Minimal mechanism for cyclic templating of length-controlled copolymers under isothermal conditions

Jordan Juritz, Jenny M. Poulton and Thomas E. Ouldridge

View Online   Export Citation   CrossMark

ARTICLES YOU MAY BE INTERESTED IN

The Journal of Chemical Physics
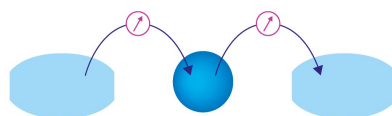
AIP Publishing

Minimal mechanism for cyclic templating of length-controlled copolymers under isothermal conditions

View Online   Export Citation   CrossMark

Jordan Juritz,[1] [iD] Jenny M. Poulton,[2] [iD] and Thomas E. Ouldridge[1,a)] [iD]

AFFILIATIONS

[1] Department of Bioengineering and Centre for Synthetic Biology, Imperial College London, London SW7 2AZ, United Kingdom
[2] Foundation for Fundamental Research on Matter (FOM), Institute for Atomic and Molecular Physics (AMOLF),
1098 XE Amsterdam, The Netherlands

[a)] Author to whom correspondence should be addressed: t.ouldridge@imperial.ac.uk

ABSTRACT

The production of sequence-specific copolymers using copolymer templates is fundamental to the synthesis of complex biological molecules and is a promising framework for the synthesis of synthetic chemical complexes. Unlike the superficially similar process of self-assembly, however, the development of synthetic systems that implement templated copying of copolymers under constant environmental conditions has been challenging. The main difficulty has been overcoming product inhibition or the tendency of products to adhere strongly to their templates—an effect that gets exponentially stronger with the template length. We develop coarse-grained models of copolymerization on a finite-length template and analyze them through stochastic simulation. We use these models first to demonstrate that product inhibition prevents reliable template copying and then ask how this problem can be overcome to achieve cyclic production of polymer copies of the right length and sequence in an autonomous and chemically driven context. We find that a simple addition to the model is sufficient to generate far longer polymer products that initially form on, and then separate from, the template. In this approach, some of the free energy of polymerization is diverted into disrupting copy–template bonds behind the leading edge of the growing copy copolymer. By additionally weakening the final copy–template bond at the end of the template, the model predicts that reliable copying with a high yield of full-length, sequence-matched products is possible over large ranges of parameter space, opening the way to the engineering of synthetic copying systems that operate autonomously.

I. INTRODUCTION

Copolymers—polymers formed from two or more types of monomer unit—are ubiquitous in biology. DNA, RNA, and proteins are biological copolymers for which the information pertaining to the function and/or the structure of the molecule is encoded into the sequence of the copolymer.[1] There is a huge diversity of copolymer sequences to be found in biology; the human proteome, for instance, is comprised of roughly 25 000 different proteins, each consisting of linear chains assembled from the 20 amino acid residues,[2] with a median length of 375 amino acids.[3]

How are these complex molecules formed? In general, it is impossible to encode and consistently assemble thousands of distinct and essentially arbitrary macromolecules, each with ~375 units, through the self-assembly of just 20 types of building blocks.[4] Instead, these copolymers are produced by copying copolymer templates, such as DNA and mRNA. In templated copying processes, complementary interactions between monomers and the template direct the assembly of the product so that distinct, arbitrary copolymer sequences can be reliably created from a common set of monomers.[5] In extant organisms, templated copying is supported by a large amount of the cellular resources and is aided by ensembles of complex molecular machines, such as DNA polymerases involved in DNA replication,[6,7] RNA polymerases in transcription,[8–10] and ribosomes in translation.[11–16] The physics of templated copolymer copying therefore lies at the

heart of the synthesis of the diverse and complex molecules of biology.[1]

While recent decades have seen the engineering of remarkably complex self-assembling systems made of DNA,[17–21] proteins,[22,23] and combinations of DNA and proteins,[24] advances in the field of synthetic templating have been comparatively slow. This disparity suggests a poor understanding of templating relative to self-assembly and represents a missed opportunity to harness biology's most important mechanism of producing chemical complexity—both to generate synthetic sequence-controlled polymers[25] and in the context of combinatorial molecular discovery.[26,27] In this work, we simulate coarse-grained models to investigate a class of templated copolymerization reactions that may enable enzyme-free copying of templates analogously to the copying processes found in biology.

We look to biological systems to set the requirements and conditions of operation for the non-enzymatic, synthetic template-copying system we wish to engineer. First, biological copying systems are capable of the accurate reproduction of copolymer sequences from arbitrary but specific lengths of the copolymer template whether by replication,[6,7] transcription,[8–10] or translation.[11–16] Second, each template can be reused many times to produce many copies; the copying process net generates complex molecules that persist separately from their template.[5,28–30] This fact is crucial, since templates themselves are necessarily complex molecules.[5,28–30] If each new copy consumes—or remains bound to—a template, then a new template must be generated from scratch for each copy formed. Third, natural copying systems are capable of operating in spatio-temporally constant environmental conditions, without perturbations, exploitable spatial gradients, or other external interventions; they are driven solely by the chemical free energy of the dissolved building blocks and "fuel" molecules, such as ATP.[31] We describe copying that does not rely on such external factors as *autonomous*.

The phenomenon of "product inhibition" presents a major challenge to achieving reliable, repeatable production of specific copolymers by templating under autonomous conditions.[32] As monomers polymerize on the template, the product's length increases and the binding free energy between the product and template typically grows linearly. Therefore, the likelihood that a product spontaneously detaches from the template is exponentially suppressed with length, inhibiting subsequent copying of the template. As we explore in Sec. III A, products of varying—but short—lengths result when copying long templates in such a setting. Due to product inhibition, long copies may fail to be released from the template, sequestering the template and slowing the copying rate, while short fragmented copies may be released rapidly. Therefore, product inhibition poses significant challenges for the reliability of both length-control and template reuse.

Michaelis *et al.*[32] emphasized that in non-enzymatic systems restricted to isothermal operation, strategies that reduce the affinity between products and the template are required to achieve high turnovers (defined as the final ratio of products to templates), even for dimeric templates.[33] Enforcing separation by reducing the affinity between the copy and template, without compromising the template's reuse, has also proved challenging. Osuna Gálvez and Bode recently reported a templated reaction in which the

dimerization of the reactants and the disruption of *both* the reactant–template bonds occurred simultaneously, drastically reducing the product-template affinity.[34] However, this came at the cost of scarring the template, which was was unable to promote further reactions.[34]

Instead, time-varying, non-autonomous environmental conditions are commonly used to drive cyclic templating and product release. In polymerase chain reactions, for instance, cycles of heating dissociate the product from the template and enable further rounds of amplification.[35] Various other time-varying strategies have also been used to drive enzyme-free, dimeric,[36,37] and longer template-copying and replicating systems.[38–42]

Others have employed non-chemical energy or exploited spatial gradients to engineer systems that spontaneously separate copies from the template, thereby favoring the cyclic assembly and separation of longer copies.[35,43–47] In Schulman's replicator, shear flow was used to fragment the layers of information-bearing DNA tile-based crystals, revealing more reusable templating surfaces, which enabled exponential self-replication of the crystal.[43] Braun *et al.* exploited the convective currents generated by thermal gradients, inspired by oceanic thermal vents,[48] with templating and polymerization occurring in cool regions and separation in hot regions.[35,44,45] Similar environments with spatially and/or temporally varying conditions are thought to have played an important role in the emergence of life,[49–57] though these conditions are not required for the operation of modern biological copying enzymes.

Recently, a novel DNA strand displacement motif, handhold-mediated strand displacement, was used to autonomously drive dimer formation directed by templates.[5] Here, binding between monomers on a template weakens the connection of one monomer to the template, limiting product inhibition. In principle, the mechanism would allow efficient dimerization without template scarring. However, it is unclear whether the mechanism can scale to overcome product inhibition and reliably produce length- and sequence-controlled copies on longer templates.

From a theoretical perspective, many authors have considered the permanent deposition of a copy on a template,[58–65] though the effects of the subsequent separation of the product from the template have not been a focus until recently. Moreover, while these previous works have considered mechanisms of sequence-control without copy separation, mechanisms for precise length-control have been neglected.

More recently, models of templated copolymerization have been used to study self-replicating molecules under prebiotically plausible conditions.[57,66] Tupper and Higgs argued that a rolling-circle mechanism could have overcome product inhibition to promote non-enzymatic RNA replication in an RNA world.[66] In this high-level model, the chemical details of the directional polymerization mechanism were not considered, and nor was precise length control. In a recent theoretical and experimental work on the pre-enzymatic templated assembly, Rosenberger *et al.* demonstrated that polymer aggregates with increased length can be generated from short building blocks under isothermal conditions, though these polymers were bound up in complexes, not separated from their templates.[57] Moreover, observing the gradual lengthening of copolymers on average was the priority,

not precise length control, as required for perfect information copying.

The thermodynamic constraints that copy–template separation place on sequence-control were considered in Ref. 28, though non-chemical means were invoked to separate the template and copy. In Refs. 29 and 30, the thermodynamic and kinetic consequences of an isothermal mechanism for generating sequence-controlled copolymers with product separation were considered, but the actual mechanistic details were implicit. Nonetheless, a key theoretical result of these works is that separating an accurate copy from its template necessitates producing a state that is extremely far from equilibrium. This thermodynamic argument also applies to length control: in the absence of residual interactions with a template, an ensemble of copolymers of a specific (but arbitrary and template-selected) length is extraordinarily far from the equilibrium of a broad exponential distribution of lengths. It is this fundamental physical principle that is practically manifested as the challenge of product inhibition: How can interactions be tuned to allow templates to act as a catalyst for the production of a specific far-from-equilibrium product state while avoiding a stable equilibrium of copy–template complexes?

In this paper, we argue that decoupling the length of a copy from its affinity for the template, by channeling the free energy released in polymerization to the disruption of copy–template bonds as achieved by handhold-mediated DNA strand displacement (HMSD),[5] would be sufficient to overcome product inhibition even on long templates in an isothermal environment. Furthermore, we hypothesize that additionally weakening the copy–template connection at the final template site would result in the selective release of complete copies rather than shorter fragments. We explore these arguments through simulations of coarse-grained models.

In Sec. II A, we introduce a basic model of isothermal templated copolymerization with separation. In Sec. III A, we demonstrate that cooperative binding to the template prohibits the release of long polymers under a basic growth mechanism, demonstrating that it is necessary to have more complex interactions between the copy and template in order to promote reliable separation in a constant environment. In Sec. III B, we propose and investigate a mechanism that could alleviate the cooperative effect of product inhibition resulting in an increase in the mean polymer length. Extending this mechanism in Sec. III C, we bias the production of complete polymers by selectively weakening the final site on the template. Finally, in Sec. III D, we demonstrate that long and accurate copies of the template can be generated if the correct and incorrect monomers have varying binding rates.

II. MODEL AND METHODS

A. Model

1. State space

As shown in Fig. 1, we consider a copolymer template $T = t_1 t_2 t_3, \ldots, t_{L_T}$, where t_i is an integer indicating the monomer type that can take values $0, 1, \ldots, \alpha_T - 1$. α_T is the number of distinct types of monomeric unit in the template. A single template copolymer is suspended in a large-volume, well-mixed bath of a second distinct type of monomers that interact with the template. The
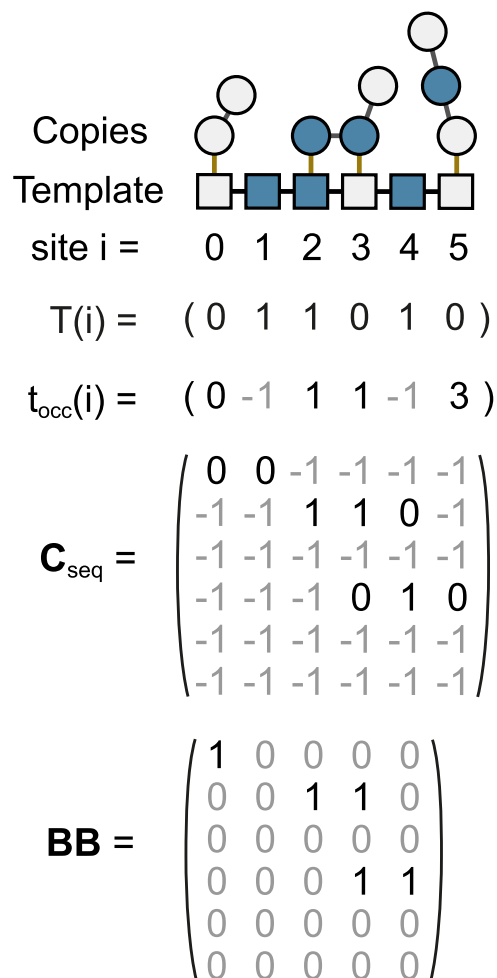

FIG. 1. The state of the system is described by t_{occ}, $\mathbf{C}_{seq}$, and $\mathbf{BB}$. Sites on the template are indexed with i. The array $T(i)$ stores the template sequence. t_{occ} is an array containing the polymer label of each copy unit attached to the template. The label is the row of $\mathbf{C}_{seq}$ in which the polymer's sequence is stored. $\mathbf{BB}$ stores the backbone bonds. In this example, three polymers (labeled 0, 1, and 3) are attached by one, two, and one bond/s to the template, respectively.

identity of these monomers is labeled by an integer that can take values $0, 1, \ldots, \alpha_C - 1$. There are α_C distinct types of copy monomer, the ith of which has a concentration $[\mathrm{M}]_i$. These monomers can bind onto the template at any unoccupied site on the template. In this paper, we consider both homogeneous copy–template systems in which there is a single monomer type, $\alpha_T = \alpha_C = 1$, and also information-bearing binary systems in which there are two monomer types for each of the copy and template monomers $\alpha_T = \alpha_C = 2$. Following Ref. 29, we assume that copy/template interactions are symmetric with respect to interchange of 0 and 1, giving identical dynamics for all template sequences. We shall use the uniform template $T = 000, \ldots, 0$ in all simulations. When the copy–template alphabets are binary, $\alpha_T = \alpha_C = 2$, monomers of types 0 and 1 in the copy copolymer can then simply be interpreted

as "correct" and "incorrect" matches, respectively. We will refer to all molecules as polymers unless their nature as copolymers is important.

