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

pubs.acs.org/JPCL

Letter

Linear-Decoupling Enables Accurate Speed and Accuracy Predictions for Copolymerization Processes

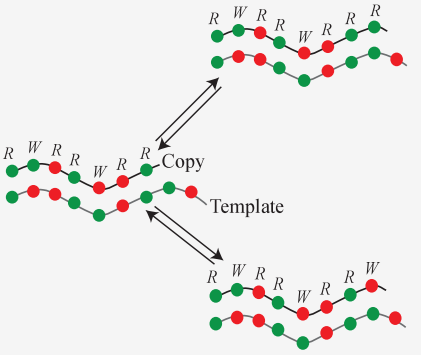
Tripti Midha, Anatoly B. Kolomeisky,*,# and Oleg A. Igoshin*,#

Cite This: J. Phys. Chem. Lett. 2024, 15, 9361−9368

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Biological processes exhibit remarkable accuracy and speed and can be theoretically explored through various approaches. The Markov-chain copolymerization theory, describing polymer growth kinetics as a Markov chain, provides an exact set of equations to solve for error and speed. Still, due to nonlinearity, these equations are hard to solve. Alternatively, the enzyme-kinetics approach, which formulates a set of linear equations, simplifies the biological processes as transitions between discrete chemical states, but generally, it might not be accurate. Here, we show that the enzyme-kinetic approach can lead to inaccurate fluxes, even for first-order polymerization processes. To address the problem, we propose a simplified linear-decoupling approximation for steady-state probabilities of higher-order copolymer chains under biologically relevant conditions. Our findings demonstrate that the stationary speed and error rate obtained from the linear-decoupling method align closely with exact values from the Markov-chain (nonlinear) approximation. Extending the technique to higher-order processes with proofreading and internal states shows that it works equally well to describe trade-offs between speed and accuracy for DNA replication and transcription elongation. Our work underscores the proposed linear-decoupling approximation's efficacy in addressing the nonlinear behavior of the Markov-chain approach and the enzyme-kinetic approach's limitations, ensuring accurate predictions for high-fidelity biological processes.



Biological information processing exhibits exceptional accuracy achieved through the precise replication of substrates that match the template strand. This fidelity, however, may need to be balanced against processing speed.[1,2] Such properties are fundamental to all fundamental biological processes involved in the transfer of genetic information.[3−7] Inspired by single-molecule experiments to investigate replication and transcription processes in living systems, numerous theoretical approaches have been presented to examine their physicochemical properties. One such framework for understanding the fidelity and speed of information processing involves analyzing the kinetics of templated copolymerization. This refers to the chemical reaction of combining different monomeric units to form copolymer sequences that correspond to the template, thereby encoding genetic information.[8−11]

Recent studies have analyzed the kinetics of templated copolymerization with the so-called k^{th}-order neighbor's effects. In the k^{th}-order neighbor effect, a monomer's attachment and detachment rates at the growing end of a copolymer chain depend on the identity of the previous k subunits. Various approaches, such as Markov-chain copolymerization theory, have addressed these higher-order effects under steady-state conditions. In these methods, the templated copolymerization process with k^{th}-order neighbor's effects is described as a k^{th}-order Markov chain. This framework treats the sequence probability distribution of copolymer chains of length up to size k exactly, while a nonlinear combination of stationary probabilities of shorter chains approximates longer chains. These methods yield exact expressions for error rates and processing speeds.[12−16] However, for realistic higher-order biological processes, these calculations become increasingly complex, often necessitating advanced numerical solutions, which significantly complicates the molecular understanding of the underlying processes.

Alternatively, an enzyme-kinetic approach has been employed to analyze the molecular mechanisms and fidelity of biological information processing.[17−19] This method assumes that after a correct or incorrect product is formed, the enzyme resets to its initial state, resulting in a periodic finite-state kinetic model.[20−22] Unlike the Markov-chain copolymerization theory, this method results in the set of chemical-kinetic equations that are linear in terms of the probabilities of individual chemical states.[19−29] These equations can often be solved analytically to determine stationary probabilities, fluxes, and error rates, defined as the ratio of the

Received: July 19, 2024
Revised: August 27, 2024
Accepted: August 29, 2024

https://doi.org/10.1021/acs.jpclett.4c02132
J. Phys. Chem. Lett. 2024, 15, 9361−9368

stationary flux forming an incorrect product to the total flux forming either product. A recent study comparing error rates from both approaches for first-order neighbor effects found an excellent agreement, provided the error is computed using a modified formula that accounts for nearest-neighbor correlations.[30]

However, the ability of the enzyme-kinetic approach to accurately predict the speed of copolymerization processes with higher-order neighbor effects remains uncertain. Preliminary findings suggest that stationary fluxes from the enzyme-kinetic approach may not always agree with those computed using copolymerization theory for first-order neighbor effects.[30] A comprehensive analysis comparing speed estimates from both approaches is still lacking. If the enzyme-kinetic approach's speed predictions do not align with the results from exact simulations or copolymerization theories, it is unclear whether a simpler, accurate method relying on linear equations can be developed. Our goal here is to present a method that can overcome the deficiencies of both theoretical approaches.

To address the limitations of the enzyme-kinetic approach and the overwhelming complexity of the copolymerization method, we propose a linear-decoupling approximation method for investigating the kinetics of copolymerization under biologically relevant conditions. For copolymerization processes with k^{th}-order neighbor effects, the method approximates the stationary probabilities of higher-order copolymer chains ($n > k$) by using a linear factorization of stationary probabilities for chains of length k and less. This approximation reduces the steady-state kinetic equations to a system of linear algebraic equations, which can be solved analytically or numerically exactly. We apply the linear-decoupling framework to analyze realistic biological processes, such as transcription elongation with second-order neighbor effects, dinucleotide cleavage mechanisms, and DNA replication with proofreading and intermediate states for exonuclease activity. We also investigate and compare speed−accuracy trade-off curves with variations in relevant biological parameters, using linear-decoupling and the nonlinear (Markov-chain) copolymerization methods.

