




Accuracy and speed of elongation in a minimal model of DNA replicationM. Sahoo ^{1,2,*}, Arsha N.,¹ P. R. Baral ² and S. Klumpp ^{3,†}¹*Department of Physics, University of Kerala, Kariavattom Campus 695581, India*²*School of Physics, Indian Institute of Science Education and Research, Thiruvananthapuram-695551, India*³*Institute for the Dynamics of Complex Systems, University of Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany*

(Received 13 July 2021; accepted 8 September 2021; published 28 September 2021)

Being a dual purpose enzyme, the DNA polymerase is responsible for elongation of the newly formed DNA strand as well as cleaving the erroneous growth in case of a misincorporation. The efficiency of replication depends on the coordination of the polymerization and exonuclease activity of DNA polymerase. Here, we propose and analyze a minimal kinetic model of DNA replication and determine exact expressions for the velocity of elongation and the accuracy of replication. We first analyze the case without exonuclease activity. In that case, accuracy is determined by a kinetic competition between stepping and unbinding, with discrimination between correct and incorrect nucleotides in both transitions. We then include exonuclease activity and ask how different modes of additional discrimination in the exonuclease pathway can improve the accuracy while limiting the detrimental effect of exonuclease on the speed of replication. In this way, we ask how the kinetic parameters of the model have to be set to coordinate the two activities of the enzyme for high accuracy and high speed. The analysis also shows that the design of a replication system does not universally have to follow the speed-accuracy trade-off rule, although it does in the biologically realized parameter range. The accuracy of the process is mainly controlled by the crucial role of stepping after erroneous incorporation, which has impact on both polymerase and exonuclease activities of DNA polymerase.

DOI: [10.1103/PhysRevE.104.034417](https://doi.org/10.1103/PhysRevE.104.034417)**I. INTRODUCTION**

The genetic information of a cell or an organism is stored in its DNA (deoxyribonucleic acid) and transmitted to the next generation through the process of DNA replication [1,2]. DNA replication is a complex process that involves multiple enzymes including helicases, nucleases, and topoisomerases in [3–7]. At its core, however, it is DNA polymerase (DNAP), which catalyzes the actual copying of the genetic information [8]. During the replication process, the DNAP acts as a molecular motor and moves step by step on the template DNA strand converting chemical energy into directed motion [9–11]. At the same time it copies the genetic information of the DNA sequence along which it moves into a newly synthesized DNA molecule with complementary sequence following the Watson-Crick complementary base pairing rule [2]. The movement of DNAP on the template occurs in stochastic deoxy-nucleotide triphosphate (dNTP) base steps. One single base step of the DNAP corresponds to a single-nucleotide elongation of the new DNA strand.

DNAP acts as a dual-purpose enzyme having polymerase as well as exonuclease activity during the replication process [12–15]. While its primary function is to elongate the newly formed DNA strand, it can also cleave the newly formed DNA strand and remove the last nucleotide when this has been transferred to the exonuclease site of the enzyme. This activity

functions as a proofreading mechanism: Upon misincorporation (incorporation of a wrong dNTP), the DNAP switches its functionality and the new strand is transferred to the exonuclease site where the wrong nucleotide is cleaved from the new strand through hydrolysis [16], before the DNA strand returns to the active site of polymerization for elongation to continue [14]. This mechanism represents an intrinsic proofreading mechanism known as exonucleolytic proofreading [1,2,16–19] as misincorporated nucleotides are excised, thus resetting the incorporation process, such that the errors can be corrected by a second attempt at correct incorporation. Polymerase and exonuclease sites of the DNAP are separated from each other by a distance of 3–4 nm [19–22], and the transfer of the new DNA strand between these two sites may involve one or many intermediates [23–25]. The dynamics of the transfer reaction is not understood in detail.

Just like wrong nucleotides can erroneously get incorporated during polymerization, compromising the fidelity of replication [26], correctly incorporated nucleotides can erroneously get transferred to the exonuclease site and be cleaved, resetting the incorporation process to the start and hence compromising the speed of elongation. Thus, the coordinated action of the polymerase and exonuclease activity of DNAP is crucial for an effective and faithful replication mechanism and determines the velocity and accuracy of replication.

It is usually argued that an enhancement of accuracy results in a slow down of the elongation process, resulting in a speed-accuracy trade-off [6,27–29]. The high fidelity or enhancement of accuracy is believed to be as a matter of proof-reading mechanism during the exonuclease activity

