Many Will Enter, Few Will Win: Cost and Sensitivity of Exploratory Dynamics

Elena F. Koslover,[1] Milo M. Lin,[2] and Rob Phillips[3]

[1] *Department of Physics, University of California, San Diego, La Jolla, CA 92093, USA*
[2] *Green Center for Systems Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390*
[3] *Department of Physics and Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125*

A variety of biomolecular systems rely on exploratory dynamics to reach target locations or states within a cell. Without a mechanism to remotely sense and direct motion towards a target, the system must sample over many paths, often including resetting transitions back to the origin. We explore how exploratory dynamics can confer an important functional benefit: the ability to respond to small changes in parameters with large shifts in the steady-state behavior. However, such enhanced sensitivity comes at a cost: resetting cycles require energy dissipation in order to push the system out of its equilibrium steady state. We focus on two concrete examples: translational proofreading in the ribosome and microtubule length control via dynamic instability to illustrate the trade-offs between energetic cost and sensitivity. In the former, a thermodynamically driven activation step enhances the ability to distinguish between substrates and decoys with small binding energy differences. In the latter, resetting cycles enable catalytic control, with the steady-state length distribution modulated by sub-stoichiometric concentrations of a reusable catalyst. Synthesizing past models of these well-studied systems, we show how path-counting and circuit mapping approaches can be used to address fundamental questions such as the number of futile cycles inherent in translation and the steady-state length distribution of a dynamically unstable polymer. In both cases, a limited amount of thermodynamic driving is sufficient to yield a qualitative transition to a system with enhanced sensitivity, enabling accurate discrimination and catalytic control at a modest energetic cost.

It is our great pleasure to contribute to this special issue dedicated to the life and work of Prof. Erich Sackmann. Sackmann was a pioneer in the broad field dedicated to using the tools of physics to understand the fascinating phenomena of life. One of the most intriguing aspects of living organisms that makes them so different from their inanimate counterparts is the expenditure of energy to maintain nonequilibrium steady states. Many of the phenomena that exploit such energy consumption share features of exploratory dynamics: sampling many paths with occasional resetting en route to a target state. Common examples include biological polymerization in processes such as replication, translation, and cytoskeletal filament dynamics. In this paper, we examine two case studies (protein translation and microtubule growth) linking the energetic cost and functional benefits of exploratory dynamics.

INTRODUCTION

One of the fundamental challenges faced by living cells is the need to carry out their functions while 'blind' – lacking a centralized omniscient organizer that can direct cellular components where to go and which interactions to perform in what order. Above the cellular scale, biological dynamics can be at least somewhat directed: individual cells follow gradients of chemotactic signals to find food or escape danger [1], morphogen patterns regulate tissue rearrangements and cellular differentiation [2], organisms make use of sensory organs to find distant targets. Within a cell, however, exploratory dynamics reigns supreme [3]. Whether searching through physical space or through chemical state-space, cells rely on dynamics that follow a characteristic pattern: blindly sampling through the available states, fixing and amplifying the targets when a trajectory stumbles across them, and resetting unsuccessful trajectories to try again.

Such exploratory dynamics was first noted in the context of dynamically unstable microtubules that engage in search-and-capture cycles of random growth and rapid depolymerization to find chromosomes when constructing the mitotic spindle [4, 5]. The same principle applies to other spatial search problems in the cell. For example, the ability of neurons to localize mitochondria in regions of high energy demand [6] relies on 'sushi-belt' dynamics [7], where a motile population of organelles constantly cycles via motor-driven transport through neuronal projections, with regions of enhanced stopping encoded through local increases in calcium [8] or glucose concentrations [9]. In the endocytic pathway, endosomes bearing activated receptors must wander through the cellular periphery until they encounter other organelles that trigger phosphoinositide conversion and deactivate the signal [10, 11].

Analogous exploratory processes occur on the molecular scale, particularly in the context of sensing and proofreading. First introduced to explain the accuracy of polymerization-based information transfer [12, 13], kinetic proofreading relies on chaining together intermediate states that the system must pass through before reaching a target. By providing multiple opportunities to reset back to the origin from each intermediate state, such proofreading can amplify the likelihood of following a pathway involving 'right' versus 'wrong' interactions. Thus, a ribosome searches for the next amino acid to add

to a growing polypeptide chain by exploring through intermediate states that might involve the right or wrong tRNA. The higher rate of resetting (release of the tRNA) for the wrong amino acid allows a greater probability that the final step of peptide elongation is reached only with the correct amino acid. Similar exploration through intermediate states shows up in sensory systems, such as T-cell activation [14] or chemotactic signaling [15]. In both cases, incorrect ligands that bind weakly to receptors are more likely to be released, resulting in resetting during each intermediate step. Proofreading through resetting is also thought to contribute to the self-assembly of large multimeric structures, including sequence-specific RecA filament formation on DNA [16] and viral RNA packaging [17].

Protein quality control systems provide additional examples of exploratory dynamics that leverage resetting to accurately sort components into distinct pathways. The ubiquitinating enzyme APC is able to distinguish its substrates among myriads of decoy proteins by sequentially marking multiple lysine groups on the target protein, eventually triggering the degradation of the substrate [18]. The resulting difference in APC binding affinity on ubiquitinated substrates versus non-ubiquitinated decoys makes it more likely that only the correct proteins are targeted for degradation. In the secretory pathway, newly translated proteins are tagged by the addition of glycan chains that facilitate binding to chaperones which help fold the proteins [19]. Multiple cycles of glycosylation in the ER allow chaperones to make several attempts at folding a nascent protein before it proceeds towards the terminal pathways of export or degradation [20].

Given the prevalence of resetting dynamics in intracellular systems, a natural question is why this approach is so common and what advantages it might offer to the cell. The effects of resetting on speed in subcellular exploration are considered in previous work [21, 22] and addressed in a cohesive framework in another article within this issue [23].

Instead of considering the temporal advantages, here we illustrate how exploratory dynamics enhances sensitivity: the ability of the cell to respond to small differences in system parameters. We contend that seemingly distinct functional objectives, including concerted activation of specific signaling molecules, accurate discrimination of targets from decoys, and regulation of molecular assembly size, are all manifestations of the same phenomenon: namely that exploratory dynamics magnifies the effect of changing system parameters (inputs) on steady-state observables (outputs). Such sensitivity comes at a cost of energy expenditure because exploratory dynamics with resetting necessarily involves cycles that break detailed balance. Using kinetic proofreading and microtubule length regulation as examples, we show how signal gain depends on energy dissipation and the concomitant driving of systems away from equilibrium detailed balance toward non-equilibrium exploratory dynamics. Importantly, we show that energy expenditure through driving one transition can qualitatively change the response function to parameters elsewhere in the system.

One manifestation of this phenomenon is catalytic control – a ubiquitous feature of biomolecular regulatory systems, where a reusable catalyst (such as a kinase) alters the state of its substrates. Because a catalyst modifies transition barriers only, the input parameter (catalyst concentration) can have no effect on steady-state output under equilibrium conditions. However, in the presence of driven exploratory dynamics, the steady-state probability of different substrate states can in fact become dependent on the level of catalyst present. This feature enables a small (sub-stoichiometric) number of regulatory molecules to trigger large-scale changes in cellular state. Examples include the whole-sale ubiquitination and degradation of specific classes of proteins during mitosis [18] and the catalytic regulation of microtubule length during cell division [24]. Such catalytic control requires energy dissipation somewhere in the system, even when the catalysis step itself does not couple to an external energy source.

In this paper, we investigate how exploratory dynamics with resetting allows biochemical systems to enhance their sensitivity, enabling accuracy in the presence of decoys and catalytic control. For concreteness, we focus on two key biological examples that illustrate these features: single-step kinetic proofreading in ribosomal translation, and multi-step resetting in microtubules undergoing dynamic instability. Along the way, we highlight two pedagogically useful approaches to describing such systems: path-counting (which intuitively incorporates exploratory dynamics), and circuit mapping (which clarifies the relation between energetic driving and steady-state distributions). These systems highlight both the functional benefits of resetting dynamics and the concomitant cost in energy dissipation.

KINETIC PROOFREADING IN THE RIBOSOME

The notion of accuracy can be defined for stochastic reaction systems where there are multiple terminal states of which only a particular subset is considered 'correct'. In many biological systems, accuracy is enhanced through the use of multiple intermediate states, with the probability of resetting to the origin higher for pathways leading to the wrong terminal state – a process that has been termed 'kinetic proofreading' [12, 14, 25–27]. At equilibrium, the error rate (ratio of wrong to right pathways selected) is bounded by the difference in free energy change from the initial to the final state [28]. By contrast, systems with actively driven resetting cycles can surpass this bound, combining the energy differences in multiple intermediate states to achieve higher fidelity [29].

The distinction between kinetic and thermodynamic proofreading [28, 30] can be concretely illustrated by two distinct biomolecular copying processes. In the case of

DNA replication, the polymerase need only distinguish between four options for the next nucleic acid to be added to the chain. DNA polymerases that are deficient in proofreading exonuclease activity (which removes mismatched nucleotides after they are incorporated into the chain) are nevertheless able to achieve error rates on the order of 10^{-5}. Their fidelity relies on a combination of steric shape-matching for the correct nucleotide in the polymerase binding pocket and the classic base-pairing hydrogen bonds [31, 32]. By contrast, there are 20 possible tRNAs that can bind to the ribosome during translation, only one of which is cognate to a given codon on the mRNA and would lead to incorporation of the correct amino acid. The difference in binding energies for distinct anti-codons relies on the formation of 1-2 hydrogen bonds and is thus limited to only a few k_bT [33]. For a 20-fold excess of wrong tRNAs, this implies an error rate well above 10%. On the scale of a 300-amino-acid long peptide, being able to build an error-free chain even half of the time would require error rates below 0.2%, necessitating the introduction of energy-consuming kinetic proofreading cycles.

A number of theoretical works have sought to elucidate the connections between energy dissipation, speed, and accuracy of a proofreading system [27, 29, 34]. Even for a simple single-step enzyme, accurately distinguishing between cognate and non-cognate substrates requires product formation to be slow compared to the binding-unbinding equilibration [35]. From the same principle, the forward step in multi-stage proof-reading pathways must be arbitrarily slow to reach the minimal possible error rate, highlighting a trade-off between speed and accuracy [36]. Furthermore, because kinetic proofreading relies on resetting, each such resetting cycle implies a cost in terms of additional time required to reach the target. For multi-step pathways, the time to reach a target scales exponentially with the number of states if resetting is more likely than forward stepping, and linearly otherwise [27]. Since each intermediate state provides an extra opportunity for proofreading, accuracy comes at a cost in speed. This trade-off has also been noted in models of T-cell activation [14] and chemosensory receptor arrays [15, 34].

The relevant energetic cost for a proofreading system depends on both the dissipation per turn of the resetting cycle and the typical number of such cycles before reaching the target state. Thus, for translational proofreading we could ask how many tRNAs are released from the activated state before an amino acid is successfully incorporated into the growing polypeptide chain. In other words, given the large excess (i.e.≈20-fold) concentration of wrong versus right tRNAs, how many GTP must be hydrolyzed per elongation event? This question has been formulated in terms of the total entropy production (dissipation rate normalized by the incorporation rate) [28], and in terms of the number of futile cycles [29]. Some fundamental bounds have been proposed for relating the dissipation (including waste) and the error rate of a proofreading system. At equilibrium, the error rate is equal to the ratio of Boltzmann factors for the right versus wrong products [37]. For non-equilibrium proofreading systems, the error rate can be driven up or down by an exponential factor incorporating both the total entropy production and the 'excess work' put into the system beyond the overall free energy difference for incorporation [28]. For multistage proofreading schemes, the minimal energy cost necessary to sustain a particular error rate decreases with both the number of intermediate states and the right-vs-wrong discrimination factor for resetting from each intermediate state [29].

These generalized results, however, can be difficult to apply to specific proofreading schemes, particularly if there are constraints on certain kinetic rates in the system. In particular, the reverse transitions associated with resetting can be so rare that they are never observed experimentally and cannot be directly measured. Despite being in some sense unphysical, models with irreversible arrows often provide a complete description of actual quantitative measurements for a reaction circuit [38]. Irreversible arrows then imply that the reverse transitions are infrequent enough to have little effect on the partition ratios of the system [29]: the choice of transition at each step in the exploratory dynamics.

Below we seek to provide a pedagogically helpful analysis of a classic single-intermediate model of translational proofreading. We begin with a model incorporating irreversible transitions, using a path-counting approach to compute the number of excess GTP hydrolyzed per elongation event. More complex reaction schemes can also be reduced to this simple model. We then expand the path-counting approach to incorporate fully reversible steps and demonstrate how increasing energetic driving causes the system to transition between distinct regimes, with intermediate driving strength but large numbers of futile cycles required to approach the greatest accuracy. Overall, active driving enhances the sensitivity of translational proofreading to small changes in tRNA binding energy, enabling it to accurately discriminate between correct and wrong amino acids for incorporation into the peptide chain.

Irreversible Model

As shown in Fig. 1, we represent translational elongation via a classic simplified reaction scheme, as proposed by Hopfield in his seminal work on kinetic proofreading [12]. This scheme begins with an empty ribosomal binding site (denoted as the R state). A reversible binding step allows a tRNA loaded with the correct (C) or wrong (W) amino acid to interact with the template strand. The wrong amino acids are expected to be present in excess of the correct ones, by a factor g. On average, we would expect an excess on the order of $g \approx 20$ (BNID 108611,105274, 105273). For simplicity, we assume that the tRNA arrival is diffusion-limited, so that

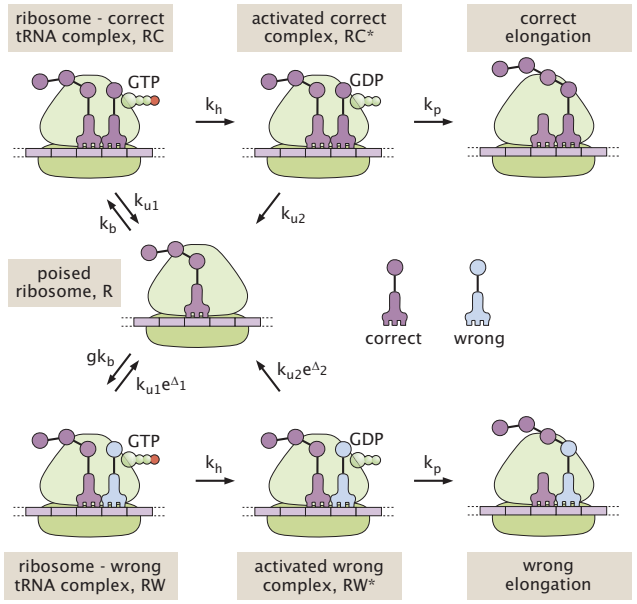


FIG. 1. **Schematic of the classic 2-step kinetic proofreading model for translation.** The states are: R = empty ribosome, RC = ribosome with correct amino acid bound, RW = ribosome with wrong amino acid bound, RC^*, RW^* = ribosome with correct or wrong amino acid, after GTP hydrolysis.

the binding rate for the correct amino acid is k_b and the rate for the wrong amino acid is gk_b. Even in the absence of proofreading, we would expect some discrimination between tRNAs carrying right and wrong amino acids based on their different binding energies. The thermodynamic limit on accuracy [28] is then set by the equilibrium ratio of wrong versus correct tRNAs bound to the ribosome, expressed as $f_{\rm passive} = ge^{-\beta\Delta\varepsilon}$, where $\Delta\varepsilon = \varepsilon_w - \varepsilon_c$ is the difference in binding energies, and we adopt the notation $\beta = 1/k_BT$. For much of what follows, we will use the convention of defining dimensionless energies such as $\Delta_1 = \beta\Delta\varepsilon$. Given these conventions, we set the off-rates for the reversible interaction to be k_{u1} and $k_{u1}e^{\Delta_1}$ for the correct and wrong amino acids, respectively.

