Error correction during DNA replication

Ajeet K. Sharma and Debashish Chowdhury*
Department of Physics, Indian Institute of Technology, Kanpur 208016, India
(Received 11 January 2012; revised manuscript received 4 June 2012; published 16 July 2012)

DNA polymerase (DNAP) is a dual-purpose enzyme that plays two opposite roles in two different situations during DNA replication. It plays its a normal role as a *polymerase* catalyzing the elongation of a new DNA molecule by adding a monomer. However, it can switch to the role of an *exonuclease* and shorten the same DNA by cleavage of the last incorporated monomer from the nascent DNA. Just as misincorporated nucleotides can escape exonuclease causing a replication error, the correct nucleotide may get sacrificed unnecessarily by erroneous cleavage. The interplay of polymerase and exonuclease activities of a DNAP is explored here by developing a minimal stochastic kinetic model of DNA replication. Exact analytical expressions are derived for a few key statistical distributions; these characterize the temporal patterns in the mechanical stepping and the chemical (cleavage) reaction. The Michaelis-Menten-like analytical expression derived for the average rates of these two processes not only demonstrate the effects of their coupling, but are also utilized to measure the extent of *replication error* and *erroneous cleavage*.

DOI: 10.1103/PhysRevE.86.011913 PACS number(s): 87.16.ad, 87.16.dj

I. INTRODUCTION

DNA polymerase (DNAP) replicates a DNA molecule; the sequence of the nucleotides, the monomeric subunit of DNA, on the product of polymerization is dictated by that on the corresponding template DNA through the Watson-Crick complimentary base-pairing rule [1]. DNAP moves step by step along the template strand utilizing chemical energy input and, therefore, these are also regarded as a molecular motor [2,3].

A unique feature of DNAP is that it is a dual-purpose enzyme that plays two opposite roles in two different circumstances during DNA replication. It plays its normal role as a polymerase catalyzing the *elongation* of a new DNA molecule. However, upon committing an error by the misincorporation of a wrong nucleotide, it switches its role to that of an exonuclease catalyzing the *shortening* of the nascent DNA by cleavage of the misincorporated nucleotide at the growing tip of the elongating DNA [4]. The two distinct sites on the DNAP where, respectively, polymerization and cleavage are catalyzed are separated by 3–4 nm [5]. The nascent DNA is transferred back to the site of polymerization after cleaving the incorrect nucleotide from its growing tip. The elongation and cleavage reactions are thus *coupled* by the transfer of the DNA between the sites of polymerase and exonuclease activity of the DNAP. However, the physical mechanism of this transfer is not well understood [6].

In this paper we develop a minimal kinetic model of DNA replication (more precisely, that of the "leading strand" which can proceed continuously) that captures the coupled polymerase and exonuclease activities of a DNAP within the same theoretical framework. From this model, we derive the exact analytical expressions for (i) the dwell time distribution (DTD) of the DNAP at the successive nucleotides on the template DNA, and (ii) the distribution of the turnover times (DTTs) of the exonuclease (i.e., the time intervals between the successive events of cleavage of misincorporated nucleotide from the nascent DNA). The mean of these two distributions characterizes the average rates of elongation and cleavage, respectively; we show that both can be written as Michaelis-Menten-like expressions for enzymatic reactions which reveals the effect of coupling explicitly.

In our model, the kinetic pathways available to the correct and incorrect nucleotides are the same. However, it is the ratio of the rate constants that makes a pathway more favorable to one species than to the other. A similar assumption was made by Galas and Branscomb [7] in one of the earliest models of replication. Therefore, in spite of the elaborate quality control system, some misincorporated nucleotides can escape cleavage; such a *replication error* in the final product is usually about 1 in 10^9 nucleotides. Moreover, occasionally a correct nucleotide is erroneously cleaved unnecessarily; such "futile" cycles slow down replication [8].

We define quantitative measures of these two types of error and derive their exact analytical expressions from our model for "wild type" DNAP. Using special cases of these analytical expressions, we also examine the effects of two different mutations of the DNAP [5]: (i) the "exodeficient" mutant that is incapable of exonuclease activity, and (ii) the "transfer-deficient" mutant on which the rate of transfer to the exonuclease site is drastically reduced.

II. MODEL

Almost all DNAPs share a common "right-hand-like" structure. Binding of the correct dNTP substrate triggers closing of the "hand" which is required for the formation of the diester bond between the recruited nucleotide monomer and the elongating DNA molecule. The kinetic scheme of our stochastic model of replication is shown in Fig. 1. The rate constants for the correct and incorrect nucleotides are denoted by ω and Ω, respectively; the same subscript is used in both the cases for the same transition.

Let us begin with the situation where the DNAP is ready to begin its next elongation cycle; this mechanochemical state is labeled by the integer index 1. In principle, the transition $1 \rightarrow 2$ consists of two steps: the binding of the dNTP substrate

*Corresponding author: debch@iitk.ac.in

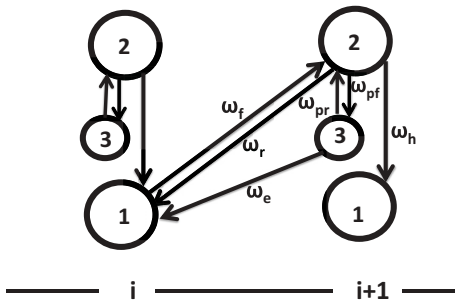

FIG. 1. The kinetic model of DNA replication. The chemomechanical states of a single DNAP and the allowed transitions are shown along with the corresponding transition probabilities per unit time (the rate constants). The corresponding rate constants for the incorrect nucleotide are denoted by the symbol Ω (see the text for details).

and the formation of the diester bond. The overall rate of this step is ω_f for a correct substrate and Ω_f for an incorrect substrate.

