

Simone Pigolotti · Pablo Sartori

Protocols for copying and proofreading in template-assisted polymerization

Received: date / Accepted: date

Abstract We discuss how information encoded in a template polymer can be stochastically copied into a copy polymer. We consider four different stochastic copy protocols of increasing complexity, inspired by building blocks of the mRNA translation pathway. In the first protocol, monomer incorporation occurs in a single stochastic transition. We then move to a more elaborate protocol in which an intermediate step can be used for error correction. Finally, we discuss the operating regimes of two kinetic proofreading protocols: one in which proofreading acts from the final copying step, and one in which it acts from an intermediate step. We review known results for these models and, in some cases, extend them to analyze all possible combinations of energetic and kinetic discrimination. We show that, in each of these protocols, only a limited number of these combinations leads to an improvement of the overall copying accuracy.

Keywords: kinetic proofreading; biological copying; polymerization, speed-accuracy trade-off.

S. Pigolotti
Universitat Politècnica de Catalunya,
Edif. GAIA, Rambla Sant Nebridi 22,
08222 Terrassa, Barcelona, Spain.
Tel.: +34 93 739 8573
Fax: +34 93 739 8000
E-mail: simone.pigolotti@upc.edu

P. Sartori
Max Planck Institute for the Physics of Complex Systems.
Noethnitzer Strasse 38 , 01187, Dresden, Germany.

1 Introduction

Copying of information constitutes one of the key operations that biological systems need to perform. Information is cyclically and reliably transferred from a template to a copy during DNA duplication, its transcription into mRNA, and also during protein synthesis. In all these examples, copies must be accurate reproductions of the template, as copy errors may compromise biological functionality.

Accuracy of copying is easy to estimate close to thermodynamic equilibrium. Because of detailed balance, the probability of a copy error η is $\eta_{\text{eq}} \sim \exp[-\beta(\Delta E^w - \Delta E^r)]$, where $\Delta E^w - \Delta E^r$ is the energy difference between a right and a wrong incorporated monomer and $\beta = (k_B T)^{-1}$ is the inverse thermal energy. Already in the 50s, before detailed molecular studies of copying reactions were possible, Pauling pointed out that structural differences between some amino acids are small, so that their binding energy difference can not possibly account for the accuracy required in protein synthesis [25]. Physically, this implies that high accuracy can only be achieved by producing the copies via a far-from-equilibrium thermodynamic process.

More recent experimental studies discovered that, indeed, most biological copying reactions are composed of many out-of-equilibrium intermediate steps, that can be identified in detail in some cases. An important hallmark in the understanding of such complex copying processes was the proposal by Hopfield [13] and Ninio [22] of the mechanism of *kinetic proofreading*. In kinetic proofreading, a discriminatory enzymatic reaction is complemented by an irreversible, non-discriminating step that brings the complex to a high-energy state. Kinetic proofreading was shown to reduce the minimum error down to the *square* of the equilibrium error, and further down to higher powers if multiple proofreading steps are considered [9]. This idea had a deep impact as a paradigm of error correction in biochemical reactions and helped understanding the mechanisms underlying the high fidelity of many copying systems, including DNA duplication [17], translation [17], and other cellular tasks requiring high specificity such as T-cells recognition [18, 8]. Recently, a generalization of kinetic proofreading to arbitrarily wired networks has been explored by Murugan, Huse and Leibler [20, 21]

A parallel approach to biological copying was pioneered by Charles Bennett, when he proposed a model in which a copy polymer is elongated to match a template [3] (see also section 3). An important conceptual aspect of Bennett's model is that it describes the copy of a whole polymer, i.e. a process in which a large number of monomers are sequentially copied. This is in contrast to stoichiometric models, such as Hopfield's proofreading model, which describe copy of a single monomer. In Bennett's model, each copy event is simply modeled as a single-step process, and energy differences between right and wrong matches are neglected. This simplicity, together with the steady-state nature of the solution, made the model particularly appropriate to analyze the issue of irreversibility of the copying process. However, Bennett's work received relatively little attention until recently, when a renewed interest in the physics of biological information processing has led to further analyses in this direction [1, 7].

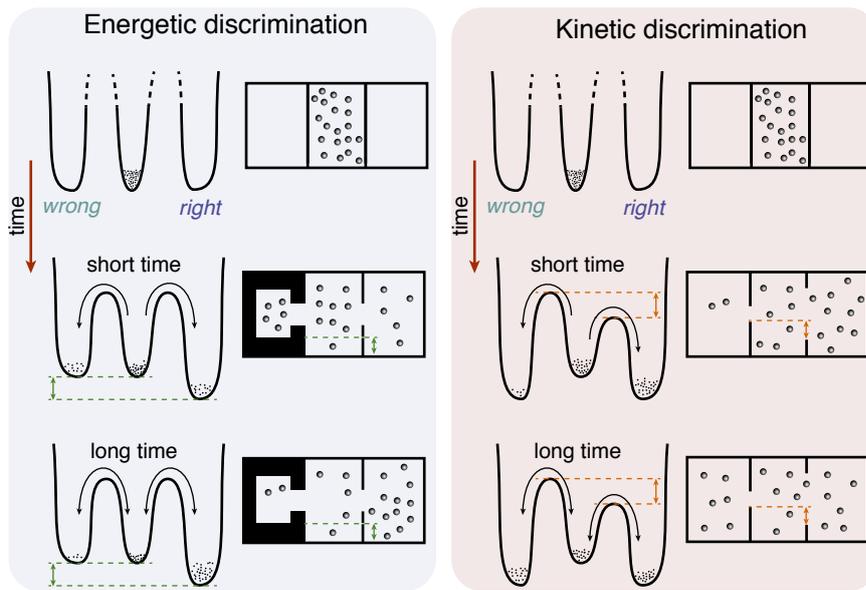


Fig. 1 Energetic and Kinetic discrimination. In both panels (Energetic and Kinetic discrimination), the first column shows an energy landscape for a copy process. In particular, in the central (null) state, no monomer is bound to the copying machine. The left and right states represent respectively a bound monomer which do not match (wrong) or do match (right) the template. The second column presents an analogy in terms of a gas of diffusing particles, where the three states correspond to different chambers. In *Energetic discrimination*, the wrong and right states have different energy, while the kinetic barriers to reach them from the null state are the same. In the gas analogy, the openings leading to the right/wrong chambers have the same width but the two chambers have different volumes. If the system is prepared in the null state, at short times the probabilities of being in the right and wrong states are dictated by the energy barriers (opening widths) and are therefore approximately equal. At longer times, the system equilibrates and the error reduces the energy differences (chamber sizes). In *Kinetic discrimination*, the wrong and right states have the same energy, while the kinetic barriers to reach them from the null state differ. In the gas analogy, the openings leading to the right/wrong chambers have different widths but the two chambers have equal volumes. Preparing the system in the null state, at short times the probabilities of being in the right and wrong states are determined by the energy barriers (opening widths), leading to a low error probability. At longer times, the system equilibrates, so that the probability of being in the right/wrong states are the same.

To understand a fundamental difference between equilibrium and non-equilibrium copying, it is useful to introduce the concepts of *energetic* and *kinetic* discrimination, see Fig. (1). In both panels, these two concepts are illustrated by means of two different conceptual examples. The first is an energy landscape for a simple three-state copying model. The three states represent, from left to right, the state in which a wrong monomer is bound, the one in which no monomer is bound, and the one in which the right monomer is bound. The second example is a mechanical model in which the system is a gas of diffusing particles, where the three chambers are the

analogous of the three states of the first example. In all cases, the system is initially prepared in the null central state, top panes in Fig. (1). We are interested in the evolution of the error, defined as the ratio of probabilities of being in the wrong vs. right state as a function of time.

