RESEARCH ARTICLE SUMMARY

MOLECULAR BIOLOGY

Evolution of error correction through a need for speed

Riccardo Ravasio†, Kabir Husain†, Constantine G. Evans, Rob Phillips, Marco Ribezzi-Crivellari, Jack W. Szostak, Arvind Murugan*

Full article and list of author affiliations: https://doi.org/10.1126/science.adt1275

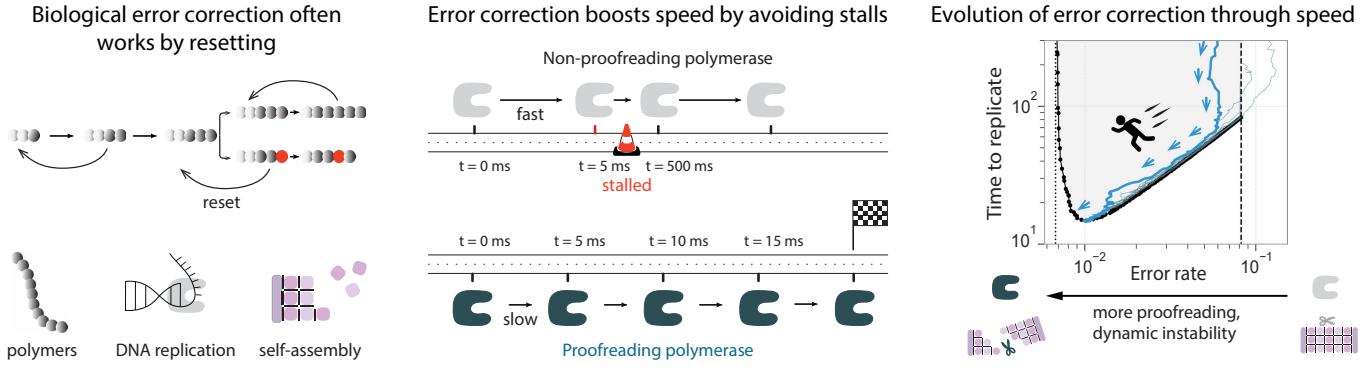
INTRODUCTION: Multistep assembly processes in biology are prone to errors, and as a result, diverse error correction strategies have evolved. For example, polymerases can excise an incorrect nucleotide by proofreading, transcription complexes can backtrack, and incorrectly assembled partial structures can be disassembled before trying again. These error-correcting mechanisms take time, leading to the widely held view that increased accuracy comes at the inevitable cost of slower replication or assembly. Error-correcting mechanisms are therefore thought to have evolved because errors in sequence or assembly are so deleterious that they must be corrected despite the additional time it takes to correct them.

RATIONALE: Standard accounts of error correction have ignored a ubiquitous effect called stalling, in which an uncorrected error, for example, one that occurs during DNA replication, can markedly slow subsequent steps, even if later steps are correct. Stalling increases the time it takes to complete a replication cycle or an assembly process, in proportion to the frequency of errors and the magnitude of the stalling effect. We show that because of stalling, error correction can speed up replication and assembly processes, despite the time taken to correct errors. We propose that a selective pressure for speed alone can lead, under certain conditions, to the evolution of error-correcting mechanisms.

RESULTS: We first developed a model of kinetic proofreading that includes stalling so that we could compare the time cost of stalling with the time cost of proofreading. Both theoretical analyses and in silico evolution show that in stall-dominated regimes, proofreading yields a net time benefit. In this context, selection for speed alone leads to increased proofreading even when errors carry no direct penalty. We predicted a preference for dissipative error correction over simpler reversible mechanisms. Second, we asked whether real biological systems satisfy these conditions. We found that stalling is widespread, from nonenzymatic and ribozyme-based replication to complex DNA and RNA polymerases, and is generally stronger in the latter. Experimental results on extensive mutagenesis of a DNA polymerase, spanning orders of magnitude in error rate and a proxy for activity, are consistent with speed selection that favors increased proofreading. Third, we generalized beyond templated replication to multicomponent assembly. Here, we found that selecting for faster assembly favors the evolution of disassembly mechanisms. These disassembly mechanisms are a by-product of a need for speed and serve to clear kinetically trapped intermediates.

CONCLUSION: In typical biological settings where accuracy is also under selection, our results predict that error-correcting mechanisms evolve more readily than expected because they confer a speed advantage rather than a time cost. This coupling can explain continued error-rate declines below that needed to maintain a genome of a given size, implying a complexity ratchet. Preexisting error correction allows genomes to increase further in size and thereby encode additional functions. This principle likely extends beyond central-dogma enzymes to the assembly of complex molecular machines and to other processes where errors slow the completion of a task. □

*Corresponding author. Email: amurugan@uchicago.edu †These authors contributed equally to this work.



Error correction evolves from a selection for speed. (Left) Many biological systems fix mistakes by going backward, such as proofreading during DNA replication or reassembling macromolecular complexes. These corrections cost time. (Center) Without correction, mistakes cause long stalls; slow error-correcting systems end up finishing first. (Right) Under selection for faster completion alone (blue), evolution drives biology to a trade-off front (black) dictated by stalling, where gaining speed requires evolving error correction. *t*, time.

MOLECULAR BIOLOGY

Evolution of error correction through a need for speed

Riccardo Ravasio[1]†, Kabir Husain[1,2]†, Constantine G. Evans[3], Rob Phillips[4], Marco Ribezzi-Crivellari[5], Jack W. Szostak[6], Arvind Murugan[1]*

Kinetic proofreading is a class of error-correcting mechanisms in biology that expend energy to avoid mistakes during replication, transcription, and translation. Proofreading is typically assumed to evolve when selection for fidelity outweighs costs in energy and the speed of replication. We show that when stalling after misincorporations is accounted for, proofreading can instead speed up replication. Consistent with data on polymerase mutagenesis, our results suggest that proofreading can evolve under selection for speed alone. We generalize to multicomponent self-assembly and show that analogous error-correcting processes, such as dynamic instability, can likewise emerge purely from selection for rapid assembly. Thus, nonequilibrium error correction can evolve from selection for speed, even without direct fidelity advantages. We discuss implications for mutation-rate evolution, molecular assembly processes, and models of early life.

Error-correcting mechanisms are ubiquitous in biology and are found whenever a process must select a correct substrate over alternative, incorrect substrates. For example, every step of DNA synthesis requires the polymerase to incorporate the correct, complementary nucleotide from the set of four canonical nucleotides (*1*). In tRNA aminoacylation and ribosomal peptide synthesis, one correct substrate must be chosen from the set of 20 amino acids or aminoacylated tRNAs (*2*). In ribosomal assembly (*3*), every one of the more than 50 constituent ribosomal proteins and ribosomal RNA must be added in the right order and location to achieve a functional molecular complex.

