Title

Efficient fidelity control by stepwise nucleotide selection in polymerase elongation Abstract: Polymerases select nucleotides

Permalink

https://escholarship.org/uc/item/98h7x1st

Journal

Computational and Mathematical Biophysics, 2(1)

ISSN

2544-7297

Author

Yu, Jin

Publication Date

2014

DOI

10.2478/mlbmb-2014-0010

Peer reviewed

Research Article

Open Access

Jin Yu*

Efficient fidelity control by stepwise nucleotide selection in polymerase elongation

Abstract: Polymerases select nucleotides according to a template before incorporating them for chemical synthesis during gene replication or transcription. Efficient selection to achieve sufficiently high fidelity and speed is essential for polymerase function. Due to multiple kinetic steps detected in a polymerase elongation cycle, there exist multiple selection checkpoints to allow different strategies of fidelity control. In our current work, we examined step-by-step selections in an elongation cycle that have conformational transition rates tuned one at a time, with a controlled differentiation free energy between the right and wrong nucleotides at each checkpoint. The elongation is sustained at non-equilibrium steady state with constant free energy input and heat dissipation. It is found that a selection checkpoint in the later stage of a reaction path has less capability for error reduction. Hence, early selection is essential to achieve an efficient fidelity control. In particular, for an intermediate state, the selection through the forward transition inhibition has the same capacity for error reduction as the selection through the backward rejection. As with respect to the elongation speed, an initial screening is indispensible for maintaining high speed, as the wrong nucleotides can be removed quickly and replaced by the right ones at the entry. Overall, the elongation error rate can be repeatedly reduced through multiple selection checkpoints. This study provides a theoretical framework to guide more detailed structural dynamics studies, and to support rational redesign of related enzymes and devices.

Keywords: Polymerase elongation; Nucleotide selection; Fidelity; Non-equilibrium steady state

1 Introduction

Polymerases are essential enzymes responsible for gene replication or transcription [1]. A polymerase moves along DNA or RNA while synthesizing a new strand of nucleic acid, largely according to Watson-Crick base pairing with a template strand. Without the polymerase, the template-based polymerization can also happen, but at an extremely low speed and with a low fidelity. A polymerase not only catalyzes the polymerization reaction but it also increases the fidelity. While the first function is expected from any inert enzyme, the second function suggests a role of a non-equilibrium chemistry. Furthermore, being a nanometer-sized molecular machine, the polymerase works under high viscosity and significant thermal noise. The achievement of high fidelity at a sufficiently high elongation speed is thus key to the biological functions of polymerase.

Previous experimental measurements suggest that the polymerase moves as a Brownian ratchet along the nucleic acid track [2–7]. Upon binding and insertion of the incoming nucleotide, backward translocation of the polymerase is inhibited. The nucleotide insertion often accompanies substantial conformational changes of the polymerase [8, 9]. Following the insertion, the nucleotide covalently links to the newly synthesized chain through phosphoryl transfer reaction (see Fig. 1a). The catalytic reaction is followed by pyrophosphate ion (PPi) release, which concludes the enzymatic cycle, and the polymerase translocates forward and recruits the next nucleotide. From the recruitment to the end of the catalysis, the nucleotide can be selected at multiple kinetic checkpoints. Here one assumes that both a cognate (*right*) and a non-cognate (*wrong*) nucleotide

*Corresponding Author: Jin Yu: Beijing Computational Science Research Center, No. 3 He-Qing Road, Hai-Dian District, Beijing, P. R. China, 100084, E-mail: jinyu@csrc.ac.cn; tel.:+86-10-82687087

follow the same kinetic path of incorporation. Hence, a forward or backward transition that is biased against the wrong nucleotide incorporation is defined as a 'checkpoint'.

Indeed, a wrong nucleotide can be selected against by the polymerase, either via enhancing the backward transition rate (*rejection*), or via lowering the forward transition rate (*inhibition*) along the reaction path (see Fig. 1b), comparing to the incorporation of a right nucleotide. Such nucleotide selections are common to fidelity control of all polymerases. The error rate achieved by the selection can reach as low as one in tens of thousands to one in a million ($10^{-4} \sim 10^{-6}$) [10]. After the catalysis, or once the nucleotide is covalently added, the error can be further corrected through an exo- or endo-nuclease reaction, which excises wrong nucleotides and serves for proofreading. In general, the fidelity of the polymerization is controlled through both the selection and proofreading [8, 9, 11–14]. The error rate can be lowered further by one to three orders of magnitudes through the proofreading, to as low as one in tens of millions or even lower [10].
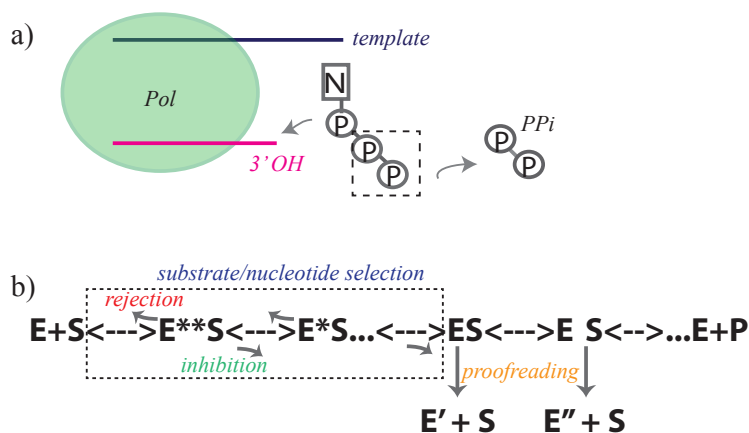
Figure 1: Schematics of polymerase reaction and stepwise nucleotide selection. **(a)** Polymerase (*Pol*) enzyme (**E**) catalyzes a phosphoryl transfer reaction E*NA$_n$ + NTP $\leftrightarrow$ E*NA$_{n+1}$ + PPi. The incoming NTP is incorporated according to the template strand. **(b)** The kinetic scheme for incorporating the nucleotide substrate during polymerase elongation. Nucleotide or substrate selection can happen at any checkpoint prior to chemical catalysis that leads to formation of **ES**, through backward *rejection* or forward *inhibition* (see text). Proofreading happens after the formation of **ES**, and before the final product formation (**E**+**P**).

The kinetic proofreading had been widely discussed in the context of genetic control [15–19]. In order to achieve high specificity or fidelity, the enzyme and substrate can form multiple intermediate states before generating the final product. The intermediate states are made through driven reactions (breaking the detailed balance) with free energy sources. Each kinetic proofreading procedure is implemented through a branching or looping reaction that breaks one of the intermediates back into the *apo* enzyme and substrate (also see Fig. 1b). The free energy for differentiation between the right and wrong substrates at individual steps or checkpoints can thus be accumulated through the proofreading procedure to achieve high fidelity. The proofreading related activities of the polymerases have been detected at the single molecule level in recent years [3, 20, 21], which inspired further modeling studies [22–25].

The proofreading-free selection, however, has attracted less attention as it appears simple. Early work suggested that the error ratio of the selection (i.e., the ratio between the wrong and right substrate incorporation) cannot be lower than $\exp(-\Omega_{max}/k_BT)$, which is determined by a maximum free energy differentiation Ω_{max} between the wrong and right substrate along the reaction path [17]. Individual steps of the selection had not been further considered. The overall characterization seemed to suffice as the selection details were not experimentally detectable. Recently, parameters for the stepwise kinetics of transcription has become accessible, for example, from the structure-based mutagenesis and kinetic analysis [26]. In this study, the

mechanism of fidelity is examined, and contributions to the overall fidelity from the individual selection steps have been characterized. At the same time, modeling and computation technologies allow protein structural dynamics to be captured at the atomic scale, providing opportunities to characterize detailed selection mechanisms. Motivated by these two recent developments, here we present a model framework on the stepwise nucleotide selection during the polymerase elongation. The selection relies on multiple kinetic intermediate states prior to the end of the catalysis. Either the backward transition to the previous state is enhanced, when the intermediate structure is bound with the wrong substrate, or the transition toward the next state is inhibited. Each modulated transition constitutes an "elementary selection" addressed below. The elementary selection happens between two consecutive states along the reaction path (see Fig. 1b), *without* branching back to the *apo* state as that in the proofreading. It is important to point out that in order to be able to actually achieve the multi-step selection, the overall reaction has to be sustained at a non-equilibrium steady state (NESS), which is indeed the case as long as the polymerase elongates at a nonzero speed [27, 28].

The key question we want to address in this study is how a stepwise selection is achieved *efficiently* during the polymerase elongation. Being 'efficient' is loosely defined here as achieving a sufficiently low error rate at a relatively high speed, when the free energy differentiation is limited and controlled. With *given* kinetic parameters for incorporating the *right* substrates, and Ω_{max} as a control parameter for the maximum free energy of differentiation, we want to find relatively 'efficient' selection strategies: to find selection parameter sets for the *wrong* substrates that lower the elongation error rate without necessarily lowering the speed much, in comparison with polymerization without selection. Indeed, one can dissect Ω_{max} into individual terms $\{\Delta_i^{\ddagger}\}$ at the multiple selection checkpoints, with $\Sigma\Delta_i^{\ddagger} = \Omega_{max}$ (i starts from m, the index of the substrate binding state, to $m + n$-1, with $m \in [1, M]$ for M kinetic states of a cycle, and $2n(n \geq 1)$ selection checkpoints exist prior to the end of the catalysis). One can see in this work how the elongation speed and error rate vary as $\{\Delta_i^{\ddagger}\}$ are allocated differently among the checkpoints along the reaction path.

In earlier work, an 'efficiency-accuracy' tradeoff was discovered in substrate selection [19, 29, 30]. The tradeoff means that a selection system operates close to its maximal accuracy when the enzyme efficiency approaches zero (the definitions of 'accuracy' and 'efficiency' are addressed later). The maximal accuracy is determined by $\exp(\Omega_{max}/k_BT)$, and both the accuracy and enzyme efficiency vary depending on the kinetic rates of the system. The tradeoff shows a limit of the selection as the system kinetics parameters varies under experimentally designed conditions, and helps to extract kinetic information of the system [19]. In the current study, however, we consider how the error rate and speed (related to but not the same as the 'accuracy' and 'efficiency') vary among different selection strategies, without varying kinetics for the right substrate incorporation.

In this work, we adopted polymerase elongation schemes in general, while using data from a prototypical single-subunit RNA polymerase (RNAP) from bacteriophage T7 [4, 7, 31] to demonstrate numerical results. T7 RNAP elongates at an error rate $\sim 10^{-4}$ *without* proofreading activities detected [32]. It is an ideal system to study the nucleotide selection. We analyze first a generic three-state kinetic scheme, building connections with previous quantitative studies. The basic findings are re-examined in a more specific elongation scheme with five states. The kinetic schemes apply to most polymerases, though different rate-limiting steps happen in different cases. Accordingly, how the findings vary as the rate-limiting step varies is also addressed. In addition, implementations of the present framework to two typical polymerases (both the RNA and DNA polymerases from T7) are introduced.

