Study on the Fidelity of Biodevice T7 DNA Polymerase

Ming Li[1], Zhong-Can Ou-Yang[2] and Yao-Gen Shu[1,2]
[1]*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*
[2]*CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*

Keywords: Biodevice, DNA Polymerase, Fidelity, 1st-order Terminal Effects.

Abstract: We proposed a comprehensive kinetic model of steady-state copolymerization and obtain analytical solution of the high replication fidelity of the biodevice DNA polymerase. Our analytical calculations definitively show that the neighbor effects are the key factor of the overall fidelity. These analytical results were further demonstrated by T7 DNAp whose fidelity (10^6) is well described by the 1st-order neighbor effect.

1 INTRODUCTION

DNA polymerase (DNAp) is an amazing biodevice, its template-directed DNA replication is the most important reaction in cells, and high replication fidelity is crucial to maintain the genetic stability of cells. The replication process is catalyzed by DNAp which has two domains. One is a polymerase (P site) which can add correct units (nucleotides forming Waston-Crick base pair with the template) to the reactive end of the growing DNA chain with a much higher efficiency than incorrect ones. The other domain is an exonuclease (E site) which can excise the ending unit of the growing chain once it's peeled off the template and transferred from P to E. It is believed that both domains contribute to the overall fidelity of the copolymerization process significantly. But how they cooperate is not yet quantitatively understood.

The kinetic proofreading mechanism correctly points out that the replication fidelity of the biodevice is not determined thermodynamically by the free energy difference, but kinetically by the incorporation rate difference, between the match and the mismatch. Though the detailed matching is very complex, DNA replication can be approximately regarded as a binary copolymerization process of matched nucleotides (denoted as A for convenience) and mismatched nucleotides (denoted as B). Based on the kinetics of steady-state copolymerization with higher-order terminal effect(Shu et al., 2015) which is a hot topic in macromolecule due to the alternate depolymerization step of two monomers such as A and B, we have expanded above theory into template-

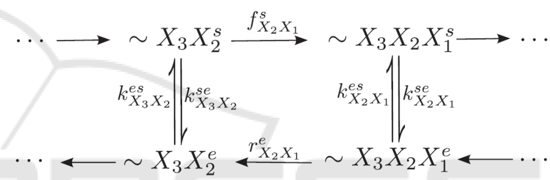

Figure 1: The minimal scheme of the first-order proofreading model(Song et al., 2017). X^s, X^e represents the state of DNAp when the primer terminus is in the synthesis (s) site or the exonuclease (e) one respectively. When the primer terminus is in the exonuclease site, one does not need to distinguish between $\sim$A^e($\sim$B^e). However, it is still convenient to use $\sim$A^e($\sim$B^e) to denote the immediate state when the terminus switches back to the polymerase site. By setting all the excision rates equal to r^e, we obtain the models for real DNAp. Under the steady-state conditions, the dNTP addition rate can be expressed as $f^s_{X_2X_1} = k_{X_2X_1}[X_1]$, where $k_{X_2X_1}$ is a pseudo-first-order rate constant, $[X_1]$ is the concentration of the incoming dNTP (to calculate the intrinsic fidelity, one often sets [A]=[B]). The rates of sliding of the primer terminus X_2X_1 into the exonuclease and polymerase active site are designated $k^{se}_{X_2X_1}$ and $k^{es}_{X_2X_1}$ respectively.

copolymerization, such as DNA replication, including higher order neighbor effect and proofreading(Song et al., 2017). The quantitative understanding of high fidelity DNA polymerase was highlighted by https://jphysplus.iop.org/2017/01/26/a-quantitative-understanding-of-high-fidelity-dna-polymerase/, however, the mathematical derivation in this 15 pages′ article is too sophisticated to be understood by chemist and biologist, as well as there are too many assumptions as emphasized in Section 3.2 (bio-relevant conditions).

