The individual and combined benefits of different non-equilibrium proofreading mechanisms.

Adélaïde A. Mohr[1], Daniel M. Busiello [1,2], Stefano Zamuner[1], Paolo De Los Rios[1,3],

[1] Institute of Physics, École Polytechnique Fédérale de Lausanne - EPFL, Lausanne, Switzerland;
[2] Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany;
[3] Institute of Bioengineering, École Polytechnique Fédérale de Lausanne - EPFL, Lausanne, Switzerland.

Abstract

Genome duplication, transcription and translation are among many crucial cellular processes that need to be performed with high fidelity. However those extremely low error rates cannot be explained with simple equilibrium thermodynamic considerations. They instead require considering irreversible, energy consuming reactions in the overall mechanism. We develop here a model of substrates selection comprising energy consuming steps and which aims at selecting right substrates among wrong ones. With this model, we investigate the impact of energy consumption on the accuracy and the speed of the selection, as well as different selection strategies. The model presented here encompasses the classic kinetic proofreading scheme and a different mechanism whereby the rates of the energy consuming step are modulated by the nature of the substrate. We show that, in our framework, the fastest and most accurate selection strategy relies on a combination of both mechanisms. A structurally and biochemically informed coarse-grained description of real biological processes such as DNA replication and protein translation, traditionally used as examples of kinetic proofreading at work, shows that, as a matter of fact, a combination of both mechanisms explored here is exploited.

Introduction

The ability to discriminate between correct and wrong substrates is essential for many biological processes, and in particular when information is transferred during transcription and translation. Indeed, from DNA replication to protein synthesis, the ability to read information and select the correct nucleotide or anti codon is crucial for the survival of all living organisms and of their progeny. From a thermodynamic perspective, high fidelity is difficult to achieve because the free-energy differences $\Delta G = G_R - G_W$ that should allow choosing the right (R) rather than the wrong (W) substrate are often in the range of only a few k_BT. Therefore, it becomes impossible for the error rate of these processes to be lower than one wrong substrate in a few hundreds, according to ratio of their Boltzmann probabilities, $\exp(-\Delta G/k_BT)$. Yet, the error rate in translation has been measured to be as low as $\sim 10^{-4}$ [1], and a staggering $\sim 10^{-10}$ for DNA replication [2]. To overcome this conundrum, it has been proposed, consistent with biochemical observations, that some irreversible, energy-consuming steps must be integral parts of these highly accurate processes, thus unshackling them from the bounds of equilibrium thermodynamics.

Hopfield and Ninio [3, 4] independently proposed, almost five decades ago, a proofreading scheme able to reproduce these exceptionally low error rates. By repeating the selection through n independent steps, separated by irreversible reactions, the error rate can be reduced to $(\exp(-\Delta G/k_BT))^n$. Several subsequent studies have chosen similar approaches to investigate different selection methods [5], more complex kinetic pathways [6] or different characteristics of these proofreading schemes [7, 8, 9, 10].

Although many of these studies acknowledge the need for energy dissipation, possibly in the form of nucleotide (ATP or GTP, henceforth NTP for generality) consumption for proper functioning, few of them explicitly respect simple thermodynamics constraints, such as detailed balance in equilibrium conditions.

In this work, we propose a general model of proofreading that explicitly considers the energy available to the system as a function of the NTP and NDP concentration (NDP is the product of the hydrolysis of NTP). This framework encompasses the classical kinetic proofreading scheme as well as other, not mutually exclusive, means of selection. We compare the behaviour of these selection methods in different energy conditions, discuss their limitations and advantages. We find that Hopfield's kinetic proofreading scheme is greatly limited by required slow transitions implying a strong trade-off between the accuracy and the speed of the selection. However, its combination with another selection mechanism allows to lift that limitation and leads to an improvement of both the accuracy and the velocity. Finally, the study of two different real biological systems highlights the use of similar combinations of selection methods.

Proofreading model

A model capturing the basic proofreading scheme, while allowing the correct accounting of the energy that is available to the process, is represented in Fig. 1. The enzyme responsible for the substrate selection and

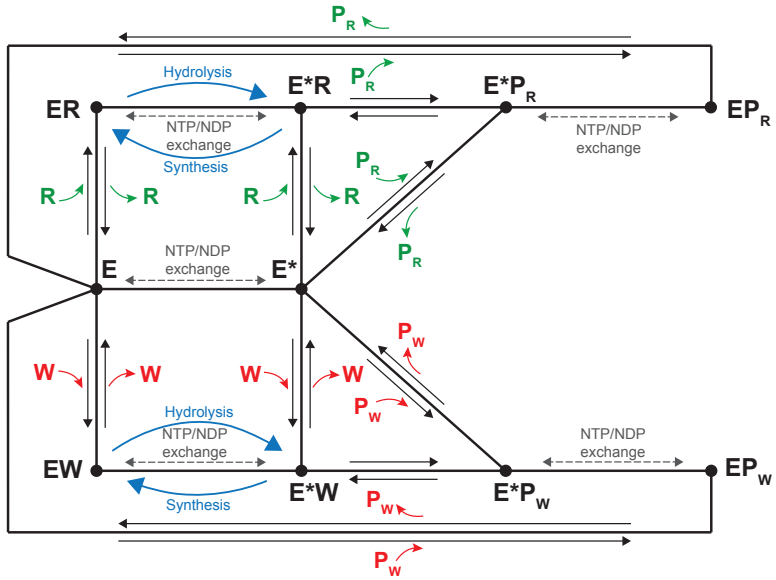

Figure 1: **Thermodynamically consistent proofreading scheme.** E is the enzyme in its NTP-state and E^* the enzyme in its NDP-state, R and W correspond to the right and wrong substrate and P_R and P_W the products associated to the respective substrate. Transition between NTP and NDP-states enzymes can either happen through nucleotide exchange (gray dashed arrows) or via NTP hydrolysis or NDP synthesis (blue arrows) when the enzyme is bound to the substrate (R or W). This is due to the approximation of very slow hydrolysis and synthesis, negligible compared to exchange, unless stimulated by a substrate. Other transitions in the scheme simply corresponds to binding or unbinding to substrates or products.

product formation can either be in an NTP bound-state, E, or in an NDP bound-state E^*. It original state E can either bind to the right, R, or wrong, W, substrate. After substrate binding, the enzyme can evolve into an intermediate state, E^*S (S is either R or W) that is reached either through an energy-consuming reaction (here hydrolysis of NTP into NDP; blue arrows in the figure) or by nucleotide exchange (release of NTP and binding of NDP marked by the black transition lines between E and E^* states). If the substrate is released ($E^*S \to E^*$), the enzyme can then exchange NDP for NTP. Otherwise the substrate is turned into a product ($E^*S \to E^*P_S$), which can then unbind before or after NDP is exchanged for NTP.

In this model, we assume the hydrolysis and synthesis to be very slow processes which can be neglected compared to the nucleotide exchange. Only the presence of a substrate, right or wrong, is able to stimulate enough these reactions, explaining why those transitions are only present between ES and E^*S.