Occasionally, because of the random fluctuation of the hand between the "open" and "closed" conformations, the dNTP may escape even before the formation of the diester bond; this takes place with the rate constant ω_r. If the recruited dNTP is incorrect, the hand remains open most of the time and the rate constant for the rejection of the dNTP is Ω_r ($\Omega_r \gg \omega_r$). Note that dNTP selection through $1 \rightleftharpoons 2$ involves a discrimination between the correct and incorrect dNTP substrate on the basis of free energy gained by complementary base pairing with the template.

The transition $2(i+1) \rightarrow 1(i+1)$ corresponds to the *relaxation* of the freshly incorporated nucleotide to a conformation that allows the DNAP to be ready for the next cycle. The rate constants for this step are ω_h and Ω_h, respectively, for correctly and incorrectly incorporated nucleotides. Alternatively, while in the state $2(i+1)$, the DNAP can transfer the growing DNA molecule to its exonuclease site; this transfer takes place at a rate ω_{pf} (Ω_{pf}) if the selected nucleotide is correct (incorrect). Since $\Omega_h \ll \omega_h$, and $\Omega_{pf} \gg \omega_{pf}$, the misincorporated nucleotide most often gets transferred to the exonuclease site whereas relaxation, rather than transfer, is the most probable pathway when the incorporated nucleotide is correct.

The actual cleavage of the diester bond that severs the nucleotide at the growing tip of the DNA is represented by the transition $3(i+1) \rightarrow 1(i)$. For a correct nucleotide, $\omega_{pr} \gg \omega_e$ indicating that the DNA is likely to be transferred back to the polymerase site without the unnecessary cleavage of the correct nucleotide. In contrast, for an incorrect nucleotide, $\Omega_{pr} \ll \Omega_e$ which makes error correction a highly probable event. Moreover, $\Omega_{pr} \ll \omega_{pr}, \Omega_e \gg \omega_e$. Since the trimmed DNA is transferred to the polymerase site extremely rapidly [4], each of the rate constants ω_e and Ω_e incorporate both the trimming and transfer.

Interestingly, the full kinetic scheme in Fig. 1 can be viewed as a *coupling* of a purely polymerase-catalyzed reaction (shown in the left panel of Fig. 2) and a purely exonuclease-catalyzed reaction [shown in the right panel of Fig. 2]; the transition with the rate ω_{pr} couples these two reactions.

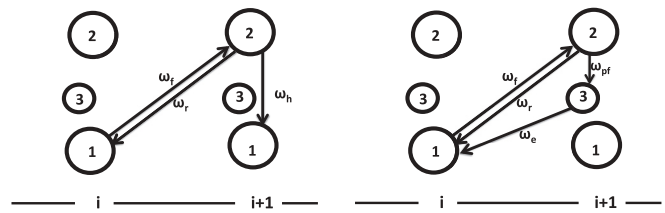

FIG. 2. "Purely" polymerizing reaction (left panel) and "pure" exonuclease reaction (right panel).

Strictly speaking, for an exodeficient DNAP [5], $\omega_e = 0 = \Omega_e$ although the rates of forward and reverse transfer between the sites of polymerase and exonuclease activities may not be necessarily negligible. Similarly, either $\omega_{pf} = 0 = \Omega_{pf}$ or $\omega_{pr} = 0 = \Omega_{pr}$ (or, both) can be the cause of transfer deficiency of the DNAP.

III. RESULTS AND DISCUSSION

A. Distribution of dwell time

The DTD considered here arises from intrinsic stochasticity and not caused by any sequence inhomogeneity of the mRNA template [9]. For a molecular motor that is allowed to step backward as well as forward, we use positive (+) and negative (−) signs to represent the forward and backward steps, respectively. For example, $\psi_{+-}(t)$ is the *conditional* DTD (cDTD) [10] when a forward step is followed by a backward step and p_{+-} is the probability of such a transition. Therefore, the DTD can be written as

$$\psi(t) = p_{+-}\psi_{+-}(t) + p_{++}\psi_{++}(t) + p_{-+}\psi_{-+}(t) + p_{--}\psi_{--}(t). \quad (1)$$

In our model two consecutive backward steps are forbidden, which implies that $p_{--}\psi_{--}(t) = 0$. We calculate the cDTD following the standard method [11] that has been used successfully earlier for the calculation of cDTD for some other motors (see, for example, Ref. [10]).

Let $P_\mu(j,t)$ be the probability of finding the DNAP in the μth ($\mu = 1,2,3$) chemical state at the jth site [i.e., at the discrete position x_j ($j = -\infty, \ldots -1,0,1,\ldots \infty$)]. Then master equations for $P_\mu(j,t)$ are

$$\frac{dP_1(j,t)}{dt} = -\omega_f P_1(j,t) + \omega_r P_2(j+1,t) + \omega_h P_2(j,t) + \omega_e P_3(j+1,t), \quad (2)$$

$$\frac{dP_2(j,t)}{dt} = \omega_f P_1(j-1,t) - (\omega_r + \omega_{pf} + \omega_h)P_2(j,t) + \omega_{pr} P_3(j,t), \quad (3)$$

$$\frac{dP_3(j,t)}{dt} = \omega_{pf} P_2(j,t) - (\omega_e + \omega_{pr})P_3(j,t). \quad (4)$$