2 Methods

For the template-based nucleotide incorporation, we note that the polymerases recognize the nucleotides either as cognate/right (*r*) or non-cognate/wrong (*w*). Below we use free energy profiles $G^{r/w}(x)$ (along the reaction coordinate x; with '*w*' labeling for the wrong and '*r*' labeling for the right) for incorporating both the right and wrong nucleotides to characterize the stepwise selection (see Fig. 2). Right before the nucleotide binds at state m (local free energy minimum with reaction coordinate x_m), we set $G^w(x_{m-1}) = G^r(x_{m-1})$ at the

previous state $m - 1$ (with coordinate x_{m-1}; when $m = 1$, m-1 is cyclically reset to M). The continuous free energy differentiation is then defined along the free energy profile as $\Omega(x) \equiv G^w(x) - G^r(x)$ for $x > x_{m-1}$.

For each transition checkpoint starting from the state i, we label it by index i($m \leq i \leq m + n - 1$, with $m + n$ the index of the product state right after catalysis; when $m + n = M+1$, it is cyclically reset to 1), plus a sign of $-/+$ for the rejection /inhibition. Correspondingly, one can define an *elementary selection* S_i^- or S_i^+ as a rejection or inhibition. If we label the free energy barrier or maximum in between state i and $i+1$ as $i^\star$, then we can define the *selection differentiation free energy* at checkpoint i as Δ_i^- or Δ_i^+, as the transition barrier difference between the right and wrong for the backward or forward transition:

$$\Delta_i^- = [G^r(x_{i-1\star}) - G^r(x_i)] - [G^w(x_{i-1\star}) - G^w(x_i)] = [G^w(x_i) - G^r(x_i)] - [G^w(x_{i-1\star}) - G^r(x_{i-1\star})]$$
$$= \Omega(x_i) - \Omega(x_{i-1\star}) = \Omega_i - \Omega_{i-1\star} \tag{1a}$$

$$\Delta_i^+ = [G^w(x_{i\star}) - G^w(x_i)] - [G^r(x_{i\star}) - G^r(x_i)] = [G^w(x_{i\star}) - G^r(x_{i\star})] - [G^w(x_i) - G^r(x_i)]$$
$$= \Omega(x_{i\star}) - \Omega(x_i) = \Omega_{i\star} - \Omega_i \tag{1b}$$

The schematic illustrations of the above definitions are shown in Fig. 2. One can see that S_i^- selects against the wrong substrates by lowering the backward transition barrier from i to $i-1$ by $\Delta_i^- (>0)$ in comparison with that of the right substrate. One can also characterize the selection strength as the ratio between the backward rates (k_{i-}) of the wrong and right: $\eta_i^- = k_{i-}^w / k_{i-}^r = e^{\Delta_i^-/k_B T}(>1)$. A strong selection against the wrong substrate corresponds to a large η ($\gg 1$). The selection denoted S_i^+, however, inhibits the wrong nucleotides from transitioning from state i to $i+1$, by raising the forward free energy barrier by $\Delta_i^+ (>0)$. Correspondingly, $\eta_i^+ = k_{i+}^r / k_{i+}^w = e^{\Delta_i^+/k_B T}$. Following S_i^+ and similar to S_i^-, the selection denoted S_{i+1}^- (with a strength η_{i+1}^-) selects against wrong nucleotides by reducing the backward transition barrier from $i+1$ to i by Δ_{i+1}^-. In order to compare different elementary selections, we set the selection differentiation free energy for one elementary selection as a constant $\Delta > 0$ (or a control parameter) and its strength $\eta = e^{\Delta/k_B T}$; at the same time, we turn off the other selections to $\eta = 1$ when one elementary selection applies.
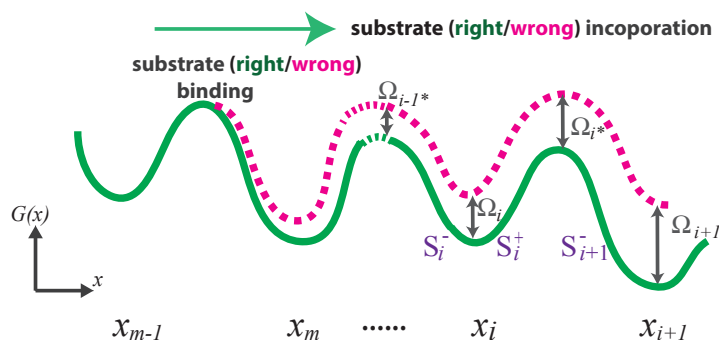
Figure 2: The schematics of selection free energy profiles along the reaction path. The free energy profile $G^{r/w}(x)$ for the right/wrong substrate incorporation is depicted in green/red solid/dashed line. Free energy minima (stable states) and maxima (intermediate states) are labeled at the corresponding reaction coordinates. The continuous free energy differentiation $\Omega(x) = G^w(x) - G^r(x)$ is indicated at those free energy minima and maxima. Substrate binding is achieved at state m, from where the selection can start. Exemplary selections S_i^-, S_i^+, and S_{i+1}^- are also indicated.

By defining the elementary selections along the reaction path, one can construct any substrate selection strategy as a combination of those elementary selections. For example, in Fig. 2, the overall selection can be regarded as a combination of elementary selections S_i^- with $\Delta_i^- = \Omega_i - \Omega_{i-1\star}$, S_i^+ with $\Delta_i^+ = \Omega_{i\star} - \Omega_i$, and S_{i+1}^- with $\Delta_{i+1}^- = \Omega_{i+1} - \Omega_{i\star}$. The continuous free energy differentiation $\Omega(x)$ indeed measures the *accumulative* selection differentiation free energies from individual checkpoints along the reaction path: for example, when $i = m$, $\Omega_{i-1\star} = 0$, $\Omega_{i\star} = \Delta_i^- + \Delta_i^+$, and $\Omega_{i+1} = \Delta_i^- + \Delta_i^+ + \Delta_{i+1}^-$. For an efficient selection, $\Delta_i^+ \geq 0$, and the

continuous free energy differentiation $\Omega(x)$ grows along the reaction path. Accordingly, the maximum free energy differentiation $\Omega_{\max} \equiv Max\Omega(x_i) = \sum \Delta_i^{\ddagger}$ is actually the accumulative free energy differentiation for all the elementary selections (as $\Delta_i^{\ddagger} \geq 0$).

Below, we focus on how an elementary selection impacts the polymerization/elongation speed and the error rate. Any selection in general applies as the elementary selections work together. Independent of enzymatic activities, the free energy input (>0) upon incorporating a right nucleotide and a wrong one is denoted ΔG_c^r and ΔG_c^w, respectively. The overall difference is $\Delta G_c^r - \Delta G_c^w = \delta_G$ (≥ 0), which reflects an equilibrium difference in the overall affinity, and thus determines the equilibrium accuracy. The enzyme activities modulate the intermediate stabilities or barriers along the reaction path, and drive the polymerization process far from equilibrium. Kinetically, the right substrate incorporation is greatly accelerated, while the wrong substrate incorporation is comparatively deterred. As a result, the fidelity rises largely above the equilibrium level, while the equilibrium free energy constraints ΔG_c^r, ΔG_c^w and δ_G remain unchanged. In order to keep a constant $\Delta G_c^r - \Delta G_c^w = \delta_G$, a 'reset' of the overall free energy difference to δ_G (>0) after each elementary selection is implemented at the end of the catalysis (see illustrations later).

3 Results

3.1 Three-state elongation scheme

We start with a generic three-state kinetic scheme (Fig. 3a) to compare two basic nucleotide selection strategies in the elongation cycle, the *rejection* and the *inhibition*. The scheme consists of the pre-translocated (I), post-translocated (II), and substrate state (III). Upon translocation ($I \rightarrow II$), an incoming NTP diffuses and binds to the polymerase ($II \rightarrow III$), prior to being recognized as right or wrong. A catalytic step then follows ($III \rightarrow I$). Correspondingly, recognition and selection of the nucleotide happen (from $m = III$) either upon substrate binding and unbinding ($III \rightarrow II$) through the rejection, or at the catalytic stage ($III \rightarrow I$) through the inhibition ($n = 1$, with a total of 2 selection checkpoints). Fig. 3b illustrates schematically the selection strategies on the free energy profile, with solid and dashed lines for incorporating the right and wrong species, respectively.

Under the initial rejection S_{III}^-, the unbinding or off-rate of the wrong nucleotide (k_{III-}^w) becomes larger than that of the right nucleotide (k_{III-}^r). One can quantify the selection strength as $\eta_{III}^- = k_{III-}^w / k_{III-}^r = e^{\Delta/k_B T}$, with Δ (>0) measuring the selection free energy differentiation between the wrong and right species detected at this checkpoint ($III \rightarrow II$).

Alternatively, the catalytic inhibition S_{III}^+ raises the activation barrier for the catalysis ($III \rightarrow I$) of the wrong nucleotide above that of the right. One can quantify the selection strength by $\eta_{III}^+ = k_{III+}^r / k_{III+}^w = e^{\Delta/k_B T}$.

To consider reaction fluxes for both the right and wrong species in the three-state scheme, one can define a population vector $\Pi = (P_I, P_{II}, P_{III}^r, P_{III}^w)^T$ to represent the probability distributions of states I, II and III (for both right and wrong species). The master equation for the distributions is:

$$\frac{d}{dt}\Pi = M\Pi \tag{2}$$

where M is a 4×4 transition matrix defined as

$$\left\{ \begin{matrix} -(1-Err)k_{I-} - Err \cdot k_{I-}\frac{\eta_G}{\eta_{III}^-\eta_{III}^+} - k_{I+} & k_{II-} & K_{III+} & \frac{k_{III+}}{\eta_{III}^+} \\ k_{I+} & -k_{II-} - k_{II+} & k_{III-} & k_{III-}\eta_{III}^- \\ (1-Err)k_{I-} & i_r k_{II+} & -k_{III-} - k_{III+} & 0 \\ Err \cdot k_{I-}\frac{\eta_G}{\eta_{III}^-\eta_{III}^+} & (1-i_r)k_{II+} & 0 & -k_{III-}\eta_{III}^- - \frac{k_{III+}}{\eta_{III}^+} \end{matrix} \right\} \tag{3}$$