A copy unit with no neighboring backbone bonds is a monomer. As shown in Fig. 1, copy polymers are chains of copy units that share a "backbone" linking one unit to the next. Each copy monomer may form up to two backbone bonds, one with a copy unit ahead and one with a copy unit behind with the boundary condition that backbone bonds can only be formed ahead at site 0 and behind at site $L_T - 1$. As can be seen in Fig. 1, copy polymers can occupy contiguous stretches of template sites and may have "tails," portions of their length that are not directly bound to the template but are indirectly tethered to the template by bound units with which they share a backbone. These tails may be present at either end of the polymer. Copy polymer "bridges," unbound stretches between bound units within the same copy polymer, are prohibited. It is assumed that such behavior is strongly suppressed due to the constrained close proximity enforced by the bound units, just as internal bubbles are strongly suppressed between complementary ++ nucleic acid duplexes.[67]

We assume that the concentration of the monomer species $[M_i]$ in the bath is buffered to be constant. In the simulations that follow, we will initiate the system with no polymers in solution and with the template empty (all sites unoccupied). Under these conditions and, again, given that the volume of the bath is large, we further assume that the concentration of copy polymers in the bath remains at a negligible level compared to the concentrations of copy monomers. That is to say, once polymers fully unbind from the template, they diffuse into the bath and do not bind to the template again. We, therefore, neglect possible product inhibition due to very high concentrations of products, a secondary problem relative to product inhibition due to products that simply never detach from the template. In the models presented here, polymers cannot bind to the template from solution and bridges cannot form loops that cause the contour length of a copy polymer to exceed the length of the template, L_T, and therefore, only copy polymers with length $L \leq L_T$ can be produced.

The state of the system is thus uniquely specified by the sequence of monomers in all growing copy polymers, including those with only one unit, and by specifying which site each monomer is attached to on the template (if it is attached to no template site, the monomer exists in a tail) as outlined in Fig. 1. We do not consider the conformations of the system within each of these macrostates.

As we indicate in Fig. 1, sites in the system are indexed with i running from 0 to $L_T - 1$. The ith site on the template can be unoccupied or occupied by a copy unit: either a monomer or part of a polymer. Each isolated monomer or polymer on the template is given a unique label. If two copy units share a backbone, even indirectly via other monomers, they must be part of the same polymer, and therefore, they share the same label. The template can be occupied by at most L_T distinct monomers or polymers at any moment; hence, polymer labels run from 0 to $L_T - 1$. We introduce this label to enable easy accounting when extending, fragmenting, or recombining polymers. The template occupation labels are stored in a vector t_{occ}, where the ith component, $t_{occ}(i)$, is the label l of the copy unit bound to the template at site i or -1 if the site is empty.

The sequence of all polymers—including units in tails—and monomers in the system is stored in an $L_T \times L_T$ matrix, $\mathbf{C}_{seq}$, in which the lth row contains the sequence of the copy unit with label l, buffered with -1 s before and after the sequence. Therefore, $\mathbf{C}_{seq}(t_{occ}(i), i)$ is the identity of the copy unit occupying the template at site i.

Each copy unit may form up to two backbone bonds, one with a copy unit ahead and one with a copy unit behind, with the boundary condition that backbone bonds can only be formed ahead at site 0 and behind at site $L_T - 1$. The $L_T \times L_T - 1$ matrix $\mathbf{BB}$ stores the state of all backbone bonds in the system. The ith column of the lth row of $\mathbf{BB}$ is 1 if there is a backbone bond between copy units with label l at sites i and $i + 1$ (not necessarily connected to the template) and 0 otherwise.

2. Transition rules

The transition rules that define the permitted state changes are listed in Fig. 3 and illustrated in Fig. 2. All transitions in this model are microscopically reversible, apart from the final detachment of copolymers, since the concentration of products in solution, and hence the rate of polymer rebinding, is assumed to be zero. The transitions that feature in this basic model are as follows:

1. A copy monomer in solution may diffuse close enough to an unoccupied site i on the template for an attractive interaction to cause it to bind there.
2. The reverse reaction may also occur in which a monomer bound at site i on the template may unbind and freely diffuse away from the template.
3. A unit of a copy polymer that is either bound at the end of a polymer or is the last bound unit before a tail may unbind from the template. This increases the length of the tail by 1.
4. In reverse, the first unit in the tail of a polymer hovering over site i (which must have a template bound and backbone-linked neighbor at either site $i - 1$ or site $i + 1$) may bind to site i if the site is not already occupied.
5. If the last bond between a polymer and the template breaks, the polymer is released. This liberated polymer mixes with the large volume and never returns to bind to the template.
6. Polymerization can only occur between available template-bound neighbors. Specifically, if a bound copy unit at site i can make a forward backbone bond to site $i + 1$ and a bound copy unit at site $i + 1$ is capable of receiving a backbone bond from site i, then a backbone bond may form.
7. Depolymerization, the breaking of backbone bonds, can only occur between template-bound copy units that share a backbone bond, a consequence of this being that units in copolymer tails cannot spontaneously depolymerize.

We assume that polymerization or depolymerization of molecules not attached to the template in the baths or on the end of free copolymer tails is negligible.

3. Concentrations and transition energies

The concentration of monomer units in the baths are set to be equal, $[M]_i = [M]$, for all monomer types i, defined as a
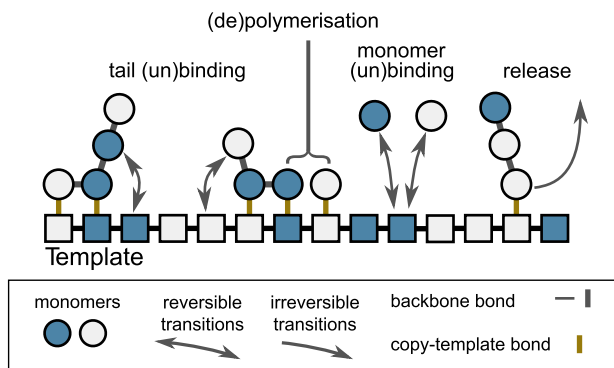
FIG. 2. Simple transition rules between coarse-grained states of a templated polymerization process. Each site on the template can be occupied by one monomeric unit. Monomers may bind to and unbind from the template from solution and from the tails of copolymers. Neighboring units on the template may polymerize or depolymerize. All transitions, aside from the final release of a polymer from the template, are reversible and can occur at any site on the template.

dimensionless quantity relative to an arbitrary reference concentration. We are interested in copying systems in which any template can be copied with the same efficiency as any other template from the same set of building blocks. Templates with an over-abundance of a certain monomer type could be accurately copied more easily if the corresponding copy monomer was over-represented in the pool too.[30] However, a biased monomer pool would, on average, provide no advantage for copying an arbitrary template sequence.

The monomeric units in the tails of the polymers are constrained to a small volume in closer proximity to the template and neighboring monomers than the free monomers in the baths. We, therefore, define an effective concentration $[M_{eff}] \gg [M]$ in order to parameterize the free energy change of

polymerization and the rebinding rate of monomers in the tail to the template.

Bond formation is parameterized through the standard free-energy change of reactions at the reference concentration ($[M] = 1$). In this work, we define all free energies as dimensionless quantities relative to $kT = 1$. The formation of the backbone bond between monomers in isolation is associated with a standard free-energy change of $\Delta G_{BB} + \ln[M_{eff}]$; ΔG_{BB} reflects the chemical bond strength and $\ln[M_{eff}]$ reflects the loss of entropy associated with binding from solution. More negative values of ΔG_{BB} favor polymer growth.

Initially, in Sec. III A, we consider a simple model of cooperative copy–template interactions that is depicted in Fig. 3. In this picture, individual monomers bind to the template with a standard free energy change of ΔG_{spec}—called "specific" because it depends on whether the copy and template monomers are complementary. The standard free-energy change for binding of a copy polymer of length l forming exactly l bonds with the template is

$$\sum_{i=0}^{l-1} \Delta G_{spec}^i - (l-1)\ln([M_{eff}]/[M]), \qquad (1)$$

where ΔG_{spec}^i is the specific bond free energy of the ith copy–template pair. Here, the factor $(l-1)\ln[M_{eff}]$ captures the cooperative nature of the bond between a long template and a long copy that is responsible for product inhibition. This formulation is an extremely simple model of cooperativity: a copolymer of length l gains l favorable bonds of the same strength but for $l-1$ of those bonds, the entropic cost is reduced enormously relative to binding from solution. In Sec. III B, we extend the model to allow for less cooperative behavior, and in Sec. III C, we also allow sites at the end of the template to bond more weakly with the copy.

In the majority of this work, we consider a uniform template where there is one type of copy and template monomer, and hence,

#	Transition	Representation	Free energy change, ΔG	Rates
1	Monomer binding	R_{bind} ⇌ R_{unbind}	$\Delta G_{spec} - \ln[M] =$ $\ln(R_{unbind}/R_{bind})$	$R_{bind} = k_0[M]$
2	Monomer unbinding			$R_{unbind} = k_0 e^{\Delta G_{spec}}$
3	Tail binding	$R_{tail\ bind}$ ⇌ $R_{tail\ unbind}$	$\Delta G_{spec} - \ln[M_{eff}] =$ $\ln(R_{tail\ unbind}/R_{tail\ bind})$	$R_{tail\ bind} = k_0[M_{eff}]$
4	Tail unbinding			$R_{tail\ unbind} = k_0 e^{\Delta G_{spec}}$
5	Release	$R_{release}$ →	Irreversible, $\Delta G = -\infty$	$R_{release} = k_0 e^{\Delta G_{spec}}$
6	Polymerisation	R_{polym} ⇌ $R_{depolym}$	$\Delta G_{BB} = \ln(R_{depolym}/R_{polym})$	$R_{polym} = k$
7	Depolymerisation			$R_{depolym} = k e^{\Delta G_{BB}}$

FIG. 3. Seven types of transition constitute the basic model of templated polymerization. Each transition modifies the system at a single site or, in the case of polymerization, between neighboring sites. The free energy change associated with the transition constrains the log of the ratio of backward to forward transitions according to the principle of local detailed balance, which reduces the number of free parameters in our model and assures that the model is thermodynamically self-consistent. Final release of a polymer is assumed to be irreversible and hence is associated with an infinite free energy change.

the specific bond strength, ΔG_{spec}, takes one value. In Sec. III D, we consider a model with a binary copy monomer alphabet with $\alpha_C = 2$.

4. Parameterization of transition rates

We assume that each transition described in Sec. II A 2 and depicted in Fig. 3 is well described as an instantaneous process with an average rate. Since the thermodynamics of copy production is an important feature underlying its physics,[28] we ensure that the thermodynamics of the system is self-consistent by applying the principle of local detailed balance to set the transition rates.[68] The principle of local detailed balance states that the ratio of forward to backward rates between any pair of states is constrained by the chemical free-energy change associated with the forward transition ΔG,

$$\frac{R_{forward}}{R_{reverse}} = e^{-\Delta G}. \tag{2}$$

We use this equation to parameterize the transition rates between states in our system, as shown in Fig. 3.

Monomer binding and unbinding is associated with a free-energy change of $\Delta G = \Delta G_{spec} - \ln[M]$, since one specific bond is formed in the process of monomer binding. The ratio of the monomer binding rate R_{bind} to the monomer unbinding rate R_{unbind} is given by

$$\frac{R_{bind}}{R_{unbind}} = [M]e^{-\Delta G_{spec}}. \tag{3}$$

We begin by assuming that monomers bind to the template sites with mass action kinetics, giving $R_{bind} \propto [M]$ and that the monomers unbind from the template with a rate that is exponentially dependent on the strength of the specific bond ΔG_{spec}, and hence, we arrive at

$$R_{bind} = k_0 [M] \text{ and } R_{unbind} = k_0 \, e^{\Delta G_{spec}}. \tag{4}$$

In Sec. III D, we consider the competition between two types of monomer and compare the effects of two different ways of parameterizing the unbinding and binding rates in which either the unbinding rate or the binding rate is exponentially dependent on ΔG_{spec}. Prior to Sec. III D, we maintain the parameterization given above for simplicity.

The free-energy change associated with copy units in the tails at either end of polymers binding to a template site is $\Delta G = \Delta G_{spec} - \ln[M_{eff}]$. Therefore, the rates of tail binding and unbinding are constrained by

$$\frac{R_{tail\,bind}}{R_{tail\,unbind}} = [M_{eff}]e^{-\Delta G_{spec}}. \tag{5}$$

Following a similar argument to that for monomers binding from solution, we set the rate that a unit in a copolymer tail rebinds to an available site on the template to be

$$R_{tail\,bind} = k_0 [M_{eff}]. \tag{6}$$

The rate at which polymer units unbind from a template site to become part of a tail is then identical to the rate at which a monomer in the same position would detach,

$$R_{tail\,unbind} = k_0 \, e^{\Delta G_{spec}}. \tag{7}$$

In this model, the final detachment and release of a copolymer is irreversible. Release occurs when the last copy–template bond is broken, and therefore, we set the rate of release to be equal the rate at which monomers unbind from the template, giving

$$R_{release} = k_0 \, e^{\Delta G_{spec}}. \tag{8}$$

Polymerization between units attached to the template at site i and $i+1$ is associated with a free energy change of ΔG_{BB}. Therefore,

$$\frac{R_{polym}}{R_{depolym}} = e^{-\Delta G_{BB}}. \tag{9}$$

We choose $R_{polym} = k$, and therefore, $R_{depolym} = ke^{\Delta G_{BB}}$. Under this parameterization, increasing the backbone bond strength increases the average lifetime of a backbone.

With these assignments for the free-energy change of each reaction, the overall free-energy change of incorporating a single monomer into a polymer in solution is

$$\Delta G_{pol} = \Delta G_{BB} - \ln \frac{[M]}{[M_{eff}]}, \tag{10}$$

as required.

