs. (6) and (7). The constant $c' = [\pi^r_{\epsilon,0} + (1-\eta)\pi^r_{1-\epsilon,1} + \pi^w_{\epsilon,0} + \eta\pi^w_{1-\epsilon,1}]^{-1}$ normalizes the probabilities exiting from the central node.

where $c' = [\pi^r_{\epsilon,0} + (1-\eta)\pi^r_{1-\epsilon,1} + \pi^w_{\epsilon,0} + \eta\pi^w_{1-\epsilon,1}]^{-1}$ is a normalization constant. We now define the matrices,

$$J^N = \frac{1}{3}\begin{bmatrix} 0 & -1 & +1 & -1 & +1 \\ +1 & 0 & -1 & 0 & 0 \\ -1 & +1 & 0 & 0 & 0 \\ +1 & 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & +1 & 0 \end{bmatrix}, \tag{C4a}$$

$$J^\mathcal{T} = \begin{bmatrix} 0 & \langle d\tau \rangle^r_{0,\epsilon} & \langle d\tau \rangle^r_{1,1-\epsilon} & \langle d\tau \rangle^w_{0,\epsilon} & \langle d\tau \rangle^w_{1,1-\epsilon} \\ \langle d\tau \rangle^r_{\epsilon,0} & 0 & \langle d\tau \rangle^r_{\epsilon,1-\epsilon} & 0 & 0 \\ \langle d\tau \rangle^r_{1-\epsilon,1} & \langle d\tau \rangle^r_{1-\epsilon,\epsilon} & 0 & 0 & 0 \\ \langle d\tau \rangle^w_{\epsilon,0} & 0 & 0 & 0 & \langle d\tau \rangle^w_{\epsilon,1-\epsilon} \\ \langle d\tau \rangle^w_{1-\epsilon,1} & 0 & 0 & \langle d\tau \rangle_{1-\epsilon,\epsilon} & 0 \end{bmatrix}, \tag{C4b}$$

which contain the contribution of each transition to the heteropolymer length N and the polymerization time $\mathcal{T}$. The time increments $\langle d\tau \rangle^x_{y,\tilde{y}}$ in $J^\mathcal{T}$ are the first-passage times from $\tilde{y}$ to y [31]. In particular, we have that

$$\langle \tau \rangle^x_{0,\epsilon} = \frac{1}{D}\left(\Phi^x_\rightarrow(0) - \frac{\int_0^1 \Phi^x_\rightarrow(y)dy}{\int_0^1 \psi^x(y)dy} \right)\epsilon + O(\epsilon^2), \tag{C5a}$$

$$\langle \tau \rangle_{1-\epsilon,\epsilon} = \frac{1}{D}\int_0^1 \Phi^x_\leftarrow(y)dy + O(\epsilon), \tag{C5b}$$

$$\langle \tau \rangle_{\epsilon,1-\epsilon} = \frac{1}{D}\int_0^1 \Phi^x_\rightarrow(y)dy + O(\epsilon), \tag{C5c}$$

$$\langle \tau \rangle^x_{1,1-\epsilon} = \frac{1}{D}\left[\Phi^x_\leftarrow(1) - \left[\exp\left(\frac{\Delta G^x(1)}{k_B T} \right) \right]\frac{\int_0^1 \Phi^x_\leftarrow(y)dy}{\int_0^1 \psi^x(y)dy} \right]\epsilon$$

$$+ O(\epsilon^2), \tag{C5d}$$

with

$$\Phi^x_\rightarrow(y) = \frac{\psi^x(y)\int_y^1 \int_u^1 \frac{\psi^x(z)}{\psi^x(u)}dz\,du}{\int_0^1 \psi^x(y)dy}, \tag{C6a}$$

$$\Phi^x_\leftarrow(y) = \frac{\psi^x(y)\int_0^y \int_0^u \frac{\psi^x(z)}{\psi^x(u)}dz\,du}{\int_0^1 \psi^x(y)dy}. \tag{C6b}$$

The remaining first-passage times $\langle d\tau \rangle^r_{\epsilon,0}$, $\langle d\tau \rangle^r_{1-\epsilon,1}$, $\langle d\tau \rangle^w_{\epsilon,0}$, and $\langle d\tau \rangle^w_{1-\epsilon,1}$ are assumed equal to zero for simplicity. Physically, this assumption is justified when binding and unbinding of monomers is much faster than processing a monomer into a finalized incorporation.