In many cases, the fidelity achieved by biology exceeds that expected from equilibrium thermodynamics, that is, from differences in binding energy between correct and incorrect substrates. In the 1970s, Hopfield (*4*) and Ninio (*5*) developed the theory of kinetic proofreading to explain how biology exploits dissipative, energy-consuming mechanisms to increase the fidelity of an assembly process. The fundamental idea of kinetic proofreading is that if the conversion of a substrate into product is coupled to a second irreversible reaction, the overall process can be made much more accurate. For example, if a substrate must remain bound to the enzyme for a characteristic timescale t, set by a second reaction such as adenosine triphosphate (ATP) hydrolysis, then the correct substrate will be more likely to remain bound and react, whereas an incorrect substrate is more likely to dissociate before it has a chance to react. Many specific kinetic schemes that can lead to increased accuracy in this way have been described, but these schemes share a common feature: Checking the identity of a substrate takes time, whereas the coupled irreversible step costs energy (*6*).

[1]Department of Physics, The University of Chicago, Chicago, IL, USA. [2]Department of Physics and Astronomy, and Laboratory for Molecular Cell Biology, University College London, London, UK. [3]Hamilton Institute, Maynooth University, Maynooth, Ireland. [4]Department of Physics, California Institute of Technology, Pasadena, CA, USA. [5]Laboratoire de Biochimie, Chimie Biologie et Innovation, ESPCI Paris, Université PSL, Paris, France. [6]Howard Hughes Medical Institute, Department of Chemistry, The University of Chicago, Chicago, IL, USA. ***Corresponding author. Email: amurugan@uchicago.edu**
†These authors contributed equally to this work.

A distinct question from the functioning of error-correcting mechanisms is their evolutionary origin: What selective pressures drive the evolution of error correction? The fact that biology makes wide use of proofreading, and other error-correcting mechanisms, is often taken to imply that the benefits of increased accuracy are greater than the associated costs in the energy and time required (*6*) to, for example, preserve a functional genome sequence in the face of the otherwise inevitable decay of this information because of mutation (*7*).

Here, we suggest another possibility, derived initially from observations on DNA and RNA synthesis. In many cases, the incorporation of an incorrect, mismatched nucleotide during template-directed synthesis makes the incorporation of the next nucleotide much more difficult. This post error slowdown is referred to as stalling (Fig. 1A). Because of this stalling effect, the time required to replicate a genome increases in proportion to the error frequency and magnitude of stalling, which can be quite large. This effect, not considered in classical proofreading theory, implies that error correction can lead to faster replication by removing incorrect stalled nucleotides if stalling effects are large enough.

Consequently, we suggest that complex, nonequilibrium error correction can evolve purely from the need for speed. The evolutionary advantage of fast replication has been widely discussed in the molecular evolution literature (*8*–*10*), though its effect is often assumed to reduce the information content of a genome, as in Spiegelman's monster (*11*, *12*). By contrast, in our scenario, we argue that selection for speed leads to an increase in copying fidelity, independent of any benefit derived from the accurate maintenance of genetic information. Selection can be solely for faster replication, a selection implicit in any population of self-replicating agents (*9*, *10*). We show that evolution will favor error-correction mechanisms not to preserve genomic information, but to speed up replication by relieving stalling-induced slowdowns.

We generalized our results beyond templated replication to a broad class of assembly processes in which mistakes lead to a stalling of further assembly. At a different scale, we found that in the self-assembly of a large multiprotein complex, the irreversible removal of an incorrectly added component, or even complete disassembly and restarting from the beginning, can lead to an overall increase in assembly rate. As an unselected consequence, the yield of error-free complexes increased. Thus, in both templated and template-free assembly processes, selection for rapid assembly implies a selective pressure for the evolution of error-correcting mechanisms and a concomitant increase in fidelity, even when that fidelity is never directly selected for.

Templated replication

We first delineated the contributions of proofreading and stalling to replication time. Without either effect, the time to replicate a genome of length L nucleotides is $T_{\mathrm{rep}} \approx Lt_{\mathrm{nuc}}$, where t_{nuc} is the average time required to add a single nucleotide. In the absence of any proofreading mechanism, this process is expected to occur with an error rate μ_0.

We next considered proofreading that operates by passing the growing end of the chain from the polymerase to an exonuclease domain, with a higher probability of transfer for a mismatched nucleotide than for a correct nucleotide, such that incorrect nucleotides are preferentially removed. Correcting these errors takes time, first for the mismatched base to be excised and then for that nucleotide to be resynthesized by the polymerase. The total time cost is further augmented because the proofreading activity will also, on occasion, excise a correctly incorporated nucleotide, which must also be resynthesized. In general, the time required to reduce the average error probability per nucleotide from μ_0 to μ_{ec}, the post–error correction error rate, will increase as μ_{ec} decreases, because more time is spent excising and resynthesizing nucleotides. Such proofreading increases the time required to replicate