In a somewhat whimsical analogy (illustrated in Fig. 2), we can describe translational elongation as entry of visitors to a particularly selective clubhouse. The first reversible binding step might then correspond to a swinging door with a passive sign declaring who can come in. This passive filter allows for some discrimination, but does not completely keep out unwelcome visitors who might sneak through the swinging door. As in protein translation, a subsequent active step to check the visitors' identity is needed.

The next step in the translational elongation pathway involves the hydrolysis of GTP (in the EF-Tu cofactor associated with the tRNA). This hydrolysis serves as a tightly-coupled energy source, transitioning the ribosome to an activated state (RC* or RW*). We assume the same hydrolysis rate k_h regardless of which amino acid is present. Because of the large free energy change associated with GTP hydrolysis and phosphate release, this step is taken to be effectively irreversible.

The activated state provides an opportunity for proofreading in that the tRNA can again disassociate from the ribosome, essentially serving as a reset in the overall exploratory dynamics. The release process is also discriminatory, with the correct tRNA falling off at rate k_{u2} and the wrong one at rate $k_{u2}e^{\Delta_2}$. The quantity Δ_2 can correspond to either the difference in binding energy between the right and wrong tRNAs, or to a difference in the barrier heights for dissociation [30]. Either way, we assume the activated state is so high on the energy landscape that the dissociation process is irreversible. For the dissociated tRNA to return to the ribosome, it must again pass through the reversibly bound state.

In our analogy, the proofreading step corresponds to an active identity check at the inner door of the clubhouse (Fig. 2, rightmost panels). Such a step is costly in that it requires an energy-consuming "Maxwell's Demon" to open the inner door selectively for the correct visitors. However, it has the advantage of more accurately vetting which visitors are allowed to enter the clubhouse. The sequential passive then active filtering steps, allow for high accuracy to be achieved without overworking the demon, since the number of undesired visitors attempting to sneak through the swinging door is already reasonably low.

Finally, there is an irreversible step for forming the peptide bond to incorporate the new amino acid into the growing peptide chain. We assume this step occurs with rate k_p, regardless of the amino acid identity. In our analogy, this corresponds to the final step of visitors being permanently sworn into the exclusive club.

Model solution

The kinetic scheme illustrated in Fig. 1 represents a network of Markovian transitions between states. Steady-state fluxes along the one-way arrows could be found by solving the corresponding master equations [29]. However, for purposes of conceptual clarity, we take the alternate approach of treating this as a stochastic fluctuating system [39], imagining the ribosome hopping between the different states. We are interested in the average cost per amino acid added to the nascent polypeptide chain, in particular as a result of the large excess in "wrong" tRNA/amino acid pairs. More precisely, starting in the empty state (R), how many hydrolysis events must occur before the system reaches a terminal state, elongating the peptide chain by an additional amino acid? This provides a measure of the energetic cost per amino acid for building a nascent peptide chain. The energetic cost can then be compared to the fidelity of translation, expressed as the error rate f (ratio of wrong versus correct amino acids incorporated into the chain).

As the system hops between the discrete states, there is a choice at each step of which outward arrow to follow.

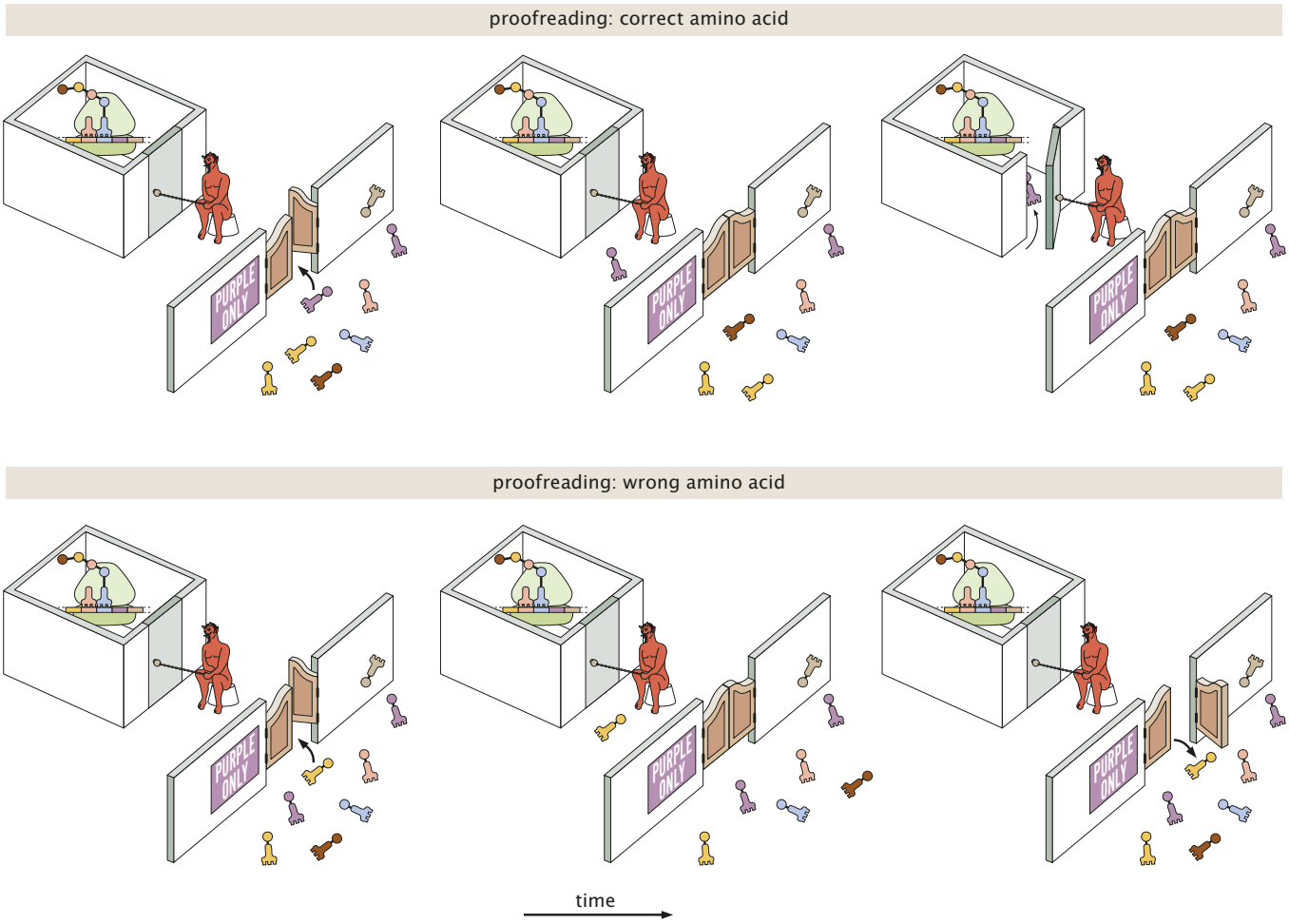


FIG. 2. **Pictorial description of the two-step process whereby correct and wrong tRNA-amino acid pairs are distinguished.** The first step is passive and involves a preference for one type of tRNA-amino acid pair over all the others. The second energetically costly active step provides a second chance to distinguish correct and wrong tRNA-amino acid pairs.

The splitting probabilities at each state can be obtained from the ratio of rates on the outward arrows. Starting from the poised R state, the probabilities of binding the correct tRNA (p_{bc}) or the wrong tRNA (p_{bw}) are given by

$$p_{bc} = \frac{1}{1+g}, \quad p_{bw} = \frac{g}{1+g}. \tag{1}$$

The probabilities of unbinding (p_{u1c}, p_{u1w}) or undergoing hydrolysis (p_{hc}, p_{hw}) from the RC and RW states, respectively, are given in turn by

$$p_{u1c} = \frac{k_{u1}}{k_{u1}+k_h}, \quad p_{u1w} = \frac{k_{u1}e^{\Delta_1}}{k_{u1}e^{\Delta_1}+k_h}, \tag{2a}$$

$$p_{hc} = \frac{k_h}{k_{u1}+k_h}, \quad p_{hw} = \frac{k_h}{k_{u1}e^{\Delta_1}+k_h}. \tag{2b}$$

Similarly, from the activated states, the probabilities of unbinding (p_{u2c}, p_{u2w}) or peptide elongation (p_{pc}, p_{pw}) are given by

$$p_{u2c} = \frac{k_{u2}}{k_{u2}+k_p}, \quad p_{u2w} = \frac{k_{u2}}{k_{u2}e^{\Delta_2}+k_p}, \tag{3a}$$

$$p_{pc} = \frac{k_p}{k_{u2}+k_p}, \quad p_{pc} = \frac{k_p}{k_{u2}e^{\Delta_2}+k_p}. \tag{3b}$$

We note that these are splitting probabilities for a ribosome assumed to be starting in a specific state; they do not directly include the steady-state probability of being in that state or the resulting flux along any given arrow. Because the system is Markovian, we can define a probabilistic weight for any multi-step path (from a given starting state) by multiplying the probabilities of all the steps. The weights of different paths can then be added together.

An individual "interaction event" between a ribosome and a tRNA consists of any path that leaves from the R state and then either returns to R or proceeds to elongation, with no intervening visits to R. The probability that an interaction includes a hydrolysis event is the sum of

two terms, corresponding to paths with correct or wrong binding and results in the expression

$$p_{\rm hyd} = p_{bc}p_{hc} + p_{bw}p_{hw}. \tag{4}$$

Similarly, the probability of a ribosome-tRNA interaction resulting in elongation is given by

$$p_{\rm el} = p_{bc}p_{hc}p_{pc} + p_{bw}p_{hw}p_{pw}. \tag{5}$$

We want to calculate the average number of hydrolysis events preceding an elongation step. This can be done quite simply using conditional probabilities, a unifying concept that has been recently highlighted for its utility in clarifying the exploratory dynamics of biological processes [40]. The problem at hand is equivalent to a classic probability question: "If you throw a fair 6-sided die until you first get a 6, how many throws of the dice do you need on average if you only count the throws with even faces?". Since only even throws are counted, we need the probability of success (rolling a 6), conditional on the roll being even. This conditional probability is given by $p = P(6|\text{even}) = 1/3$. The average number of even rolls to reach a 6 (including the last one) is then $1/p = 3$.

For the proofreading case, the conditional probability of a ribosome-tRNA interaction leading to elongation, given that a hydrolysis event occurs, is $p = P(\text{elongation}|\text{hydrolysis}) = p_{\rm el}/p_{\rm hyd}$. We therefore compute the average number of futile hydrolysis events (not counting the one that successfully leads to elongation) as

$$\begin{aligned}\langle n\rangle = \frac{1}{p} - 1 &= \frac{p_{bc}p_{hc}(1-p_{pc}) + p_{bw}p_{hw}(1-p_{pw})}{p_{bc}p_{hc}p_{pc} + p_{bw}p_{hw}p_{pw}} \\ &= \frac{p_{bc}p_{hc}p_{u2c} + p_{bw}p_{hw}p_{u2w}}{p_{bc}p_{hc}p_{pc} + p_{bw}p_{hw}p_{pw}}.\end{aligned} \tag{6}$$

Notably, this average number of futile cycles can also be intuitively expressed as the ratio of unsuccessful to successful hydrolysis events: $\langle n\rangle = (1-p)/p = P(\text{not elongation}|\text{hydrolysis})/P(\text{elongation}|\text{hydrolysis})$. Plugging in Eq. 1- 3 gives the result in terms of kinetic parameters:

$$\langle n\rangle = \frac{k_{u2}}{k_p}\left[\frac{\left(k_{u1}e^{\Delta_1}+k_h\right)\left(k_{u2}e^{\Delta_2}+k_p\right) + ge^{\Delta_2}\left(k_{u1}+k_h\right)\left(k_{u2}+k_p\right)}{\left(k_{u1}e^{\Delta_1}+k_h\right)\left(k_{u2}e^{\Delta_2}+k_p\right) + g\left(k_{u1}+k_h\right)\left(k_{u2}+k_p\right)}\right] \tag{7}$$

The error rate can also be expressed as the ratio of probabilistic weights for paths with the wrong tRNA leading to elongation versus paths with the correct tRNA:

$$f = \frac{p_{bw}p_{hw}p_{pw}}{p_{bc}p_{hc}p_{pc}} = \frac{g(k_{u1}+k_h)(k_{u2}+k_p)}{(k_{u1}e^{\Delta_1}+k_h)(k_{u2}e^{\Delta_2}+k_p)}. \tag{8}$$

We note that this specific proofreading scheme can be mapped to a more general model that incorporates transitions between multiple intermediate states. The general scheme (see Appendix, Fig. 11) includes any system that has a single energy-consuming irreversible transition between a set of initial states and a set of activated states, followed by either return to the initial state or an irreversible elongation. In particular, this includes both the Hopfield model in Fig. 1 and certain more detailed models that incorporate codon recognition, accommodation, etc [38]. The probabilistic approach here can be applied after computing splitting probabilities and mean first passage times between the coarse-grained sets of states in the general system.

Accuracy and cost in irreversible model

The energetic cost of translational elongation can be computed from the number of hydrolysis events, as $\Delta E = (\langle n\rangle + 1)\epsilon_{\rm GTP}$ where $\epsilon_{\rm GTP}$ is the change in free energy associated with hydrolyzing one ATP molecule. We note some features of this energy cost are apparent from Eq. 7. First of all, the energy cost is not dependent on the binding rate k_b of the tRNAs, as expected since binding initiates all interactions, whether they are futile or not.