A centerpiece of this model is the reversibility of all reactions, which allows deriving a correct relation between the ability of the system to discriminate right substrates from wrong ones and the available energy. Irreversibility or, to be more correct, directionality of reactions emerges from the unbalance between hydrolysis/synthesis and nucleotide exchange. This unbalance is tuned by the parameter $\alpha = [\text{NTP}]/[\text{NDP}]$ representing the amount of energy available to the system. At equilibrium $\alpha = \alpha_{\text{eq}} = 10^{-6}$, corresponding to the equilibrium ratio between NTP and NDP. For any value of $\alpha > \alpha_{\text{eq}}$, the system is pushed out of equilibrium, breaking the detailed balance and leading to directionality in the system reactions. In this work, we usually compare the equilibrium case, $\alpha = \alpha_{\text{eq}}$, with the physiological one, $\alpha = 10$ [11].

In what follows, we consider a fixed concentration of substrate and products. In particular, $[P_W] = [P_R] = 0$ as we assume that products are immediately extracted from the solution. As a consequence, there is a constant net flux flowing through the network even at equilibrium due to these chemostatted concentrations.

Both at equilibrium and out of equilibrium, the system reaches a steady-state. In this state, we compute the flux of wrong and correct products, J_w and J_r, released by the enzyme. To measure the fidelity of the process, we define, as in [3], the error rate $f := \dfrac{J_w}{J_r}$. The quantities J_r and f will be used to assess the speed and accuracy of different selection strategies. Further mathematical details on the model are given in the appendices.

Kinetic proofreading: Trade-off between speed and accuracy

The original Hopfield-Ninio selection scheme, represented in Fig. 2, is called **kinetic proofreading** and is based on the presence of the two discriminating steps separated by an irreversible reaction. The initial discriminating step acts as the usual selection step of an enzyme in the classic Michaelis-Menten description, while the second discriminating step allows to reject wrong substrates, acting thus as a proofreading mech-

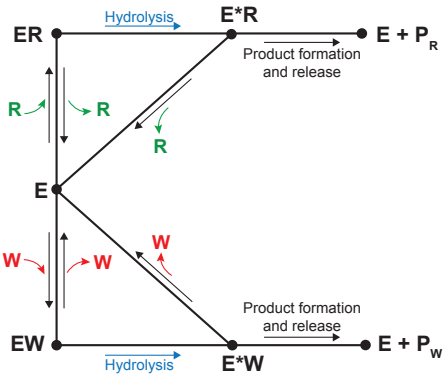

Figure 2: Hopfield's proofreading scheme. Under Hopfield's hypothesis the system is symmetric in both the hydrolysis and product formation rates and those rates are slow. It is worth noting the unidirectionality of the transitions in hydrolysis, product formation and release and the release of the substrate in state E^*S

anism. The kinetic proofreading scheme can be obtained from the model introduced above by considering that the synthesis and exchange reactions between ES to E^*S are infinitely slow, while the exchange reaction between E^* and E is infinitely fast, and $\alpha \to \infty$. Whereas at equilibrium (*i.e.* no irreversible step), the error rate would have a simple linear dependence on the ratio of the dissociation constants, the presence of an irreversible step allows the system to probe the substrates twice, resulting in

$$f_{Hop} := \frac{J_w}{J_r} \geq \frac{[W]}{[R]} \left(\frac{K_R^d}{K_W^d} \right)^2 \tag{1}$$

a relation that can be simply derived within a Michaelis-Menten approach.
Since the Hopfield model represented in Fig.2 is a special case of the one presented in this work (Fig. 1), its outcome can be recovered in a suitable limit, with the advantage that, being it thermodynamically consistent, the interplay between energy consumption, fidelity and velocity can be fully appreciated.
To do so, we assume our system to be completely symmetric between the correct and incorrect substrates except for the unbinding rates: $k_W^{\text{off}} \neq k_R^{\text{off}}$ and $k_{W^*}^{\text{off}} \neq k_{R^*}^{\text{off}}$. In agreement with Hopfield's assumption, we set these unbinding rates as large as $k_{W^*}^{\text{off}}, k_{R^*}^{\text{off}} \gg k_{SP}$, where k_{SP} is the product production rate for both correct and wrong substrates.

In Fig.3A we show the behavior of the error rate f as a function of the hydrolysis rate (k_S^{hy}) for various values of the available energy.
If $\alpha = \alpha_{eq}$ (here $\alpha_{eq} = 10^{-6}$), the error rate does not depend the hydrolysis rate, and is equal to the equilibrium predictions [3], as expected,

$$f_{eq} = \frac{[W]}{[R]} \frac{K_R^d}{K_W^d} \tag{2}$$

As the energy available to the system increases (increasing values of α), three regions can be highlighted. For very low values of the hydrolysis rate (while keeping all other rates, including exchange, fixed), the system cannot improve the error rate with respect to its equilibrium value. As a matter of fact, in this case the enzyme spends most of its time binding/unbinding the substrate while in either the E or E^* state (bound to NTP and NDP respectively) with very few hydrolysis-induced transitions between the two, as shown in Fig. 4A. Thus, no "double-checking" of the substrate takes place, *i.e.* no kinetic proofreading. As k_S^{hy} increases, there is a region where the error rate decreases, and the Hopfield result ($f = 0.01$ given the present choice of parameters) is the lower-bound that can be attained for increasing available energy. Indeed, the system is able to fully exploit the two selection steps, as highlighted in Fig. 4B. For extremely high values of the hydrolysis rate, the equilibrium error rate is again the dominant behavior, irrespective of the value of α. In this case, the extremely high rate of the $ES \to E^*S$ transition prevents the equilibration of the ES complex (Fig. 4C), and thus prevents one of the two substrate-testing steps necessary for proofreading, as predicted by Hopfield [3].

In addition to improving accuracy, energy consumption can also allow the process to be faster (Fig.3B). At equilibrium, $\alpha = \alpha_{eq}$, the production rate is expectedly constrained to the one due to the imposed imbalance between the substrate and product concentrations. Away from equilibrium ($\alpha > \alpha_{eq}$), J_r can be even slower than at equilibrium for small hydrolysis rates: the increasing concentration of NTP pushes the enzyme back to E by nucleotide exchange, and product production is slowed. Yet, as k_S^{hy} increases, hydrolysis prevails over exchange and product catalysis correspondingly increases.

Within this Hopfield-like scheme, minimizing the error rate and simultaneously maximizing the production rate is not possible at the same time. Slowing down the hydrolysis allows the system to fully exploit the two selection steps proposed by Hopfield but also slows down the passage of the substrate in the proofreading pathway. This implies a high accuracy but a production rate (red dots on Fig. 3A and B) which is not the optimal one, by orders of magnitude, in agreement with [5]. Increasing the hydrolysis rate improves the production rate up to a maximum value (black dot on Fig. 3B), at the expense of proofreading fidelity. Such trade-off between speed and accuracy, sometimes with energy dissipation, are commonly observed in information processing systems like proofreading or error correction [6, 8, 10, 12, 13], sensory adaptation [14] or polymer synthesis [15, 16, 17].