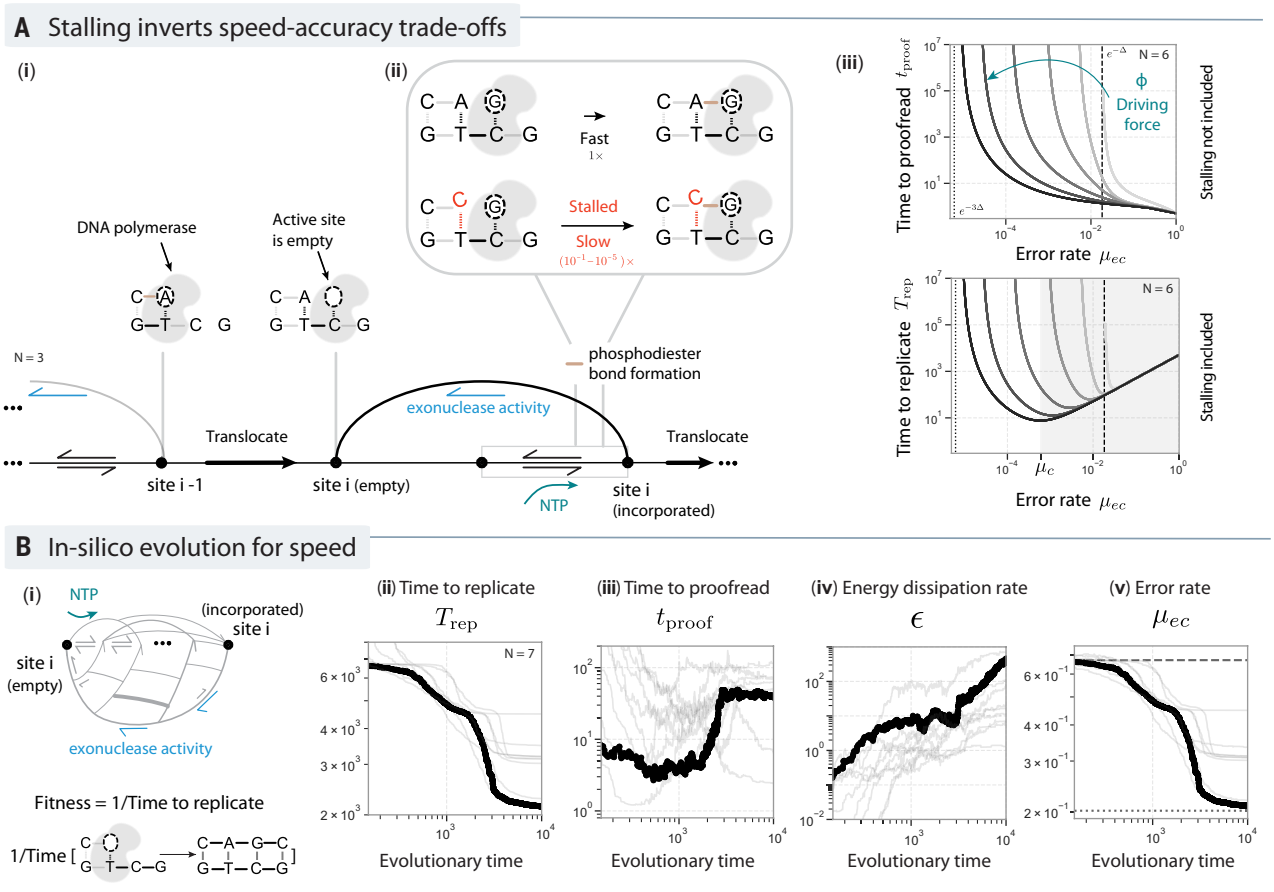


Fig. 1. Stalling leads to the evolution of kinetic proofreading owing to selection for replication speed. (**A**) We extended canonical models (i) of kinetic proofreading for polymerases by including the stalling effect (ii): Incorporation of the correct nucleotide (G) at site i is slowed [up to 10^5 times (*18*–*20*)] if site $i-1$ has an incorrect base (red C) because of a misaligned 3′ hydroxyl of C and 5′ phosphate of G. (iii) Average time to copy a strand (of length $L = 1$ nucleotide) as a function of error rate in proofreading models without and with stalling. Curves correspond to different nonequilibrium driving potentials ϕ in the proofreading network, for example, due to the coupling to NTP hydrolysis. In models with stalling, higher accuracy is linked to higher speed (gray region). Proofreading networks contain N nodes; mismatches are modeled as reactions with off rates increased by e^{Δ} relative to matched nucleotides. The maximal discrimination achieved at equilibrium gives an error rate of $e^{-\Delta}$ (dashed line), and proofreading enhances it to $e^{-N\Delta/2}$ (dotted line). See materials and methods for details. (**B**) (i) In silico evolution of a random network model of polymerase-substrate dynamics that incorporates stalling. The fitness function is set by the time to copy a long strand, with no consideration of accuracy. (ii to v) Evolution of strand replication time T_{rep}, time to add a base t_{proof} without accounting for stalling, rate of energy ϵ dissipated by network, and resulting error rate μ_{ec} during in silico evolution, where evolutionary time denotes Monte Carlo steps (see materials and methods). One of 10 trajectories is shown in bold.

each nucleotide, on average, from t_{nuc} to $t_{proof}(\mu_0 \rightarrow \mu_{ec})$, with the total time to replicate the genome then increasing to $T_{rep} \approx Lt_{proof}(\mu_0 \rightarrow \mu_{ec})$.

Here, the replication time and error rate become linked: Reducing the error rate requires a more frequent use of the exonuclease domain, which increases the replication time T_{rep}. A simple mathematical model, following numerous published models (*4*, *5*, *13*–*17*), quantifies this intuitive "higher accuracy, lower speed" trade-off, as shown in Fig. 1A, iii.

We now incorporate a key experimentally observed phenomenon, stalling (*18*–*25*), which is rarely considered in proofreading models (*26*) but completely inverts this intuitive trade-off. The incorporation of an incorrect nucleotide slows the incorporation of the next base, thus slowing the overall process of replication. This biophysical effect is of a geometric nature (*21*) and arises from mispositioning of the prior mismatched base on the template, that is, the 3′ end is misaligned. This effect is intrinsic to the nature of templated replication and is independent of proofreading or enzymatic activity; indeed, stalling has been observed even in nonenzymatic RNA replication (*27*).

If this error-induced stalling is not relieved by an error-correction mechanism such as proofreading, the average time to replicate a strand of length L increases with the number of mutations made, $\mu_0 L$. Unlike the effect of proofreading, the time to replicate then increases with the mutation rate, becoming $T_{rep} \approx Lt_{nuc} + \mu_0 L\tau_{stall}$, where τ_{stall} is the average time penalty incurred for each stall.

Error correction can thus, in principle, speed up replication by reducing the time spent in a stalled configuration, though the time cost of proofreading included in t_{proof} must also be accounted for. Combining these competing effects gives

$$T_{rep} \approx Lt_{proof}(\mu_0 \rightarrow \mu_{ec}) + \mu_{ec} L\tau_{stall} \quad (1)$$

where $\mu_{ec}L$ is the average number of uncorrected errors per copied strand, each of which causes a stall of time τ_{stall}.

The combination of error-induced stalling and error-correcting proofreading reveals a new regime (Fig. 1A, iii). For large-enough error rates $\mu_{ec} > \mu_c$, with μ_c being a crossover value defined in Fig. 1A, iii, the trade-off between speed and accuracy is reversed and error-correcting polymerases with higher accuracy are faster at replicating the strand.

Based on an analytic model provided in supplementary text section 1.1.3, we found that the "accurate is faster" effect occurs over a wide range of mutation rates $\mu_{ec} > \mu_c$ when the stalling time τ_{stall} or nonequilibrium driving force increases. In Fig. 1A, we quantify this driving force ϕ as the extent of detailed balance breaking due to the coupling to nucleoside triphosphate (NTP) hydrolysis.

Indeed, this simplified model underestimates the time spent in a stalled configuration, as even errors that are eventually corrected can transiently stall replication. In supplementary text section 1.2.3, we developed a more realistic model that allows for base excision after stalling. Because the stall time depends on both μ_0 and μ_{ec}, the "accurate is faster" regime can persist even when the observed error rate μ_{ec} is low.

This new regime has evolutionary consequences. Although the usual theory of kinetic proofreading suggests that the evolution of error correction must overcome an associated loss in speed, our results imply the opposite: Error correction can evolve without any speed cost and can in fact confer a speed advantage. Our results also suggest that selection for speed alone could lead to the evolution of complex, dissipative error-correction mechanisms, with an increase in fidelity, even if fidelity is not directly selected for, to maintain functional sequences.

