THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL Letter

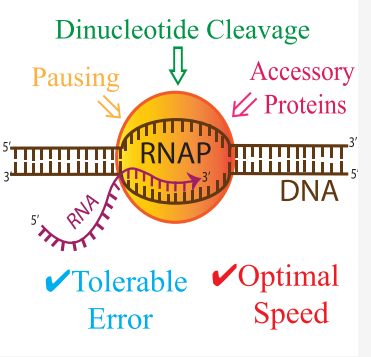
Synergy among Pausing, Intrinsic Proofreading, and Accessory Proteins Results in Optimal Transcription Speed and Tolerable Accuracy

Tripti Midha, Joel D. Mallory, Anatoly B. Kolomeisky,* and Oleg A. Igoshin*

Cite This: *J. Phys. Chem. Lett.* 2023, 14, 3422−3429 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Cleavage of dinucleotides after the misincorporational pauses serves as a proofreading mechanism that increases transcriptional elongation accuracy. The accuracy is further improved by accessory proteins such as GreA and TFIIS. However, it is not clear why RNAP pauses and why cleavage-factor-assisted proofreading is necessary despite transcriptional errors *in vitro* being of the same order as those in downstream translation. Here, we developed a chemical−kinetic model that incorporates most relevant features of transcriptional proofreading and uncovers how the balance between speed and accuracy is achieved. We found that long pauses are essential for high accuracy, whereas cleavage-factor-stimulated proofreading optimizes speed. Moreover, in comparison to the cleavage of a single nucleotide or three nucleotides, RNAP backtracking and dinucleotide cleavage improve both speed and accuracy. Our results thereby show how the molecular mechanism and the kinetic parameters of the transcriptional process were evolutionarily optimized to achieve maximal speed and tolerable accuracy.



The ability of cells to read and transfer molecular information requires discrimination among chemically similar substrates.[1] For instance, each enzyme involved in processes such as DNA replication, transcription, and translation must accurately distinguish correct versus incorrect substrates in each biochemical step.[2−4] Many of these processes display remarkable accuracy; e.g., DNA polymerases on average incorporate fewer than one incorrect base in the polynucleotide chain for every 10^7 steps.[2] To achieve such accuracy, these enzymes employ kinetic proofreading (KPR), a mechanism that reduces the number of errors below the equilibrium thermodynamic limit with additional non-equilibrium steps that are coupled to energy dissipation (e.g., ATP hydrolysis).[5,6] As a result, higher accuracy comes at the expense of increased energy costs and slower kinetics. Recent theoretical models have clarified how the free energy landscapes of enzymes evolved to optimize speed−accuracy−dissipation trade-offs for DNA replication, translation, tRNA charging, and coronavirus RNA genome replication.[7−14] For instance, evolution tuned the kinetic parameters of DNA polymerase and ribosome toward maximizing speed while still maintaining the the number of errors at tolerable levels. Structural investigations as well as *in vitro* and *in vivo* enzymatic studies of RNA polymerase (RNAP) showed that transcription in bacterial and eukaryotic cells also evolved to employ KPR.[15,16,18−23] However, the molecular details of the proofreading mechanisms are more complex than a classical Hopfield−Ninio picture.[5,6,24]

The proofreading mechanism for transcription by RNAP involves several unique features: pauses, backtracking, and endonucleolytic cleavage of multiple bases.[25−28] Pauses are the intervals where transcriptional elongation has temporarily ceased after the addition of a noncognate nucleotide. Backtracking, which sometimes follows pausing, involves an upstream translocation of RNAP by one or several bases. Following the backtracking, endonucleolytic cleavage of the most recently added ribonucleotides occurs.[22,28] Backtracking by one base pair, which is known to be energetically more favorable, leads to the endonucleolytic dinucleotide cleavage event.[29−31] The cleavage of two nucleotides at once provides two checkpoints for the cleavage of each misincorporated nucleotide. The first checkpoint enables the recently misincorporated nucleotide to be removed along with the previously incorporated nucleotide. The second checkpoint allows for the cleavage of the misincorporated nucleotide along with the next incorporated nucleotide. Furthermore, additional accessory proteins, such as GreA in the bacterium *Escherichia coli*, TFIIS in eukaryotes, and TFS in archaea, can increase transcriptional fidelity by preferentially stimulating the cleavage of misincorporated nucleotides.[28,32] *In vitro* experiments have

Received: February 6, 2023
Accepted: March 30, 2023

https://doi.org/10.1021/acs.jpclett.3c00345
J. Phys. Chem. Lett. 2023, 14, 3422−3429

shown that in eukaryotes, the RNA polymerase II subunit Rpb9 improves transcriptional fidelity by decreasing error propagation rates and increasing the error excision rate, which is facilitated by TFIIS.[33,34] Although Rpb9 does not directly contribute to the intrinsic cleavage activity of pol II, it is nevertheless required for the optimal function of the transcript cleavage factor TFIIS.[33] The mutation in Rpb9 can decrease the duration of pause, leading to a larger error.[34−36] Several studies have also argued that TFIIS can simultaneously decrease the duration of pauses.[37,38] A similar role is played by the accessory protein GreB in bacterial cells,[28] though not all bacterial species have its homologues.[39]

A comprehensive understanding of the roles of pauses, backtracking, and endonucleolytic cleavage is critically important for elucidating the molecular mechanisms of KPR for transcription. For this purpose, several kinetic models for RNAP transcription based on the sequence-dependent motion of RNAP and the experimentally determined standard free energies of the underlying steps have been formulated.[40−43] One of the models primarily focused on the initial substrate selection steps and, therefore, included only one proofreading checkpoint.[42] Another recently developed model considered the second step of proofreading but did not include the pauses at that step,[43] contrary to the experimental observations.[34] Moreover, none of the existing models take cleavage-factor-assisted proofreading into account, which limits their applications for *in vivo* systems. Thus, a complete understanding of how evolution optimizes the speed and accuracy of RNAP transcription is still lacking.

In this work, we address several key questions about the microscopic details of proofreading by RNAP. Which kinetic steps affect the speed and accuracy of the transcription, and are there trade-offs between these for all kinetic parameters? How does pausing after misincorporation affect transcriptional accuracy and speed? How much do the cleavage factors contribute to the transcription accuracy and speed? To answer these questions, we constructed a kinetic model that takes into account the most relevant features and applied the forward master equation formalism to analytically compute the error, speed, and the steady-state probabilities of distinct elongation states. We varied the rate constants of the underlying chemical reactions around their experimentally determined values to observe their effect on speed and accuracy. Furthermore, to understand the importance of backtracking and dinucleotide cleavage, we compared the predictions of our model with those of an alternative KPR model with no backtracking and single- or triple-nucleotide cleavage. The results allow us to formulate evolutionary design principles of the transcriptional process, i.e., uncover how it was optimized for the balance of speed and accuracy.