In this paper, we focus on the simplest situation, the 1st-order neighbor effect, and derive a general

Li, M., Ou-Yang, Z-C. and Shu, Y-G.
Study on the Fidelity of Biodevice T7 DNA Polymerase.
DOI: 10.5220/0006636101350139
In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018) - Volume 3: BIOINFORMATICS*, pages 135-139
ISBN: 978-989-758-280-6
Copyright © 2018 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

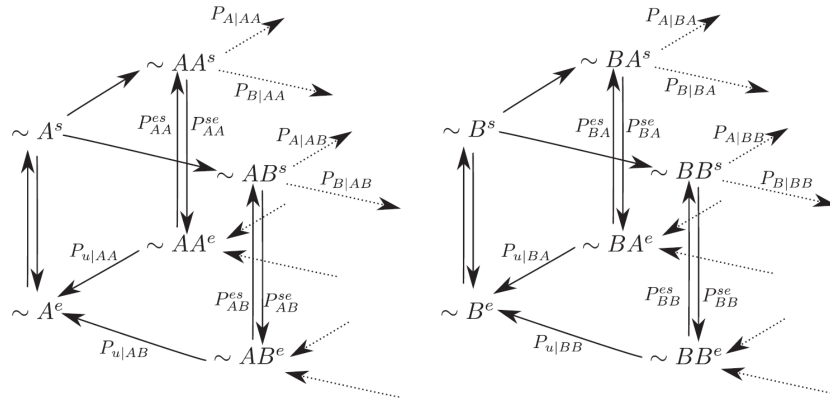

Figure 2: Branching model for the first-order polymerization and excision(Song et al., 2017).

formula of the high replication fidelity of the biodevice such as T7 DNAp. For readability of the paper, we first simply introduce the first-order proofreading model(Song et al., 2017) in Section 2, then derive the general formula of the DNAp fidelity in first-order proofreading model in section 3, and estimate the fidelity of T7 DNAp in section 4. Finally, we will discuss the results in section 5 and hope the general formula is instructive for experimental fidelity verification of other DNAp.

2 FIRST-ORDER PROOFREADING MODEL

The match between the incoming nucleotide dNTP and the template (i.e. the canonical Watson-Crick base pairings such as A-T and G-C) in the replication process plays a central role for any organism to maintain its genome stability, whereas mismatch (8 non-canonical Watson-Crick base pairings such as A-A, A-C, A-G, T-T, T-C, T-G, C-C and G-G) may introduce genetic mutation, and thus the error rate of replication must be kept very low. In living cells, the replication fidelity is controlled mainly by DNAp which catalyzes the template-directed DNA synthesis. For simplification, matched or mismatched dNTP is represented by A or B respectively throughout this paper. The superscript s or e means that the primer terminus is in the polymerase (i.e. synthesis) site or the exonuclease site, respectively.

In this section, we will discuss the general first-order proofreading model Fig.1 to demonstrate the basic ideas of our approach. Following the same logic of steady-state copolymerization kinetics in a two-component (A, B) system(Shu et al., 2015), we use $P^s_{X_n\cdots X_1}$ and $P^e_{X_n\cdots X_1}$ to denote the occurrence probability of the terminal sequence $X_n\cdots X_1$ in the synthesis (polymerase) and excision (exonuclease) site respectively, $X_i = \mathrm{A},\mathrm{B}$. $N_{X_n\cdots X_1}$ is defined as the total number of sequence $X_n\cdots X_1$ appearing in the primer chain.

The overall incorporation rate of sequence $X_n\cdots X_2X_1 (n \geq 2)$ is defined as $\dot{N}_{X_n\cdots X_2X_1} \equiv J_{X_n\cdots X_2X_1} = J^s_{X_n\cdots X_2X_1} + J^e_{X_n\cdots X_2X_1}$, where $J^s_{X_n\cdots X_2X_1} = f^s_{X_2X_1}P^s_{X_n\cdots X_2}$, $J^e_{X_n\cdots X_2X_1} = -r^e_{X_2X_1}P^e_{X_n\cdots X_2X_1}$. The kinetic equations of $P^m_{X_n\cdots X_2X_1} (n \geq 1, m = s, e)$ can be written as,

$$\left.\begin{aligned} \dot{P}^s_{X_n\cdots X_2X_1} &= J^s_{X_n\cdots X_2X_1} - \tilde{J}^s_{X_n\cdots X_2X_1*} - J^{se}_{X_n\cdots X_2X_1} \\ \dot{P}^e_{X_n\cdots X_2X_1} &= J^e_{X_n\cdots X_2X_1} - \tilde{J}^e_{X_n\cdots X_2X_1*} + J^{se}_{X_n\cdots X_2X_1} \end{aligned}\right\} \quad (1)$$