To test these predictions, we carried out in silico evolution of a DNA polymerase. In our model, fitness is agnostic to the copying fidelity $1/\mu_{ec}$ and is instead entirely set by the time T_{rep} required to replicate a genome of length L. Note that T_{rep} includes both the time cost of proofreading as well as the stalling time, as in Eq. 1.

We modeled the process of polymerization as a network of N states (Fig. 1B), corresponding, for example, to conformational states of the polymerase bound to the template. We initialized the transition rates between states k_{ij} at random but subject to detailed balance, that is, with no nonequilibrium driving force. However, each transition can evolve to break detailed balance, modifying its rate k_{ij}^{eq} to $k_{ij} = k_{ij}^{eq} e^{\frac{\phi}{k_B T}}$, where ϕ denotes the extent of detailed balance breaking arising from, for example, NTP hydrolysis as the nonequilibrium driving force; k_B is the Boltzmann constant; and T is temperature (see materials and methods and fig. S1).

The results of the evolution for speed of a network of size $N = 7$ are summarized in Fig. 1B. Over cycles of mutation and selection, the replication time T_{rep} falls as expected but so does the error rate μ_{ec}, despite the absence of direct selection on fidelity. Further, the time cost to replicate each base t_{proof}, and the overall energy dissipation, increases (Fig. 1B).

We conclude that templated replication, under selection for fast replication, can evolve a proofreading mechanism that costs energy and is slower per base (i.e., higher t_{proof}). However, the reduction in errors and thus stalling ultimately pays off in faster replication (i.e., lower T_{rep}).

Experimental evidence

Our results thus far suggested that if the time cost of stalling, τ_{stall}, is sufficiently high, selection for speed alone can lead to the evolution of nonequilibrium error correction. In Fig. 2A, we show stalling factors compiled from the literature for processes across the central dogma. As shown, mismatches lead to stalling in every system studied to date, varying from large factors of $>10^5$ for extant proofreading DNA polymerases to ≈10 to 100 for RNA replication without proteins (i.e., nonenzymatic or driven by ribozymes). The effect of stalling is therefore potentially relevant to genome replication during the origin of life.

Quantitatively, our model suggests that stalling will lead to a "more accurate is faster" regime if the time spent in a stalled state is greater than the time spent in correcting errors. Even the modest stalling factors reported above may be sufficient to invert the speed-accuracy trade-off for processes with high intrinsic error rates, such as enzymatic and nonenzymatic RNA replication ($\mu_{ec} = \mu_0 \sim 10^{-1}$ to 10^{-2}). In supplementary text section 1.2, we present a more detailed model of base excision by exonuclease domains in DNA polymerases,

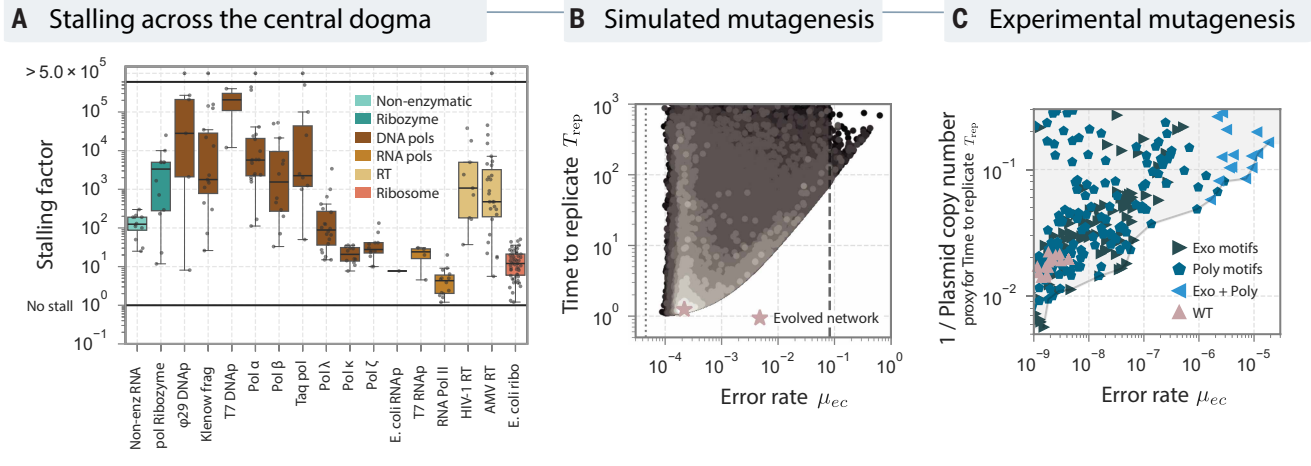


Fig. 2. Stalling measurements and exhaustive mutagenesis support speed-through-error correction. (**A**) Measurements of stalling factors across central dogma processes, curated from the literature; see supplementary text section 1.2.2 for references. Stalling factors are computed as the ratio between the rate of extending a mismatch and the rate of extending a matched base ($\sim\tau_{stall}/t_{proof}$). The upper line indicates values of the stalling factors $>5 \times 10^5$. In the box-and-whisker plot, the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values. Pols, polymerases; RT, reverse transcriptase. (**B** and **C**) Comprehensive mutagenesis reveals the relationship between speed and accuracy; see fig. S3. Shown are simulated (B) and experimental (C) measurements of speed and accuracy for mutants of a templated replication system. Simulated data are mutants of an evolved network model from Fig. 1, where the shading represents the strength of the perturbation from the evolved network. Experimental data are from a mutational screen of a DNA polymerase that copies a cytosolic plasmid in yeast (*29*). Data points of different shape indicate mutations in the exonuclease, polymerase domain, or both. Error rate μ is in substitutions per base. DNA plasmid copy number maintained at steady state by the dedicated DNA polymerase is a measure of its total activity, which combines speed and processivity; see materials and methods. The vertical dotted and dashed lines in (B) correspond to the minimal error rate achieved with proofreading and equilibrium discrimination, respectively. WT, wild type.

showing that the relevant mutation rate is not the post–error correction error rate, μ_{ec}, that is typically measured in experiments. Instead, the relevant rate is the error rate μ_0, before accounting for base excision, because the stalling time scales with this uncorrected error rate. This upgraded model extends stalling effects to much lower measured mutation rates μ_{ec} such as those relevant for DNA polymerases.

An orthogonal test of the relevance of our framework is to perturb the replication machinery, for example, by chemical conditions or mutagenesis, and then to measure the resulting change in both mutation rate μ_{ec} and replication time T_{rep}. If a decrease in μ_{ec} leads to a decrease in replication time T_{rep}, it implies that selection for faster replication will lower μ_{ec}, even if it does not directly implicate stalling as the mechanistic cause. In Fig. 2B, we show the result of a simulated mutagenesis experiment in a polymerase model evolved only for speed, as in Fig. 1B. We observed a clear trade-off: Error-prone polymerases are necessarily slow, whereas the faster polymerases are necessarily more accurate. We discuss a more complex model, specific to extant DNA polymerases with low mutation rates, in supplementary text section 1.2.3.