A more detailed description of RNAP transcription along the DNA template in the RNA synthesis is presented in Figure S1A. The process involves the reversible phosphodiester bond formation step and the proofreading step comprising backtracking and dinucleotide cleavage. In the presence of noncognate nucleotides, each phosphodiester bond formation can lead to two outcomes. The rate of each addition depends on whether the nucleotide that binds the elongation complex is cognate or noncognate to the current residue in the DNA template.[17,44] Moreover, if the noncognate nucleotide is incorporated at the last or second-to-last steps of elongation, RNAP pauses; i.e., the subsequent phosphodiester bond formation step is slower. In the same manner, the rate of dinucleotide cleavage also depends on the nature (i.e., cognate or noncognate) of the last two incorporated nucleotides.[27,34] Thus, to develop a kinetic model for transcription elongation, we have to explicitly account for the branching at every step. However, as explained in the Supporting Information, with periodic boundary conditions and with the assumption that noncognate incorporations are rare, one can describe the process with a simplified kinetic scheme with only six different elongation states as shown in Figure 1. Although the proposed

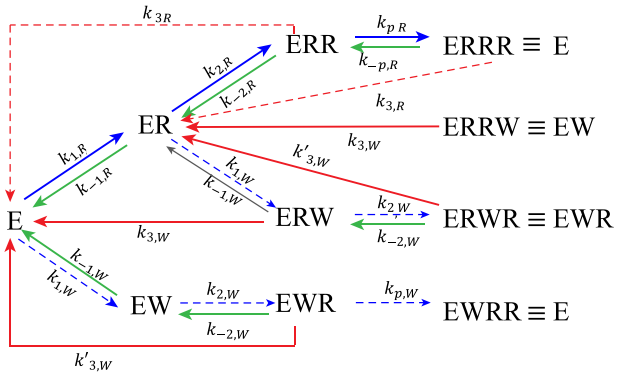


Figure 1. Schematic representation of the kinetic proofreading (KPR) model for transcription with six possible steady-state elongation states. State E represents the elongation complex in the middle of elongation after incorporation of two right (R) nucleotides, and the possibility of misincorporation at one of the two subsequent steps is accounted for. After the third forward elongation step, the resulting state is mapped to one of the preceding states as indicated. The forward and reverse elongation reactions are highlighted with blue and green arrows, respectively. Dinucleotide cleavage reactions are shown with red arrows. The dashed arrows indicate the reactions for which the rates are significantly smaller than the reaction rates represented by the solid lines of the same color.

model in Figure 1 is an approximate model of the true biochemical network of states, it is composed of the most relevant steps on the enzymatic pathways for which the rate constants have been determined experimentally. The rates corresponding to each reaction in the model are listed in Table S1. We relate the rate constants $k_{\pm i,\mathrm{W}}$ to the rate constants $k_{\pm i,\mathrm{R}}$ by introducing the discrimination factor $f_{\pm i} = \frac{k_{\pm i,\mathrm{W}}}{k_{\pm i,\mathrm{R}}}$, for i = 1, 2, 3, p. Factors f_i provide the kinetic discrimination between the R and W pathways. For example, pausing is characterized by discrimination factor f_2. It distinguishes the rate of extension after a mismatch $k_{2,\mathrm{W}}$, with the match extension rate $k_{2,\mathrm{R}}$.[34] A smaller value of f_2 implies a smaller mismatch extension rate and thus a longer duration of pauses after the addition of the wrong nucleotide.

To compute the transcriptional characteristic properties such as speed and error, we use the forward master equation formalism. It allows us to exactly compute the steady-state probabilities of the six elongation states (see the Supporting Information for the mathematical derivations). Notably, the steady-state probabilities of E, ER, and ERR are practically the same. Similarly, states EW and ERW have the same steady-state probability, as both represent the state with the first-time encounter with the wrong nucleotide. These steady-state probabilities can be used to obtain the stationary fluxes for each transition. We define the error rate $\eta = J_{\mathrm{W}}/J_{\mathrm{R}}$, i.e., as the ratio of stationary fluxes to create the wrong and right products

https://doi.org/10.1021/acs.jpclett.3c00345
J. Phys. Chem. Lett. 2023, 14, 3422−3429

(J_W and J_R, respectively). The RNAP speed (V) is defined as the sum of the stationary fluxes forming the final product states; i.e., $V = J_R + J_W \approx J_R$, as $J_W \ll J_R$.

We now explore how different features of the proofreading mechanism affect the speed−accuracy trade-off and the transcriptional dynamics. We analyze the error−speed trade-off plots (see, e.g., Figure 2D), generated by varying a single

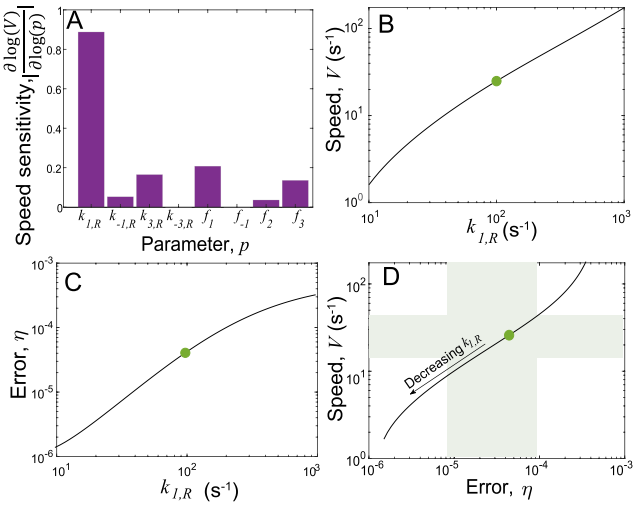


Figure 2. (A) Sensitivity of the reaction speed (V) to kinetic parameters and discrimination factors. (B) Variation of speed (V) with the phosphodiester bond formation rate $k_{1,R}$ (parameter with a maximum sensitivity value), with all other parameters fixed. (C) Variation of error η with $k_{1,R}$. (D) Trade-off plot that demonstrates the interplay of error and speed with changes in $k_{1,R}$. The green circle denotes the predicted value of speed and error at the experimentally determined kinetic parameters. The light green shaded region represents the experimentally measured ranges of error (vertical) and speed (horizontal), in the absence of accessory proteins.

rate constant $k_{i,R}$ or discrimination factor f_i to see how evolution changes the speed and error while keeping all of the other kinetic parameters fixed at their experimentally determined values (see Table S1). We say that there is a trade-off between accuracy and speed if an increase in either speed or accuracy corresponds to a decrease in the other.

