Nonequilibrium correlations in minimal dynamical models of polymer copying

Jenny M. Poulton[a], Pieter Rein ten Wolde[b], and Thomas E. Ouldridge[a,c,1]

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

Edited by David J. Schwab, Northwestern University, and accepted by Editorial Board Member Curtis G. Callan Jr. December 11, 2018 (received for review May 24, 2018)

Living systems produce "persistent" copies of information-carrying polymers, in which template and copy sequences remain correlated after physically decoupling. We identify a general measure of the thermodynamic efficiency with which these nonequilibrium states are created and analyze the accuracy and efficiency of a family of dynamical models that produce persistent copies. For the weakest chemical driving, when polymer growth occurs in equilibrium, both the copy accuracy and, more surprisingly, the efficiency vanish. At higher driving strengths, accuracy and efficiency both increase, with efficiency showing one or more peaks at moderate driving. Correlations generated within the copy sequence, as well as between template and copy, store additional free energy in the copied polymer and limit the single-site accuracy for a given chemical work input. Our results provide insight into the design of natural self-replicating systems and can aid the design of synthetic replicators.

biophysics | stochastic processes | thermodynamics | information transmission

The copying of information from a template into a substrate is fundamental to life. The most powerful copying mechanisms are persistent, autonomous, and generic. A persistent copy retains the copied data after physically decoupling from its template (1, 2). An autonomous copy process does not require systematically time-varying external conditions (2), making it more versatile. Finally, a generic copy process is able to copy arbitrary data. DNA replication and both steps of gene expression necessarily exhibit all three characteristics.

Unlike natural systems, synthetic polymer copying mechanisms developed hitherto have not incorporated all three features. Early work focused on using template polymers to synthesize specific daughter polymers, but failed to adequately demonstrate subsequent separation of copy and template (3, 4). We describe such a process as templated self-assembly (TSA), by analogy with molecular systems that assemble into a well-defined structure determined by highly specific interactions that are retained in the final state.

Due to cooperativity, the tendency of such copies to remain bound to templates grows with template length (5–7). Consequently, generic copying of long polymers [as opposed to dimers and trimers (5, 8, 9)] has proved challenging. One tactic is to consider environments in which the system experiences cyclically varying conditions, with assembly of the copy favored in one set of conditions and detachment from the template favored in another (10–12). A more subtle approach is to use a spatially nonuniform environment, so that individual molecules undergo cyclic variation in conditions (13). While these experiments may indeed reflect early life (14, 15), they do not demonstrate copying in a truly autonomous context.

We also contrast the copying of a generic polymer sequence with the approach in refs. 7 and 16. Here, the information is propagated between successive units of a single self-assembling polymer, rather than between a template and a daughter polymer, limiting information transmission. Externally induced mechanical stress on long length scales severs the polymers, leading to exponential growth of the number of polymers.

These challenges suggest that a full understanding of the basic biophysics of copying is lacking. Recently, we outlined fundamental thermodynamic constraints imposed by persistence (1), but did not propose a dynamical mechanism for autonomous copying. Previous dynamical models fall into two major categories: those that remain agnostic about the distinction between TSA and copying by considering thermodynamically self-consistent models for only part of the polymerization process (17) and those that explicitly address TSA (18–25).

In this work, we analyze a family of model systems that generate persistent copies in an autonomous and generic way. We introduce a metric for the thermodynamic efficiency of copying, and investigate the accuracy and efficiency of our models. We highlight the profound consequences of requiring persistence, namely, that correlations between copy and template can only be generated by pushing the system out of equilibrium. Previous work has considered self-assembly (26–28) or TSA (18–25, 28) in nonequilibrium contexts; in these cases, however, the nonequilibrium driving merely modulates a nonzero equilibrium specificity. Alongside the effect on copy–template interactions, we find that intra–copy-sequence correlations arise naturally. These correlations store additional free energy in the copied polymer, which does not contribute toward the accuracy of copying.

Significance

The ordering of chemical units within DNA, RNA, and proteins carries information about how living cells operate, and copying these sequences accurately is vital. We have a limited understanding of the fundamental physical underpinnings of these processes, since important mechanistic constraints due to the need to separate daughter sequences from their templates have never been investigated in detail. By considering the simplest models that incorporate these constraints, we highlight their profound consequences in terms of the effort that must be expended to make accurate copies. These insights will help us to understand not only life today but also how early replicators may have functioned in the past, and how we might develop synthetic copiers in the future.

Author contributions: T.E.O. designed research; J.M.P. performed research; J.M.P., P.R.t.W., and T.E.O. analyzed data; and J.M.P., P.R.t.W., and T.E.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.J.S. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. Email: t.ouldridge@imperial.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1808775116/-/DCSupplemental.

Published online January 18, 2019.

Models and Methods

Model Definition. We consider a copy polymer $M = M_1, \ldots, M_l$, made up of a series of subunits or monomers M_x, growing with respect to a template $N = N_1, \ldots, N_L$ ($l \leq L$). Inspired by transcription and translation, we consider a copy that detaches from the template as it grows; Fig. 1*B* shows the simplest model of this type. We consider whole steps in which a single monomer is added or removed, encompassing many individual chemical substeps (21, 24). After each step, there is only a single interpolymer bond at position l, between M_l and N_l. As a new monomer joins the copy at position $l+1$, the bond position l is broken, contrasting with previous models of TSA (20–25) (Fig. 1*A*). Importantly, as explained in the next paragraphs of this section, each step therefore depends on both of the two leading monomers, generating extra correlations within the copy sequence.