Several prior studies have examined speed-accuracy trade-offs by considering only the impact of a few mutations or one biophysical perturbation, for example, changing Mg^{2+} levels (*28*). However, a limited set of perturbations to a well-adapted enzyme can suggest misleading trade-offs (see fig. S2). Here, we compared our theoretical prediction to a large library of DNA polymerase mutants built recently

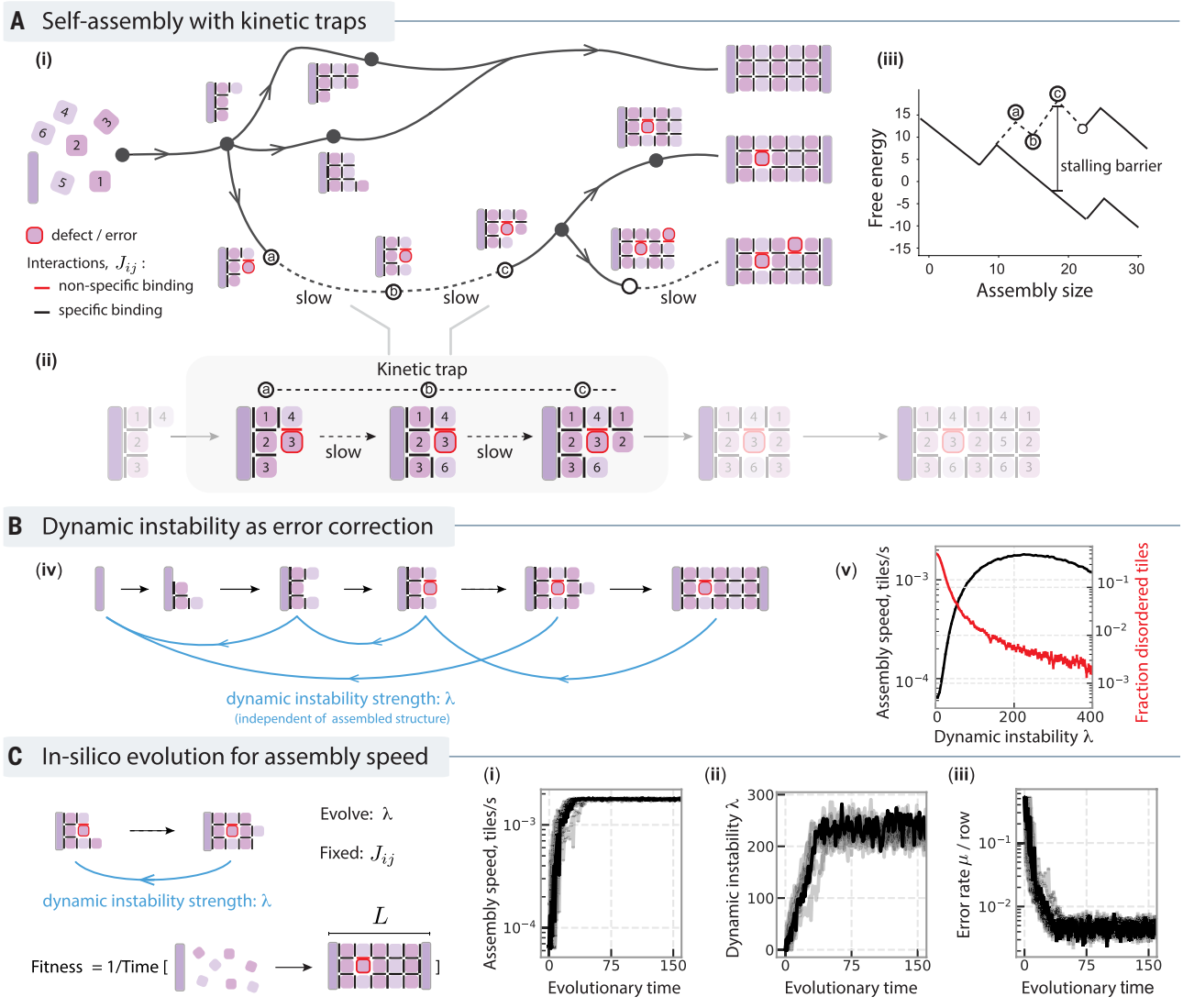


Fig. 3. Speeding up self-assembly selects for dynamic instability–based error correction. (**A**) We considered a system of N molecular species ($N = 6$ in this figure) with interactions J_{ij} that enable self-assembly of long (ii) ribbon-like structures. (i and ii) For a wide class of structures (*35*,*36*), misincorporations due to random nonspecific interactions can lead to kinetic traps, that is, geometrically frustrated partial assemblies that stall further growth. The free energy of partially assembled structures as a function of the assembly size (iii) shows free-energy barriers (dashed line) that must be overcome after misincorporations in order to continue growth. (**B**) (iv) We allowed for the evolution of nonequilibrium dynamic instability that induces frequent irreversible disassembly during assembly. (v) In silico simulations show that stronger dynamic instability results in a higher speed of assembly while also decreasing the fraction of disordered tiles. (**C**) In silico evolution with fitness set by the speed of assembly for any structure of length L, independent of component ordering. Molecular interactions J_{ij} are held fixed, but the dynamic instability strength λ is allowed to evolve. Despite selecting for higher assembly speed alone (i) and not on ordered assembly, (ii) dynamic instability strength λ increases and, concurrently, (iii) structural errors, defined as the variation in component ordering in structures of size L, decrease.

(*29*) during the creation of the OrthoRep directed evolution platform (*30*). We plotted the data generated by (*29*) for 213 single and multiple mutant variants of the pGKL1 DNA polymerase, a viral-origin family-B DNA polymerase with exonuclease activity that is homologous to the well-studied ϕ29 DNA polymerase. We took the copy number of the p1 plasmid, which is exclusively replicated by the pGKL1 DNA polymerase, to be a proxy for the overall polymerase activity (see fig. S3 and supplementary text section 1.3).

Similar to the trade-off seen in the simulated mutagenesis plot (Fig. 2B), the experimental data in Fig. 2C show that the "accurate is faster" relationship holds across four orders of magnitude in error rate and two orders of magnitude in replication activity. The absence of variants in the fast and inaccurate (lower right) region suggests that, starting from a slow, error-prone polymerase [e.g., $\mu_{ec} = 10^{-5}$ substitutions per base (spb) in Fig. 2C], selection for replication speed alone could favor the evolution of proofreading and thus increase accuracy. Comparable trends have been noted in other DNA polymerase screens, although with a smaller dynamic range or additional caveats (*31*, *32*) (see supplementary text section 1.4.2).

Self-assembly

The "evolution of error correction through selection for speed" mechanism above might seem limited to template-driven copying, because stalling arises from direct physical interactions between template and copy. However, analogous effects can emerge for other physical reasons even without a template, generalizing our result beyond templated replication. Here, we extend our results to the evolution of error correction during the self-assembly of multicomponent structures.