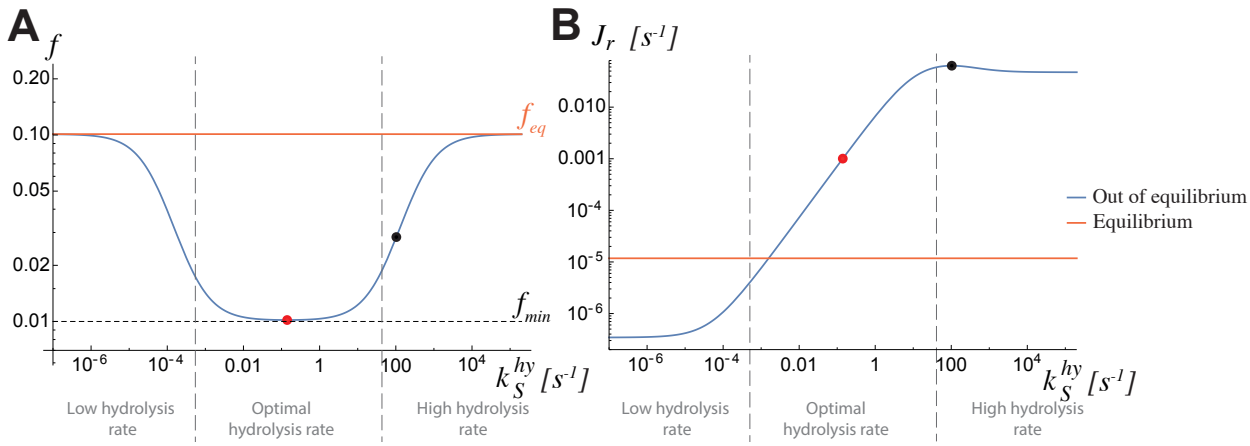

Figure 3: Hopfield-like selection system highly depends on available energy, α, and on the transition rate between ES and E^*S, k_S^{hy}. **A** Error rate and **B** production rate in function of the hydrolysis rate out of equilibrium in physiological conditions *i.e.* $\alpha = 10$ (blue) and at equilibrium $\alpha = \alpha_{eq}$ (orange). Red dots: minimum in error rate under physiological condition ($\alpha = 10$). Black dots: maximum in production rate under physiological condition ($\alpha = 10$). Dashed horizontal line: theoretical limit for the error rate under Hopfield's assumption (1).

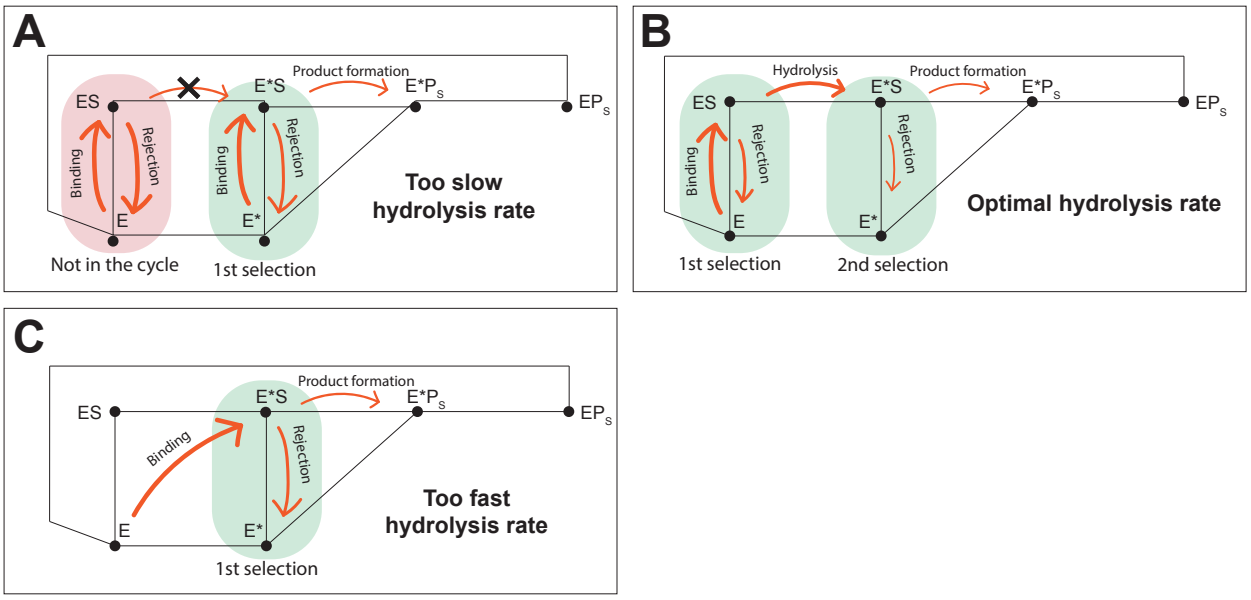

Figure 4: Scheme of the selection pathways following Hopfield kinetics for different values of the hydrolysis rate. **A** the slow hydrolysis rate prevent substrate from entering the selection pathway from the E state. Therefore the system is using only one selection step (between E^* and E^*S, implying a high error rate. **B** the optimal hydrolysis rate allows substrate to enter the selection pathway from the E state and make the most of the two selection steps leading to a low error rate. **C** with a fast hydrolysis rate, the substrate entering the selection is directly brought to the state E^*S where it can be rejected or directly proceed to product formation. This prevents the system from benefiting from the two selection steps, leading to a higher error rate.

Catalytic discrimination: Active recognition of the substrate leads to fast processes.

In a kinetic proofreading scheme, accuracy is based on the repetition of energy-based discriminating steps. However, the system is forced to slow down to exploit its selection mechanism to the fullest. Indeed, the Hopfield scheme is optimal for relatively slow hydrolysis, so that the system has time to equilibrate in both ES and E^*S states. The price is thus an inevitably low flux of product.

A different selection strategy could instead rely on the ability of the correct substrate to act as a catalyst for the selection process: The system would progress faster for correct substrates than for wrong ones, resulting in their **catalytic** discrimination. This selection mechanism is similar to the kinetic discrimination proposed in [5] (we here use the word catalytic instead as kinetic to avoid any confusion with the kinetic proofreading) and to the forward discrimination in [17] .

This selection strategy is possible if the enzyme first recognizes the nature of the substrate (through binding) and then accelerate its uptake in the selection scheme. This approach is similar to a Maxwell Demon, whereby an appropriate action is taken only if a specific event occurs [18, 19].

In the present model, this is captured by allowing hydrolysis to be faster when the bound substrate is the correct one. At the molecular level, this scheme would require the enzyme to be able to access two different states (*e.g.* ES_1 and ES_2) that are characterized by different hydrolysis rates [20]. Substrate binding would then, allosterically, modulates the equilibrium constant between the two states in such a way that binding of the right substrate should stabilize the fast hydrolysis state more than the wrong substrate. Here, this phenomenon is coarse-grained and we are only considering the effective rate due to the inbalance between ES_1 and ES_2, namely k_S^{hy}, with $S = R$ or W.

In the same spirit of adapting the scheme in Fig. 1 to Hopfield's kinetic proofreading, in this case we set all the rates of the reactions involving the right and wrong substrates to be equal, apart for the hydrolysis rates (and consequently the synthesis rates) of the $ES \to E^*S$ reactions. Selection is therefore only based on **catalytic discrimination** during the hydrolysis step and can thus be effective only in non-equilibrium conditions (Fig. 5). This also implies that the dissociation constants of the wrong and right substrates for the enzyme (both E and E^*) are the same and no discrimination would be possible at equilibrium ($f = 1$ if $[R] = [W]$).