*jolly.iopb@gmail.com

†stefan.klumpp@phys.uni-goettingen.de

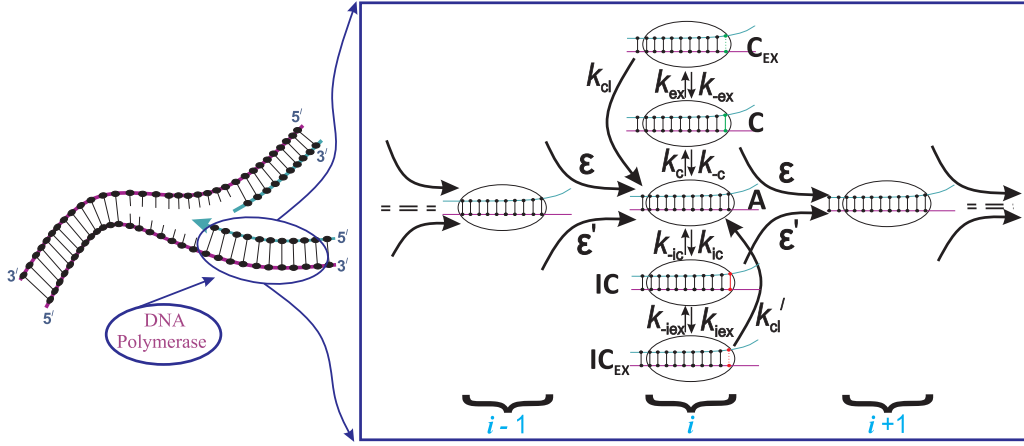


FIG. 1. Minimal model for the dynamics of DNA polymerase (DNAP). Left panel: Schematic depiction of the replication process with the leading DNA strand passing through the DNA polymerase (encircled area). Right panel: Kinetic model for DNA replication. Three sites on the DNA are labeled as $i-1$, i , and $i+1$ and for the i th site, the states of the DNAP are labeled as A , C , IC , C_{EX} , and IC_{EX} , respectively. Binding of a correct nucleotide takes the DNAP to the state C , whereas binding of an incorrect nucleotide takes the DNAP to the state IC . Actively elongating DNAP takes a forward step with rate ϵ from the state C and with rate ϵ' from the state IC . Exonuclease activity is described by the states C_{EX} and IC_{EX} , depending on whether it is initiated from a correct or from an incorrect nucleotide. The arrows indicate the transition from one state to the other state with their corresponding transition rates as mentioned. The bond marked in red in state IC corresponds to the nucleotide mismatch and thus to a weaker bond in state- IC_{EX} .

[17,30]. However, a recent study [31] claims that the speed-accuracy trade-off is not universal, but rather depends on the kinetic parameters of the enzyme. In this work we study a simple kinetic model of DNA replication, for which we can exactly calculate the velocity and accuracy of replication to investigate the interplay of the two functions of DNAP. Based on that analysis we compare different possible schemes of discrimination between correct and incorrect nucleotides and address how an accurate replication mechanism can be achieved, while also limiting the negative effect on the speed of elongation.

II. MODEL

A. Model with parallel pathways for correct and incorrect dNTP incorporation

To study the interplay of polymerization and exonuclease activity, we use a minimal stochastic model, which is depicted in Fig. 1 as a schematic diagram. The DNA template is considered as a one-dimensional lattice of certain length and each site of that lattice represents an individual base of the DNA template. The DNA polymerase (DNAP) moves on the DNA template by single base steps, simultaneously extending the copy by one deoxy-nucleotide triphosphate (dNTP) in each step. As a minimal description, we describe this process by Michaelis-Menten kinetics, reversible binding of the dNTP followed by irreversible incorporation into the new DNA strand. We note that this kinetics have been characterized in more detail and exhibit additional substeps, but here we aim at a minimal model. In the stochastic model, these two substeps are represented by transitions from the active state with a free binding site (A) to the occupied state (denoted by C) and back, where there are two pathways for the enzyme to transition back to the free state, either by unbinding of the dNTP or by incorporation and stepping, in which case the active and free state is reached at the subsequent position and with a

one based longer new DNA strand. The rates of the three transitions are denoted as k_c , k_{-c} , and ϵ , respectively.

To account for misincorporations and thus for errors in replication, the model includes a second pathway with Michaelis-Menten kinetics that describes exactly the same steps as the previous one, however for binding and incorporation of an incorrect, i.e., noncognate dNTP. The corresponding occupied state is denoted by IC and the rates, which may generally differ from those in the correct pathway, with k_{ic} , k_{-ic} , and ϵ' , respectively.

B. Proofreading pathway

This basic model is extended to include the exonuclease activity of DNA polymerase. For that, the polymerase transitions to a state with the newly added dNTP in the exonuclease site. The exonuclease state is denoted by C_{EX} or state- IC_{EX} , depending on whether the new nucleotide is a correct or an incorrect one. The transition to that state occurs from state- C (or state- IC) with rate k_{ex} (or k_{iex}), respectively. The reverse transition is also included in the model (with rates k_{-ex} and k_{-iex}). Finally, the exonuclease reaction occurs from state C_{EX} or state- IC_{EX} with the cleavage rate k_{cl} or k'_{cl} , respectively. In this transition, the last incorporated nucleotide is cleaved off and the enzyme goes back to the free state A .

We note that the reaction scheme described here differs from the classical kinetic proofreading scheme [3], as the proofreading step requires a transition away from the main pathway, namely the transfer of the new DNA strand from the polymerization site to the exonuclease site. A reaction scheme that is very similar to ours (and also describes the off-pathway proofreading step) has been used by Sharma and Chowdhury [13]. Their focus however was on the stochasticity of DNA elongation by calculating dwell time distributions rather than on the different mechanisms of discrimination between the correct and incorrect nucleotides and on how they affect the

TABLE I. Estimated values of the model parameters

Parameter	Symbol	Value	Refs
Forward stepping rate	ϵ	$300s^{-1}$	[8,12,33,34]
Binding rate (correct dNTP)	k_c	$10^{-1}\epsilon$	[34]
Binding rate (incorrect dNTP)	k_{ic}	$10^{-2} - 10^{-1}\epsilon$	[15]
Unbinding rate (correct dNTP)	k_{-c}	$10^{-2}\epsilon$	[15]
Unbinding rate (incorrect dNTP)	k_{-ic}	$10^{-2} - 10^{-1}\epsilon$	[3,15]
Erroneous stepping rate	ϵ'	$10^{-3} - 10^{-2}\epsilon$	[8,33,34]
Exonuclease site transfer rate (correct nucleotide)	k_{ex}	$10^{-3}\epsilon$	[34]
Reverse transfer rate (correct nucleotide)	k_{-ex}	$<700s^{-1}$	[33], further comments
Exonuclease site transfer rate (incorrect nucleotide)	k_{iex}	$10 * k_{ex}$	[34]
Reverse transfer rate (incorrect nucleotide)	k_{-iex}	$10^{-4}\epsilon$	[34]
Cleavage rate (correct nucleotide)	k_{cl}	$10^{-2}\epsilon$	[34]
Cleavage rate (incorrect nucleotide)	k'_{cl}	$10^{-1}\epsilon$	[3]

design of accurate replication and the speed-accuracy trade-off. There are also some smaller differences between these two models. In their model, they include a transition to the exonuclease site before incorporation, which does not affect the accuracy (and is not included here), and do not explicitly represent the correct and incorrect incorporations as parallel pathways. We also note that a related, but considerably more complex proofreading scheme is found in RNA polymerase: RNA polymerase transitions to an off-pathway state for proofreading, the backtracking state, where it however performs diffusive motion and can cleave longer pieces of the newly synthesized RNA [32,35].