Numerous error-correcting mechanisms, including microtubule dynamic instability and ribosomal assembly checkpoints, achieve accuracy by disassembling incorrectly formed intermediate structures (*3*, *33*). These mechanisms can be viewed as analogous to exonuclease-based proofreading in that both disassemble incorrect intermediates (*34*). Although these disassembly mechanisms are usually thought to enhance accuracy at a cost to speed, we show here that such error-correcting disassembly pathways can instead spontaneously evolve as a result of selection for faster assembly alone.

We modeled the evolution of error correction for a self-assembling structure composed of N molecular species with both cognate (specific) and noncognate (nonspecific) interactions between them. The cognate interactions allow the molecules to spontaneously assemble into a desired structure; however, the noncognate interactions can lead to misassembled structures with components out of place (see fig. S4). Error correction is modeled by introducing a microtubule-inspired dynamic instability (*33*): a layer of nonequilibrium dynamics that allows structures to be partially or entirely disassembled irreversibly at some frequency λ (see fig. S5 and materials and methods for details of the model).

Notably, however, we did not impose a fitness cost for the erroneous structure; instead, fitness was set entirely by the time taken to build any structure of size L. This assumption is a theoretical device intended to isolate the most extreme consequences of "evolution of error correction through selection for speed." In practice, as in nucleic acid replication, both the speed and fidelity of self-assembly are likely to be under selection.

Selection for fast self-assembly produced two notable trends (Fig. 3C). First, the frequency of disassembly through dynamic instability λ increased (Fig. 3C, ii), despite selecting for fast assembly. Second, despite not selecting for ordered assembly, structural variation in fully assembled structures decreased (Fig. 3C, iii), that is, assembled structures tended to be ordered in a specific way.

These results can be explained by realizing that assembly in two and three dimensions has an emergent behavior, due to kinetic traps (*35*), that is analogous to stalling in nucleic acid polymerization. Defects in structure often slow further growth through geometric frustration effects (*36*). For example, as shown in Fig. 3, after a component is misplaced because of nonspecific interactions, further growth is possible but will be slower because the next components cannot find two specific partners to bind to. The resulting delay, analogous to stalling, has been studied in both tile assemblies (*36*) similar to that shown here and in more complex geometric models where similar effects arise from misplaced components in three-dimensional space (*37*).

Given this natural geometric mechanism that slows down assembly after misincorporations, dynamic instability speeds up assembly by disassembling misformed structures. In the absence of dynamic instability, such misformed leading edges form frequently and serve as kinetic traps that might take a long time to spontaneously dissociate. By disassembling these traps, dynamic instability, at an intermediate rate, can effectively speed up assembly (*38*) while also making final structures error free.

Note that we modeled a simple dynamic instability mechanism that does not directly distinguish correct from incorrect structures, but because incorrect intermediate structures have a longer lifetime than correctly formed intermediates, dynamic instability is more likely to

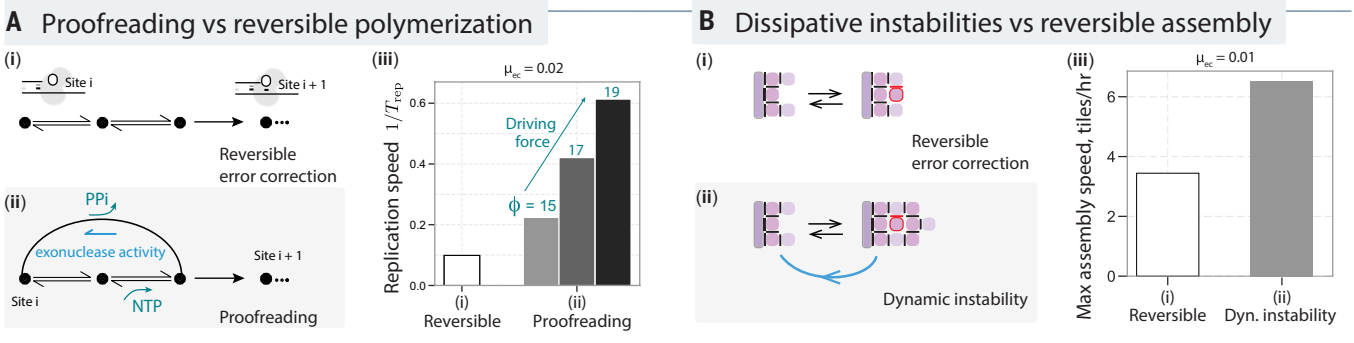


Fig. 4. A need for speed preferentially selects for more dissipative error-correcting mechanisms. (**A**) We compared polymerases (i) that remove errors by microscopically reversing polymerization pathways (here, a linear network) and (ii) with exonuclease-based error correction, that is, using alternative irreversible pathways (a network with loops). Upon selecting for speed, both networks evolve toward a lower error rate. However, when compared at the same error rate $\mu_{ec} = 0.02$, more dissipative mechanisms have higher replication speed, where the amount of dissipation is controlled by the nonequilibrium driving force ϕ. Stalling time $\tau_{stall} = 5$. (**B**) We compared reversible and irreversible error correction in self-assembly (i) through near-equilibrium self-assembly that exploits the microscopic reversal of assembly pathways and (ii) through dynamic instability that uses distinct irreversible disassembly pathways. Both mechanisms reduce defects while speeding up assembly, but the assembly speed of more dissipative mechanisms is higher when compared at the same defect rate $\mu_{ec} = 0.01$. See materials and methods for more details.

disassemble incorrect structures. In the cell, checkpoints can further bias disassembly to selectively target incorrect structures, further enhancing the "evolution of error correction through selection for speed" result shown here in a minimal model.

Preferential evolution of nonequilibrium mechanisms

At the molecular level, there are usually multiple mechanisms of error correction available. For example, besides exonuclease-based proofreading, the fidelity of a DNA polymerase can be improved by operating near the reversible equilibrium limit of polymerization (*6*). In this regime, nucleotides are frequently removed by the reverse of the polymerization reaction itself, rather than by a separate, irreversible proofreading pathway. Errors are still corrected because incorrect nucleotides tend to stall the polymerase and are more likely to be removed through this reversible process.

Our framework predicts that the need for speed favors more complex exonuclease-based proofreading over more reversible error-correcting mechanisms, even when both are compared at the same error rate (see Fig. 4 for numerical results). A mathematical derivation for a family of proofreading models is presented in the materials and methods.

Intuitively, highly irreversible mechanisms provide higher gains in fidelity $\Delta\mu_{ec}$ for a given amount of time Δt_{proof} spent on error correction (fig. S6). Because the overall speed is a balance between the time cost of error correction and the time benefit of avoiding stalling, faster error-correcting methods are preferred, even when they carry a direct cost in terms of energy expenditure (fig. S7). Similar results hold for self-assembly; an alternative to dynamic instability as an error-correction mechanism is simply carrying out assembly closer to equilibrium, where errors can dissociate more easily and allow forward assembly to continue (*39*). As detailed in the materials and methods, a need for speed preferentially selects for the more dissipative mechanism of dynamic instability instead.