With these assumptions and with infinite available energy ($\alpha \to \infty$ and $\alpha_{eq} \to 0$), the error rate is:

$$f_{th} = \frac{W}{R}\frac{k_W^{\mathrm{hy}}}{k_R^{\mathrm{hy}}}\frac{(k_R^{\mathrm{hy}} + k_S^{\mathrm{off}})}{(k_W^{\mathrm{hy}} + k_S^{\mathrm{off}})} \quad . \tag{3}$$

which is indeed the asymptotic limit in Fig.5 and very similar to the error rate found in [17] when considering similar discrimination mechanisms. Here k_S^{off} is the unbinding rate for both W and R, and k_W^{hy}, k_R^{hy}, are the hydrolysis rate for EW, respectively ER. In a non-equilibrium setting, necessary to perform any kind of selection, the ability of the system to be accurate largely depends on the choice of k_R^{hy}, k_W^{hy}, k_S^{off} as expected from (3) and observed on Fig. 6. Clearly, when both hydrolysis rates are equal, slow or fast, no selection is possible (blue and orange line on Fig. 6A), and fast hydrolysis rates bring a higher production rate compared to slow ones. To trigger proofreading, the system needs to decrease k_W^{hy} with respect to k_R^{hy}, but only if the unbinding rate is set accordingly. Indeed the system starts exhibiting a significant reduction of the error rate when $k_S^{\mathrm{off}} > k_W^{\mathrm{hy}}$ (dashed line on Fig. 6A). Overall, the minimal error rate is reached for $k_S^{\mathrm{off}} \gg k_W^{\mathrm{hy}}$. Therefore, the selectivity depends greatly on the ability to measure quickly (high k_S^{off}) and to have an effective feedback after measurement (slower hydrolysis for wrong substrate, *i.e.* small k_W^{hy}).

On intuitive grounds, the system will work best when the right pathway rapidly proceeds toward the product after substrate binding, while the wrong pathway prefers to unbind the substrate before it is processed. This translates into $k_R^{\mathrm{hy}} > k_S^{\mathrm{off}} > k_W^{\mathrm{hy}}$.

Moreover, at variance with Hopfield's kinetic proofreading scheme, error rate minimization could be achieved alongside maximum production rate (Fig. 6), and no compromise between accuracy and speed has to be made.

Kinetic proofreading for accuracy, catalytic discrimination for speed

As we described above, the kinetic proofreading scheme is based on the difference in free energy between two different states. In terms of substrate binding, this implies a difference in affinity where the enzyme is more likely to be bound to the right substrate compared to the wrong one, which is discarded. To amplify this mechanism, multiple selection steps can be added separated by energy dependant steps ensuring the unidirectionality of the process. However, a downside of this strategy is the incompatibility between high production rate and low error rate. On the other hand, catalytic discrimination relies on a difference in the activation energy between forming two states of equivalent free energy. This translates into a faster process for the right substrate compared to the wrong one. This selection mechanism is only efficient in strongly non-equilibrium conditions, but without requiring a strong trade-off between speed and accuracy.

In our model, we represent the kinetic proofreading scheme as based on the difference between the unbinding rates of the right and wrong substrate with the enzyme in two different steps: before ($k_W^{\mathrm{off}} > k_R^{\mathrm{off}}$) and after hydrolysis ($k_{W^*}^{\mathrm{off}} > k_{R^*}^{\mathrm{off}}$). Catalytic discrimination is implemented during the hydrolysis step where

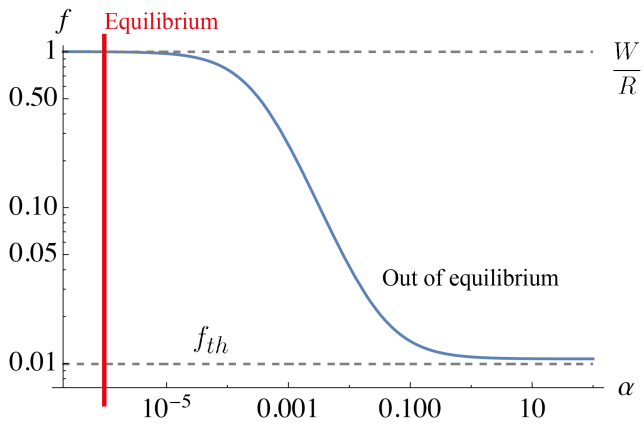

Figure 5: Error rate f in a catalytic discrimination scheme as a function of the available energy, α. Catalytic discrimination is only effective away from equilibrium. The upper dashed line corresponds to the error rate at equilibrium where no selection is effective, while the lower dashed line is the error rate limit (3) when the available energy is approaching ∞. The red vertical line represents the equilibrium condition with $\alpha = \alpha_{eq}$.

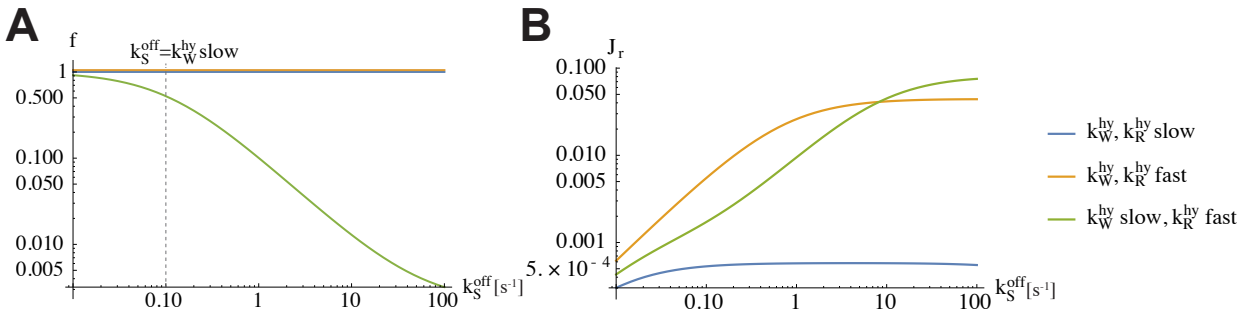

Figure 6: Error rate (A) and production rate (B) for slow hydrolysis rate for W and R (blue), fast hydrolysis rates for both W and R and slow hydrolysis rate for W and fast hydrolysis rate for R. For a selection mechanism based of catalytic discrimination, the fidelity is dictated by a slow hydrolysis rate for incorrect substrate, k_W^{hy}, a fast hydrolysis rate for correct substrate k_R^{hy}, and a fast enough unbinding rate k_S^{off}.

we assume that the right substrate accelerates the hydrolysis compared to the wrong one. This means that the process is much faster for the right substrate than for the wrong one ($k_R^{\text{hy}} \gg k_W^{\text{hy}}$).

To compare these two selection strategies, we explored the behavior of the model represented in Fig. 1 by means of two parameters that allow transitioning from one selection mechanism to the other. The first, $\omega = \frac{k_R^{\text{hy}}}{k_W^{\text{hy}}}$, corresponds to the acceleration of the hydrolysis rate in the presence of the right substrate, k_R^{hy}, compared to the hydrolysis rate with the wrong substrate, k_W^{hy}, which controls the catalytic discrimination. The second, $\lambda = \frac{k_W^{\text{off}}}{k_R^{\text{off}}} = \frac{k_{W*}^{\text{off}}}{k_{R*}^{\text{off}}}$, controls the kinetic proofreading *à la* Hopfield. All other parameters are completely symmetric between the two branches of the reaction network.