with k_{I+} and k_{II-} the forward and backward translocation rate, k_{II+} and k_{III-} the binding ($\propto [NTP]$) and unbinding rate of the nucleotide, and k_{III+} and k_{I-} the catalytic and its reverse rate. i_r is the portion of right nucleotides from solution at 'input' ($i_r = 1/4$ by default for four equally mixed nucleotides in solution). *Err*
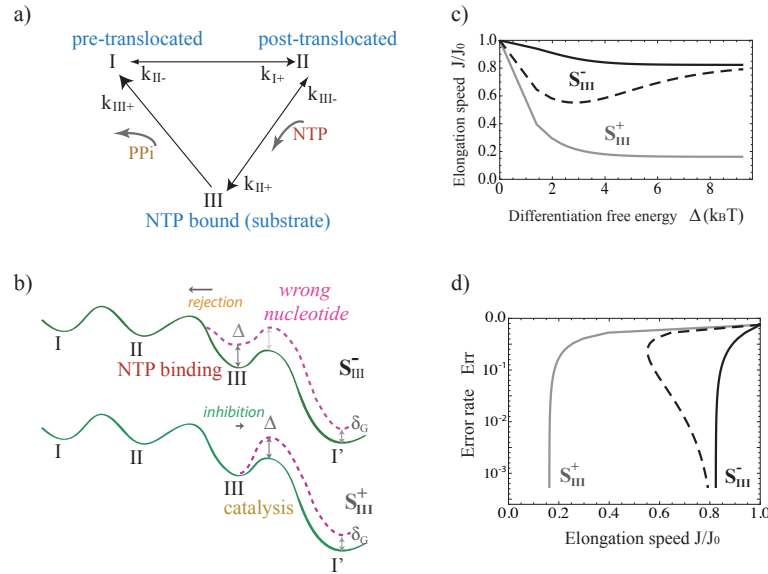
Figure 3: Nucleotide selections in the three-state elongation scheme. (a) The three-state scheme consists of translocation, NTP binding, and catalysis / PPi release. (b) Elementary selections demonstrated on the free energy profile (incorporating the right/wrong nucleotide in solid /dashed line). The initial screening or rejection is through S_{III}^-, while S_{III}^+ selects through catalytic inhibition. $\Delta G_c^r - \Delta G_c^w = \delta_G$ is set in the end. (c) The elongation rate or speed vs. the activation differentiation free energy Δ for respective selections S_{III}^- (dark line) and S_{III}^+ (gray). The elongation speed is normalized as J/J_0, with J_0 the elongation rate without the selection. A combined selection strategy with S_{III}^- and S_{III}^+ at equal strength $\eta = e^{\Delta/2k_BT}$ is also shown (dashed line). (d) The error rate *Err* vs. the elongation speed under S_{III}^-, S_{III}^+, and the combined strategy, as Δ varies.

is the 'output' or elongation error rate at the end of the cycle, after nucleotide selection. $\eta_G \equiv e^{\delta_G/k_BT}$ is to keep the overall free energy difference between the right and wrong nucleotide incorporation to δ_G ($\eta_G = 10$ is used by default).

Using the steady state solution for Eq. 2, at $k_{I-} \to 0$ for simplicity, one obtains the probability flux or the elongation rate J (or the speed $v = l_0 J$ with $l_0 = 1$ bp),

$$K = \frac{\frac{k_{max}^0}{1+\Gamma}[NTP]_{total}}{\frac{\Lambda}{1+\Gamma}K_M^0 + [NTP]_{total}} \tag{4}$$

where k_{max}^0 and K_M^0 are the maximal rate and Michaelis constant when there is *no* nucleotide selection ($\eta_{III}^- = \eta_{III}^+ = 1$). The modulation constants Γ (>0) and $\Lambda(1 < \Lambda_{<1/ir})$ depend on the selection strength η_{III}^- and η_{III}^+. The expressions of the constants can be found in Appendix I.

3.2 Speed modulation by the selection

When there is no selection, $J = J_0 \equiv \frac{k_{max}^0[NTP]_{total}}{K_M^0+[NTP]_{total}}$, ($\Gamma \sim 0$ and $\Lambda = 1$). Below we show how the selections affect the polymerization rate/speed for three typical cases.

– When *only* the initial rejection S_{III}^- works ($\eta_{III}^- > 1$, $\eta_{III}^+ = 1$). In this case, $J \to J_1 \equiv \frac{k_{max}^0[NTP]_{right}}{K_M^0+[NTP]_{right}}$ with $[NTP]_{right} = i_r[NTP]_{total}$. In other words, when strong selection applies through NTP unbinding, the polymerization rate or speed converges to that for a single species of right nucleotides. Usually $J_1 < J_0$ as $[NTP]_{right} < [NTP]_{total}$. When the NTP concentration is sufficiently high, $J_1 \to k_{max}^0$ and $J_1 \sim J_0$, so that the speed remains high.

- When *only* the catalytic inhibition S_{III}^+ works ($\eta_{III}^- = 1$, $\eta_{III}^+ > 1$)). As the inhibition strength increases, $J \to \frac{k_{max}^0 [NTP]_{right}}{K_M^0 + (1+\Gamma)[NTP]_{right}} < J_1$. At a high NTP concentration, $J \to k_{max}^0/(1+\Gamma)$, the rate is significantly reduced below J_0 as long as the translocation is not rate limiting (see Appendix I).

  In Fig. 3c, we show the relative speed (J/J_0) vs. the selection free energy differentiation (Δ) under selections S_{III}^- and S_{III}^+, respectively. By default, translocation is set fast (e.g. over thousands of times back and forth per second) as it is reported or commonly assumed [3, 4]. Accordingly, one sees that the strong nucleotide selection through the catalytic inhibition (S_{III}^+) lowers the polymerization rate significantly, while S_{III}^- does not.

- When *both* selections S_{III}^- and S_{III}^+ work ($\eta_{III}^- > 1$ and $\eta_{III}^+ > 1$). As both selections become strong, $J \to \frac{k_{max}^0 [NTP]_{right}}{K_M^0 + [NTP]_{right}} = J_1$. That says, under a *combined* selection strategy, the elongation rate can still remain high, approaching to that under S_{III}^- alone. In Fig. 3c, the combined selection strategy relies equally on S_{III}^- and S_{III}^+, i.e., with the same differentiation free energy $\Delta/2$ for each. One sees that even a small free energy differentiation ($1 \sim 2\,k_B T$) at the initial screening can keep the relative speed (J/J_0) above 0.5. Indeed, the higher the contribution from S_{III}^- in the combined strategy, the faster the polymerization rate converges to J_1.

3.3 Error rate reduction by the selection

Based on the above calculations, one obtains the output error rate *Err*, as the polymerization flux of the wrong over that of the total (right and wrong):

$$E_{rr} \equiv \frac{J^w}{J} = 1 - i_r \Lambda = \frac{1 - i_r}{1 + i_r \frac{k_{III-}}{k_{III+} + k_{III-}}(\eta_{III}^- \eta_{III}^+ - 1)} \tag{5}$$

From Eq. 5 above, one sees $E_{rr} \sim \frac{1-i_r}{i_r}(1 + \kappa)e^{-2\Delta/k_B T}$ at $\eta_{III}^- = \eta_{III}^+ = e^{\Delta/k_B T} \gg 1$, with $\kappa \equiv k_{III+}/k_{III-}$. In other words, the error rate deceases exponentially with the *accumulate* activation differentiation free energy as the selection gets strong. Interestingly, selections S_{III}^- and S_{III}^- perform *equally* well on the error reduction: An equal value of the individual selection strength η_{III}^- or η_{III}^+ gives an equal error rate *Err*. The error rate is independent of the overall free energy input as the elongation considered is under the strong non-equilibrium limit [27, 28].

To connect with conventions that use enzyme *efficiency* k_{max}/K_M, we obtain an overall efficiency under selection as $\zeta = \zeta_0/\Lambda (\zeta_0 \equiv k_{max}^0/K_M)$, and the efficiencies for the right and wrong nucleotides as $\zeta^r = \zeta_0$ and $\zeta^w = \frac{1 - i_r \Lambda}{(1 - i_r)\Lambda}\zeta_0$, respectively (see Appendix I). In other words, the efficiency for incorporating the right nucleotide is fixed in the current study, while the efficiency for incorporating the wrong approaches zero as the nucleotide selection becomes strong ($\Lambda \to 1/i_r$). Accordingly, the *accuracy A*, defined as the efficiency of incorporating the right over that of the wrong, becomes $A = \frac{(1-Err)/Err}{i_r/(1-i_r)}$. Hence, *A* is to quantify the proportion of right nucleotides at *output* relative to that at *input*. In contrast, the error rate *Err* only counts the percentage of wrong nucleotides at output. In Appendix I, we also derived the efficiency-accuracy tradeoff ($\zeta^r \propto \frac{e^{\Omega_{max}/k_B T} - A}{e^{\Omega_{max}/k_B T} - 1}$).

In Fig. 3d, we show the error rate *Err* vs. speed J/J_0 for the respective selections S_{III}^- and S_{III}^+, and for the combined selection, as the selection grows strong (Δ increases). Basically, one sees that by increasing the selection strength, the error rates can be continuously lowered, while the elongation speeds/rates converge to different values and do not change further upon the strong selection. At high NTP concentration, S_{III}^- keeps the converged speed high, S_{III}^+ lowers the speed significantly, while the combined selection leads to the speed as that under S_{III}^- alone. In terms of the error reduction, S_{III}^- and S_{III}^+ (and the combined) perform equally well.

Experimentally, the individual selection strength can be quantified. When *purely wrong* nucleotides are supplied in the solution ($i_r = 0$), $J \to \frac{\frac{k_{max}^0}{\eta_{III}^+}[NTP]_{right}}{\eta_{III}^- K_M^0 + [NTP]_{right}}$. By measuring how k_{max} and K_M are modulated comparing assays of purely wrong and purely right nucleotides, one is able to determine η_{III}^- and η_{III}^+ in the three-state elongation scheme.

4 Five-state elongation scheme

Next, we use a slightly more specific scheme with five states, as that from T7 RNAP [7, 31], to describe the polymerase elongation cycle (see Fig. 4a and 4b). Comparing to the three-state scheme, the essential difference is that there are two kinetic steps instead of just one ($II \rightarrow III$ and $III \rightarrow IV$) proceeding to the chemical catalysis. In T7 RNAP and some other polymerases, the two steps are regarded as nucleotide pre-insertion and insertion [33, 34], respectively. In particular, the nucleotide insertion likely happens slowly [8, 9, 31], so we take it as a rate-limiting step by default. Variation of the rate-limiting step in the scheme will be addressed later.
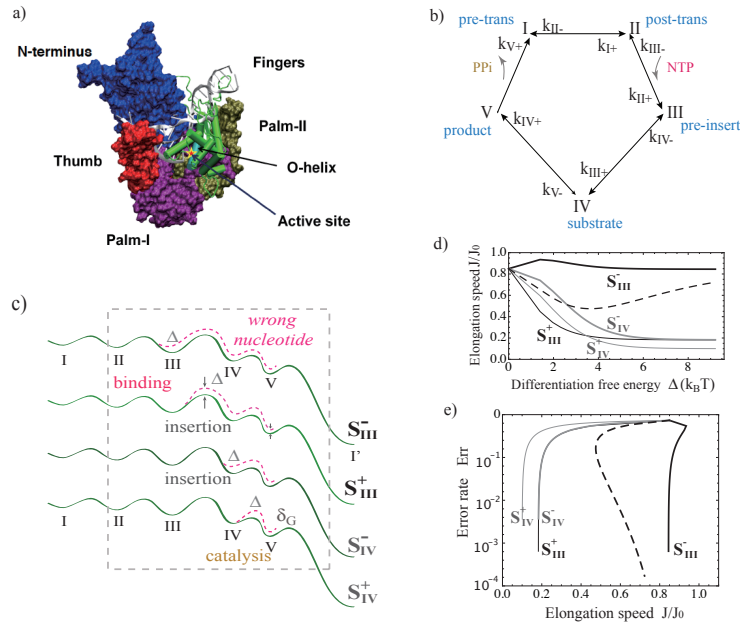
Figure 4: Nucleotide selections in the five-state elongation scheme. (a) A molecular view of T7 RNAP structure. The single subunit polymerase looks similar to a hand grabbing onto the DNA. (b) The five-state scheme consists of translocation, nucleotide pre-insertion (binding), insertion, catalysis, and PPi release. (c) Four elementary selections demonstrated on the free energy profile (for right / wrong nucleotides in solid / dashed line). S_{III}^- applies an immediate rejection on the wrong nucleotides. S_{III}^+ inhibits the insertion of the wrong nucleotides. S_{IV}^- rejects the wrong nucleotides upon insertion, while S_{IV}^+ inhibits the catalysis of the wrong. $\Delta G_c^r - \Delta G_c^w = \delta_G$ is set at the end. (d) The elongation speed vs. the activation differentiation free energy for selection S_{III}^- (dark line), S_{III}^+ (dark thin line), S_{IV}^- (gray line) and S_{IV}^+ (gray thin line). A combined selection strategy with all four elementary selections at equal strength $\eta = e^{\Delta/3k_B T}$ is also shown (dashed line). (e) The error rate vs. elongation speed under S_{III}^-, S_{III}^+, S_{IV}^-, S_{IV}^+, and the combined.