B. Methods

1. Gillespie simulation

We simulate the dynamics of this system using Gillespie's kinetic Monte Carlo simulation method[69] with two levels of stochastic sampling. The first level of sampling selects the template site to update, with a probability proportional to the sum of transition rates at that site, and the second selects the transition rule to apply at that site with a probability proportional to its individual transition rate. This two-stage sampling method is advantageous because the valid transitions of the system only change the rates for subsequent transitions in a localized region of the system. For instance, a monomer binding the template at site 3 will not change the conformation of the system at site 10 or 100 and hence does not change the valid transitions around site 10 or 100. Given the locality of the model we have developed, we need only need to recalculate the valid transitions and sum over the valid rates at the few sites either side of the site that was last updated at each step.

2. Parameters, initial conditions, and stopping criteria

We use dimensionless units by setting $k_B T = 1$ and $k_0 = 1$. We also use $[M_{eff}] = 100$ throughout. We vary the free parameters k, $[M]$, ΔG_{spec}, ΔG_{BB}, ΔG_{gen}, and ΔG_{end} that will be introduced later.

Unless stated otherwise, the initial condition of the system in each simulation is an empty template. We simulate systems with templates of length $L_T = 10, 30$, and 100. We calculate statistics of the copying process, such as the mean polymer length and, later when considering systems with two distinct monomer pools, the error rate (defined as the number of mistakes contained within a polymer divided by its length), over the first fixed number of polymers (polymers with length ≥ 2) that form upon and are released from the template. We take averages of quantities such as the mean polymer length by running multiple simulations with the same parameters but different random number seeds. Where the number of polymers produced by the system (the sample size) generated within the allotted compute-time for each simulation is less than 10, the data are excluded as they are unrepresentative of the steady state. In each case, we run between $O(10)$ independent repeats for each set of parameter values, with each independent simulation stopping after producing $O(1000)$ polymers, though this varied from case to case. The code and input files required to reproduce the data presented in this work are freely accessible through the link provided under Data Availability.

III. RESULTS

A. A simple model of copy–template binding and copy–copy polymerization cannot reliably produce long polymers

We begin by asking whether long copy polymers can consistently be produced from a system of monomers governed by the simple model introduced in Sec. II A 4, with transitions parameterized as in Fig. 3. In Fig. 4, we analyze the polymers produced by the system on a template with length $L_T = 30$. We show the mean polymer length $\langle L \rangle$ against the free energy driving backbone formation on the template, ΔG_{BB}, as we vary the polymerization rate $k = [10^{-4}, 10^{-2}, 1, 10^2, 10^4]$, the monomer concentration $[M] = [0.01, 0.1, 1]$, the specific copy–template bond strength $\Delta G_{spec} = [-10, -6, -2, 2, 6]$, and backbone strength ΔG_{BB} ranging from -4.6 to -20. Five independent simulations were run for each parameter value, with a target of 2000 products. When parameter values where extreme, long embedded Markov processes were observed that were time consuming to simulate. Refer to Sec. S1 of the supplementary material for a discussion of these parameter values.

As shown in Fig. 4, the mean length of polymers produced by the system remains low, $2 \lesssim \langle L \rangle \lesssim 4$, across the parameter space. We now explore particular regions of parameter space to explain why. To help with this analysis, we consider a simple cut through the parameter space and also probe the actual state of the template during simulations. To do so, we consider the specific case with $\Delta G_{spec} = -4$, $[M] = 1$, and $k = 1$, keeping other parameters the same as in Fig. 4, and average over five independent repeats of 1000 polymers at each data point as we vary the backbone bond strength ΔG_{BB}. The results are reported in Fig. 5, alongside characteristic snapshots from simulations.

Parameter values where a template-bound polymer is unstable with respect to either the dissociation into monomers free in solution $(\Delta G_{BB} + \Delta G_{spec} - \ln[M] > 0)$ or unconnected
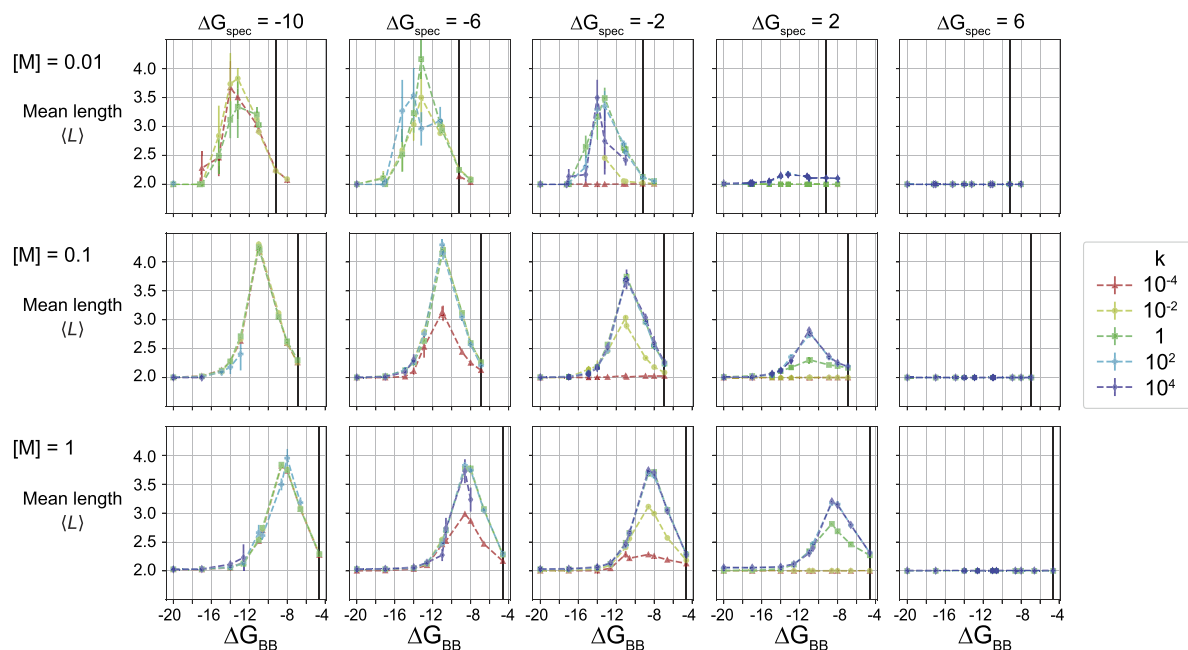

FIG. 4. For the basic model of Fig. 3, a wide search over multiple axes of the parameter space shows that the mean polymer length is always low. Each graph shows an average of the mean polymer length produced by the system, $\langle L \rangle$, with SEM error bars, on a template of length $L_T = 30$, against ΔG_{BB}, with $-\Delta G_{BB}$ being the driving force behind backbone bond formation. The monomer concentration $[M]$ is varied between rows and the specific bond strength, ΔG_{spec}, is varied between between the columns, while the polymerization rate, k, is varied and represented with different colors and marker shapes. Error bars are smaller than linewidth where not visible. A black line gives the value of ΔG_{BB} at which $\Delta G_{pol} = 0$. In all cases, $\langle L \rangle \ll L_T$, although a moderate peak in $\langle L \rangle$ is observed at moderate values of ΔG_{BB}.
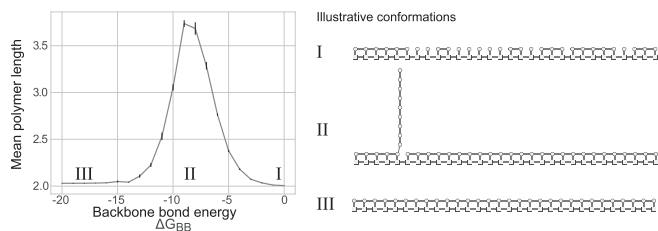
FIG. 5. Typical configurations help to explain the observed mean length $\langle L \rangle$ of copy products in the simple model of templating. We plot $\langle L \rangle$ against ΔG_{BB} for a template of length 30, $\Delta G_{spec} = -4$, $[M] = 1$, and $k = 1$, showing a moderate peak surrounded by regions of $\langle L \rangle = 2$. We extract typical configurations at the labeled values of ΔG_{BB}: (I) weak backbone, many small fragments; (II) moderate backbone strength, a small number of longer fragments; and (III) strong backbone, a single long polymer covers the whole template.

template-bound monomers $(\Delta G_{BB} > 0)$ do not feature heavily in Figs. 4 and 5 but are consistent with $\langle L \rangle \to 2$ as ΔG_{BB} gets more positive as the only "polymers" produced are transiently bound dimers.

As ΔG_{BB} becomes less positive/more negative, a significant number of longer polymers appear on the template. An example configuration is shown in Fig. 5(I). However, $\langle L \rangle \approx 2$ remains true because although longer copies are present on the template in significant numbers, they tend to detach more slowly than dimers due to the cooperative bond with the template. Therefore, dimers dominate the distribution of products.

At extremely negative values of ΔG_{BB}, configurations such as Fig. 5(III) are obtained. In this limit, a single long polymer covers the whole template. This configuration is extremely stable due to the difficulty in breaking the backbone bond and the cooperativity with which the long copy binds to the template. However, this cooperativity does not prevent the ends of the long polymer from "fraying" (undergoing transitory detachment from the template), which allows dimers to form under the raised tails; these dimers are subsequently released into solution must faster than the longer polymer and are thus registered as the only product (see Fig. 6).

For some sets of parameters, a moderate peak in $\langle L \rangle$ is observed between these limits. In this case, although ΔG_{BB} is negative, backbone bonds within the copies are broken at an appreciable rate,
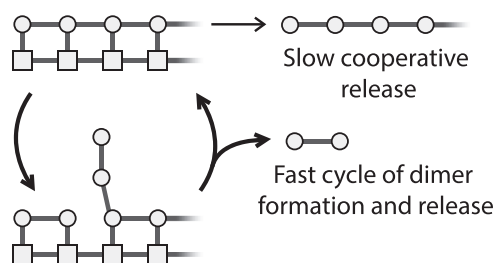

FIG. 6. Production of dimers in the limit of strong backbone bonds. Long polymers form on the template when the backbone bond ΔG_{BB} is sufficiently strong. Cooperative binding slows the complete release of the long polymer, while dimers are formed and released under its fraying ends.

resulting in relatively long fragments on the template. In some cases, via a mechanism of fraying and polymerization at the junction between fragments, one fragment can displace the other from the template in a step-by-step fashion. An intermediate state in this process is shown in Fig. 5(II). We observe that this peak occurs at slightly negative values of ΔG_{pol}, which can be rationalized by noting that adding a single monomer to the tail of template-attached polymer is associated with a free-energy change of ΔG_{pol}, and so states with long tails are thermodynamically unfavorable for $\Delta G_{pol} > 0$.

Although this mechanism can provide some increase in $\langle L \rangle$, a longer polymer is more likely to push a shorter one off the template, and no complete polymers are produced. Importantly, therefore, even the systems that give $\langle L \rangle \neq 2$ still produce outputs that are much smaller than the template and do not have a well-defined length. Either template-bound polymer formation is too unfavorable or cooperative template binding prevents the release of long stable polymers (product inhibition). Putative displacement mechanisms can be initiated at any location and produce fragmented polymers with a bias toward shorter products. These results suggest that for systems that are well-described by the basic model of Sec. II A 4, the basic templating mechanism is incapable of generating and, crucially, releasing long polymers reliably. Indeed, in Ref. 57, in a bottom-up, coarse-grained model of a replicating system that obeyed similar dynamics to that described by the model in Fig. 3, the authors observed the formation of only long polymers that remained bound in complexes under isothermal conditions.

In Sec. III B, we extend the model to allow for a more complex polymerization mechanism. This scheme can be viewed simultaneously as a way to avoid product inhibition, and a way to favor effective displacement mechanisms, by disrupting the cooperativity of copy–template interactions in a directional manner.

B. A polymerization mechanism that disrupts copy–template binding allows for long copies

In the context of simple catalysts, the authors of Ref. 70 argued that product inhibition can be reduced if part of the free energy of product formation is diverted into destabilizing the interaction of the catalyst and product. With that idea in mind, we develop a model in which backbone bonds in the copy can only be formed at the expense of breaking a bond between the copy and template. The scheme, illustrated in Fig. 7, could describe a range of chemical systems with competitive bond formation but, in particular, resembles the handhold-mediated strand displacement mechanism introduced by Cabello-Garcia et al.[5]

Formally, we implement this mechanism by splitting the copy–template bond into two parts, one of which must be broken for a backbone bond to form. The free-energy change of monomers binding to the template from solution is now given by $\Delta G = \Delta G_{spec} + \Delta G_{gen} - \ln[M]$, where the generic bond ΔG_{gen} is not dependent on the match between the copy and template. We assume that the polymers have a directional asymmetry (as is typical in macromolecular polymers such as DNA and RNA[31]), and we represent this in Fig. 7 by drawing the specific bond (dashed) on the left and the generic bond (solid) on the right. This asymmetry will be retained throughout this paper.