For discrimination between correct and incorrect nucleotides, we generally expect the following relations to hold: In the initial binding step, binding of incorrect nucleotides should be suppressed compared to binding of correct nucleotides. This may be due to a lower binding rate and/or a higher unbinding rate, reflecting a higher affinity for correct than for incorrect binding. The discrimination achieved in this step is restricted thermodynamically, by the difference in binding energy between the correct and incorrect dNTPs [3]. Additional discrimination can be expected based on the irreversible step (translocation of the enzyme), typically with $\epsilon' < \epsilon$. In contrast to binding, discrimination by the translocation rate is kinetic rather than thermodynamic and thus is not restricted by the binding energies but rather by the difference in the rates or in the corresponding energy barriers. Finally, the proofreading pathway is expected to also contribute kinetically to the accuracy, as transfer to the exonuclease site competes kinetically with the translocation of the enzyme. If the transfer for incorrect nucleotides is faster than that for correct nucleotides (i.e., for $k_{iex} > k_{ex}$), while at the same time if the translocation is slower ($\epsilon' < \epsilon$), one can expect the typical fate of an incorrect nucleotide to be excision and that of a correct one translocation. This tendency will be further enhanced if for correct nucleotides in the exonucleotide site, return to the polymerization site is more likely than cleavage, $k_{-ex} > k_{cl}$, while it is just the opposite for the incorrect nucleotides in the exonuclease site, $k_{-iex} < k'_{cl}$. Estimates of the rates from the literature are listed in Table I. Note that in some cases, these are estimated from more detailed kinetic schemes. All rates in this table are scaled with respect to the forward stepping rate ϵ .

C. Steady state solution

In the following, we will analyze the elongation speed and the accuracy of replication for two cases. In the first case, we consider a purely polymerizing scenario, where the DNA polymerase does not undergo the transition to the exonuclease state (i.e., $k_{ex} = k_{iex} = 0$). In the second case, we include the exonuclease activity. In both cases, we calculate the accuracy and velocity of elongation from the steady-state fluxes between the states of the discrete stochastic model described so far. The approach is similar to that used in our earlier work for backtracking of RNA polymerase [35] and provides the mean values of these quantities. Consideration of the fluctuations as in Ref. [13] requires time dependent methods. Effectively, our system is a stochastic five-state model. In the construction of our model, we have assumed that there is no sequence dependence of the rates [4,36,37], so all the rates considered here should be considered averages over a possible sequence-dependence. To determine steady-state fluxes, we solve the five state model shown in Fig. 1 and first determine the steady state probabilities of the five different states of the DNAP. These probabilities are denoted by P_i , where $i = A$ (active state), C (correct state), IC (incorrect state), C_{EX} (correct exonuclease state), and IC_{EX} (incorrect exonuclease state), respectively. We calculate the steady state probabilities by equating the total incoming flux and the total outgoing flux for each state, which can be expressed as

$$k_c P_A + k_{-ex} P_{C_{EX}} - \epsilon P_C - k_{ex} P_C - k_{-c} P_C = 0, \quad (1)$$

$$k_{ic} P_A + k_{-iex} P_{IC_{EX}} - \epsilon' P_{IC} - k_{iex} P_{IC} - k_{-ic} P_{IC} = 0, \quad (2)$$

$$k_{ex} P_C - k_{cl} P_{C_{EX}} - k_{-ex} P_{C_{EX}} = 0, \quad (3)$$

and

$$k_{iex} P_{IC} - k'_{cl} P_{IC_{EX}} - k_{-iex} P_{IC_{EX}} = 0. \quad (4)$$

Together with the normalization condition for the probabilities,

$$P_A + P_C + P_{IC} + P_{C_{EX}} + P_{IC_{EX}} = 1, \quad (5)$$

the steady state probability of all states can be found. The explicit expressions are included in the Appendix.

III. RESULTS AND DISCUSSIONS

Using the steady state probabilities of individual states, the velocity of elongation (V) can be calculated as

$$V = \epsilon' P_{IC} + \epsilon P_C. \quad (6)$$

Here, $\epsilon' P_{IC}$ and ϵP_C represent the contributions to the elongation velocity by the forward stepping to the next site after correct incorporation with stepping rate ϵ and after incorrect incorporation with rate ϵ' , respectively. Plugging P_{IC} and P_C into Eq. (6), an explicit expression for V can be obtained, which is given in the Appendix.

Further, the accuracy (A) of DNA replication can be calculated as the ratio of the correct flux and the total flux as

$$A = \frac{\epsilon P_C}{\epsilon P_C + \epsilon' P_{IC}} = \frac{1}{1 + (\frac{\epsilon'}{\epsilon}) \frac{P_{IC}}{P_C}}. \quad (7)$$

By substituting the expression for the steady state probabilities, (P_C and P_{IC}) in the above expression, we can obtain the accuracy A of replication, which is given in the Appendix.