In terms of the Fourier transform

$$\bar{P}_\mu(q,t) = \sum_{j=-\infty}^{\infty} P_\mu(x_j,t)e^{-iqx_j} \quad (5)$$

of $P_\mu(x_j,t)$, the master equations can be written as a matrix equation,

$$\frac{d}{dt}\bar{\mathbf{P}}(q,t) = \mathbf{M(q)}\bar{\mathbf{P}}(q,t), \tag{6}$$

where $\bar{\mathbf{P}}(q,t)$ is a column vector whose three components are $\bar{P}_1(q,t), \bar{P}_2(q,t), \bar{P}_3(q,t)$ and

$$\mathbf{M(q)} = \begin{bmatrix} -\omega_f & \omega_h + \omega_r\rho_-(q) & \omega_e\rho_-(q) \\ \omega_f\rho_+(q) & -(\omega_h + \omega_r + \omega_{pf}) & \omega_{pr} \\ 0 & \omega_{pf} & -(\omega_e + \omega_{pr}) \end{bmatrix} \tag{7}$$

with $\rho_+(q) = e^{-iqd}$ and $\rho_-(q) = e^{iqd}$; d being the step size, i.e., $x_{j+1} - x_j = d$. Taking Laplace transform of (6) with respect to time,

$$\tilde{P}_\mu(q,s) = \int_0^\infty \bar{P}_\mu(q,t)e^{-st}, \tag{8}$$

the solution of the master equation in the Fourier-Laplace space is

$$\tilde{\mathbf{P}}(q,s) = \mathbf{R}(q,s)^{-1}\tilde{\mathbf{P}}(0), \tag{9}$$

where

$$\mathbf{R}(q,s) = s\mathbf{I} - \mathbf{M}(q), \tag{10}$$

and $\tilde{\mathbf{P}}(0)$ is the column vector corresponding to the initial probabilities.

Now we define

$$\tilde{P}(q,s) = \sum_{i=1}^{3} \tilde{P}_i(q,s), \tag{11}$$

which can be calculated from

$$\tilde{P}(q,s) = \frac{\sum_{i,j=1}^{3} \mathcal{C}_{j,i} P_j(0)}{|\mathbf{R}(q,s)|}, \tag{12}$$

where $\mathcal{C}_{j,i}$ are the cofactors of the $\mathbf{R}(q,s)$.

The determinant of the matrix $\mathbf{R}(q,s)$ is a third order polynomial of s and can be expressed as

$$|\mathbf{R}(q,s)| = s^3 + \alpha s^2 + \beta(q)s + \gamma(q). \tag{13}$$

Note that α is independent of q whereas β and γ are the functions of q. For the explicit form (7) of $\mathbf{M}$ the coefficients α, $\beta(q)$, and $\gamma(q)$ are given below:

$$\alpha = \omega_e + \omega_f + \omega_h + \omega_{pf} + \omega_{pr} + \omega_r; \tag{14}$$

β can be expressed as

$$\beta(q) = \beta(0) + \beta_+[1 - \rho_+(q)] + \beta_{+-}[1 - \rho_+(q)\rho_-(q)], \tag{15}$$

where

$$\beta(0) = \omega_e\omega_f + \omega_e\omega_h + \omega_e\omega_{pf} + \omega_f\omega_{pf} + \omega_f\omega_{pr} + \omega_h\omega_{pr} + \omega_e\omega r + \omega_{pr}\omega_r, \tag{16}$$

$$\beta_+ = \omega_f\omega_h, \tag{17}$$

and

$$\beta_{+-} = \omega_f\omega_r. \tag{18}$$

Similarly,

$$\gamma(q) = \gamma_+[1 - \rho_+(q)] + \gamma_{+-}[1 - \rho_+(q)\rho_-(q)] \tag{19}$$

where

$$\gamma_+ = \omega_e\omega_f\omega_h + \omega_f\omega_h\omega_{pr} \tag{20}$$

and

$$\gamma_{+-} = \omega_e\omega_f\omega_{pf} + \omega_e\omega_f\omega_r + \omega_f\omega_{pr}\omega_r. \tag{21}$$

For convenience, we define the 2×2 diagonal matrix

$$\rho(\mathbf{q}) = \begin{bmatrix} \rho_+(q) & 0 \\ 0 & \rho_-(q) \end{bmatrix}, \tag{22}$$

the column vector

$$\mathbf{\Psi}(s) = \frac{1}{s}\begin{bmatrix} 1 - p_{++}\psi_{++}(s) - p_{+-}\psi_{+-}(s) \\ 1 - p_{-+}\psi_{-+}(s) - p_{--}\psi_{--}(s) \end{bmatrix}, \tag{23}$$

and the 2×2 matrix

$$\psi(\mathbf{s}) = \begin{bmatrix} p_{++}\psi_{++}(s) & p_{+-}\psi_{+-}(s) \\ p_{-+}\psi_{-+}(s) & p_{--}\psi_{--}(s) \end{bmatrix}, \tag{24}$$

where $\psi_{\pm\pm}(s)$ are the Laplace transforms of the cDTDs $\psi_{\pm\pm}(t)$.

$\tilde{P}(q,s)$ and cDTD are related [11] by the equation

$$\tilde{P}(q,s) = \mathbf{p}_0^T[\mathbf{I} - \psi(s)\rho(q)]^{-1}\mathbf{\Psi}(s), \tag{25}$$

where $\mathbf{p}_0$ is the vector of initial conditions. For example, $\mathbf{p}_0^T = (10)$ corresponds to the given condition that the motor has taken the initial step in the forward $(+)$ direction.