First, let us start by considering the copolymerization process, where polymer molecules are produced by pairing monomers with complementary monomeric units at the template. If the pairing is correct, we label the monomer in the copolymer sequence as R; otherwise, the monomer is denoted as W (see Figure 1A). It is assumed that the attachment and detachment of monomers at the tip of the copolymer chain occur under the first-order neighbor effects; that is, the corresponding rates depend only on the chemical nature of the last added monomer to the chain.[12,14,31] We denote the rate of adding a monomer Y to any copolymer chain ending with monomer X, by k_{XY}, while the rate for the reverse reaction, i.e., the removal of Y, is given by k_{-XY}, where X and Y can be right R or wrong W subunits. Figure 1A shows the branching network of all of the possible states in the system.

The Markov-chain copolymerization theory describes the probability of the copolymer chain ending with sequence $i_1 i_2 \cdots i_n$ at time t as $Q_{i_1 i_2 \cdots i_n}(t)$. For the processes with the first-order neighbor effect, the chemical-kinetic equations that govern the temporal evolution of these probabilities are as follows:[14]

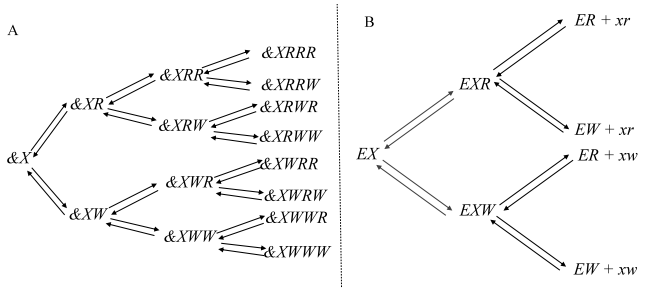


Figure 1. (A) Branching network of growing copolymer chains from the original sequence with monomer X at the end. (B) Enzyme-kinetic description of a binary copolymerization with rates depending on the previous monomer. Here, E represents the enzyme involved in the polymerization process, and rr, rw, wr, and ww are dimer products formed in each cycle. X represents either a right (R) or a wrong (W) monomer.

$$\frac{dQ_{ij}}{dt} = k_{ij}Q_i + \sum_{l=\{R,W\}} k_{-jl}Q_{ijl} - (k_{jR} + k_{jW} + k_{-ij})Q_{ij} \tag{1}$$

Each term of the right-hand side eq 1 has a simple physical meaning. The first term describes the attachment of monomer j to the copolymer chain ending with monomer i. The second term describes the monomeric unit l detachment from the copolymer sequence ending with trimer ijl. The last term describes the loss events occurring to the copolymer chain ending with dimer ij itself due to the addition of subunits R, W or removal of the subunit j. Note that the conservation of probability implies that $Q_R + Q_W = 1$. Also, the definition of $Q_{i_1 i_2 \cdots i_n}$ implies that $Q_i = Q_{Ri} + Q_{Wi}$ and $Q_{ij} = Q_{Rij} + Q_{Wij}$, for $i, j, \in \{R, W\}$.

For the copolymerization processes, the error, η, is defined as the ratio of stationary fluxes flowing out from the chains ending with the wrong monomer (J_W) to the sum of the stationary fluxes flowing out from the chains ending with monomer R (J_R) and monomer W (J_W),[14,32] yielding

$$\eta = \frac{J_{WR} + J_{WW}}{J_{RR} + J_{RW} + J_{WR} + J_{WW}} \tag{2}$$

Here, the stationary fluxes from copolymer chains ending with monomer X to the chains ending with dimers XY are defined as

$$\begin{aligned} J_{XY} &= k_{XY}Q_X - k_{-XY}Q_{XY} \\ &= k_{XY}(Q_{RX} + Q_{WX}) - k_{-XY}Q_{XY} \end{aligned} \tag{3}$$

The polymerization speed can be defined as the total stationary flux for adding the right monomer to the copolymer chains, which is equivalent to the total stationary flux flowing out of the chains ending with monomer R,[14,32] and is given by

$$V = J_{RR} + J_{WR} \tag{4}$$

For the special case when reactions in the binary copolymerization process are irreversible (see Figure 1A), the set of kinetic equations given by eq 1, in the stationary limit ($\frac{dQ_{ij}}{dt} = 0$), reduces to four uncoupled algebraic equations with only Q_{ij}. These equations can be solved exactly to obtain the probabilities satisfying the normalization condition $\sum Q_{ij} = 1$ (see the Supporting Information for details). The error and speed for the irreversible process can be obtained by

https://doi.org/10.1021/acs.jpclett.4c02132
J. Phys. Chem. Lett. 2024, 15, 9361−9368

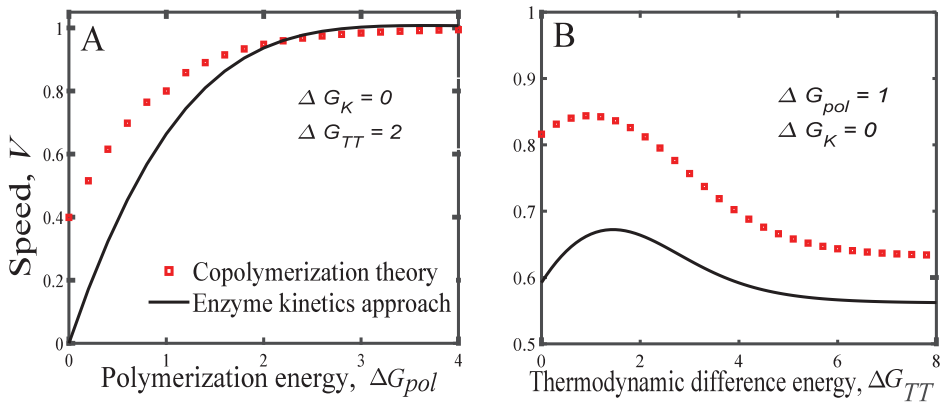


Figure 2. Polymerization speed V (A) as a function of polymerization energy, ΔG_{pol}, with fixed $\Delta G_K = 0$, and $\Delta G_{TT} = 2$; (B) as a function of thermodynamic energy ΔG_{TT}, with fixed $\Delta G_K = 0$, and $\Delta G_{pol} = 1$. Energies are in the unit of k_BT. Solid lines correspond to the estimates from the enzyme-kinetics approach, and symbols correspond to the exact estimates from the nonlinear copolymerization theory.

substituting the stationary probabilities and kinetic rates into eqs 2−4.