A. Purely polymerizing scenario

To analyze the discrimination between correct and incorrect nucleotide and the effect of errors in the velocity, we first consider the purely polymerizing scenario, as observable in mutants without exonuclease activity. To that end, we set $k_{ex} = k_{iex} = 0$, i.e., the DNA polymerase does not make a transition to the exonuclease state, neither from the correct binding state nor from the incorrect binding state. For this case, the elongation velocity V_{pp} can be calculated as

$$V_{pp} = \frac{\epsilon k_c (\epsilon' + k_{-ic}) + \epsilon' k_{ic} (\epsilon + k_{-c})}{(\epsilon + k_{-c})(\epsilon' + k_{ic} + k_{-ic}) + k_c (\epsilon' + k_{-ic})}. \quad (8)$$

Similarly, the accuracy A_{pp} of replication can be calculated as

$$A_{pp} = \frac{\epsilon k_c (\epsilon' + k_{-ic})}{\epsilon \epsilon' k_c + \epsilon \epsilon' k_{ic} + \epsilon k_c k_{-ic} + \epsilon' k_{ic} k_{-c}}. \quad (9)$$

Figures 2(a) and 2(b) shows the plot of velocity V_{pp} and the accuracy A_{pp} for a constant rate of correct nucleotide binding, varying the rate of incorrect nucleotide binding (k_{ic}). It is observed that the possibility of errors obviously decreases the accuracy whereas the influence of such errors may increase or decrease the velocity of elongation. V_{pp} may increase or decrease with k_{ic} , depending on the value of stepping rate ϵ' after an erroneous incorporation. Similarly Figs. 2(c) and 2(d) shows V_{pp} and A_{pp} as a function of ϵ' for different values of k_{-ic} . V_{pp} has a finite value at $\epsilon' = 0$ and for any fixed value of k_{-ic} . This nonzero value of V_{pp} is due to the forward stepping of the DNAP from the correct state to the next site with rate ϵ . For $\epsilon' = 0$, the DNAP usually goes back to the active state for a nonzero value of k_{-ic} and eventually incorporation of a correct nucleotide (and stepping) takes place. Thus for small ϵ' , the typical fate of an incorrect nucleotide is to unbind, making the binding of an incorrect nucleotide a dead-end branch off the main pathway and rapid unbinding beneficial for high speed. By contrast, for large ϵ' , incorrect nucleotides are incorporated and contribute to the velocity of polymerization. Moreover, interestingly we observe that for a critical value of ϵ' , (i.e., at $\epsilon' = \epsilon'_c$), at which the transition between

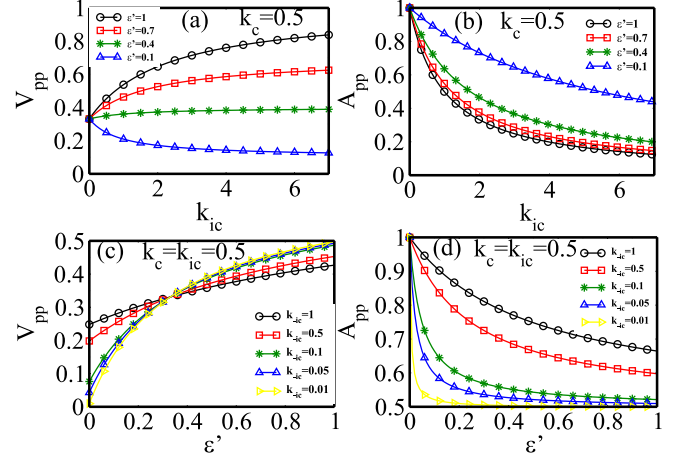


FIG. 2. Velocity V_{pp} and accuracy A_{pp} in the purely polymerizing scenario, both (a), (b) as a function of the binding rate of incorrect nucleotides, k_{ic} and (c), (d) as function of the stepping rate ϵ' . Other parameters are varied as indicated in the legends, the parameters $\epsilon = 1$ and $k_{-c} = 0.01$ are fixed.

these two regimes occurs, the velocity is independent of k_{-ic} . The accuracy [Fig. 2(d)], on the other hand, always increases when the unbinding rate is increased, independent of the value of ϵ' . Just like decreasing the binding rate of incorrect nucleotides, increasing the unbinding rate k_{-ic} always increases the accuracy, but may increase or decrease the velocity.

The critical stepping rate ϵ'_c at which V_{pp} is independent of k_{-ic} is exactly calculated as $\epsilon'_c = \frac{\epsilon k_c}{\epsilon + k_c + k_{-c}}$ when $k_c = k_{ic}$. From this expression one can notice that ϵ'_c increases as a function of k_c and for large k_c ($k_c \rightarrow \infty$), it approaches ϵ , as plotted in Fig. 3. This implies that ϵ'_c can be controlled by fine tuning the k_c values.

The scenario considered so far, is the simplest exonuclease-deficient scenario. Another exonuclease-deficient and therefore purely polymerizing scenario is when transfer to the exonuclease site is possible, but with no cleavage (i.e., with $k_{cl} = k'_{cl} = 0$). Velocity and accuracy in this case are plotted as a function of k_{ic} together with the corresponding results

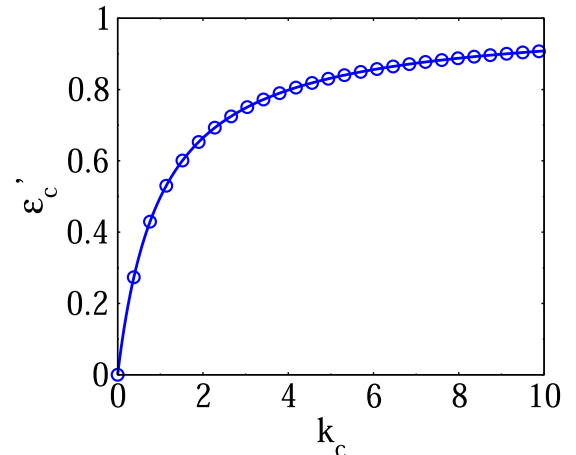


FIG. 3. Critical stepping rate ϵ'_c as a function of the binding rate k_c . The fixed parameters are $\epsilon = 1$ and $k_{-c} = 0.01$.