Discussion

In this work, we have shown that selection for speed can drive the evolution of complex, nonequilibrium error-correcting mechanisms. We developed models of templated replication that incorporate post error stalling and showed that error correction can increase replication speed by preventing long stall events, thereby inverting the commonly assumed speed-fidelity trade-off. Our results are consistent with data on stalling effects across diverse biochemical processes and data on a "more accurate is faster" trade-off in a large mutational screen of a DNA polymerase. These observations predict that complex, nonequilibrium error-correcting mechanisms could evolve from a selection for speed alone, for example, without any selection for the preservation of the function encoded by parental sequences. By comparison with distinct processes of multicomponent self-assembly, we generalized these results beyond templated replication to argue that the evolution of error correction through selection for speed is a generic feature of biological assembly processes.

We used selection for speed alone as a theoretical device to illustrate the most notable consequences of stalling, but not necessarily as a literal evolutionary assumption. However, our results also have implications for components of the central dogma (*40*), which are likely under selection for both fidelity and speed. Although these two traits are often assumed to be in tension with each other, our framework predicts the opposite, that is, mutations may improve both fidelity and speed at the same time. This prediction will inform both studies of natural variation as well as efforts to engineer enzymes of higher specificity without loss of speed (*41*, *42*). In self-assembly, our results suggest that disassembly pathways can provide time-efficient error correction when combined with misincorporation-induced pauses seen in natural systems such as ribosomal assembly checkpoints (*3*) and synthetic systems (*36*, *37*, *43*). Our results suggest, counterintuitively, that classic annealing protocols (*44*) for reducing defects in crystal growth can also increase the net growth rate under some conditions.

A central implication of our work is that error-correcting mechanisms can arise as a by-product of selection for speed, rather than by being directly selected to preserve sequence-encoded information. This perspective can explain mutation rates that appear low from the standpoint of mutational load (*45*). For example, the mutation rate of the pGKL1 polymerase in Fig. 2 is 10^{-9} spb, whereas the toxin-antitoxin expressing plasmid it copies has a length $L = 10^4$ nucleotides. Such a low μ_{ec} (after error correction) can instead be explained by a selection for faster replication because, in this system, faster replication results in a higher plasmid copy number and thereby increases toxin-antitoxin expression. Assuming that plasmid copy number reflects overall polymerase activity (see caveats in supplementary text section 1.3), our framework predicts that this selection alone could be sufficient to maintain proofreading activity and the associated low mutation rate. See supplementary text section 4 for a discussion of when our results are expected to apply and when they might not.

Our work raises, but does not answer, the question of whether stalling is an evolvable trait or an unavoidable biophysical necessity. Figure 2 shows that more efficient enzymes exhibit substantially larger stalling factors (~10^5) than ribozyme-based or nonenzymatic systems (~10^1 to 10^2). This pattern is consistent with the hypothesis that selection for catalytic efficiency favors tighter active sites that amplify mismatch-induced stalling (*46*), a prediction that can be tested in future work.

In the context of the origins of life, the speed-first route to error correction suggests a way around the chicken-and-egg problem for fidelity. Large genomes need error correction to avoid the error catastrophe, yet canonical error-correcting machinery itself seems to require sizable, information-rich genomes. Our results suggest a resolution of this tension by showing that simple error-correcting mechanisms could evolve before the evolution of complex ribozymes such as an RNA replicase. The key mechanism is that misincorporations stall growth, so that variants with even rudimentary error correction replicate faster and become more abundant, even before other information-rich function-encoding sequences emerge.

Once such speed-selected error correction mechanisms are in place, the resulting incidental increase in fidelity can be co-opted to maintain and evolve increasingly information-rich sequences. For example, high-fidelity RNA replication would enable the evolution of new advantageous ribozymes, creating a secondary selective pressure to retain and refine the error-correcting mechanisms that initially evolved "for free" (*10*, *47*). This two-step dynamic—speed advantage first, functional payoff later—is an example of a complexity ratchet (*48*, *49*): Nonequilibrium machinery becomes entrenched not because complexity was needed, but because selection for speed favors it and subsequent functions lock it in place. Accordingly, the evolutionary path to error-correcting ribozymes and, by extension, to large, information-bearing genomes may be smoother than previously thought (*24*, *27*).

REFERENCES AND NOTES

1. T. A. Kunkel, K. Bebenek, *Annu. Rev. Biochem.* **69**, 497–529 (2000).
2. H. S. Zaher, R. Green, *Nature* **457**, 161–166 (2009).
3. Z. A. Sanghai, R. Piwowarczyk, A. V. Broeck, S. Klinge, *Nat. Struct. Mol. Biol.* **30**, 594–599 (2023).
4. J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4135–4139 (1974).
5. J. Ninio, *Biochimie* **57**, 587–595 (1975).
6. C. H. Bennett, *Biosystems* **11**, 85–91 (1979).
7. M. Lynch, W. Gabriel, *Evolution* **44**, 1725–1737 (1990).
8. J. F. Crow, M. Kimura, *An Introduction to Population Genetics Theory* (The Blackburn Press, 2010).
9. I. A. Chen, R. W. Roberts, J. W. Szostak, *Science* **305**, 1474–1476 (2004).
10. G. Laurent, T. Göppel, D. Lacoste, U. Gerland, *PRX Life* **2**, 013015 (2024).
11. S. Spiegelman, I. Haruna, I. B. Holland, G. Beaudreau, D. Mills, *Proc. Natl. Acad. Sci. U.S.A.* **54**, 919–927 (1965).