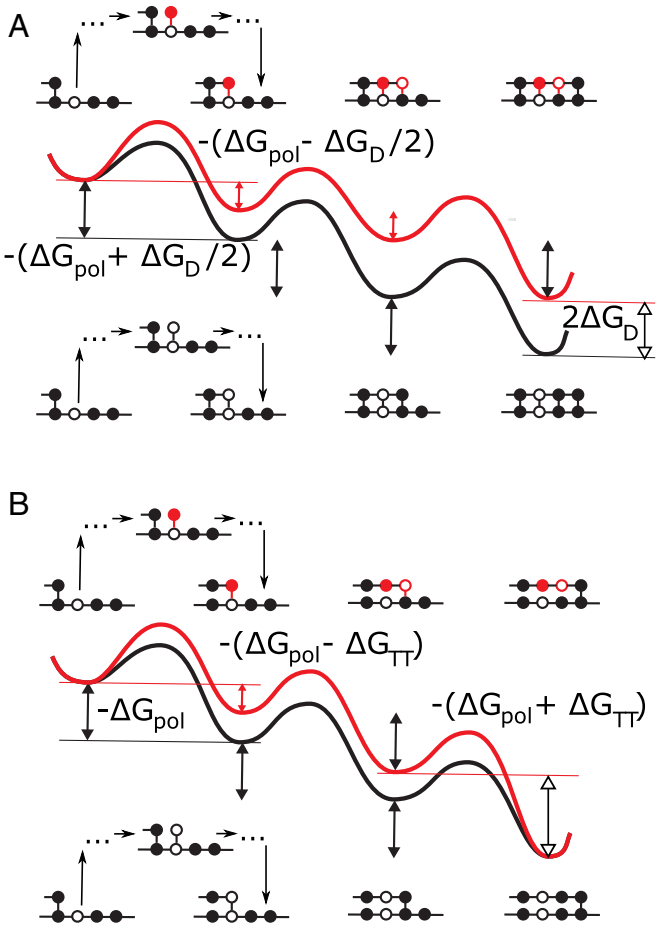

Fig. 1. Free-energy landscapes for simple examples of (*A*) TSA, in which the monomers remain bound to the template during the copy process and (*B*) persistent copying, in which the monomers detach from the template after they have been incorporated into the polymer. Both diagrams show the addition of three monomers to a growing polymer, driven by a chemical free energy of backbone polymerization ΔG_{pol}. In each subfigure, two scenarios are considered: the addition of two incorrect monomers, followed by a correct one (*Top*), and the addition of three correct monomers (*Bottom*). Local minima in the landscape represent macrostates following complete incorporation of monomers; intermediate configurations, illustrated schematically for the first transition, are part of the effective barriers. In TSA, the chemical free-energy cost of previously incorporated mismatches is retained as the daughter grows (20–25). Thus, in *A*, each mismatch in the daughter increases the chemical free energy by ΔG_{D} relative to the perfect match. In persistent copying (in *B*), the chemical free-energy penalty for incorporating wrong monomers is only temporary; it arises when the incorrect monomer is added to the growing polymer, but is lost when that monomer subsequently detaches from the template. As a result, the overall chemical free-energy change of creating an incorrect polymer is the same as that for a correct one. Analyzing the consequences of this constraint, which is a generic feature of copying but does not arise in TSA, is the essence of this work. The figure also shows that, in our specific model, incorporating a wrong monomer after a correct one tends to reduce the chemical free-energy drop to $\Delta G_{\text{pol}} - \Delta G_{\text{TT}}$, and incorporating a correct monomer after an incorrect one tends to increase it to $\Delta G_{\text{pol}} + \Delta G_{\text{TT}}$; however, adding a wrong monomer to a wrong one, and adding a correct monomer to a correct one, does not change the free-energy drop ΔG_{pol}.

Model Thermodynamics. Following earlier work, we assume that both polymers are copolymers, and that the two monomer types are symmetric (20–25). Thus, the relevant question is whether monomers M_l and N_l match; we ignore the specific sequence of N and describe M_l simply as right or wrong. Thus, $M_l \in r, w$, with example chain $M = rrwwrrrrrwrr$. An excess of r indicates a correlation between template and copy sequences. Breaking this symmetry would favor specific template sequences over others, disfavoring the accurate copying of other templates and compromising the generality of the process.

Given the model's state space, we now consider state free energies (which must be time-invariant for autonomy). We treat the environment as a bath of monomers at constant chemical potential (20–25). By symmetry, extending the polymer while leaving the copy–template interaction unchanged involves a fixed polymerization free energy. We thus define $-\Delta G_{\text{pol}}$ as the chemical free-energy change for the transition between any specific sequence $m_1, \ldots, m_l$ and any specific sequence $m_1, \ldots, m_{l+1}$, ignoring any contribution from interactions with the template. We then define ΔG_{TT} as the effect of the free-energy difference between r and w interactions with template. This bias can be describes as "temporary thermodynamic" (TT), since it only lasts until that contact is broken.

Overall, each forward step makes and breaks one copy–template bond. There are four possibilities: either adding r or w at position $l+1$ to a template with $M_l = r$ or adding r or w in position $l+1$ to a template with $M_l = w$. The first and last of these options make and break the same kind of template bond, so the total free-energy change is $-\Delta G_{\text{pol}}$. For the second case, there is a r bond broken and a w bond added, implying a free-energy change of $-\Delta G_{\text{pol}} + \Delta G_{\text{TT}}$. Conversely, for the third case, there is a w bond broken and a r bond added, giving a free-energy change of $-\Delta G_{\text{pol}} - \Delta G_{\text{TT}}$. These constraints are highlighted in Fig. 1*B*; the contribution of this work is to study the consequences of these constraints. Models of TSA (Fig. 1*A*) of equivalent complexity can be constructed, but they are not bound by these constraints, and hence the underlying results and biophysical interpretation are distinct.