Speedy transcription elongation is important to ensure the timely synthesis of RNA and rapid changes in gene expression profiles. In our model, this elongation rate is controlled by the kinetic parameters of different steps (Figure 1). To understand which steps are the most relevant in affecting the speed locally (i.e., around the native parameter values), we performed the sensitivity analysis of speed with respect to all independent kinetic parameters. The results (Figure 2A) indicate that the reaction speed is most sensitive to changes in the native phosphodiester bond formation rate, $k_{1,R}$. Other parameters, such as cleavage rate $k_{3,R}$, can also affect the speed but to a much lesser degree. To understand the effect more globally, we compute the speed with $k_{1,R}$ varying over 2 orders of magnitude (Figure 2B). The results imply that the speed of the system can monotonically increase with the phosphodiester bond formation rate. Notably, the experimentally estimated $k_{1,R}$ value corresponds to an intermediate value of speed. To determine why the transcription elongation does not occur at a higher phosphodiester bond formation rate, we analyzed how the error rate is affected by changes in this rate. The results depicted in Figure 2C demonstrate that increasing the polymerization rate also monotonically increases the error. Thus, a change in the phosphodiester bond formation rate can accelerate the transcription at the cost of a higher error or can increase the transcriptional accuracy at a cost of slower elongation. This trade-off is visualized in Figure 2D. Therefore, the naturally selected value of this rate seems to be determined by these competing considerations, leading to suboptimal speed and error rates.

Given the large effect of the phosphodiester bond formation rate on speed, it is important to understand why this rate is occasionally reduced during transcription elongation. This sporadic deceleration can be classified into two categories: sequence-dependent and sequence-independent pauses. Sequence-dependent pauses occur due to RNA hairpin formation[45] with a frequency of ≤1 pause per 100 nucleotides.[46] These pauses are generally short and are not associated with backtracking leading to oligonucleotide cleavage.[47] On the contrary, sequence-independent pauses are randomly distributed and occur once per 1000 nucleotides, which is a rate equivalent to the nucleotide misincorporation rate. These pauses are associated with backtracking leading to the cleavage of recently misincorporated nucleotides.[28] In our model, we incorporate sequence-independent pauses only after a misincorporation because of their role in proofreading.[48]

For the pauses caused by misincorporation, the reduction of the rate after mismatch is described by discrimination factor f_2. Therefore, we analyze the effects of f_2 on the speed−accuracy trade-off curve. The increase in this parameter ($f_2 \to 1$) would accelerate phosphodiester bond formation after the misincorporation eliminating the pauses. The results depicted in Figure 3A show that with these changes both the error and the speed monotonically increase. However, the magnitude of the increase in the speed is quite low compared to changes in the error: while the error corresponding to the native system with pauses (green circle) is approximately 1 order of magnitude

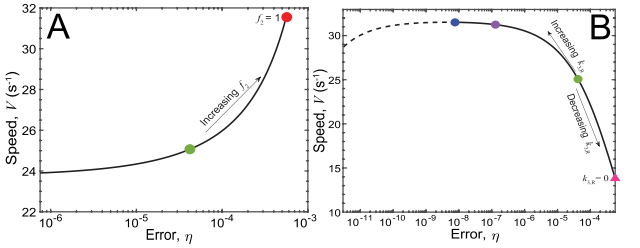


Figure 3. (A) Effect of varying the duration of pauses on the error and speed by changing mismatch extension discrimination factor f_2. The green color represents the native system with pausing after incorporation. As f_2 decreases, the duration of pauses decreases. The red circle denotes the system with no pausing after misincorporation, i.e., $f_2 = 1$. (B) Error−speed (η−V) interplay for the dinucleotide cleavage step. The native system with intrinsic proofreading (green circle) lies on the non-trade-off branch (solid line) of the η−V curve. To the left of the optimal (maximum) speed lies the trade-off branch (dotted line). The pink triangle denotes the system in the absence of cleavage ($k_{3,R} \to 0$). The purple circle represents the system for *E. coli* bacteria with GreA increasing the native intrinsic value of $k_{3,R}$ by 25 times, while the blue circle characterizes the system for eukaryotes in the presence of the accessory protein TFIIS that increases the intrinsic cleavage rate by only 100 times.

https://doi.org/10.1021/acs.jpclett.3c00345
J. Phys. Chem. Lett. 2023, 14, 3422−3429

less than that of the system without pauses (red circle), the speed changes by <20%. Thus, the pauses before the cleavage improve the accuracy significantly with only a minor effect on the speed.

Our theoretical analysis also suggests (Figure 2A) that the speed is also moderately sensitive to changes in cleavage rate $k_{3,R}$. Therefore, we investigated how the changes in $k_{3,R}$ affect speed and accuracy. Interestingly, the results depicted in Figure 3B show that there is no trade-off between accuracy and speed with changes in the cleavage rate. With an increase in $k_{3,R}$, the error decreases and the speed increases. In other words, the native system with intrinsic proofreading (green circle) lies on the non-trade-off curve in Figure 3B. The complete elimination of dinucleotide cleavage [that is, $k_{3,R} = 0$ (pink triangle)] will result in an error that is ∼1 order of magnitude higher and a speed that is ∼80% slower than those of the native case. This can be understood from the fact that the increase in the rate of cleavage decreases the relative time the enzyme spends in the wrong product states: EW, ERW, and EWR. This is evident from the decreased return of their corresponding probabilities $P_W = P_{ERW}$ and P_{EWR}, as shown in Figure S4. Thus, higher rates of cleavage increase the speed by releasing the paused enzyme from states EW, ERW, and EWR and improve the accuracy by reducing the probability of reaching the final wrong product state. However, a further increase in the cleavage rate can eventually lead to a decrease in the speed as the probability of futile cleavage of the correct nucleotide becomes significant (trade-off branch in Figure 3B, dashed line to the left of the blue circle). Therefore, there is a value of the cleavage rate that optimizes the speed (blue circle in Figure 3B). Interestingly, the native value corresponding to the intrinsic proofreading is significantly lower than this optimal value, resulting in a higher error and lower speed. Therefore, the mechanisms for further accelerating dinucleotide cleavage are expected to be beneficial.