where $\tilde{J}^s_{X_n\cdots X_1*} = J^s_{X_n\cdots X_1A} + J^s_{X_n\cdots X_1B}$, $\tilde{J}^e_{X_n\cdots X_1*} = J^e_{X_n\cdots X_1A} + J^e_{X_n\cdots X_1B}$, $J^{se}_{X_n\cdots X_2X_1} = k^{se}_{X_2X_1}P^s_{X_n\cdots X_2X_1} - k^{es}_{X_2X_1}P^e_{X_n\cdots X_2X_1}$. And $P^s_{X_i\cdots X_1} = P^s_{AX_i\cdots X_1} + P^s_{BX_i\cdots X_1}$, $J^s_{X_i\cdots X_1} = J^s_{AX_i\cdots X_1} + J^s_{BX_i\cdots X_1} (i \geq 1)$ and so on.

The steady state is defined as $\dot{P}^m_{X_n\cdots X_2X_1} = 0$ for any $n \geq 1$. To analytically solve these coupled equations, we extend the logic of steady-state copolymerization kinetics and propose the following factorization conjecture(Shu et al., 2015):

$$P^m_{X_n\cdots X_2X_1} = \prod_{i=3}^{n} P^s_{X_iX_{i-1}} \left[\prod_{i=3}^{n} P^s_{X_{i-1}}\right]^{-1} P^m_{X_2X_1} \quad (2)$$

where $n \geq 3$, $m = s, e$. By this factorization conjecture, the original unclosed equations can be reduced to the following closed equations of the eight basic

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

variables $P^m_{X_2X_1}(m = s, e)$(Shu et al., 2015):

$$\left.\begin{aligned} & J^e_{\mathrm{BA}} - J^e_{\mathrm{AB}} = J^{se}_{\mathrm{B}},\ \ J^s_{\mathrm{BA}} - J^s_{\mathrm{AB}} = J^{se}_{\mathrm{A}} \\ & \frac{J^s_{\mathrm{AA}} - J^{se}_{\mathrm{AA}}}{J^s_{\mathrm{BA}} - J^{se}_{\mathrm{BA}}} = \frac{P^s_{\mathrm{AA}}}{P^s_{\mathrm{BA}}}\ \ \frac{J^s_{\mathrm{AB}} - J^{se}_{\mathrm{AB}}}{J^s_{\mathrm{BB}} - J^{se}_{\mathrm{BB}}} = \frac{P^s_{\mathrm{AB}}}{P^s_{\mathrm{BB}}} \\ & \frac{J^e_{\mathrm{AA}} + J^{se}_{\mathrm{AA}}}{J^e_{\mathrm{BA}} + J^{se}_{\mathrm{BA}}} = \frac{P^s_{\mathrm{AA}}}{P^s_{\mathrm{BA}}}\ \ \frac{J^e_{\mathrm{AB}} + J^{se}_{\mathrm{AB}}}{J^e_{\mathrm{BB}} + J^{se}_{\mathrm{BB}}} = \frac{P^s_{\mathrm{AB}}}{P^s_{\mathrm{BB}}} \\ & J^{se}_{\mathrm{A}} + J^{se}_{\mathrm{B}} = 0,\quad \sum_{X,Y=\mathrm{A,B}} (P^s_{XY} + P^e_{XY}) = 1. \end{aligned}\right\} \tag{3}$$

3 THE FIDELITY OF DNA REPLICATION WITHIN FIRST-ORDER PROOFREADING

Here, we only discuss the kinetic-based fidelity, since it can be rigorously defined and calculated within the framework of our basic theory. We define the fidelity as $\varphi = N_{\mathrm{A}}/N_{\mathrm{B}}$. N_{A} is the total number of incorporated matches in the primer, N_{B} is the total number of mismatches. Once the steady-state kinetic equations such as equations (3) are solved numerically or analytically, the total flux $J_{\mathrm{A}}(= J^s_{\mathrm{A}} + J^e_{\mathrm{A}})$,$J_{\mathrm{B}}(= J^s_{\mathrm{B}} + J^e_{\mathrm{B}})$ can be calculated. Since $\dot{N}_{\mathrm{A}} = J_{\mathrm{A}}$, $\dot{N}_{\mathrm{B}} = J_{\mathrm{B}}$ and $\mathrm{d}(N_{\mathrm{A}}/N_{\mathrm{B}})\mathrm{d}t = 0$ (in steady state), we can calculate the replication fidelity exactly by $\varphi = N_{\mathrm{A}}/N_{\mathrm{B}} = J_{\mathrm{A}}/J_{\mathrm{B}}$. However, it is often impossible to solve the kinetic equations analytically. To circumvent this problem, we introduce below an alternative method, the infinite-state Markov chain method, to calculate φ.