Essentially, one can identify four elementary selections along the reaction path (shown schematically in Fig. 4c) prior to the product (V) formation ($m = III$ and $n = 2$ in this case). Splitting the PPi release from the catalysis after the production formation has no impact on the selection checkpoints.

The first elementary selection, denoted S_{III}^-, rejects wrong nucleotides immediately upon binding to the pre-insertion state III (at the strength $\eta_{III}^- = k_{III-}^w / k_{III-}^r$). The next selection strategy, denoted S_{III}^+, inhibits wrong nucleotides from *inserting* into the active site (from III to IV, $\eta_{III}^+ = k_{III+}^r / k_{III+}^w$). The third selection strategy S_{IV}^- destabilizes the wrong nucleotides after being inserted at state IV ($\eta_{IV}^- = k_{IV-}^w / k_{IV-}^r$). The last selection strategy S_{IV}^+ inhibits catalytic reaction of the wrong nucleotides ($\eta_{IV}^+ = k_{IV+}^r / k_{IV+}^w$).

4.1 Error and speed control by the selection

To compare the elongation speeds and error rates under individual elementary selections, each selection is assumed to work at the same strength η, or to use the same amount of differential free energy $\Delta = k_B T \ln \eta$. The polymerization rate J/J_0 vs. Δ and the error rate Err are plotted in Fig. 4 d-e. Similar to the three-state scheme, one can see that the selection S_{III}^- maintains the highest elongation speed at a high NTP concentration (*e.g.* $J/J_0 \sim 0.8$). The elongation speeds under the selection S_{III}^+, S_{IV}^- and S_{IV}^+ all approach to similarly low values when the selections become strong (*e.g.* $J/J_0 \sim 0.2$), similar to that under S_{III}^+ in the three-state scheme.

    One can also write down the output error rate, for simplicity, at the irreversible product release condition ($k_{I-} \to 0$) as

$$Err = \left\{(1 - i_r)[k_{III-}k_{IV-}k_{V+} + k_{III-}k_{IV-}k_{V+} + k_{III-}k_{IV+}k_{V+} + k_{III+}k_{IV+}k_{V+}]\right\} \Big/ \left\{k_{III-}k_{IV-}k_{V-}(1 - i_r + i_r\eta_G) + \right.$$

$$\left. k_{V+}[k_{III-}k_{IV-} + k_{III-}k_{IV+} + k_{III+}k_{IV+} + i_r k_{III-}(k_{IV-}\eta_{III}^-\eta_{III}^+\eta_{IV}^-\eta_{IV}^+ + k_{IV+}\eta_{III}^-\eta_{III}^+ - k_{IV-} - k_{IV+})]\right\} \tag{6}$$

    One sees from Eq. 6 that the first two elementary selections S_{III}^- and S_{III}^+ reduce the error rate through a joint term $\eta_{III}^-\eta_{III}^+$, so that the same error rates are obtained for respective selections (when applied individually at the same strength). Similarly, the latter two selections S_{IV}^- and S_{IV}^+ also lead to the same error rates, respectively, and thus we say that they perform equally well in the error reduction. However, the performance of the latter two is inferior to the former two: When only the first two selections work and are equally strong ($\eta_{III}^- = \eta_{III}^+ \gg 1, \eta_{IV}^- = \eta_{IV}^+ = 1$),

$$Err = \frac{1 - i_r}{i_r}\left[1 + \frac{k_{III-}k_{IV-}k_{IV+} + k_{III+}k_{IV+}k_{V+}}{k_{III-}(k_{IV-} + k_{IV+})k_{V+}}\right]\frac{1}{\eta_{III}^-\eta_{III}^+} = \frac{1 - i_r}{i_r}[1 + \kappa_{III}]e^{-2\Delta/k_B T} \tag{7}$$

In contrast, when only the two latter selections work and are equally strong ($\eta_{III}^- = \eta_{III}^+ = 1, \eta_{IV}^- = \eta_{IV}^+ \gg 1$),

$$Err = \frac{1 - i_r}{i_r}\left[1 + \frac{k_{III-}k_{IV-}k_{IV+} + (k_{III-} + k_{III+})k_{IV+}k_{V+}}{k_{III-}k_{IV-}k_{V+}}\right]\frac{1}{\eta_{III}^-\eta_{III}^+} = \frac{1 - i_r}{i_r}[1 + \kappa_{IV}]e^{-2\Delta/k_B T} \tag{8}$$

    Since $\kappa_{III} < \kappa_{IV}$, a lower error rate is achieved under the first two selections than that under the latter two. This feature is due to the linear reaction topology and is independent of the kinetic parameters (for example, see a six-state scheme in Appendix III). When the last transition or the full scheme is reversible ($k_{I-} > 0$), S_{III}^- and S_{III}^+ still perform equally well in the error reduction, and outperform S_{IV}^- and S_{IV}^+. In Fig. 4e, we see that the lowest error rate reached by S_{III}^- or S_{III}^+ is $\sim 6 \times 10^{-4}$ for $\Delta \sim 9$ k$_B$T. The error rate reached by S_{IV}^- or S_{IV}^+ is $\sim 4 \times 10^{-3}$ for the same Δ. If one combines four elementary selections together, with each selection using a small differentiation free energy e.g., $\Delta/3=3$ k$_B$T, one obtains a fairly low error rate $\sim 2 \times 10^{-4}$, and the speed is still kept high as $J/J_0 \sim 0.75$.

4.2 Kinetic impact and variation of the rate-limiting step

Following the derivation (see Eq. 6 – **??** and Eq. A10), one can write down the elongation error rate in a general form (with an irreversible last transition in the scheme) as:

$$Err = \frac{(1 - i_r)(\kappa_a + \sum_{j=m}^{N} a_j)}{(1 - i_r)(\kappa_a + \sum_{j=m}^{N} a_j) + i_r(\kappa_b + \sum_{j=m}^{N} a_j \prod_{i=m}^{j} \eta_i^-\eta_i^+)} \tag{9}$$

where κ_a, κ_b and a_M are combinations of the kinetic parameters. In the three-state scheme, $m = N = III$; in the five-state scheme, $m = III$, $N = IV$). As each elementary selection grows strong ($\eta_i^\pm = e^{\Delta_i^\pm/k_B T} \gg 1$), $Err \sim \frac{1 - i_r}{i_r}(1 + \kappa)e^{-\Sigma\Delta_i^\pm/k_B T}$. That is to say, the error rate deceases exponentially with the *accumulative selection differentiation free energy* along the reaction path. If one lowers the rates of forward transitions involved in the selection (e.g. k_{III+}, k_{IV+} or both), or raises the rates of backward transitions, one can reduce the value

of κ and thus lower the error rate. Indeed, this type of accuracy improvement is at the price of lowering the speed, as described in the efficiency-accuracy tradeoff.

One can also see how the speed and error rate control vary when the rate-limiting step changes in the elongation cycle (see Fig. A4 in Appendix IV). Here the rate-limiting step is determined according to the kinetics of the right substrate. First, if one lowers the NTP concentration, the nucleotide binding can become rate limiting. We see that all selections significantly lower the speed. The first selection S_{III}^-, as it happens after the slow NTP binding, cannot recover the speed as the NTP binding is already too slow. Next, if the catalysis becomes slow such that the catalytic rate is much smaller than the reversal rate of the nucleotide insertion (e.g. $k_{IV+} \ll k_{IV-}$), then the error rate achieved under the former two selections (S_{III}^- and S_{III}^+) becomes almost identical to that under the latter two (S_{IV}^- and S_{IV}^+). Additionally, if the translocation after the product release happens slowly, then the elongation rate is largely determined by the translocation rate, and cannot be reduced much by nucleotide selection. In brief, the error and speed control patterns persist but become more or less pronounced at different rate-limiting conditions.

5 Discussion

In the current work, we show how stepwise nucleotide selection proceeds for efficient fidelity control in polymerase elongation. Basically, we want to identify selection strategies that achieve comparatively low error rates for a certain selection differentiation free energy, without significantly lowering the polymerization speed as in the absence of the selection. From previous studies, various ways of nucleotide selection had been reported for different polymerases [12]. As rate-limiting steps also vary among different systems, it is hard to identify common selection mechanisms. In this study, we demonstrate some general features of the selection systems, as summarized below. During evolution, polymerases did *not* necessarily evolve to be highly efficient in the selectivity. Functional developments of the polymerases have to meet various internal and external requirements. Nevertheless, it is worthwhile identifying those essential features that make the selection system efficient in fidelity control, so as to recognize and quantify how the fidelity control is achieved in a particular system.

To characterize stepwise nucleotide selection, one needs to consider the *selection differentiation free energy* (Δ) between the right and wrong substrates at every transition checkpoint along the reaction path. The differentiation free energy relies on physical properties of the enzyme and ambient conditions. For example, to differentiate the substrate species, some structural or electrostatic characteristics of the protein have to be developed, while water molecules need to be more or less excluded, and certain ions may also be required for coordination [11, 35, 36]. Accordingly, the selectivity demands very specific fine-tuning of molecular interactions, and the differentiation capacity is thus restricted at any one checkpoint (i.e., Δ cannot be very large).

An elementary selection is either to inhibit the forward transition or to enhance the backward transition in the wrong substrate incorporation, by modulating the transition activation barrier by Δ in comparison to that of the right species. When Δ is fixed, while the system kinetics varies in controlled conditions, the selection accuracy can be improved by compensation of the reaction efficiency, as pointed out by the efficiency-accuracy tradeoff [19].