If the release step is indiscriminate between right and wrong amino acids ($\Delta_2 \to 0$), then the number of excess hydrolysis events becomes k_{u2}/k_p (the rate of falling out of the activated state, relative to the rate of peptide bond formation). This is expected since each time a tRNA enters either activated state there is a constant probability of p_{pc} of proceeding to elongation versus falling off. In this case, the calculation of $\langle n\rangle$ is equivalent to asking how many times a biased coin (with probability p_{pc} of showing heads) must be flipped before the first heads is obtained ($\langle n\rangle = 1/p_{pc} - 1 = k_{u2}/k_p$). In this limit, the energetic cost of elongation can be minimized by preventing release of the activated state ($k_{u2} \to 0$) or effectively removing the indiscriminate proofreading step from the system. The same limit is obtained when there are no wrong amino acids present in the system ($g \to 0$).

We next consider the limit where hydrolysis is much slower than the unbinding rate ($k_h \ll k_{u1}$), allowing the initial binding step to equilibrate before each hydrolysis occurs. This limit was also assumed in past analyses of kinetic proofreading processes [12, 27]. In this case, the average count of excess hydrolyses, and the error rate

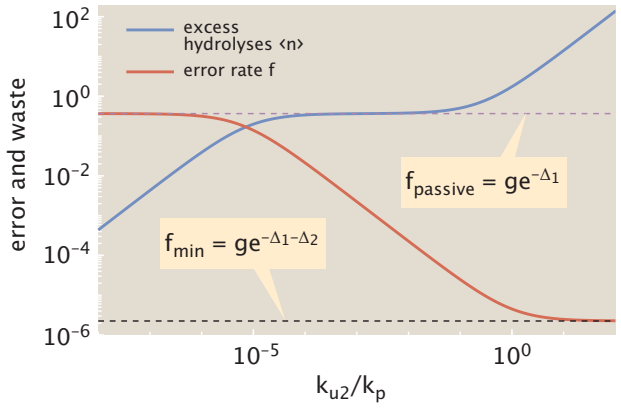


FIG. 3. **Proofreading error is decreased at the cost of additional futile cycles** Error rate (red, f) decreases and the cost in excess hydrolysis events (blue, $\langle n \rangle$) increases as the ratio of release to elongation rates (k_{u2}/k_p) is raised. Dotted purple line marks the passive error rate in the absence of proofreading. This is also equal to the excess hydrolysis count in the intermediate regime where wrong tRNAs are likely to dissociate during the proofreading step while correct tRNAs are likely to proceed to elongation. Dotted black line marks the fundamental fidelity limit for the proofreading system, which can only be achieved reached with an infinite number of hydrolyses per elongation. Results shown are in the limit $k_h \to 0$, with $\Delta_1 = 4, \Delta_2 = 12, g = 20$.

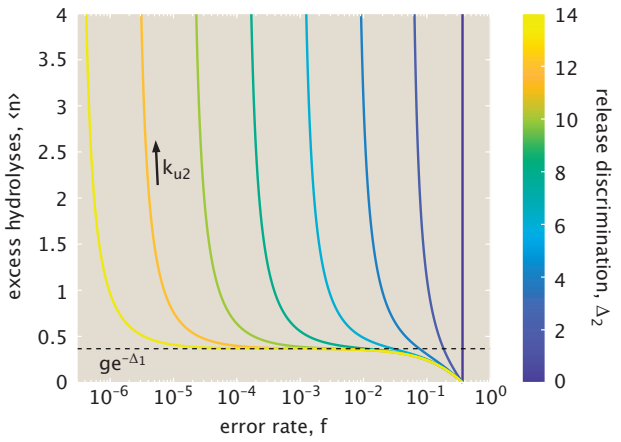


FIG. 4. **Limited energetic cost allows for near-optimal fidelity.** The average number of excess hydrolysis events $\langle n \rangle$ is plotted versus the error rate f. Each curve corresponds to a fixed value of Δ_2, with increasing release rates k_{u2} moving to the left along the curve. Dashed black line marks the thermodynamic error rate for the hydrolysis itself, corresponding to the probability that the wrong tRNA reaches the activated state. Large values of the release discrimination energy Δ_2 allow for the plateau value of excess hydrolysis to extend to low overall error rates. Results shown are in the limit $k_h \to 0$, with $\Delta_1 = 4, g = 20, k_p = 1$.

simplify to

$$\langle n \rangle \xrightarrow{k_h \to 0} \frac{k_{u2}}{k_p} \left[\frac{e^{\Delta_1} \left(k_{u2} e^{\Delta_2} + k_p \right) + g e^{\Delta_2} \left(k_{u2} + k_p \right)}{e^{\Delta_1} \left(k_{u2} e^{\Delta_2} + k_p \right) + g \left(k_{u2} + k_p \right)} \right], \tag{9a}$$

$$f \xrightarrow{k_h \to 0} g e^{-\Delta_1} \left(\frac{k_p + k_{u2}}{k_p + k_{u2} e^{\Delta_2}} \right). \tag{9b}$$

In this limit, neither the energy cost nor the fidelity are dependent on the unbinding rate k_{u1}. Instead, they are determined by the discrimination energy for the initial binding step (Δ_1), and for dissociation during the proofreading step (Δ_2), as well as the fold-excess of wrong amino acids (g) and the relative rate of release during proofreading (k_{u2}/k_p). Both error rate and the excess hydrolysis count are plotted in Fig. 3, as a function of this release rate.

In the regime where the release rate during proofreading is very low ($k_{u2} \ll k_p$), there are very few excess hydrolysis events ($\langle n \rangle \to 0$). However, the system also loses its ability to proofread since no incorrect tRNAs are released after they enter the activated state. The error rate then approaches the thermodynamic limit for passive binding: $f_{\text{passive}} = g e^{-\Delta_1}$.

As the release rate increases, we enter an intermediate regime where the correct tRNA has a low chance of being released during proofreading but the wrong tRNA has a high chance of being released: $k_{u2} \ll k_p \ll k_{u2} e^{\Delta_2}$. In addition, we also assume $g e^{-\Delta_1} \ll \frac{k_{u2}}{k_p} e^{\Delta_2}$, implying that the probability of activated release for wrong amino acids is high enough to overcome the error inherent in the initial binding. Since most wrong tRNAs are then released after activation, the average number of excess hydrolysis events is equal to the (thermodynamic) error rate of the hydrolysis process itself: $\langle n \rangle \approx g e^{-\Delta_1}$. Notably, this quantity is independent of k_{u2}, giving rise to a plateau region where the excess hydrolysis count remains flat while the error rate continues to decrease (Fig. 3).

In the regime where the release rate is high for all tRNAs ($k_{u2} \gg k_p$), the error rate approaches its fundamental limit [12] of $f_{\text{min}} = g e^{-\Delta_1 - \Delta_2}$. However, in this regime the number of excess hydrolysis cycles approaches infinity as many release events precede each successful incorporation.

A direct relation between the error rate and the number of excess hydrolysis events, in the limit of $k_h \to 0$, can be written as

$$\langle n \rangle = \left(\frac{f_{\text{passive}} - f}{f - f_{\text{min}}} \right) \left(\frac{f_{\text{min}} + f f_{\text{passive}}}{f_{\text{passive}} + f f_{\text{passive}}} \right), \tag{10}$$

where the error rate is always constrained to lie in the range $f_{\text{min}} < f < f_{\text{passive}}$. For the case where $\Delta_1 = \Delta_2$, this expression is identical to that previously derived for the minimum number of futile hydrolysis cycles in a reversible proofreading system with a fixed error rate [29].

From this relation (plotted in Fig. 4), we again see the three regimes for the error rate. For the lowest error rates ($f \to f_{\text{min}}$), the proofreading process becomes exceedingly wasteful as $\langle n \rangle$ goes to infinity. For

the highest error rates ($f \to f_{\rm passive}$), the excess hydrolysis count goes to 0: there is no waste, but the system also loses the accuracy boost due to proofreading. In the intermediate plateau regime, corresponding to $f_{\rm min} \ll \{f f_{\rm passive}, f\} \ll f_{\rm passive}$, the hydrolysis count is approximated by $\langle n \rangle \approx g e^{-\Delta_1}$. The plateau becomes wider, accessing lower error rates, when there is greater discrimination for release of correct versus wrong tRNAs (higher Δ_2, or larger ratio of $f_{\rm passive}/f_{\rm min}$).

The plateau region implies that a relatively low error rate, close to the thermodynamic limit, can be achieved with only a modest cost in terms of the number of excess hydrolysis events. From published estimates of the kinetic parameters for translational elongation [33], we assume a binding discrimination energy in the range of $\Delta_1 \approx 2 - 4 k_b T$, comparable to a couple of hydrogen bonds, and a $g = 20$ fold-excess of wrong amino acids. The average excess hydrolysis count in the intermediate regime is then $\langle n \rangle \approx 0.3 - 2.7$.

The difference in release rates from the activated state is not well characterized. However, so long as its value is high enough to enable the plateau region to extend beyond the desired error rate for translational elongation, Δ_2 does not substantially affect the expected number of hydrolysis events per incorporated amino acid. For a typical error rate of $f \approx 10^{-4}$, this requires $\Delta_2 \gtrsim 9 - 11$, depending on the specific value of binding energy selected. For release discrimination energies that are more than $(\log 2) k_b T$ below this cutoff value, the desired error rate cannot be achieved at any cost of excess hydrolyses.

Overall, despite the high excess concentration of wrong amino acids, the energetic cost per elongation step is quite low – in the range of $1.3 - 3.7$ GTP molecules hydrolyzed per new amino acid incorporated. This is due largely to the fact that the binding energy Δ_1 provides an initial non-energy-consuming discrimination step, such that $30 - 80^{\%}$ of the tRNAs reaching the activated state are already bearing the correct amino acid. Without this discrimination in the binding energy, the ribosome would need to hydrolyze on average 20 GTP molecules to incorporate each amino acid, in order to achieve the necessary low error rates in translation. In the clubhouse analogy, even a rough selection of who comes through the swinging doors can greatly reduce the number of identity checks the demon has to perform.

Reversible model: proofreading on a landscape

The model schematic in Fig. 1 makes use of one-way arrows and thus does not explicitly define the energy cost associated with each hydrolysis cycle. This model is applicable when the reverse rates for both the activation (hydrolysis) step and the subsequent release step are small enough to be negligible. Specifically, they must be small compared to other outward arrows from the same states in order to not contribute to the splitting probabilities.

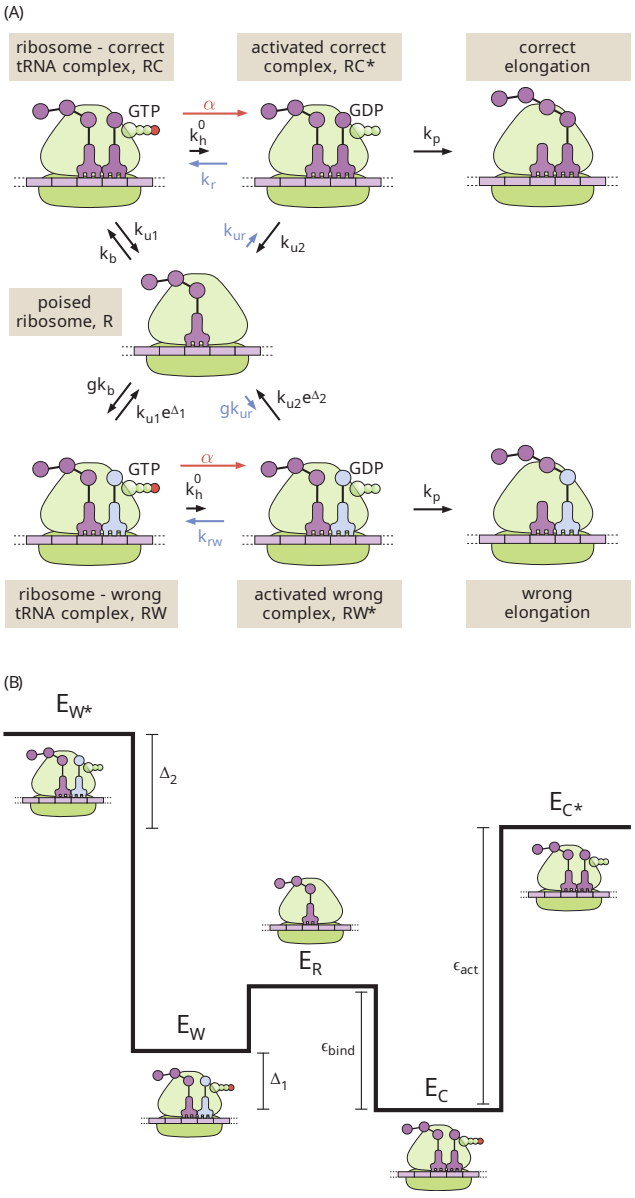


FIG. 5. **Schematic of proofreading model with reversible transitions and thermodynamic driving.** (a) Blue arrows and rates indicate reverse transitions not included in the original model. Red arrows show additional driven reaction rate along the activation step. (b) Example energy landscape describing the states involved prior to elongation. Energy levels shown here are for a single tRNA, and so do not include the additional factor of g associated with higher concentration of wrong tRNAs. All energy steps labeled $(\Delta_1, \Delta_2, \epsilon_{\rm bind}, \epsilon_{\rm act})$ are taken as positive numbers in units of $k_b T$.