Having specified model thermodynamics, we now parameterize kinetics. We assume that there are no "futile cycles," such as appear in kinetic proofreading (17). Reactions are thus tightly coupled: Each step requires a well-defined input of free energy determined by $-\Delta G_{\text{pol}}$ and $\pm\Delta G_{\text{TT}}$ (29), and no free-energy release occurs without a step.

A full kinetic treatment would be a continuous time Markov process incorporating the intermediate states shown schematically in Fig. 1*B*. To identify sequence output, however, we need only consider the state space in Fig. 1*B* and the relative probabilities for transitions between these explicitly modeled states, ignoring the complexity of nonexponential transition waiting times (21). We define propensity ψ^{+}_{xy} as the rate per unit time in which a system in state $\&x$ starts the process of becoming $\&xy$ and, propensity ψ^{-}_{xy} as the equivalent quantity in the reverse direction ($\&$ is an unspecified polymer sequence). Our system has eight of these propensities ($\psi^{\pm}_{rr}$, $\psi^{\pm}_{rw}$, $\psi^{\pm}_{wr}$, and $\psi^{\pm}_{ww}$); the simplest TSA models require four (20, 21, 23–25).

Prior literature on TSA (23) has differentiated between purely "kinetic" discrimination, in which r and w have an equal

template-binding free energy but different binding rates, and purely thermodynamic discrimination in which r and w bind at the same rate, but r is stabilized in equilibrium by stronger binding interactions. Eventually, all discrimination is kinetic for persistent copying, since there is no lasting equilibrium bias (Fig. 1*B*). However, by analogy with TSA, we do consider two distinct mechanisms for discrimination: a kinetic one, in which r is added faster than w to the growing tip, and one based on the TT bias toward correct matches at the tip of the growing polymer due to short-lived favorable interactions with the template, quantified by $\Delta G_{\mathrm{TT}} > 0$ (Fig. 1*B*). The kinetic mechanism should not be conflated with fuel-consuming "kinetic proofreading" cycles that are not considered.

We parameterize the propensities as follows. Assuming, for simplicity, that the propensity for adding r or w is independent of the previous monomer, we have $\psi^+_{rr} = \psi^+_{wr}$ and $\psi^+_{rw} = \psi^+_{ww} = 1$, also defining the overall timescale. Kinetic discrimination is then quantified by $\psi^+_{xr}/\psi^+_{xw} = \exp(\Delta G_{\mathrm{K}}/k_B T)$. Forward propensities are thus differentiated solely by ΔG_{K}; backward propensities are set by fixing the ratios ψ^+_{xy}/ψ^-_{xy} according to the free-energy change of the reaction, which follows from ΔG_{pol} and ΔG_{TT} (Fig. 1*B*). Thus, setting $k_B T = 1$,

$$\psi^+_{rr} = e^{\Delta G_{\mathrm{K}}},\ \psi^-_{rr} = e^{-\Delta G_{\mathrm{pol}}} e^{\Delta G_K}, \quad [1]$$

$$\psi^+_{rw} = 1,\ \psi^-_{rw} = e^{-\Delta G_{\mathrm{pol}}} e^{\Delta G_{\mathrm{TT}}}, \quad [2]$$

$$\psi^+_{wr} = e^{\Delta G_{\mathrm{K}}},\ \psi^-_{wr} = e^{-\Delta G_{\mathrm{pol}}} e^{\Delta G_{\mathrm{K}}} e^{-\Delta G_{\mathrm{TT}}}, \quad [3]$$

$$\psi^+_{ww} = 1,\ \psi^-_{ww} = e^{-\Delta G_{\mathrm{pol}}}. \quad [4]$$

For a given ΔG_{K}, ΔG_{pol}, and ΔG_{TT}, Eqs. **1–4** describe a set of models with distinct intermediate states that yield the same copy sequence distribution. We can thus analyze the simplest model in each set, which is Markovian at the level of the explicitly modeled states, and has $\psi^{\pm}_{xy}$ as rate constants.

Model Analysis. We use Gaspard's method to solve the system (28); we note that the underlying kinetic equations can also be mapped to models of distinct physical systems that have different constraints on the parameters (26). In this approach, the tip monomer identity probabilities $[\mu(m_l)]$, the joint tip and penultimate monomer identity probabilities $[\mu(m_{l-1}, m_l)]$, and the conditional probabilities $[\mu(m_{l-1} \mid m_l)]$ become stationary for a long polymer and can be calculated. One must first calculate the partial velocities, v_r and v_w. The quantity $v_x \mu(x)$ is the net rate at which monomers are added after an x,

$$v_x = \psi^+_{xr} - \frac{\mu(x \mid r)\mu(r)}{\mu(x)}\psi^-_{xr} + \psi^+_{xw} - \frac{\mu(x \mid w)\mu(w)}{\mu(x)}\psi^-_{xw}. \quad [5]$$

Following ref. 28, these velocities can be solved in terms of the propensities. In turn, the velocities and propensities determine tip and conditional probabilities $\mu(m_l)$ and $\mu(m_{l-1} \mid m_l)$ (28). Further details are provided in *SI Appendix*.

Gaspard's method describes the chain while it is still growing through the stochastic variables M_l and M_{l-1}, with the index l being the current length of the polymer. We, however, are interested in the identity of the monomer at position n when $l \gg n$. We label this "final" state of the monomer at position n as M^{∞}_n. As discussed in *SI Appendix*, M^{∞}_n is distinct from M_n near the tip. M^{∞}_n is described by the error probability ϵ and the conditional error probabilities ϵ_r and ϵ_w, defined as the probability that $M^{\infty}_{n+1} = w$ given that $M^{\infty}_n = r$ or $M^{\infty}_n = w$, respectively. In *SI Appendix*, we show that ϵ, ϵ_r and ϵ_w are sufficient to describe the final state, by proving that the M^{∞} is a Markov chain of r and w monomers (this requirement is distinct from the Markovian growth dynamics). We note that $\epsilon \neq \epsilon_r \neq \epsilon_w$ as a direct result of the dependence of the transition propensities on the current and previous tip monomers, which in turn arises from detachment.