12. R. Mizuuchi, T. Furubayashi, N. Ichihashi, *Nat. Commun.* **13**, 1460 (2022).
13. B. Munsky, I. Nemenman, G. Bel, *J. Chem. Phys.* **131**, 235103 (2009).
14. P. Sartori, S. Pigolotti, *Phys. Rev. Lett.* **110**, 188101 (2013).
15. A. Murugan, D. A. Huse, S. Leibler, *Phys. Rev. X* **4**, 021016 (2014).
16. K. Banerjee, A. B. Kolomeisky, O. A. Igoshin, *J. Phys. Chem. Lett.* **8**, 1552–1556 (2017).
17. K. Banerjee, A. B. Kolomeisky, O. A. Igoshin, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5183–5188 (2017).
18. F. W. Perrino, L. A. Loeb, *J. Biol. Chem.* **264**, 2898–2905 (1989).
19. L. V. Mendelman, J. Petruska, M. F. Goodman, *J. Biol. Chem.* **265**, 2338–2346 (1990).
20. M.-M. Huang, N. Arnheim, M. F. Goodman, *Nucleic Acids Res.* **20**, 4567–4573 (1992).
21. S. J. Johnson, L. S. Beese, *Cell* **116**, 803–816 (2004).
22. J. K. Ichida, A. Horhota, K. Zou, L. W. McLaughlin, J. W. Szostak, *Nucleic Acids Res.* **33**, 5219–5225 (2005).
23. A. G. Baranovskiy, N. D. Babayeva, A. E. Lisova, L. M. Morstadt, T. H. Tahirov, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2111744119 (2022).
24. T. Göppel, B. Obermayer, I. A. Chen, U. Gerland, bioRxiv 2021.08.06.455386 [Preprint] (2021); https://doi.org/10.1101/2021.08.06.455386.
25. Y. J. Matsubara, N. Takeuchi, K. Kaneko, *Phys. Rev. Res.* **5**, 013170 (2023).
26. T. Midha, J. D. Mallory, A. B. Kolomeisky, O. A. Igoshin, *J. Phys. Chem. Lett.* **14**, 3422–3429 (2023).
27. S. Rajamani *et al.*, *J. Am. Chem. Soc.* **132**, 5880–5885 (2010).
28. M. Johansson, M. Lovmar, M. Ehrenberg, *Curr. Opin. Microbiol.* **11**, 141–147 (2008).
29. A. Ravikumar, G. A. Arzumanyan, M. K. A. Obadi, A. A. Javanpour, C. C. Liu, *Cell* **175**, 1946–1957.e13 (2018).
30. A. Ravikumar, A. Arrieta, C. C. Liu, *Nat. Chem. Biol.* **10**, 175–177 (2014).
31. E. Loh, J. Choe, L. A. Loeb, *J. Biol. Chem.* **282**, 12201–12209 (2007).
32. C. S. Diercks *et al.*, *Science* **389**, 618–622 (2025).
33. T. Mitchison, M. Kirschner, *Nature* **312**, 237–242 (1984).
34. A. Murugan, D. A. Huse, S. Leibler, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12034–12039 (2012).
35. A. Jhaveri, S. Loggia, Y. Qian, M. E. Johnson, *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2403384121 (2024).
36. E. Winfree, R. Bekbolatov, *Lect. Notes Comput. Sci.* **2943**, 126–144 (2004).
37. M. Lenz, T. A. Witten, *Nat. Phys.* **13**, 110–1104 (2017).
38. T. E. Holy, S. Leibler, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5682–5685 (1994).
39. C. Rossi-Gendron *et al.*, *Nat. Nanotechnol.* **18**, 1311–1318 (2023).
40. C. Bustamante, W. Cheng, Y. X. Mejia, *Cell* **144**, 480–497 (2011).
41. D. S. Tawfik, *Curr. Opin. Chem. Biol.* **21**, 73–80 (2014).
42. A. I. Flamholz *et al.*, *Biochemistry* **58**, 3365–3376 (2019).
43. C. G. Evans, E. Winfree, in *DNA Computing and Molecular Programming*, vol. 11145, Lecture Notes in Computer Science Series, D. Doty, H. Dietz, Eds. (Springer, 2018), pp. 37–54.
44. G. Müller, *Cryst. Res. Technol.* **42**, 1150–1161 (2007).
45. M. Lynch, *Trends Genet.* **26**, 345–352 (2010).
46. V. A. Maola *et al.*, *Nat. Catal.* **7**, 1173–1185 (2024).
47. K. F. Tjhung, M. N. Shokhirev, D. P. Horning, G. F. Joyce, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 2906–2913 (2020).
48. G. K. A. Hochberg *et al.*, *Nature* **588**, 503–508 (2020).
49. A. S. Pillai *et al.*, *Nature* **581**, 480–485 (2020).
50. R. Ravasio *et al.*, Code. Zenodo (2025); https://doi.org/10.5281/zenodo.17554537.
51. R. Ravasio *et al.*, Simulation data. Zenodo (2025); https://doi.org/10.5281/zenodo.17478191.

ACKNOWLEDGMENTS

We thank J. Sutherland, I. Chen, D. Huse, M. Rust, E. Winfree, C. Liu, and members of the Murugan group and the CZI theory group for discussions. **Funding:** A.M. and J.W.S. received support from the Sloan (G-2022-19518) and Moore (11479) foundations, Matter-to-Life program. R.R., A.M., and J.W.S. were supported by the NSF Center for Living Systems (grant no. 2317138). A.M. acknowledges support from NSF PHY-2310781 and the National Institute of General Medical Sciences (NIGMS) of the NIH under award no. R35GM151211. J.W.S. acknowledges support from NSF grant 2325198. J.W.S. is an Investigator of the Howard Hughes Medical Institute. R.P. acknowledges NIH Maximizing Investigators' Research Award (MIRA) 1R35 GM118043-01. C.G.E. acknowledges support from European Research Council (ERC) grant no. 772766, Science Foundation Ireland grant nos. 18/ERCS/5746 and 20/FFP-P/8843, and the Evans Foundation for Molecular Medicine. R.R. and K.H. acknowledge support from the Yen fellowship at the University of Chicago. This work has been supported by the University of Chicago's Research Computing Center. **Author contributions:** Conceptualization: R.R., K.H., C.G.E., R.P., M.R.-C., J.W.S., A.M.; Writing and editing: R.R., K.H., C.G.E., R.P., M.R.-C., J.W.S., A.M.; Investigation and analysis: R.R., K.H., C.G.E., A.M. **Competing interests:** The authors declare no competing interests. **Data, code, and materials availability:** Code and simulation data are available at Zenodo (*50*, *51*). **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. This research was also funded in whole or in part by ERC (772766); as required, the author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adt1275
Materials and Methods; Supplementary Text; Figs. S1 to S7; References (*52*–*163*);
MDAR Reproducibility Checklist

Submitted 18 September 2024; resubmitted 27 June 2025; accepted 5 January 2026
10.1126/science.adt1275

Evolution of error correction through a need for speed

Riccardo Ravasio, Kabir Husain, Constantine G. Evans, Rob Phillips, Marco Ribezzi-Crivellari, Jack W. Szostak, and Arvind Murugan

Science **391** (6787), . DOI: 10.1126/science.adt1275

Editor's summary
Life expends substantial effort correcting errors introduced by the forces of disorder. Molecular machines, from polymerases to the ribosome, use elaborate proofreading schemes to double-check their work, reducing errors at the cost of time and energy. Such mechanisms may not evolve easily because they devote precious resources to fixing mistakes. However, mistakes themselves also impose a time penalty, an effect known as stalling. Ravasio *et al*. present a theoretical analysis showing that because of stalling, error-correction mechanisms can in fact evolve to accelerate biological processes. The researchers found experimental support for this conclusion across a wide range of systems, from genome replication to the assembly of complex molecular structures. —Di Jiang