Thus, in principle, if one can calculate $\tilde{P}(q,s)$, one can use the relation (25) to solve for $\psi_{\pm\pm}(s)$ and, then taking inverse Laplace transform, obtain $\psi_{\pm\pm}(t)$. To calculate $\tilde{P}(q,s)$, one has to use an appropriate set of initial conditions consistent with the definition of the dwell times. The set $P_1(0) = 0, P_2(0) = 1, P_3(0) = 0$ ensures that the first step is taken forward. In other words, in our calculation, we start the clock by setting it to $t = 0$ when the DNAP reaches the state 2 at j from state 1 at $j - 1$. Therefore in this case $p_{--} = 0 = p_{-+}$. Corresponding to this initial condition, we now define

$$\tilde{P}_+(q,s) = \tilde{P}(q,s)|_{\{P_1(0)=0, P_2(0)=1, P_3(0)=0\}} \tag{26}$$

and, from equation (25), we get [11]

$$\left.\frac{1}{s\tilde{P}_+(q,s)}\right|_{\{\rho_-(q)=0\}} = \frac{1 - \rho_+(q)p_{++}\tilde{\psi}_{++}(s)}{1 - p_{++}\tilde{\psi}_{++}(s) - p_{+-}\tilde{\psi}_{+-}(s)}. \tag{27}$$

Equation (27) can be re expressed as

$$\left.\frac{1}{s\tilde{P}_+(q,s)}\right|_{\{\rho_-(q)=0\}} = a_0 + a_+\rho_+(q), \tag{28}$$

where

$$a_0 = \frac{1}{1 - p_{++}\tilde{\psi}_{++}(s) - p_{+-}\tilde{\psi}_{+-}(s)},$$
$$a_+ = -\frac{p_{++}\tilde{\psi}_{++}(s)}{1 - p_{++}\tilde{\psi}_{++}(s) - p_{+-}\tilde{\psi}_{+-}(s)}. \tag{29}$$

Hence,

$$p_{++}\tilde{\psi}_{++}(s) = -\frac{a_+}{a_0} \tag{30}$$

and

$$p_{+-}\tilde{\psi}_{+-}(s) = \frac{a_0 + a_+ - 1}{a_0}. \tag{31}$$

Therefore next we obtain $\frac{1}{s\tilde{P}_+(q,s)}|_{\{\rho_-(q)=0\}}$ directly from (12) and, by comparing it with equation (28), find out the expressions for a_0 and a_+; substituting these expressions for a_0 and a_+ into Eqs. (30) and (31) we get $p_{++}\tilde{\psi}_{++}(s)$ and $p_{+-}\tilde{\psi}_{+-}(s)$, respectively.

Using the same initial condition, from Eq. (12), we get

$$\tilde{P}_+(q,s) = \frac{s^2 + s\{\alpha - \omega_r[1-\rho_-(q)]\} + \beta(0) - [1-\rho_-(q)](\omega_e\omega_{pf} + \omega_e\omega_r + \omega_{pr}\omega_r)}{s^3 + \alpha s^2 + \beta(q)s + \gamma(q)}. \tag{32}$$

Therefore,

$$\left.\frac{1}{s\tilde{P}_+(q,s)}\right|_{\{\rho_-(q)=0\}} = \frac{s^3 + \alpha s^2 + s[\beta(0) + \beta_+ + \beta_{+-}] + \gamma_+ + \gamma_{+-} - (s\beta_+ + \gamma_+)\rho_+(q)}{s^3 + s^2(\alpha - \omega_r) + s[\beta(0) - (\omega_e\omega_{pf} + \omega_e\omega_r + \omega_{pr}\omega_r)]}. \tag{33}$$

Comparing Eq. (33) with Eq. (28) we identify a_0 and a_+ and substituting these expressions for a_0 and a_+ into (30) we get

$$\begin{aligned} p_{++}\tilde{\psi}_{++}(s) &= \frac{s\beta_+ + \gamma_+}{s^3 + \alpha s^2 + s[\beta(0) + \beta_+ + \beta_{+-}] + \gamma_+ + \gamma_{+-}} \\ &= \frac{s\beta_+ + \gamma_+}{(s+\omega_1)(s+\omega_2)(s+\omega_3)}, \end{aligned} \tag{34}$$

where ω_1, ω_2, and ω_3 are roots of the following equation:

$$\omega^3 - \alpha\omega^2 + \omega[\beta(0) + \beta_+ + \beta_{+-}] - (\gamma_+ + \gamma_{+-}) = 0. \tag{35}$$

Inverse Laplace transformation of equation (34) gives the exact expression of $p_{++}\psi_{++}(t)$:

$$\begin{aligned} p_{++}\psi_{++}(t) = &\frac{e^{-\omega_1 t}(\gamma_+ - \beta_+\omega_1)}{(\omega_1-\omega_2)(\omega_1-\omega_3)} + \frac{e^{-\omega_2 t}(\gamma_+ - \beta_+\omega_2)}{(\omega_2-\omega_1)(\omega_2-\omega_3)} \\ &+ \frac{e^{-\omega_3 t}(\gamma_+ - \beta_+\omega_3)}{(\omega_3-\omega_1)(\omega_3-\omega_2)}. \end{aligned} \tag{36}$$

Similarly, using the expressions of a_0 and a_+ in Eq. (31), we get

$$\begin{aligned} &p_{+-}\tilde{\psi}_{+-}(s) \\ &= \frac{s^2\omega_r + s(\beta_{+-} + \omega_e\omega_{pf} + \omega_e\omega_r + \omega_{pr}\omega_r) + \gamma_{+-}}{s^3 + \alpha s^2 + s[\beta(0) + \beta_+ + \beta_{+-}] + \gamma_+ + \gamma_{+-}}. \end{aligned} \tag{37}$$