To calculate ϵ, ϵ_r, and ϵ_w, we define currents J_{xy} that are related to $\psi^{\pm}_{xy}$, and to ϵ, ϵ_r, and ϵ_w, separately. The current J_{xy} is the net rate per unit time at which transitions $\&x \to \&xy$ occur: $J_{xy} = \psi^+_{xy}\mu(x) - \psi^-_{xy}\mu(x, y)$. By considering the transitions in our system as a tree, as in Fig. 2, we can relate the current through a branch to the overall rate at which errors are permanently incorporated into a polymer growing at total velocity $v = v_r\mu(r) + v_w\mu(w)$,

$$J_{rr} = (1-\epsilon)(1-\epsilon_r)v = \mu(r)\psi^+_{rr} - \mu(r, r)\psi^-_{rr}, \quad [6]$$

$$J_{rw} = (1-\epsilon)\epsilon_r v = \mu(r)\psi^+_{rw} - \mu(r, w)\psi^-_{rw}, \quad [7]$$

$$J_{wr} = \epsilon(1-\epsilon_w)v = \mu(w)\psi^+_{wr} - \mu(w, r)\psi^-_{wr}, \quad [8]$$

$$J_{ww} = \epsilon\epsilon_w v = \mu(w)\psi^+_{ww} - \mu(w, w)\psi^-_{ww}. \quad [9]$$

Eliminating ϵ from the simultaneous Eqs. **6–9** yields ϵ_r and ϵ_w in terms of known quantities. To find ϵ, note that the final sequence itself is a Markov chain with a transition matrix parameterized by ϵ_r and ϵ_w, with the overall error ϵ given by its dominant eigenvector. As detailed in *SI Appendix*, we obtain $\epsilon = \epsilon_r/(1 + \epsilon_r - \epsilon_w)$. From ϵ, ϵ_r, and ϵ_w, we calculate copy properties in terms of $\psi^{\pm}_{xy}$ and thus the free energies. We corroborate the results with simulation (see *SI Appendix*).

Results

General Thermodynamic Bounds. The free energy of the combined bath and polymer system decreases over time. There are two contributions to the free-energy change per added monomer: the chemical free energy ΔG_{pol} and a contribution from the uncertainty of the final polymer sequence (1). The latter is quantified by the entropy rate H (30, 31),

$$H(M^{\infty}) = \lim_{n\to\infty} \frac{1}{n} H(M^{\infty}_1, M^{\infty}_2, \ldots, M^{\infty}_n), \quad [10]$$

which, in our case, is given by (31)

$$H(M^{\infty}) = -\epsilon\,(\epsilon_w \ln \epsilon_w + (1-\epsilon_w)\ln(1-\epsilon_w)) - (1-\epsilon)\,(\epsilon_r \ln \epsilon_r + (1-\epsilon_r)\ln(1-\epsilon_r)). \quad [11]$$

The overall free-energy change per added monomer is then $\Delta G_{\mathrm{tot}} = -\Delta G_{\mathrm{pol}} - H$, which must be negative for growth: $H \geq -\Delta G_{\mathrm{pol}}$. Since copy–template interactions are not extensive in the copy length, they do not contribute. Given that $H \geq 0$, growth is possible in the region where $\Delta G_{\mathrm{pol}} < 0$, corresponding to the "entropically driven" regime (20, 25).

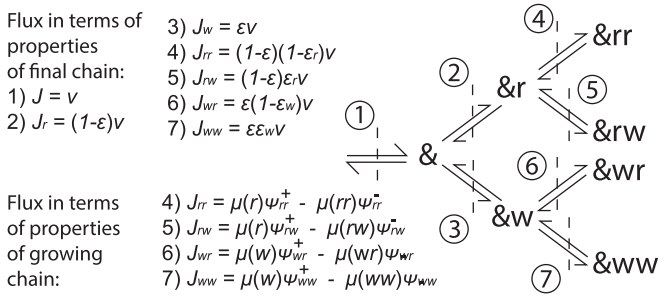

Fig. 2. Transitions of an arbitrary polymer &. To relate the final chain to the growing chain, it is useful to consider fluxes through interfaces in this transition diagram. Using the tip and combined probabilities, along with relative propensities, it is possible to describe the fluxes through interfaces 4 to 7 in terms of properties of the growing chain. Equally, by considering errors and conditional errors and taking fractions of the overall growth velocity, it is possible to find the fluxes through interfaces 4 to 7 in terms of properties of the final chain and growth velocity.

The entropy rate is bounded by the single-site entropy $H \leq H_{ss} = -\epsilon \ln \epsilon - (1-\epsilon)\ln(1-\epsilon)$. H_{ss} quantifies the desired correlations between copy and template. For previous models of TSA with uncorrelated monomer incorporation, $H = H_{ss}$ (20–25, 28). In our model, the necessary complexity of $\psi^{\pm}_{xy}$ generates correlations within the copy, as well as between copy and template. A stronger constraint on the single site entropy, and hence accuracy, then follows: $H_{ss} \geq H \geq -\Delta G_{\rm pol}$.