In energetic discrimination, the wrong and right states are separated by the same kinetic barrier from the null state; however, the wrong and right states have different energies. In the gas analogy, the openings to reach the right and wrong chambers have the same width, but the two chambers have different volumes. At short times, the error is mostly determined by the kinetic barriers and is therefore large. At longer times, the system equilibrates leading to a reduction of the error. Conversely, in kinetic discrimination, the kinetic barrier to reach the wrong state is higher than that to reach the right state, but the right and wrong states are isoenergetic. In the mechanical analogy, the openings leading to the right and wrong chambers have different widths, but the two chambers have the same volume. At short times, the error is determined by the kinetic barrier difference and therefore is low. At longer times, the system equilibrates and the error grows due to the lack of energy differences.

We recently discussed how kinetic and energetic discrimination are mutually exclusive in single-step protocols [27]: it is impossible to improve accuracy at intermediate times by combining their effects. Indeed, most models in the literature consider only one form of discrimination at a time: for example, Hopfield’s model [13] does not consider kinetic discrimination, while Bennett’s model [3] does not consider energetic differences.

In this paper, we discuss how kinetic and energetic discrimination in different steps determine the minimum error in multi-step copy protocols. To this aim, we both present a systematic review of existing results and generalize some of them, for example by systematically studying all possible discriminating regimes of proofreading. We begin by reviewing the main features of Bennett’s copolymerization model [3] in the general case in which both energetic and kinetic discrimination are present [27], see Fig. 2A. We then generalize Bennett’s model, which can be seen as the simplest possible copy protocol, [3] to the biologically relevant cases of multi-step protocols (see also [28]). In particular, our second copy protocol is made up of a linear chain of two states, Fig. 2B, while the third copy protocol, Fig. 2C, combines a linear chain of two states with a proofreading reaction acting from the last state. The last model, proofreading/accomodation, is a variation of model C in which proofreading acts from the intermediate state, see Fig. 2D. We shall see how this small variants changes quite dramatically the dependence of the error on the kinetic barriers of the three reactions. All these protocols can be seen as “building blocks” of the mRNA translation pathway, as represented in Fig. 2. By analytically solving these models we find that, at given energy landscape, accurate copying can be achieved only in a limited number of regimes, that we fully characterize.

The paper is organized as follows. Section 2 introduces our modeling framework, which has already been used in [28] to study the non-equilibrium thermodynamics of copying. In section 3, we present the analytical solution for protocol A in Fig. 2. We study in particular the different regimes in which

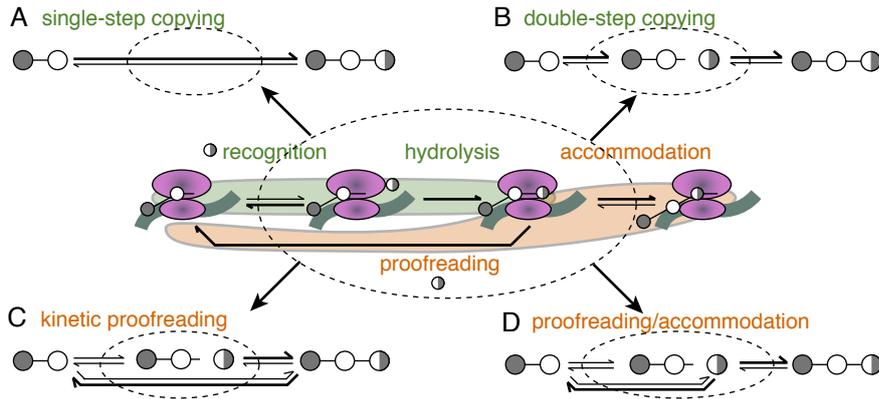


Fig. 2 Protocols of template-assisted polymerization. Panels A - D show the different copy protocols that we consider in this paper. **A:** single-step copolymerization without intermediate state (equivalent to Bennett’s model [3]). **B:** double-step copying protocol. This can be seen as a variant of Bennett’s model with the addition of an intermediate state. **C:** kinetic proofreading and **D:** proofreading/accommodation. The central panel is a schematic of mRNA translation by the ribosome, and illustrates how models A-D can be seen as building blocks of this complex biological copying scheme, see e.g [14, 15, 32]. In all panels, full gray and full white circles represent two different kinds of monomers. Half gray/half white circles represent monomers being either gray or white.

error correction can be achieved. In section 4, we show how to solve more complex template-assisted polymerization models with intermediate states. Sections 5, 6, and 7 present the solution and analysis of the double-step copying, proofreading, and proofreading/accommodation protocols respectively. Section 8 is devoted to conclusions and future perspectives.

2 General model of template assisted polymerization

Our modeling framework describes cyclic stochastic copying of a long polymer. The system is made up of three parts: a pre-existing long template polymer, which is not modified during the dynamics; a copy polymer, which is assembled during the dynamics; and a copy machine, which assists growth of the copy polymer and, in particular, tries to match monomers incorporated in the copy with those in the template (see Fig. 3A). For simplicity, we consider the monomer pool to be an infinite reservoir.

We focus on the case in which the physics does not depend on the type of monomer, but only on whether it is a right match r (black-black or white-white in Fig. 3A) or a wrong match w (black-white or white-black). The states of the system and the network of possible stochastic transitions among them is then that in Fig. 3B. After finalizing incorporation of a monomer, the system is in a *main states*. Main states are represented in the figure as rectangles with a string of r/w inside. To process a candidate monomer, the machine performs transitions through an arbitrarily wired network of *intermediate*

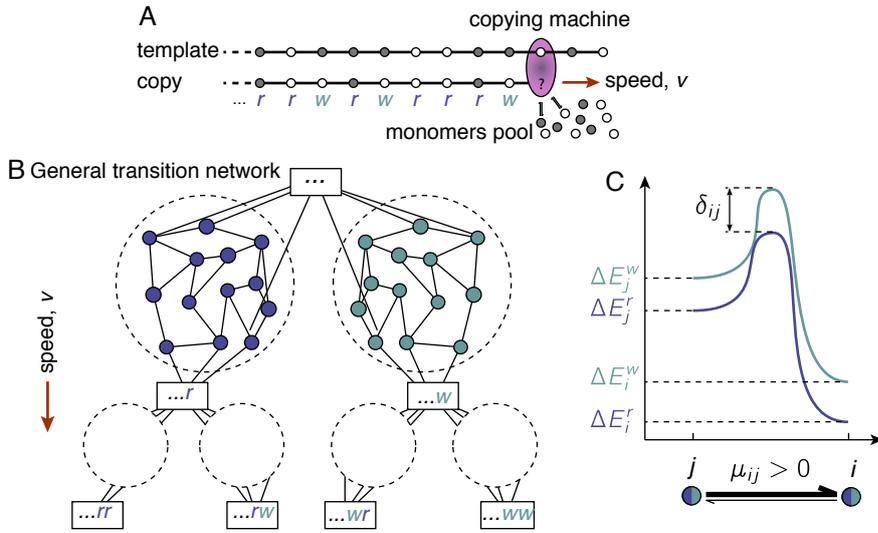


Fig. 3 Schematic summary of template-assisted polymerization. A. Sketch of the model in physical space. A molecular copying machine aids elongation of a copy polymer, by trying to match candidate monomers with those on a pre-existing template polymer. **B.** Transition network: rectangles represent the *main states*, and circles the *intermediate states* leading to monomer incorporation. **C.** Energy landscape corresponding to each transition, see Eq. (1). State i can be either a main state (if $i = 0$ or $i = n + 1$) or an intermediate state (if $0 < i \leq n$) and the same for j .

states. These represent different conformational states of the copying machine, and allow for complex error-correcting protocols. Intermediate states are represented as green and blue circles in Fig. 3B, depending on whether the candidate monomer corresponds to a wrong or right match. Large dashed circles enclose each sub-network leading to incorporation/rejection of the candidate monomer.