A more general form of the model with reversible transitions can be used to explore the relation between the thermodynamic driving force and the fidelity of the system. Such a model requires defining an energy landscape for the system, as illustrated in Fig. 5. To constrain the space of possible schemes, we make a few key assumptions. First, we assume that the final elongation step,

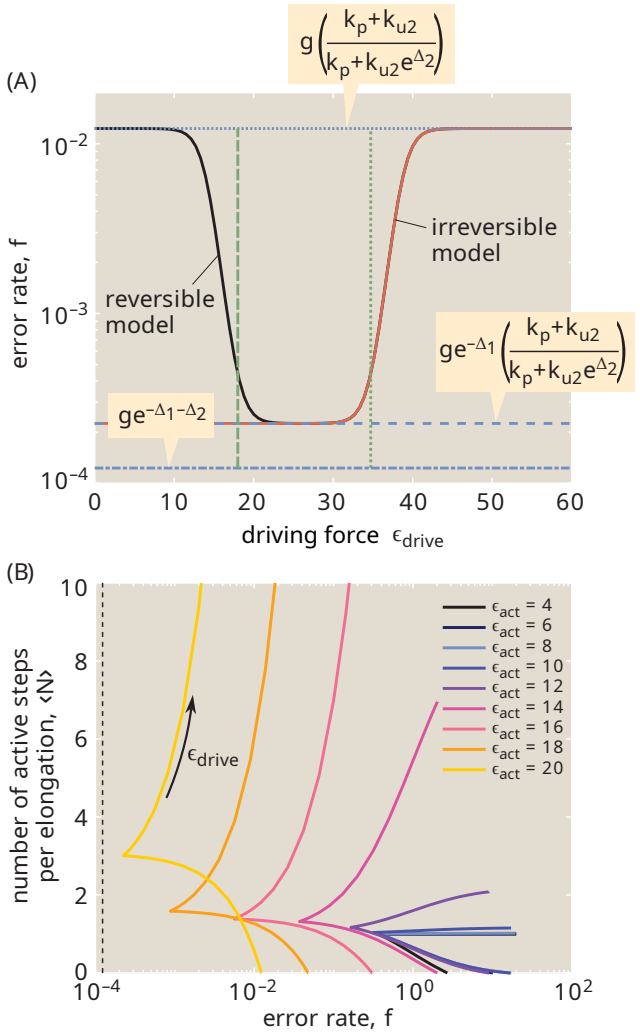


FIG. 6. **Trade-off between accuracy, driving, and cost for reversible proofreading system.** (a) Error rate f is plotted against the thermodynamic driving force $\epsilon_{\rm drive}$. Black curve: reversible model. Red curve: irreversible limit with total activation rate set to $k_h = k_h^0 e^{\epsilon_{\rm drive}}$. Dashed blue line is the limit for the irreversible model with small k_h. Dotted blue line is the limit where discrimination occurs entirely in the release step. Dash-dotted blue line shows the thermodynamic limit for the minimal possible error rate. The green lines mark the region of optimal driving: $\epsilon_{\rm act} - \epsilon_{\rm bind} + \Delta_1 + \log\frac{k_{ur}}{k_r} < \epsilon_{\rm drive} < \epsilon_{\rm act} - \epsilon_{\rm bind} + \log\frac{k_b}{k_r}$. Activation barrier is set to $\epsilon_{\rm act} = 20$. (b) Average number of active transitions to reach elongation, plotted versus the error rate. Each curve corresponds to a different activation barrier $\epsilon_{\rm act}$, with the driving force $\epsilon_{\rm drive}$ increased along each curve. Dashed black line shows thermodynamic minimal error rate $ge^{-\Delta_1-\Delta_2}$. Parameters are: $\Delta_1 = 4, \Delta_2 = 8, \epsilon_{\rm bind} = 6, k_r = 10^{-6}, k_{ur} = 10^{-6}, k_b = 10^4, g = 20$.

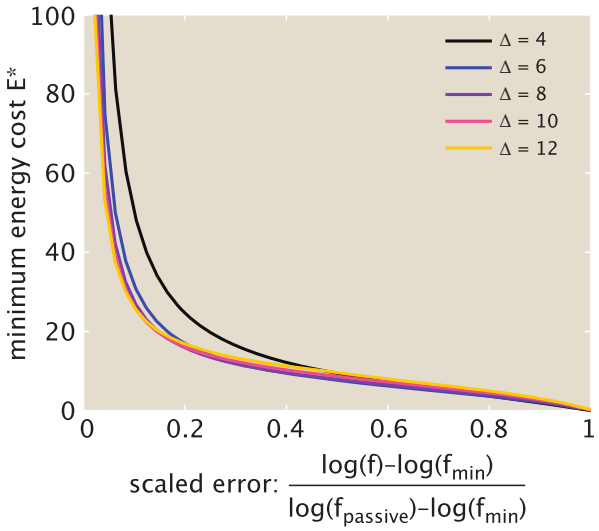


FIG. 7. **Minimal energy cost to reach a desired error rate** Minimal energy per elongation E^* is plotted as a function of the rescaled error rate, defined as an interpolation between the minimum value $f_{\rm min} = ge^{-2\Delta}$ and the passive value $f_{\rm passive} = ge^{-\Delta}$, on a logarithmic scale. Both binding energy differences are assumed to be the same, with each curve corresponding to a fixed value of $\Delta_1 = \Delta_2 = \Delta$. The energetic cost is minimized over all values of $\epsilon_{\rm bind}, \epsilon_{\rm act}, k_r, k_{ur}$, with fixed parameters $k_b = 10^4, g = 20$.

which proceeds at rate k_p is still effectively irreversible. The elongated chain thus serves as an absorbing state for the system, and we focus our attention on the kinetics of the transitions preceding this state.

For the energy landscape shown in Fig. 5b, there is an energy drop of $\epsilon_{\rm bind}$ upon the binding of the correct tRNA, and an energy difference of Δ_1 between the bound state with the wrong versus the correct tRNA. We assume that any external thermodynamic driving is localized specifically to the activation transitions ($RC \to RC^*, RW \to RW^*$). We would then expect the energies of the activated states (E_{C*}, E_{W*}) to be high, so that the release from these states can occur with substantial probability without any additional driving. We define the energy increase from the bound to the activated state for the correct tRNA to be $\epsilon_{act} = E_{C*} - E_C$. The difference in activated energies for wrong versus correct tRNAs is given by $\Delta_2 = E_{W*} - E_{C*}$, corresponding to the additional discrimination that allows for proofreading.

As in the original model, we assume that the rates of release from the activated state are k_{u2} for correct tRNAs and $k_{u2}e^{\Delta_2}$ for wrong ones. Correspondingly, the reverse transitions for an individual tRNA, regardless of identity must be $k_{ur} = k_{u2}e^{-(\epsilon_{\rm act}-\epsilon_{\rm bind})}$. Because the wrong tRNAs are present in g-fold excess, we assume this reverse process to have rate gk_{ur} for the wrong tRNA.

The activation transition is assumed to have a single low basal rate of k_h^0 in the absence of driving. The corresponding reverse rates for correct and wrong tRNAs are then given by $k_r = k_h^0 e^{\epsilon_{\rm act}}, k_{rw} = k_h^0 e^{\epsilon_{\rm act}+\Delta_2-\Delta_1}$. Given the much higher energies of the activated states, the equilibrium probability of those states (and concomitantly the elongation flux) will be very low. An additional driven process is needed (red arrow in Fig. 5a) to push the

system towards elongation. This process could represent the hydrolysis of GTP and/or release of GDP from the Ef-Tu elongation factor. We assume the rate α associated with this driving is the same regardless of the tRNA identity. The corresponding thermodynamic driving force can then be expressed as $\epsilon_{\rm drive} = \log(1 + \alpha/k_h^0)$ [41, 42].

The model with reverse transitions can be solved as before by combining the splitting probabilities to compute the error rate f of elongating with the wrong versus the correct tRNA, and the average number of transitions through the driven activation step ($\langle N \rangle$) to reach elongation. The final expressions (with derivation provided in Appendix) are:

$$f = \frac{p_{pw}(p_{bw}p_{hw} + p_{urw})/(1 - p_{rw}p_{hw})}{p_{pc}(p_{bc}p_{hc} + p_{urc})/(1 - p_{rc}p_{hc})}, \tag{11}$$

$$\langle N \rangle = \left[\frac{(\frac{p_{bc}+p_{urc}p_{rc})p_{hc}}{1-p_{rc}p_{hc}} + \frac{(p_{bw}+p_{urw}p_{rw})p_{hw}}{1-p_{rw}p_{hw}}}{\frac{(p_{bc}p_{hc}+p_{urc})p_{pc}}{1-p_{rc}p_{hc}} + \frac{(p_{bw}p_{hw}+p_{urw})p_{pw}}{1-p_{rw}p_{hw}}} \right] (1 - e^{-\epsilon_{\rm drive}}) \tag{12}$$

where p_{urc}, p_{urw} are the splitting probabilities for going directly from the empty R state into an activated state along the reverse release pathway, p_{rc}, p_{rw} are splitting probabilities for the reverse activation transition, and p_{hc}, p_{hw} are splitting probabilities of going towards the activated state from the bound state (along either the basal or the active arrow).

As plotted in Fig. 6A, increasing the driving force for activation has a non-monotonic effect on the accuracy of the system. When there is no driving, the basal activation rate k_h^0 is so low that the system primarily reaches the activated state through the reverse-release pathway (rates k_{ur}, gk_{ur}). In this case, the difference in binding energy Δ_1 becomes irrelevant and the error rate approaches the known value for a simple substrate-selective enzymatic reaction [35, 36]:

$$f \to g\frac{k_p + k_{u2}}{k_p + k_{u2}e^{\Delta_2}}. \tag{13}$$

As the driving force rises, the binding and activation pathway begins to dominate, and the error rate decreases until the activation transition becomes effectively irreversible. The only splitting probabilities in Eq. 11, 12 that depend on the driving force are p_{hc}, p_{hw}. The transition to the irreversible system occurs when $p_{bw}p_{hw} > p_{urw}$, or equivalently $\epsilon_{\rm drive} > \epsilon_{\rm act} - \epsilon_{\rm bind} + \Delta_1 + \log(k_{ur}/k_r)$. At that point, the error rate reaches the value given in Eq. 9b for the irreversible system, and further driving does not improve the accuracy. If the driving force becomes much higher, the rapid activation transition prohibits the system from sensing the difference in binding energy Δ_1 and it again approaches the limit in Eq. 13 where the only discriminating step is release from the activated state. The transition to this increased error rate occurs when $k_h^0 + \alpha > k_{u1}$ or equivalently when $\epsilon_{\rm drive} > \epsilon_{\rm act} - \epsilon_{\rm bind} + \log(k_b/k_r)$.

In the reversible model, the energetic cost for elongating the peptide by one amino acid can be expressed as $E^* = \langle N \rangle \epsilon_{\rm drive}$: the product of the thermodynamic driving force for each activation cycle and the number of such cycles required per elongation event. The tradeoff between futile cycles and accuracy is evident in Fig. 6b. Approaching closer to the minimal possible error rate requires a higher energy for the activated state ($\epsilon_{\rm act}$), which in turn raises the release rate and the number of futile cycles, as well as the concomitant energy cost.

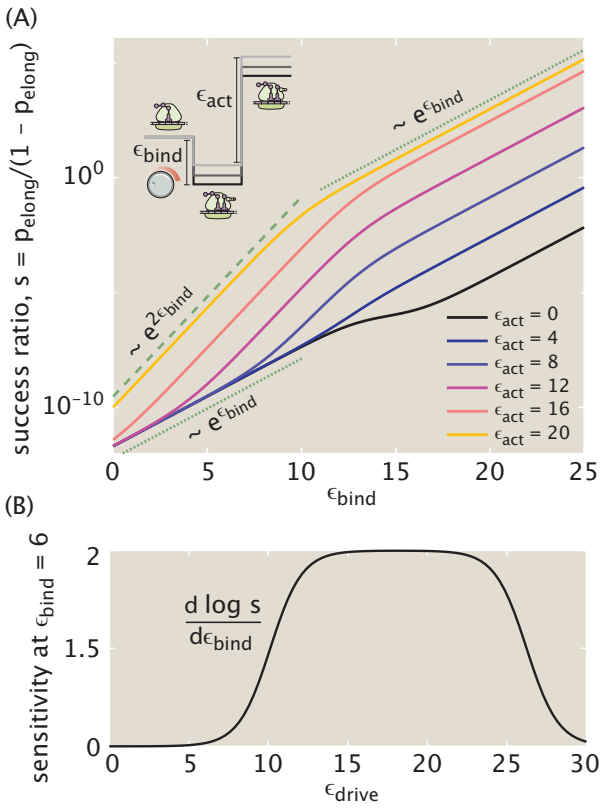


FIG. 8. **Energetic driving enhances sensitivity to substrate binding** (a) For a single type of tRNA ($g = 0$), probability that an interaction will successfully result in elongation is plotted against the binding energy. The activation barrier is held fixed at $\epsilon_{\rm act} = 20$. Each curve corresponds to a different driving force. Dashed and dotted green lines show two different exponential scalings. (b) Sensitivity to the binding energy (defined as $d \log p_{\rm elong}/d\epsilon_{\rm bind}$), as $\epsilon_{\rm bind} = 6$ is plotted as a function of driving force.

The range of possible error rates, and the total energy E^* needed to achieve a certain error, depend on the binding energy differences Δ_1, Δ_2 distinguishing correct versus wrong tRNAs. The minimal possible error for the active system is given by $f_{\rm min} = ge^{-\Delta_1-\Delta_2}$. A passive equilibrium system can only achieve the error of $f_{\rm passive} = ge^{-\max(\Delta_1,\Delta_2)}$. By numerically minimizing over all the other kinetic parameters, we can compute the minimal cost for sliding between these two error limits, as shown in Fig. 7. Errors above $f_{\rm passive}$ can be achieved at zero cost. Pushing towards the minimal possible value of $f_{\rm min}$ requires an infinite energetic cost. For reasonable

values of the binding energy difference Δ (corresponding to a few hydrogen bonds), an energy cost on the order of $10 - 20k_bT$ per elongation step is sufficient for approaching close to the minimal error, after which the cost begins to grow steeply.

Energy dissipation in the translational proofreading system allows it to more accurately discriminate among tRNAs with similar binding energies. This property can be couched in terms of sensitivity or signal gain: the input signal is the binding energy of a particular tRNA and the output is the likelihood that the tRNA will successfully transfer its amino acid to the peptide chain each time it interacts with the ribosome. In Fig. 8, we quantify this output by plotting the success ratio $s = p_{\text{elong}}/(1 - p_{\text{elong}})$ for a system with only one type of tRNA present (*i.e.* $g = 0$). The more sensitive this ratio is to the binding energy, the more capable the system will be of distinguishing between tRNAs with small binding energy differences. For a passive system with $\epsilon_{\text{drive}} = 0$, the success ratio scales exponentially with the binding strength in both the strong binding and weak binding limits. In the presence of strong energetic driving and weak binding, the success ratio exhibits a quadratically steeper scaling, implying greater sensitivity of the system. When binding is very strong, then unbinding becomes vanishingly unlikely and the system loses its ability to stack multiple binding energy differences, reverting back to the lower sensitivity. The classic definition of sensitivity as the derivative of the logarithm of the output [43] is plotted in Fig. 8B, demonstrating that intermediate driving confers the greatest sensitivity values.

As summarized in Fig. 9 this simple proofreading system exhibits three regimes with increasing driving force, which can be seen in both Fig. 6A and Fig. 8B. At very low driving, hydrolysis is extremely unlikely, and elongation can only be achieved when the tRNA bypasses the binding and hydrolysis step to enter the active state directly. The error rate is then determined entirely by the binding energy difference in the activated state (Δ_2), corresponding to a sensitivity of 1. At intermediate driving, the hydrolysis pathway dominates, and the irreversible model becomes an adequate description of the system. Within this regime, the error rate is determined multiplicatively by two factors that each correspond to the error rate of a single Michaelis-Menten enzyme. Each factor depends on the binding energy difference and involves a balance between the release rates and the rate of transitioning forward to the next state. In the limit where release dominates, the error rate scales exponentially with the sum of both binding energy differences ($\Delta_1 + \Delta_2$), but the requisite number of futile cycles and the concomitant energetic cost approaches infinity. This limit corresponds to a sensitivity of 2. When excess driving is applied, the tRNA has no chance to unbind before hydrolysis and the error rate is again determined only by a single binding energy difference (Δ_2), with sensitivity approaching 1.