The first-order proofreading scheme can be rewritten as a branching model shown in Fig.2. The steady-state growth of primer can be completely characterized by four groups of transition probabilities:

$$\begin{aligned} P_{X|X_2X_1} &\equiv \frac{\delta_{X_1X}}{(\delta_{X_1\mathrm{A}} + \delta_{X_1\mathrm{B}})(1 + \beta_{X_2X_1})}, \\ P^{se}_{X_2X_1} &\equiv \frac{\beta_{X_2X_1}}{1 + \beta_{X_2X_1}}, \\ P^{es}_{X_2X_1} &\equiv \frac{\alpha_{X_2X_1}}{1 + \alpha_{X_2X_1}}, \\ P_{u|X_2X_1} &\equiv 1 - P^{es}_{X_2X_1} = \frac{1}{1 + \alpha_{X_2X_1}}. \end{aligned}$$

where $\beta_{X_2X_1} = k^{se}_{X_2X_1}/[f^s_{\mathrm{AA}}(\delta_{X_1\mathrm{A}} + \delta_{X_1\mathrm{B}})]$ with $\delta_{X_2X_1} = f^s_{X_2X_1}/f^s_{\mathrm{AA}}$, while $\alpha_{X_2X_1} = k^{es}_{X_2X_1}/r^e_{X_2X_1}$. Since any incorporated nucleotide (either A or B) has a chance to be excised, only those not being excised account for the final composition of the primer. Thus the fidelity for the first-order terminal model can be defined as,

$$\varphi \equiv \frac{Q_{\mathrm{AA}} + Q_{\mathrm{BA}}}{Q_{\mathrm{AB}} + Q_{\mathrm{BB}}} \tag{4}$$

where$Q_{X_2X_1}$ is the probability that X_1 is added to the terminal X_2 and never being excised, satisfying $Q_{\mathrm{AA}} + Q_{\mathrm{AB}} + Q_{\mathrm{BA}} + Q_{\mathrm{BB}} = 1$. $Q_{X_2X_1}$ can be explicitly expressed as $Q_{X_2X_1} \equiv \hat{P}_{X_2X_1} P_{nuX_2X_1}$, where $\hat{P}_{X_2X_1}$ is the probability that adding X_1 to the terminal X_2, $P_{nuX_2X_1}$ is the probability of the terminal X_2X_1 never being excised. The absolute values of $\hat{P}_{X_2X_1}$ are not known *a prior*, but the following equalities obviously hold:

$$\left.\begin{aligned} \frac{\hat{P}_{\mathrm{AA}}}{\hat{P}_{\mathrm{AB}}} = \frac{P_{\mathrm{A|AA}}}{P_{\mathrm{B|AA}}} = \frac{P_{\mathrm{A|BA}}}{P_{\mathrm{B|BA}}} = \frac{f^s_{\mathrm{AA}}}{f^s_{\mathrm{AB}}}, \\ \frac{\hat{P}_{\mathrm{BA}}}{\hat{P}_{\mathrm{BB}}} = \frac{P_{\mathrm{A|AB}}}{P_{\mathrm{B|AB}}} = \frac{P_{\mathrm{A|BB}}}{P_{\mathrm{B|BB}}} = \frac{f^s_{\mathrm{BA}}}{f^s_{\mathrm{BB}}}, \end{aligned}\right\} \tag{5}$$

Considering the fact that the number of AB should equal to the number of BA in the copolymer chain, we have the following intrinsic constraint:

$$Q_{\mathrm{AB}}(= \hat{P}_{\mathrm{AB}} P_{nu}\mathrm{AB}) = Q_{\mathrm{BA}}(= \hat{P}_{\mathrm{BA}} P_{nu}\mathrm{BA}). \tag{6}$$