Fundamentally, a persistent copy is a high free-energy state, as the entropic cost of copy–template correlations cannot be counteracted by stabilizing copy–template interactions. Thus, the process moves a system between two high free-energy states, converting chemical work into correlations. In general, only a fraction of the chemical work done by the monomer bath is retained in the final state, implying dissipation, and so it is natural to introduce an efficiency. The overall free energy stored in the polymer has contributions both from the creation of an equilibrium (uncorrelated) polymer and from correlations within the copy and with the template. We are interested only in the contributions above equilibrium. The efficiency η is then the proportion of the additional free energy expended in making a copy above the minimum required to grow a random equilibrium polymer that is successfully converted into the nonequilibrium free energy of the copy sequence rather than being dissipated. In our simple case,

$$\eta \equiv \frac{H_{\rm eq} - H}{H_{\rm eq} + \Delta G_{\rm pol}} \leq 1. \qquad [12]$$

Here, $\Delta G_{\rm pol} + H_{\rm eq} = \Delta G_{\rm pol} + \ln 2$ is the extra chemical work done by the buffer above that required to grow an equilibrium polymer, $\Delta G^{\rm eq}_{\rm pol} = -H_{\rm eq} = -\ln 2$. The free energy stored in the copy sequence, above that stored in an equilibrium system, is $H_{\rm eq} - H$; $\eta \leq 1$ follows from $\Delta G_{\rm pol} + H \geq 0$. Similarly, since $H_{\rm ss} \geq H$, we can define a single-site efficiency,

$$\eta_{\rm ss} \equiv \frac{H_{\rm eq} - H_{\rm ss}}{H_{\rm eq} + \Delta G_{\rm pol}} \leq \eta \leq 1. \qquad [13]$$

Unlike η, the single-site efficiency $\eta_{\rm ss}$ discounts the free energy stored in "useless" correlations within the copy.

Behavior of Specific Systems. We consider three representative models consistent with Eqs. **1**–**4**. First, we consider the purely kinetic mechanism obtained by setting $\Delta G_{\rm TT} = 0$ and $\Delta G_{\rm K} = \Delta G$ in Eqs. **1**–**4**. Originally proposed by Bennett (20) for TSA, it is coincidentally a limiting case of persistent copying, since there is no equilibrium bias. We also consider two other mechanisms: pure "TT discrimination," with $\Delta G_{\rm K} = 0$ and $\Delta G_{\rm TT} = \Delta G$, and a "combined discrimination mechanism," in which both template binding strengths and rates of addition favor r monomers: $\Delta G_{\rm K} = \Delta G_{\rm TT} = \Delta G$.

All three mechanisms have two free parameters, the overall driving strength $\Delta G_{\rm pol}$ and the discrimination parameter ΔG. We plot error probability against $\Delta G_{\rm pol}$ for various ΔG in Fig. 3. Also shown is the thermodynamic lower bound on ϵ implied by $H_{\rm ss} \geq H \geq -\Delta G_{\rm pol}$. All three cases have no accuracy ($\epsilon = 0.5$) in equilibrium ($\Delta G_{\rm pol} \to -\ln 2$), since an accurate persistent copy is necessarily out of equilibrium (1). By contrast, TSA allows for accuracy in equilibrium (19, 23, 24).

The TT mechanism is always the least effective. It has no accuracy for high $\Delta G_{\rm pol}$, as the difference between r and w is only manifest when stepping backward, and, for high $\Delta G_{\rm pol}$, back steps are rare (23, 24). More interestingly, TT discrimination is also inaccurate as $\Delta G_{\rm pol} \to -\ln 2$, when the system takes so many back steps that it fully equilibrates. Low ϵ only occurs when $\Delta G_{\rm pol}$ is sufficient to inhibit the detachment of r monomers, but not the detachment of w monomers. This trade-off region grows with ΔG. By contrast, both the combined case and the kinetic case have accuracy in the limit of $\Delta G_{\rm pol} \to \infty$, since they allow r to bind faster than w. Considering $\Delta G_{\rm pol} \leq 0$ closely (Fig. 3*B*) shows the combined case to be superior.

Intriguingly, all three mechanisms are far from the fundamental bound on ϵ implied by $H_{\rm ss} \geq H \geq -\Delta G_{\rm pol}$ as $\Delta G_{\rm pol} \to -\ln 2$, and there is an apparent cusp in ϵ at $\Delta G_{\rm pol} \approx 0.48$ as $\Delta G \to \infty$ in the combined case. The performance relative to the bound is quantified by $\eta_{\rm ss}$ (Fig. 4). Surprisingly, we observe, in Fig. 4*A*, that not only does ϵ go to zero as $\Delta G_{\rm pol} \to -\ln 2$, but so does η_{ss} in all cases. For small nonequilibrium driving, none of the extra chemical work input is stored in correlations with the template. Mathematically, this inefficiency arises because $\epsilon - 0.5 \propto \Delta G_{\rm pol} - \ln 2$ as $\Delta G_{\rm pol} \to -\ln 2$ (as observed in Fig. 3), and $H_{\rm ss} - \ln 2 \propto (\epsilon - 0.5)^2$ for $\epsilon \approx 0.5$, by definition. Thus, from Eq. **13**, $\eta_{\rm ss} \propto \Delta G_{\rm pol} - \ln 2$ as $\Delta G_{\rm pol} \to -\ln 2$. That this result only depends on error probability decreasing proportionally with $\Delta G_{\rm pol}$ for small driving suggests that a vanishing η_{ss} in equilibrium may be quite general.