For the processes with reversible polymerization reactions, eq 1 in the steady state yields four coupled algebraic equations that involve Q_{ij} and Q_{ijk} probabilities. In the copolymerization theory, Q_{ijk} or other higher-order probabilities $Q_{i_1 i_2 \cdots i_n}$ are computed using the first-order Markov-chain decoupling approximation,[12−14] which is a nonlinear factorization conjecture given as

$$Q_{ijk} = \frac{Q_{ij}Q_{jk}}{Q_j} \tag{5}$$

or

$$Q_{i_1 i_2 \cdots i_n} = \frac{Q_{i_1 i_2} Q_{i_2 i_3} \cdots Q_{i_{n-1} i_n}}{Q_{i_2} Q_{i_3} \cdots Q_{i_{n-1}}} \tag{6}$$

Substituting eq 5 into eq 1 in the stationary limit, we obtain

$$k_{ij} Q_i + \sum_{l=\{R,W\}} k_{-jl} \frac{Q_{ij} Q_{jl}}{Q_j} = (k_{jR} + k_{jW} + k_{-ij}) Q_{ij} \tag{7}$$

Solving the above nonlinear equations along with the normalization condition $Q_{RR} + Q_{WR} + Q_{RW} + Q_{WW} = 1$ gives the stationary probabilities Q_{RR}, Q_{RW}, Q_{WR}, and Q_{WW} for any given set of rates. Substituting these probabilities into the equations for error and speed, given by eqs 2−4, we obtain the error and speed denoted by η_{NL} and V_{NL}, where the subindex reflects that these results correspond to the nonlinear copolymerization theory.

The enzyme-kinetic approach is an alternative method to analyze the molecular mechanisms of information-copying biological processes.[19−29] For the first-order polymerization processes, i.e., processes with first-order neighbor effects, this approach yields a discrete-state kinetic model as shown in Figure 1B. It accounts for all possible discrete states of the enzyme E that might occur before a product formation, which in this case is defined by one of four possible dimers: *rr*, *rw*, *wr*, *ww*. The addition of monomer R and W to the enzyme E leads to the states ER and EW, respectively. Then from EX, four different enzyme states EXY can be achieved by adding another subunit Y, for X, $Y \in \{R, \text{and } W\}$. Assuming that after the dimer product formation the enzyme resets to the state corresponding to the identity of the last incorporated monomer, the system can be described as a six-state kinetic model.

Notably, in contrast to the copolymerization theory, here, the six states, with probabilities P_i for $i \in \{ER, EW, ERR, ERW, EWR, EWW\}$, are independent states leading to the following normalization condition:

$$P_{ER} + P_{EW} + P_{ERR} + P_{ERW} + P_{EWR} + P_{EWW} = 1 \tag{8}$$

The advantage of using the enzyme-kinetics approach to describe copolymerization is that the dynamics in the system can be viewed as a set of linear transitions between several discrete states with constant rates (as we assume that monomers and cofactors are at fixed concentrations). This allows us to utilize powerful methods of either backward master equations or forward master equations to compute stationary fluxes, probabilities, errors, and speed (see the Supporting Information for details). Note that the stationary fluxes from the enzyme-kinetic approach are given by

$$J_{XY} = k_{XY} P_{EX} - k_{-XY} P_{EXY} \tag{9}$$

and the polymerization speed V is defined as the total right product formation flux, which is given as

$$V = 2(J_{RR} + J_{RW}) \tag{10}$$

Here, the factor 2 appears due to the formation of a dimer product in each enzymatic cycle.

We now compare the polymerization speed from the enzyme-kinetic approach with the speed from the copolymerization theory for the first-order polymerization processes with kinetic rates defined by the free energy landscape, depicted in Figure S1.[30] This landscape is described by 3 parameters: the chemical driving (polymerization) energy for nucleotide addition (ΔG_{pol}), the kinetic barrier heights between correct and incorrect substrates (ΔG_K), and the thermodynamic difference in binding energies between these substrates (ΔG_{TT}). Figure 2A reflects the speed V as a function of polymerization energy ΔG_{pol}, keeping ΔG_K and ΔG_{TT} constant for the two theories. The results revealed a discrepancy between the speed predicted by the enzyme-kinetic approach (solid line) and that obtained from the nonlinear copolymerization theory (symbols). Notably, these discrepancies were more pronounced at lower driving energies (Figure 2B). To determine which theory accurately predicts the speed, we compared the copolymerization theory's speed

https://doi.org/10.1021/acs.jpclett.4c02132
J. Phys. Chem. Lett. 2024, 15, 9361−9368