On the other hand, if the reaction kinetics of the right substrates is given, while Δ is allowed to increase at one checkpoint, the error rate can be continuously lowered while the speed converges to a constant value. Depending on which selection checkpoint is exploited on the reaction path, the error rate and elongation speed vary. Our study shows that early selections on the reaction path outperform the late ones on the error reduction, and the initial selection is indispensible for maintaining a high elongation speed. Here, the initial selection starts right upon the substrate binding, and a very late one ends as the catalysis finishes. We show essentially that the error rate can be repeatedly lowered through the stepwise selection. That is to say, multiple kinetic checkpoints do improve the fidelity level as the selection free energies for differentiation at these checkpoints accumulate along the reaction path. Mathematically, this property is similar to the amplification

effects in kinetic proofreading, as the elongation cycle is essentially maintained at the NESS, breaking the detailed balance.

In previous sections, we compared error reduction and speed modulation of the elementary nucleotide selections in the three- and five-step elongation scheme. The three-state scheme is characterized by one-step NTP binding and two kinetic checkpoints, while two-step NTP insertion and four checkpoints apply in the five-state scheme. For mathematical simplicity, we assumed in both schemes that the PPi dissociation step is irreversible, as if PPi concentration is quite low around the active site of the polymerase. Similar results reveal in the fully reversible scheme on the speed and error control (see Appendix II). We also checked the four-state and six-state kinetic schemes (see Appendix III). The four-state scheme has similar results as the three-state scheme if there is only one-step NTP binding prior to the chemical catalysis; the scheme achieves similar results as the five-state case if the NTP insertion happens in two steps. In the six-state scheme, we made a three-step pre-chemical NTP insertion process, with six checkpoints in total. The variations of the kinetic scheme make no essential changes in the speed and error control in the corresponding elementary selections.

Below, we summarize crucial aspects of the efficient fidelity control, which is essentially the achievement of a low error rate with minimal lowering of the speed. Last, we apply the current framework to describe some particular polymerases.

## 5.1 Achieve low error rates – select early, properly, and repeatedly

We have examined elementary selections that tune only one transition barrier forward or backward at a time when incorporating the wrong nucleotides. The elementary selections are arranged sequentially along the reaction path of the elongation cycle. Our results highlight two interesting findings: (*i*) The error rate achieved by the selection S_i^- is always the same as that achieved by the S_i^+ at the same selection strength, and (*ii*) The error rate achieved by the selection S_i^+ is always lower than that achieved by S_{i+1}^- at the same selection strength. In brief, the error reduction performance of the selection does not improve following down the reaction path.

As any nucleotide selection strategy can be regarded as a combination of the elementary selections, the above results give some rules of thumb on identifying a *proper* selection strategy. First, as a direct consequence of (*i*) above, one cannot combine S_i^- with an 'anti-selection' $(S_i^+)^{-1}$ of strength $(\eta_i^+)^{-1} = e^{-\Delta/k_B T}(<1)$ to achieve an error reduction. Here the *anti-selection* indicates a 'selection' with $\eta<1$, which favors the wrong substrate rather than the right one. This futile strategy is illustrated in Fig. 5a *left*; the state *i* is destabilized for the wrong species, while *both* forward and backward activation barriers are reduced by Δ (>0): Thus, the operations S_i^- and $(S_i^+)^{-1}$ simply cancel each other.

On the other hand, another selection strategy illustrated in Fig. 5a *middle*, for example, does work well for error reduction. It can be regarded as a combination of S_i^+ and an *anti-selection* $(S_{i+1}^-)^{-1}$ of strength $(\eta_{i+1}^-)^{-1} = e^{-\Delta/k_B T}$. S_i^+ outperforms S_{i+1}^- in the error control as from (*ii*) above, so they do not fully cancel.

Consequently, the strategy shown in Fig. 5a *right* works as a combination of S_{i+1}^+ of strength $\eta_{i+1}^+ = e^{\Delta/k_B T}$, S_i^- of strength $\eta^- = e^{(\Delta'-\Delta)/k_B T}$ ($\eta_i^->1$ as $\Delta' >\Delta$), $(S_i^+)^{-1}$ of strength $(\eta_i^+)^{-1} = e^{(\Delta-\Delta')/k_B T}$ (<1), and $(S_{i+1}^-)^{-1}$ of strength $(\eta_{i+1}^-)^{-1} = e^{(\Delta-\Delta')/k_B T}$. Since S_i^- and $(S_i^+)^{-1}$ cancel each other, the overall strategy actually works as a combination of S_{i-1}^+ and $(S_{i+1}^-)^{-1}$. Hence, the error rate achieved under this strategy cannot be lower than that achieved under a single elementary selection S_{i-1}^+. It shows that an improperly combined selection strategy cannot improve the fidelity over that of an elementary selection.

Indeed, we put a general expression of the elongation error rate as a function of individual selection strength (see Eq. 9). The properties (*i*) and (*ii*) summarized above are the natural consequences of this expression. Importantly, one obtains $Err \sim \frac{1-i_r}{i_r}(1 + \kappa)e^{-\Sigma\Delta_i^{\ddagger}/k_B T}$ for any combined selection when individual selections are sufficiently strong ($e^{\Delta_i^{\ddagger}/k_B T} >> 1$ or $\Delta_i^{\ddagger}>\sim 3$ k_BT, above the thermal fluctuation level). It indicates that the error reduction can be amplified through multiple kinetic steps along the reaction path. Since $\kappa>0$, it is easy to see that the error rate cannot be lower than $\frac{1-i_r}{i_r}e^{-\Sigma\Delta_i^{\ddagger}/k_B T}$, or the accuracy cannot be higher than $e^{\Sigma\Delta_i^{\ddagger}/k_B T}$. Lowering the forward rate or raising the backward rate for the involved transition in the selection
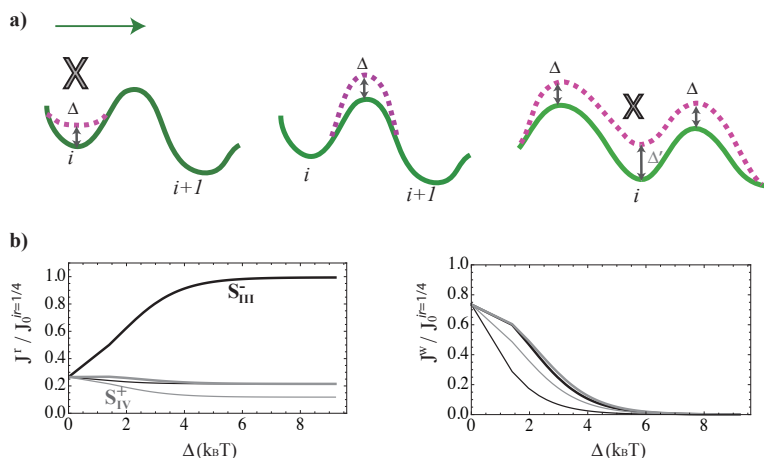
Figure 5: Error reduction strategies and flux properties. (a) Examples on futile and effective selection strategies. The one on the left is a futile strategy, which combines S_i^- and $(S_i^+)^{-1}$ that cancel each other for error reduction. The one in the middle combines S_i^+ and $(S_{i+1}^-)^{-1}$ together and works for error reduction. The one on the right is a combination of four elementary selections; the overall strategy is equivalent to S_{i-1}^+ and $(S_{i+1}^-)^{-1}$ together for error reduction. Note that S^{-1} selects against the right instead of the wrong nucleotides. (b) The elongation fluxes or speeds for the right (*left*) and wrong (*right*) nucleotide species vs. the differentiation free energy Δ, calculated from the five-state scheme. $J_0^{1/4}$ is the flux in the absence of the nucleotide selection at $i_r = 1/4$.

can lower the value of κ so that the error rate reduces as the reaction slows down. When every elementary selection participates properly in the combined selection ($\Delta_i^\ddagger > 0$), $e^{\Sigma \Delta_i^\ddagger / k_B T}$ becomes the maximum accuracy as suggested previously [17]. Note that when $\Delta_i^\ddagger < 0$ exists, $\Sigma \Delta_i^\ddagger = \Omega_{max}$ does not coincide with the seemingly largest free energy difference between the right and wrong (e.g. Δ' is the largest stepwise free energy difference in Fig. 5a *right*, but is indeed irrelevant to the selectivity). Hence, an overall description of the selection using only the maximum free energy differentiation can be insufficient or even misleading.

5.2 Maintain high polymerization speed – initial screening is indispensible

Our results indicate that the overall elongation speed is more or less reduced upon any nucleotide selection to increase the fidelity. For a constant energy differentiation capacity, however, varying the selection checkpoint can possibly improve both the fidelity and speed. For the very first selection, it outperforms the late selections not only in achieving a low error rate, but also in maintaining a high speed.

Fig. 5b shows the respective polymerization speeds of the right and wrong nucleotides, as Δ increases for various elementary selections (as in the five-state scheme). When there is no selection ($\Delta = 0$), the relative flux for the right and wrong species ($J^r / J_0^{1/4}$ and $J^w / J_0^{1/4}$) are 0.25 and 0.75, respectively (as $i_r = 1/4$ or the input error rate is ¾). When Δ increases, the wrong fluxes uniformly decrease to zero. On the other hand, the right fluxes diminish to small values for all but the initial screening. The more stringent the initial screening, the higher the polymerization flux of the right nucleotides, as most of the wrong species are expelled away soon at entry and replaced by the right species. However, the initial free energy differentiation can be quite limited: When the nucleotide is just recruited, the site is relatively open (e.g. in the pre-insertion state of T7 RNAP [34] and water solvent may not be well excluded. Hence, it is unlikely to achieve nucleotide selection largely at the beginning. Nevertheless, when initial screening is *combined* with selections performed later in the cycle, the elongation speed would still remain high, approaching that under the initial screening alone. This is because the presence of the initial screening makes it efficient to throw away the wrong substrates in the proofreading-free system. The properties highlight the importance of *including* the substrate screening at the beginning for an efficient selection system.

When the nucleotide concentration is very low, however, even the initial selection can lower the elongation rate significantly. On the other hand, if the rate-limiting step happens behind all the selection checkpoints, such as at the translocation, then the selection hardly impacts the speed. In brief, for polymerase elongation at high nucleotide concentration, the initial selection always helps to maintain a high speed, while later selections lower the speed. This gives some clues to determine essential amino acids for the initial screening through experimental mutagenesis: For mutations that adversely affect both the accuracy and speed, the original amino acids at the mutation sites likely contribute to the initial screening; for those that lower the accuracy but not the speed, either the original amino acids do not contribute to the initial selections, or the mutations still exhibit the initial screening function.

5.3 The nucleotide selection in some exemplary polymerase systems

In this work, we have used elongation kinetic data of single-subunit T7 RNAP [4, 7, 31] for numerical demonstrations (see kinetic parameters in Table A1 from Appendix V). Though the three-state scheme had been employed in early work for experimental data fitting [4], later on studies supported a five-state elongation scheme in this system [31]. In the five-state scheme, the pre-chemical nucleotide insertion following the initial nucleotide binding/pre-insertion is regarded as rate limiting [31]. Our recent molecular dynamics simulations show that substantial nucleotide selection happens prior to the full insertion of the nucleotide into the active site in T7 RNAP (*in preparation*). That is, both the initial screening S_{III}^- upon the NTP per-insertion and the following selection S_{III}^+ during the NTP insertion play essential roles in the nucleotide selection. Hence, T7 RNAP seems to be a quite efficient selection system that can fully employ the early selections on the reaction path. Since the error rate achieved by T7 RNAP is $\sim 10^{-4}$ [32], one can estimate the maximum or accumulate selection free energy differentiation at ~ 10 k$_B$T. Actually, T7 RNAP achieves the error rate without proofreading detected. This likely explains why the nucleotide selection has to be efficient in T7 RNAP.