To calculate $P_{nuX_2X_1}$, we define $P_{euX_2X_1} \equiv 1 - P_{nuX_2X_1}$ as the probability of the terminal X_2X_1 ever being excised. $P_{euX_2X_1}$ satisfy the following iterative equations(Song et al., 2017):

$$P_{euX_2X_1} = \frac{\hat{P}_{u|X_2X_1}}{P^{se}_{X_2X_1} P^{es}_{X_2X_1}} \left(\frac{1}{G_{X_2X_1}} - \frac{1}{T_{X_2X_1}} \right). \tag{7}$$

where

$$\begin{aligned} G_{X_2X_1} &\equiv 1 - P^{es}_{X_2X_1}(\hat{P}_{A|X_2X_1} P_{euX_1A} + \hat{P}_{B|X_2X_1} P_{euX_1B}) \\ &= 1 - \frac{\alpha_{X_2X_1}(\delta_{X_1\mathrm{A}} P_{euX_1\mathrm{A}} + \delta_{X_1\mathrm{B}} P_{euX_1\mathrm{B}})}{(1 + \alpha_{X_2X_1})(1 + \beta_{X_2X_1})(\delta_{X_1\mathrm{A}} + \delta_{X_1\mathrm{B}})} \\ T_{X_2X_1} &\equiv \frac{1}{1 - P^{se}_{X_2X_1} P^{es}_{X_2X_1}} = \frac{\rho_{X_2X_1} + \alpha_{X_2X_1}\beta_{X_2X_1}}{\rho_{X_2X_1}}, \end{aligned}$$

where $\rho_{X_2X_1} = 1 + \alpha_{X_2X_1} + \beta_{X_2X_1}$. and $\hat{P}_{u|X_2X_1} = P_{u|X_2X_1} P^{se}_{X_2X_1} T_{X_2X_1}$, $\hat{P}_{X|X_2X_1} = P_{X|X_2X_1} T_{X_2X_1}$. Because of $G_{X_2X_1} = 1 - \Delta$ (see DISCUSSION section in detailed),

$$\begin{aligned} P_{euX_2X_1} &= \frac{T_{X_2X_1}}{\alpha_{X_2X_1}} \left(\frac{1}{G_{X_2X_1}} - \frac{1}{T_{X_2X_1}} \right) \\ &\approx \frac{T_{X_2X_1} - 1}{\alpha_{X_2X_1}} + \frac{T_{X_2X_1}(\delta_{X_1\mathrm{A}} P_{euX_1\mathrm{A}} + \delta_{X_1\mathrm{B}} P_{euX_1\mathrm{B}})}{(\rho_{X_2X_1} + \alpha_{X_2X_1}\beta_{X_2X_1})(\delta_{X_1\mathrm{A}} + \delta_{X_1\mathrm{B}})} \\ &= \frac{\beta_{X_2X_1}}{\rho_{X_2X_1}} \left[1 + \frac{\delta_{X_1\mathrm{A}} P_{euX_1\mathrm{A}} + \delta_{X_1\mathrm{B}} P_{euX_1\mathrm{B}}}{\beta_{X_2X_1}(\delta_{X_1\mathrm{A}} + \delta_{X_1\mathrm{B}})} \right]. \end{aligned} \tag{8}$$