#	Transition	Representation	Free energy change, ΔG	Rates
1	Monomer binding	R_{bind}	$\Delta G_{spec} + \Delta G_{gen} - \ln[M]$	$R_{bind} = k_0[M]$
2	Monomer unbinding	R_{unbind}	$= \ln(R_{unbind}/R_{bind})$	$R_{unbind} = k_0 e^{\Delta G_{gen} + \Delta G_{spec}}$
3	Leading tail binding	$R_{lead\ tail\ bind}$	$\Delta G_{spec} + \Delta G_{gen} - \ln[M_{eff}]$	$R_{lead\ tail\ bind} = k_0[M_{eff}]$
4	Leading tail unbinding	$R_{lead\ tail\ unbind}$	$= \ln(R_{lead\ tail\ unbind}/R_{lead\ tail\ bind})$	$R_{lead\ tail\ unbind} = k_0 e^{\Delta G_{gen} + \Delta G_{spec}}$
3'	Tail binding	$R_{tail\ bind}$ $R_{tail\ unbind}$	$\Delta G_{spec} - \ln[M_{eff}]$	$R_{tail\ bind} = k_0[M_{eff}]$
4'	Tail unbinding	$R_{tail\ bind}$ $R_{tail\ unbind}$	$= \ln(R_{tail\ unbind}/R_{tail\ bind})$	$R_{tail\ unbind} = k_0 e^{\Delta G_{spec}}$
5	Leading release	$R_{lead\ release}$	Irreversible, $\Delta G = -\infty$	$R_{lead\ release} = k_0 e^{\Delta G_{gen} + \Delta G_{spec}}$
5'	Release	$R_{release}$	Irreversible, $\Delta G = -\infty$	$R_{release} = k_0 e^{\Delta G_{spec}}$
6	Polymerisation	R_{polym}	$\Delta G_{BB} - \Delta G_{gen}$	$R_{polym} = k$
7	Depolymerisation	$R_{depolym}$	$= \ln(R_{depolym}/R_{polym})$	$R_{depolym} = k\,e^{\Delta G_{BB} - \Delta G_{gen}}$

FIG. 7. The generic bond model is an extension of the basic model presented in Fig. 3. Here, there are two kinds of copy–template bond; the specific bond (dashed red line) that is potentially sensitive to the match between copy and template units and the generic bond (solid black line) that is not sensitive to the match. The generic bond is formed during monomer binding and breaks upon polymerization and vice versa for the reverse reactions. Units at the leading edge of copy polymers (at the right-hand side edge) bind to the template more strongly than other monomers in copies due to the modified polymerization mechanism. The generic bond energy, ΔG_{gen}, like the specific bond energy, ΔG_{spec}, does not contribute to the overall free energy change of polymer extension ΔG_{pol}.

As shown in reactions 6 and 7 of Fig. 7, the generic bond in the leftmost monomer breaks when polymerization occurs between neighboring monomers, leaving the leftmost monomer less tightly bound to the template. In visual representations, polymerization will always disrupt the bond of the left or "lagging" monomer; the generic bond of the right or "leading" monomer is unaffected. The total free-energy change of the template-attached polymerization step is then $\Delta G = \Delta G_{BB} - \Delta G_{gen}$. ΔG_{gen}, like ΔG_{spec}, does not contribute to the overall free energy change of extending a polymer tail (or a polymer in solution), $\Delta G_{pol} = \Delta G_{BB} - \ln[M]/[M_{eff}]$. We assume that the disruption of the generic bond during polymerization happens at the same time as the polymerization reaction, i.e., the bond is transferred from the template to polymer via an "attack" or an "invasion" mechanism, as occurs in Ref. 5.

Having modified the thermodynamics of the model, we modify the kinetics as follows. As shown in Fig. 7, generic bonds are only formed during the binding of monomers from solution when the leading monomer in a copy polymer rebinds to the template from a tail state or when the backbone between two template-attached monomers is disrupted. For simplicity, we assume that all unbinding reactions that break a generic bond have rates that scale as $R_{unbind} \propto e^{\Delta G_{gen}}$ (so ΔG_{gen} does not appear in the binding rates) and that $R_{depolym} = ke^{\Delta G_{BB} - \Delta G_{gen}}$, $R_{polym} = k$.

In Fig. 8, we present multiple contour plots of the averaged mean length of polymers produced via a template of length $L_T = 30$ as the backbone bond strength ΔG_{BB} and the generic bond strength ΔG_{gen} take values in the range $[0, -1, \ldots, -17]$. For each point sampled, we averaged the mean length of 1000 polymers each produced by five independent simulations. Data are excluded for simulations that did not produce this 1000 polymers within the allowed window. We set $k = 1$ and consider $[M] = [1, 10]$ and $\Delta G_{spec} = [-4, 2, 8]$. Additionally, we provide snapshots showing the wide variety of conformations reached by the system at a selected point in the parameter space when the simulation time passed 200 000 units.

In Fig. 8, we observe a high mean length of around $L_T/2 = 15$ over large regions of parameter space. This behavior corresponds to a fully occupied template, with a "brush" of polymers attached to it (regions exemplified by III, III', IV and V). In all of these regimes, the coupling of polymerization to the disruption of generic bonds with the template has two important effects. First, provided the disrupted bond is strong enough, the mechanism can overcome the cooperative effect that leads to product inhibition by a single long polymer. The leading monomer of a copy binds strongly, but the lagging monomers bind much more weakly. Undesirable confirmations with a single long polymer bound to the template are recovered (see configuration II in Fig. 8) when ΔG_{gen} is not large and negative,
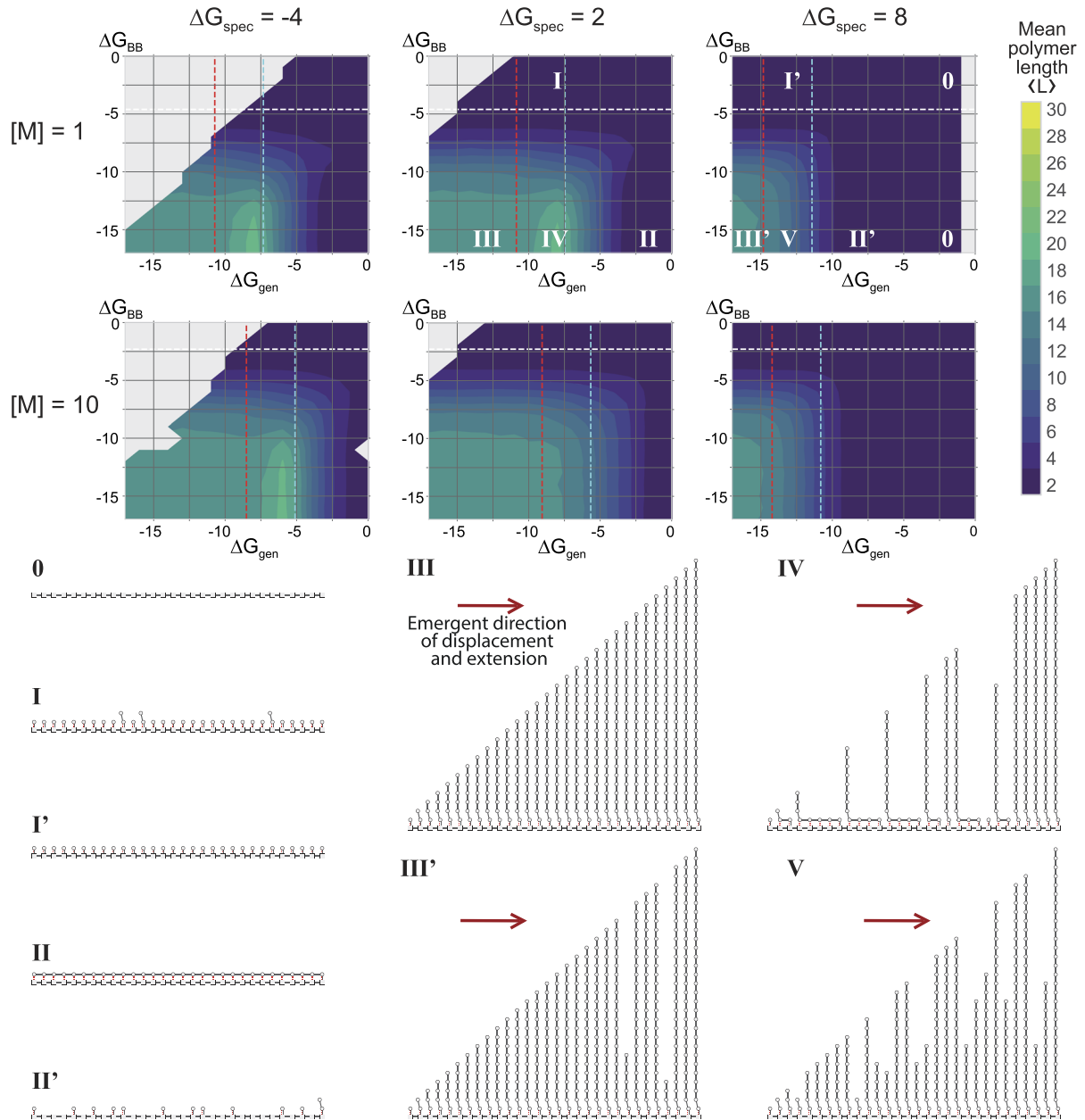
FIG. 8. Coupling polymerization to disrupting copy–template bonds can generate copies with average length $\langle L \rangle \sim L_T/2$. Above, we show surface plots of the mean polymer length, $\langle L \rangle$, against the backbone bond strength ΔG_{BB} and the generic bond strength ΔG_{gen} for monomer concentrations $[M] = 1, 10$ and $\Delta G_{spec} = -4, 2, 8$ on a template of length $L_T = 30$. Below, we provide snapshots of the system for the indicated points within the parameter space as the simulation time passed $t = 200\,000$. White dashed line constraint: $\Delta G_{BB} = \ln([M]/[M_{eff}])$. Above the white line, the template is either occupied by monomers and dimers (I and I′) or is empty 0 because the backbone bond is too weak. The system is only capable of producing dimers $\langle L \rangle \approx 2$. Dashed cyan line constraint: $\Delta G_{gen} < -\Delta G_{spec} + \ln \frac{2k_0}{L_T \langle \tau_{disp} \rangle}$. To the left of this line, the generic bond is strong enough that a lagging polymer tends to step forward faster than it detaches. To the right of the cyan constraint, we see either an empty template at 0 (since everything detaches rapidly) or, when the specific bonds are stronger, a template with a fully bound copy, since invading copies cannot force it off the template (configuration II). Here, the system still produces dimers $\langle L \rangle \approx 2$. Dashed red line constraint: $\Delta G_{gen} < -\Delta G_{spec} + \ln \frac{2k_0}{L_T^2 \langle \tau_{disp} \rangle}$. To the left of this constraint, lagging polymers step forward fast enough to create a dense, orderly polymer brush, as in III and III′. A uniform product length distribution is observed with a mean length $\langle L \rangle = L_T/2$. In between the red and cyan constraints, at IV, slower advancement of the lagging polymers leads either to a less dense brush with some tails occupying the template (IV) or the spontaneous appearance of gaps that encourage the initiation of shorter polymers within the brush (V). In configurations such as IV, $\langle L \rangle \gtrsim L_T/2$ is observed, whereas for parameters that lead to configurations like V, the mean length is slightly lower $\langle L \rangle < L_T/2$.

but the specific bond is strong. Second, the asymmetry introduced by the generic bond solves the problems associated with the unbiased strand displacement mechanisms mentioned in Sec. III A. The leading edge of a lagging polymer out-competes the tail of the polymer ahead by binding to the template with greater strength at their junction, allowing the system to act as a ratchet in which biased strand displacement causes shorter lagging polymers to push longer polymers ahead off the template. The emergent direction of displacement and extension is indicated in Fig. 8 by a red arrow. As shown in configuration III of Fig. 8, the result is a brush of polymers connected to the template by only one strong bond across the length of the template. The random detachment of polymers from this brush produces a distribution with an average polymer length $\langle L \rangle \sim L_T/2 = \sim 15$.

Some simple physical reasoning can be used to explain the regions of parameter space in which various behaviors are observed. For long polymer copies to form easily, incorporating monomers into copy polymers must be thermodynamically favorable.[30] Thus, $\Delta G_{\text{pol}} = \Delta G_{\text{BB}} - \ln([M]/[M_{\text{eff}}]) < 0$, and we require a sufficiently strong backbone,

$$\Delta G_{\text{BB}} \ll \ln([M]/[M_{\text{eff}}]). \tag{11}$$

The line representing this constraint is shown in white in Fig. 8. Above this line, the system exhibits either an empty template as in configuration 0 if the copy–template interaction if weak ($\Delta G_{\text{spec}} + \Delta G_{\text{gen}} > \ln[M_{\text{eff}}]$) or a template largely covered in monomers as in I and I′ if the copy–template interaction is strong ($\Delta G_{\text{spec}} + \Delta G_{\text{gen}} < \ln[M_{\text{eff}}]$). When the constraint of Eq. (11) is not satisfied, the system is only capable of producing short polymers, much for the same reasons we presented in Sec. III A and Fig. 5, as long polymers are unstable structures.

Given a sufficiently stable backbone, long polymers can form. However, the strength of the generic bond ΔG_{gen} can drastically change the conformations that these long polymers settle into and thereby change the typical length of polymers that are produced by the system. When the generic bond is weak, as in II where $\Delta G_{\text{gen}} = -2$, we tend to observe single, long polymers stuck on template, and dimers are created and released under the fraying end of the polymer as in IIIA. Note that when $\Delta G_{\text{gen}} = 0$, the system reverts back to the model presented in Sec. III A. When the generic bond is strong and takes very negative values, as in III and III′, we observe conformations where the polymer tails form a dense brush, and the mean length is roughly half the template length $\langle L \rangle = L_T/2$. We also observe regions with configurations such as IV and V in which the copy polymers form a less dense, less regular brush.

The key criterion for whether a brush of long polymers forms relates to two timescales. The first is the time taken for a polymer to polymerize forward into a space that is occupied by the lagging edge of the polymer in front. We call this time $\langle \tau_{\text{disp}} \rangle$, since the lagging tail of the polymer in front is effectively displaced (although in many cases its binding is weak to begin with). The second is the timescale on which the leading edge of a polymer detaches, $\langle \tau_{\text{unbind}} \rangle$. We can estimate $\langle \tau_{\text{disp}} \rangle$ with the simple discrete-state, continuous-time model in Fig. 9. We depict the four states involved in a single forward step alongside their transition rates.
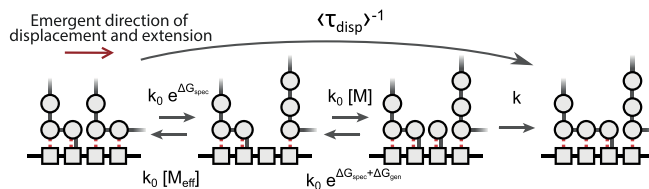

FIG. 9. Model to estimate the time, $\langle \tau_{\text{disp}} \rangle$, taken for a lagging polymer to extend by one unit, displacing the tail of the polymer ahead.