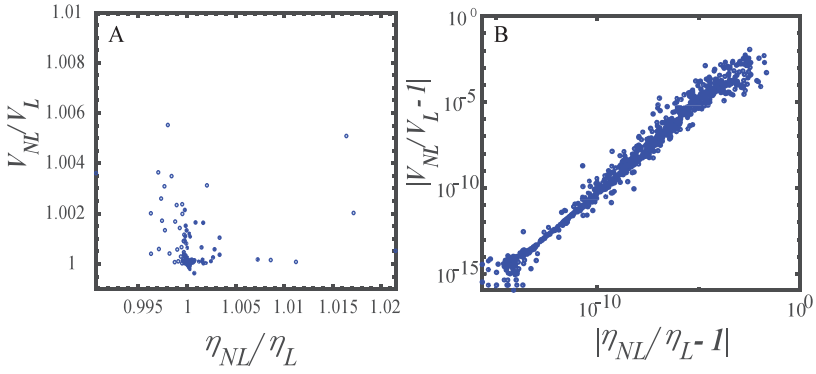


Figure 3. (A) Scatter plot for the ratio of errors (η_{NL}/η_L) versus the ratio of speeds (V_{NL}/V_L). (B) Scatter plot for $|(\eta_{NL}/\eta_L) - 1|$ vs $|(V_{NL}/V_L) - 1|$ on the logarithmic scale. The subscripts*NL* and *L* reflect the results corresponding to the nonlinear copolymerization theory and linear decoupling method, respectively. The error and speed values are obtained for the 1000 randomly distributed set of points in the parameter space defined by ΔG_K, ΔG_{TT}, and ΔG_{pol}. Specifically, $\Delta G_K \in [0, 10]$, $\Delta G_{TT} \in [-5, 10]$, $\Delta G_{pol} \in [0.5, 10]$.

estimates with those from Monte Carlo simulations (see the Supporting Information), and they were found to be in exact agreement, as shown in Figure S2. This implies that the speed estimates from the enzyme-kinetic approach are not always correct. This motivates the development of an alternative approach, which utilizes simpler steady-state linear algebraic equations similar to the enzyme-kinetic approach but can accurately predict error rates and speeds.

To overcome the drawback of the enzyme-kinetic approach in measuring correct fluxes, we propose an alternative method, namely, the linear-decoupling method, which approximates in a linear fashion the probabilities $Q_{i_1 i_2 \cdots i_n}$, for $n > 2$, involved in the chemical-kinetic equations (see eq 1). This linear approximation is relevant for biologically relevant conditions. More specifically, it relies on the experimentally measured kinetic parameters of DNA and RNA polymerases involved in DNA replication and transcription elongations.[4,33,34] For the first-order polymerization processes, these conditions can be described intuitively as the following. First, the successive addition of right subunits *R* dominates in high-fidelity biological processes, which means that $k_{RR} \gg k_{WR}$. This would also imply that $Q_{WR} \ll Q_{RR}$, where Q_{ij} denotes the stationary probability of copolymer chain &*ij*. Second, the successive additions of wrong subunits *W* are very improbable. This would imply that $Q_{WW} \ll Q_{WR}$, Q_{RW}.

Under these conditions, the exact nonlinear factorization for probabilities Q_{ijk} (eq 5) can be well approximated by the following linear factorization conjecture (see the Supporting Information for more details):

$$Q_{WRR} \approx Q_{WR} \Rightarrow Q_{RRR} \approx Q_{RR} - Q_{WR} \tag{11}$$

$$Q_{WRW} \approx 0 \Rightarrow Q_{RRW} \approx Q_{RW} \tag{12}$$

$$Q_{WWR} \approx 0 \Rightarrow Q_{RWR} \approx Q_{WR} \tag{13}$$

$$Q_{WWW} \approx 0 \Rightarrow Q_{RWW} \approx Q_{WW} \tag{14}$$

These relations reduce the stationary kinetic equations for Q_{ij} (eq 1) for the first-order polymerization process to the system of linear algebraic equations $\mathbf{M} \cdot \mathbf{Q} = \mathbf{0}$, where vector $\mathbf{Q} = [Q_{RR}, Q_{WR}, Q_{RW}, Q_{WW}]$, $\mathbf{0}$ is a 4×1 zero vector, and matrix $\mathbf{M}$ is given as

$$\mathbf{M} = \begin{pmatrix} -k_{RW} & k_{RR} - k_{-RR} & k_{-RW} & 0 \\ 0 & -(k_{RR} + k_{RW} + k_{-WR} - k_{-RR}) & k_{WR} & k_{WR} \\ k_{RW} & k_{RW} + k_{-WR} & -(k_{WR} + k_{WW} + k_{-RW}) & k_{-WW} \\ 0 & 0 & k_{WW} & -(k_{WR} + k_{-WW}) \end{pmatrix} \tag{15}$$

Solving system $\mathbf{M} \cdot \mathbf{Q} = \mathbf{0}$, under the condition $\mathbf{Q}^T \cdot \mathbf{1} = 1$, where $\mathbf{1}$ is a 4×1 unit vector, yields the stationary probabilities, Q_{ij}. The substitution of these probabilities into eqs 2 and 4, respectively, produces the error (η_L) and the speed (V_L) under the linear-decoupling approximation.

We can now test the linear-decoupling method by comparing its predictions of the error (η_L) and the speed (V_L) with the corresponding exact values obtained from the nonlinear copolymerization theory for the first-order polymerization processes with rates defined in terms of three parameters: ΔG_{pol}, ΔG_K, and ΔG_{TT}. Figure 3 shows the scatter plot of the ratio of errors (η_{NL}/η_L) and ratio of speeds (V_{NL}/V_L) computed at 1000 randomly distributed points in the parameter space defined by ΔG_K, ΔG_{TT}, and ΔG_{pol}. Specifically, $\Delta G_K \in [0, 10]$, $\Delta G_{TT} \in [-5, 10]$, and $\Delta G_{pol} \in [0.5, 10]$. The majority of the points cluster around (1, 1), indicating that for a large part of the parameter space, the results from the linear-decoupling approximation closely match the exact values from nonlinear copolymerization theory. Note that for better visualization of the results in Figure 3A, we scatter plot the absolute difference between the ratio of errors and ratio of speed, i.e., $|(\eta_{NL}/\eta_L) - 1|$ vs $|(V_{NL}/V_L) - 1|$, on the logarithmic scale, as shown in Figure 3B. The results in Figure 3 suggest that the linear-decoupling approximation is effective not only under biologically relevant conditions, specified by large driving energy and low errors, but also under other conditions, such as low driving energies ($\Delta G_{pol} \approx 1\ k_BT$), where the enzyme-kinetic approach fails to predict the correct fluxes.