Let us focus on a specific copy protocol, specified by the topology of the network of its n intermediate states. We denote with $\dots rwr_i$ the i -th intermediate state leading to incorporation of a right monomer and analogously for a wrong one, where the preceding string of r/w represents the already incorporated monomers. It is convenient to identify, with a slight abuse of notation, $i, j = 0$ with the preceding main state and $i, j = n + 1$ with the main state in which the monomer is incorporated, i.e. $\dots rwr_0 \equiv \dots rw$ and $\dots rwr_{n+1} \equiv \dots rwr$. Irrespective of the existing r/w string, transitions from state j to state i occur at rate k_{ij}^r and k_{ij}^w for right and wrong candidates respectively. To better understand the physics of the copying process, it is useful to think that the stochastic transition rates are determined by a) the energy of the states, b) chemical driving potentially fueling the different reactions, and c) activation barriers separating the different states, as shown in Fig. 3C. In terms of these quantities, the rates can be expressed as

$$\begin{aligned}
k_{ij}^r &= w_{ij} \exp[(\Delta E_j^r + \mu_{ij} + \delta_{ij})/T] \\
k_{ij}^w &= w_{ij} \exp[(\Delta E_j^w + \mu_{ij})/T] \\
k_{ji}^r &= w_{ij} \exp[(\Delta E_i^r + \delta_{ij})/T] \\
k_{ji}^w &= w_{ij} \exp[(\Delta E_i^w)/T].
\end{aligned} \tag{1}$$

In Eqs. (1), ΔE_j^r is the energy of right state j , and similarly for wrong matches. We take as reference the initial main state, so that $\Delta E_0^w = \Delta E_0^r = 0$. T is the temperature, measured from now on in units of the Boltzmann constant ($k_B = 1$). The parameter δ_{ij} accounts for a difference in the activation barrier between state j and i for right and wrong monomers. In particular, a positive value of δ_{ij} means that the corresponding transition is kinetically facilitated for right monomers with respect to wrong ones. The chemical driving $\mu_{ij} > 0$ favors the reaction $j \rightarrow i$, irrespective of whether the match is right or wrong. Finally, ω_{ij} is the inverse characteristic time of the transition between states j and i . The convention of Eqs. (1) differs slightly from that used in [27] and, in particular, simplifies picturing the energy landscape in multi-step copy protocols. Note that, in this case, the energies do not affect only the backward rates but potentially also the forward rates and thus kinetic discrimination.

The main focus of this paper is to study error-correcting regimes by computing the error rate η and speed v for different specific copy protocols. In all cases, we shall assume that the parameters characterizing the difference between right and wrong monomers (i.e. the ΔE_i^r , ΔE_i^w , and δ_{ij}) are given, while the chemical drivings μ_{ij} and the inverse timescales ω_{ij} can be tuned to optimize the error.

3 Protocol A: single-step copolymerization

We begin by considering a protocol in which copying occurs without intermediate states. Although this case has been studied in a number of works (see e.g. [3,1,6,7,27]), it is useful to review the main results in order to introduce the basic concepts that we will apply to more complex protocols.

The master equations governing this polymerization process can be written by separating main states ending with a right incorporated monomer from those ending with a wrong one:

$$\begin{aligned}
\dot{P}(\dots r) &= k_{10}^r P(\dots) + k_{01}^r P(\dots rr) + k_{01}^w P(\dots rw) \\
&\quad - (k_{10}^r + k_{10}^w + k_{01}^r) P(\dots r) \quad , \\
\dot{P}(\dots w) &= k_{10}^w P(\dots) + k_{01}^r P(\dots wr) + k_{01}^w P(\dots ww) \\
&\quad - (k_{10}^r + k_{10}^w + k_{01}^w) P(\dots w) \quad .
\end{aligned} \tag{2}$$

The system of equations (2) can be solved at steady state by using an *ansatz* of uncorrelated errors. This amounts to assume that the steady-state probability of an arbitrary template-copy sequence containing N^r right matches and N^w wrong matches is given by

$$P(\dots) \propto (1 - \eta)^{N^r} \eta^{N^w} \quad , \tag{3}$$

where η is the error rate to be determined *a posteriori*. Substituting (3) in the master equations (2) and imposing steady-state leads to a condition of the form $\eta/(1-\eta) = v^w(\eta)/v^r(\eta)$, where $v^w(\eta) = k_{10}^w - k_{01}^w\eta$ and $v^r(\eta) = k_{10}^r - k_{01}^r(1-\eta)$ are the average incorporation speeds of wrong and right monomers: the stationary solution represents a non-equilibrium steady state characterized by a net growth speed (see e.g [4]). Substituting the expressions of the rates given by Eq. (1) and Fig. 4A finally yields

$$\frac{\eta}{1-\eta} = e^{-\delta_{10}/T} \frac{e^{\mu_{10}/T} - \eta e^{\Delta E_1^w/T}}{e^{\mu_{10}/T} - (1-\eta)e^{\Delta E_1^r/T}}. \quad (4)$$

Once the error is known, this speed is simply given by $v = \eta v^w + (1-\eta)v^r$. By rearranging terms, Eq. (4) becomes a second-order equation for the error η that can be easily solved. It is instructive to discuss the properties of this solution in some depth and, in particular, how varying the chemical driving μ_{10} affects the error rate.

A copy is performed quasi-statically when the growth speed vanishes, $v = 0$. This requires that $v^w = 0$ and $v^r = 0$, i.e. both numerator and denominator of Eq. (4) vanish. Imposing this condition, one finds that the error reaches its equilibrium value $\eta_{\text{eq}} = 1/(1 + e^{(\Delta E_1^w - \Delta E_1^r)/T})$. To reach this limit, the chemical driving μ_{10} must be tuned to its stall value $\mu_{\text{st}} \equiv -T \log(e^{-\Delta E_1^w/T} + e^{-\Delta E_1^r/T})$. As originally observed by Bennett, the stall chemical driving is *negative*, since to arrest polymerization it needs to oppose the entropic force driving growth of the chain [3, 1, 7, 27]. This entropic force can be physically understood by thinking that the number of possible states grows exponentially as the copy polymer elongates. In the following, we shall focus only on the case in which the copying speed is non-negative, $\mu_{10} \geq \mu_{\text{st}}$ (see [2] for considerations on the case of negative speed).

Upon increasing the chemical driving from its stall value, the speed v increases monotonically. The error approaches, also monotonically, its large-driving irreversible limit $\eta_{\text{irr}} = e^{-\delta_{10}/T}/(1 + e^{-\delta_{10}/T})$. This implies that the error is an increasing or decreasing function of the chemical driving depending on whether η_{irr} is larger or smaller than η_{eq} , respectively.