Next, we examined selection kinetics of T7 DNAP as its kinetic rates for incorporating both the right and wrong nucleotides had been reported [14] (see Table A2 in Appendix V). The chemical catalysis proceeds more slowly than the nucleotide insertion in T7 DNAP, while the reversal of the nucleotide insertion happens extremely slowly. The selection strengths are identified as: $\eta_{III}^- \sim 7$ for S_{III}^-, $\eta_{III}^+ \sim 3$ for S_{III}^+, $\eta_{IV}^- \sim 263$ for S_{IV}^-, and $\eta_{IV}^+ \sim 1200$ for S_{IV}^+, giving the differentiation activation free energies $\Delta_i/$k$_B$T as {1.9, 1.1, 5.6, 7.1}. The DNAP conducts proper nucleotide selection by combining all four elementary selections. However, the initial screening does not seem to be strong enough to support very high speed (see Fig. A5 in Appendix V); the first two selections also do not appear strong enough to make the overall selection highly efficient. The full selection gives an error rate $\sim 10^{-3}$ in T7 DNAP. The performances seem to leave room for further improvements by proofreading, which is indeed required and substantial in T7 DNAP.

In multi-subunit RNAPs, recent mutagenesis studies nicely show that discrimination against wrong nucleotides proceeds via a stepwise mechanism, and each step contributes differently to the overall fidelity [26]. In these systems, at least two steps happen prior to the chemical catalysis [37]. In particular, the non-complementary NTPs are discriminated efficiently through both the first and second checkpoints, at the open active center and through a trigger loop folding process, respectively. The discrimination of the deoxy-NTPs does not happen until after the first checkpoint, and hence, appears less efficient. Indeed, the regulation of the enzyme activities is largely controlled through the trigger loop folding, providing possibilities that several selection checkpoints are coordinated coupled in the fidelity control. It is not clear which step is rate limiting in the multi-subunit RNAPs. Likely one slow event takes place before or during the catalytic stage, and another slow event is around the translocation stage [38]. In that case, the speed modulation of the nucleotide selection may not be significant.

6 Conclusions

In this work we have studied how stepwise nucleotide selection could proceed efficiently during the template-based polymerase elongation. Basically, the selection happens at multiple checkpoints prior to the end of chemical catalysis or the product formation. At each state, conformational transition of the enzyme backward or forward along the reaction path is enhanced or inhibited when the enzyme is bound with a wrong nucleotide. The selection through a single backward or forward transition is regarded as an elementary selection, and any selection in general can be regarded as a combination of the elementary selections. An efficient selection strategy takes advantage of multiple selection checkpoints to reduce the error rate repeatedly, and selects early along the reaction path. At the same time, any selection checkpoint is subject to structural and energetic constraints to differentiate the nucleotide species. The efficient selection strategy achieves a relatively low error rate with a limited amount of selection differentiation free energy accumulated along the reaction path, while mildly perturbing the overall elongation speed.

We found that at sufficiently high nucleotide concentrations, the initial screening against wrong nucleotides perturbs the elongation speed slightly, while selections thereafter on the reaction path can significantly diminish the elongation speed. Notably, combing the initial screening with selections afterwards keeps the speed high, similarly to that under the initial screening alone. Hence, for polymerases that need high elongation speeds, the initial screening is indispensible, and even a small differentiation upon nucleotide entry can help. Importantly, we found that the early selections along the reaction path outperform the late ones in the error reduction, as lower error rates are achieved under the early rather than the late selections for the same activation free energy for differentiation. In particular, for a pair of neighboring elementary selections, one rejects the wrong substrate state back to the previous state and one inhibits the same wrong substrate state toward the next state give the same error rate at the same differentiation free energy. These properties persist but become more or less pronounced at different rate-limiting conditions in the elongation process.

Based on this framework, we compared a proofreading-free T7 RNAP with a proofreading T7 DNAP. We found T7 RNAP to be an efficient selection system while T7 DNAP does not seem to be so. The current work on the stepwise nucleotide selection supports further quantitative research to reveal underlying mechanisms of the selection. It may further help molecular engineering design of efficient selection systems.

APPENDIX

Appendix I: The generic three-state scheme and the efficiency-accuracy tradeoff

In Eq. 4 in main text, the polymerization flux or rate J is obtained in the Michaelis-Menten form. The maximum rate constant $k_{\max}^0$ and the Michaelis constant K_M^0 are written as:

$$k_{\max}^0 = \frac{k_{I+}k_{III+}}{k_{I+} + k_{III+}} \tag{A1}$$

$$K_M^0 = \frac{k_{I+} + k_{II-}}{k_{I+} + k_{III+}} \frac{k_{III+} + k_{III-}}{k_T^0} \tag{A2}$$

where k_T^0 is the NTP binding rate constant ($k_{II-} = k_T^0[NTP]$). Correspondingly, the exact forms of Γ and Λ are:

$$\Gamma = \frac{(1 - i_r)(\eta_{III}^+ - 1)(k_{III+} + k_{III-})}{k_{III+} + (1 - i_r\eta_{III}^-\eta_{III}^+)k_{III-}} \cdot \frac{k_{I+}}{k_{I+} + k_{III+}} \approx \frac{(1 - i_r)(\eta_{III}^+ - 1)(k_{III+} + k_{III-})}{k_{III+} + (1 - i_r + i_r\eta_{III}^-\eta_{III}^+)k_{III-}} \tag{A3}$$

$$\Lambda = \frac{k_{III+} + k_{III-}\eta_{III}^-\eta_{III}^+}{k_{III+} + (1 - i_r + i_r\eta_{III}^-\eta_{III}^+)k_{III-}} \tag{A4}$$

The approximation in Eq. A3 is taken for $k_{III+} \ll k_{I+}$, that is, the translocation rate k_{I+} is much larger than the catalytic rate k_{III+}. It has been measured that the translocation is much faster than other kinetic steps in the elongation cycle [4, 39]. If the translocation slows down, the value of Γ will decrease, and the impact from Γ weakens. As $k_{III+} \gg k_{I+}$, $\Gamma \to 0$, the selection strength η can only modulate K_M rather than $k_{\max}$ in the Michaelis-Menten form of the elongation rate/speed. In that case, the nucleotide insertion happens very fast and the selection cannot significantly affect the elongation speed at fairly high nucleotide concentration.

To make it clear how the individual selection strength $\eta = e^{\Delta/k_B T}$ or the differentiation free energy Δ affects Γ and Λ to modulate the rate and efficiency of the polymerization, Fig. A1 below shows Γ, Λ, $k_{\max} = \frac{k_{\max}^0}{1+\Gamma}$, $K_M = \frac{\Lambda}{1+\Gamma}K_M^0$, and $\zeta \equiv \frac{k_{\max}}{K_M} = \frac{1}{\Lambda}\zeta_0$ vs. Δ under the two elementary selections S_{III}^- and S_{III}^+ in the three-state scheme. Note that the two selections give the same curves of Λ and $\zeta = k_{max}/K_M$. Indeed, Λ, ζ, and the error rate Err all depend only on the accumulate selection strength $\eta_{III}^- \eta_{III}^+$ rather than the individual terms.
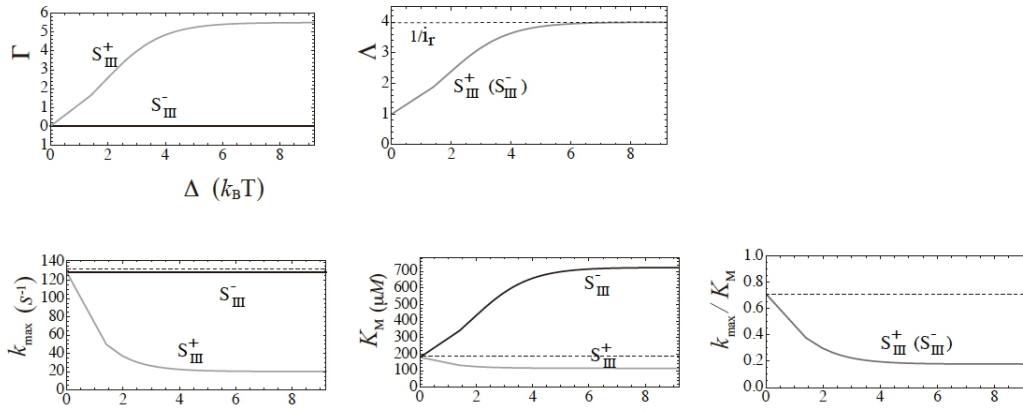
Figure A1: Variations of Γ, Λ, k_{max}, K_M, and $\zeta = k_{max}/K_M$ upon the variation of the differentiation free energy Δ under respective selections S_{III}^- (black) and S_{III}^+ (gray) in the three-state elongation scheme.

In order to compare the polymerization rates for the right and wrong nucleotides as they compete for binding, one writes $J = J^r + J^w$ with

$$J^r = i_r \Lambda J = \frac{\frac{\Lambda k_{max}^0}{1+\Gamma}[NTP]_{right}}{\frac{\Lambda}{1+\Gamma}K_M^0 + [NTP]_{right} + [NTP]_{wrong}} \tag{A5}$$

and

$$J^w = (1 - i_r \Lambda)J = \frac{\frac{(1-i_r\Lambda)k_{max}^0}{(1-i_r)(1+\Gamma)}[NTP]_{wrong}}{\frac{\Lambda}{1+\Gamma}K_M^0 + [NTP]_{right} + [NTP]_{wrong}} \tag{A6}$$

From Eq. A5 and A6, one can obtain the polymerase efficiencies for the right and wrong nucleotides as $\zeta^r = \zeta_0$ and $\zeta^w = \frac{1-i_r\Lambda}{(1-i_r)\Lambda}\zeta_0$, respectively (note $\zeta_0 \equiv k_{max}^0/K_M^0$). The accuracy A can be written then as:

$$A \equiv \frac{\zeta^r}{\zeta^w} = \frac{(1-i_r)\Lambda}{1-i_r\Lambda} = \frac{k_{III-}\eta_{III}^-\eta_{III}^+}{k_{III+} + k_{III-}} \tag{A7}$$

As in the efficiency-accuracy tradeoff discussed early [21, 23, 40], we can calculate straightforward the efficiency ζ^r for incorporating the right nucleotide substrates:

$$\zeta^r = \frac{k_{cat}^r}{K_M^r} = \frac{\frac{\Lambda}{1+\Gamma}k_{max}^0}{\frac{\Lambda}{1+\Gamma}K_M^0} = \zeta^0 = \kappa_a \cdot \frac{k_{III+}}{k_{III+} + k_{III-}} \tag{A8}$$

where $\kappa_a \equiv \frac{k_T^0 k_{I+}}{k_{I+} + k_{II-}} = \frac{k_T^0}{1 + \frac{k_{II-}}{k_{I+}}}$ is the effective binding constant in the three-state cycle, $\kappa_a \sim k_T^0/2$ as polymerases translocate in a Brownian ratchet fashion, with equal forward and backward rates ($k_{I+} \sim k_{II-}$) [6].