https://doi.org/10.1021/acs.jpclett.4c02132
J. Phys. Chem. Lett. 2024, 15, 9361−9368

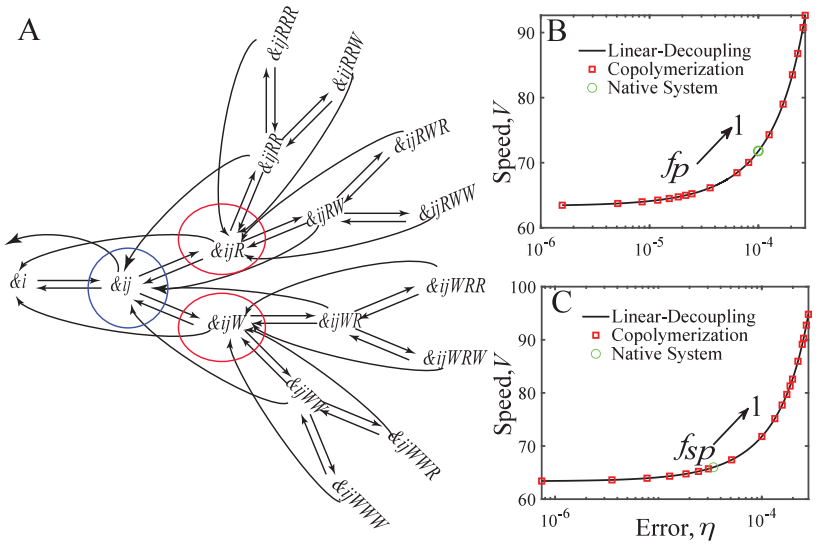


Figure 4. (A) Branching model for the polymerization and proofreading in the transcription elongation. (B) Speed−accuracy trade-off curve with changes in the first-order pausing discrimination parameter f_p, and other parameters are fixed as in Table S2. (C) Speed−accuracy trade-off curve with changes in the second-order pausing parameter f_{sp}, i.e., $k_{RWR} = k_{WWR} = k_{WRR} = f_{sp}k_{RRR}$ and other parameters fixed as in Table S3. Symbols and the solid lines represent the results from the nonlinear copolymerization theory and linear-decoupling method, respectively. The green circles in panels B and C denote the predicted speed−error values from the simulations at the kinetic parameters given in Tables S1 for the first-order and for parameters given in Tables S2 for the second-order polymerization-proofreading model at the kinetic parameters given in Tables S1 and S2, respectively. Arrows and lines within the blue circle and red circles, respectively, denote the flux coming into and out of the coarse-grained state &*ij* for the first-order model, and &*ijR*, &*ijW* for the second-order model for the transcription elongation.

However, the linear-decoupling may not predict correct results under some conditions significantly that are very different from typical biological situations, such as when the polymerization energy ΔG_{pol} is quite small (i.e., $<1k_BT$) (see Figure S3 in the Supporting Information).

To theoretically understand the validity condition of linear-decoupling method in comparison to nonlinear copolymerization theory, we found the condition on ΔG_{pol}, denoted by ΔG_{pol}^{min}, above which the error obtained from the linear decoupling method matches with the error from the nonlinear copolymerization theory within a significant accuracy range (see the Supporting Information for details). The density plot of the lower limit of ΔG_{pol} as a function of changes in the thermodynamic difference energy (ΔG_{TT}) and kinetic difference energy (ΔG_K) is shown in Figure S4 in the Supporting Information. The results suggest that the linear-decoupling method is applicable not only for small error systems but also for systems with moderate driving energies.

The advantage of the linear-decoupling method is that it can be applied to any information-processing biological process undergoing kinetic proofreading (KPR) schemes. We specifically applied the method to two biological systems.

Transcription elongation by RNA polymerase (RNAP) is a crucial system for which we can apply our method. During transcription elongation, RNAP selects the correct nucleotide (R) over incorrect (W) ones, pauses after incorporation (W), frequently backtracks, and cleaves dinucleotides[16,24,34−40] (see Figure 4A). When pausing after misincorporation has only a nearest-neighbor (first-order) effect,[16,35,39,40] the elongation process can be theoretically modeled using a first-order polymerization and dinucleotide cleavage model (Figure S4 and Figure 4A). In this model, the incorporation rate of a correct nucleotide is reduced if an incorrect monomer is present at the copolymer chain's terminus. The system's dynamic properties are obtained by solving the chemical-kinetic equations for the probabilities of copolymer chains &*ij* for $i, j \in \{R, W\}$ in the stationary limit, as shown in the Supporting Information. Using the linear-decoupling approximation and Markov-chain copolymerization theory, stationary probabilities, error rates, and speeds can be explicitly computed (see the Supporting Information).

There are some recent reports of a second-order pausing effect during biological polymerization,[24,41,42] where the incorporation rate of a correct nucleotide is slowed if an incorrect monomer is present at the terminal or penultimate position of the copolymer chain. The second-order polymerization and cleavage model is shown in Figure S5 and Figure 4A. Here, using our theoretical approach, we can obtain the system's dynamic properties by solving the chemical-kinetic equations for the probabilities of copolymer chains &*ijR* and &*ijW* for $i, j \in \{R, W\}$ in the stationary limit. Again, stationary probabilities, error rates, and speeds are computed using the linear-decoupling approximation and Markov-chain copolymerization theory (Supporting Information).