This finding is summarized in Fig. 4B, where we represent the minimum possible error η_{min} as a function of the energy difference $\Delta E_1^w - \Delta E_1^r$ and the kinetic barrier δ_{10} . Depending on which is the larger between these two quantities, we can distinguish between two different regions in parameter space:

- **Kinetic region.** This region is defined by the condition $\delta_{10} > \Delta E_1^w - \Delta E_1^r$ in Fig. 4B, which implies $\eta_{\text{irr}} < \eta_{\text{eq}}$. In this region, the minimum error is equal to η_{irr} , and is achieved in the limit of a very large chemical driving $\mu_{10} \gg \mu_{\text{st}}$.
- **Energetic region.** This is the region $\delta_{10} < \Delta E_1^w - \Delta E_1^r$. The minimum error is the equilibrium error η_{eq} , which depends only on the energy difference $\Delta E_1^w - \Delta E_1^r$. The minimal error is achieved when the chemical driving tends to its stall value μ_{st} .

Symmetric points on different sides of the diagonal in Fig. 4B are characterized by the same minimum error. However, being in one or the other region

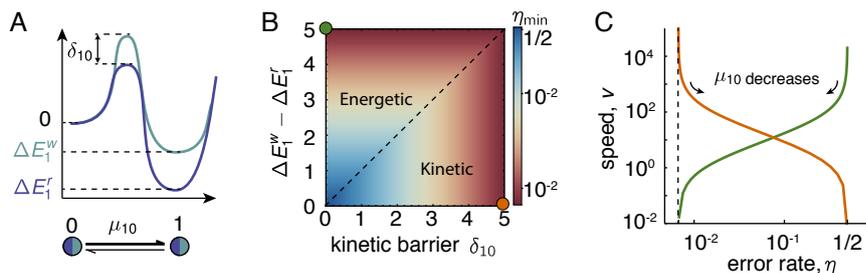


Fig. 4 Single-step protocol. **A.** Energy diagram . The dark blue curve corresponds to the landscape for a right match, and light blue for a wrong one. Below the states topology. **B.** Heat map of the minimum error η_{\min} as a function of the kinetic barrier δ_{10} and the energy difference $\Delta E_1^w - \Delta E_1^r$. A dashed line distinguishes the kinetic from the energetic regions. **C.** Speed-error tradeoffs in the energetic and kinetic regions (see Eq. 17 in the Appendix for the expression of the speed, and circles in panel B for values). The speed is measured in units of ω_{10} .

affects the way the system responds to a change in the chemical driving. On the other hand, increasing the chemical driving always increases the copying speed. This means that the error is a decreasing function of the chemical driving and the copying speed in the kinetic region, and an increasing function in the energetic region. These two opposite tradeoffs are shown in Fig. 4C.

4 Solution to a general protocol

The approach used in the previous section can be generalized for arbitrarily complex models [28]. We now discuss how to find the steady-state solution for a general protocol including intermediate states.

We start by writing the general master equations. It is convenient to define the probability fluxes $\mathcal{J}_{ij}^r(\dots) = k_{ij}^r P(\dots r_j) - k_{ji}^r P(\dots r_i)$ between pairs of states. The master equations for the intermediate states then read

$$\dot{P}(\dots r_i) = \sum_{j=0}^{n+1} \mathcal{J}_{ij}^r(\dots) \quad \text{and} \quad \dot{P}(\dots w_i) = \sum_{j=0}^{n+1} \mathcal{J}_{ij}^w(\dots) \quad , \quad (5)$$

while the master equations for the main states are

$$\begin{aligned} \dot{P}(\dots w) &= \sum_{j=0}^{n+1} [\mathcal{J}_{n+1j}^w(\dots) - \mathcal{J}_{j0}^r(\dots w) - \mathcal{J}_{j0}^w(\dots w)] \quad , \\ \dot{P}(\dots r) &= \sum_{j=0}^{n+1} [\mathcal{J}_{n+1j}^r(\dots) - \mathcal{J}_{j0}^r(\dots r) - \mathcal{J}_{j0}^w(\dots r)] \quad . \end{aligned} \quad (6)$$

Note that Eqs. (6) correctly reduce to Eqs. (2) in the case of $n = 0$.

The general system of equations above can be solved at steady state for any choice of the copy protocol. A solution can be obtained also in this case

by assuming uncorrelated errors, as in Eq. (3). However, in this case, we have to further assume that steady-state probabilities of intermediate states are proportional to the probability of the main state before attempted monomer incorporation:

$$P(\dots r_i) = P(\dots) p_i^r \quad , \quad (7)$$

where we introduced the *occupancies* p_i^r , and similarly for wrong monomers. Note that Eq. (3) implies $p_0^r = p_0^w = 1$, $p_{n+1}^r = (1 - \eta)$ and $p_{n+1}^w = \eta$.

For any given network of intermediate states, one can impose the steady-state condition and the solution *ansatz* in Eqs. (3) and (7). Doing so, Eq. (5) becomes a linear system of equations for the occupancies of the intermediate states. Solving this linear systems yields explicit expressions for the occupancies in terms of the transition rates and the yet unknown error rate η . Substituting the occupancies into Eqs. (6) results into equations formally identical to Eqs. (2), with “renormalized” rates of incorporation/removal of right/wrong monomers that take into account the effect of the intermediate states. In short, after having performed steady-state elimination of the intermediate states, any arbitrarily complex template-assisted polymerization protocol can be formally reduced to the simple model without intermediate states described in section 3. As a consequence, the equation for the error rate η can be always put in the form

$$\frac{\eta}{1 - \eta} = \frac{v^w(\eta)}{v^r(\eta)} \quad , \quad (8)$$

where, thanks to the steady-state elimination of the occupancies, $v^w(\eta)$ and $v^r(\eta)$ can be explicitly expressed in terms of the rates as in the case of Eq. (4). Also in this case, Eq. 8 is an algebraic equation for the only unknown η and its solution completes the solution of the model. We stress once more that, while in the following we will focus on simple protocols with only one intermediate state, the solution strategy described in this section is general: the difficulties of solving larger networks of intermediate states are only of algebraic nature.

5 Protocol B: double-step copying

The second-simplest possible copy protocol is a linear chain of states where final monomer accommodation is preceded by an intermediate state, see Fig. 2. In the most general case both steps can discriminate kinetically and energetically, and the energy landscape is that shown in Fig. 5A. Following the general solution strategy described in previous subsection we can determine the error rate η and speed v of this complex copying pathway. Solving at steady-state Eqs. (5) gives the occupancies of the intermediate states, $p_1^r = [k_{10}^r + (1 - \eta)k_{12}^r]/(k_{01}^r + k_{21}^r)$ and $p_1^w = (k_{10}^w + \eta k_{12}^w)/(k_{01}^w + k_{21}^w)$. Using further the expressions for v^w and v^r (see Appendix), the equation for the error rate can be written as

$$\frac{\eta}{1 - \eta} = \frac{p_1^w k_{21}^w - k_{12}^w \eta}{p_1^r k_{21}^r - k_{12}^r (1 - \eta)} \quad . \quad (9)$$