Treating this subprocess in isolation, we obtain a rate matrix $\mathbf{K}$ with entries,

$$\mathbf{K} = \begin{pmatrix} -k_0 e^{\Delta G_{\text{spec}}} & k_0[M_{\text{eff}}] & 0 & 0 \\ k_0 e^{\Delta G_{\text{spec}}} & -k_0[M_{\text{eff}}] - k_0[M] & k_0 e^{\Delta G_{\text{gen}}+\Delta G_{\text{spec}}} & 0 \\ 0 & k_0[M] & -k_0 e^{\Delta G_{\text{gen}}+\Delta G_{\text{spec}}} - k & 0 \\ 0 & 0 & k & 0 \end{pmatrix}.$$

The mean first-passage time $\langle \tau_{\text{disp}} \rangle$ from state 0 to state 3 is given by the $(3, 0)$ element of the Drazin inverse of $\mathbf{K}$, which takes a value

$$\langle \tau_{\text{disp}} \rangle = \left(k \left(1 + ([M] + [M_{\text{eff}}]) e^{-\Delta G_{\text{spec}}} \right) \right.$$
$$\left. + k_0 \left(e^{\Delta G_{\text{gen}}+\Delta G_{\text{spec}}} + [M] + e^{\Delta G_{\text{gen}}}[M_{\text{eff}}] \right) \right) / (k k_0[M]). \tag{12}$$

$\langle \tau_{\text{unbind}} \rangle$ is simply given by the inverse of the monomer unbinding rate, $\langle \tau_{\text{unbind}} \rangle = (R_{\text{unbind}})^{-1} = \frac{1}{k_0} e^{-\Delta G_{\text{gen}}-\Delta G_{\text{spec}}}$. To produce copies of length $\sim L_T/2$, it is necessary that $\langle \tau_{\text{disp}} \rangle L_T/2 < \langle \tau_{\text{unbind}} \rangle$; otherwise, the growing polymer will detach too rapidly. By rearranging this inequality, we arrive at

$$\Delta G_{\text{gen}} < -\Delta G_{\text{spec}} + \ln \frac{2k_0}{L_T \langle \tau_{\text{disp}} \rangle}, \tag{13}$$

which can be implicitly solved for ΔG_{gen}. This constraint is represented by the cyan dashed line in Fig. 8. When this constraint is not satisfied, newly initiated polymers tend to detach before they extend. When the total copy–template affinity $\Delta G_{\text{spec}} + \Delta G_{\text{gen}}$ is strong, as at II, we observe a single, long polymers stuck to the template, and when when the affinity is weak, at 0, we observe empty templates. Polymer brush configurations III, III′, IV, and V, which all produce relatively long distributions, all lie to the right of this constraint.

For a dense, ordered brush conformation as in III and III′ to be stable, $\sim L_T/2$ sequential extension and displacement steps must occur on a timescale shorter than that at which any of the L_T polymers fall off, enabling the system to heal the brush after any detachment events. Therefore, a dense, brush-like conformation will be reached when $\langle \tau_{\text{disp}} \rangle L_T/2 < \langle \tau_{\text{unbind}} \rangle / L_T$. Rearranging, we obtain

$$\Delta G_{\text{gen}} < -\Delta G_{\text{spec}} + \ln \frac{2k_0}{L_T^2 \langle \tau_{\text{disp}} \rangle}, \tag{14}$$

represented by the red dashed line in Fig. 8. In a maximally displaced, brush-like conformation, all polymers are only attached by their leading edge. Hence, detachment can occur equally from any point and the product length distribution is uniform, with a mean length very close to $\langle L \rangle = L_T/2$.

When ΔG_{gen} lies between the constraints [Eqs. (13) and (14)], the generic bond is strong enough to enable long polymers to form but insufficient to reach a maximally displaced, brush-like conformation. If the specific bond strength is strong enough that lagging tails tend to bind back to the template and are only removed via displacement, we see conformations similar to IV. Here, we see gaps between the leading edges of polymers, with weakly-bound tails occupying the intervening sites. Some shorter polymers occupy large swathes of the template, while some longer polymers have been pushed to the end of the template where they can advance no further. Here, we see a slight inversion of the cooperative effect; longer polymers that bunch at the end of the template tend to have fewer template connections than shorter polymers, which skews the product length distribution slightly toward long polymers giving $\langle L \rangle > L_T/2$.

This effect is not observed when the specific bond is weak enough that polymer tails can spontaneously detach from the template without the need for displacement, as seen at V where $\Delta G_{\mathrm{spec}} = 8 > \ln[M_{\mathrm{eff}}]$. Here, gaps appear quicker than they can be filled by extension of polymers already on the template, and monomers that bind within these gaps can sometimes form dimers. The tails of the dimers may spontaneously detach from the template, preventing the incorporation of the dimer into a polymer behind and initiating a new copy from the center of the template. As a consequence, the ordered polymer brush conformation of III or III$'$ is disrupted by shorter polymers that have incorrectly initiated in the middle of the template. Consequently, the system produces polymers with a mean length $\langle L \rangle < L_T/2$.

We have not observed any parameter values in which the template sites are typically unoccupied but in which long copies are frequently produced. To get long polymers, monomer binding in isolation must be stable $(R_{\mathrm{bind}}/R_{\mathrm{unbind}} > 1)$ since we have deliberately disrupted cooperative binding. Therefore, the only way that the template sites can typically be available is if both (a) lagging tails tend to unbind and (b) stepping forward is too fast for monomers to fill in behind. However, a monomer filling in happens at least as quickly as a polymer can step forward, by definition, and much faster if the polymer runs into a traffic jam of other polymers. So empty sites tend to be a rare commodity unless binding to the template is just pathologically unstable.

We have found that applying the two constraints of Eqs. (11) and (13) is generally sufficient to identify the region in which long copies are produced (Fig. 8). We note that in the upper left triangle of Fig. 8, where $G_{BB} \gg G_{\mathrm{gen}}$, the system failed to produce 1000 polymers during the allotted run-time and the data were excluded. When $G_{BB} \gg G_{\mathrm{gen}}$, the depolymerization step in the model becomes very fast, causing repeated polymerization and depolymerization events. In this region, the embedded process of polymer growth becomes very long and requires ever longer run-times to simulate. In the under-sampled regions, we have no evidence to suggest that the polymer length distributions produced by the system deviate from the simple arguments presented above.

By introducing a mechanism that channels the free energy released during polymerization into destabilizing the interactions of polymers tails and the template, we have shown that far longer polymers can be produced in a simple model of templating. However, for true copying, the polymers produced by the system must have identical lengths to the template. In Sec. III C, we bias the production complete polymers simply by weakening the copy–template interaction strength at end of the template.

C. Weakening the final site on the template biases the production of complete polymers

We now propose a simple adjustment to the model presented in Sec. III B that allows for the systematic release of complete polymers from the template when the system is in a brush-like conformation. Here, the copy–template bond at the final template site is weakened with a free-energetic factor $\Delta G_{\mathrm{end}} > 0$. This change is somewhat analogous to a stop codon in translation.[11] Crucially, since the end of the template is a unique site, this change can be made in a way that does not interfere with the copy–template interactions at other sites.

The effect of this adjustment is illustrated in the free energy profiles shown in Fig. 10. In Sec. III B, we saw that disrupting the copy–template bond upon polymerization could allow the lagging tail of polymers to detach from the template, meaning that, although polymerization was thermodynamically "down-hill," the binding strength of the copy to the template did not increase as the copy extends; the template binding free energy landscape was flat. Here, by weakening the copy–template interaction strength at the end of the template, ΔG_{end}, we expected to destabilize the binding of completed polymers and promote their release over incomplete polymers.
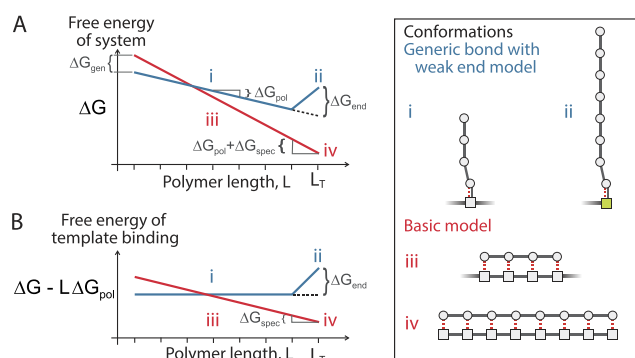
FIG. 10. (a) The free energy of a polymer, ΔG, decreases with length L when $\Delta G_{\mathrm{pol}} < 0$ in all models. (b) The component of the free energy due to the copy–template bonds is given by $\Delta G - L\Delta G_{\mathrm{pol}}$. The copy–template free energy decreases with length in the basic model (red line and conformations iii and iv) which suppresses the release of long polymers. In the maximally displaced conformations (i and ii) generated under the generic bond model (blue line), all polymers are only connected to the template at their leading edge, and hence, the free energy landscape is flat. Weakening the final site on the template by an amount ΔG_{end} destabilizes polymers with length $L = L_T$ (ii), leading to their selective release.

The thermodynamics and kinetics of the modified system are identical to the model presented in Sec. III B and Fig. 7, apart from at the very end of the template. In Fig. 11, we outline the thermodynamic and kinetic changes to the model. The free-energy change of binding at the end of the template is now given by $\Delta G = \Delta G_{\text{spec}} + \Delta G_{\text{gen}} + \Delta G_{\text{end}} - \ln[M]$, and the free-energy change of polymer tail binding at the end of the template is now given by $\Delta G = \Delta G_{\text{spec}} + \Delta G_{\text{gen}} + \Delta G_{\text{end}} - \ln[M_{\text{eff}}]$. The rates of monomer unbinding, leading-edge polymer unbinding, and polymer termination at the end of the template all take the value $R = k_0 e^{\Delta G_{\text{spec}} + \Delta G_{\text{gen}} + \Delta G_{\text{end}}}$; binding rates are unchanged relative to Sec. III B.

In Fig. 12, we plot the probability of producing a full-length copy polymer, p_{complete}, against the destabilizing free-energy penalty of the last site, ΔG_{end}, at a few values of the generic bond strength ΔG_{gen}, the monomer concentration $[M]$, and the template length L_T as a function of ΔG_{end}. We set $k = 1$, $\Delta G_{\text{spec}} = -4$, and $\Delta G_{\text{BB}} = -20$, which puts the system into a dense brush steady state conformation when $\Delta G_{\text{end}} = 0$ as seen in Sec. III B. For each parameter point sampled we ran five independent repeated simulations that would produce up to 2000 polymers or truncated after 4 h run-time. In Fig. 13, we show the full product length distribution for $[M] = 1$, $\Delta G_{\text{spec}} = -4$, and $L_T = 10$.

Figure 12 shows that weakening the end of the template can increase the proportion of complete products, p_{complete}, and that $p_{\text{complete}} \approx 1$ is possible under certain conditions. When $\Delta G_{\text{end}} = 0$, the polymer brushes encountered in Sec. III B tend toward a jammed state, as depicted in Fig. 14(a), with detachment events occurring at any point on the template. Weakening the final site on the template creates a release point that allows this polymer "traffic" to flow through, with unbinding predominantly occurring at the final site where complete polymers are attached by a weakened final copy–template bond, as seen in the binding free energy landscape in Fig. 10.

To probe this behavior in more detail, we construct a simple model to explain the proportion of complete polymers in the product pool, p_{complete}, produced by the system when a dense conformation has been reached. This model considers only the rate of formation and release of complete polymers and the rate of a competing unwanted process of release of a uniform distribution of polymers from the body of the template. In the densely packed regime, as depicted in Fig. 14(a), all polymers have one bond with the template. The timescale on which one of the $L_T - 1$ incomplete polymers are released from a template of length L_T is approximately

$$\langle \tau_{\text{uniform}} \rangle = (L_T - 1)^{-1} R_{\text{lead release}}^{-1}$$
$$\approx (L_T - 1)^{-1} k_0 e^{-\Delta G_{\text{gen}} - \Delta G_{\text{spec}}}. \quad (15)$$

Next, we estimate the time it takes the system to complete a growing polymer and release it from the template. To extend the polymer attached to the penultimate site, a free monomer must occupy the final site. The fraction of time that this final site is occupied (as opposed to empty), g, can be approximated by the binding equilibrium of the monomer,

$$g \approx \frac{k_0 [M]}{k_0 [M] + k_0 e^{\Delta G_{\text{gen}} + \Delta G_{\text{spec}} + \Delta G_{\text{end}}}}. \quad (16)$$

Given a polymerization rate k, the average timescale on which a polymer with length $L_T - 1$ is completed by polymerizing with a monomer at the end site on the template is

$$\langle \tau_{\text{form}} \rangle = 1/gk \approx \frac{[M] + e^{\Delta G_{\text{gen}} + \Delta G_{\text{spec}} + \Delta G_{\text{end}}}}{[M] k}. \quad (17)$$

In this simple model, we assume that the time it takes this newly completed polymer to fall off the template is rate limited by either

#	Transition	Representation	Free energy change, ΔG	Rates
1	End monomer binding	$R_{\text{end bind}}$ $\rightleftharpoons$ $R_{\text{end unbind}}$	$\Delta G_{\text{spec}} + \Delta G_{\text{gen}} + \Delta G_{\text{end}} - \ln[M] = \ln(R_{\text{end unbind}} / R_{\text{end bind}})$	$R_{\text{bind}} = k_0 [M]$
2	End monomer unbinding			$R_{\text{unbind}} = k_0 e^{\Delta G_{\text{gen}} + \Delta G_{\text{spec}} + \Delta G_{\text{end}}}$
3	End leading tail binding	$R_{\text{end lead bind}}$ $\rightleftharpoons$ $R_{\text{end lead unbind}}$	$\Delta G_{\text{spec}} + \Delta G_{\text{gen}} + \Delta G_{\text{end}} - \ln[M_{\text{eff}}] = \ln(R_{\text{end lead unbind}} / R_{\text{end lead bind}})$	$R_{\text{end lead bind}} = k_0 [M_{\text{eff}}]$
4	End leading tail unbinding			$R_{\text{end lead unbind}} = k_0 e^{\Delta G_{\text{gen}} + \Delta G_{\text{spec}} + \Delta G_{\text{end}}}$
5	End leading release	$R_{\text{end release}}$ $\rightarrow$	Irreversible, $\Delta G = -\infty$	$R_{\text{end release}} = k_0 e^{\Delta G_{\text{gen}} + \Delta G_{\text{spec}} + \Delta G_{\text{end}}}$