Both bacterial and eukaryotic cells produce accessory proteins such as GreA in *E. coli* and TFIIS in eukaryotes, which bind to the polymerase during transcription and stimulate intrinsic cleavage. In *E. coli*, GreA is known to increase the cleavage efficiency of the RNAP enzyme by 25−35-fold.[49] The purple circle in Figure 3B represents the error and speed corresponding to the 25-fold increase in $k_{3,R}$. This assisted proofreading not only decreases the error by 2 orders of magnitude but also increases the speed by approximately 24%.[50,51] This increased speed is very close to the maximal value (blue circle). For eukaryotes, the protein factor TFIIS increases the cleavage efficiency by almost 100-fold (blue circle, Figure 3B). If that were the only effect of TFIIS, we would expect a further order of magnitude decrease in error and the optimal speed to be achieved. However, TFIIS not only increases the cleavage efficiency but also decreases the duration of pauses.[37,38]

To analyze the complete effect of TFIIS on the speed and accuracy of transcription, we need to understand the interplay between the two roles of these protein factors: reduction of pause duration and acceleration of cleavage. However, the magnitude of its effect on the duration of pauses following misincorporation has not been quantified. For the sake of simplicity, we assume that it decreases the duration of the pauses by the same magnitude as it increases the rate of cleavage. To model that, we introduce a stimulating factor α, by which both the cleavage rate and the mismatch extension rate are increased, and investigate the effect of these changes on speed and accuracy. Figure 4A shows that increasing

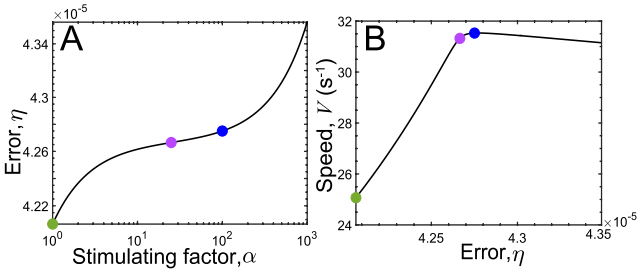


Figure 4. (A) Effect of simultaneously increasing both mismatch extension and cleavage rate by a stimulating factor α on the error rate η. (B) Change in the error−speed curve for stimulating factor α. The green circle represents the native system, corresponding to an α value of 1. The purple and blue circles correspond to α values of 25 and 100, respectively.

stimulating factor α does not significantly change the error. This is in contrast to the results in which these parameters are varied individually. When only the mismatch extension rate is increased, the error monotonically increases (Figure 3A). When the cleavage rate is increasing, the error significantly decreases (Figure 3B). Thus, two opposite effects of the mismatch extension rate and cleavage rate on error cancel one another when these rates are increased simultaneously. The error−speed curve indicates that while a 100-fold increase in these rates only slightly increases the error (blue circle, Figure 4A), it brings the speed to a nearly optimal value (blue circle, Figure 4B). Thus, we conclude that the primary role of TFIIS is the improvement of speed rather than accuracy.

Unlike the KPR mechanism for other biological processes, such as DNA replication, which involves the cleavage of a single misincorporated nucleotide, RNAP backtracks and cleaves two or more nucleotides at a time. However, the reason for this feature of the error correction mechanism by RNAP is still unknown. To investigate this, we analyzed how transcription accuracy and speed are affected if RNAP does not backtrack and cleaves only a single nucleotide. To this end, we developed a KPR scheme, as shown in Figure S2, consisting of the pausing but no backtracking and hence only single-nucleotide cleavage with the rates chosen to match with the corresponding rates for the model in Figure 1, given in Table S2. Two versions of the model without backtracking are analyzed. The first (Figure S2A) includes pauses after the first and second checkpoints, i.e., at both EW/ERW and EWR states. The second model (Figure S2B) includes pausing only at the first checkpoint, i.e., only after the most recent misincorporation (EW/ERW state). We computed the error and the speed for both models using the same methodology (see the Supporting Information). We first analyzed the model in the absence of accessory proteins. Interestingly, we found that, for the model with pausing at both checkpoints, error rate η (=1.8 × 10^{-4}) is ∼3.3 times higher than the error for the system with backtracking and dinucleotide cleavage [4.2 × 10^{-5} (green circle in Figure 1)]. Moreover, the speed [V = 22 nucleotides (nts)/s] is approximately 10% smaller than the speed computed for the model in Figure 1 ($V \approx$ 25 nts/s). Similarly, for the second model, the error increased to 2 × 10^{-4}, which is ∼3.8 times higher than that of the dinucleotide model. However, as the pausing at the second cleavage step is missing, it increases the speed to 26 nts/s. This speed is just 4% more than the speed for the dinucleotide cleavage model with pausing at both checkpoints (see Figure 5). Moreover, the

https://doi.org/10.1021/acs.jpclett.3c00345
J. Phys. Chem. Lett. 2023, 14, 3422−3429

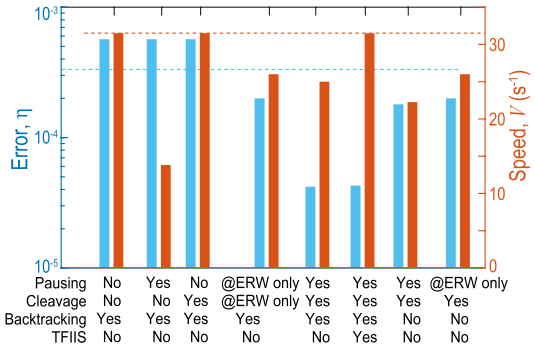


Figure 5. Pausing, intrinsic cleavage, backtracking, and accessory proteins (TFIIS) affecting the error (blue) and speed (red) of transcription by RNAP. The yes or no against each pausing, cleavage, backtracking, and TFIIS indicates their presence or absence in the model. The dashed blue line indicates the downstream translation error as an estimate of an error rate that is acceptable for the transcription process, whereas the dashed red line represents the maximum transcription speed in the absence of proofreading and pausing.

results hold even in the presence of TFIIS, which simultaneously affects the duration of pause and the intrinsic single-nucleotide cleavage by 100-fold. In that case, the error for the single-nucleotide cleavage model is still ∼3.5-fold larger than the error for the dinucleotide cleavage model, while the speeds for all models are almost the same.