From Eq. A7, we write $\frac{d-A}{d-1} = \frac{\eta_{III}^-\eta_{III}^+-(k_{III-}\eta_{III}^-\eta_{III}^++k_{III+})/(k_{III-}+k_{III+})}{\eta_{III}^-\eta_{III}^+-1} = \frac{k_{III+}}{k_{III-}+k_{III+}}$, where $d \equiv \eta_{III}^-\eta_{III}^+ = e^{\Omega \max /k_B T}$ is the accumulate selection strength, or the 'maximum accuracy'.

Hence, we obtain the linear efficiency-accuracy tradeoff relationship as that in Eq. 1 from [19].

$$\zeta^r = \frac{k_{cat}^r}{K_M^r} = \kappa_a \frac{d - A}{d - 1} \tag{A9}$$

Appendix II: In the reversible three-state scheme of polymerase elongation

One can solve Eq. 2 at the steady state $\frac{d}{dt}\Pi = 0$ *without* using the approximation $k_{I-} \to 0$. The error rate *Err* can be obtained iteratively. From the expression of *Err* (not shown), one can see it contains *only* $\eta_{III}^-\eta_{III}^+$ but not the individual term η_{III}^- or η_{III}^+. Hence, the same strength of the individual selection always leads to the same error rate. In Fig. A2 below, we show the impact of the selections on the speed and error rates:
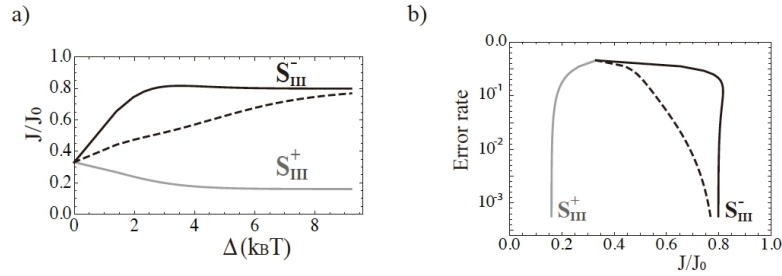
Figure A2: The polymerization rates/speeds and error rates under the two elementary selections in the fully reversible three-state scheme. (a) The polymerization speed (normalized J/J₀) vs. the differentiation free energy $\Delta = k_B T \ln \eta$. (b) The error rate vs. the polymerization speed as Δ varies. See main Fig. 3 (c-d).

From the results above, we see that the conclusions still hold as: (1) The initial selection S_{III}^- leads to a polymerization speed close to J_0, much higher than that under S_{III}^+. Under the combined selection, the polymerization speed also approaches to that under S_{III}^- alone as the selection becomes strong. (2) S_{III}^- and S_{III}^+ perform equally well in the error control, giving the same error rate at the same differentiation free energy or selection strength.
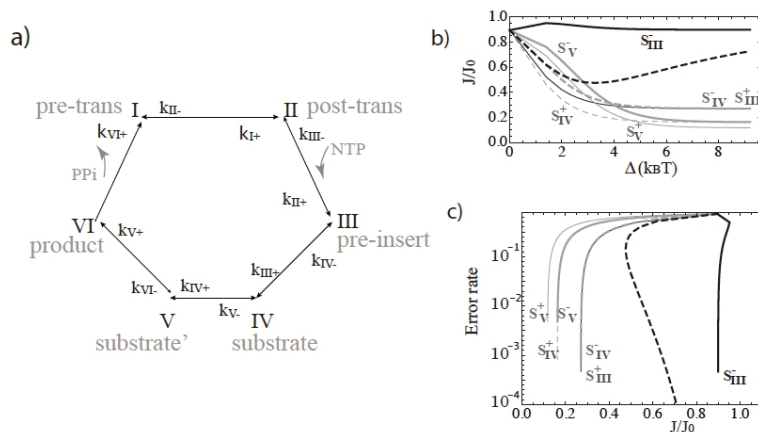
Appendix III: A six-state scheme with three pre-chemistry transitions

One can also build a six-state kinetic scheme by putting another 'insertion' step (see Fig. A3a below) prior to the chemical catalysis, in addition to the pre-insertion and insertion step in the five-state scheme. Solving an equation similar to Eq. 2 but in a ten-state vector space: $\Pi = (P_I, P_{II}, P_{III}^r, P_{III}^w, P_{IV}^r, P_{IV}^w, P_V^r, P_V^w, P_{VI}^r, P_{VI}^w)^T$ ('r' and 'w' labeling for probabilities of the wrong and right nucleotide bound states, respectively), one obtains the error rate:

$$Err = \cfrac{1}{1 + \cfrac{i_r[k_{VI+}(k_{III+}k_{IV+}k_{V+} + k_{III-}\eta_{III}^-\eta_{III}^+(k_{IV+}k_{V+} + k_{IV-}\eta_{IV}^-\eta_{IV}^+(k_{V+} + k_{V-}\eta_V^-\eta_V^+))) + k_{III-}k_{IV-}k_{V-}\eta_G]}{(1 - i_r)[k_{III-}k_{IV-}k_{V-}k_{VI-} + k_{VI+}(k_{III+}k_{IV+}k_{V+} + k_{III-}(k_{IV+}k_{V+} + k_{IV-}k_{V-} + k_{IV-}k_{V+}))]}} \tag{A10}$$

The diagrams on polymerization speeds and error rates under the selections are:

From the above results we see that the conclusions listed in the previous section still hold. In particular, one sees that two neighboring selections against wrong nucleotides, S_i^- and S_i^+ from the same kinetic state i, give the same error rates at the same selection strength (as $\eta_i^- = \eta_i^+$).

Figure A3: The polymerization speeds and error rates under nucleotide selections in the six-state scheme. (a) The six-state elongation scheme with three pre-chemistry steps: pre-insertion (II→III), two insertions (III→IV and IV→V) (b) The polymerization speed (normalized J/J₀) vs. $\Delta = k_B T \ln \eta$. (c) The error rate vs. the polymerization speed as Δ varies. See also Fig. 4 (d-e).

Appendix IV: Error rate vs. speed at different rate-limiting conditions

In main (*IV.2 Kinetic impact and variation of the rate-limiting step*) we address how variation of the rate-limiting step affects the selection modulation on the speed and error rate. Below, we show in Fig. A4 the error rate vs. speed in the five-state elongation scheme under the four elementary selections, at different rate-limiting conditions.
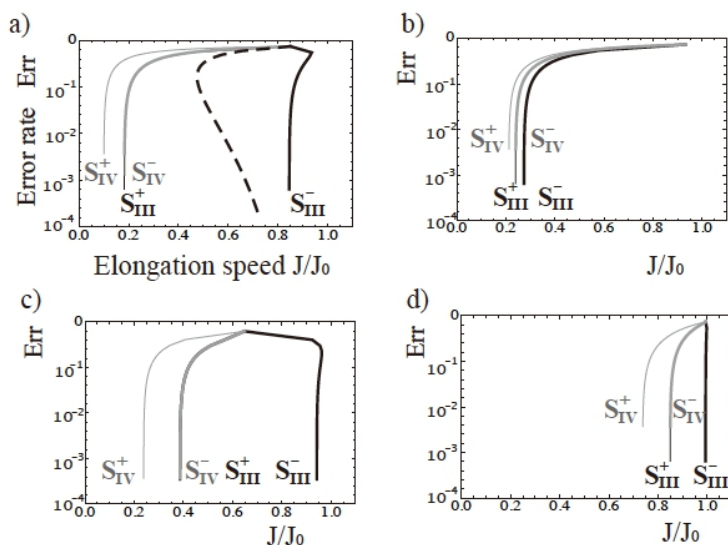
Figure A4: The error rate vs. elongation speed in the five-state elongation scheme. (a) The same as that in main Fig. 4d for easy comparison. In this case, NTP insertion (from III->IV) is rate limiting; (b) NTP concentration is low so that NTP binding (II->III) is rate limiting; (c) The catalytic process is slow (IV->V), and the rate is much lower than the reverse rate of the nucleotide insertion; (d) The translocation (I<->II) is rate limiting.

In the case that NTP binding is slow (Fig. A4b), all selections lower the speed significantly. In case that catalytic rate is very low (and is much lower than the reverse rate of the NTP insertion, Fig. A4c), all selections

perform similarly well in error reduction as almost the same error rates are obtained at the same selection strength. When the translocation is low (Fig. A4d), selections do not impact on the polymerization speed.

Appendix V: T7 DNAP in five-state scheme

For comparison, kinetic data utilized for T7 RNAP [7, 31] are listed in Table A1 below; kinetic data for T7 DNAP [14] are listed in Table A2, for incorporating both right and wrong nucleotides.

Table A1: Kinetic rates for T7 RNAP used in current work for numerical demonstration by default. All rates listed above are in units of s^{-1}. Data in bold are from transient state kinetics measured [31]. Other were numerically tuned or used for convenience [7].

Translocation		NTP binding/ pre-insertion		NTP insertion		Chemical catalysis		PPi dissociation	
k_{t+}	k_{t-}	$k_{b+} = k_o^T[NTP]$	k_{b-}	k_{i+}	k_{i-}	k_{c+}	k_{c-}	k_{d+}	k_{d-}
5000	5000	2*588	2 × **80**=160	**220**	210	1000	135	**1200**	0.01
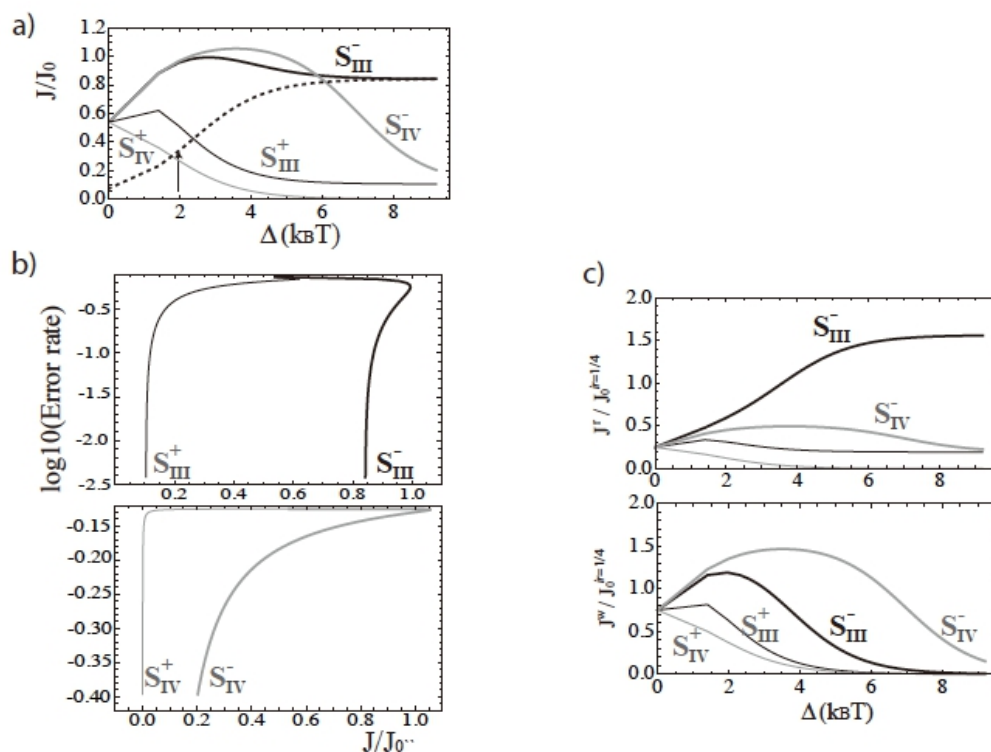
Figure A5: The polymerization speeds and error rates under nucleotide selections in the five-state scheme of T7 DNAP. (a) The polymerization speed (normalized J/J_0) vs. $\Delta = k_B T \ln \eta$. The dashed line is the combined nucleotide selection, as $\eta_{III}^+ = 3$, $\eta_{IV}^- = 262.5$, $\eta_{IV}^+ = 1200$ while Δ varies for S_{III}^-. The arrow indicates $\Delta \sim 2k_B T (\eta_{III}^- = 7.1)$ for the initial screening selection. (b) The error rate vs. the polymerization speed as Δ varies. Shown in two diagrams for clarity. One can compare the results (a, b) with that in main Fig. 4 (d-e). (c) The polymerization speeds for the right and wrong nucleotide species in T7 DNAP, in comparison with that of T7 RNAP in main Fig. 5b.