In all cases, the single-site efficiency $\eta_{\rm ss}$ increases from 0 at $\Delta G_{\rm pol} = -\ln 2$ to a peak near $\Delta G_{\rm pol} = 0$, with $\eta_{\rm ss} \to 1$ as

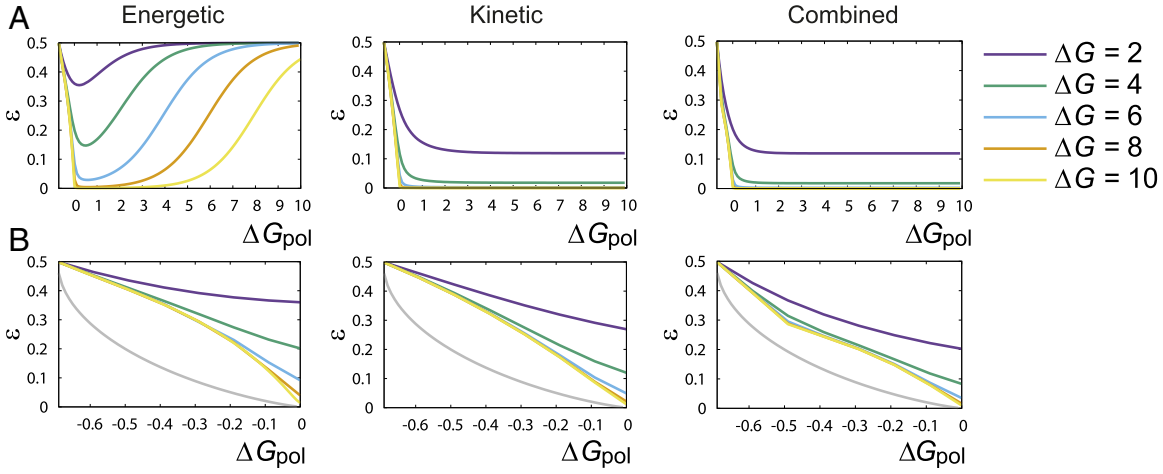

Fig. 3. Error probability ϵ as a function of $\Delta G_{\rm pol}$ for all three mechanisms: (*A*) over a wide range of $\Delta G_{\rm pol}$ and (*B*) within the entropy-driven region $\Delta G_{\rm pol} \leq 0$. The TT mechanism is always the least accurate, and the combined mechanism is the most accurate. All mechanisms have no accuracy in the limit of $\epsilon \to 0$, and are far from the fundamental bound on single-site accuracy imposed by $H_{\rm ss} = -\epsilon \ln \epsilon - (1-\epsilon)\ln(1-\epsilon) \geq -\Delta G_{\rm pol}$, except at $\Delta G_{\rm pol} \approx 0$.

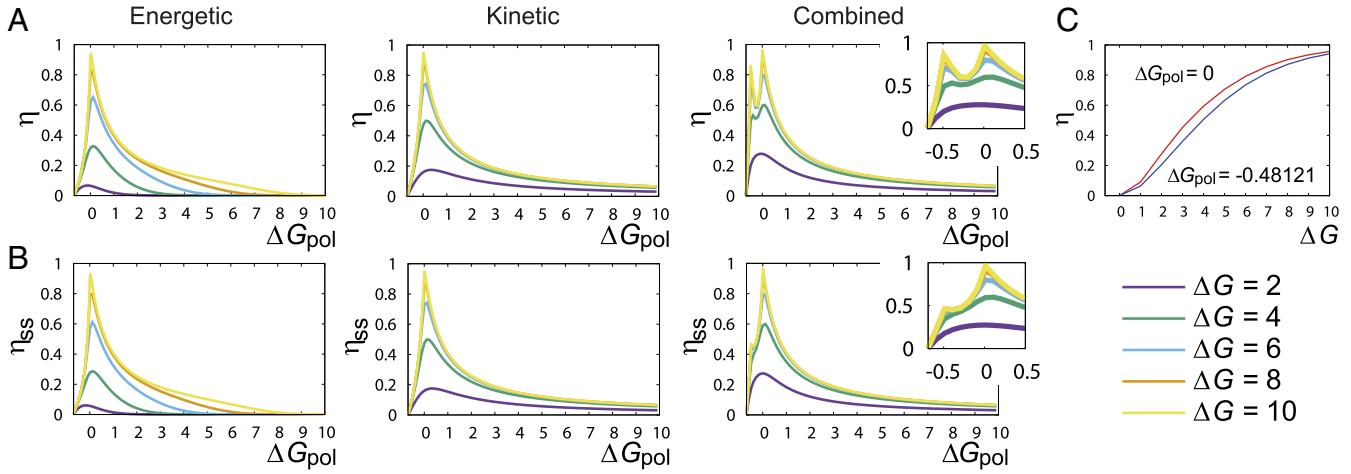

Fig. 4. Efficiencies (*A*) η_{ss} and (*B*) η plotted against ΔG_{pol} show sharp peaks at $\Delta G_{pol} = 0$ as $\Delta G \to \infty$ in all three cases. In the combined case, we see a second peak in η, and a shoulder in η_{ss} at $\Delta G_{pol} = -0.48121$. (*C*) Both of these peaks in η tend to unity at $\Delta G \to \infty$.

$\Delta G \to \infty$. Beyond this peak, η_{ss} drops as the stored free energy is bounded by $\ln 2$ per monomer. To understand the peak, note that, for every $\Delta G_{pol} \leq 0$, there is a hypothetical highest accuracy copy with ϵ fixed by $H_{ss} = -\Delta G_{pol}$ that is marginally thermodynamically permitted. However, this copy is not usually kinetically accessible. At $\Delta G_{pol} = 0$, the marginal copy has 100% accuracy, and, unusually, all three mechanisms can approach it kinetically, causing a peak. The efficiency approaches its limit of unity even for moderate values of ΔG. We note that, as $\Delta G \to \infty$, growth is slow for $\Delta G_{pol} \leq 0$: The total number of steps taken diverges.