The reason for the larger error for the models with no backtracking is the small chance of the removal of a misincorporated nucleotide from the second checkpoint EWR as compared to the dinucleotide cleavage model. This is evident from the steady-state probabilities of state EWR ($P_{EWR} \approx 0.14$) for the single-nucleotide cleavage model (Figure S2A), which is approximately 3-fold more than the P_{EWR} of ≈0.036 for the dinucleotide cleavage model (see Figure S3). Also, the probability of state E (the same as ER and ERR) ($P_E = P_{ER} = P_{ERR} \approx 0.23$) for the single-nucleotide cleavage model is 11% less than the corresponding probabilities ($P_E = P_{ER} = P_{ERR} \approx 0.26$) for the dinucleotide cleavage model (see Figure S3). Thus, the misincorporated nucleotide spends more time in state EWR and further tends to form a wrong product instead of being cleaved back to EW and then to E, hence decreasing the speed. Therefore, we conclude that for better accuracy and faster speed, backtracking and cleavage of dinucleotides are much better than single-nucleotide cleavage.

To further analyze if the cleavage of two nucleotides at once is optimal during transcription elongation, we constructed a KPR model with trinucleotide cleavage (Figure S5) with the corresponding kinetic parameters shown in Table S3. We observed that, in the absence of accessory proteins, and with pausing present at both checkpoints, the error for the system is 1.7×10^{-4}, which is 3-fold greater than that of the dinucleotide cleavage model. The speed for the trinucleotide cleavage model ($V \approx 13$ nts/s) is ∼93% less than that of the dinucleotide cleavage model. However, if we take pausing with the three neighbors into account, the error becomes 4.9×10^{-5}, which is approximately 16% more than that of the dinucleotide cleavage model. However, the speed is still $V \approx 13$ nts/s, which is 93% more than that of the dinucleotide cleavage model. This analysis suggests that the cleavage of two nucleotides at once during the transcription elongation proofreading mechanism is optimal for improving the transcriptional speed and accuracy.

The speed and accuracy of transcription are very important. The speed of transcription can directly contribute to the fitness of organisms;[52−54] e.g., in bacterial ribosome synthesis, the transcription of rRNA is the rate-limiting step and thereby affects the cell growth rate.[55] With respect to accuracy, transcriptional mistakes can be costly if mistranscribed RNA cannot perform its function or encodes a wrong amino acid. For mRNA, additional mistakes can be introduced during translation, so the error rate for these steps (∼3×10^{-4},[56,57]) can be used as an estimate of what is physiologically acceptable. To assess how different unique features of the RNAP proofreading mechanism affect speed and accuracy, we constructed a kinetic proofreading model that can account for the dinucleotide cleavage and the existence of the two checkpoints for proofreading. At each checkpoint state, the error correction depends on the competition between the extension rate and the cleavage rate. At the first checkpoint state following the attachment of the incorrect (W) nucleotide, dinucleotide "RW" is removed. If that does not happen, there is a second chance to remove the wrong nucleotide after another extension step with the cleavage of dinucleotide "WR". We aimed to keep the model quite generic so that it could be applied to both eukaryotic and prokaryotic transcription but nevertheless used realistic parameter estimates from the literature for the parameter values. The results indicate that pausing, backtracking, and cleavage assisted by accessory proteins allow for maximal transcriptional speed with tolerable accuracy (Figure 5).

In the model, RNAP pauses after a misincorporation event.[28] These pauses are sequence-independent and associated with backtracking leading to dinucleotide cleavage. Therefore, these pauses are expected to affect the error of the proofreading mechanism unlike the more frequent but shorter sequence-dependent pauses.[46,47] Our findings suggest that pauses following each checkpoint step are important for achieving high transcriptional accuracy. This allows the dinucleotide cleavage rate to effectively compete with the extension rate, but only when the wrong substrate is incorporated in one of the last two steps. Elimination of pauses increases the transcription speed but leads to a several-fold increase in the error. Complete elimination of proofreading (i.e., no pausing and cleavage) would lead to error and speed values very similar to those of the case with no pauses. These results indicate the essential role of pauses in the intrinsic proofreading mechanism. On the contrary, the presence of pausing in the absence of cleavage would only slow transcription without improving the accuracy. Furthermore, with only the first step of proofreading, the error decreases almost 3-fold with an ≈20% decrease in speed. The second step of proofreading further reduces the error by 5-fold at the cost of an only ≈3% loss of speed. This is in contrast to the result found in ref 43, where the contribution of the second step of proofreading in enhancing accuracy is quite low. The reason for the discrepancy is that the authors of ref 43 did not account for the pauses during the second checkpoint step. These pauses have been experimentally observed[28,34] and are essential to ensure that dinucleotide cleavage can even occur before the next elongation step.

To further understand the evolutionary selection of the kinetic parameters, we varied these parameters around their estimated values to evaluate their effects on the speed−

https://doi.org/10.1021/acs.jpclett.3c00345
J. Phys. Chem. Lett. 2023, 14, 3422−3429

accuracy trade-off behavior. The results showed that the trade-off between speed and accuracy is not universal, and its occurrence depends on the specific values of the kinetic rates. Observing that the transcriptional speed is most sensitive to the variation of the phosphodiester bond formation rate (see Figure 2A), we varied this rate around its native value and found that a higher speed can be achieved by increasing the phosphodiester bond formation rate at the cost of a higher error rate. Similarly, a higher accuracy can be achieved by decreasing the phosphodiester formation rate only at the price of a lower speed. Motivated by a study that found that the absence of the Rpb9 subunit in RNAP II could disrupt the fidelity of the transcription process by increasing the mismatch extension rate by 2−3-fold,[34] we analyzed the speed−accuracy trade-off behavior for the mismatch extension step. Again, a trade-off was found. Increasing the mismatch extension rate (i.e., reducing the duration of pauses) significantly increases the error of the system at the cost of a slightly lower speed. On the contrary, for the variation of the dinucleotide cleavage rate, there is a partitioning of the error−speed curve into trade-off and non-trade-off branches. Interestingly, near the experimentally determined value of the dinucleotide cleavage rate, the speed−accuracy curve exhibits non-trade-off behavior. Indeed, an optimal speed with a simultaneously higher accuracy can be observed for the significantly increased cleavage rate. A similar effect of increasing the cleavage has also been reported in ref 24.