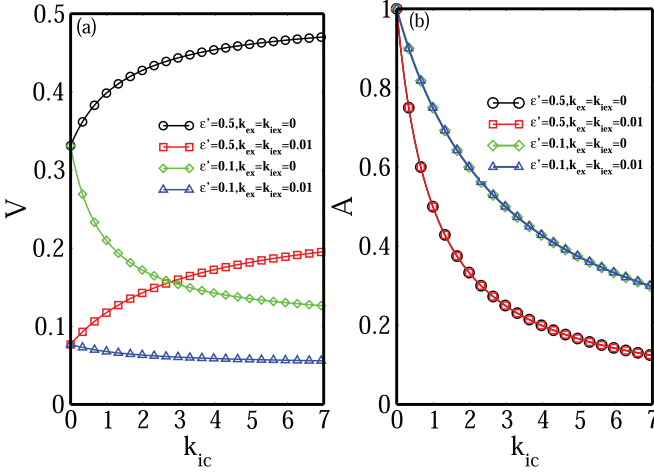


FIG. 4. Scenario with transfer to the exonuclease site, but without exonuclease activity: Velocity V and accuracy A as a function of the binding rate k_{ic} of incorrect nucleotides for different values of the stepping rate ϵ' and the exonuclease site transfer rates k_{ex} and k_{iex} . The other parameters are fixed: $k_c = k_{ic} = 0.5$, $\epsilon = 1$, $k_{-c} = k_{-iex} = 0.01$, and $k_{-ex} = 0.001$.

for the case without transfer to the exonuclease site (i.e., for $k_{ex} = k_{iex} = 0$) in Fig. 4. The velocity is generally reduced by the transfer to the (inactive) exonuclease site, whereas the accuracy is seen to be the same in both scenarios.

B. Polymerization with exonuclease activity

Including the exonuclease activity, we have exactly calculated the elongation velocity, V as well as the accuracy of replication, A using Eqs. (16) and (19), respectively. We expect that the typical effect of a proofreading mechanism is a reduction of the velocity and an increase in accuracy. However, the following consideration shows that this is not generally the case: Without discrimination between the correct and incorrect nucleotides in the exonuclease transfer step ($k_{ex} = k_{iex}$), we observed that both velocity as well as accuracy decrease with increasing k_{ex} [Figs. 5(a) and 5(b)]. The decrease in accuracy can be understood as follows: Once a nucleotide, correct or incorrect, is bound, it can either be incorporated irreversibly by the forward step or be removed again by one of two pathways, unbinding or exonuclease transfer and cleavage. If the cleavage rate is sufficiently high, then the cleavage follows almost deterministically upon exonuclease transfer, and the effective rate for the removal of the nucleotide is $k_{-c} + k_{ex}$ (or $k_{-ic} + k_{iex}$). If the first term discriminates between the correct and the incorrect nucleotides, but the second does not, increasing the second will inevitably reduce the accuracy rather than increasing it, as expected for a proofreading pathway. Thus, additional discrimination in the exonuclease pathway is required for proofreading.

We also note that the decrease of the velocity with increasing k_{ex} is also not universal. When there is no discrimination between the correct and incorrect nucleotides in the exonuclease transfer step, for very small value of ϵ' , the velocity initially increases with k_{ex} for very small values of k_{-c} , shows

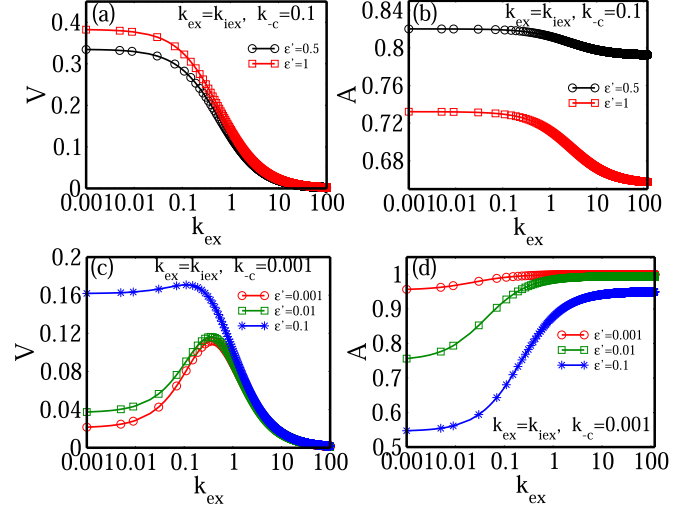


FIG. 5. Variety of behaviors in a scenario with exonuclease activity: Velocity V and accuracy A as a function of the exonuclease transfer rate k_{ex} for two different values of k_{-c} [in (a), (b) and in (c), (d), respectively]. In all cases, there is no discrimination between correct and incorrect nucleotides in the exonuclease pathway. The parameters are $k_c = k_{ic} = 0.5$, $\epsilon = 1$, $k_{cl} = k_{-ex} = k_{-iex} = 0.1$, and $k'_{cl} = k_{cl} * \exp(3)$. $k_{-ic} = k_{-c} * \exp(3)$. Note that both increase and decrease of velocity and accuracy are seen and that the two are not generally anticorrelated.

a maximum and then decreases with k_{ex} [Fig. 5(c)]. Whereas the accuracy increases with k_{ex} [Fig. 5(d)]. The increase of velocity can be understood as follows: The presence of incorrect nucleotide reduces the velocity due to the corresponding slow stepping. Exonuclease can circumvent that slow stepping. However for high exonuclease transfer rates, the velocity is decreased as exonuclease transfer now also successfully compete with the rapid stepping after correct incorporation. The presence of a maximum in the velocity is understood as preventing slow stepping after incorrect nucleotide binding, but not rapid stepping after correct binding, suggests one principle how the exonuclease transfer rate should be set, namely between the two stepping rates (see below).