Inverse Laplace transformation gives the exact expression of $p_{+-}\psi_{+-}(t)$,

$$\begin{aligned} p_{+-}\psi_{+-}(t) = &\frac{e^{-\omega_1 t}\left(\gamma_{+-} - c_1\omega_1 + \omega_1^2\omega_r\right)}{(\omega_1-\omega_2)(\omega_1-\omega_3)} \\ &+ \frac{e^{-\omega_2 t}\left(\gamma_{+-} - c_1\omega_2 + \omega_2^2\omega_r\right)}{(\omega_2-\omega_1)(\omega_2-\omega_3)} \\ &+ \frac{e^{-\omega_3 t}\left(\gamma_{+-} - c_1\omega_3 + \omega_3^2\omega_r\right)}{(\omega_3-\omega_1)(\omega_3-\omega_2)}. \end{aligned} \tag{38}$$

Note that by putting $s = 0$ in Eqs. (34) and (37) we get the "branching probabilities"

$$p_{++} = \frac{\gamma_+}{\gamma_+ + \gamma_{+-}}, \tag{39}$$

$$p_{+-} = \frac{\gamma_{+-}}{\gamma_+ + \gamma_{+-}}, \tag{40}$$

which satisfy the normalization condition $p_{++} + p_{+-} = 1$. The cDTDs $\psi_{+-}(t)$ and $\psi_{++}(t)$ are plotted in Figs. 3 and 4, respectively, for a few different values of the parameters ω_r and ω_f. In both the figures, the most probable dwell time increases with decreasing ω_r and decreasing ω_f.

In the same matrix-based formalism, the average velocity of a DNAP is given by the general expression [11]

$$V_p = -i\frac{\dot{\gamma}(0)}{\beta(0)}, \tag{41}$$

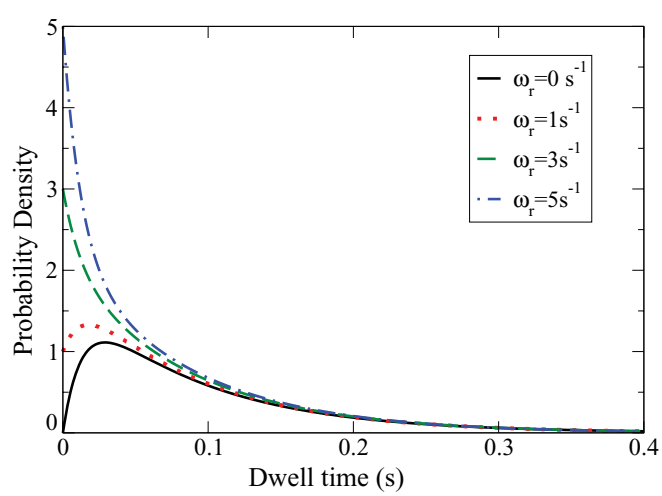

FIG. 3. (Color online) Probability density of conditional dwell time, ψ_{+-}(t), is plotted for a few different values of parameter ω_r. The values of other parameters are (all in s^{-1}) $\omega_f = 20.0$, $\omega_{pf} = 30.0$, $\omega_{pr} = 15.0$, $\omega_h = 40.0$, and $\omega_e = 4.0$.

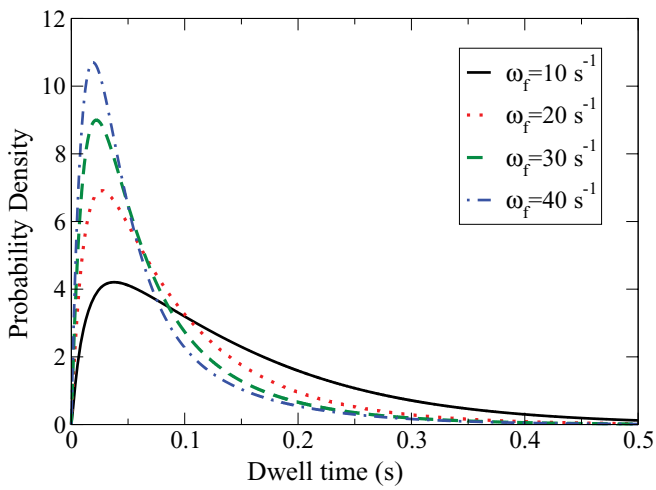

FIG. 4. (Color online) Probability density of conditional dwell time, ψ_{++}(t), is plotted for a few different values of parameter ω_f. The values of other parameters are (all in s^{-1}) $\omega_r = 5.0$, $\omega_{pf} = 30.0$, $\omega_{pr} = 15.0$, $\omega_h = 40.0$, and $\omega_e = 4.0$.

where the dot indicates the derivative with respect to q. The right-hand side of Eq. (41) can be evaluated for our model of DNAP using the explicit expressions (19) and (16) for $\gamma(q)$ and $\beta(0)$, respectively. For the purpose of showing the close relation of V_p with the Michaelis-Menten (MM) equation for the average rates of enzymatic reaction, we now assume that dNTP binding is rate limiting (the general framework of our theory does not need this assumption). Under this assumption, we can write

$$\omega_f = \omega_f^0[\mathrm{dNTP}_c] \text{ and } \Omega_f = \Omega_f^0[\mathrm{dNTP}_w] \tag{42}$$

where $[\mathrm{dNTP}_c]$ and $[\mathrm{dNTP}_w]$ are the concentrations of the correct and incorrect substrates, respectively, and that $\omega_f^0 \gg \Omega_f^0$. In this case, the average velocity of the DNAP, i.e., the average rate of polymerization, can be expressed in a MM-like form,

$$V_p^{(c)} = \frac{\tilde{K}_{\mathrm{cat}}[\mathrm{dNTP}_c]}{\tilde{K}_M + [\mathrm{dNTP}_c]}, \tag{43}$$

for the correct nucleotides, where

$$\tilde{K}_{\mathrm{cat}} = \frac{\omega_h(\omega_e + \omega_{pr})}{\omega_{pr} + \omega_e + \omega_{pf}} \tag{44}$$

and the effective Michaelis constant is

$$\tilde{K}_M = \frac{(\omega_{pr} + \omega_e)(\omega_h + \omega_r) + \omega_{pf}\omega_e}{\omega_f^0(\omega_{pr} + \omega_e + \omega_{pf})}. \tag{45}$$