An increase in the cleavage rate can be achieved with the help of cleavage-stimulating proteins such as GreA in *E. coli* bacteria and TFIIS in eukaryotes as well as subunit Rpb9 of RNAP II.[35] These factors yield a nearly optimal transcription speed (Figure 3B). Even when the acceleration of cleavage is accompanied by the removal of pauses, as is the case with TFIIS, the system still obtains the maximum possible speed but the error is almost the same as that with only intrinsic proofreading (Figure 4). This is consistent with the experimental finding that TFIIS is necessary *in vitro* to enhance the transcription elongation rate.[15,21,58] Moreover, these accessory proteins do not change the qualitative trade-off behavior seen for the phosphodiester bond formation rate (see Figure S6). Thus, it seems that evolution has optimized the kinetic parameters involved in the transcription by RNA polymerase to achieve the maximum possible speed with the desired tolerable accuracy.

It is known that the proofreading step requires the consumption of free energy through the hydrolysis of energy-rich nucleotide triphosphate (NTP) molecules and therefore dissipates energy through futile cycles.[7,59,60] To avoid an unnecessary waste of energy, the biological systems aim to keep the amount of energy dissipated small.[61] For example, for translation during the tRNA charging step and selection step during protein synthesis by an *E. coli* ribosome, a notable increase in the expenditure of energy from proofreading prevented evolution from completely optimizing the speed and accuracy of these processes.[7,8,10] In both of these examples, the proofreading cost was significant with 3−13% of the nucleotide hydrolysis not associated with product formation.[7,10] Given the large cost, the system avoided increasing the proofreading cost to further increase the speed.[62] This raises the question of why the transcription process achieves its optimal speed via cleavage-factor-assisted proofreading if it could increase the proofreading dissipation energy. To understand this, we computed the proofreading cost for transcription and showed that its value is 2.7×10^{-4} for intrinsic proofreading and 1.3×10^{-3} for the cleavage-assisted case. TFIIS-assisted cleavage increases the speed by 26% at the proofreading cost of ∼1 order of magnitude. This implies that RNAP II could bear this proofreading cost for a significantly large increase in speed.

All experimental studies for proofreading by RNAP during transcription elongation have claimed the cleavage of at least two nucleotides. The studies never observed the cleavage of a single nucleotide. To understand why RNAP does not cleave a single nucleotide, unlike in other proofreading processes by different enzymes,[63] we also constructed a kinetic proofreading model with single-nucleotide cleavage. It was found that the error for the single-nucleotide cleavage model with pausing having either a one-neighbor effect or a two-neighbor effect is always larger (∼3.8-fold) than in our proposed system with the dinucleotide cleavage. This high error for the single-nucleotide cleavage model remains unaffected in the presence of accessory proteins, affecting both the duration of the pause and cleavage rate. However, the speed for the single-nucleotide cleavage model is less than the speed for our proposed model in the absence of accessory proteins, and when they are present, all models operate at an optimal speed. This implies that the requirement to achieve high accuracy prevents RNAP from cleaving single nucleotides.

In summary, our study presents a coherent understanding of why RNAP pauses, backtracks, and cleaves dinucleotides based on how these processes affect accuracy and speed. Our theoretical analysis shows how the ultimate balance between speed and accuracy is achieved by adjusting various kinetic rates and through the effects of accessory proteins. It remains to be seen how these parameters vary among different organisms to fine-tune the speed−accuracy trade-off to the specific demands.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.3c00345.

Detailed description of the KPR model for transcription elongation by RNAP and quasi-equilibrium approach to eliminate intermediate steps; experimentally determined kinetic rates and discrimination factors; forward master equation formalism for computing steady-state probabilities for the dinucleotide cleavage model with backtracking and the single-nucleotide cleavage model without backtracking; definitions of error rate η, speed V, and proofreading cost C in terms of stationary fluxes; steady-state probabilities and characteristic properties for the KPR models computed with the native experimental kinetic parameters; sensitivity of the speed to various kinetic parameters of the model; and experimental data for the error and speed of mRNA transcription by RNA polymerase (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Anatoly B. Kolomeisky − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States; Department of Chemistry, Department of Chemical and Biomolecular Engineering, and Department of Physics and Astronomy, Rice University, Houston, Texas 77005,*

https://doi.org/10.1021/acs.jpclett.3c00345
J. Phys. Chem. Lett. 2023, 14, 3422−3429

United States; orcid.org/0000-0001-5677-6690; Email: tolya@rice.edu

Oleg A. Igoshin − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States; Department of Chemistry, Department of Bioengineering, and Department of Biosciences, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0002-1449-4772; Email: igoshin@rice.edu

Authors

Tripti Midha − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States*

Joel D. Mallory − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States;* Present Address: J.D.M.: Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.3c00345

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful to Gene-Wei Li, Robert Raphael, and Yang Gao for their comments on the draft of the manuscript. This work was supported by Center for Theoretical Biological Physics National Science Foundation (NSF) Grant PHY-2019745. O.A.I. also acknowledges support from Welch Foundation Grant C-1995. A.B.K. also acknowledges support from the Welch Foundation (C-1559) and the NSF (CHE-1953453 and MCB-1941106).

■ REFERENCES

(1) Alberts, B.; Bray, D.; Hopkin, K.; Johnson, A. D.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Essential cell biology*; Garland Science, 2015.

(2) Kunkel, T. A.; Bebenek, K. DNA replication fidelity. *Annu. Rev. Biochem.* **2000**, *69*, 497.

(3) Zaher, H. S.; Green, R. Fidelity at the molecular level: lessons from protein synthesis. *Cell* **2009**, *136*, 746−762.

(4) Reynolds, N. M.; Lazazzera, B. A.; Ibba, M. Cellular mechanisms that control mistranslation. *Nat. Rev. Microbiol* **2010**, *8*, 849−856.

(5) Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. U. S. A* **1974**, *71*, 4135−4139.

(6) Ninio, J. Kinetic amplification of enzyme discrimination. *Biochimie* **1975**, *57*, 587−595.

(7) Banerjee, K.; Kolomeisky, A. B.; Igoshin, O. A. Elucidating interplay of speed and accuracy in biological error correction. *Proc. Natl. Acad. Sci. U. S. A* **2017**, *114*, 5183−5188.