It is known that pausing might significantly impact the transcription elongation process.[43−45] Therefore, it is important to analyze the effect of pausing on the transcriptional accuracy and speed. We first analyze the effect of first-order pausing on transcriptional accuracy and speed using both linear and nonlinear decoupling approaches (see the Supporting Information). The reduction in polymerization rate after a mismatch is quantified by the discrimination factor f_p, defined as $f_p = \frac{k_{WR}}{k_{RR}}$. One can analyze the effect of f_p on the speed−accuracy trade-off curve, while keeping all other kinetic parameters fixed (Table S1, Figure 4B). As $f_p \to 1$, both error and speed increase monotonically, indicating a trade-off between accuracy and speed. The error−speed trade-off curve obtained from the linear-decoupling approximation under biologically relevant conditions matches well with the corresponding curve from the nonlinear copolymerization

https://doi.org/10.1021/acs.jpclett.4c02132
J. Phys. Chem. Lett. 2024, 15, 9361−9368

theory (Figure 4B), validating our theoretical method. Additionally, the accuracy and speed derived from both theories, using the kinetic parameters in Table S1, agree well with the simulation results for the native system parameter values given in Table S1 (green circle, Figure 4B).

Next, we examine the more realistic scenario of second-order pausing effects on the RNAP speed and accuracy. By varying the factor f_{sp}, such that $k_{RWR} = k_{WWR} = k_{WRR} = f_{sp}k_{RRR}$, we account for the reduction in polymerization rate after a mismatch at the terminal or penultimate position, or both (see Figure 4C and Figure S5). The results from the linear-decoupling approximation are consistent with the nonlinear copolymerization theory results. We observe that the speed−accuracy trade-off persists with second-order pausing. Furthermore, the accuracy and speed obtained from both theories agree well with the simulation results at the native system parameter values given in Table S2 (green circle, Figure 4C). Notably, second-order pausing enhances the accuracy by nearly an order of magnitude with a slight speed reduction. This is consistent with the results from the enzyme-kinetic approach found recently in ref 24; see the Supporting Information for details.

DNA replication by DNA polymerase (DNAP) is another critical biological copying process that can be regarded as a binary copolymerization process of matched nucleotides (R), and mismatched nucleotides (W). The fidelity and speed of the DNA replication process are controlled by the two protein domains, polymerase (p) and exonuclease (e).[6,15,25,31,33] The polymerase domain binds the incoming dNTP molecule, whereas the exonuclease domain proofreads the just-incorporated mismatched nucleotide, transferred from the polymerase domain, as shown in the schematic picture for DNA replication in Figure 5A. In this model, the attachment rate (k_{ij}) and detachment rate (k_{-ij}) of a nucleotide (j) not only depend on the nucleotide j itself but also on the previous nucleotide i bound to the polymerase at the growing end ($\&i^p$). The rates

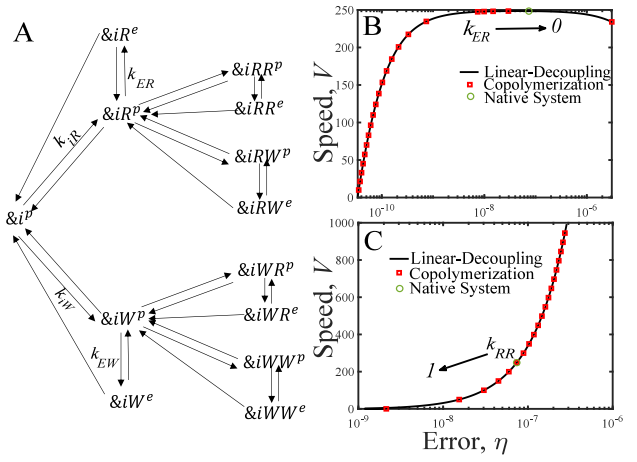


Figure 5. (A) Branching model for the first-order polymerization and excision for the DNA replication process. (B) Speed−accuracy trade-off curve with changes in the pol-exo sliding rate k_{ER}, with fixed f_E, and other kinetic parameters of Table S3. (C) Speed−accuracy trade-off curve with changes in the polymerization rate k_{RR}, with fixed f_R, and other kinetic parameters of Table S3. Symbols and solid lines represent the results from nonlinear and linear factorization of probabilities, respectively. The green circle denotes the predicted error−speed values from the simulations at the kinetic parameters given in Table S3.

of sliding between the polymerase domain and exonuclease domain, $\&ij^p \leftrightarrow \&ij^e$, are denoted by $k_{\pm Ej}$, whereas the rates of cleavage of nucleotide j from the exonuclease domain is k_{Cj} (see Figure 5A). The kinetic parameters of the model are based on the experimental data of Wong et al.,[33] and they are listed in Table S3.

The kinetics of the DNAP is described in terms of the probabilities, $Q^p_{ij}(t)$, and $Q^e_{ij}(t)$ corresponding to the copolymer chains, $\&ij$, ending with dimers ij in the polymerase and exonuclease domain, respectively, at time t. Detailed chemical-kinetic equations governing the process are provided in the Supporting Information. Those equations can be solved for the steady-state probabilities under both nonlinear copolymerization theory and the linear-decoupling method. The stationary probabilities are further utilized to compute the errors and the speeds of the replication process (see the Supporting Information for details).

Now, one can test the linear-decoupling method for the model in Figure 5A by comparing its predictions with the ones obtained from the exact nonlinear copolymerization theory. We start by observing the effect of the sliding rate of DNAP between the polymerase and exonuclease domains, which is a critical step in the proofreading mechanism. The error−speed ($\eta - V$) trade-off curve as a function of the pol-exo sliding rate k_{ER}, with fixed f_E, where $f_E = \frac{k_{EW}}{k_{ER}}$, and other kinetic rates as given in Table S3 are plotted in Figure 5B,C. The predictions from the linear-decoupling approximation closely match the exact results from the nonlinear approach. Notably, the minimum error is asymptotically approached at high k_{ER} values, consistent with findings from the previous enzyme-kinetic studies employing backward master equation formalism for the chemical-kinetic model in ref 25 (Figure S10). The speed−accuracy trade-off emerges after k_{ER} exceeds the value corresponding to the peak speed. The native system (green circle) at the experimental value of k_{ER} lies on the nontrade-off branch of the $\eta - V$ curve. This indicates that speed is prioritized in the DNA replication process as long as a reasonable level of accuracy is maintained.[25] In the Supporting Information, we also solve the chemical-kinetic model (Figure S10) for the DNA replication process using the forward master equation formalism, yielding the linear kinetic equations for discrete states in the model. Interestingly, we found that the error−speed trade-off curve obtained from the forward master equation formalism for the enzyme-kinetic approach agrees very well with the corresponding trade-off curve obtained from the linear-decoupling method at and near the kinetic parameters value for the system, given in Table S3 (see Figure S11). This implies that the enzyme-kinetic approach works well in computing the stationary properties such as error and speed at and near the biologically relevant kinetic parameters.