FIG. 11. The dynamics of the system at the weakened end (green square) of the template. We extend the generic bond model presented in Fig. 7 and modify the copy–template interaction strength at the end of the template with an energetic term ΔG_{end}. When $\Delta G_{\text{end}} > 0$, the bond made between copy units and the final template site is weakened. As before, the right-hand side edge of the copy polymer is the leading edge.
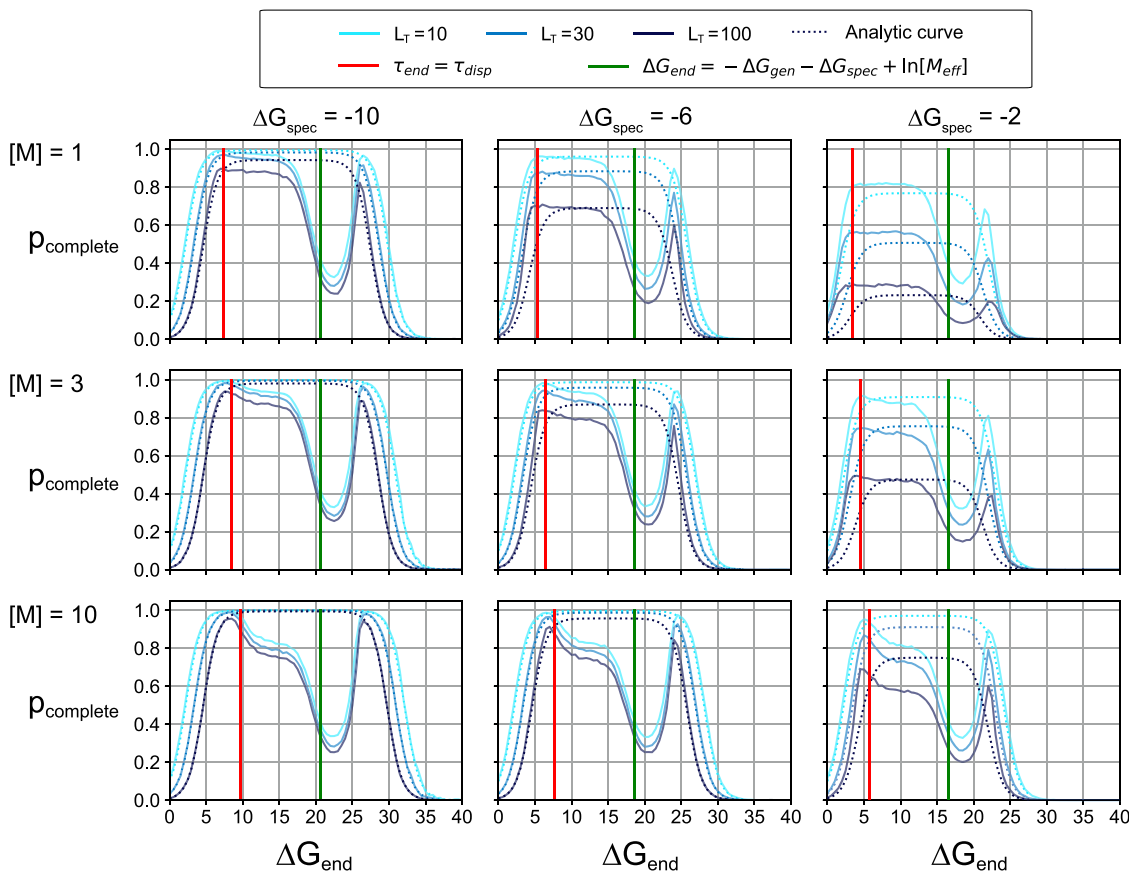
FIG. 12. Destabilizing the last site on the template with ΔG_{end} increases the probability of producing a complete polymer $p_{complete}$ when the configuration of copies forms a dense brush as seen in conformation III in Fig. 8. We plot the averaged probability of producing a complete polymer, $p_{complete}$, against the destabilizing energy penalty, ΔG_{end}, for a range of system parameters. We compare the data to a simple analytical model that accounts for the flux of complete polymers and uniformly distributed polymers from the template and also indicate the points at which $\langle \tau_{disp} \rangle = \langle \tau_{end} \rangle$ (red line) and $\Delta G_{end} = -\Delta G_{gen} - \Delta G_{spec} + \ln[M_{eff}]$ (green line) to guide interpretation, as explained in the text.

the timescale on which the bond with the end of the template is broken,

$$\langle \tau_{end} \rangle = 1/k_0\, e^{\Delta G_{gen}+\Delta G_{spec}+\Delta G_{end}}, \tag{18}$$

or the time it takes for the polymer unit at the penultimate template site to be displaced by a polymer behind, which is given by $\langle \tau_{disp} \rangle$ as calculated in Eq. (12) in Sec. III B.

Taken together, the timescale on which complete polymers fall off the template is approximately

$$\langle \tau_{fall} \rangle \approx \langle \tau_{end} \rangle + \langle \tau_{disp} \rangle, \tag{19}$$

and the total timescale of producing and releasing a complete polymer is

$$\langle \tau_{complete} \rangle \approx \langle \tau_{form} \rangle + \langle \tau_{fall} \rangle. \tag{20}$$

We can then approximate the steady state value of $p_{complete}$ as

$$p_{complete} \approx \frac{1/\langle \tau_{complete} \rangle}{1/\langle \tau_{complete} \rangle + 1/\langle \tau_{uniform} \rangle}. \tag{21}$$

With no free fitting parameters, this simple model—that accounts only for the production rate of the end and uniform components of the product distribution—captures remarkably well the range of ΔG_{end} over which $p_{complete}$ is high and the maximum value of $p_{complete}$ over the whole range of ΔG_{end}, as seen in Figs. 12 and S5. This fact suggests that the physics incorporated into this simple model explains the majority of the behavior.

Specifically, in Fig. 12, at $\Delta G_{end} = 0$, there is no bias for completed polymers to fall off the template. Products are dominated by dissociation from the body of the template, and we see a roughly uniform product length distribution as encountered in Sec. III B and $p_{complete} \approx 1/L_T$. At small values of ΔG_{end}, breaking the copy–template bond at the end of the template remains the rate-limiting step in producing a full-length copy and $\langle \tau_{complete} \rangle \approx \langle \tau_{end} \rangle$. As ΔG_{end} increases, the time to break the end copy–template bond, $\langle \tau_{end} \rangle$, decreases exponentially, and therefore,
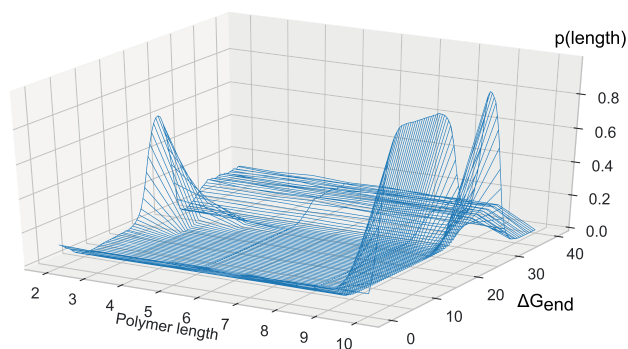
FIG. 13. The full length distributions of the data shown in Fig. 12 for [M] = 1 and $\Delta G_{\text{spec}} = -4$ and a template of length $L_T = 10$, averaged over five independent simulations at each value of ΔG_{end}. At $\Delta G_{\text{end}} = 0$, we see an approximately uniform product distribution. At low ΔG_{end}, we see a sharp increase in the probability of producing complete polymers with ΔG_{end} before a plateau is reached. At higher ΔG_{end}, a peak in the production of dimers is observed. As ΔG_{end} is increased further, there is a final peak in the production of full length products, after which the system produces a uniform distribution of products excluding complete polymers.

p_{complete} increases exponentially. Eventually, however, p_{complete} saturates when $\langle \tau_{\text{end}} \rangle < \langle \tau_{\text{disp}} \rangle$; at this point, $\langle \tau_{\text{complete}} \rangle \approx \langle \tau_{\text{disp}} \rangle$ and breaking the copy–template bond at the end is no longer rate-limiting. Here, the model suggests that no further increase in p_{complete} can be achieved by increasing ΔG_{end}; the value of the saturating level is determined by the relative timescales of displacement $\langle \tau_{\text{disp}} \rangle$ and spontaneous detachment from the body of the template $\langle \tau_{\text{uniform}} \rangle$.

At very high values of ΔG_{end}, the simple model accurately predicts the drop in p_{complete} to zero shown in Fig. 12. Here, the end bond is so unstable that the monomer occupancy of the final site on the template, g, tends to zero. Consequently, $\langle \tau_{\text{form}} \rangle$ diverges and no full length copies form. Instead, the system effectively operates as a uniform template of length $L_T - 1$ and generates a uniform distribution of polymers with lengths $L \leq L_T - 1$ but no polymers with length $L = L_T$. This behavior is particularly clear in the full length distribution shown in Fig. 13 in which at high ΔG_{end} no polymers are produced with $L = L_T = 10$.

The simple model presented here assumes that the system operates in the dense, ordered brush regime. However, in Fig. 12, where [M] = 1 and $\Delta G_{\text{gen}} = -8$, the constraint given in Eq. (14), which predicts whether the dense conformation is reached, $\langle \tau_{\text{disp}} \rangle L_T / 2 < \langle \tau_{\text{unbind}} \rangle / L_T$, is not satisfied. Here, the model's estimates of the level of the plateau in p_{complete} are less accurate. The biggest failure of the simple model, however, is not predicting the dips in p_{complete} that occur at intermediate values of ΔG_{end}. In Fig. 12, a single dip in p_{complete} is observed for [M] = 1, and a dip and an intermediate plateau are observed for [M] = 10. From looking at the full polymer length distributions shown in Figs. 13 and S2–S4, we can conclude that these dips correspond to a large spike in dimer production. Moreover, from snapshots of simulations, we observe that this dimer formation occurs at the final two sites on the template underneath the fraying end of completed but not yet detached-polymers [Fig. 14(e)].

Despite significant effort, a simple, quantitative model that predicts these "dimer dips" could not be constructed; however, a qualitative understanding of the factors that cause dimer production has been reached. The production of dimers occurs at
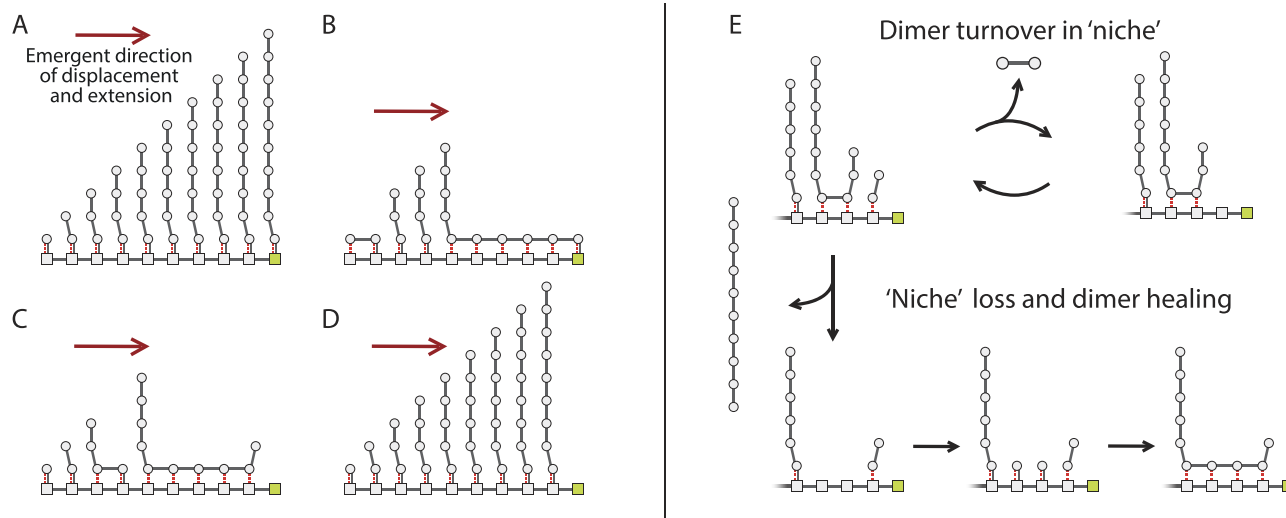

FIG. 14. Weakening the end of the template with ΔG_{end} affects the typical conformations of the system on the templates. We illustrate typical configurations from the simulations reported in Fig. 13. (a) At $\Delta G_{\text{end}} = 0$, the system reaches a dense brush. (b) At $\Delta G_{\text{end}} = 10$, completed polymers are released sooner than displacement can compensate, and so the brush of copies becomes less dense. (c) At $\Delta G_{\text{end}} = 22$, the final template site has low occupancy. Here, dimers form under the fraying ends of completed polymers. (d) At $\Delta G_{\text{end}} = 35$, the occupancy of the final site is so low that incomplete polymers detach sooner than they can complete by polymerizing at the end of the template. (e) Niche formation. The leading edge of completed polymers may fray, creating a protective niche that promotes the cyclical turnover of dimers. If the completed polymer falls off faster than a dimer can form, then the dimer can be incorporated into the preceding polymer.

fairly high values of ΔG_{end}, which are certainly not necessary for this the copying to be effective (it would be sufficient to just weaken the connection by a few units of thermal energy) and which may be experimentally inaccessible. Nonetheless, we offer an explanation of how this unexpected phenomenon arises in the model.