These results shed some light on the question of a trade-off between velocity and accuracy, which is generally expected for proofreading mechanisms and has been reported in a number of studies [6,27,28]. Our results show that such trade-off is also seen in our minimal model, provided the kinetic parameters are in the range expected for DNA polymerases. The observation of a trade-off is independent of the mechanisms of discrimination in the proofreading pathway as discussed below. However, Fig. 5 shows that speed-accuracy trade-off is not universal, consistent with a recent study [31], but rather dependent on the kinetic model parameters. This is a typical feature of kinetic proofreading mechanism and can be due to the discrimination of free energy associated with different reactions.

Since additional discrimination between correct and incorrect nucleotides is needed for the exonuclease activity to result in an increase in accuracy, we next compared the three possibilities for discrimination in exonuclease pathway: Using realistic parameters for the polymerization pathway, includ-

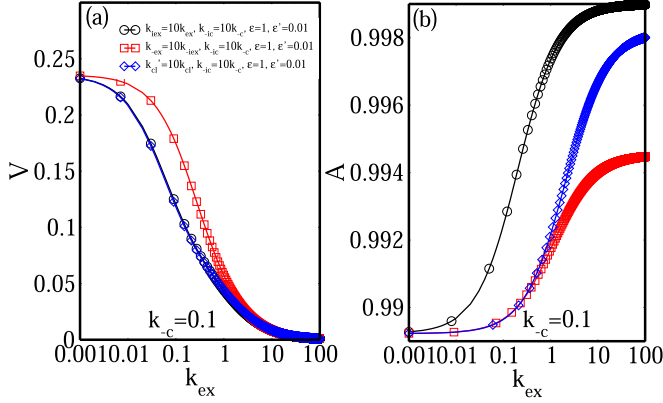


FIG. 6. Three scenarios with discrimination between correct and incorrect nucleotides in the exonuclease pathway: (a) Velocity V and (b) accuracy A as a function of k_{ex} . The three cases depict discrimination in the exonuclease transfer rate, the reverse transfer rate and the cleavage rate, respectively. The fixed parameters are $k_c = k_{ic} = 0.5$.

ing discrimination in unbinding ($k_{-ic} > k_{-c}$) and in stepping ($\epsilon' < \epsilon$), we determined the velocity and the accuracy for (i) $k_{iex} > k_{ex}$, (ii) $k_{-iex} < k_{-ex}$, and for (iii) $k'_{cl} > k_{cl}$. In all these three different cases, all other rates of the exonuclease pathway are the same for correct and incorrect nucleotides, and we take the rates that are different to differ by a factor of 10. The latter factor is of the order of $e^{\frac{\Delta G}{k_B T}}$ with a mismatch energy $\Delta G \sim 3k_B T$.

Figure 6 shows the comparison between the three cases. In all cases, the accuracy increases with increasing exonuclease transfer rate, while the velocity decreases. However there are some marked differences between the scenarios: Discrimination in the transfer rate or in the cleavage rate show an almost identical decrease in the velocity and a very similar increase in accuracy. However, the increase in accuracy is shifted towards higher values of k_{ex} for the case of discrimination in the transfer rate as compared to the discrimination in the cleavage rate. Thus, for the same transfer rate (and for the same decrease in velocity), the discrimination in transfer results in a higher accuracy than discrimination in cleavage. Discrimination in the reverse transfer rate, by comparison, shows an overall smaller increase in accuracy as well as smaller decrease in velocity.

IV. CONCLUSIONS

In this paper, we have studied a simple kinetic model of DNA replication and have exactly calculated the accuracy and elongation velocity in the absence as well as in the presence of exonuclease activity. In general, the fidelity of DNA replication is based on the nucleotide selectivity of the DNA polymerase during incorporation, its exonucleolytic proofreading activity and the postreplicative DNA mismatch repair. The latter contribution is not included in our model, in which an error remains uncorrected once the new DNA strand is elongated and DNA polymerase has made a forward step. During the exonucleolytic proofreading, an erroneously incorporated nucleotide is cleaved off, hence the exonuclease activity improves the overall accuracy of replication, but at

the same time it also typically reduce the speed of elongation by "resetting" the replication process to the beginning of the last step. Similarly the escape of errors (via erroneous stepping) speeds up the elongation mechanism but compromises the overall accuracy. The trade-off between speed and accuracy however is not universal, but depends on the kinetic parameters, in agreement with the earlier results [31]. Our minimal model provides a simple picture of the design of an accurate replication system. Through the modulation of the model parameters, we can identify several kinetic principles. Discrimination between the correct and incorrect nucleotides arises both from the polymerization pathway as well as from the proofreading pathway and is based on the kinetic competition in both the cases. For both, slow forward stepping after incorporation of an incorrect nucleotide (slower than for a correct nucleotide) is crucial for the accuracy, giving time for unbinding and transfer to the exonuclease site, respectively. However, additional discrimination in the exonuclease pathway is needed and provided most efficiently by the transfer reaction. Thus the accuracy depends on the coordinated action of the polymerase and exonuclease activity of DNA polymerase.