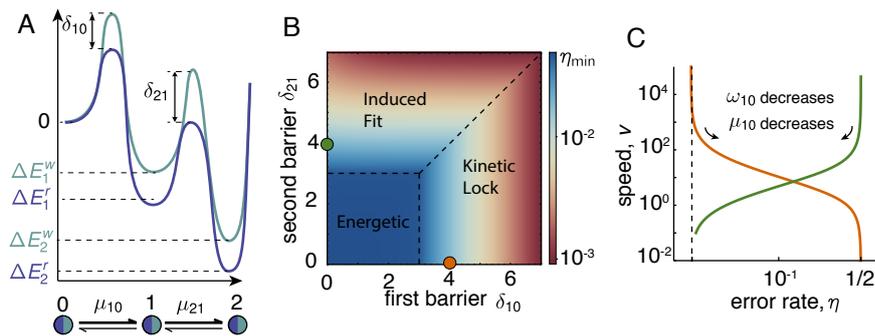


Fig. 5 Double-step copying. **A.** Energy landscape showing kinetic and energetic differences in both steps. **B.** Minimum error and regions. Parameters are $\Delta E_2^w = \Delta E_1^r = 0$, $\Delta E_2^r = \Delta E_1^w = 3T$. The chemical drivings and the relative time scales of the different reactions have been determined by numerically minimizing the error. **C.** Speed-error tradeoff for points in the Induced fit and kinetic lock regimes, see B. Speed is in units of ω_{10} .

Substituting the expressions of the occupancies as well as the parametrization of the rates yields

$$\frac{\eta}{1-\eta} = e^{-\frac{(\delta_{10}+\delta_{21})}{T}} \frac{\left[e^{\frac{\delta_{10}}{T}} + r e^{\frac{\delta_{21}+\mu_{21}}{T}} \right]}{\left(1 + r e^{\frac{\mu_{21}}{T}} \right)} \frac{\left(e^{\frac{\mu_{10}+\mu_{21}}{T}} - \eta e^{\frac{\Delta E_2^w}{T}} \right)}{\left[e^{\frac{\mu_{10}+\mu_{21}}{T}} - (1-\eta) e^{\frac{\Delta E_2^r}{T}} \right]} \quad (10)$$

where we introduced the dimensionless ratio of time scales of the two reactions $r = \omega_{21}/\omega_{10}$.

The above equation allows to calculate the error rate η given the discrimination barriers δ_{10} and δ_{21} , the energy differences ΔE_2^w and ΔE_2^r , the chemical drivings μ_{10} and μ_{21} , and the ratio of rates r . In this case, optimizing the error at fixed energy landscape and discrimination barriers is more subtle, as there are three free parameters in the system: the two chemical drivings and the time scale ratio. By taking different limits, we can identify three regions in parameter space:

- **Energetic region.** This region is defined by $\Delta E_2^w - \Delta E_2^r > \delta_{10}$ and $\Delta E_2^w - \Delta E_2^r > \delta_{21}$. In this region, the equilibrium error $\eta_{\text{eq}} = e^{-\Delta E_2^w/T} / (e^{-\Delta E_2^w/T} + e^{\Delta E_2^r/T})$ is the minimal error of the system. It is approached on the quasi-static limit, as can be derived from Eq. (10) by taking the limit of vanishing speed. Also in this case, to achieve equilibrium, the total chemical driving $\mu_{10} + \mu_{21}$ must be negative and equal to the stall value μ_{st} .
- **Kinetic Lock.** This is the region $\delta_{10} > \Delta E_2^w - \Delta E_2^r$ and $\delta_{10} > \delta_{21}$. The minimum error is equal to η_{irr} defined in the previous section. It is achieved in a strongly-driven regime in which $\mu_{10}, \mu_{21} \rightarrow \infty$. Note that, since the dynamics is completely irreversible, only the first barrier influences the error. The system is thus kinetically locked after the first step.

- **Induced fit.** In this region the parameters satisfy $\delta_{21} > \Delta E_2^w - \Delta E_2^r$ and $\delta_{21} > \delta_{10}$. The minimum error rate is $\eta \approx e^{-\delta_{21}/T}$. It can be achieved by taking first the limit $r \rightarrow 0$ and then $\mu_{10}, \mu_{21} \rightarrow \infty$. We term this regime “induced fit” as it is reminiscent of the mechanism in enzyme kinetics where discrimination occurs via a kinetic mechanism acting after initial binding, see e.g. [23].

Summarizing, copies can be produced near equilibrium, by using an energy difference; or out of equilibrium, by using either the first or the second kinetic barrier. Thus, also in the double-step copying protocol energetic and kinetic discrimination strategies can not be combined. Moreover, it is impossible to exploit both kinetic barriers: in particular, if the first reaction is strongly driven, no mechanism acting afterwards can alter the overall accuracy.

These three regions are shown in Fig. 5B, where we numerically minimized the error at fixed values of the energies parametrically varying the two barrier differences δ_{10} and δ_{21} . Finally, due to the complex dependence of the error on parameters, also in this case one can find negative and positive speed-error relations depending on which parameter is altered to tune the error and in which region the model is operating, see example in Fig. 5C.

6 Protocol C: Kinetic Proofreading

We now discuss a proofreading protocol in the spirit of Hopfield’s model, where a double-step copying scheme is assisted by a proofreading reaction $2 \rightarrow 0$ which is driven backward and tries to correct wrongly copied monomers, see Fig. 2 and the general energy landscape in Fig. 6A. Proceeding as before, we obtain the equation for the error rate

$$\frac{\eta}{1 - \eta} = \frac{p_1^w k_{21}^w + k_{20}^w - (k_{12}^w + k_{02}^w)\eta}{p_1^r k_{21}^r + k_{20}^r - (k_{12}^r + k_{02}^r)(1 - \eta)}, \quad (11)$$

where p_1^w and p_1^r have the same expression as in the previous section. Substituting the formula of the rates yields a lengthy expression (see Appendix). To simplify it, we exploit the fact that, in an efficient error-correction scheme, the second step and the proofreading reaction are irreversible, so that k_{12}^r , k_{12}^w , k_{20}^r , and k_{20}^w can be neglected. These assumptions are the same used to simplify the reaction scheme in Hopfield’s original model [13]. In this approximation, the error satisfies

$$\frac{\eta}{1 - \eta} = \frac{\left[e^{\frac{\delta_{10}}{T}} + r e^{\frac{\delta_{21} + \mu_{21}}{T}} \right]}{\left(1 + r e^{\frac{\mu_{21}}{T}} \right)} \frac{K - \eta e^{\frac{\Delta E_2^w}{T}} (1 + r)}{K e^{\frac{\delta_{21} + \delta_{10}}{T}} - (1 - \eta) e^{\frac{\Delta E_2^r + \delta_{02}}{T}} \left(e^{\frac{\delta_{10}}{T}} + r e^{\frac{\delta_{21}}{T}} \right)}, \quad (12)$$

where we defined $r = \omega_{21}/\omega_{10}$ as previously and $K = (\omega_{21}/\omega_{02}) e^{\frac{\mu_{10} + \mu_{21} - \mu_{02}}{T}}$. Increasing the proofreading driving μ_{02} (i.e. decreasing K) improves the accuracy as more and more copies are proofread. However, this also reduces the copying speed; the limiting error is then obtained in the limit of vanishing