The region of the values of the parameters that we explored is $\omega \in [1, 1000]$ and $\lambda =\in [1, 100]$ (as a reference, KP indicates the selection model based on Hopfield's kinetic proofreading scheme for $\omega = 1$ and $\lambda = 100$, while C indicates catalytic discrimination for $\omega = 1000$ and $\lambda = 1$).

Fig. 7 reports the phase space associated to the error and production rates (f and J_r respectively) obtained by varying ω and λ in three different energy conditions ($\alpha = \alpha_{eq}, 0.005, 10$).

At equilibrium, the phase space is simply a point that captures the error rate at equilibrium, trivially related to the ratio of the dissociation constants of the right and wrong substrates. Changing the kinetic parameters cannot play any role as we saw in Fig. 5.

Increasing α opens a larger region of the parameter space where proofreading by both methods take place. Furthermore, it is possible to recognize the presence of a combination of KP and C (KP+C), characterized by a large difference between the hydrolysis rates ($\omega = 1000$) and an important difference in the unbinding rates such that $\lambda = 100$, where the best features of both schemes coexist, namely low error rate and high production rate.

This combination of selection strategies can also be found in one of the protocols investigated in [21]. However, in their paper they only focused on obtaining the minimum error rate in particular configurations without comparing them to each other. Here we are able to show how Hopfield's kinetic proofreading

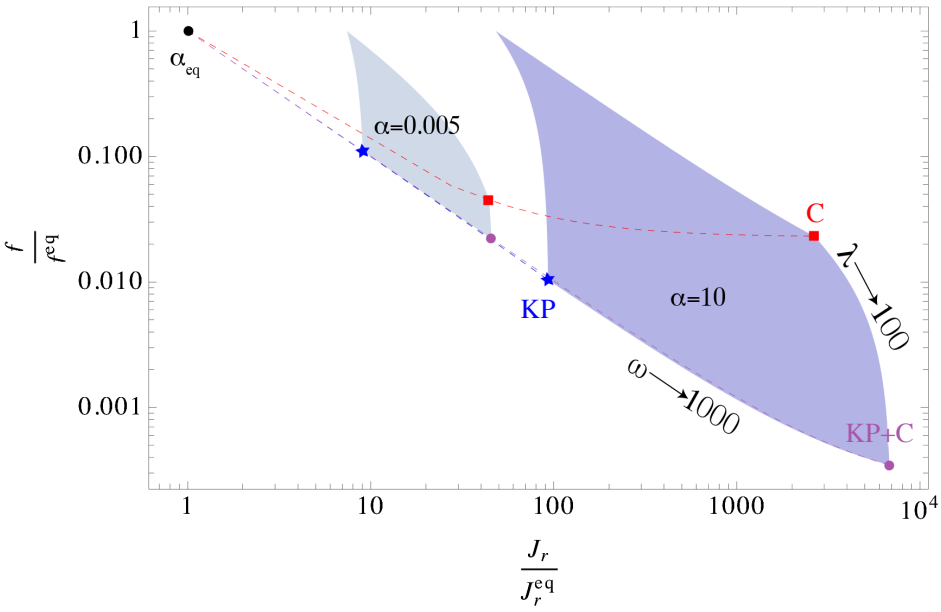

Figure 7: Error and production rates phase space for $\omega = \frac{k_R^{hy}}{k_W^{hy}} \in [1, 1000]$, $\lambda = \frac{k_W^{off}}{k_R^{off}} = \frac{k_{W*}^{off}}{k_{W*}^{off}} \in [1, 100]$ and different energy conditions (equilibrium: $\alpha = 10^{-6}$, non-equilibrium: $\alpha = 0.005$ and $\alpha = 10$, physiological conditions). Blue star: KP scheme ($\lambda = 100$ and ω=1), red square: catalytic discrimination C ($\lambda = 1$ and $\omega = 1000$) and purple dot: KP + catalytic discrimination C ($\lambda = 100$, $\omega = 1000$).

combined with catalytic discrimination is beneficial for both the accuracy and the velocity of the process.

Finally, it is worth noting the precise order of those selection mechanisms, i.e. a catalytic discrimination step in-between the two selection steps of the kinetic proofreading scheme. This order is based on biological and theoretical constraints. An energy consuming step, such as NTP hydrolysis, had to be placed in the kinetic proofreading scheme to reproduce Hopfield's results. Furthermore, in some enzyme, the ATPase activity largely depends on substrate binding via the induced fit mechanism [22]. Hence the choice of a catalytic discrimination step during the hydrolysis. However, other combinations are possible and have been studied [5, 21].

Selection in biological systems

The mechanisms we highlighted here are relevant to real biological systems, that crucially need both accuracy and velocity. By focusing on two of such systems, namely protein translation by ribosomes and DNA replication by polymerases, we suggest that indeed the combination of kinetic proofreading and catalytic discrimination underlies these biological processes.

Protein synthesis

Protein synthesis denotes the set of reactions allowing the ribosome to synthesize a protein by assembling amino acids in a precise order determined by the corresponding mRNA sequence. The ability to form a functioning protein highly depends on the ribosome accuracy in selecting the correct aminoacyl-tRNA (aa-tRNA) complex, corresponding to the targeted mRNA anti-codon. This selection is a multiple steps process.

First the ribosome binds to the ternary complex formed by aa-tRNA, the elongation factor Tu (EF-Tu) and GTP. This part is described as an initial selection step where only cognate and near cognate codons are accepted. Following the stabilization of the codon, configurational changes lead to GTPase activation allowing hydrolysis of the GTP and the release of EF-Tu. After hydrolysis, depending on the interaction between the tRNA codon and the associated anti-codon [23], the amino acid can either be added to the peptide chain or rejected. This second selection step is called proofreading.

Actually, experimental observations of the tRNA selection pathway kinetics are in agreement with the model proposed here. The ribosome is able to differentiate between cognate, near-cognate and non-cognated codons during initial selection and the proofreading step [23, 24, 25] via a two steps energy-based discrimination (kinetic proofreading), but also during GTP-hydrolysis by catalytic discrimination. Indeed, the interactions between the cognate codon and the correct anti-codon induce changes in the complex configuration which in the end lead to a fast GTPase activation followed by GTP hydrolysis [23]. This recognition, by induced-fit mechanism [24], allows to accelerate the process for cognate codon and slows it down for the near-cognate codon giving more time for the ribosome to release it. Therefore, the high accuracy is due to a

discriminating process where kinetic proofreading and catalytic discrimination coexist, similar to the KP+C model.