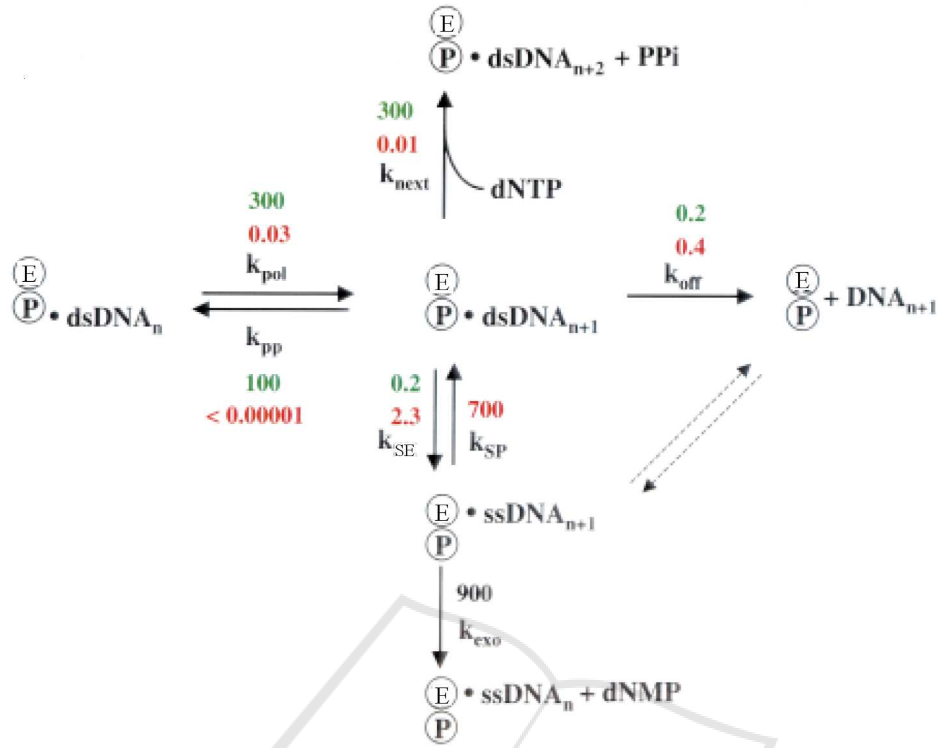

Figure 3: Kinetic scheme for proofreading by T7 DNAp(Kunkel and Bebenek, 2000). The rates (all per second) are shown in green for correct base pairs and in red for incorrect base pairs. The rates of sliding into the exonuclease and polymerase active site are designated k_{SE} and k_{SP}, respectively, where E and P correspond to the exonuclease and polymerase active sites. The k_{exo} value is the rate of excision of ssDNA. Comparing with the the minimal scheme of the first-order proofreading model as shown in Fig.3, the corresponding kinetic key parameters are: $f^s_{\mathrm{AA}} \approx 300$ s^{-1}, $f^s_{\mathrm{AB}} \approx 0.03$ s^{-1}, $f^s_{\mathrm{BA}} \approx 0.01$ s^{-1}, $f^s_{\mathrm{BB}} \approx 0.0$ s^{-1}, $k^{se}_{\mathrm{AB}} \approx 2.3$ s^{-1}, $k^{se}_{\mathrm{AA}} \approx 0.2$ s^{-1}, $k^{es}_{\mathrm{A}X_1} \approx 700$ s^{-1}, and $r^e_{X_2X_1} \approx 900$ s^{-1}, X_i =A,B. So that $\beta_{\mathrm{AA}} \sim 7\times 10^{-4} \ll 1$, while $\beta_{\mathrm{AB}} \sim 230 \gg 1$.

Then

$$\begin{aligned}
P_{eu\mathrm{AA}} &= \frac{\beta_{\mathrm{AA}}(\delta_{\mathrm{AA}}+\delta_{\mathrm{AB}})+\delta_{\mathrm{AB}}P_{eu\mathrm{AB}}}{(\alpha_{\mathrm{AA}}+\beta_{\mathrm{AA}})\delta_{\mathrm{AA}}+\rho_{\mathrm{AA}}\delta_{\mathrm{AB}}} && (9)\\
P_{eu\mathrm{BB}} &= \frac{\beta_{\mathrm{BB}}(\delta_{\mathrm{BA}}+\delta_{\mathrm{BB}})+\delta_{\mathrm{BA}}P_{eu\mathrm{BA}}}{(\alpha_{\mathrm{BB}}+\beta_{\mathrm{BB}})\delta_{\mathrm{BB}}+\rho_{\mathrm{BB}}\delta_{\mathrm{BA}}} && (10)\\
P_{eu\mathrm{AB}} &= \frac{\beta_{\mathrm{AB}}}{\rho_{\mathrm{AB}}}\left[1+\frac{\delta_{\mathrm{BA}}P_{eu\mathrm{BA}}+\delta_{\mathrm{BB}}P_{eu\mathrm{BB}}}{\beta_{\mathrm{AB}}(\delta_{\mathrm{BA}}+\delta_{\mathrm{BB}})}\right]\\
&= \frac{\lambda_{\mathrm{A}}+\varepsilon_{\mathrm{BA}}P_{eu\mathrm{BA}}}{\rho_{\mathrm{AB}}} && (11)\\
P_{eu\mathrm{BA}} &= \frac{\beta_{\mathrm{BA}}}{\rho_{\mathrm{BA}}}\left[1+\frac{\delta_{\mathrm{AA}}P_{eu\mathrm{AA}}+\delta_{\mathrm{AB}}P_{eu\mathrm{AB}}}{\beta_{\mathrm{BA}}(\delta_{\mathrm{AA}}+\delta_{\mathrm{AB}})}\right]\\
&= \frac{\lambda_{\mathrm{B}}+\varepsilon_{\mathrm{AB}}P_{eu\mathrm{AB}}}{\rho_{\mathrm{BA}}} && (12)
\end{aligned}$$