Replacing ω by Ω and $[\mathrm{dNTP}_c]$ by $[\mathrm{dNTP}_w]$ we get the average rate of polymerization V_p^w for the wrong nucleotides. In the limit of negligible exonuclease activity, the kinetic diagram shown in Fig. 1 reduces to the scheme shown in the left panel of Fig. 2 which is the standard MM scheme with a single intermediate complex; in this limit the expressions for $\tilde{K}_{\mathrm{cat}}$ and $\tilde{K}_M$ are consistent with those for the standard MM scheme [12].

B. Distribution of turnover time for exonuclease

In this section we derive the DTT for *exonuclease* activity of the DNAP. We insert a hypothetical state P_1^* such that

$$P_3 \xrightarrow{\omega_e} P_1^* \xrightarrow{\delta} P_1, \tag{46}$$

where in the limit $\delta \to \infty$, P_1 and P_1^* become identical and we recover our original model.

For the simplicity of notation, in this subsection we drop the site index without loss of any information. The master equations for $P_\mu(t)$ ($\mu = 1,2,3$) and that for $P_1^*(t)$ are

$$\frac{dP_1(t)}{dt} = -\omega_f P_1(t) + (\omega_r + \omega_h)P_2(t), \tag{47}$$

$$\frac{dP_2(t)}{dt} = \omega_f P_1(t) - (\omega_r + \omega_{pf} + \omega_h)P_2(t) + \omega_{pr} P_3(t), \tag{48}$$

$$\frac{dP_3(t)}{dt} = \omega_{pf} P_2(t) - (\omega_e + \omega_{pr})P_3(t), \tag{49}$$

$$\frac{dP_1^*(t)}{dt} = \omega_e P_3(t). \tag{50}$$

For the calculation of DTT, we impose the initial condition $P_1(0) = 1$, and $P_2(0) = P_3(0) = P_1^*(0) = 0$. Suppose that $f(t)$ denotes the DTT. Then, $f(t)\Delta t$ is the probability that one exonuclease cycle is completed between t and $t + \Delta t$, i.e., the DNAP was in state 3 at time t and made a transition to the state $1*$ between t and $t + \Delta$. Obviously, $f(t)\Delta t = \omega_e P_3(t)$ and, hence,

$$f(t) = \omega_e P_3(t). \tag{51}$$

Using a compact matrix notation, Eqs. (47)–(49) can be written the form

$$\frac{d}{dt}\mathbf{Q}(t) = \mathbf{N}\mathbf{Q}(t), \tag{52}$$

where

$$\mathbf{N} = \begin{bmatrix} -\omega_f & \omega_h + \omega_r & 0 \\ \omega_f & -(\omega_h + \omega_r + \omega_{pf}) & \omega_{pr} \\ 0 & \omega_{pf} & -(\omega_e + \omega_{pr}) \end{bmatrix} \tag{53}$$

and

$$\mathbf{Q} = \begin{bmatrix} P_1(t) \\ P_2(t) \\ P_3(t) \end{bmatrix}. \tag{54}$$

Solution of Eq. (52) in Laplace space

$$\tilde{\mathbf{Q}}(s) = \mathbf{S}(s)^{-1}\tilde{\mathbf{Q}}(0), \tag{55}$$

where

$$\mathbf{S}(s) = s\mathbf{I} - \mathbf{N}. \tag{56}$$

Solution (55) for the assumed initial conditions provides

$$\tilde{P}_3(s) = \frac{(-1)^{1+3}\mathcal{C}_{13}}{|\mathbf{S}(s)|} \tag{57}$$

which, explicitly in terms of the rate constants, takes the form

$$\tilde{P}_3(s) = \frac{\omega_f\omega_{pf}}{s^3 + \alpha' s^2 + \beta' s + \gamma'}, \tag{58}$$

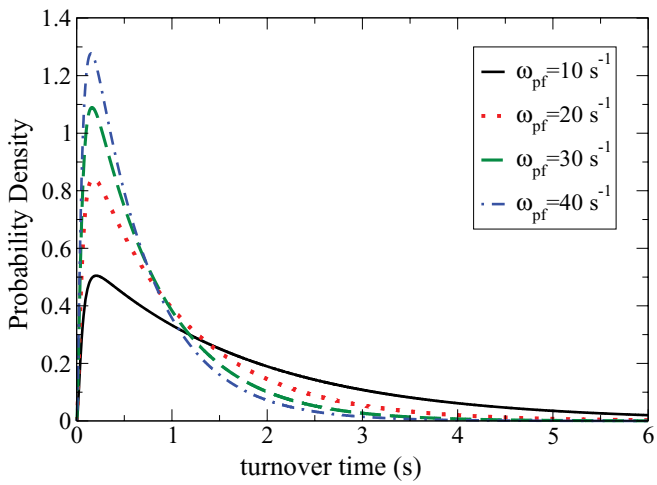

FIG. 5. (Color online) Probability density of turnover time for unproductive exonuclease mode of enzyme; $f(t)$ is plotted for a few different values of parameter ω_{pf}. The values of the other parameters are (all in s^{-1}) $\omega_f = 20.0$, $\omega_r = 5.0$, $\omega_{pr} = 15.0$, $\omega_h = 40.0$, and $\omega_e = 4.0$.