Next, we computed the error−speed trade-off curve as a function of the polymerization rate k_{RR}, with fixed f_R, where $f_R = \frac{k_{RW}}{k_{RR}}$, $f_p = \frac{k_{WR}}{k_{RR}}$, and other kinetic rates in Table S3. The error−speed values under the linear-decoupling approximation align well with the exact values from the nonlinear copolymerization theory over a broad range of parameters. The system (green circle) at the kinetic parameters in Table S3 lies on the trade-off branch of the $\eta - V$ curve and is significantly distant from the minimum error point. Achieving the minimum error would result in a substantial loss of speed, which may not be acceptable. This suggests that speed is

https://doi.org/10.1021/acs.jpclett.4c02132
J. Phys. Chem. Lett. 2024, 15, 9361−9368

prioritized in DNA replication as long as an acceptable upper level of η is maintained.

A new theoretical method for evaluating the dynamic properties of biological information transfer processes is developed. More specifically, we concentrated on analyzing copolymerization phenomena when a new biopolymer molecule (nucleic acid or protein) is formed by replicating the substrates in an already existing template strand. Stimulated by observations that existing theoretical methods are too complex for understanding the underlying molecular mechanisms (copolymerization theory) or do not always work well (enzyme-kinetics approach), a new theoretical framework that is precise and relatively simple is presented. It is based on the linear-decoupling approximation for biological conditions with either small errors or large driving energies. The method simplifies the chemical-kinetic equations into a set of linear algebraic equations, facilitating analytical calculations. It not only avoids the nonlinear complexities arising in the copolymerization theory but also overcomes the limitation of the enzyme-kinetic approach in predicting the correct stationary fluxes. It is found that the error rates and polymerization speeds can be reliably estimated not only under the biologically relevant conditions but also under a very broad range of kinetic parameters. Moreover, we applied the method to various biological processes, such as transcription elongation and DNA replication processes. In all systems, the predictions of our theoretical approach agreed well with the numerically exact results, showing the robustness of the proposed method. Thus, our approach provides a convenient theoretical tool for investigating the molecular mechanisms of complex biological information-copying processes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.4c02132.

Computation of stationary probabilities for irreversible polymerization process with first-order neighbor effects; forward master equations formalism for the stochastic model from enzyme-kinetic approach; linear-decoupling approximation for first-order copolymerization processes; validity condition for the linear decoupling method for the case of large drive; application of copolymerization theory and linear-decoupling to the transcription elongation with first-order polymerization and dinucleotide cleavage; application of copolymerization theory and linear-decoupling to the transcription elongation with second-order polymerization and dinucleotide cleavage; computation of error and speed for the model of transcription elongation with second-order pausing; application of copolymerization theory and linear-decoupling to DNA replication with first-order polymerization and kinetic proofreading; computation of error and speed for the chemical-kinetic model of DNA replication; Monte Carlo simulations (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Anatoly B. Kolomeisky − *Center for Theoretical Biological Physics, Department of Chemistry, Department of Chemical and Biomolecular Engineering, and Department of Physics and Astronomy, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0001-5677-6690; Email: tolya@rice.edu

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

Author

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

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.4c02132

Author Contributions

#A.B.K. and O.A.I. contributed equally and are the authors to whom the correspondence is to be addressed.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Center for Theoretical Biological Physics supported by the NSF (PHY-2019745). O.A.I. acknowledges support from the Welch Foundation Grant C-1995. A.B.K. also acknowledges the support from the Welch Foundation (C-1559). This work was performed in part at Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452.

■ REFERENCES

(1) Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences* **1974**, *71*, 4135−4139.

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

(3) Crick, F. Central dogma of molecular biology. *Nature* **1970**, *227*, 561−563.

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

(5) Reynolds, N. M.; Lazazzera, B. A.; Ibba, M. Cellular mechanisms that control mistranslation. *Nature Reviews Microbiology* **2010**, *8*, 849−856.

(6) Johnson, K. A. Conformational coupling in DNA polymerase fidelity. *Annual review of biochemistry* **1993**, *62*, 685−713.

(7) Kunkel, T. A.; Bebenek, K. DNA replication fidelity. *Annual review of biochemistry* **2000**, *69*, 497−529.

(8) Mayo, F. R.; Walling, C. Copolymerization. *Chemical Reviews* **1950**, *46*, 191−287.

(9) Berlin, A. A.; Volfson, S. A.; Enikolopian, N. S. Kinetics of polymerization processes. *Polymerization Processes* **1981**, *38*, 89−140.

(10) Boulding, N. A.; Millican, J. M.; Hutchings, L. R. Understanding copolymerisation kinetics for the design of functional copolymers via free radical polymerisation. *Polymer Chemistry* **2019**, *10*, 5665−5675.

(11) Qureshi, B.; Juritz, J.; Poulton, J. M.; Beersing-Vasquez, A.; Ouldridge, T. E. A universal method for analyzing copolymer growth. *The Journal of Chemical Physics* **2023**, *158*, 104906.

(12) Gaspard, P.; Andrieux, D. Kinetics and thermodynamics of first-order Markov chain copolymerization. *The Journal of chemical physics* **2014**, *141*, 044908.