Using Eq. (C4), we define the tilted matrix B with components,

$$B_{i,j} = A_{i,j}\exp\left[q_N J^N_{i,j} + q_\tau J^\tau_{i,j} \right], \tag{C7}$$

and dummy variables q_N and q_τ. For large values of ζ, the largest eigenvalue of B coincides with the scaled cumulant

generating function of N and $\mathcal{T}$, see Ref. [44]. The implicit function theorem then implies

$$N \approx -\zeta \frac{\partial_{q_N} \det[B - \lambda I]}{\partial_\lambda \det[B - \lambda I]}\bigg|_{q_N = q_\tau = 0,\, \lambda = 1}, \quad \text{(C8a)}$$

$$\mathcal{T} \approx -\zeta \frac{\partial_{q_\tau} \det[b - \lambda I]}{\partial_\lambda \det[B - \lambda I]}\bigg|_{q_N = q_\tau = 0,\, \lambda = 1}, \quad \text{(C8b)}$$

where $\det[B - \lambda I]$ is the characteristic polynomial of B. To compute v, we finally use that

$$v = \frac{N}{\mathcal{T}} = \frac{\partial_{q_N} \det[B - \lambda I]}{\partial_{q_\tau} \det[B - \lambda I]}\bigg|_{q_N = q_\tau = 0,\, \lambda = 1}, \quad \text{(C9)}$$

which is equivalent to Eq. (18). Substituting Eqs. (6), (7), (C3), (C4), and (C7) into Eq. (C9) and then taking the leading order for small ϵ, yields Eq. (20), where

$$\frac{\langle \tau \rangle D}{c} = [p_\to^r - (1 - \eta) p_\leftarrow^r] \left(\int_0^1 [\Phi_\leftarrow^r(y) - \Phi_\to^r(y)] dy \right)$$

$$+ (p_\to^w - \eta p_\leftarrow^w) \left(\int_0^1 [\Phi_\leftarrow^w(y) - \Phi_\to^w(y)] dy \right)$$

$$+ \frac{\epsilon}{3} [\Phi_\to^r(0) + \Phi_\to^w(0) + (1 - \eta)\Phi_\leftarrow^r(1) + \eta \Phi_\leftarrow^w(1)]$$

$$+ O(\epsilon^2), \quad \text{(C10)}$$

and c is defined as in Eq. (13).

APPENDIX D: EFFECTIVE INCORPORATION AND REMOVAL PROBABILITIES FOR THE KINETIC PROOFREADING EXAMPLE

To compute the incorporation and removal probabilities for the kinetic proofreading case, we mimic the procedure that leads to Eq. (9). We consider the probabilities $P_h(\rho)$, $P_{hx^*}(\rho)$, and $P_{hx}(\rho)$ to obtain the reactants h, hx^*, and hx after ρ subreactions of Eq. (25). These probabilities evolve according to the Markov chain,

$$P_h(\rho + 1) = p_\leftarrow^{1,x} P_{hx^*}(\rho) + [1 - (p_\to^{1,x} + p_\to^{3,x})]P_h(\rho)$$

$$+ \text{external fluxes}, \quad \text{(D1a)}$$

$$P_{hx^*}(\rho + 1) = p_\to^{1,x} P_h(\rho) + p_\leftarrow^{2,x} P_{hx}(\rho) + p_\to^{3,x} P_h(\rho)$$

$$+ [1 - (p_\leftarrow^{1,x} + p_\to^{2,x} + p_\leftarrow^{3,x})]P_{hx^*}(\rho), \quad \text{(D1b)}$$

$$P_{hx}(\rho + 1) = p_\to^{2,x} P_{hx^*}(\rho) + [1 - p_\leftarrow^{2,x}]P_{hx}(\rho)$$

$$+ \text{external fluxes}, \quad \text{(D1c)}$$

where the *external fluxes* are the probability fluxes of the other subreactions entering the nodes $y = 0$ and $y = 1$. At steady state, we simplify Eq. (D1) with adiabatic elimination [33]: We impose $P_{hx^*}(\rho + 1) = P_{hx^*}(\rho)$ into Eq. (D1b), solve it for $P_{hx^*}(\rho)$, and substitute the solution in Eqs. (D1a) and (D1c). This yields, after some rearrangements,

$$P_{h+x}(\rho + 1) = p_\leftarrow^x P_{hx}(\rho) + [1 - p_\to^x]P_{h+x}(\rho), \quad \text{(D2a)}$$

$$P_{hx}(\rho + 1) = p_\to^x P_{h+x}(\rho) + [1 - p_\leftarrow^x]P_{hx}(\rho), \quad \text{(D2b)}$$

with effective incorporation and removal probabilities $p_\to^x$ and $p_\leftarrow^x$ defined as in Eq. (30).

[1] N. M. Reynolds, B. A. Lazazzera, and M. Ibba, Nat. Rev. Microbiol. **8**, 849 (2010).