A related argument explains the apparent cusp in the error ϵ for the combined mechanism at $\Delta G_{pol} \approx -0.48$ and high ΔG. On the plot of η_{ss} (Fig. 4 *A*), this cusp manifests as a shoulder. The full efficiency η (Fig. 4 *B*) has a prominent second peak. Uniquely, the combined mechanism's kinetics strongly disfavor chains of consecutive ws for high ΔG. A final copy with no consecutive ws has $\epsilon_w = 0$ but $\epsilon_r \neq 0$. Maximizing the entropy rate of such a Markov chain over ϵ_r gives $H_{max} = 0.48121$; $\Delta G_{pol} = -H_{max}$ matches the location of our peak/cusp. Thus, the combined mechanism initially eliminates consecutive ws, and, at $\Delta G_{pol} = -0.48121$, a state with $\eta_w = 0$ is thermodynamically permitted for the first time. For large ΔG, this polymer is kinetically accessible and grows with thermodynamic efficiency approaching unity (Fig. 4 *C*). In this limit, the overall entropy generation is zero.

The above behavior is a striking example of correlations being generated within the copy sequence, as well as with the template. Notably, while η approaches unity at this point, η_{ss} does not (Fig. 4). Correlations within the copy sequence limit the chemical work that can be devoted to improving the single-site accuracy of the copy polymer, since they prevent the bound $H_{ss} \geq H \geq -\Delta G_{pol}$ from being saturated.

Conclusion

The thermodynamic constraint on copying that underlies this work is deceptively simple: Unlike TSA, the overall chemical contribution to the free energy of a copy must be independent of the match between template and copy sequences. By studying the simplest mechanisms satisfying this constraint, however, we can draw important conclusions for copying mechanisms and thermodynamics more generally.

For copying, the most immediate contrast with previous work on TSA (20–25, 28) is that accuracy is necessarily zero when the copy assembles in equilibrium, since equilibrium correlations between physically separated polymers are impossible (1). Consequently, unlike self-assembly, no autonomous copying system can rely on relaxation to near-equilibrium; fundamentally different paradigms are required.

A direct result of the temporary nature of thermodynamic discrimination in persistent copying is that relying solely on the strong binding of correct copies is ineffective in ensuring accuracy. At $\Delta G = 6 k_B T$, comparable to the cost of a mismatched base pair (32), the TT discrimination model never improves upon $\epsilon = 0.0285$, which is more than 10 times the equilibrium error probability based on energetic discrimination obtainable in TSA, $1/(1+\exp(\Delta G))$. This performance would degrade further if many competing monomers were present. Mechanisms for copying must therefore be more carefully optimized than those for TSA. Either some degree of direct kinetic discrimination (with correct monomers incorporated more quickly), or, as an alternative, fuel-consuming proofreading cycles, appear necessary. It is well established that proofreading cycles can enhance discrimination above equilibrium in TSA (18, 20, 24), and the challenges in achieving direct kinetic discrimination in diffusion-influenced reactions via the details of the microscopic substeps may explain the ubiquity of such cycles in true copying systems.

Correlations within the copy, as well as between copy and template, arise naturally in persistent mechanisms. Indeed, in one case, pairs of mistakes are eliminated well before individual mistakes. These correlations contribute to the nonequilibrium free energy of the final state, reducing the single-site copy accuracy achievable for a given chemical work input. Biologically, however, it is arguably the accuracy of whole sequences, rather than individual monomers, that matters. In this case, positive correlations could advantageously increase the number of 100% correct copies for a given average error rate. It remains to be seen whether positive correlations, which may arise in real systems (33), can feasibly be used in this way. Regardless, we predict that within-copy correlations may be significant, particularly in simple systems with low accuracy. These correlations may change significantly if the requirement to remain bound exactly one bond, and exploiting correlations to extend functionality beyond simple copying is an intriguing prospect.

Relaxing this requirement also raises the possibility of early copy detachment. This risk is likely to be worse for genome replication than for transcription and translation, which may explain why the latter proceed by mechanisms analogous to our model, while DNA replication does not: Here, the copy of a single DNA strand from the double helix is first completely assembled on the template, and separated only much later at the next round of cell division.

Thermodynamically, a persistent copying mechanism converts the high free energy of the input molecules into a high free-energy copy state; we have defined a general efficiency of this free-energy transduction for copying. In typical physical systems with tight coupling of reactions, high efficiency occurs when the load is closely matched to the driving, either in autonomous systems operating near the stall force or in quasistatically manipulated systems. For the polymer copying mechanisms studied here, however, we find that the thermodynamic efficiency of information transfer, and not just the accuracy, approaches zero as polymer growth stalls. We predict that this result is general, since the alternative would require a sublinear convergence of the error rate on 50% as thermodynamic driving tends toward the stall point.

Fundamentally, the copy process transduces free energy into a complex system with many degrees of freedom (the sequence), and not just the polymer length. To be accurate, the sequence must be prevented from equilibrating. Thus, while weak driving leads to polymer growth with little overall entropy generation, it does a poor job of pushing the polymer sequence out of equilibrium. We predict that similar behavior will arise whenever an output must have a subset of its degrees of freedom out of equilibrium.

Away from the equilibrium limit, the efficiency shows one or more peaks as the polymerization free energy is varied. At these peaks, the system transitions between two nonequilibrium states with remarkably little dissipation. These particular values of ΔG_{pol} are sufficient to stabilize nonequilibrium distributions that happen to be especially kinetically accessible, rendering the true equilibrium particularly inaccessible. This alignment of kinetic and thermodynamic factors is most evident in the combined mechanism that efficiently produces a state with few adjacent mismatches at $\Delta G_{\mathrm{pol}} = -0.48121$. These results slightly qualify the prediction of ref. 1 that accurate copying is necessarily entropy-generating, since entropy generation can be made arbitrarily small, while retaining finite copy accuracy, by taking $\Delta G \to \infty$ at these specific values of ΔG_{pol}.