Overall, translational proofreading serves as an illustrative case study of the tradeoff between energetic driving, total cost of futile cycles, and the accuracy of the system in distinguishing between substrates with small differences in binding energy.

CATALYTIC CONTROL OF DYNAMIC INSTABILITY

In the simple proofreading scheme described above, active driving enables the system to better differentiate the right vs wrong substrate, despite the fact that the driven rate itself is not substrate-dependent. Instead, proofreading makes the system more sensitive to the pre-existing difference in release rates for the two substrates. We proceed to further explore this feature of exploratory dynamics with resetting: the ability to regulate system behavior by tuning passive transition rates.

Specifically we focus on control by reusable catalysts present in much smaller quantities than the reactants themselves. At thermodynamic equilibrium, the presence of a catalyst can only alter the transition barriers but not change the steady-state distribution of a system. However, many biomolecular systems engage in catalytic control, where a catalyst modulates the probability distribution at steady state. Unlike allosteric control via ligand binding [44], catalytic control implies that the regulator molecule can be reused over and over again, while the substrates maintain some memory of their interaction with it. Consequently, the catalyst can be present at substoichiometric concentrations compared to the substrate, enabling the cell to carry out its tasks with limited resources. For example, some kinases are more than three orders of magnitude lower in concentration than their target substrates [45]. The prevalence of catalytic control relieves spatial crowding constraints when multiple regulatory proteins are necessary to tune protein activity or assembly.

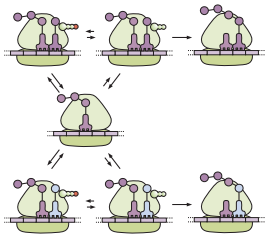
Catalytic control requires the system to be driven out of equilibrium. Recent work demonstrated that the change in probability distribution due to the addition of a catalyst is bounded by twice the total applied thermodynamic force [42]. The presence of driving can make the steady state of a system sensitive to catalysts facilitating other transitions that are not themselves actively driven. Such systems are qualitatively distinct from molecular machines that dissipate energy to directly push the system towards a target state. Instead, they rely on exploratory dynamics that sample many pathways, enabling multiple points of regulation by different catalysts present in sub-stoichiometric concentrations.

Microtubule length control model

Here, we illustrate a concrete manifestation of this phenomenon in the context of microtubule length regulation by catalytic factors that destabilize the end cap at the tip

fidelity scheme	splitting probabilities	accuracy	driving force	cost

1 No driving force. Visitors bypass outer doors and arrive at inner door directly. All visitors enter, but the wrong ones are more likely to be ejected.



$p_{bw}p_{hw} < p_{urw}$

$p_{bc}p_{hc} < p_{urc}$

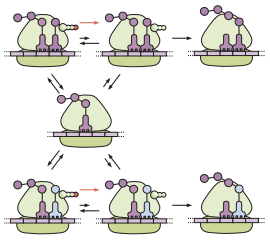
$f \to g\left(\frac{k_p+k_{u2}}{k_p+k_{u2}e^{\Delta_2}}\right)$

accuracy of a classic Michaelis-Menten enzyme with one binding step

$\epsilon_{drive} << \epsilon_{act} - \epsilon_{bind} + \Delta_1 + \log\frac{k_{ur}}{k_r}$

$E^* \to 0$

2 Sufficient driving force. Visitors verified through the outer door first, and only then energy expended to open the inner door. Only a few visitors successfully enter.



$p_{bw}p_{hw} > p_{urw}$

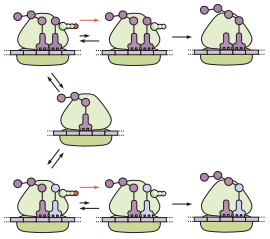
$p_{bc}p_{hc} > p_{urc}$

$f \to g\left(\frac{k_h+k_{u1}}{k_h+k_{u1}e^{\Delta_1}}\right)\left(\frac{k_p+k_{u2}}{k_p+k_{u2}e^{\Delta_2}}\right)$

$\epsilon_{act} - \epsilon_{bind} + \Delta_1 + \log\frac{k_{ur}}{k_r} << \epsilon_{drive}$

$\epsilon_{drive} << \epsilon_{act} - \epsilon_{bind} + \log\frac{k_b}{k_r}$

2a Same as (2) above. Inner door wide open, all the visitors pass through inner door. No chance to eject them.

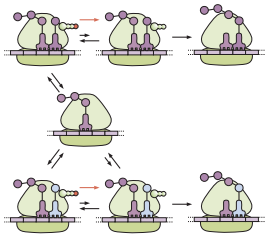


$p_{pc} \to 1, p_{pw} \to 1$

$f \to g\left(\frac{k_h+k_{u1}}{k_h+k_{u1}e^{\Delta_1}}\right)$

$E^* \to 0$

2b Same as (2) above. After entering the outer door, the wrong visitors get ejected but the right ones are allowed to stay.

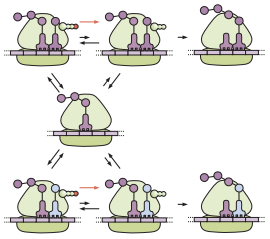


$p_{pc} \to 1, p_{pw} \to 0$

$f \to g\left(\frac{k_h+k_{u1}}{k_h+k_{u1}e^{\Delta_1}}\right)\frac{k_p}{k_{u2}}e^{-\Delta_2}$

$E^* \to (ge^{-\Delta_1}+1)\epsilon_{drive}$

2c Same as (2) above. Eject both right and wrong visitors after letting them through the outer door.



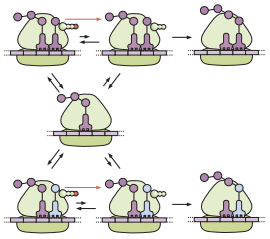
$p_{hc} \to 0, p_{hw} \to 0$

$p_{pc} \to 0, p_{pw} \to 0$

$f \to ge^{-\Delta_1-\Delta_2}$

$E^* \to \infty$

3 Excess driving force. Open the outer door too fast, so visitors get in whether they respect the sign or not.



$p_{hc} \to 1, p_{hw} \to 1$

$f \to g\left(\frac{k_h+k_{u2}}{k_h+k_{u2}e^{\Delta_2}}\right)$

$\epsilon_{drive} >> \epsilon_{act} - \epsilon_{bind} + \log\frac{k_b}{k_r}$

$E^* \to \infty$

FIG. 9. **Summary of regimes in translational proofreading model.** Limiting cases are shown for (1) No driving, (2) Sufficient driving to approach the irreversible model, and (3) Excess driving. The splitting probability limits, the error rate f, the driving force $\epsilon_{\mathrm{drive}}$, and the total energetic cost are listed for each regime.

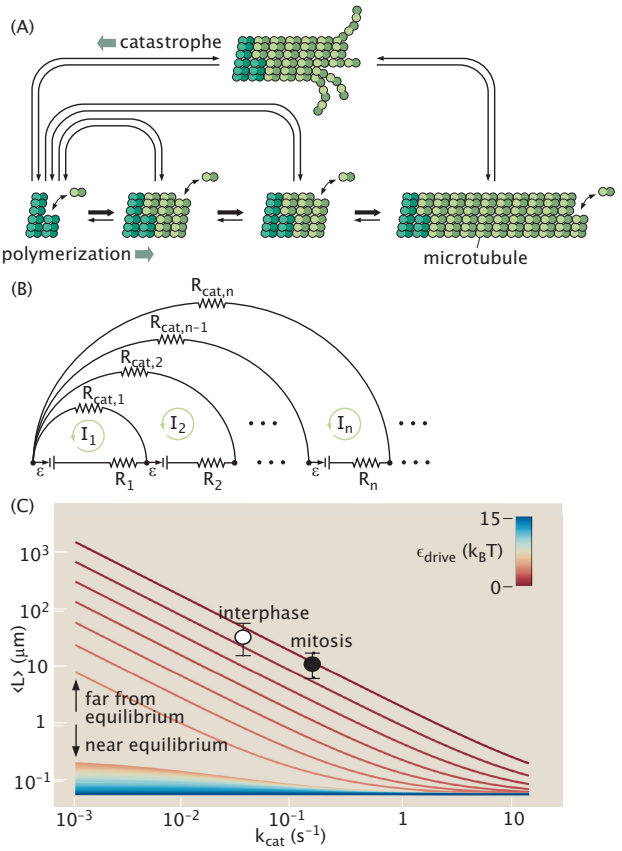


FIG. 10. **Catalytic regulation of microtubule length.** Self-assembly of tubulin subunits into filaments, with complete disassembly events (catastrophes) (A). This process can be mapped to a circuit diagram (B). Eq. 1 gives the mean length as a function of catastrophe rate and thermodynamic force, which can be interpreted as the GTP concentration (color bar; C). At equilibrium (blue line), catastrophe rate has no influence on mean length. At physiological GTP concentration (red), the predicted $k_{\rm cat}$ dependence is in excellent agreement with the measured mean microtubule length at interphase and mitosis, which differ only in the catastrophe rate (C). The catastrophe rate linearly tunes mean microtubule length above a critical thermodynamic force (GTP concentration). Parameters and measured lengths are taken from Ref.[46].

of the growing microtubule. The elementary steps constituting microtubule self-assembly are shown in Fig. 10A in the absence of rescue from catastrophe [47]. The state space contains an infinite number of possible states corresponding to increasing lengths of the microtubule. The forward elongation process in this simple model is followed by the hydrolysis of GTP to GDP, whose rate determines the size of the GTP-containing microtubule end-cap.

In a system where GTP and GDP are allowed to fully equilibrate, the reversible assembly of tubulin onto the filament occurs with forward and reverse rate constants k_f and k_r. Thus, $k_f/k_r = e^{-\beta G}$, where G is the equilibrium dimer binding free energy. In cells, GTP is kept at high concentration in excess of its equilibrium level, effectively giving rise to an additional forward rate constant α, which is proportional to excess [GTP] up to a saturation concentration. The thermodynamic driving force is $\epsilon_{\rm drive} = \ln[1 + \alpha/k_f]$.

Assembly is counteracted by a catalyzed catastrophe process with rate constant $k_{\rm cat}$, allowing for the complete disassembly of the microtubule in a regulatable manner [48, 49]. Catastrophe is triggered by the stochastic disruption of the growing microtubule cap [50], which allows cap-modifying substrates to act as sub-stoichimetric catalysts of microtubule shrinkage. Microtubule-associated proteins that trigger catastrophe include both ATP-consuming motors in the kinesin family [51] and passive factors such as Op18/Stathmin [52]. Notably, stathmin levels in the cell are estimated to be sub-micromolar [53], while tubulin can reach concentrations in the hundreds of micromolar [54], indicating that this enzyme must act as a reusable catalyst.

Microtubule catastrophe serves as a resetting step for the exploratory dynamics of the microtubule length. The dynamic instability steady state is reached when catastrophe balances net dimer addition, resulting in a length distribution that is distinct from the equilibrium steady state of the system [47] and is dependent on the level of catalysist present.

Although these processes have been modeled mathematically [55] and via computational simulations [56], the complexity of the dynamical system consisting of numerous reversible reactions, has limited our quantitative understanding of how system parameters control microtubule length distributions. Previous work has established the intrinsic speed-up of non-equilibrium polymer reorganization kinetics compared to equilibrium reorganization [57]. An article within the current issue [23] highlights how the exploratory dynamics of growing microtubules enable them to rapidly find targets within the cell. The resetting catastrophe process thus provides clear benefits to the speed of the system. Here we highlight the additional advantage of non-equilibrium driving in enabling steady-state length regulation via a catalyst.

Length distribution depends on catalytic rate

Equilibrium theory teaches that catalytic rate constants cannot affect the mean value of any observable. In contrast, the microtubule length probability distribution $P(L)$ reaches a steady-state where the mean length is known to depend explicitly on the catalytic rate $k_{\rm cat}$. In the limit of strong driving, with near-irreversible catastrophe and forward-biased growth, the mean length has previously been computed as $\langle L\rangle = (\alpha + k_{\rm f} - k_{\rm r})/k_{\rm cat}$ [58]. In this regime, a catalyst which only decreases the energy barrier to catastrophe leads to a proportional change in the mean length, in violation of the equilibrium rule. Such catalytic regulation is known to occur during the eukaryotic cell cycle, where increased $k_{\rm cat}$ causes the decrease in microtubule length necessary for cell division

[59]. However, the switching on of catalytic control as a function of thermodynamic driving has not been established.

The reaction scheme in Fig. 10A is a generalization of the single-step proofreading circuit shown in Fig. 1. In the appendix, we demonstrate how the method of counting weighted paths can be applied to a simplified system where the catastrophe process is irreversible. We use this approach to compute the distribution of lengths at which catastrophe occurs. For catastrophe to be nearly irreversible in an equilibrium system, the free energies of longer states must be much higher, and most microtubules can only reach a very short length before undergoing catastrophe. For a driven system however, it is possible to extend this distribution to arbitrarily long lengths by raising the probability of stepping forward rather than reversing or undergoing catastrophe at each state. Because this probability depends on the catastrophe rate $k_{\rm cat}$, the resulting system is necessarily sensitive to the level of catalytic enzyme. This sensitivity is analogous to proofreading fidelity, allowing the system to accurately convert different levels of catalyst to different responses.

We note that the path-counting approach becomes prohibitively tedious for complex reaction systems, including when reverse catastrophe transitions are included. The approach can be automated in the form of matrix algebra [60], or replaced with approaches that rely on solving the chemical master equation [29] or on graph-theoretic methods that count spanning trees across the network [61]. However, to maximize our intuition regarding the role of energetic driving, we turn to an alternate technique that involves mapping the system to an effective circuit framework [41] (Fig. 10B), with batteries representing driven transitions ($\mathcal{E} \propto \alpha/k_f$).

By leveraging techniques for simplifying electronic circuits, we can then compute a closed-form expression for the steady-state length distribution of microtubules (See Appendix C):

$$P(L)/P(1) = \frac{k_{\rm cat}}{k_{\rm cat} - \alpha(e^{\beta G} - 1)} e^{-\beta G(L-1)} + \frac{\alpha(e^{\beta G} - 1)}{\alpha(e^{\beta G} - 1) - k_{\rm cat}} e^{-D(L-1)}, \quad (14)$$

where $P(1)$ is the monomer fraction, and $D \equiv -\ln[1 - \frac{\sqrt{(\alpha+k_{\rm cat}+k_f-k_r)^2+4k_{\rm cat}k_r}-(\alpha+k_{\rm cat}+k_f-k_r)}{2k_r}]$. When catastrophe is equally likely from all states, this expression can also be used to compute the distribution of lengths at which catalysis occurs: $P_{\rm cat}(L) = P(L)/[1 - P(1)]$, matching the results obtained by counting paths.