[2] C. H. Bennett, BioSystems **11**, 85 (1979).

[3] S. Pigolotti and P. Sartori, J. Stat. Phys. **162**, 1167 (2016).

[4] T. A. Kunkel and K. Bebenek, Annu. Rev. Biochem. **69**, 497 (2000).

[5] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **71**, 4135 (1974).

[6] J. Ninio, Biochimie **57**, 587 (1975).

[7] M. O'Donnell, L. Langston, and B. Stillman, Cold Spring Harbor Perspect. Biol. **5**, a010108 (2013).

[8] T. A. Kunkel and D. A. Erie, Annu. Rev. Genet. **49**, 291 (2015).

[9] M. V. Rodnina and W. Wintermeyer, Annu. Rev. Biochem. **70**, 415 (2001).

[10] D. Andrieux and P. Gaspard, Proc. Natl. Acad. Sci. USA **105**, 9516 (2008).

[11] M. Johansson, M. Lovmar, and M. Ehrenberg, Curr. Opin. Microbiol. **11**, 141 (2008).

[12] P. Gaspard and D. Andrieux, J. Chem. Phys. **141**, 044908 (2014).

[13] K. Banerjee, A. B. Kolomeisky, and O. A. Igoshin, Proc. Natl. Acad. Sci. USA **114**, 5183 (2017).

[14] P. Sartori and S. Pigolotti, Phys. Rev. X **5**, 041039 (2015).

[15] Y. Savir and T. Tlusty, Cell **153**, 471 (2013).

[16] P. Sartori and S. Pigolotti, Phys. Rev. Lett. **110**, 188101 (2013).

[17] A. Murugan, D. A. Huse, and S. Leibler, Proc. Natl. Acad. Sci. USA **109**, 12034 (2012).

[18] F. Cady and H. Qian, Phys. Biol. **6**, 036011 (2009).

[19] H. Kramers, Physica **7**, 284 (1940).

[20] R. Rao and L. Peliti, J. Stat. Mech.: Theory Exp. (2015) P06001.

[21] J. A. Wagoner and K. A. Dill, Proc. Natl. Acad. Sci. USA **116**, 5902 (2019).

[22] L. V. Bock, M. H. Kolář, and H. Grubmüller, Curr. Opin. Struct. Biol. **49**, 27 (2018).

[23] P. V. Banushkina and S. V. Krivov, WIREs Comput, Mol. Sci. **6**, 748 (2016).

[24] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, J. Chem. Phys. **104**, 5860 (1996).

[25] R. B. Best and G. Hummer, Phys. Rev. Lett. **96**, 228104 (2006).

[26] R. Zwanzig, Phys. Rev. **124**, 983 (1961).

[27] A. M. Berezhkovskii and A. Szabo, J. Phys. Chem. B **117**, 13115 (2013).

[28] J. Lu and E. Vanden-Eijnden, J. Chem. Phys. **141**, 044109 (2014).

[29] S. V. Krivov and M. Karplus, J. Phys. Chem. B **110**, 12689 (2006).

[30] D. K. Klimov and D. Thirumalai, Phys. Rev. Lett. **79**, 317 (1997).

[31] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, Springer Series in Synergetics (Springer, Berlin, Heidelberg, 2009).

[32] S. Iyer-Biswas and A. Zilman, *Advances in Chemical Physics* (Wiley, Hoboken, NJ, 2016), Vol. 160, pp. 261–306.

[33] S. Pigolotti and A. Vulpiani, J. Chem. Phys. **128**, 154114 (2008).

[34] C. Bender and S. Orszag, *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*, Advanced Mathematical Methods for Scientists and Engineers (Springer, New York, 1978).

[35] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability (Springer, Berlin, Heidelberg, 2011).

[36] S. V. Krivov, J. Chem. Theory Comput. **14**, 3418 (2018).

[37] J. N. E. Lucero, A. Mehdizadeh, and D. A. Sivak, Phys. Rev. E **99**, 012119 (2019).

[38] A. K. S. Kasper and D. A. Sivak, arXiv:1905.10640.

[39] I. Neri, M. Lopez-Suarez, D. Chiuchiú, and L. Gammaitoni, Europhys. Lett. **111**, 10004 (2015).

[40] M. López-Suárez, I. Neri, and L. Gammaitoni, Nat. Commun. **7**, 12068 (2016).

[41] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, Nature (London) **483**, 187 (2012).

[42] Y. Jun, M. Gavrilov, and J. Bechhoefer, Phys. Rev. Lett. **113**, 190601 (2014).

[43] D. Chiuchiú, Europhys. Lett. **109**, 30002 (2015).

[44] H. Touchette, Phys. Rep. **478**, 1 (2009).