where

$$\alpha' = \omega_e + \omega_f + \omega_h + \omega_{pf} + \omega_{pr} + \omega_r, \tag{59}$$

$$\beta' = \omega_f\omega_{pf} + \omega_e\omega_h + \omega_e\omega_{pf} + \omega_f\omega_{pf} + \omega_f\omega_{pr} + \omega_h\omega_{pr} + \omega_e\omega_{pr} + \omega_{pr\omega_r}, \tag{60}$$

$$\gamma' = \omega_f\omega_{pf}\omega_e. \tag{61}$$

Since in the Laplace space Eq. (51) becomes

$$\tilde{f}(s) = \omega_e \tilde{P}_3(s), \tag{62}$$

we get

$$\tilde{f}(s) = \frac{\omega_e\omega_f\omega_{pf}}{s^3 + \alpha' s^2 + \beta' s + \gamma'} = \frac{\omega_e\omega_f\omega_{pf}}{(s+\upsilon_1)(s+\upsilon_2)(s+\upsilon_3)} \tag{63}$$

as the DTT in the Laplace space.

Taking the inverse Laplace transform of Eq. (63), we get the DTT,

$$f(t) = \left[\frac{\omega_f\omega_{pf}\omega_e}{(\upsilon_1-\upsilon_2)(\upsilon_1-\upsilon_3)}\right]e^{-\upsilon_1 t} + \left[\frac{\omega_f\omega_{pf}\omega_e}{(\upsilon_2-\upsilon_1)(\upsilon_2-\upsilon_3)}\right] \times e^{-\upsilon_2 t} + \left[\frac{\omega_f\omega_{pf}\omega_e}{(\upsilon_3-\upsilon_1)(\upsilon_3-\upsilon_2)}\right]e^{-\upsilon_3 t}, \tag{64}$$

where υ_1,υ_2,υ_3 are solution of the following equation:

$$\upsilon^3 - (\omega_e + \omega_f + \omega_h + \omega_{pf} + \omega_{pr} + \omega_r)\upsilon^2 + (\omega_f\omega_{pf} + \omega_e\omega_h + \omega_e\omega_{pf} + \omega_f\omega_{pf} + \omega_f\omega_{pr} + \omega_h\omega_{pr} + \omega_e\omega_{pr} + \omega_{pr\omega_r})\upsilon - (\omega_f\omega_{pf}\omega_e) = 0. \tag{65}$$

This DTT is plotted in Fig. 5. Plots are consistent with the intuitive expectation that increasing ω_{pf} leads to a decrease of the turnover time.

Suppose $\langle t\rangle$ denotes the mean time gap between the completion of the successive *exonuclease* reactions catalyzed by the DNAP. Then the average rate $V_e = 1/\langle t\rangle$ of the exonuclease reaction can be expressed in a MM-like form [13],

$$V_e^{(c)} = \frac{K_{\rm cat}[{\rm dNTP}_c]}{K_M + [{\rm dNTP}_c]} \text{ with } K_{\rm cat} = \frac{\omega_{pf}\omega_e}{\omega_e + \omega_{pf} + \omega_{pr}}, \tag{66}$$

for the correct nucleotides where $K_M = \tilde{K}_M$. Replacing ω by Ω and $[{\rm dNTP}_c]$ by $[{\rm dNTP}_w]$ in (66) we get $V_e^{(w)}$ for the wrong nucleotides. In the limit $\omega_h \to 0, \omega_{pr} \to 0$, the kinetic diagram shown in Fig. 1 reduces to the simpler scheme shown in the right panel of Fig. 2 which is essentially a generalized MM-like scheme with two intermediate states. Not surprisingly, in this limit, the average rate of the exonuclease reaction is consistent with that of the MM-like scheme with two intermediate states [12].

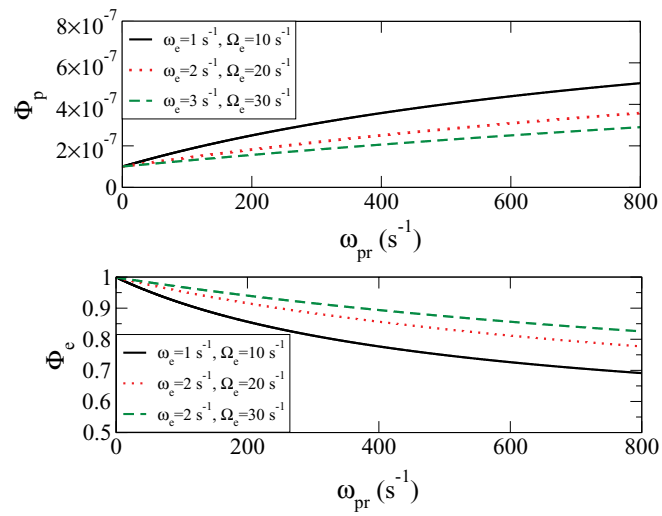

FIG. 6. (Color online) Φ_p and Φ_e plotted against ω_{pr} while the ratio $\Omega_{pr}/\omega_{pr} = 0.1$ is kept fixed. The three curves correspond to $\omega_e = 1.0, 2.0, 3.0\ s^{-1}$. The values of the other parameters are (all in s^{-1}): $\omega_f = 1.0$, $\Omega_f = 10^{-5}$, $\omega_{pf} = 0.1$, $\Omega_{pf} = 10.0$, $\omega_r = .1$, $\Omega_r = 1.0$, $\omega_h = 10.0$, $\Omega_h = 0.1$.