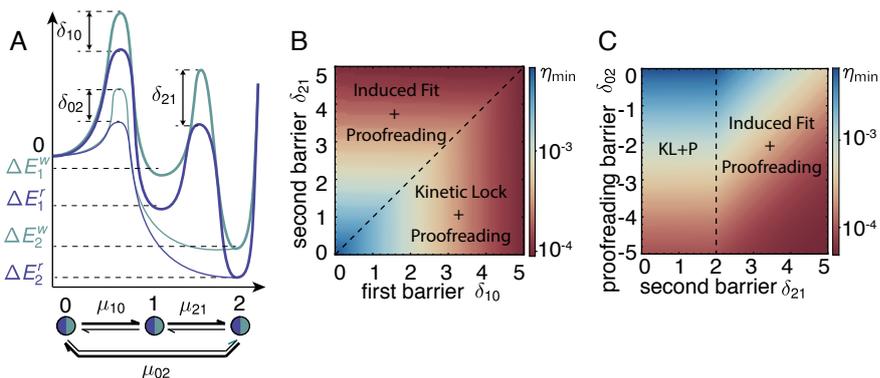


Fig. 6 Kinetic Proofreading. **A:** Energy landscape showing a generic double copy landscape together with the proofreading reaction which discriminates kinetically through the barrier $\delta_{02} < 0$. **B.** & **C.** Regions and minimum error. Parameters are $\Delta E_2^r = \Delta E_1^f = 0$, $\Delta E_2^w = \Delta E_1^w = 3T$. In panel B, we fixed $\delta_{02} = -2T$ while in panel C we fixed $\delta_{10} = 2T$. The chemical drivings and the relative time scales of the different reactions have been determined by numerically minimizing the error.

speed. Determining K in this way and substituting back in the numerator of Eq. (12) results in a general expression for the minimal error

$$\eta_{\min}(v \rightarrow 0) \approx e^{\frac{\Delta E_2^r - \Delta E_2^w + \delta_{02}}{T}} \frac{\left(e^{\frac{\delta_{10}}{T}} + r e^{\frac{\delta_{21}}{T}} \right)}{(1+r) e^{\frac{\delta_{21} + \delta_{10}}{T}}}, \quad (13)$$

where the irreversible assumption explicitly excludes the trivial energetic region. But even with this assumption, the expression above is richer than that of the irreversible error η_{irr} in previous sections.

The first factor in Eq. (13) represents the discrimination due to proofreading, and includes the energy difference $\Delta E_2^r - \Delta E_2^w < 0$ and the kinetic barrier $\delta_{02} < 0$. The following fraction represents the discrimination of the copying reaction, and through its dependence on the ratio r , different operating regions can be distinguished. As Eq. (13) is always monotonic in r , we can identify two different regions depending on whether the error decreases or increases with r :

- **Kinetic Lock + Proofreading.** This is the region in which $\delta_{10} > \delta_{21}$. The minimum error is obtained in the limit $r \gg 1$ of Eq. (13), so that the discrimination factor of the copying reaction, i.e. the fraction in Eq. (13), tends to $\exp(-\delta_{10}/T)$.
- **Induced Fit + Proofreading.** In this region the parameters satisfy $\delta_{21} > \delta_{10}$. To obtain the minimum error, one should take the limit $r \ll 1$ of Eq. (13), so that the fraction in Eq. (13) tends to $\exp(-\delta_{21}/T)$.

The existence of these two regions is shown in Fig. 6B and 6C, where we plot the minimum error as a function of the kinetic barrier of the second copy reaction δ_{21} and the proofreading barrier δ_{02} . According to Eq. 13, the

minimum error always decreases exponentially with δ_{02} , consistently with the figure. On the other hand, the minimum error depends on the maximum between δ_{10} and δ_{21} , so that the error in the figure depends on δ_{21} only when $\delta_{21} > \delta_{10}$.

This protocol is a generalization of previous proofreading models as those proposed by Hopfield [13] and Bennett [3]. In particular, Hopfield's model corresponds to the choice of $\delta_{10} = \delta_{02} = 0$, $\Delta E_2^w = \Delta E_1^w$ and $\Delta E_2^r = \Delta E_1^r$. Moreover, as the forward rates of the second reaction are equal for right and wrong incorporations, one should fix $\delta_{21} = \Delta E_2^w - \Delta E_2^r$. With this choice, the minimum error in Eq. (13) correctly becomes $\eta_{\min} \approx \exp[-2(\Delta E_2^w - \Delta E_2^r)/T]$, i.e. the square of the equilibrium error. On the other hand, Bennett's model of proofreading can be seen as the case in which accuracy in copying is achieved via kinetic lock with no energetic differences, $E_2^w = E_2^r$. In this case we have $\eta_{\min} \approx \exp[-(\delta_{21} - \delta_{02})/T]$.

7 Protocol D: Proofreading/accommodation

In this protocol, inspired by the later stage of the mRNA translation pathway (see Fig. 2), proofreading acts from the intermediate step of the reaction. The energy landscape is shown in Fig. 7A. Notice that in this case there are two sets of rates connecting state 0 with state 1, corresponding to either the first copying step or the proofreading reaction. To resolve this ambiguity, we call k_{10}^r the binding rate of right monomer for the copying reaction and \bar{k}_{10}^r the corresponding rate of the proofreading reaction, and similarly for the other rates. We adopt the same notation for the energies. The occupancies of the intermediate states then are $p_1^r = [k_{10}^r + \bar{k}_{10}^r + (1 - \eta)k_{12}^r]/(k_{01}^r + \bar{k}_{01}^r + k_{21}^r)$ and $p_1^w = (k_{10}^w + \bar{k}_{10}^w + \eta k_{12}^w)/(k_{01}^w + \bar{k}_{01}^w + k_{21}^w)$. The solution in terms of the occupancies is the same as in the double-copying model

$$\frac{\eta}{1 - \eta} = \frac{p_1^w k_{21}^w - k_{12}^w \eta}{p_1^r k_{21}^r - k_{12}^r (1 - \eta)} . \quad (14)$$

Substituting the expressions of the rates results in Eq. (21) in the Appendix. If we assume that all drivings are large, and that proofreading compensates copying so that the net speed vanishes as before, we obtain from the previous expression that the minimum error is

$$\eta_{\min}(v \rightarrow 0) \approx e^{-\frac{\delta_{10} + E_2^r - E_2^w + \delta_{10}}{T}} , \quad (15)$$

i.e. the minimum error associated to the first barrier, times the proofreading error reduction factor. To exploit the second barrier, we should take the opposite limit of large speed. In this case, the error tends to

$$\eta_{\min}(v \rightarrow \infty) \approx e^{-\frac{\delta_{10} - \delta_{21}}{T}} \frac{\left[e^{\frac{\delta_{10}}{T}} + \bar{r} e^{\frac{\bar{\delta}_{10}}{T}} + r e^{\frac{\delta_{21} + \mu_{21}}{T}} \right]}{\left(1 + \bar{r} + r e^{\frac{\mu_{21}}{T}} \right)} \quad (16)$$

with $\bar{r} = \bar{\omega}_{01}/\omega_{10}$. Assuming, as before, that the proofreading barrier is non-positive $\bar{\delta}_{10} \leq 0$, that the other two barriers are non-negative and that

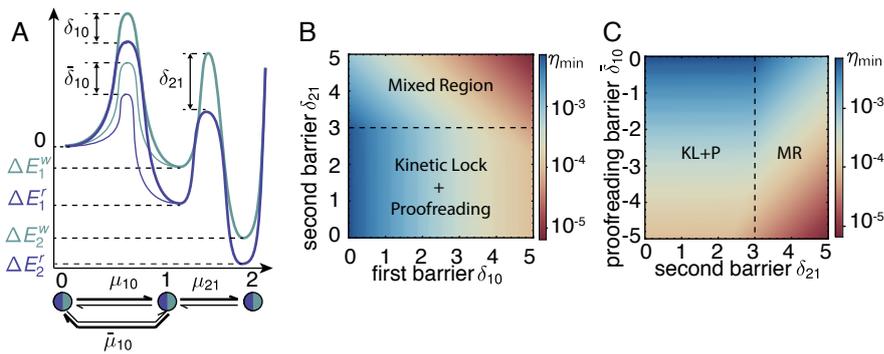


Fig. 7 Proofreading/accommodation. **A** Energy landscape **B.** & **C.** Regions and minimum error. Parameters are $\Delta E_2^r = \Delta E_1^r = 0$, $\Delta E_2^w = \Delta E_1^w = 3T$, $\delta_{01} = 2T$. The chemical drivings and the relative time scales of the different reactions have been determined by numerically minimizing the error.