(8) Mallory, J. D.; Kolomeisky, A. B.; Igoshin, O. A. Trade-offs between error, speed, noise, and energy dissipation in biological processes with proofreading. *J. Phys. Chem. B* **2019**, *123*, 4718−4725.

(9) Chiuchiú, D.; Tu, Y.; Pigolotti, S. Error-speed correlations in biopolymer synthesis. *Phys. Rev. Lett.* **2019**, *123*, 038101.

(10) Yu, Q.; Mallory, J. D.; Kolomeisky, A. B.; Ling, J.; Igoshin, O. A. Trade-offs between speed, accuracy, and dissipation in tRNAIle aminoacylation. *J. Phys. Chem. Lett.* **2020**, *11*, 4001−4007.

(11) Mallory, J. D.; Mallory, X. F.; Kolomeisky, A. B.; Igoshin, O. A. Theoretical analysis reveals the cost and benefit of proofreading in coronavirus genome replication. *J. Phys. Chem. Lett.* **2021**, *12*, 2691−2698.

(12) Boeger, H. Kinetic Proofreading. *Annu. Rev. Biochem.* **2022**, *91*, 423.

(13) Zhang, D.; Ouyang, Q. Nonequilibrium thermodynamics in biochemical systems and its application. *Entropy* **2021**, *23*, 271.

(14) Sahoo, M.; N, A.; Baral, P.; Klumpp, S. Accuracy and speed of elongation in a minimal model of DNA replication. *Phys. Rev. E* **2021**, *104*, 034417.

(15) Thomas, M. J.; Platas, A. A.; Hawley, D. K. Transcriptional fidelity and proofreading by RNA polymerase II. *Cell* **1998**, *93*, 627−637.

(16) Erie, D. A.; Hajiseyedjavadi, O.; Young, M. C.; von Hippel, P. H. Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science* **1993**, *262*, 867−873.

(17) Sydow, J. F.; Brueckner, F.; Cheung, A. C.; Damsma, G. E.; Dengl, S.; Lehmann, E.; Vassylyev, D.; Cramer, P. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell* **2009**, *34*, 710−721.

(18) Rosenberger, R.; Foskett, G. An estimate of the frequency of in vivo transcriptional errors at a nonsense codon in Escherichia coli. *Molecular and General Genetics MGG* **1981**, *183*, 561−563.

(19) Erie, D. A.; Yager, T. D.; von Hippel, P. H. The single-nucleotide addition cycle in transcription: a biophysical and biochemical perspective. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 379−415.

(20) Gout, J.-F.; Thomas, W. K.; Smith, Z.; Okamoto, K.; Lynch, M. Large-scale detection of in vivo transcription errors. *Proc. Natl. Acad. Sci. U. S. A* **2013**, *110*, 18584−18589.

(21) Jeon, C.; Agarwal, K. Fidelity of RNA polymerase II transcription controlled by elongation factor TFIIS. *Proc. Natl. Acad. Sci. U. S. A* **1996**, *93*, 13677−13682.

(22) Surratt, C. K.; Milan, S. C.; Chamberlin, M. J. Spontaneous cleavage of RNA in ternary complexes of Escherichia coli RNA polymerase and its significance for the mechanism of transcription. *Proc. Natl. Acad. Sci. U. S. A* **1991**, *88*, 7983−7987.

(23) Izban, M. G.; Luse, D. S. The RNA polymerase II ternary complex cleaves the nascent transcript in a 3′—-5′direction in the presence of elongation factor SII. *Genes Dev.* **1992**, *6*, 1342−1356.

(24) Sahoo, M.; Klumpp, S. Backtracking dynamics of RNA polymerase: pausing and error correction. *J. Condens. Matter Phys.* **2013**, *25*, 374104.

(25) Orlova, M.; Newlands, J.; Das, A.; Goldfarb, A.; Borukhov, S. Intrinsic transcript cleavage activity of RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A* **1995**, *92*, 4596−4600.

(26) Epshtein, V.; Nudler, E. Cooperation between RNA polymerase molecules in transcription elongation. *Science* **2003**, *300*, 801−805.

(27) Zenkin, N.; Yuzenkova, Y.; Severinov, K. Transcript-assisted transcriptional proofreading. *Science* **2006**, *313*, 518−520.

(28) Shaevitz, J. W.; Abbondanzieri, E. A.; Landick, R.; Block, S. M. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* **2003**, *426*, 684−687.

(29) Wang, D.; Bushnell, D. A.; Huang, X.; Westover, K. D.; Levitt, M.; Kornberg, R. D. Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* **2009**, *324*, 1203−1206.

(30) Sydow, J. F.; Cramer, P. RNA polymerase fidelity and transcriptional proofreading. *Curr. Opin. Struct. Biol.* **2009**, *19*, 732−739.

(31) Bochkareva, A.; Yuzenkova, Y.; Tadigotla, V. R.; Zenkin, N. Factor-independent transcription pausing caused by recognition of the RNA−DNA hybrid sequence. *EMBO journal* **2012**, *31*, 630−639.

(32) Lange, U.; Hausner, W. Transcriptional fidelity and proof-reading in Archaea and implications for the mechanism of TFS-induced RNA cleavage. *Mol. Microbiol.* **2004**, *52*, 1133−1143.

(33) Walmacq, C.; Kireeva, M. L.; Irvin, J.; Nedialkov, Y.; Lubkowska, L.; Malagon, F.; Strathern, J. N.; Kashlev, M. Rpb9 Subunit Controls Transcription Fidelity by Delaying NTP Sequestration in RNA Polymerase II. *J. Biol. Chem.* **2009**, *284*, 19601−19612.

(34) Knippa, K.; Peterson, D. O. Fidelity of RNA polymerase II transcription: Role of Rbp9 in error detection and proofreading. *Biochem.* **2013**, *52*, 7807−7817.