The behavior of the efficiency in these models emphasizes the importance, in both natural and synthetic copying systems, of kinetically preventing equilibration. Our work emphasizes that this paradigm should be applied not only to highly evolved systems with kinetic proofreading mechanisms (17) but also the most basic mechanisms imaginable.

Extending our analysis to consider fuel-consuming proofreading cycles would be natural. However, these cycles will not change the fundamental result that the entropy of the copy sequence is thermodynamically constrained, in this case by $H_{\mathrm{ss}} \geq H \geq -\Delta G_{\mathrm{pol}} - \Delta G_{\mathrm{fuel}}$, where the final term is the additional free energy expended per step to drive the system around proofreading cycles. We predict that nonequilibrium proofreading cycles, by their very nature, are unlikely to approach efficiencies of unity.

ACKNOWLEDGMENTS. We thank Charles Bennett for instructive conversations. T.E.O. is supported by the Royal Society, and P.R.t.W. is supported by the Netherlands Organization for Scientific Research.

1. Ouldrige TE, ten Wolde PR (2017) Fundamental costs in the production and destruction of persistent polymer copies. *Phys Rev Lett* 118:158103.
2. Ouldridge TE, Govern CC, ten Wolde PR (2017) Thermodynamics of computational copying in biochemical systems. *Phys Rev X* 7:021004.
3. Tjivikua T, Ballester P, Rebek J, Jr (1990) Self-replicating system. *J Am Phys Soc* 112:1249–1250.
4. Feng Q, Park TK, Rebek J (1992) Crossover reactions between synthetic replicators yield active and inactive recombinants. *Science* 256:1179–1180.
5. Vidonne A, Douglas P (2009) Making molecules make themselves–The chemistry of artificial replicators. *Eur J Org Chem* 2009:593–610.
6. Orgel LE (1992) Molecular replication. *Nature* 358:203–209.
7. Colomb-Delsuc M, Mattia E, Sadownik JW, Otto S (2015) Exponential self-replication enabled through a fibre elongation/breakage mechanism. *Nat Commun* 6: 7427.
8. Sievers D, Von Kiedrowski G (1994) Self-replication of complementary nucleotide-based oligomers. *Nature* 369:221–224.
9. Lincoln TA, Joyce GF (2009) Self-sustained replication of an RNA enzyme. *Science* 323:1229–1232.
10. Wu N, Willner I (2016) ph-stimulated reconfiguration and structural isomerization of origami dimer and trimer systems. *Nano Lett* 16:6650–6655.
11. Wang T, et al. (2011) Self-replication of information-bearing nanoscale patterns. *Nature* 478:225–228.
12. Li T, Nicolaou KC (1994) Chemical self-replication of palindromic duplex DNA. *Nature* 369:218–221.
13. Mast CB, Braun D (2010) Thermal trap for DNA replication. *Phys Rev Lett* 104: 188102.
14. Orgel LE (1998) The origin of life—A review of facts and speculations. *Trends Biochem Sci* 23:491–495.
15. Martin W, Baross J, Kelley D, Russell MJ (2008) Hydrothermal vents and the origin of life. *Nat Rev Microbiol* 6:805–814.
16. Schulman R, Bernard Y, Winfree E (2012) Robust self-replication of combinatorial information via crystal growth and scission. *Proc Natl Acad Sci USA* 109:6405–6410.
17. Hopfield JJ (1974) Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci USA* 71:4135–4139.
18. Ehrenberg M, Blomberg C (1980) Thermodynamic constraints on kinetic proofreading in biosynthetic pathways. *Biophys J* 31:333–358.
19. Johansson M, Lovmar M, Ehrenberg M (2008) Rate and accuracy of bacterial protein synthesis revisited. *Curr Opin Microbiol* 11:141–147.
20. Bennett CH (1979) Dissipation-error tradeoff in proofreading. *Biosystems* 11:85–91.
21. Cady F, Qian H (2009) Open-system thermodynamic analysis of DNA polymerase fidelity. *Phys Biol* 6:036011.
22. Andrieux D, Gaspard P (2008) Nonequilibrium generation of information in copolymerization processes. *Proc Natl Acad Sci USA* 105:9516–9521.
23. Sartori P, Pigolotti S (2013) Kinetic versus energetic discrimination in biological copying. *Phys Rev Lett* 110:188101.
24. Sartori P, Pigolotti S (2015) Thermodynamics of error correction. *Phys Rev X* 5:041039.
25. Esposito M, Lindenberg K, Van den Broeck C (2010) Extracting chemical energy by growing disorder: Efficiency at maximum power. *J Stat Mech Theory Exp* 2010: P01008.
26. Whitelam S, Schulman R, Hedges L (Dec 2012) Self-assembly of multicomponent structures in and out of equilibrium. *Phys Rev Lett* 109:265506.
27. Nguyen M, Vaikuntanathan S (2016) Design principles for nonequilibrium self-assembly. *Proc Nat Acad Sci*, 113:14231–14236.
28. Gaspard P, Andrieux D (2014) Kinetics and thermodynamics of first-order Markov chain copolymerization. *J Chem Phys* 141:044908.
29. Wachtel A, Rao R, Esposito M (2018) Thermodynamically consistent coarse graining of biocatalysts beyond Michaelis–Menten. *New J Phys* 20:042002.
30. Boyd AB, Mandal D, Crutchfield JP (2016) Identifying functional thermodynamics in autonomous Maxwellian ratchets. *New J Phys* 18:023049.
31. Cover TM, Thomas JA (2012) *Elements of Information Theory* (John Wiley, New York).
32. SantaLucia J, Jr, Hicks D (2004) The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct* 33:415–440.
33. Rao R, Peliti L (2015) Thermodynamics of accuracy in kinetic proofreading: Dissipation and efficiency trade-offs. *J Stat Mech Theor Exp* 2015:P06001.