Although mean filament length has been calculated using generating functions [62], this is the first time that the full distribution $P(L)$ has been solved and the role of the thermodynamic force isolated. Interestingly, $P(L)$ is a superposition of two exponential functions, corresponding to the equilibrium and nonequilibrium contributions, respectively. The double exponential explains why previous attempts to fit $P(L)$ generated from numerical simulations to a single exponential distribution led to poor fits [62].

Fig. 10C shows the mean microtubule length as a function of catastrophe rate as predicted by Eq. 14 using measured rate constants [46], for varying α corresponding to different GTP concentrations. As expected, if $\alpha = 0$ (blue line) then Eq. 14 reduces to the equilibrium single-exponential distribution, which is independent of $k_{\rm cat}$. However, as the system is driven from equilibrium, the length distribution jumps between two distinct regimes with qualitatively different dependence on $k_{\rm cat}$. The jump occurs when α exceeds $k_r - k_f$. In the strongly-driven regime, for which $(\alpha + k_{\rm f} - k_r)/k_{\rm cat} \gg 1$, Eq. 14 simplifies to $\langle L \rangle_{\rm strong} = (\alpha + k_{\rm f} - k_{\rm r})/k_{\rm cat}$, which is the well-known formula cited above. At physiological GTP concentrations, the predicted mean length is in excellent agreement with measured lengths [59] in both mitosis and interphase (circles in Fig. 10C). In the weakly-driven regime ($-(\alpha + k_{\rm f} - k_r)/k_{\rm cat} \gg 1$), Eq. 14 simplifies to $\langle L \rangle_{\rm weak} = -\ln[\frac{\alpha+k_f}{kr} + \frac{k_{\rm cat}(\alpha+k_f)}{k_r(\alpha+k_f-k_r)}]^{-1}$; the mean length is only marginally sensitive to $k_{\rm cat}$ in this regime. The thermodynamic force, as parameterized by α or [GTP], controls the transition between the near and far-from-equilibrium regimes, whose sharpness is inversely proportional to $k_{\rm cat}$(Fig. 10C). Therefore, a uniquely non-equilibrium feature (catalytic regulation of an ensemble-averaged observable) is turned on in a switch-like manner when the system is driven beyond the threshold level.

DISCUSSION

In this work we highlight how exploratory dynamics with resetting conveys the advantage of sensitivity to biochemical pathways. In the case of translational proofreading, release from an activated state increases the ability of the system to select the correct tRNA among an excess of decoys, rendering it more sensitive to a small difference in binding energies. For the case of microtubule length control, resetting through catastrophe allows the system to be responsive to sub-stoichiometric concentrations of a destabilizing enzyme. In both cases, sensitivity comes at an energetic cost, requiring GTP hydrolysis to drive the resetting cycles.

We compute the energetic cost associated with translational proofreading by starting with the classic Hopfield and Ninio model, which assumes a single GTP is hydrolyzed at each irreversible activation step. An intuitively simple probabilistic path-counting approach gives an expression for the number of excess activation cycles. The resulting total energetic cost increases monotonically with the release rate from the activated state, while the error rate of the system decreases. Notably,

there is a broad plateau region for intermediate release rates where the number of excess activation cycles is well-approximated by the equilibrium error rate for the initial binding step: $\langle n \rangle \approx g e^{-E_1}$. The plateau spans the parameter regime where most correct tRNAs proceed towards elongation while most wrong ones are released. In this case, futile activation cycles occur only when the wrong tRNA passes through to the activation step.

A key consequence of this plateau is that, despite the excess of wrong tRNAs, the energetic cost for elongation remains quite low. The system capitalizes on equilibrium discrimination during the initial binding to limit the frequency of wrong tRNAs proceeding through activation. Since correct tRNAs are less likely to be released, this means that only a small number of GTP hydrolysis events are needed per incorporated amino acid. Conceptually this is akin to letting visitors self-filter through a passive set of swinging doors before allowing them to proceed to an energy-consuming identity check. Partially accurate discrimination in the first step implies that only a few identity checks are needed before an acceptable visitor is permitted to enter.

Models with irreversible transitions are, in principle, unphysical, requiring an infinite input of energy to completely preclude reverse transitions. In practice, however, such models are meant to represent schemes where the reverse transition is so unlikely that it does not contribute to the splitting probabilities of the system. As shown for the translational proofreading example, increasing energetic driving can push a system towards the irreversible limit; further driving beyond that necessary to reach this limit can actually hinder the sensitivity of the system. Thus, irreversible models constitute a useful limiting case for quantifying the efficiency of a system undergoing exploratory dynamics.

In kinetic proofreading, energy dissipation at one point in the system increases the sensitivity to small differences in release rates elsewhere, enabling accurate discrimination in the presence of excess decoys. The sensitivity jumps sharply when the thermodynamic driving force exceeds a certain critical value. Analogously, we showed that the sensitivity of mean microtubule length to the rate of catastrophe undergoes a switch from logarithmic dependence near equilibrium to linear dependence far from equilibrium. Energetic driving of systems undergoing exploratory dynamics can thus trigger a qualitative transition in their input-output response functions.

Overall, our results parallel past work [27, 29**?**] linking the accuracy, speed, and energy efficiency of active proofreading systems. We focus on the specific cases of translational proofreading and catalytic control to concretely illustrate biochemical systems that face a tradeoff in energetic cost versus function. Two pedagogically useful approaches are demonstrated to analyze these systems: probabilistic path-counting and mapping to an electrical circuit system. Both case studies highlight the importance of driven resetting steps for enhancing sensitivity in conditions where decoy substrate concentrations may be high or regulator concentrations are limited.

We hypothesize that many other examples of intracellular exploratory dynamics, including quality control pathways, signaling cascades, cell-cycle associated transitions, and organelle rearrangements may be analyzed in an analogous manner to link the energetic cost with the sensitivity to various control parameters. Future exploration of such systems may help elucidate the many functions of the homeostatic energy consumption defining living cells.

ACKNOWLEDGEMENTS

We are grateful to the CZI Theory Institute Without Walls for supporting our joint efforts. RP is grateful to the NIH for support through award numbers DP1OD000217 (Director's Pioneer Award) and NIH MIRA 1R35 GM118043-01. MML is grateful for support from the Sloan Foundation Matter-to-Life grant G-2024-22449 and the NIH R01GM125748. EFK acknowledges support from NSF grant 2034482 and from a UCSD Academic Senate research grant.

[1] H. C. Berg, *E. coli in motion* (Springer, New York, 2004).
[2] L. Wolpert, Positional information and the spatial pattern of cellular differentiation, J Theor Biol **25**, 1 (1969).
[3] G. J. and K. M., *Cells, Embryos and Evolution* (Blackwell Science, Malden, Mass., 1997).
[4] M. Kirschner and T. Mitchison, Beyond Self-Assembly - from Microtubules to Morphogenesis, Cell **45**, 329 (1986).
[5] R. Heald and A. Khodjakov, Thirty years of search and capture: The complex simplicity of mitotic spindle assembly, J Cell Biol **211**, 1103 (2015).
[6] T. Misgeld and T. L. Schwarz, Mitostasis in neurons: maintaining mitochondria in an extended cellular architecture, Neuron **96**, 651 (2017).
[7] A. H. Williams, C. O'Donnell, T. J. Sejnowski, and T. O'Leary, Dendritic trafficking faces physiologically critical speed-precision tradeoffs, eLife **5**, e20556 (2016).
[8] A. F. MacAskill, J. E. Rinholm, A. E. Twelvetrees, I. L. Arancibia-Carcamo, J. Muir, A. Fransson, P. Aspenstrom, D. Attwell, and J. T. Kittler, Miro1 is a calcium sensor for glutamate receptor-dependent localization of mitochondria at synapses, Neuron **61**, 541 (2009).
[9] G. Pekkurnaz, J. C. Trinidad, X. Wang, D. Kong, and T. L. Schwarz, Glucose regulates mitochondrial motility via milton modification by O-GlcNAc transferase, Cell **158**, 54 (2014).
[10] H. M. York, K. Joshi, C. S. Wright, L. Z. Kreplin, S. J. Rodgers, U. K. Moorthi, H. Gandhi, A. Patil, C. A.

Mitchell, S. Iyer-Biswas, *et al.*, Deterministic early endosomal maturations emerge from a stochastic trigger-and-convert mechanism, Nat Commun **14**, 4652 (2023).
[11] R. Villaseñor, Y. Kalaidzidis, and M. Zerial, Signal processing by the endosomal system, Curr Opin Cell Biol **39**, 53 (2016).
[12] J. J. Hopfield, Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity, P Natl Acad Sci **71**, 4135 (1974).
[13] J. Ninio, Kinetic amplification of enzyme discrimination, Biochimie **57**, 587 (1975).
[14] W. Cui and P. Mehta, Identifying feasible operating regimes for early T-cell recognition: The speed, energy, accuracy trade-off in kinetic proofreading and adaptive sorting, PloS one **13**, e0202331 (2018).
[15] D. Hathcock, Q. Yu, and Y. Tu, Time-reversal symmetry breaking in the chemosensory array reveals a general mechanism for dissipation-enhanced cooperative sensing, Nat Commun **15**, 8892 (2024).
[16] R. Bar-Ziv, T. Tlusty, and A. Libchaber, Protein–DNA computation by stochastic assembly cascade, P Natl Acad Sci **99**, 11589 (2002).
[17] I. Mizrahi, R. Bruinsma, and J. Rudnick, Spanning tree model and the assembly kinetics of RNA viruses, Phys Rev E **106**, 044405 (2022).
[18] Y. Lu, W. Wang, and M. W. Kirschner, Specificity of the anaphase-promoting complex: a single-molecule study, Science **348**, 1248737 (2015).
[19] B. M. Adams, M. E. Oster, and D. N. Hebert, Protein quality control in the endoplasmic reticulum, Protein J **38**, 317 (2019).
[20] A. I. Brown and E. F. Koslover, Design principles for the glycoprotein quality control pathway, Plos Comput Biol **17**, e1008654 (2021).
[21] M. R. Evans and S. N. Majumdar, Diffusion with stochastic resetting, Phys Rev Lett **106**, 160601 (2011).
[22] M. R. Evans, S. N. Majumdar, and G. Schehr, Stochastic resetting and applications, Journal of Physics A: Mathematical and Theoretical **53**, 193001 (2020).
[23] J. Kondev, M. Kirschner, H. Garcia, and R. Phillips, Biological Processes as Exploratory Dynamics, arXiv (2025).
[24] L. D. Belmont, A. A. Hyman, K. E. Sawin, and T. J. Mitchison, Real-time visualization of cell cycle-dependent changes in microtubule dynamics in cytoplasmic extracts, Cell **62**, 579 (1990).
[25] H. Mellenius and M. Ehrenberg, Transcriptional accuracy modeling suggests two-step proofreading by RNA polymerase, Nucleic Acids Res **45**, 11582 (2017).
[26] K. Banerjee, A. B. Kolomeisky, and O. A. Igoshin, Elucidating interplay of speed and accuracy in biological error correction, P Natl Acad Sci **114**, 5183 (2017).
[27] A. Murugan, D. A. Huse, and S. Leibler, Speed, dissipation, and error in kinetic proofreading, P Natl Acad Sci **109**, 12034 (2012).
[28] P. Sartori and S. Pigolotti, Thermodynamics of error correction, Phys Rev X **5**, 041039 (2015).
[29] Q. Yu, A. B. Kolomeisky, and O. A. Igoshin, The energy cost and optimal design of networks for biological discrimination, J Roy Soc Interface **19**, 20210883 (2022).
[30] P. Sartori and S. Pigolotti, Kinetic versus energetic discrimination in biological copying, Phys Rev Lett **110**, 188101 (2013).
[31] E. T. Kool and H. O. Sintim, The difluorotoluene debate—a decade later, Chem Commun , 3665 (2006).
[32] T. A. Kunkel and K. Bebenek, Dna replication fidelity, Annu Rev Biochem **69**, 497 (2000).
[33] J. M. Ogle, F. V. Murphy, M. J. Tarry, and V. Ramakrishnan, Selection of tRNA by the ribosome requires a transition from an open to a closed form, Cell **111**, 721 (2002).
[34] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. H. Tu, The energy-speed-accuracy trade-off in sensory adaptation, Nat Phys **8**, 422 (2012).
[35] D. S. Tawfik, Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency?, Curr Opin Chem Biol **21**, 73 (2014).
[36] M. Johansson, M. Lovmar, and M. Ehrenberg, Rate and accuracy of bacterial protein synthesis revisited, Curr Opin Microbiol **11**, 141 (2008).
[37] R. Phillips, J. Kondev, J. Theriot, and H. Garcia, *Physical biology of the cell* (Garland Science, 2012).
[38] M. V. Rodnina and W. Wintermeyer, Ribosome fidelity: tRNA discrimination, proofreading and induced fit, Trends Biochem Sci **26**, 124 (2001).
[39] E. Franco and J. J. Velázquez, A Stochastic Version of the Hopfield–Ninio Kinetic Proofreading Model, Siam J Appl Math **85**, 576 (2025).
[40] O. Hachmo and A. Amir, Conditional probability as found in nature: Facilitated diffusion, Am J Phys **91**, 653 (2023).
[41] M. M. Lin, Circuit reduction of heterogeneous nonequilibrium systems, Phys Rev Lett **125**, 218101 (2020).
[42] E. Arunachalam and M. M. Lin, Information gain limit of biomolecular computation, Phys Rev Lett **134**, 148401 (2025).
[43] J. A. Owen and J. M. Horowitz, Size limits the sensitivity of kinetic schemes, Nat Commun **14**, 1280 (2023).
[44] R. Phillips, *The molecular switch: Signaling and Allostery* (Princeton University Press, 2020).
[45] B. M. Martins and P. S. Swain, Ultrasensitivity in phosphorylation-dephosphorylation cycles with little substrate, Plos Comput Biol **9**, e1003175 (2013).
[46] H. Bowne-Anderson, M. Zanic, M. Kauer, and J. Howard, Microtubule dynamic instability: a new model with coupled gtp hydrolysis and multistep catastrophe, Bioessays **35**, 452 (2013).
[47] T. Mitchison and M. Kirschner, Dynamic instability of microtubule growth, Nature **312**, 237 (1984).
[48] F. A. Piedra, T. Kim, E. S. Garza, E. A. Geyer, A. Burns, X. Ye, and L. M. Rice, GDP-to-GTP exchange on the microtubule end can contribute to the frequency of catastrophe, Molecular Biology of the Cell **27**, 3515 (2016).
[49] H. Bowne-Anderson, A. Hibbel, and J. Howard, Regulation of microtubule growth and catastrophe: unifying theory and experiment, Trends Cell Biol **25**, 769 (2015).
[50] L. Brun, B. Rupp, J. J. Ward, and F. Nédélec, A theory of microtubule catastrophes and their regulation, P Natl Acad Sci **106**, 21173 (2009).
[51] C. E. Walczak, S. Gayek, and R. Ohi, Microtubule-depolymerizing kinesins, Annu Rev Cell Dev Biol **29**, 417 (2013).
[52] L. D. Belmont and T. J. Mitchison, Identification of a protein that interacts with tubulin dimers and increases the catastrophe rate of microtubules, Cell **84**, 623 (1996).
[53] D. N. Ringhoff and L. Cassimeris, Stathmin regulates centrosomal nucleation of microtubules and tubu-