When ΔG_{end} is weakened sufficiently that $\langle \tau_{\text{end}} \rangle < \langle \tau_{\text{disp}} \rangle$ (shown as a red line in Fig. 12), the leading edge of completed polymers are released faster than the tails are fully displaced, leading to less dense conformations, such as Figs. 14(b) and 14(c). Completed polymers tend to be displaced from the template from their lagging edge and can spontaneously fray from their leading edge at the end of the template. When $\Delta G_{\text{end}} > -\Delta G_{\text{gen}} - \Delta G_{\text{spec}} + \ln[M_{\text{eff}}]$ (equality at green line in Fig. 12), frayed configurations at the leading edge, such as Fig. 14(c), are more likely than un-frayed alternatives, such as Fig. 14(b). This constraint on ΔG_{end} locates the dimer dip. If a completed polymer frays far enough to expose the penultimate site on the template, a monomer may occasionally bind there with a strong combined generic and specific bond. This penultimate monomer is unable to polymerize into the completed chain and instead dimerizes with monomers that occasionally bind to the final template site. Completed polymers can be understood to create a "protective niche" for the formation of dimers. Upon dimerizing, the dimer is rapidly released. This cycle is shown in Fig. 14(e).

If the timescale of formation and release of dimers in the protective niche is less than the average timescale on which the completed polymers are released, then multiple dimers may be released for each completed polymer. When ΔG_{end} is increased further, the monomer occupancy of the final site reduces and the time to form a dimer increases. If the timescale of formation and release of dimers in the protective niche is greater than the timescale on which the completed polymers are released, then the completed polymers fall off, removing the protective niche. As in Fig. 14(e), without the protective niche, the dimer is able to polymerize into the preceding polymer, which dramatically reduces the number of dimers that are released into the product pool. Hence, after the dimer dip, we observe a sharp increase in p_{complete} again, close to the saturating value that the simple model predicts, before the eventual drop off to $p_{\text{complete}} \approx 0$ at the highest values of ΔG_{end}. At high monomer concentrations, such as $[M] = 10$ in Fig. 12, we observe an intermediate plateau in p_{complete} as well as the dimer dip, which occurs when $\langle \tau_{\text{end}} \rangle < \langle \tau_{\text{disp}} \rangle$ shortly after the model predicts saturation. Due to the increased monomer concentration, fraying events at the end of the template result in more monomer invasions at the penultimate site, causing more frequent formation of dimers.

D. On-rate discrimination enables the reliable production of copies with the correct length and accurate sequences

In Sec. III C, we identified a mechanism for producing polymers of a single monomer type with a specific template-determined length in a reliable fashion. In this section, we consider whether the mechanisms that grant control over the length distribution are compatible with mechanisms that generate accurate copies from a pool of two types of monomer. We extend the models presented in

Secs. III B and III C and consider the growth of binary copies. The parameterization is given in Fig. S6 in the supplementary material. It is identical to that in Fig. 11, except that (i) ΔG_{spec} now depends on the match between the template and copy monomer and (ii) we allow the specific binding free energy to appear either in the binding or the unbinding rates.

In general, we now have four specific bond strengths, ΔG_{00}, ΔG_{01}, ΔG_{10}, and ΔG_{11}, where the first and second indices give the copy unit and template unit, respectively. We make the simplifying assumption that all copy monomer concentrations are equal $[M_0] = [M_1] = [M]$ and that all interactions are symmetric (all mismatches are equal, as are all matches) following Ref. 29. If the correct pairings, 00 and 11, have the same free energy, $\Delta G_{00} = \Delta G_{11} = \Delta G_r$, and if the incorrect pairings are also equal, $\Delta G_{01} = \Delta G_{10} = \Delta G_w$, we can describe the specific bond with two energy parameters, $\Delta G_{\text{spec,r}} = \Delta G_r$ or $\Delta G_{\text{spec,w}} = \Delta G_w$, the standard free energy of forming a right-match or wrong-match copy template bond. Given these simplifying assumptions, we can describe the process of the copy sequence simply in terms of whether the monomers are matched (right, r) or unmatched (wrong, w) with respect to the template, allowing us to solve the problem in a way that is independent of the specific template sequence in question.

Again, we use the principle of local detailed balance constraint the forward and backward transition rates between any pair of states. The ratio of the monomer binding rates $R_{\text{bind,r/w}}$ to the monomer unbinding rates $R_{\text{unbind,r/w}}$ for right or wrong template matches is proportional to the exponential of the specific bond strength (among other factors in Fig. 11),

$$\frac{R_{\text{bind,r/w}}}{R_{\text{unbind,r/w}}} \propto e^{-\Delta G_{r/w}}. \tag{22}$$

We compare the effects of two contrasting parameterizations of the model, detailed in Fig. S6 of the supplementary material. If the unbinding rates are dependent on the match between the copy and template, $R_{\text{unbind,r/w}} \propto e^{\Delta G_{r/w}}$, as was the case in the models presented in Secs. III A–III C, then the system has "off-rate" discrimination, and the monomer binding rates are constant. If the monomer binding rates are dependent on the match between the copy and template, $R_{\text{bind,r/w}} \propto e^{-\Delta G_{r/w}}$, then the system has "on-rate" discrimination and the unbinding rates are constant. As shown in Fig. S6 of the supplementary material, we apply the ΔG_{spec} sensitivity consistently across all binding and unbinding transitions for on- and off-rate discrimination. As in Sec. III C, all unbinding rates for units at the end of the template are increased by a factor $e^{\Delta G_{\text{end}}}$.

We take parameter values that produced complete polymers in Sec. III C. To generate the data shown in Fig. 15, we set $k = 1$, $\Delta G_{\text{BB}} = -20$, $\Delta G_{\text{gen}} = -12$, $\Delta G_{\text{end}} = 7$, and $[M_0] = [M_1] = 1$. We set $\Delta G_{\text{spec,w}} = \Delta G_w = -1$ for wrong-match monomers and $\Delta G_{\text{spec,r}} = \Delta G_r = \{-1, -2, \ldots, -8\}$ for right-match monomers. As described in Sec. II B, to generate each data point in Fig. 15, we ran five independent repeats of simulations that each generate up to 5000 copolymers. In addition to calculating the average of p_{complete}, we also calculate the average error rate per monomer (the number of wrong units in the copolymer divided by the polymer length) for complete polymers.

In Fig. 15(a), it is evident that on-rate discrimination maintains a high proportion of complete polymers even at long lengths
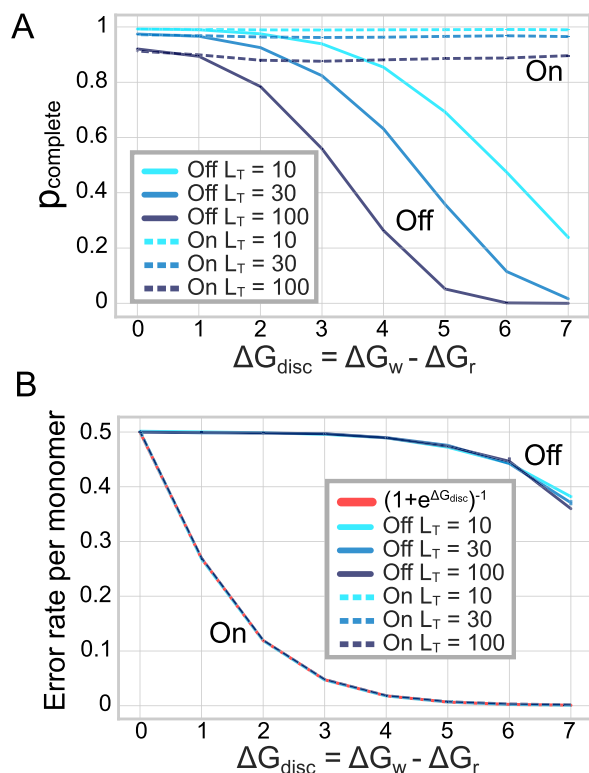
FIG. 15. Models with on-rate discrimination can yield complete polymers with high probability, $p_{complete} \approx 1$, and low error rates per monomer, whereas models with off-rate discrimination cannot reliably generate complete polymers with low error rates per monomer. We present (a) the mean of $p_{complete}$ and (b) the mean error rate per monomer over five repeats at each value of the discrimination free energy, $\Delta G_{disc} = \Delta G_w - \Delta G_r$, for simulations using the model in Fig. S6 of the supplementary material with an alphabet size of two. Simulations are performed for $L_T = 10$, 30, and 100 using $k = 1$, $\Delta G_{BB} = -20$, $\Delta G_{gen} = -12$, $\Delta G_{end} = 7$, and $[M_0] = [M_1] = 1$ for models with on- and off-rate discrimination, shown with dotted and solid lines, respectively. Standard Error of the Mean (SEM) is shown as error bars, which are generally smaller than linewidth. In (b), the error rates per monomer for different template lengths overlap one another and lie very close to the limiting line $(1 + \exp(\Delta G_{disc}))^{-1}$.

and large values of $\Delta G_{disc} = \Delta G_w - \Delta G_r$. Off-rate discrimination, however, fails to produce long polymers as the discrimination free-energy difference, $\Delta G_{disc} = \Delta G_0 - \Delta G_1$, grows larger. The generic bond and weak end mechanisms presented in Sec. III C, which generate and selectively release long polymers, require the leading edge of polymers to remain bound to the template until the end of the template is reached. However, if the system discriminates on the off-rates, then the incorporation of incorrect monomers into a polymer can cause it to be released prematurely, thereby reducing $p_{complete}$.

In Fig. 15(b), on-rate discrimination delivers low error rates per monomer, close to the equilibrium ratio implied by the discrimination free energy $1/(1 + \exp(\Delta G_{disc}))$, even though the products are far from equilibrium after separation from the template. Off-rate discrimination is inaccurate for moderate discrimination energies as

a strong ΔG_{gen} allows incorrect monomers to be incorporated into copolymers. For off-rate discrimination, there is a trade-off between accuracy and $p_{complete}$ for the mechanisms we have considered here. Hence, on-rate discrimination is needed to gain control of the length and accuracy in these minimal models of autonomous templated copying.

IV. CONCLUSION

Unlike the template-copying machinery found in biological systems that can cyclically copy templates under constant environmental conditions,[6–16] state-of-the-art synthetic molecular copying systems require external intervention or non-chemical driving[35–38,43–46] to operate cyclically. Here, we have developed coarse-grained models of isothermal, enzyme-free, templated copolymerization processes in which the copy copolymers can spontaneously detach from the template in a constant environment, driven by chemical free energy alone.

The central challenge of producing polymer copies is overcoming product inhibition, which is exacerbated by the natural cooperativity of the interactions between two polymers. We have shown that a surprisingly simple mechanism—using the free-energy of polymerization to weaken the interaction between the monomers and the template behind the leading edge of the copy—is sufficient to generate long polymers reliably. Combining this mechanism with a weakened copy–template bond at final site on the template, as in Sec. III C, is sufficient to reliably generate copies of the full template length under a wide range of conditions.

These proposed mechanisms require the following: (1) strong template–monomer bonds (so that the leading edge of growing polymers remain bound to the template for long enough to extend along its full length); (2) strong backbone bonds between monomeric units within polymers (so that formed polymers do not spontaneously fragment); (3) polymers upon formation cause major disruption to the copy–template bond of the "back" monomer (so that the long cooperatively binding tails of polymers can be displaced from the template by other polymers); and (4) a moderate destabilization of the copy–template bond at the end of the template (to enable the selective release of fully formed polymers). To copy specific sequences in addition to producing polymers of a fixed length, it is necessary to also have "on-rate" discrimination in which matching sequences bind to the template faster.

The conditions stated above can be met by DNA-based chemical reaction systems that exploit handhold-mediated DNA strand displacement (HMSD).[5] In HMSD, polymerization between monomers on the template results in the weakening of the copy–template connection in the "lagging" monomer through the transfer of bonds via strand displacement. Moreover, templating via HMSD has the potential to achieve effective on-rate discrimination during monomer binding using a mechanism in which a transient match-specific toehold forms first, with the strong generic bond only forming after an additional (toehold-mediated) strand displacement reaction.[71,72]

The models we have introduced have a surprisingly rich range of behaviors, and their dynamics did not match our initial expectations. We expected that optimal copying would involve isolated copies on a template, as is frequently observed in transcription[8–10] and translation[11–16] and as reflected in previous models of

templated copolymerization.[29] In such a setting, it would be necessary to prohibit copies from starting at random locations on the template, thereby generating truncated products. However, the natural tendency of brush-like configurations observed here to absorb isolated monomers into growing copies at all points on the template solves that challenge, which otherwise might require complex activation and deactivation of template sites ahead of and behind the growing copy. Moreover, the observed displacement behavior means that copying is possible even when the tails of polymers have moderate affinity with the template; indeed, this regime was more effective in producing longer polymers than regimes in which tail affinity was so low that they would detach without displacement, as described in Sec. S3 and Fig. S5 of the supplementary material.

Extant biological template-copying processes, such as transcription[8–10] and translation,[11–16] can operate in sparse regimes in which the copies do not fully occupy the template, although reasonably dense regimes are also observed.[73,74] The sparse regime of operation is possible because the initiation of polymer growth only occurs at specific locations on the template, at ribosome binding sites in translation (for instance, Ref. 11), in tightly controlled ways that involve complex enzymatic machinery. Similarly, the enzymatic machinery of transcription and translation provides a larger moving window in which copies are bound to their template.[11–16] This factor enables off-rate discrimination between the incoming monomers without causing partially formed polymers to fall off prematurely. It is challenging to see how to engineer this in a simple setting without the interplay of enzymes that can define a scale for this template-attached region, but such a scheme may facilitate additional mechanisms, such as kinetic proofreading to increase the accuracy above the limit set by the free energetic discrimination between monomer types.