(35) Nesser, N. K.; Peterson, D. O.; Hawley, D. K. RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. *Proc. Natl. Acad. Sci. U. S. A* **2006**, *103*, 3268−3273.
(36) Kireeva, M. L.; Nedialkov, Y. A.; Cremona, G. H.; Purtov, Y. A.; Lubkowska, L.; Malagon, F.; Burton, Z. F.; Strathern, J. N.; Kashlev, M. Transient Reversal of RNA Polymerase II Active Site Closing Controls Fidelity of Transcription Elongation. *Mol. Cell* **2008**, *30*, 557−566.
(37) Fish, R. N.; Kane, C. M. Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta* **2002**, *1577*, 287−307.
(38) Imashimizu, M.; Kireeva, M. L.; Lubkowska, L.; Gotte, D.; Parks, A. R.; Strathern, J. N.; Kashlev, M. Intrinsic translocation barrier as an initial step in pausing by RNA polymerase II. *J. Mol. Biol.* **2013**, *425*, 697−712.
(39) China, A.; Mishra, S.; Nagaraja, V. A transcript cleavage factor of Mycobacterium tuberculosis important for its survival. *PLoS One* **2011**, *6*, e21941.
(40) Bai, L.; Shundrovsky, A.; Wang, M. D. Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *J. Mol. Biol.* **2004**, *344*, 335−349.
(41) Guajardo, R.; Sousa, R. A model for the mechanism of polymerase translocation. *J. Mol. Biol.* **1997**, *265*, 8−19.
(42) Mellenius, H.; Ehrenberg, M. DNA template dependent accuracy variation of nucleotide selection in transcription. *PloS one* **2015**, *10*, e0119588.
(43) Mellenius, H.; Ehrenberg, M. Transcriptional accuracy modeling suggests two-step proofreading by RNA polymerase. *Nucleic Acids Res.* **2017**, *45*, 11582−11593.
(44) Yuzenkova, Y.; Bochkareva, A.; Tadigotla, V. R.; Roghanian, M.; Zorov, S.; Severinov, K.; Zenkin, N. Stepwise mechanism for transcription fidelity. *BMC Biol.* **2010**, *8*, 54.
(45) Shundrovsky, A.; Santangelo, T. J.; Roberts, J. W.; Wang, M. D. A Single-Molecule Technique to Study Sequence-Dependent Transcription Pausing. *Biophys. J.* **2004**, *87*, 3945−3953.
(46) Kireeva, M. L.; Kashlev, M. Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A* **2009**, *106*, 8900−8905.
(47) Neuman, K. C.; Abbondanzieri, E. A.; Landick, R.; Gelles, J.; Block, S. M. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **2003**, *115*, 437−447.
(48) James, K.; Gamba, P.; Cockell, S. J.; Zenkin, N. Misincorporation by RNA polymerase is a major source of transcription pausing in vivo. *Nucleic Acids Res.* **2016**, *45*, 1105−1113.
(49) Hogan, B. P.; Hartsch, T.; Erie, D. A. Transcript cleavage by thermus thermophilus RNA polymerase: effects of GreA and anti-GreA factors. *J. Biol. Chem.* **2002**, *277*, 967−975.
(50) Gamba, P.; Zenkin, N. Transcription fidelity and its roles in the cell. *Curr. Opin. Microbiol.* **2018**, *42*, 13−18.
(51) Bubunenko, M. G.; Court, C. B.; Rattray, A. J.; Gotte, D. R.; Kireeva, M. L.; Irizarry-Caro, J. A.; Li, X.; Jin, D. J.; Court, D. L.; Strathern, J. N.; et al. A Cre Transcription Fidelity Reporter Identifies GreA as a Major RNA Proofreading Factor in Escherichia coli. *Genetics* **2017**, *206*, 179−187.
(52) Maslon, M. M.; Braunschweig, U.; Aitken, S.; Mann, A. R.; Kilanowski, F.; Hunter, C. J.; Blencowe, B. J.; Kornblihtt, A. R.; Adams, I. R.; Cáceres, J. F. A slow transcription rate causes embryonic lethality and perturbs kinetic coupling of neuronal genes. *EMBO J.* **2019**, *38*, e101244.
(53) Yuzenkova, Y.; Gamba, P.; Herber, M.; Attaiech, L.; Shafeeq, S.; Kuipers, O. P.; Klumpp, S.; Zenkin, N.; Veening, J.-W. Control of transcription elongation by GreA determines rate of gene expression in Streptococcus pneumoniae. *Nucleic Acids Res.* **2014**, *42*, 10987−10999.
(54) Falcon, K. T.; Watt, K. E.; Dash, S.; Zhao, R.; Sakai, D.; Moore, E. L.; Fitriasari, S.; Childers, M.; Sardiu, M. E.; Swanson, S.; et al. Dynamic regulation and requirement for ribosomal RNA transcription during mammalian development. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, e2116974119.
(55) Zhu, M.; Mu, H.; Jia, M.; Deng, L.; Dai, X. Control of ribosome synthesis in bacteria: the important role of rRNA chain elongation rate. *Sci. China Life Sci.* **2021**, *64*, 795−802.
(56) Stelzl, U.; Connell, S.; Nierhaus, K. H.; Wittmann-Liebold, B. Ribosomal proteins: role in ribosomal function. *Encyclopedia of Life Sciences* **2001**, *2001*, 1−12.
(57) Mori, N.; Funatsu, Y.; Hiruta, K.; Goto, S. Analysis of translational fidelity of ribosomes with protamine messenger RNA as a template. *Biochem.* **1985**, *24*, 1231−1239.
(58) Agarwal, K.; Baek, K.; Jeon, C.; Miyamoto, K.; Ueno, A.; Yoon, H. Stimulation of transcript elongation requires both the zinc finger and RNA polymerase II binding domains of human TFIIS. *Biochem.* **1991**, *30*, 7842−7851.
(59) Qian, H.; Kjelstrup, S.; Kolomeisky, A. B.; Bedeaux, D. Entropy production in mesoscopic stochastic thermodynamics: nonequilibrium kinetic cycles driven by chemical potentials, temperatures, and mechanical forces. *J. Condens. Matter Phys.* **2016**, *28*, 153004.
(60) Libby, R. T.; Nelson, J. L.; Calvo, J.; Gallant, J. A. Transcriptional proofreading in Escherichia coli. *EMBO journal* **1989**, *8*, 3153−3158.
(61) Sartori, P.; Pigolotti, S. Kinetic versus energetic discrimination in biological copying. *Phys. Rev. Lett.* **2013**, *110*, 188101.
(62) Yu, Q.; Kolomeisky, A. B.; Igoshin, O. A. The energy cost and optimal design of networks for biological discrimination. *J. R. Soc., Interface* **2022**, *19*, 20210883.
(63) Mallory, J. D.; Igoshin, O. A.; Kolomeisky, A. B. Do we understand the mechanisms used by biological systems to correct their errors? *J. Phys. Chem. B* **2020**, *124*, 9289−9296.

https://doi.org/10.1021/acs.jpclett.3c00345
J. Phys. Chem. Lett. 2023, 14, 3422−3429