$\Delta E^w - \Delta E^r \geq 0$, the minimum error is found in one of the following two regions:

- **Kinetic Lock + Proofreading.** This region is defined by the condition $\delta_{21} < \Delta E_2^w - \Delta E_2^r$ and is analogous to the one in the proofreading model of the previous section. The expression for the minimum error is thus given by Eq. (15).
- **Mixed Region.** In this region, the parameters satisfy $\delta_{21} > \Delta E_2^w - \Delta E_2^r$. The minimum error is obtained by taking the limit $\bar{r} \gg r$ and $\bar{r} \gg 1$ in Eq. (16) and is equal to $\eta_{\min} = \exp[(\bar{\delta}_{10} - \delta_{10} - \delta_{21})/T]$. As proofreading is performed from the intermediate state, in this model it is possible to exploit all three kinetic barriers at the same time. This is at variance with models B and C where only one barrier, either δ_{10} or δ_{21} , affects the minimum error in each region.

Also in this case we confirmed the existence of these two regions by numerically minimizing the error for different values of the discrimination parameters. The results are summarized in Fig. 7B and 7C.

8 Conclusions

In this paper, we considered four copy protocols inspired by the building blocks of the mRNA translation pathway. We have shown that, despite the relatively large number of parameters, their copying accuracy can be analyzed in full generality. Our analysis reveals that each of these copy protocols can function in several operating regimes, which generalize of the kinetic and energetic regimes of the single-step copy protocol (see Fig. 1). It remains an open problem to find generic rules by which kinetic and energetic discriminating steps can be combined in an arbitrarily complex copying protocol.

As for the single-step copy, the regimes of a complex copy protocol are also characterized by a distinctive minimal error and a strategy to achieve

it. Each regime can be associated to particular tradeoffs between chemical driving, speed, and error. It is interesting to explore how these trade-offs are resolved in actual biological systems. For example, it has been argued that initial energetic differences that could allow for energetic discrimination if equilibrated are not fully exploited in translation [30, 11, 15, 14]. This constitutes an example where, in the presence of a positive error-speed tradeoff, optimizing the speed seems to be more relevant.

In this paper, we focused on optimizing the error rate alone, and found that it can be minimized in different regimes corresponding to very different values of the speed. One might also wonder whether the error rate is determined by some general optimization principle. Some ideas in this direction have been recently explored. By looking at the first two steps of translation, Savir and Tlusty [29] concluded that the energy landscape is sculpted to optimize decoding of information. More recently, Rao and Peliti proposed a measure of optimality of copying reaction which includes a cost for dissipated work [26].

The modeling framework discussed in this paper is very general. Given the possibility of analytically solving any copying protocol, as discussed in section 4, it constitutes a natural starting point to study the dynamics of many complex copying machines. A possible further generalization would be to include the possibility of performing stochastic transitions *among* different sub-networks. This possibility is important to study backtracking, i.e. the possibility of copying enzymes to error-correct entire strings of the copy polymer containing a high error fraction, see e.g. [10, 31, 19].

From the point of view of building up a general theory [24], it would also be interesting to understand the relationship between copying and other information-processing tasks performed by biological systems. For example, an interesting analogy exists with sensing, where cells have to detect a (noisy) external signal and internally store its value [16, 5]. Some aspects of this analogy have already been explored, for example with the concept of kinetic proofreading [12]. It is however possible that this analogy can be pushed further. For example, thinking in terms of kinetic and energetic discrimination, as discussed in this paper, could help understanding regimes of sensing and other information-processing tasks.

Acknowledgements This work was supported by the Ministerio de Economía y Competitividad (Spain) and FEDER (European Union), under project FIS2012-37655-C02-01.

Appendix: exact expressions

In this Appendix, we report the exact expression for the solutions of the proofreading models, that have been omitted or simplified in the Results section for ease of reading.

 Template-assisted polymerization without intermediate states

The exact equation for the error is in the Results section. The copying speed is equal to

$$v = k_{10}^w + k_{10}^r - k_{01}^w \eta - k_{01}^r (1 - \eta). \quad (17)$$

Double-step copying

Also in this case, the exact equation for the error is in the Results section. The expression for the speed is

$$v = \mathcal{N}[p_1^w k_{21}^w + p_1^r k_{21}^r - k_{12}^w \eta - k_{12}^r (1 - \eta)] \quad (18)$$

where we introduced the normalization factor for the occupancies \mathcal{N} , which in this case is equal to $\mathcal{N} = (1 + p_1^w + p_1^r)^{-1}$.

Kinetic proofreading

The error rate of the proofreading model satisfies the equation

$$\begin{aligned} \frac{\eta}{1 - \eta} &= \frac{\left[e^{\frac{\delta_{10}}{T}} + r e^{\frac{\delta_{21} + \mu_{21}}{T}} \right]}{\left(1 + r e^{\frac{\mu_{21}}{T}} \right)} \times \\ &\times \frac{r e^{\frac{\mu_{10} + \mu_{21}}{T}} + r_p \left(1 + r e^{\frac{\mu_{21}}{T}} \right) - \eta e^{\frac{\Delta E_2^w}{T^2}} \left[r + (1 + r) r_p e^{\frac{\mu_{02}}{T}} \right]}{r e^{\frac{\mu_{10} + \mu_{21} + \delta_{21} + \delta_{10}}{T}} + r_p e^{\frac{\delta_{20}}{T}} \left(e^{\frac{\delta_{10}}{T}} + r e^{\frac{\mu_{21} + \delta_{21}}{T}} \right) - (1 - \eta) e^{\frac{\Delta E_2^r}{T}} A} \end{aligned} \quad (19)$$

where $r_p = \omega_{02}/\omega_{10}$ and $A = \left(r e^{\frac{\delta_{10} + \delta_{21}}{T}} + r_p e^{\frac{\delta_{10} + \delta_{02} + \mu_{02}}{T}} + r r_p e^{\frac{\delta_{21} + \delta_{02} + \mu_{02}}{T}} \right)$.

The copying speed is equal to

$$v = \mathcal{N}[p_1^w k_{21}^w + k_{20}^w + p_1^r k_{21}^r + k_{20}^r - (k_{12}^w + k_{02}^w) \eta - (k_{12}^r + k_{02}^r)(1 - \eta)] \quad (20)$$

where \mathcal{N} is defined as previously.