ACKNOWLEDGMENTS

M.S. acknowledges the INSPIRE Faculty award (IFA 13 PH-66) by the Department of Science and Technology, India and the Faculty Recharge Program (FRP-56055) of UGC Faculty Recharge Program research grant, UGC, India.

APPENDIX

In this Appendix, we give the explicit expressions for the steady-state probabilities obtained as the solution to Eqs. (1)–(5):

$$P_C = k_c(k_{-ex} + k_{cl})\{k_{iex}k'_{cl} + (k_{-iex} + k'_{cl})(\epsilon' + k_{-ic})\}/T, \quad (A1)$$

$$P_{IC} = k_{ic}(k_{-iex} + k'_{cl})\{k_{ex}k_{cl} + (k_{-ex} + k_{cl})(\epsilon + k_{-c})\}/T, \quad (A2)$$

$$P_{C_{EX}} = k_{ex}k_c\{k_{iex}k'_{cl} + (k_{-iex} + k'_{cl})(\epsilon' + k_{-ic})\}/T, \quad (A3)$$

$$P_{I_{C_{EX}}} = k_{iex}k_{ic}\{k_{ex}k_{cl} + (k_{-ex} + k_{cl})(\epsilon + k_{-c})\}/T, \quad (A4)$$

and

$$P_A = 1 - P_{C_{EX}} - P_{I_{C_{EX}}} - P_{IC} - P_C. \quad (A5)$$

In the above expressions, T is given by

$$\begin{aligned} T = & k_{iex}\{k_{ex}k'_{cl}k_c + k_{ex}k_{cl}(k_{ic} + k'_{cl}) \\ & + (k_{-ex} + k_{cl})((k_{ic} + k'_{cl})(\epsilon + k_{-c}) + k'_{cl}k_c)\} \\ & + (k_{-iex} + k'_{cl})\{k_{ex}k_{ic}k_{cl} + k_{ex}(k_c + k_{cl})(\epsilon' + k_{-ic}) \\ & + (k_{-ex} + k_{cl})(k_c(\epsilon' + k_{-ic}) \\ & + (k_{ic} + (\epsilon' + k_{-ic}))(\epsilon + k_{-c}))\}. \end{aligned} \quad (A6)$$

We also give explicit expressions for the velocity and accuracy as obtained from Eqs. (6) and (7): The velocity V

can be expressed as

$$V = \frac{S}{U} \quad (\text{A7})$$

with S and U are given by

$$S = \epsilon' k_{ic} (k'_{cl} + k_{-iex}) \{k_{cl} k_{ex} + (\epsilon + k_{-c})(k_{-ex} + k_{cl})\} \\ + \epsilon k_c (k_{cl} + k_{-ex}) \{k'_{cl} k_{iex} + (\epsilon' + k_{-ic})(k'_{cl} + k_{-iex})\} \quad (\text{A8})$$

and

$$U = (k'_{cl} + k_{-iex}) \{ (k_{cl} + k_{-ex}) ((\epsilon + k_{-c})(\epsilon' + k_{ic} + k_{-ic}) \\ + k_c(\epsilon' + k_{-ic})) + k_{ex}(\epsilon' + k_{-c})(k_c + k_{cl}) + k_{ic} k_{cl} k_{ex} \}$$

$$+ k_{iex} \{ (k_{cl} + k_{-ex})(k_c k'_{cl} + (\epsilon + k_{-c})(k_{ic} + k'_{cl})) \\ + k_{cl} k_{ex} (k_{ic} + k'_{cl}) + k_c k'_{cl} k_{ex} \}. \quad (\text{A9})$$

The accuracy A is found to be

$$A = \frac{1}{1 + \frac{\epsilon'}{\epsilon} N}, \quad (\text{A10})$$

where

$$N = \frac{k_{ic}(k_{-iex} + k'_{cl}) [k_{ex} k_{cl} + (k_{-ex} + k_{cl})(\epsilon + k_{-c})]}{k_c(k_{-ex} + k_{cl}) [k_{iex} k'_{cl} + (k_{-iex} + k'_{cl})(\epsilon' + k_{-ic})]}. \quad (\text{A11})$$