DNA replication

During DNA replication, DNA polymerases are responsible for the synthesis of new DNA strands by assembling correct nucleotide bases according to the chosen DNA template.
The DNA polymerase first binds a free nucleotide (dNTP) with its polymerase domain in an open conformation. This initial binding step is depicted as a multi-steps complex mechanism where accuracy of the base is controlled [26, 27]. This verification promotes conformational changes in the polymerase domain which, in turn, allows the formation of the phosphodiester bond and then the incorporation of the nucleotide. Finally, an exonuclease, or the DNA polymerase exonuclease domain, can correct mistakes by removing wrong nucleotide after their incorporation. This last step is widely regarded as a proofreading activity and was already used as a justification in the original work of Hopfield to explain the high fidelity in DNA replication [3]. Indeed, without exonuclease activity, DNA replication fidelity is decreasing from one error every 10^8 to 10^{10} copied base-pairs [28] to one every 10^3 to 10^6 [29] depending on the conditions. However, this also shows that the polymerase activity is actually reaching a very high accuracy by itself. To explain this low error rate, multiple successive selection steps have been proposed. First, the initial binding step is able to reject wrong nucleotide due to a 10 to 100-fold higher affinity for the right nucleotide [30, 29]. This selection is understood to depend more on a "geometric fit" than on differences in the free energy between binding to the right and wrong nucleotide [31], but still follows the principle of the kinetic proofreading scheme. The following conformational changes leading to chemistry step of phosphodiester bond formation are promoted by an induced-fit mechanism [26], which is more that 10000 times slower for wrong nucleotide [29]. Finally, there is evidence of selection during the chemistry step [32], where the phosphodiester bond would be formed faster in presence of the correct nucleotide. While it is not yet clear if both the induced-fit mechanism and the chemistry step are able to discriminate between right and wrong nucleotide during DNA replication [26], we can consider both of them as catalytic discrimination steps, where only the initial one matters [21]. In conclusion, we once again find a strong hint that a selection mechanism relying on the same process and in the same order than our KP+C model is at work.

Conclusion

In this work, we proposed a non-equilibrium, thermodynamically consistent model of substrate selection and processing. This allows rigorously accounting for the energy which is available and which is driving the system away from equilibrium. In this framework, we tested selection based either on Hopfield's kinetic proofreading [3] or on catalytic discrimination [5, 12], or both.

The kinetic proofreading is very effective in terms of fidelity, however the optimal fidelity comes at the price of slowing down the process by orders of magnitude. This can be compensated by the catalytic-based selection process, whose fidelity is not as high as the kinetic proofreading, but has no intrinsic velocity limitations. Including both these selection mechanisms in our model allows it to benefit from both of their advantages. Indeed, their combination allows simultaneously attaining high fidelity and fast production rate.
Given the ability of evolution to explore a vast array of possibilities, and retain the advantageous ones, it may not come as a surprise that some cellular processes exploit such a combination. Indeed, both protein translation and DNA replication, when coarse-grained to their essential selection steps, display both kinetic proofreading and catalytic discrimination steps in the same order as we suggest in our theoretical framework.

Appendices

1 Mathematical development

1.1 Energy dependence and equilibrium conditions

Tuning our model out of equilibrium allows us to take advantage of the imposed directionality of reactions in the our model due to the energy consuming steps. To highlight how (and how much) the system becomes directional upon energy consumption, it is instructive to focus on two cycles from Fig. 1: $E \to ER \to E^*R \to E^* \to E$ and its reverse, $E \to E^* \to E^*R \to ER \to E$. If the step between ER and E^*R takes place over the exchange reactions (dashed double arrows in figure Fig. 1) no NTP molecule is transformed into a NDP molecule, hence no energy is consumed and the cycle must be balanced. If instead that particular step of the cycle is associated with the hydrolysis/synthesis reactions, energy consumption is at play. The unbalance between the two different ways to travel the cycle gives rise to clockwise (or anti-clockwise) directionality. This argument is made more precise using explicitly the reaction rates.
Indeed, by considering the nucleotide-less state of the enzyme ($\bar{E}$ on Fig. A.1) as very short-lived and poorly populated (this approximation is useful to somewhat simplify the reaction scheme, but is not necessary for the conclusions of this work), the rates of the exchange reactions can be written as

$$\begin{aligned} k_X^{TD}(\alpha) &= k_X^{T-} \frac{k_X^{D+}[NDP]}{k_X^{D+}[NDP] + k_X^{T+}[NTP]} \\ &= k_X^{T-} \frac{k_X^{D+}}{k_X^{D+} + \alpha k_X^{T+}} \qquad \text{(A.1)} \\ k_X^{DT}(\alpha) &= k_X^{D-} \frac{k_X^{T+}[NTP]}{k_X^{D+}[NDP] + k_X^{T+}[NTP]} \\ &= k_X^{D-} \frac{\alpha k_X^{T+}}{k_X^{D+} + \alpha k_X^{T+}} \qquad \text{(A.2)} \end{aligned}$$

where k_X^{T-} and k_X^{T+} are the unbinding and binding rates of NTP, respectively (and similarly for the binding/unbinding rates of NDP), and $\alpha = [NTP]/[NDP]$. The subscript X indicate the presence of a substrate ($X = R, W$), a product ($X = P_r, P_w$) or nothing.
Their ratio is then

$$\frac{k_X^{DT}(\alpha)}{k_X^{TD}(\alpha)} = \alpha \frac{k_X^{T+} k_X^{D-}}{k_X^{T-} k_X^{D+}} \quad . \qquad \text{(A.3)}$$

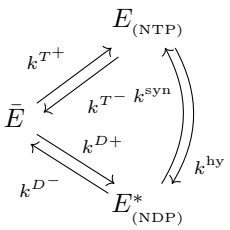

Figure A.1: Details of the transitions between the enzyme in a NTP state (E) and in an NDP state (E^*) with an intermediate state $\bar{E}$ where the enzyme is not bound to any nucleotide (apo-state).

The ratio between the synthesis and hydrolysis rate can be obtained instead by considering the binding and unbinding of NTP and NDP to the enzyme, and their catalytic interconversion while bound, in an equilibrium solution of NTP and NDP: $([NTP]/[NDP])_{eq} = \alpha_{eq}$ (in the case of ATP and ADP, depending on the solution condition $\alpha_{eq} \simeq 10^{-8} - 10^{-6}$, whereas in cellular conditions $\alpha \simeq 1 - 10$ [11]). Applying detailed balance, it can be derived that

$$\frac{k_S^{\mathrm{syn}}}{k_S^{\mathrm{hy}}} = \alpha_{eq} \frac{k_S^{T+} k_S^{D-}}{k_S^{T-} k_S^{D+}} \quad , \qquad \text{(A.4)}$$

which also gives $k_S^{DT}(\alpha)/k_S^{TD}(\alpha) = (\alpha/\alpha_{eq})\, k_S^{\mathrm{syn}}/k_S^{\mathrm{hy}}$. Here we are using the subscript $S = R, W$ as the hydrolysis and synthesis only happen when the enzyme is bound to a substrate.
In equilibrium conditions ($\alpha/\alpha_{eq} = 1$) the currents flowing from ES to E^*S along the exchange and hydrolysis/synthesis branches are expectedly equal and opposite, so that there is no net current. Away from equilibrium ($\alpha/\alpha_{eq} > 1$), as in cellular conditions, and if k_S^{hy} is high (as for most NTPase enzymes), the current over the

hydrolysis/synthesis branch exceeds the opposite exchange current, and a net flux, which captures directionality, is then present.