C. Quantitative measures of error

Note that $\Phi_p = \tilde{V}_p^{(w)}/(\tilde{V}_p^{(w)} + \tilde{V}_p^{(c)})$ is the fraction of nucleotides *misincorporated* in the final product of replication. Similarly, the fraction $\Phi_e = V_e^c/(V_e^c + V_e^w)$ is a measure of the *errorneous severings*, i.e., fraction of the cleaved nucleotides that were incorporated correctly into the growing DNA. Since ω_{pr} is the strength of the "coupling" between the two different enzymatic activities, we plot Φ_p and Φ_e against ω_{pr} in Fig. 6 for a few typical sets of values of the model parameters.

Decreasing Φ_e with increasing ω_{pr} is a consequence of the escape route via ω_{pr} for the correctly incorporated nucleotides that get transferred unnecessarily to the exonuclease site. It is the increasing number of such correctly incorporated nucleotides recused from the exonuclease site that leads to the raising of Φ_p with increasing ω_{pr}. The limiting values of Φ_p and Φ_e in the limit of large ω_{pr} are determined by the corresponding limiting expressions $\tilde{V}_p^{(c)} \simeq (\omega_f\omega_h)/(\omega_h + \omega_f + \omega_r)$, and $V_e^{(c)} \simeq (\omega_f\omega_{pf}\omega_e)/[\omega_{pr}(\omega_h + \omega_f + \omega_r)]$. (Expressions for $\tilde{V}_p^{(w)}$ and $V_e^{(w)}$ are similar in the limit of large Ω_{pr}.)

IV. SUMMARY AND CONCLUSION

Here we have theoretically investigated the effects of the coupling of two different modes of enzymatic activities of a DNAP; in one of these it elongates a DNA whereas in the other it shortens the same DNA. We have addressed fundamental questions here in the context of DNA replication [14]. The effects of tension on the polymerase and exonuclease activities, which have been the main focus of the earlier works [14], will be reported elsewhere [15]. The mechanism of error correction by DNAP is somewhat different from the mechanism of transcriptional proofreading which is intimately coupled to "back tracking" of the RNA polymerase [16,17].

We have derived exact analytical formulas for the cDTD and DTT which will be very useful in analyzing experimental data in single DNAP biophysics, particularly its stepping patterns and enzymatic turnover. In spite of their coupling, the average rates of both the enzymatic activities are MM like; the analytical expressions for the effective MM parameters explicitly display the nature of the coupling of the two kinetic processes. We have also reported exact analytical expressions for the fractions Φ_p and Φ_e which measure *replication error* and *erroneous cleavage*; these expressions can be used for analyzing data from both single molecule [5] and bulk [18] experiments on wild-type and mutant DNAPs.

ACKNOWLEDGMENTS

We thank S. Klumpp for useful comments. This work has been supported at IIT Kanpur by the Dr. Jag Mohan Garg Chair Professorship (D.C.) and a CSIR Fellowship (A.K.S.). This research (D.C.) has been supported in part also by the Mathematical Biosciences Institute at the Ohio State University and the National Science Foundation under Grant No. DMS 0931642.

[1] A. Kornberg and T. Baker, *DNA Replication* (Freeman, San Francisco, 1992).
[2] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer Associates, Sunderland, MA, 2001).
[3] A. B. Kolomeisky and M. E. Fisher, Annu. Rev. Phys. Chem. **58**, 675 (2007).
[4] L. J. Reha-Krantz, Biochimica et Biophysica Acta **1804**, 1049 (2010).
[5] B. Ibarra, Y. R. Chemla, S. Plyasunov, S. B. Smith, J. M. Lazaro, M. Salas, and C. Bustamante, EMBO J. **28**, 2794 (2009).
[6] P. Xie, J. Theor. Biol. **259**, 434 (2009).
[7] D. J. Galas and E. W. Branscomb, J. Mol. Biol. **124**, 653 (1978).
[8] A. R. Fersht, J. W. K. Jones, and W. C. Tsui, J. Mol. Biol. **156**, 37 (1982).
[9] J. J. Schawrtz and S. R. Quake, Proc. Natl. Acad. Sci. USA **106**, 20294 (2009).
[10] A. Garai and D. Chowdhury, Europhys. Lett. **93**, 58004 (2011).
[11] Y. R. Chemala, J. R. Moffitt, and C. Bustamante, J. Phys. Chem. B **112**, 6025 (2008).
[12] M. Dixon and E. C. Webb, *Enzymes* (Academic, New York, 1979).
[13] S. C. Kou, B. J. Cherayil, W. Min, B. P. English, and X. S. Xie, J. Phys. Chem. B **109**, 19068 (2005).
[14] A. Goel, R. D. Astumian, and D. Herschbach, Proc. Natl. Acad. Sci. USA **100**, 9699 (2003).
[15] A. K. Sharma and D. Chowdhury (unpublished).
[16] M. Voliotis, N. Cohen, C. Molina-Paris, and T. B. Liverpool, Phys. Rev. Lett. **102**, 258101 (2009).
[17] M. Sahoo and S. Klumpp, Europhys. Lett. **96**, 60004 (2011).
[18] I. Wong, S. S. Patel, and K. A. Johnson, Biochem. **30**, 526 (1991).