(13) Gaspard, P. Growth and dissolution of macromolecular Markov chains. *J. Stat. Phys.* **2016**, *164*, 17−48.

(14) Shu, Y.-G.; Song, Y.-S.; Ou-Yang, Z.-C.; Li, M. A general theory of kinetics and thermodynamics of steady-state copolymerization. *Journal of Physics: Condensed Matter* **2015**, *27*, 235105.

(15) Song, Y.-S.; Shu, Y.-G.; Zhou, X.; Ou-Yang, Z.-C.; Li, M. Proofreading of DNA polymerase: a new kinetic model with higher-order terminal effects. *Journal of Physics: Condensed Matter* **2017**, *29*, 025101.
(16) Fu, W.; Li, Q.; Song, Y.; Shu, Y.; Ouyang, Z.; Li, M. Theoretical analysis of RNA polymerase fidelity: a steady-state copolymerization approach. *Commun. Theor. Phys.* **2022**, *74*, 015601.
(17) Segel, I. H. *Enzyme kinetics: behavior and analysis of rapid equilibrium and steady state enzyme systems*; Wiley: New York, 1975; Vol. *115*.
(18) Engel, P. C. *Enzyme kinetics: the steady-state approach*; Springer: New York, NY, 2013.
(19) Bel, G.; Munsky, B.; Nemenman, I. The simplicity of completion time distributions for common complex biochemical processes. *Physical biology* **2010**, *7*, 016003.
(20) de Oliveira, L. P.; Hudebine, D.; Guillaume, D.; Verstraete, J. J. A review of kinetic modeling methodologies for complex processes. *Oil & Gas Science and Technology—Revue d'IFP Energies nouvelles* **2016**, *71*, 45.
(21) Laidler, K. J.; Keith, J. et al. *Chemical kinetics*; McGraw-Hill: New York, 1965; Vol. *2*.
(22) Cao, J. Michaelis- Menten Equation and Detailed Balance in Enzymatic Networks. *The Journal of Physical Chemistry B* **2011**, *115*, 5493−5498.
(23) Mallory, J. D.; Igoshin, O. A.; Kolomeisky, A. B. Do we understand the mechanisms used by biological systems to correct their errors? *The Journal of Physical Chemistry B* **2020**, *124*, 9289−9296.
(24) Midha, T.; Mallory, J. D.; Kolomeisky, A. B.; Igoshin, O. A. Synergy among Pausing, Intrinsic Proofreading, and Accessory Proteins Results in Optimal Transcription Speed and Tolerable Accuracy. *The Journal of Physical Chemistry Letters* **2023**, *14*, 3422−3429.
(25) Banerjee, K.; Kolomeisky, A. B.; Igoshin, O. A. Elucidating interplay of speed and accuracy in biological error correction. *Proceedings of the National Academy of Sciences* **2017**, *114*, 5183−5188.
(26) 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. *The Journal of Physical Chemistry Letters* **2021**, *12*, 2691−2698.
(27) Yu, Q.; Mallory, J. D.; Kolomeisky, A. B.; Ling, J.; Igoshin, O. A. Trade-offs between speed, accuracy, and dissipation in tRNAIle aminoacylation. *The journal of physical chemistry letters* **2020**, *11*, 4001−4007.
(28) Iwahara, J.; Kolomeisky, A. B. Discrete-state stochastic kinetic models for target DNA search by proteins: Theory and experimental applications. *Biophysical chemistry* **2021**, *269*, 106521.
(29) Park, W.; Moon, I. A discrete multi states model for the biological production of hydrogen by phototrophic microalga. *Biochemical engineering journal* **2007**, *36*, 19−27.
(30) Midha, T.; Kolomeisky, A. B.; Igoshin, O. A. Insights into Error Control Mechanisms in Biological Processes: Copolymerization and Enzyme-Kinetics Revisited. *The Journal of Physical Chemistry B* **2024**, *128*, 5612−5622.
(31) Gaspard, P. Kinetics and thermodynamics of living copolymerization processes. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2016**, *374*, 20160147.
(32) Poulton, J. M.; ten Wolde, P. R.; Ouldridge, T. E. Nonequilibrium correlations in minimal dynamical models of polymer copying. *Proceedings of the National Academy of Sciences* **2019**, *116*, 1946−1951.
(33) Wong, I.; Patel, S. S.; Johnson, K. A. An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry* **1991**, *30*, 526−537.
(34) Zenkin, N.; Yuzenkova, Y.; Severinov, K. Transcript-assisted transcriptional proofreading. *Science* **2006**, *313*, 518−520.
(35) Thomas, M. J.; Platas, A. A.; Hawley, D. K. Transcriptional fidelity and proofreading by RNA polymerase II. *Cell* **1998**, *93*, 627−637.
(36) 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.
(37) Gamba, P.; Zenkin, N. Transcription fidelity and its roles in the cell. *Curr. Opin. Microbiol.* **2018**, *42*, 13−18.
(38) 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.
(39) Mellenius, H.; Ehrenberg, M. DNA template dependent accuracy variation of nucleotide selection in transcription. *PloS one* **2015**, *10*, e0119588.
(40) Mellenius, H.; Ehrenberg, M. Transcriptional accuracy modeling suggests two-step proofreading by RNA polymerase. *Nucleic Acids Res.* **2017**, *45*, 11582−11593.
(41) 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.
(42) Knippa, K.; Peterson, D. O. Fidelity of RNA polymerase II transcription: role of Rbp9 in error detection and proofreading. *Biochemistry* **2013**, *52*, 7807−7817.
(43) Lee, D. N.; Phung, L.; Stewart, J.; Landick, R. Transcription pausing by Escherichia coli RNA polymerase is modulated by downstream DNA sequences. *J. Biol. Chem.* **1990**, *265*, 15145−15153.
(44) 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.
(45) 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.

https://doi.org/10.1021/acs.jpclett.4c02132
J. Phys. Chem. Lett. 2024, 15, 9361−9368