1.2 Resolution of the model

Let N_i be the i-th state of our system. The master equation describing the time evolution of its concentration is given by:

$$\frac{d[N_i]}{dt} = \sum_{j \neq i} k_{ji}[N_j] - k_{ij}[N_i] \tag{A.5}$$

where the k_{ij} are the transition rate from N_i to N_j, for all j.
The whole set of equations can therefore be written as

$$\frac{d\vec{N}}{dt} = M\vec{N} \tag{A.6}$$

where $\vec{N}$ is a column vector containing the concentration of each state and M is the rate matrix. At steady state, the concentrations of each states are found by solving

$$M\vec{N}_{ss} = 0 \tag{A.7}$$

From the steady state solution, we can compute the production fluxes

$$J_r = [EP_r]k^{\text{off}}_{P_r} + [E^*P_r]k^{\text{off}}_{P^*_r} - [E][P_r]k^{\text{on}}_{P_r} - [E^*][P_r]k^{\text{on}}_{P^*_r} \tag{A.8}$$

$$J_w = [EP_w]k^{\text{off}}_{P_w} + [E^*P_w]k^{\text{off}}_{P^*_w} - [E][P_w]k^{\text{on}}_{P_w} - [E^*][P_w]k^{\text{on}}_{P^*_w} \tag{A.9}$$

where k^{on}_X and k^{off}_X correspond to the binding and unbinding rates of the products, $X = P_r$ or P_w.
and the error rate

$$f = \frac{J_w}{J_r} \tag{A.10}$$

J_r and f characterize the system accuracy and velocity, giving us a base to compare different selection methods.

2 Rates for the simulations

All rates in this work were carefully chosen to respect detailed balance at equilibrium ($\alpha = \alpha_{eq}$), as stated above. Therefore, among the following rates, some were freely chosen while others were constrained by the equilibrium condition.

2.1 Kinetic proofreading case

Numerical values to obtain results on Fig. 3.

Conc. [μM]		On rates [$s^{-1}\mu\text{M}^{-1}$]		Off rates [s^{-1}]					
$[R]$	$[W]$	k^{on}_X	$k^{\text{on}}_{X^*}$	k^{off}_R	k^{off}_W	$k^{\text{off}}_{R^*}$	$k^{\text{off}}_{W^*}$	$k^{\text{off}}_{P_s}$	$k^{\text{off}}_{P^*_s}$
1	1	95	10^{-3}	45	450	84	840	10^{-5}	10^3

Table A.1: Values of the different on and off rates used in Hopfield's like scheme. $X = R, W, P_r, P_w$ and $P_s = P_r, P_w$.

Production rates [s^{-1}]		"Reverse production" rates [s^{-1}]		Hydrolysis rate [s^{-1}]	Synthesis rate [s^{-1}]
k_{RP}	k_{WP}	k_{PR}	k_{PW}	k^{hy}_S	k^{syn}_S
1	1	1.2 $\times 10^{-5}$	1.2 $\times 10^{-6}$	varies	varies

Table A.2: Rates of products formation and destruction and NTP hydrolysis and synthesis in a Hopfield's like scheme simulation

Enzyme bound to	"On rates $[s^{-1}\mu\mathrm{M}^{-1}]$ " k^{D+}	k^{T+}	"Off rates $[s^{-1}]$" k^{D-}	k^{T-}
$\varnothing$	10	10	10	10
S	0.0018	5.7×10^{4}	10^{-5}	0.0018
P_s	2×10^{-5}	4.6×10^{6}	10^{-10}	2×10^{-5}

Table A.3: Rates of binding and unbinding NTP and NDP for Hopfield's kinetic proofreading scheme

2.2 Catalytic discrimination case

Numerical values to obtain results on Figs. 5 and 6.

Conc. $[\mu\mathrm{M}]$		On rates $[s^{-1}\mu\mathrm{M}^{-1}]$		Off rates $[s^{-1}]$			
$[R]$	$[W]$	k_X^{on}	$k_{X^*}^{\mathrm{on}}$	k_S^{off}	$k_{S^*}^{\mathrm{off}}$	$k_{P_s}^{\mathrm{off}}$	$k_{P_s^*}^{\mathrm{off}}$
1	1	95	10^{-3}	varies	84	10^{-5}	1000

Table A.4: Values of the different on and off rates used in the catalytic discrimination scheme. $X = R, W, P_r, P_w$ and $P_s = P_r, P_w$.

Production rates $[s^{-1}]$		"Reverse production" rates $[s^{-1}]$		Hydrolysis rate $[s^{-1}]$		Synthesis rate $[s^{-1}]$	
k_{RP}	k_{WP}	k_{PR}	k_{PW}	k_R^{hy}	k_W^{hy}	k_R^{syn}	k_S^{syn}
10	10	1.2×10^{-4}	1.2×10^{-4}	varies	varies	$\frac{(7.98k_R^{hy})}{k_S^{\mathrm{off}}}$	$\frac{(7.98k_W^{hy})}{k_S^{\mathrm{off}}}$

Table A.5: Rates of products formation and destruction and NTP hydrolysis and synthesis for simulation of the catalytic discrimination

Enzyme bound to	"On rates $[s^{-1}\mu\mathrm{M}^{-1}]$ " k^{D+}	k^{T+}	"Off rates $[s^{-1}]$" k^{D-}	k^{T-}
$\varnothing$	10	10	10	10
S	$5\times10^{-4}(k_S^{\mathrm{off}})^{1/3}$	$2\times10^{-5}(k_S^{\mathrm{off}})^{-1/3}$	10^{-5}	$5\times10^{-4}(k_S^{\mathrm{off}})^{1/3}$
P_s	2×10^{-5}	4.6×10^{6}	10^{-10}	2×10^{-5}

Table A.6: Rates of binding and unbinding NTP and NDP for simulation of the catalytic discrimination

2.3 Comparison

Numerical values for the comparison between the two selection schemes on Fig. 7.

Conc. $[\mu\mathrm{M}]$		On rates $[s^{-1}\mu\mathrm{M}^{-1}]$		Off rates $[s^{-1}]$					
$[R]$	$[W]$	k_X^{on}	$k_{X^*}^{\mathrm{on}}$	k_R^{off}	k_W^{off}	$k_{R^*}^{\mathrm{off}}$	$k_{W^*}^{\mathrm{off}}$	$k_{P_s}^{\mathrm{off}}$	$k_{P_s^*}^{\mathrm{off}}$
1	1	100	10^{-3}	4.5	$k_R^{\mathrm{off}} \times \lambda$	8.4	$k_{R^*}^{\mathrm{off}} \times \lambda$	10^{-5}	1000

Table A.7: Values of the different on and off rates used for the comparison between kinetic proofreading and catalytic discrimination. $X = R, W, P_r, P_w$ and $P_s = P_r, P_w$. λ varies from 1 to 100, tuning the kinetic proofreading steps.

Production rates $[s^{-1}]$		"Reverse production" rates $[s^{-1}]$		Hydrolysis rate $[s^{-1}]$		Synthesis rate $[s^{-1}]$	
k_{RP}	k_{WP}	k_{PR}	k_{PW}	k_R^{hy}	k_W^{hy}	k_R^{syn}	k_W^{syn}
100	100	0.012	$0.012/\lambda$	$k_W^{hy} \times \omega$	0.1	$0.19\ k_R^{hy}$	0.019

Table A.8: Rates of products formation and destruction and NTP hydrolysis and synthesis, λ varies from 1 to 100, tuning the kinetic proofreading steps and ω varies from 1 to 1000, tuning the catalytic discrimination step.