-
- [1] S. J. Benkovic, A. M. Valentine, and F. Salinas, Replisome-mediated DNA replication, *Ann. Rev. Biochemistry* **70**, 181 (2001).
- [2] A. Kornberg and T. A. Baker, *DNA Replication*, Second Edition W. H. Freeman (New York, 1992), Chap. 4, pp. 113–159.
- [3] J. J. Hopfield, Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity, *Proc. Nat. Acad. Sci. USA* **71**, 4135 (1974).
- [4] T. D. Yager and P. H. Von Hippel, A thermodynamic analysis of RNA transcript elongation and termination in escherichia coli, *Biochemistry* **30**, 1097 (1991).
- [5] I. Tubulekas and D. Hughes, Suppression of rps1 phenotypes by tuf mutations reveals a unique relationship between translation elongation and growth rate, *Molecular Microbiology* **7**, 275 (1993).
- [6] I. Wohlgenuth, C. Pohl, J. Mittelstaet, A. L. Konevega, and M. V. Rodnina, Evolutionary optimization of speed and accuracy of decoding on the ribosome, *Phil. Trans. R. Soc. B: Biological Sci.* **366**, 2979 (2011).
- [7] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, Third edition (Elsevier, Amsterdam, 2007), Vol. 1.
- [8] S. S. Patel, I. Wong, and K. A. Johnson, Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant, *Biochemistry* **30**, 511 (1991).
- [9] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer, Sunderland MA, 2001).
- [10] A. B. Kolomeisky and M. E. Fisher, Molecular motors: A theorist's perspective, *Annu. Rev. Phys. Chem.* **58**, 675 (2007).
- [11] M. Schliwa and G. Wohlke, Molecular motors, *Nature (London)* **422**, 759 (2003).
- [12] P. Xie, A possible mechanism for the dynamics of transition between polymerase and exonuclease sites in a high-fidelity DNA polymerase, *J. Theor. Biology* **259**, 434 (2009).
- [13] A. K. Sharma and D. Chowdhury, Error correction during DNA replication, *Phys. Rev. E* **86**, 011913 (2012).
- [14] L. J. Reha-Krantz, DNA polymerase proofreading: Multiple roles maintain genome stability, *Biochim. Biophys. Acta - Protein Proteomics* **1804**, 1049 (2010).
- [15] K. A. Johnson, The kinetic and chemical mechanism of high-fidelity DNA polymerases, *Biochim. Biophys. Acta - Protein Proteomics* **1804**, 1041 (2010).
- [16] T. A. Kunkel and K. Bebenek, DNA replication fidelity, *Ann. Rev. Biochem.* **69**, 497 (2000).
- [17] A. Bębenek and I. Ziuzia-Graczyk, Fidelity of DNA replication – a matter of proofreading, *Current Genetics* **64**, 985 (2018).
- [18] T. A. Kunkel, DNA replication fidelity, *J. Biological Chem.* **279**, 16895 (2004).
- [19] B. Ibarra, Y. R. Chemla, S. Plyasunov, S. B. Smith, J. M. Lazaro, Margarita Salas, and C. Bustamante, Proofreading dynamics of a processive DNA polymerase, *EMBO J.* **28**, 2794 (2009).
- [20] P. S. Freemont, J. M. Friedman, L. S. Beese, M. R. Sanderson, and T. A. Steitz, Cocystal structure of an editing complex of klenow fragment with DNA, *Proc. Natl. Acad. Sci. USA* **85**, 8924 (1988).
- [21] J. Wang, A. K. M. A. Sattar, C. C. Wang, J. D. Karam, W. H. Konigsberg, and T. A. Steitz, Crystal structure of a pol α family replication DNA polymerase from bacteriophage rb69, *Cell* **89**, 1087 (1997).
- [22] S. Kamtekar, A. J. Berman, J. Wang, J. M. Lázaro, M. de Vega, L. Blanco, M. Salas, and T. A. Steitz, Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage ϕ 29, *Molecular cell* **16**, 609 (2004).
- [23] Y. Shamoo and T. A. Steitz, Building a replisome from interacting pieces: Sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex, *Cell* **99**, 155 (1999).
- [24] M. Hogg, S. S. Wallace, and Sylvie Doublé, Crystallographic snapshots of a replicative DNA polymerase encountering an abasic site, *EMBO J.* **23**, 1483 (2004).
- [25] U. Subuddhi, M. Hogg, and L. J. Reha-Krantz, Use of 2-aminopurine fluorescence to study the role of the β hairpin in the proofreading pathway catalyzed by the phage t4 and rb69 DNA polymerases, *Biochemistry* **47**, 6130 (2008).
- [26] A. R. Fersht, J. W. Knill-Jones, and W.-C. Tsui, Kinetic basis of spontaneous mutation: Misinsertion frequencies, proofreading specificities and cost of proofreading by DNA polymerases of Escherichia coli, *J. Molecular Biology* **156**, 37 (1982).
- [27] M. Johansson, J. Zhang, and M. Ehrenberg, Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection, *Proc. Natl. Acad. Sci. USA* **109**, 131 (2012).

- [28] M. Johansson, M. Lovmar, and M. Ehrenberg, Rate and accuracy of bacterial protein synthesis revisited, *Current opinion in microbiology* **11**, 141 (2008).
- [29] M. Lovmar and M. Ehrenberg, Rate, accuracy and cost of ribosomes in bacterial cells, *Biochimie* **88**, 951 (2006).
- [30] A. Murugan, D. A. Huse, and S. Leibler, Discriminatory proof-reading regimes in nonequilibrium systems, *Phys. Rev. X* **4**, 021016 (2014).
- [31] K. Banerjee, A. B. Kolomeisky, and O. A. Igoshin, Elucidating interplay of speed and accuracy in biological error correction, *Proc. Natl. Acad. Sci. USA* **114**, 5183 (2017).
- [32] D. A. Erie, O. Hajiseyedjavadi, M. C. Young, and P. H. von Hippel, Multiple RNA polymerase conformations and GreA: Control of the fidelity of transcription, *Science* **262**, 867 (1993).
- [33] Yu.-C. Tsai and K. A. Johnson, A new paradigm for DNA polymerase specificity, *Biochemistry* **45**, 9675 (2006).
- [34] K. A. Johnson, Conformational coupling in DNA polymerase fidelity, *Ann. Rev. Biochem.* **62**, 685 (1993).
- [35] M. Sahoo and S. Klumpp, Backtracking dynamics of RNA polymerase: Pausing and error correction, *J. Phys. Condens. Matter* **25**, 374104 (2013).
- [36] L. Bai, A. Shundrovsky, and M. D. Wang, Sequence-dependent kinetic model for transcription elongation by RNA polymerase, *J. Molecular Biology* **344**, 335 (2004).
- [37] V. R. Tadigotla, D. Ó Maoiléidigh, A. M. Sengupta, V. Epshtein, R. H. Ebright, E. Nudler, and A. E. Ruckenstein, Thermodynamic and kinetic modeling of transcriptional pausing, *Proc. Natl. Acad. Sci. USA* **103**, 4439 (2006).