lin dimer/polymer partitioning, Mol Biol Cell **20**, 3451 (2009).
[54] J. Baumgart, M. Kirchner, S. Redemann, A. Bond, J. Woodruff, J. M. Verbavatz, F. Jülicher, T. Müller-Reichert, A. A. Hyman, and J. Brugués, Soluble tubulin is significantly enriched at mitotic centrosomes, J Cell Biol **218**, 3977 (2019).
[55] V. Yadav, B. Srinivas, and M. Gopalakrishnan, Microtubule catastrophe under force: mathematical and computational results from a brownian ratchet model, Phys Biol **18**, 016006 (2020).
[56] P. N. Zakharov, V. K. Arzhanik, E. V. Ulyanov, N. B. Gudimchuk, and F. I. Ataullakhanov, Microtubules: dynamically unstable stochastic phase-switching polymers, Phys-usp+ **59**, 773 (2016).
[57] D. Swanson and N. S. Wingreen, Active biopolymers confer fast reorganization kinetics, Phys Rev Lett **107**, 218103 (2011).
[58] M. Dogterom and S. Leibler, Physical aspects of the growth and regulation of microtubule structures, Physical Review Letters **70**, 1347 (1993).
[59] L. D. Belmont, A. A. Hyman, K. E. Sawin, and T. J. Mitchison, Real-time visualization of cell cycle-dependent changes in microtubule dynamics in cytoplasmic extracts, Cell **62**, 579 (1990).
[60] Z. C. Scott, A. I. Brown, S. S. Mogre, L. M. Westrate, and E. F. Koslover, Diffusive search and trajectories on tubular networks: a propagator approach, Eur Phys J E **44**, 80 (2021).
[61] F. Wong and J. Gunawardena, Gene regulation in and out of equilibrium, Ann Rev Biophys **49**, 199 (2020).
[62] P. Ranjith, D. Lacoste, K. Mallick, and J. F. Joanny, Nonequilibrium self-assembly of a filament coupled to ATP/GTP hydrolysis, Biophysical Journal **96**, 2146 (2009).
[63] M. Johansson, E. Bouakaz, M. Lovmar, and M. Ehrenberg, The kinetics of ribosomal peptidyl transfer revisited, Mol Cell **30**, 589 (2008).
[64] O. A. Igoshin, A. B. Kolomeisky, and D. E. Makarov, Coarse-graining chemical networks by trimming to preserve energy dissipation, J Phys Chem Lett **16**, 1229 (2025).
[65] D. S. Grebenkov and L. Tupikina, Heterogeneous continuous-time random walks, Phys Rev E **97**, 012148 (2018).
[66] M. Johansson, J. Zhang, and M. Ehrenberg, Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection, P Natl Acad Sci **109**, 131 (2012).
[67] M. Lovmar and M. Ehrenberg, Rate, accuracy and cost of ribosomes in bacterial cells, Biochimie **88**, 951 (2006).
[68] H. Yamakawa, Statistical mechanics of wormlike chains: Path integral and diagram methods, J Chem Phys **59**, 3811 (1973).
[69] A. J. Spakowitz and Z.-G. Wang, End-to-end distance vector distribution with fixed end orientations for the wormlike chain model, Phys Rev E **72**, 041802 (2005).

Appendix A: Fidelity and energetic cost for reversible model

For the model with reversible transitions (Fig. 5A) we can define the following splitting probabilities for correct and wrong tRNAs, respectively, as:

$$p_{bc} = \frac{k_b}{(k_b + k_{ur})(1+g)}, \qquad p_{bw} = \frac{g k_b}{(k_b + k_{ur})(1+g)} \tag{15a}$$

$$p_{urc} = \frac{k_{ur}}{(k_b + k_{ur})(1+g)}, \qquad p_{urw} = \frac{g k_{ur}}{(k_b + k_{ur})(1+g)} \tag{15b}$$

$$p_{hc} = \frac{k_h}{k_h + k_{u1}}, \qquad p_{hw} = \frac{k_h}{k_h + k_{u1w}} \tag{15c}$$

$$p_{rc} = \frac{k_r}{k_p + k_r + k_{u2}}, \qquad p_{rw} = \frac{k_r w}{k_p + k_r w + k_{u2w}} \tag{15d}$$

$$p_{pc} = \frac{k_p}{k_p + k_r + k_{u2}}, \qquad p_{pw} = \frac{k_p}{k_p + k_r w + k_{u2w}}, \tag{15e}$$

where $k_h = k_h^0 + \alpha = k_h^0 e^{\epsilon_{\mathrm{drive}}}$.

We consider individual interactions of a tRNA with a ribosome, each of which involves leaving the poised R state of the ribosome and eventually returning to it (possibly with a longer peptide chain), without any intermediate visits to that state. Each interaction can be resolved through either unbinding, release from an activated state, or elongation.

For a system that reaches the activated RC^* state, the probabilistic weight of all paths with exactly i activation transitions since the start of the interaction is:

$$w_{i,c} = p_{urc}(p_{rc}p_{hc})^i + p_{bc}p_{hc}(p_{rc}p_{hc})^{i-1}, \tag{16}$$

with an analogous expression w_{iw} for a system in the RW^* state. The probability that the interaction resolves in elongation with the correct or the wrong tRNA $(p_{el,c}, p_{el,w})$ is then

$$\begin{aligned} p_{\mathrm{el,c}} &= p_{urc}p_{pc} + \sum_{i=1}^{\infty} w_{ic}p_{pc} = \frac{(p_{urc} + p_{bc}p_{hc})p_{pc}}{1 - p_{rc}p_{hc}} \\ p_{\mathrm{el,w}} &= \frac{(p_{urw} + p_{bw}p_{hw})p_{pw}}{1 - p_{rw}p_{hw}}. \end{aligned} \tag{17}$$

The error rate is given by $f = p_{\mathrm{el,w}}/p_{\mathrm{el,c}}$, yielding Eq. 11.

The average number of activation steps for an interaction involving the correct tRNA (N_c) is found by summing over the corresponding $w_{i,c}$, multiplied by the probability that the activated state resolves with no further activation transitions: $p_{fc} = 1 - p_{rc}p_{hc}$. The result for both correct and wrong interactions is:

$$\begin{aligned} N_c &= \sum_{i=0}^{\infty} i w_{i,c} p_{fc} = \frac{(p_{bc} + p_{urc}p_{rc})p_{hc}}{1 - p_{rc}p_{hc}} \\ N_w &= \frac{(p_{bw} + p_{urw}p_{rw})p_{hw}}{1 - p_{rw}p_{hw}} \end{aligned} \tag{18}$$

Finally, we can find $\langle N \rangle$: the average number of driven activation transitions per interaction event, conditional on that event resolving to elongation. Here we multiply by the fraction of activation transitions that proceed along the driven arrow rather than the basal activation process: $\alpha/(k_h^0 + \alpha) = 1 - e^{-\epsilon_{\mathrm{drive}}}$, to yield:

$$\langle N \rangle = \left[\frac{N_c + N_w}{p_{el,c} + p_{el,w}} \right] (1 - e^{-\epsilon_{\mathrm{drive}}}), \tag{19}$$

which gives Eq. 12.

Appendix B: Coarse-graining of proofreading schemes

The path-counting approach described here can be generalized to a variety of more complex irreversible proofreading systems by mapping to the general scheme shown in Fig. 11A. In particular, this includes translational proofreading

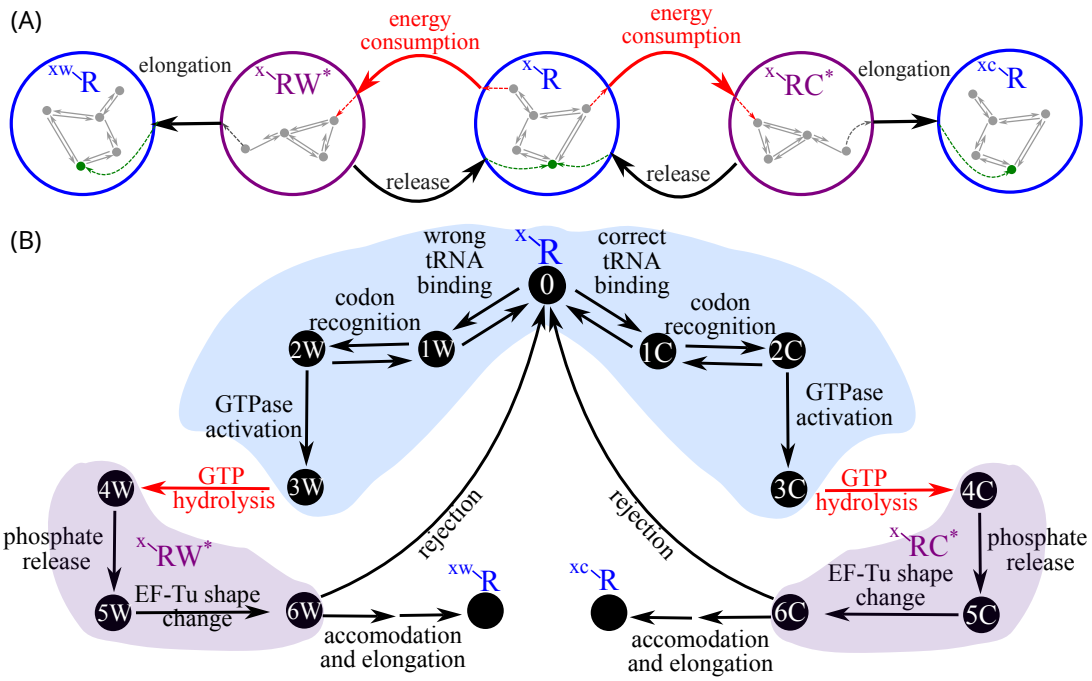


FIG. 11. **Alternate schemes for translational proofreading with a single energy-consuming step.** (A) General schematic, where each circle corresponds to a compound state containing no internal driven transitions. Red arrow represents the single energy-consuming step in the system. (B) Example of more complex proofreading scheme proposed in Ref. [38] to describe experimental measurements of ribosomal translation. Shaded regions delineate coarse-graining to the compound states in (A).

systems with additional passive transitions incorporating codon recognition, GTPase translocation, or accommodation. Such schemes (see Fig. 11B as an example) have been used to summarize experimental data on the kinetics of translational elongation in past work [38, 63]. The generalized scheme considered here uses an analogous coarse-graining procedure to that described in recently published work by Igoshin, *et al* [64], which trims states and non-energy-consuming loops using splitting probabilities and mean-first-passage times between remaining milestone states.

In Fig. 11, the one-way arrows denote reactions where the reverse rates are negligible in their effect on splitting probabilities. The red arrows mark the only energy-consuming steps (eg: GTP hydrolysis), and the energy-consumption and release arrows denote transitions along distinct pathways. Release from the activated macrostate is assumed to return the system to the same (green) microstate, regardless of which tRNA is released.

This system can be treated as a heterogeneous continuous-time random walk [65] with Markovian (albeit not constant-rate) transitions on a simplified coarse-grained network of states. We can define the splitting probabilities p_c, p_w for transitioning out of the R into the RC* or RW* states respectively. Specifically, p_c is the probability for a particle starting in the initial (green) state to first reach the compound state RC* before it ever reaches state RW*. We can also define the probability p_{pc} that a particle in the RC* macrostate will first transition to the elongated state, before a release occurs, and the probability $p_{uc} = 1 - p_{pc}$ for the opposite case.

With these definitions, we can proceed using the same analysis as before. Consider each transition out of a state as the roll of a weighted dice. The probability that the system passes through the RC* state and results in elongation is $p_c p_{pc}$. The probability that it passes through the RC* state but results in release is $p_c p_{uc}$.

The error rate f and the number of excess transitions over the energy-consuming pathway $\langle n \rangle$, can then be written as:

$$\begin{aligned} f &= \frac{p_w p_{pw}}{p_c p_{pc}}, \\ \langle n \rangle &= \frac{p_c p_{uc} + p_w p_{uw}}{p_c p_{pc} + p_w p_{pw}} \end{aligned} \tag{20}$$

We note that the scheme in Fig. 11A is equivalent to a substrate-selective Michaelis-Menten enzymatic reaction. For such reactions, the accuracy has previously been expressed as a ratio of the catalytic efficiencies $k_{\mathrm{cat}}/K_{\mathrm{m}}$ for the cognate versus noncognate substrates [66, 67]. The error rate in Eq. 20 is directly equivalent to such an expression.

Example scheme for ribosomal translation

For the specific scheme illustrated in Fig. 11B, extracted from Ref. [38], the "R" macrostate can be considered to include the initial binding and codon recognition transitions, as well as GTPase activation, with the energy-consuming exit from this macrostate corresponding to GTP hydrolysis. Since the GTPase activation is assumed to be effectively irreversible, the splitting probability out of the initial compound state is simply the probability that activation with the correct tRNA occurs before activation with the wrong tRNA (state 3C is reached before state 3W). This probability can be computed through coarse-graining of the reaction scheme as follows.

First, we find the probability $\hat{p}_{02}$ that a system starting at state 0 hits the 2C state before the 2W state. This can be done by considering each time the system leaves state 0 as the start of an independent path. Each such path must either reach state $2C$ (with probabilistic weight $p_{01}^{(c)}p_{02}^{(c)}$) or state $2W$ (with probabilistic weight $p_{01}^{(w)}p_{02}^{(w)}$) or else return to the beginning at state 0. The resulting probability of hitting $2C$ first is then:

$$\hat{p}_{02}^{(c)} = \frac{p_{01}^{(c)}p_{12}^{(c)}}{p_{01}^{(c)}p_{12}^{(c)} + p_{01}^{(w)}p_{12}^{(w)}}, \tag{21}$$

where $p_{ij}^{(c)}$ is the splitting probability from state i to adjacent state j with the correct tRNA. Similarly, we find the probability $\hat{p}_{23}$ that a system starting at state 2 hits the 3C state before returning to the 0 state. Again we consider the weight of each path leaving state 2 without returning to it, to get:

$$\hat{p}_{23}^{(c)} = \frac{p_{23}^{(c)}}{p_{23}^{(c)} + p_{21}^{(c)}p_{10}^{(c)}}. \tag{22}$$

Analogous probabilities are defined for the wrong tRNA. The desired splitting probability for leaving the macrostate entirely (through GTPase hydrolysis) is then

$$p_c = \frac{\hat{p}_{02}^{(c)}\hat{p}_{23}^{(c)}}{\hat{p}_{02}^{(c)}\hat{p}_{23}^{(c)} + \hat{p}_{02}^{(w)}\hat{p}_{23}^{(w)}}. \tag{23}$$

Note that this approach can be generalized to recursively compute splitting probabilities for any number of interchanging states arranged in a linear array.