Enzyme bound to	"On rates $[s^{-1}\mu\mathrm{M}^{-1}]$ "		"Off rates $[s^{-1}]$"	
	k^{D+}	k^{T+}	k^{D-}	k^{T-}
$\varnothing$	10	10	10	10
S	0.0018	5.7×10^{4}	10^{-5}	0.0018
P_s	2×10^{-5}	4.6×10^{6}	10^{-10}	2×10^{-5}

Table A.9: Rates of binding and unbinding NTP and NDP for comparison between kinetic proofreading and catalytic discrimination

References

[1] Hani S. Zaher and Rachel Green. Fidelity at the molecular level: lessons from protein synthesis. *Cell*, 136(4):746–762, 2009.

[2] Kenneth A Johnson. Conformational coupling in DNA polymerase fidelity. *Annual Review of Biochemistry*, 62(1):685–713, 1993.

[3] J. J. Hopfield. 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.

[4] J. Ninio. Kinetic amplification of enzyme discrimination. *Biochimie*, 57(5):587–595, 1975.

[5] Pablo Sartori and Simone Pigolotti. Kinetic versus Energetic Discrimination in Biological Copying. *Physical Review Letters*, 110(18):188101, 2013.

[6] Arvind Murugan, David A. Huse, and Stanislas Leibler. Speed, dissipation, and error in kinetic proofreading. *Proceedings of the National Academy of Sciences*, 109(30):12034–12039, 2012.

[7] Arvind Murugan, David A Huse, and Stanislas Leibler. Discriminatory Proofreading Regimes in Nonequilibrium Systems. *Physical Review X*, 4(2):021016, 2014.

[8] Kinshuk Banerjee, Anatoly B. Kolomeisky, and Oleg A. Igoshin. Elucidating interplay of speed and accuracy in biological error correction. *Proceedings of the National Academy of Sciences*, 114(20):5183–5188, 2017.

[9] William D. Piñeros and Tsvi Tlusty. Kinetic proofreading and the limits of thermodynamic uncertainty. *Physical Review E*, 101(2):022415, 2020. Publisher: American Physical Society.

[10] Joel D. Mallory, Anatoly B. Kolomeisky, and Oleg A. Igoshin. Trade-Offs between Error, Speed, Noise, and Energy Dissipation in Biological Processes with Proofreading. *The Journal of Physical Chemistry B*, 123(22):4718–4725, 2019.

[11] Ron Milo and Rob Phillips. *Cell biology by the numbers*. Garland Science, Taylor & Francis Group, New York, NY, 2016.

[12] Charles H. Bennett. Dissipation-error tradeoff in proofreading. *Biosystems*, 11(2):85–91, 1979.

[13] Riccardo Rao and Luca Peliti. Thermodynamics of accuracy in kinetic proofreading: dissipation and efficiency trade-offs. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(6):P06001, 2015.

[14] Ganhui Lan, Pablo Sartori, Silke Neumann, Victor Sourjik, and Yuhai Tu. The energy–speed–accuracy trade-off in sensory adaptation. *Nature Physics*, 8(5):422–428, 2012.

[15] Magnus Johansson, Martin Lovmar, and Måns Ehrenberg. Rate and accuracy of bacterial protein synthesis revisited. *Current Opinion in Microbiology*, 11(2):141–147, 2008.

[16] Ingo Wohlgemuth, Corinna Pohl, and Marina V Rodnina. Optimization of speed and accuracy of decoding in translation. *The EMBO Journal*, 29(21):3701–3709, 2010.

[17] Davide Chiuchiú, Yuhai Tu, and Simone Pigolotti. Error-Speed Correlations in Biopolymer Synthesis. *Physical Review Letters*, 123(3):038101, 2019.

[18] Grégory Boël, Olivier Danot, Victor de Lorenzo, and Antoine Danchin. Omnipresent Maxwell's demons orchestrate information management in living cells. *Microbial Biotechnology*, 12(2):210–242, 2019.

[19] Solange Flatt, Daniel M. Busiello, Stefano Zamuner, and Paolo De Los Rios. ABC Transporters are billion-year-old Maxwell Demons. *bioRxiv*, 2021.

[20] Wayne A. Hendrickson. Theory of Allosteric Regulation in Hsp70 Molecular Chaperones. *QRB Discovery*, 1:e7, 2020.

[21] Pablo Sartori and Simone Pigolotti. Protocols for Copying and Proofreading in Template-Assisted Polymerization. *Journal of Statistical Physics*, 162(5):1167–1182, 2016.

[22] Daniel E. Koshland Jr. The Key–Lock Theory and the Induced Fit Theory. *Angewandte Chemie International Edition in English*, 33(23-24):2375–2378, 1995.

[23] Scott C. Blanchard, Ruben L. Gonzalez, Harold D. Kim, Steven Chu, and Joseph D. Puglisi. tRNA selection and kinetic proofreading in translation. *Nature Structural & Molecular Biology*, 11(10):1008–1014, 2004.

[24] Tillmann Pape, Wolfgang Wintermeyer, and Marina Rodnina. Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *The EMBO Journal*, 18(13):3800–3807, 1999.

[25] Kirill B. Gromadski and Marina V. Rodnina. Kinetic Determinants of High-Fidelity tRNA Discrimination on the Ribosome. *Molecular Cell*, 13(2):191–200, 2004.

[26] Chithra Hariharan, Linda B. Bloom, Sandra A. Helquist, Eric T. Kool, and Linda J. Reha-Krantz. Dynamics of Nucleotide Incorporation: Snapshots Revealed by 2-Aminopurine Fluorescence Studies. *Biochemistry*, 45(9):2836–2844, 2006. Publisher: American Chemical Society.

[27] Daniel R. Garalde, Christopher A. Simon, Joseph M. Dahl, Hongyun Wang, Mark Akeson, and Kate R. Lieberman. Distinct Complexes of DNA Polymerase I (Klenow Fragment) for Base and Sugar Discrimination during Nucleotide Substrate Selection*. *Journal of Biological Chemistry*, 286(16):14480–14492, 2011.

[28] Anna Bębenek and Izabela Ziuzia-Graczyk. Fidelity of DNA replication—a matter of proofreading. *Current Genetics*, 64(5):985–996, 2018.

[29] Thomas A. Kunkel and Katarzyna Bebenek. DNA Replication Fidelity. *Annual Review of Biochemistry*, 69(1):497–529, 2000.

[30] Elizabeth Fidalgo da Silva, Subhrangsu S. Mandal, and Linda J. Reha-Krantz. Using 2-Aminopurine Fluorescence to Measure Incorporation of Incorrect Nucleotides by Wild Type and Mutant Bacteriophage T4 DNA Polymerases. *Journal of Biological Chemistry*, 277(43):40640–40649, 2002.

[31] Myron F. Goodman. Hydrogen bonding revisited: Geometric selection as a principal determinant of DNA replicationfidelity. *Proceedings of the National Academy of Sciences*, 94(20):10493–10495, 1997.

[32] Jan Florián, Myron F. Goodman, and Arieh Warshel. Computer simulation of the chemical catalysis of DNA polymerases: discriminating between alternative nucleotide insertion mechanisms for T7 DNA polymerase. *Journal of the American Chemical Society*, 125(27):8163–8177, 2003.