Mechanisms that rely on displacement of copies by the subsequent copies have been considered as solutions to product inhibition in an RNA world.[66,75] In particular, Tupper and Higgs argued that a rolling-circle copying of a circular template can alleviate product inhibition via strand displacement.[66] This rolling-circle approach eliminates the tendency of short, newly initialized copies to be out-competed by longer polymers on a linear template, as we observed in Sec. III A. The mechanism assumes directional polymerization, but polymerization is still potentially hampered by a displacement-based competition between polymer ends for binding to the template, with no intrinsic bias favoring the desired pre-polymerization configuration. The asymmetric destabilization in our model induces a directionally biased strand displacement between copy polymers, which can overcome product inhibition even on linear templates. Our mechanism may also work well in conjunction with the rolling-circle mechanism of Tupper and Higgs, providing a directional bias to resolve the competition for the template. However, the chemistry and bonding stability of individual nucleotides as explored by Tupper and Higgs[66]—as opposed to larger oligonucleotide monomers as used in processes like HMSD[5]—may not be well-suited to the mechanisms we have investigated in this work. Outside of DNA-based reactions, Osuna Gálvez and Bode's reaction—in which polymerization is directly coupled to the disruption of product-template connections—shares the most resemblance to our mechanisms, though the templates cannot be reused. It is an open question whether small organic molecules with the appropriate chemistry can be identified and used to build synthetic copying systems with the properties we have stipulated.

We have attempted to keep the models as simple and general as possible so that they can be used to guide the design of synthetic copying systems in a wide array of contexts. We have also attempted to make all chemical steps explicit, rather than invoking a chemical *deus ex machina* that resolves the central challenges of the copy processes, just as highly evolved enzymatic machinery do *in vivo*. A key assumption of our model, however, is that (de)polymerization only occurs, while monomers are attached to the template—the template acts as a catalyst for bond formation and breakage. A passive template that simply brings reactants into close proximity could certainly allow polymerization reactions that are prohibitively slow in solution to proceed on the template. However, such a mechanism would not accelerate depolymerization as well, and so unless the template has some direct chemical coupling to the bond formation mechanism, as is typical in enzymes,[6–16] it is hard to see how to justify this catalytic assumption. Indeed, as we saw in Sec. III A, simple templates are not effective catalysts in our model—they accelerate the polymerization reaction by forming a stable complex with the product, preventing cyclic copying. However, in our modified mechanism presented in Sec. III B, the competition between the polymerization bond and the generic bond with the template provides a direct mechanism for catalysis, justifying the assumption that (de)polymerization only occurs while monomers are attached to the template.

We have also assumed that polymers that detach from the template and are released into the bath do not rebind to the template. One could imagine, if the concentration of polymers in the bath grew to significant levels, that rebinding could occur. Indeed, the rebinding and subsequent elongation of products was presented as a plausible pathway to the generation of longer RNA oligomers, bridging a portion of the gap between short RNA building blocks and the long RNA molecules that make up ribozymes in a pair of papers on templated ligation in an RNA world.[56,57] However, in our mechanism, polymerization is coupled to the disruption of the copy–template bonds, reducing the interaction strength that most units in the polymer chain could rebind by. Only the leading edge of a grown polymer could potentially stick tightly to the template, possibly causing unwanted over-extension of polymers if this reattachment happened at an earlier site on the template that happened to be complementary. If one were to design a copying machine based upon the biased displacement mechanism introduced here, a special monomeric unit could be reserved to bind effectively only to the ends of templates and, thus, cap the end of completed polymers. This approach would prevent further unwanted extension and reduce the copy–template affinity for completed products, alleviating the aforementioned issues.

Another key assumption of this work is that there is no interaction between polymer tails. Our models suggest that a dense brush of polymers is beneficial for length control; however, this may introduce interactions between polymer tails, especially on long templates, in *in vitro* settings. The importance of the interaction between polymer tails and the formation of bulges on the template should be considered on a case-by-case basis in experimental settings.

Finally, in our results, we have not focused on the speed with which copies are produced, only the relative yield of longer or

full-length copies. In this setting, we observed that polymer tails that have some affinity to the template are advantageous in maximizing the likelihood of producing full length copies. However, this behavior likely comes at the cost of somewhat slowed copy production, since template sites are more likely to be blocked. The optimal balance will depend on the system in question and again will require further investigation on a case-by-case basis.

SUPPLEMENTARY MATERIAL

The supplementary material contains the following items: a description of parameter values excluded from Fig. 4 for which few polymers are produced; the full polymer length distributions accompanying Fig. 12; a demonstration that the asymmetric destabilization mechanism is more effective when polymer tails have a higher affinity for the template; and a depiction of the model used in Sec. III D, for which there are two types of monomer.

ACKNOWLEDGMENTS

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

All authors conceived the project. J.J. wrote the simulations and performed the analysis. All authors interpreted the results and co-wrote this paper.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Zenodo at http://doi.org/10.5281/zenodo.5643001.

REFERENCES

[1] F. Crick, Nature 227, 561 (1970).
[2] M. P. H. Stumpf, T. Thorne, E. de Silva, R. Stewart, H. J. An, M. Lappe, and C. Wiuf, Proc. Natl. Acad. Sci. U. S. A. 105, 6959 (2008).
[3] L. Brocchieri and S. Karlin, Nucleic Acids Res. 33, 3390 (2005).
[4] P. Sartori and S. Leibler, Proc. Natl. Acad. Sci. U. S. A. 117, 114 (2020).
[5] J. Cabello-Garcia, W. Bae, G. B. V. Stan, and T. E. Ouldridge, ACS Nano 15, 3272 (2021).
[6] L. J. Reha-Krantz, Biochim. Biophys. Acta, Proteins Proteomics 1804, 1049 (2010).
[7] C. M. Joyce and S. J. Benkovic, Biochemistry 43, 14317 (2004).
[8] A. Vannini, Biochim. Biophys. Acta, Gene Regul. Mech. 1829, 258 (2013).
[9] R. H. Ebright, J. Mol. Biol. 304, 687 (2000).
[10] Y. Hirose and J. L. Manley, Genes Dev. 14, 1415 (2000).
[11] M. V. Rodnina, Cold Spring Harbor Perspect. Biol. 10, a032664 (2018).
[12] G. E. Fox, Cold Spring Harbor Perspect. Biol. 2, a003483 (2010).
[13] A. S. Petrov, B. Gulen, A. M. Norris, N. A. Kovacs, C. R. Bernier, K. A. Lanier, G. E. Fox, S. C. Harvey, R. M. Wartell, N. V. Hud, and L. D. Williams, Proc. Natl. Acad. Sci. U. S. A. 112, 15396 (2015).
[14] J. C. Bowman, A. S. Petrov, M. Frenkel-Pinter, P. I. Penev, and L. D. Williams, Chem. Rev. 120, 4848 (2020).
[15] R. Nikolay, D. van den Bruck, J. Achenbach, and K. H. Nierhaus, "Ribosomal proteins: Role in ribosomal functions," in eLS (John Wiley & Sons, Chichester, UK, 2015), pp. 1–12.
[16] G. Caetano-Anollés and D. Caetano-Anollés, Comput. Struct. Biotechnol. J. 13, 427 (2015).
[17] N. C. Seeman, J. Theor. Biol. 99, 237 (1982).
[18] P. W. Rothemund, "Folding DNA to create nanoscale shapes and patterns," Nature 440, 297 (2006).
[19] C. Lin, Y. Liu, S. Rinker, and H. Yan, ChemPhysChem 7, 1641 (2006).
[20] Y. Ke, L. L. Ong, W. M. Shih, and P. Yin, Science 338, 1177 (2012).
[21] N. C. Seeman and H. F. Sleiman, Nat. Rev. Mater. 3, 17068 (2017).
[22] R. Divine, H. V. Dang, G. Ueda, J. A. Fallas, I. Vulovic, W. Sheffler, S. Saini, Y. T. Zhao, I. X. Raj, P. A. Morawski, M. F. Jennewein, L. J. Homad, Y. H. Wan, M. R. Tooley, F. Seeger, A. Etemadi, M. L. Fahning, J. Lazarovits, A. Roederer, A. C. Walls, L. Stewart, M. Mazloomi, N. P. King, D. J. Campbell, A. T. McGuire, L. Stamatatos, H. Ruohola-Baker, J. Mathieu, D. Veesler, and D. Baker, Science 372, eabd9994 (2021).
[23] H. Shen, J. A. Fallas, E. Lynch, W. Sheffler, B. Parry, N. Jannetty, J. Decarreau, M. Wagenbach, J. J. Vicente, J. Chen, L. Wang, Q. Dowling, G. Oberdorfer, L. Stewart, L. Wordeman, J. De Yoreo, C. Jacobs-Wagner, J. Kollman, and D. Baker, Science 362, 705 (2018).
[24] J. Jin, E. G. Baker, C. W. Wood, J. Bath, D. N. Woolfson, and A. J. Turberfield, ACS Nano 13, 9927 (2019).
[25] J.-F. Lutz, M. Ouchi, D. R. Liu, and M. Sawamoto, Science 341, 1238149 (2013).
[26] D. L. Usanov, A. I. Chan, J. P. Maianti, and D. R. Liu, Nat. Chem. 10, 704 (2018).
[27] R. K. O'Reilly, A. J. Turberfield, and T. R. Wilks, Acc. Chem. Res. 50, 2496 (2017).
[28] T. E. Ouldridge and P. R. ten Wolde, Phys. Rev. Lett. 118, 158103 (2017).
[29] J. M. Poulton, P. R. ten Wolde, and T. E. Ouldridge, Proc. Natl. Acad. Sci. U. S. A. 116, 1946 (2019).
[30] J. M. Poulton and T. E. Ouldridge, New J. Phys. 23, 063061 (2021).
[31] B. Alberts, A. Johnson, J. Lewis, D. Morgan, and M. Raff, Molecular Biology of the Cell (Taylor & Francis Group, Oxford, 2014).
[32] J. Michaelis, A. Roloff, and O. Seitz, Org. Biomol. Chem. 12, 2821 (2014).
[33] H. Abe and E. T. Kool, J. Am. Chem. Soc. 126, 13980 (2004).
[34] A. Osuna Gálvez and J. W. Bode, J. Am. Chem. Soc. 141, 8721 (2019).
[35] D. Braun, Mod. Phys. Lett. B 18, 775 (2004).
[36] X. He, R. Sha, R. Zhuo, Y. Mi, P. M. Chaikin, and N. C. Seeman, Nat. Mater. 16, 993 (2017).
[37] R. Zhuo, F. Zhou, X. He, R. Sha, N. C. Seeman, and P. M. Chaikin, Proc. Natl. Acad. Sci. U. S. A. 116, 1952 (2019).
[38] D. Núñez-Villanueva, M. Ciaccia, G. Iadevaia, E. Sanna, and C. A. Hunter, Chem. Sci. 10, 5258 (2019).
[39] D. Núñez-Villanueva and C. A. Hunter, Org. Biomol. Chem. 17, 9660 (2019).
[40] D. Núñez-Villanueva and C. A. Hunter, Acc. Chem. Res. 54, 1298 (2021).
[41] D. Núñez-Villanueva and C. A. Hunter, Chem. Sci. 12, 4063 (2021).
[42] A. Kühnlein, S. A. Lanzmich, and D. Braun, eLife 10, e63431 (2021).
[43] R. Schulman, B. Yurke, and E. Winfree, Proc. Natl. Acad. Sci. U. S. A. 109, 6405 (2012).
[44] C. B. Mast and D. Braun, Phys. Rev. Lett. 104, 188102 (2010).
[45] C. B. Mast, S. Schink, U. Gerland, and D. Braun, Proc. Natl. Acad. Sci. U. S. A. 110, 8030 (2013).
[46] M. Kreysing, L. Keil, S. Lanzmich, and D. Braun, Nat. Chem. 7, 203 (2015).
[47] M. Colomb-Delsuc, E. Mattia, J. W. Sadownik, and S. Otto, Nat. Commun. 6, 7427 (2015).
[48] J. A. Baross and S. E. Hoffman, Origins Life Evol. Biospheres 15, 327 (1985).
[49] L. E. Orgel, J. Mol. Biol. 38, 381 (1968).

[50] F. H. C. Crick, J. Mol. Biol. **38**, 367 (1968).

[51] W. Gilbert, Nature **319**, 618 (1986).

[52] G. F. Joyce, Nature **338**, 217 (1989).

[53] H. Mutschler, A. Wochner, and P. Holliger, Nat. Chem. **7**, 502 (2015).

[54] L. E. Orgel, Crit. Rev. Biochem. Mol. Biol. **39**, 99 (2004).

[55] J. W. Szostak, J. Syst. Chem. **3**, 2 (2012).

[56] P. W. Kudella, A. V. Tkachenko, A. Salditt, S. Maslov, and D. Braun, Proc. Natl. Acad. Sci. U. S. A. **118**, e2018830118 (2021).

[57] J. H. Rosenberger, T. Göppel, P. W. Kudella, D. Braun, U. Gerland, and B. Altaner, Phys. Rev. X **11**, 031055 (2021).

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

[59] M. Ehrenberg and C. Blomberg, Biophys. J. **31**, 333 (1980).

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

[61] D. Andrieux and P. Gaspard, Proc. Natl. Acad. Sci. U. S. A. **105**, 9516 (2008).

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

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

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

[65] M. Esposito, K. Lindenberg, and C. Van den Broeck, J. Stat. Mech.: Theory Exp. **2010**, P01008.

[66] A. S. Tupper and P. G. Higgs, J. Theor. Biol. **527**, 110822 (2021).

[67] J. SantaLucia and D. Hicks, Annu. Rev. Biophys. Biomol. Struct. **33**, 415 (2004).

[68] T. E. Ouldridge, Nat. Comput. **17**, 3 (2018).

[69] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).

[70] A. Deshpande and T. E. Ouldridge, Biol. Cybern. **114**, 653 (2020).

[71] D. Y. Zhang and E. Winfree, J. Am. Chem. Soc. **131**, 17303 (2009).

[72] N. Srinivas, T. E. Ouldridge, P. Šulc, J. M. Schaeffer, B. Yurke, A. A. Louis, J. P. K. Doye, and E. Winfree, Nucleic Acids Res. **41**, 10641 (2013).

[73] I. Nanikashvili, Y. Zarai, A. Ovseevich, T. Tuller, and M. Margaliot, Sci. Rep. **9**, 1703 (2019).

[74] H. Zur and T. Tuller, Nucleic Acids Res. **44**, 9031 (2016).

[75] L. Zhou, S. C. Kim, K. H. Ho, D. K. O'Flaherty, C. Giurgiu, T. H. Wright, and J. W. Szostak, eLife **8**, e51888 (2019).