where $\lambda_X \equiv \beta_{X\bar{X}} + g_{\bar{X}\bar{X}}\beta_{\bar{X}\bar{X}}$, $g_{\mathrm{BB}} \equiv \delta_{\mathrm{BB}}/[(\alpha_{\mathrm{BB}} + \beta_{\mathrm{BB}})\delta_{\mathrm{BB}} + \rho_{\mathrm{BB}}\delta_{\mathrm{BA}}]$, $\varepsilon_{\mathrm{BA}} \equiv \delta_{\mathrm{BA}}(1+g_{\mathrm{BB}})/(\delta_{\mathrm{BA}}+\delta_{\mathrm{BB}})$, $g_{\mathrm{AA}} \equiv \delta_{\mathrm{AA}}/[(\alpha_{\mathrm{AA}}+\beta_{\mathrm{AA}})\delta_{\mathrm{AA}}+\rho_{\mathrm{AA}}\delta_{\mathrm{AB}}]$ and $\varepsilon_{\mathrm{AB}} \equiv \delta_{\mathrm{AB}}(1+g_{\mathrm{AA}})/(\delta_{\mathrm{AA}}+\delta_{\mathrm{AB}})$. The key variables, $P_{euX_2X_1}$, can be calculated by combining equations (9)-(12), for example,

$$P_{eu\mathrm{AB}} = \frac{\lambda_{\mathrm{A}}\rho_{\mathrm{BA}}+\lambda_{\mathrm{B}}\varepsilon_{\mathrm{BA}}}{\rho_{\mathrm{BA}}\rho_{\mathrm{AB}}-\varepsilon_{\mathrm{AB}}\varepsilon_{\mathrm{BA}}} \qquad (13)$$

4 THE FIDELITY OF T7 DNAP

In Fig.3, we list experimental values of some kinetic parameters for T7 DNAp. The dNTP concentration appearing in the pseudo-first-order rates of dNTP incorporation (i.e. the polymerization rates, see the caption) is often set as 100 μM which is the typical value under physiological conditions. In such cases, there exists huge difference in the order of magnitudes of the parameters. For example, addition of matched nucleotide at the polymerase site is very fast, and always much faster than mismatch addition, that is,$\delta_{\mathrm{AB}} \ll \delta_{\mathrm{AA}} = 1$; once mismatch happens, the addition rate of nucleotide is also very lower, and the primer terminus AB will very rapidly slide to the ex-

onuclease, which means $\delta_{BA} \ll 1$ and $\beta_{AB} \gg 1$; especially, continuous mismatch additions are impossible, i.e., $\delta_{BB} \sim 0$ and $\beta_{BB} \to \infty$. These intrinsic characteristics of high fidelity DNAp enable us to suggest reasonable approximations to simplify the above calculation and obtain explicit mathematical expressions of φ in terms of some key parameters. for example,

$$\begin{aligned} P_{eu\mathrm{AB}} &\approx 1 - \frac{1+\alpha_{\mathrm{AB}}}{\beta_{\mathrm{AB}}} \\ P_{eu\mathrm{AA}} &= \frac{\beta_{\mathrm{AA}} + \delta_{\mathrm{AB}} P_{eu\mathrm{AB}}}{\alpha_{\mathrm{AA}}} \sim 10^{-4} \\ P_{eu\mathrm{BB}} &= \frac{\beta_{\mathrm{BB}} + P_{eu\mathrm{BA}}}{\rho_{\mathrm{BB}}} \sim 1 \end{aligned}$$

then

$$\begin{aligned} P_{nu\mathrm{AB}} &\approx \frac{1+\alpha_{\mathrm{AB}}}{\beta_{\mathrm{AB}}} \\ P_{nu\mathrm{AA}} &\approx 1 \\ P_{nu\mathrm{BB}} &\approx 0 \end{aligned}$$