Proofreading/accommodation

The error rate satisfies

$$\begin{aligned} \frac{\eta}{1 - \eta} &= e^{-\frac{\delta_{21}}{T}} \frac{\left[e^{\frac{\delta_{10}}{T}} + \bar{r} e^{\frac{\bar{\delta}_{10}}{T}} + r e^{\frac{\delta_{21} + \mu_{21}}{T}} \right]}{\left(1 + \bar{r} + r e^{\frac{\mu_{21}}{T}} \right)} \times \\ &\times \frac{e^{\frac{\mu_{21}}{T}} \left(e^{\frac{\mu_{10}}{T}} + \bar{r} \right) - \eta e^{\frac{\Delta E_2^w}{T^2}} \left(1 + \bar{r} e^{\frac{\bar{\mu}_{10}}{T}} \right)}{e^{\frac{\mu_{21}}{T}} \left(e^{\frac{\mu_{10} + \delta_{10}}{T}} + \bar{r} e^{\frac{\bar{\delta}_{10}}{T}} \right) - (1 - \eta) e^{\frac{\Delta E_2^r}{T}} \left(e^{\frac{\delta_{10}}{T}} + \bar{r} e^{\frac{\bar{\mu}_{10} + \bar{\delta}_{10}}{T}} \right)} \end{aligned} \quad (21)$$

and the copying speed is

$$v = \mathcal{N}[p_1^w k_{21}^w + p_1^r k_{21}^r - k_{12}^w \eta - k_{12}^r (1 - \eta)]. \quad (22)$$

References

1. Andrieux, D., Gaspard, P.: Nonequilibrium generation of information in copolymerization processes. *Proceedings of the National Academy of Sciences* **105**(28), 9516–9521 (2008)
2. Andrieux, D., Gaspard, P.: Information erasure in copolymers. *EPL (Europhysics Letters)* **103**(3), 30,004 (2013)
3. Bennett, C.H.: Dissipation-error tradeoff in proofreading. *BioSystems* **11**(2), 85–91 (1979)
4. Betterton, M., Jülicher, F.: A motor that makes its own track: helicase unwinding of dna. *Physical review letters* **91**(25), 258,103 (2003)
5. Bo, S., Del Giudice, M., Celani, A.: Thermodynamic limits to information harvesting by sensory systems. *Journal of Statistical Mechanics: Theory and Experiment* **2015**(1), P01,014 (2015)
6. Cady, F., Qian, H.: Open-system thermodynamic analysis of dna polymerase fidelity. *Physical biology* **6**(3), 036,011 (2009)
7. Esposito, M., Lindenberg, K., Van den Broeck, C.: Extracting chemical energy by growing disorder: efficiency at maximum power. *Journal of Statistical Mechanics: Theory and Experiment* **2010**(01), P01,008 (2010)
8. François, P., Voisinne, G., Siggia, E.D., Altan-Bonnet, G., Vergassola, M.: Phenotypic model for early t-cell activation displaying sensitivity, specificity, and antagonism. *Proceedings of the National Academy of Sciences* **110**(10), E888–E897 (2013)
9. Freter, R.R., Savageau, M.A.: Proofreading systems of multiple stages for improved accuracy of biological discrimination. *Journal of theoretical biology* **85**(1), 99–123 (1980)
10. Galburt, E.A., Grill, S.W., Wiedmann, A., Lubkowska, L., Choy, J., Nogales, E., Kashlev, M., Bustamante, C.: Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **446**(7137), 820–823 (2007)
11. Gromadski, K.B., Rodnina, M.V.: Kinetic determinants of high-fidelity trna discrimination on the ribosome. *Molecular cell* **13**(2), 191–200 (2004)
12. Hartich, D., Barato, A.C., Seifert, U.: Nonequilibrium sensing and its analogy to kinetic proofreading. *arXiv preprint arXiv:1502.02594* (2015)
13. Hopfield, J.J.: Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences* **71**(10), 4135–4139 (1974)
14. Johansson, M., Bouakaz, E., Lovmar, M., Ehrenberg, M.: The kinetics of ribosomal peptidyl transfer revisited. *Molecular cell* **30**(5), 589–598 (2008)
15. Johansson, M., Lovmar, M., Ehrenberg, M.: Rate and accuracy of bacterial protein synthesis revisited. *Current opinion in microbiology* **11**(2), 141–147 (2008)
16. Lan, G., Sartori, P., Neumann, S., Sourjik, V., Tu, Y.: The energy-speed-accuracy trade-off in sensory adaptation. *Nature physics* **8**(5), 422–428 (2012)
17. Loeb, L.A., Kunkel, T.A.: Fidelity of DNA synthesis. *Annual review of biochemistry* **51**(1), 429–457 (1982)
18. Mckeithan, T.W.: Kinetic proofreading in T-cell receptor signal transduction. *Proceedings of the national academy of sciences* **92**(11), 5042–5046 (1995)
19. Mellenius, H., Ehrenberg, M.: DNA Template Dependent Accuracy Variation of Nucleotide Selection in Transcription. *PloS one* **10**(3), e0119,588 (2015)
20. Murugan, A., Huse, D.A., Leibler, S.: Speed, dissipation, and error in kinetic proofreading. *Proceedings of the National Academy of Sciences* **109**(30), 12,034–12,039 (2012)
21. Murugan, A., Huse, D.A., Leibler, S.: Discriminatory proofreading regimes in nonequilibrium systems. *Physical Review X* **4**(2), 021,016 (2014)
22. Ninio, J.: Kinetic amplification of enzyme discrimination. *Biochimie* **57**(5), 587–595 (1975)
23. Pape, T., Wintermeyer, W., Rodnina, M.: Induced fit in initial selection and proofreading of aminoacyl-trna on the ribosome. *The EMBO Journal* **18**(13), 3800–3807 (1999)

-
24. Parrondo, J.M., Horowitz, J.M., Sagawa, T.: Thermodynamics of information. *Nature Physics* **11**(2), 131–139 (2015)
 25. Pauling, L.: *Festschrift fuer Prof. Dr. Arthur Stoll* (1958)
 26. Rao, R., Peliti, L.: Thermodynamics of accuracy in kinetic proofreading: Dissipation and efficiency trade-offs. arXiv preprint arXiv:1504.02494 (2015)
 27. Sartori, P., Pigolotti, S.: Kinetic versus energetic discrimination in biological copying. *Physical review letters* **110**(18), 188,101 (2013)
 28. Sartori, P., Pigolotti, S.: Thermodynamics of error correction. arXiv preprint arXiv:1504.06407 (2015)
 29. Savir, Y., Tlusty, T.: The ribosome as an optimal decoder: a lesson in molecular recognition. *Cell* **153**(2), 471–479 (2013)
 30. Thompson, R.C., Karim, A.M.: The accuracy of protein biosynthesis is limited by its speed: high fidelity selection by ribosomes of aminoacyl-tRNA ternary complexes containing GTP [γ S]. *Proceedings of the National Academy of Sciences* **79**(16), 4922–4926 (1982)
 31. Voliotis, M., Cohen, N., Molina-París, C., Liverpool, T.B.: Fluctuations, pauses, and backtracking in DNA transcription. *Biophysical journal* **94**(2), 334–348 (2008)
 32. Zaher, H.S., Green, R.: Fidelity at the molecular level: lessons from protein synthesis. *Cell* **136**(4), 746–762 (2009)