Using the rates provided in Ref. [38], and assuming a 20-fold excess of incorrect tRNAs, gives the values $p_c = 0.05, p_w = 0.95$. With this particular kinetic scheme there is almost no discrimination between the correct and wrong tRNAs during the accommodation step, simply because the irreversible GTPase activation rate is taken to be so high ($500/50\text{s}^{-1}$, for the cognate and non-cognate tRNAs, respectively) as compared to the release rate during codon recognition ($0.2/17\text{s}^{-1}$ for the cognate and non-cognate tRNAs). Finally, the probability of elongation (rather than release) post-hydrolysis is $p_{pc} = 0.96, p_{pw} = 0.016$. With these numbers the error rate is $f \approx 0.32$, and the average number of wasteful excess hydrolysis cycles is $\langle n \rangle \approx 15$.

The proposed kinetic scheme with the claimed rate parameters thus appears to be inconsistent with the observed low error rates for translational elongation. Furthermore, it would be highly wasteful, with 15 GTP molecules consumed per amino acid incorporated into the growing peptide chain. This is the case despite multiple steps in the pathway that discriminate between correct and wrong tRNAs, favoring the former to proceed forward and the latter to be released. The inefficiency of the scheme is due primarily to the claimed very high rate of GTP hydrolysis, highlighting again the tradeoff between speed, accuracy, and energy consumption.

Appendix C: Catalytic control from exploratory dynamics with resetting

Below we describe two distinct approaches to computing the steady-state length distribution of a simple model consisting of reversible motion along a linear set of states, with a constant (catalytically controlled) rate of resetting to the origin. The first approach uses weighted path counting, analogously to the proofreading models in the main text, to compute the length at which resetting occurs. This approach is simple enough to compute manually, but is limited to irreversible resetting transitions. The second approach maps the system to an electric circuit, which can be analyzed in its entirety using well-established matrix methods.

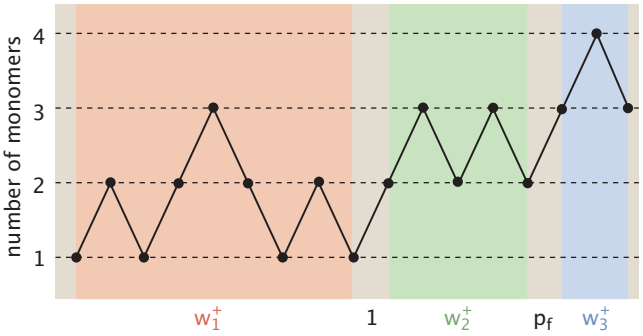


FIG. 12. Stone-fence diagram illustrating example path starting in state 1 and ending in state $i = 4$. Levels correspond to the sequential state i (eg: length of a microtubule in monomers). The path is decomposed into components that end at the last visit to each level, and the probabilistic weights of each component are given beneath.

Microtubule length control by irreversible resetting, via path-counting

Here we demonstrate how the approach of adding up probabilistically weighted paths can be extended to larger reaction systems with multiple intermediate states. The system considered here is a simple model of dynamic instability (Fig. 10A), in the limit where the catastrophe transitions are effectively irreversible. For a microtubule that initially starts at length 1, We seek to compute the probability $P_L^{\rm cat}$ that catastrophe occurs from the L state. For simplicity, we assume that the forward, reverse, and catastrophe rates are constant for all states.

We begin by defining the splitting probabilities at each state. For state $i > 1$, the probabilities of stepping forward and backward are, respectively, $p_f = (k_f + \alpha)/(k_f + \alpha + k_r + k_c), p_r = k_r/(k_f + \alpha + k_r + k_c)$. From each such state there is also a catastrophe probability $p_c = k_c/(k_f + \alpha + k_r + k_c)$. For state $i = 1$, the probability of stepping forward is simply 1. In the case where $p_r \to 0$, the system steps forward until catastrophe is reached. The probability this happens in state i is then the product of forward-stepping probabilities for $i - 1$ steps times the catastrophe probability: $P_i^{\rm cat} = p_f^{i-2}(1 - p_f)$. This probability is normalized over $i > 1$.

For the case with substantial reversals, the paths can include any number of back and forth steps, so long as they never go below 0 and end in the $i^{\rm th}$ state. These paths can be conveniently enumerated using stone-fence diagrams and continued fractions, as previously employed for computing statistics of semiflexible polymers [68, 69]. Specifically, we define w_i^+ as the total weight of all paths that start at state i and never go below it. This quantity can be expanded as

$$\begin{aligned} w_i^+ &= 1 + p_f w_{i+1}^+ p_r + (p_f w_{i+1}^+ p_r)^2 + \ldots = \frac{1}{1 - p_f p_r w_{i+1}^+}, \quad i > 1, \\ w_1^+ &= \frac{1}{1 - p_r w_{i+1}^+}. \end{aligned} \tag{24}$$

Here, the first term is a path of length 0, the second includes all paths that step up from the ith level only once (but can meander arbitrarily at $i + 1$ and above), the second term corresponds to paths that step up above the ith level twice, and so on. Because we allow the states to go infinitely high, we must have $w_i^+ = w^+$, a constant for all states $i > 1$. We can then get the closed-form expression:

$$w^+ = (1 - \sqrt{1 - 4p_f p_r})/(2p_f p_r). \tag{25}$$

Any path starting at 1 and ending at state i (with no intervening catastrophes) can be decomposed into the following sequential components: the part of the path up to its last visit at 1, then a step up to 2, the next part of the path up until its last visit at 2, then a step up to 3 and so on until you reach the last section of the path that starts and ends at i and never goes beneath it. This decomposition for an example path is illustrated in Fig. 12. The resulting total weight is then multiplied by p_c to compute the probability that catastrophe happens at state i:

$$\begin{aligned} P_i^{\rm cat} &= (w_1^+ \cdot 1 \cdot w_2^+ \cdot p_f \cdot w_3^+ \cdot p_f \ldots \cdot w_i^+) p_c \\ &= p_c w_1^+ w^+ (w^+ p_f)^{i-2} = (1 - w^+ p_f)(w^+ p_f)^{i-2}, \end{aligned} \tag{26}$$

where the last expression accounts for the normalization of the probabilities added up over all states from 2 onwards.

This geometric series skews towards longer lengths when the forward stepping probability p_f becomes high. This probability represents a balance between the forward stepping rate versus reversal or catastrophe. At equilibrium, the catastrophe can only be effectively irreversible if the energies associated with longer-length states are much higher than shorter ones, implying $k_f/k_r \ll 1$. If there is little driving in the system ($\alpha \ll k_r$) then this means $p_f \to 0$ and the distribution of lengths becomes peaked at 1, regardless of the rate of catastrophe. On the other hand, if the system is strongly driven, then $w^+ \to 1$ and we recover the limit with unidirectional stepping, discussed above. The distribution is then determined by $p_f \approx \alpha/(\alpha + k_c)$, with the average length at catastrophe given by $\langle L_{\rm cat}\rangle \to (2-p_f)/(1-p_f) \to \alpha/k_{\rm cat}$ for large α. Thus, active driving allows the steady-state microtubule length to be linearly sensitive to the level of catalyst present, ensuring catalytic control.

Finally, we note that because the catastrophe rate is the same from each state, the distribution of length upon catastrophe is directly proportional to the steady-state length distribution: $P_i = \gamma P_i^{\rm cat}$, where $\gamma = 1 - P_1$ is an appropriate normalization constant.

Microtubule catalytic control via the circuit mapping

In terms of the nth mesh current shown in Fig. 10B, the voltage equation taken along the path of the nth battery is:

$$P_{n+1}e^{\beta G_{n+1}} - P_n e^{\beta G_n} = \frac{\alpha}{k_f} P_n e^{\beta G_n} - R_n I_n. \tag{27}$$

where $R_n = e^{\beta G_n}/k_f$ and $G_n = nG$. Note that $e^{\beta G} = k_b/k_f$, where k_f and k_b are the equilibrium forward and backward rates, respectively. Using these definitions, we can solve for the probability of the $(n+1)$th state in terms of the previous state probability and current:

$$P_{n+1} = \Big(1 + \frac{\alpha}{k_f}\Big)e^{-\beta G} P_n - I_n \frac{e^{-\beta G}}{k_f} \tag{28}$$

Taking the potential difference from state $n+1$ and state 1 along the catastrophe path:

$$P_1 e^{\beta G_1} - P_{n+1} e^{\beta G_n + 1} = -R_{{\rm cat},n}(I_n - I_{n+1}), \tag{29}$$

Where $R_{{\rm cat},n} = e^{\beta G_{n+1}}/k_{\rm cat} = R_{n+1}(\frac{k_f}{k_{\rm cat}})$. Therefore, the $(n+1)$th current is:

$$I_{n+1} = I_n - k_{\rm cat} P_{n+1} + k_{\rm cat} P_1 e^{-\beta G_n} \tag{30}$$

In vector notation, the recursive probability and current equations become:

$$\begin{bmatrix} 1 & 0 \\ k_{\rm cat} & 1 \end{bmatrix} \begin{bmatrix} P_{n+1} \\ I_{n+1} \end{bmatrix} = \begin{bmatrix} \Big(1 + \frac{\alpha}{k_f}\Big)e^{-\beta G} & -\frac{e^{-\beta G}}{k_f} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_n \\ I_n \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{k_{\rm cat} P_1}{e^{\beta G_n}} \end{bmatrix} \tag{31}$$

Multiplying both sides by the inverse of the right-hand-side matrix, the recursion relation is:

$$\begin{bmatrix} P_{n+1} \\ I_{n+1} \end{bmatrix} = M \begin{bmatrix} P_n \\ I_n \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{k_{\rm cat} P_1}{e^{\beta G_n}} \end{bmatrix} \tag{32}$$

where the transition matrix M is given by:

$$M = \begin{bmatrix} \Big(1 + \frac{\alpha}{k_f}\Big)e^{-\beta G} & -\frac{e^{-\beta G}}{k_f} \\ -k_{\rm cat}\Big(1 + \frac{\alpha}{k_f}\Big)e^{-\beta G} & k_{\rm cat}\frac{e^{-\beta G}}{k_f} + 1 \end{bmatrix} \tag{33}$$

The probability and current of state n in terms of those of state 1 is thus:

$$\begin{bmatrix} P_{n+1} \\ I_{n+1} \end{bmatrix} = M^n \begin{bmatrix} P_1 \\ I_1 \end{bmatrix} + \sum_{k=0}^{n-1} (e^{\beta G} M)^k \begin{bmatrix} 0 \\ k_{\rm cat} P_1 e^{-\beta G n} \end{bmatrix} \tag{34}$$

Diagonalizing M:

$$M = V \begin{bmatrix} \lambda_- & 0 \\ 0 & \lambda_+ \end{bmatrix} V^{-1} \tag{35}$$

Where the columns of V are the eigenvectors of M and λ_- and λ_+ are the eigenvalues of M:

$$\lambda_\pm = \frac{e^{-\beta G}}{2k_f}\left(\alpha + k_f(1+e^{\beta G}) + k_{\text{cat}} \pm \sqrt{(\alpha + k_f - k_f e^{\beta G})^2 + k_{\text{cat}}(2\alpha + 2(1+e^{\beta G})k_f + k_{\text{cat}})}\right) \tag{36}$$

Which simplifies to the value given in the text:

$$D = -\ln\lambda_- = -\ln\left[1 - \frac{\sqrt{(\alpha + k_{\text{cat}} + k_f - k_r)^2 + 4k_{\text{cat}}k_r} - (\alpha + k_{\text{cat}} + k_f - kr)}{2k_r}\right] \tag{37}$$

Note that $\lambda_- \leq 1$ whereas $\lambda_+ \geq 1$.
The transfer matrix equation is then

$$\begin{bmatrix} P_{n+1} \\ I_{n+1} \end{bmatrix} = V \begin{bmatrix} \lambda_-^n & 0 \\ 0 & \lambda_+^n \end{bmatrix} V^{-1} \begin{bmatrix} P_1 \\ I_1 \end{bmatrix} + \sum_{k=0}^{n-1} e^{\beta k G} V \begin{bmatrix} \lambda_-^k & 0 \\ 0 & \lambda_+^k \end{bmatrix} V^{-1} \begin{bmatrix} 0 \\ k_{\text{cat}} P_1 e^{-\beta G n} \end{bmatrix} \tag{38}$$

Expanding this expression and taking the geometric sum yields P_n:

$$P_n = P_1 e^{-\beta G(n-1)} \frac{k_{\text{cat}}}{k_{\text{cat}} - \alpha(e^{\beta G} - 1)} + A_1 \lambda_-^{n-1} + A_2 \lambda_+^{n-1}, \tag{39}$$

where the A_1 and A_2 are explicit functions of the elementary parameters. For nonzero k_{cat} the probability monotonically decreases for larger n, thus the coefficient A_2 must be zero. Solving this boundary condition for I_1 and substituting into the expression for A_1, we obtain the length distribution (Eq. 14 in the main text):

$$P_n = \frac{k_{\text{cat}}}{k_{\text{cat}} - \alpha(e^{\beta G} - 1)} P_1 e^{-\beta G(n-1)} + \frac{\alpha(e^{\beta G} - 1)}{\alpha(e^{\beta G} - 1) - k_{\text{cat}}} P_1 e^{-D(n-1)}, \tag{40}$$

where $D = -\ln\lambda_-$. For microtubule assembly, the physiologically relevant parameters were obtained from Ref. [46]. From the expression for D, we can see that the mean length and the sensitivity of the mean length to k_{cat} is maximal in the limit that $4k_{\text{cat}}k_r/(\alpha + k_{\text{cat}} + k_f - k_r)^2 << 1$(visualized in Fig. 10C). Expanding the expression for D to first order in this ratio, we obtain:

$$D \approx -\ln\left[1 - \frac{k_{\text{cat}}}{|k_{\text{cat}} + \alpha + k_f - k_r|}\right] \tag{41}$$

In this limit, the mean length retains linear sensitivity to k_{cat} (that is, the linear approximation to the logarithm is valid) if $\alpha > k_r - k_f + \Delta$, where the minimum buffer Δ is set by the value of k_{cat} because $\frac{k_{\text{cat}}}{k_{\text{cat}}+\Delta}$ must be much less than 1. Therefore, as stated in the main text, the transition from weak (logarithmic) to strong (linear) catalytic regulation occurs when $\alpha > k_r - k_f$, with the sharpness being inversely proportional to k_{cat}.