A very simple expression of the replication fidelity can be derived:

$$\begin{aligned} \varphi &\approx 1 + \frac{Q_{\mathrm{AA}}}{Q_{\mathrm{AB}}} = 1 + \frac{1}{\delta_{\mathrm{AB}}} \frac{P_{nu\mathrm{AA}}}{P_{nu\mathrm{AB}}} \\ &\approx 1 + \left[\frac{f^s_{\mathrm{AA}}}{f^s_{\mathrm{AB}}}\right] \left[\frac{r^e}{r^e + k^{es}_{\mathrm{AB}}} \frac{k^{se}_{\mathrm{AB}}}{f^s_{\mathrm{BA}}}\right] \approx 10^6, \end{aligned} \tag{14}$$

which can be divided into two parts: the first one is contributed by the P site: $\varphi_s \equiv f^s_{\mathrm{AA}}/f^s_{\mathrm{AB}} \approx 10^4$, while the second one is contributed by E site: $\varphi_e \equiv k^{se}_{\mathrm{AB}} r^e / [(k^{es}_{\mathrm{AB}} + r^e) f^s_{\mathrm{BA}}] \approx 10^2$, thus, the overall fidelity $\varphi = \varphi_s \varphi_e \approx 10^6$, which is consistent with the experimental result in vitro.

5 DISCUSSION

We proposed a comprehensive kinetic model and obtain analytical solution of the high replication fidelity of biodevice DNA polymerase. The basic assumption is that there can be nearest (1st) neighbor interactions in the copolymerization process. Our analytical calculations definitively show that the neighbor effects (reflected in the kinetic rate parameters) are the key factor of the overall fidelity. Considering the nearest neighbor effect, if the P site can add a correct unit to the correct terminus with a much faster rate than adding an incorrect one, the ratio of these two rates can be very large, meaning that the P site contributes significantly to the overall fidelity. When an incorrect unit is incorporated, the E site may discard it if the unstable terminus is transferred from P to E quickly enough before the incorrect unit is buried by the next incorporation of correct unit. In this way, the E site can also make a significant contribution to fidelity. These analytical results were further demonstrated by T7 DNAp whose fidelity (10^6) is well described by the 1st-order neighbor effect.

It must be pointed that the high fidelity DNAp maybe has intrinsic mechanism: the addition of matched nucleotide at the polymerase site is very fast, and always much faster than mismatch addition ($\delta_{AB} \ll \delta_{AA} = 1$); once mismatch happens, the addition rate of nucleotide is also very lower, and the primer terminus AB will very rapidly slide to the exonuclease ($\delta_{BA} \ll 1$ and $\beta_{AB} \gg 1$); especially, continuous mismatch additions rarely occours ($\delta_{BB} \sim 0$ and $\beta_{BB} \to \infty$).

The parameters of k_{BA} are not needed because of the intrinsic constraint (6). The number of needed kinetic parameters shown in Fig.3 also can be cut back if the normalization $\sum Q_{X_2X_1} = 1$ is engaged.

ACKNOWLEDGEMENTS

The authors thank the financial support by the National Natural Science Foundation of China (No.11574329, 11774358, 11322543, 11105218, 11675180, 11421063, 11647601, 11675017), The Fundamental Research Funds for the Central Universities (No.2017EYT24), The Joint NSFC-ISF Research Program(No.51561145002), CAS Interdisciplinary Term(No.2060299) and QYZDY-SSW-SYS008.

REFERENCES

Kunkel, T. A. and Bebenek, K. (2000). DNA replication fidelity. *Annu. Rev. Biochem.*, 69:497–529.

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

Song, Y. S., Shu, Y. G., Zhou, X., Ou-Yang, Z. C., and Li, M. (2017). Proofreading of DNA polymerase: a new kinetic model with higher-order terminal effects. *J.Phys.: Condens. Matter*, 29:025101.