Table A2: Kinetic rates for T7 DNAP used for numerical demonstration below (upper row for the right substrate and lower row for the wrong substrate). All rates listed above are in units of s^{-1}. Data in **bold** are from reference [14]. Other data are the same as that in T7 RNAP (see Table A1) for easy comparison.

Translocation		NTP binding/ pre-insertion		NTP insertion		Chemical catalysis		PPi dissociation	
k_{t+}	k_{t-}	$k_{b+} = k_o^T[NTP]$	k_{b-}	k_{i+}	k_{i-}	k_{c+}	k_{c-}	k_{d+}	k_{d-}
5000	5000	2*588	2 × 28=56	**660**	**1.6**	**360**	320	1200	0.01
5000	5000	2*588	2 × 200=400	**220**	**420**	**0.3**	320	1200	0.01

The selections in T7 DNAP, in comparison with that of T7 RNAP (see Fig. 4 and 5 in main), are shown below in Fig. A5. The combined selection strategy for T7 DNAP is shown in Fig. A5a (dashed line). In this strategy, the strengths for three of the four elementary selections are set at $\eta_{III}^+ = 3 \; \eta_{IV}^- = 262.5 \; \eta_{IV}^+ = 1200$ for $S_{III}^+, S_{IV}^-, S_{IV}^+$, while the strength of the initial selection S_{III}^- varies. The arrow points to $\Delta \sim 2 \; k_B T$ for S_{III}^- as reading the data from Table A2, giving a polymerization speed ~ 48 nt/s ($J_0 \sim 144$ nt/s). One sees that if Δ increases upon the initial screening S_{III}^-, the polymerization speed would increase. Overall, the nucleotide selection in T7 DNAP gives an error rate about 10^{-3}. The proofreading is expected to further help the error reduction in T7 DNAP.

References

[1]   Buc, H., and T. Strick, editors. 2009. RNA polymerase as molecular motors. The Royal Society of Chemistry, Cambridge, UK.

[2]   Bar-Nahum, G., V. Epshtein, A. E. Ruckenstein, R. Rafikov, A. Mustaev, and E. Nudler. 2005. A Ratchet Mechanism of Transcription Elongation and Its Control. *Cell* 120:183-193.

[3]   Abbondanzieri, E. A., W. J. Greenleaf, J. W. Shaevitz, R. Landick, and S. M. Block. 2005. Direct observation of base-pair stepping by RNA polymerase. *Nature* 438:460-465.

[4]   Thomen, P., P. J. Lopez, and F. Heslot. 2005. Unravelling the Mechanism of RNA-Polymerase Forward Motion by Using Mechanical Force. *Physical Review Letters* 94:128102.

[5]   Guo, Q., and R. Sousa. 2006. Translocation by T7 RNA Polymerase: A Sensitively Poised Brownian Ratchet. *Journal of Molecular Biology* 358:241-254.

[6]   Wang, H., and G. Oster. 2002. Ratchets, power strokes, and molecular motors. *Applied Physics A* 75:315-323.

[7]   Yu, J., and G. Oster. 2012. A Small Post-Translocation Energy Bias Aids Nucleotide Selection in T7 RNA Polymerase Transcription. *Biophysical Journal* 102:532-541.

[8]   Johnson, K. A. 1993. Conformational coupling in DNA polymerase fidelity. *Annual Review of Biochemistry* 62:685-713.

[9]   Schlick, T., K. Arora, W. A. Beard, and S. H. Wilson. 2012. Perspective: pre-chemistry conformational changes in DNA polymerase mechanisms. *Theoretical Chemistry Accounts* 131:1287.

[10]   McCulloch, S. D., and T. A. Kunkel. 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Research* 18:148-161.

[11]   Joyce, C. M. 1997. Choosing the right sugar: How polymerase select a nucleotide substrate. *Proceedings of the National Academy of Sciences of the United States of America* 94:1619-1622.

[12]   Joyce, C. M., and S. J. Benkovic. 2004. DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry* 43:14317-14324.

[13]   Sydow, J. F., and P. Cramer. 2009. RNA polymerase fidelity and transcriptional proofreading. *Current Opinion in Structural Biology* 19:732-739.

[14]   Johnson, K. A. 2010. The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochimica et Biophysica Acta* 1804:1041-1048.

[15]   Hopfield, J. J. 1974. Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity. *Proceedings of the National Academy of Sciences of the United States of America* 71:4135-4139.

[16]   Ninio, J. 1975. Kinetic amplification of enzyme discrimination. *Biochimie* 57:587–595.

[17]   Ehrenberg, M., and C. Blomberg. 1980. Thermodynamic constraints on kinetic proofreading in biosynthetic pathways. *Biophysical Journal* 31:333-358.

[18]   Blanchard, S., R. Gonzalez, H. Kim, S. Chu, and J. Puglisi. 2004. tRNA selection and kinetic proofreading in translation. *Nature structural & molecular biology* 11:1008-1014.

[19] Johansson, M., J. Zhang, and M. Ehrenberg. 2012. Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA Selection. *Proceedings of the National Academy of Sciences of the United States of America* 109:131-136.

[20] Galburt, E. A., S. W. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, and C. Bustamante. 2007. Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* 446:820-823.

[21] Herbert, K. M., W. J. Greenleaf, and S. M. Block. 2008. Single-Molecule Studies of RNA Polymerase: Motoring Along. *Annual Review of Biochemistry* 77.

[22] Tadigotla, V. R., D. O. Maoileidigh, A. M. Sengupta, V. Espshtein, R. H. Ebright, E. Nudler, and A. E. Ruckenstein. 2006. Thermodynamic and kinetic modeling of transcriptional pausing. *Proceedings of the National Academy of Sciences of the United States of America* 103:4439-4444.

[23] Voliotis, M., N. Cohen, C. Molina-Paris, and T. B. Liverpool. 2009. Backtracking and Proofreading in DNATranscription. *Physicsal Reveiw Letters* 102:258101.

[24] Depken, M. D., E. A. Galburt, and S. W. Grill. 2009. The Origin of Short Transcriptional Pauses. *Biophysical Journal* 96:2189-2193.

[25] Violiotis, M., N. Cohen, C. Molina-Paris, and L. T.B. 2012. Proofreading of misincorporated nucleotides in DNA transcription. *Physical Biology* 93:036007.

[26] Yuzenkova, Y., A. Bochkareva, V. R. Tadigotla, M. Roghanian, S. Zorov, K. Severinov, and N. Zenkin. 2010. Stepwise mechanism for transcription fidelity. *BMC Biology* 8:54.

[27] Cady, F., and H. Qian. 2009. Open-system thermodynamic analysis of DNA polymerase fidelity. *Physical Biology* 6:036011.

[28] Andrieux, D., and P. Gaspard. 2008. Nonequilibrium generation of information in copolymerization processes. *Proceedings of the National Academy of Sciences of the United States of America* 105:9516-9521.

[29] Ehrenberg, M., and C. G. Kurland. 1984. Costs of accuracy determined by a maximal growth rate constraint. *Quarterly Reviews of Biophysics* 17:45-82.

[30] Kurland, C. G., and M. Ehrenberg. 1984. Optimization of translation accuracy. *Progress in Nucleic Acid Research and Molecular Biology* 31:191-219.

[31] Anand, V. S., and S. S. Patel. 2006. Transient State Kinetics of Transcription Elongation by T7 RNA Polymerase. *The Journal of Biological Chemistry* 281:35677-35685.

[32] Huang, J., L. G. Brieba, and R. Sousa. 2000. Misincorporation by Wild-Type and Mutant T7 RNA Polymerases: Identification of Interactions That Reduce Misincorporation Rates by Stabilizing the Catalytically Incompetent Open Conformation. *Biochemistry* 39:11571-11580.

[33] Yin, Y. W., and T. A. Steitz. 2004. The Structural Mechanism of Translocation and Helicase Activity in T7 RNA Polymerase. *Cell* 116:393-404.

[34] . Temiakov, D., V. Patlan, M. Anikin, W. T. McAllister, S. Yokoyama, and D. G. Vassylyev. 2004. Structural Basis for Substrate Selection by T7 RNA Polymerase. *Cell* 116:381-391.

[35] Sousa, R., and R. Padilla. 1995. A mutant T7 RNA polymerase as a DNA polymerase. *The EMBO Journal* 14:4609-4621.

[36] Petruska, J., L. C. Sowers, and M. F. Goodman. 1986. Comparison of nucleotide interactions in water, proteins, and vacuum: Model for DNA polymerase fidelity. *Proceedings of the National Academy of Sciences of America* 83:1559-1562.

[37] Feig, M., and Z. F. Burton. 2010. RNA Polymerase II with Open and Closed Trigger Loops: Active Site Dynamics and Nucleic Acid Translocation. *Biophysical Journal* 99:2577-2586.

[38] Dangkulwanich, M., T. Ishibashi, S. Liu, M. L. Kireeva, L. Lubkowska, M. Kashlev, and C. J. Bustamante. 2013. Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. *eLife* 2:e00971.

[39] Thomen, P., P. J. Lopez, U. Bockelmann, J. Guillerez, M. Dreyfus, and F. Heslot. 2008. T7 RNA Polymerase Studied by Force Measurements Varying Cofactor Concentration. *Biophysical Journal* 95:2423-2433.

[40] Bai, L., T. J. Santangelo, and M. D. Wang. 2006. Single-Molecule Analysis of RNA Polymerase Transcription. *Annual Review of Biophysics and Biomolecular Structure* 35:343-360.