THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

pubs.acs.org/JPCL

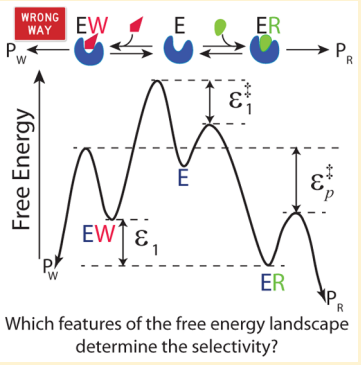
Accuracy of Substrate Selection by Enzymes Is Controlled by Kinetic Discrimination

Kinshuk Banerjee,[†] Anatoly B. Kolomeisky,*[,†,‡] and Oleg A. Igoshin*[,†,¶]

[†]Center for Theoretical Biological Physics, Rice University, PO Box 1892, MS-654, Houston, Texas, 77251-1892, United States
[‡]Department of Chemistry, Rice University, PO Box 1892, MS-60, Houston, Texas, 77251-1892, United States
[¶]Department of Bioengineering, Rice University, PO Box 1892, MS-142, Houston, Texas, 77251-1892, United States

Supporting Information

ABSTRACT: Enzymes have the remarkable ability to select the correct substrate from the pool of chemically similar molecules. The accuracy of such a selection is determined by differences in the free-energy profiles for the right and wrong reaction pathways. Here, we investigate which features of the free-energy landscape govern the variation and minimization of selectivity error. It is generally believed that minimal error is affected by both kinetic (activation barrier heights) and thermodynamic (binding stability) factors. In contrast, using first-passage theoretical analysis, we show that the steady-state selectivity error is determined only by the differences in transition-state energies between the pathways and is independent of the energies of the stable complexes. The results are illustrated for two common catalytic mechanisms: (i) the Michaelis−Menten scheme and (ii) an error-correcting kinetic proofreading scheme with tRNA selection and DNA replication as guiding biological examples. Our theoretical analysis therefore suggests that the selectivity mechanisms are always kinetically controlled.



Enzymes are biological catalysts that are essential for all processes in living organisms. They exhibit extraordinary accuracy in selecting for the right (cognate) substrate and against the wrong (near/noncognate) ones. For example, the enzymes involved in all the stages of biological information processing—replication, transcription, and translation—show remarkable fidelity.[1,2] It is widely believed that the selectivity of an enzyme is due to the difference in the free-energy profiles for two types of substrates giving desirable and undesirable products. Given finite (and often small) free-energy differences for chemically similar substrates, there must be a lower limit of the error, η, the enzymes can achieve.

Interestingly, there are contrasting views on the conditions under which the minimal error is achieved. Several studies, including the seminal work by Hopfield on enzyme selectivity,[3] suggest that the minimum error, η_{min}, is obtained when the catalytic rate tends to zero. On the other hand, some other investigations, e.g., the study on copolymerization by Bennett,[4] argue for the opposite case: the lowest error is achieved for very fast catalytic rates. A recent work by Sartori and Pigolotti[5] reconciled these opposing results by suggesting that there are two mechanisms of enzymatic discrimination: kinetic (difference in activation barrier heights dominates) and energetic (difference in binding stability of intermediates dominates). Taking the Michaelis−Menten (MM) enzyme kinetic scheme, it was shown that for kinetic discrimination, η_{min} is obtained at very fast catalytic rate, whereas for energetic discrimination, η_{min} is attained when the catalytic rate tends to zero. The Hopfield approach takes equal binding rate constants for right (R) and wrong (W) substrates but different equilibrium constants. Because there is no difference in barrier heights, the discrimination is energetic. The Bennett scheme, on the other hand, with different binding and dissociation rate constants but the same equilibrium constant (zero difference in intermediate stability), employs the kinetic discrimination. Hence, both the situations are limiting cases of a more general selectivity mechanism in which binding and dissociation rate constants as well as the equilibrium constants can be different between the R and W pathways.[5]

However, there is a crucial simplifying assumption in the study of Sartori and Pigolotti,[5] which is the equality of the catalytic rate constants for both right and wrong pathways. Such assumptions are also present in more general models of enzyme accuracy[6,7] that consider kinetic proofreading (KPR), a nonequilibrium error-correcting mechanism in biological processes. This is the mechanism that was proposed independently by Hopfield[3] and Ninio[8] to explain the strikingly low errors in different biological polymerization processes, and it was later verified experimentally in different biochemical systems.[9−11] In contrast to these theoretical views, experimental kinetic data on biological KPR networks, e.g., in DNA replication[11] and peptide chain elongation,[12] suggest that catalytic rates are significantly higher for the right pathway. Hence, the theoretical study of the selectivity mechanisms even in the basic MM scheme should be more realistic by considering these rates different. Similar considerations can

Received: February 22, 2017
Accepted: March 21, 2017
Published: March 21, 2017

DOI: 10.1021/acs.jpclett.7b00441
J. Phys. Chem. Lett. 2017, 8, 1552−1556

be applied to the classical Hopfield–Ninio (HN) KPR model[3] where only the dissociation rate constants of the intermediates are taken to be different between the two pathways. Then, for given differences in the free-energy profiles between the right and wrong pathways, we ask the following questions: (i) What features of the free-energy landscapes in such complex biochemical systems determine the value of the steady-state selectivity error? (ii) How does the error change with the kinetic parameters, and how many different patterns of error variation can be observed? (iii) Out of various theoretical possibilities, which patterns of error variation are realized in living systems?

We theoretically analyze the enzymatic selectivity by considering the overall process as a first-passage problem.[13] The standard definition of selectivity error, η, is given as the ratio of the steady-state flux of wrong product formation to that of the right product formation.[3] Given that each final catalytic step is assumed to be irreversible, i.e., end states are absorbing states, and that the enzyme is reset to the original free state following these reactions, we can compute this ratio using splitting probabilities. In this methodology, the error will be equal to the ratio of the splitting probabilities of reaching the respective end products starting from free enzyme state E (see Figure 1a). These probabilities are determined from the time-integral of the respective first-passage probability densities to reach either end before reaching the other. Given that in steady state flux to each end-state will scale with the splitting probability, the first-passage approach yields the same expression for error as obtained from the steady-state fluxes. We note that while here we focus on the steady-state error, additional dynamic properties of the system can also be obtained by studying time-evolution equations of the first-passage probability densities (also called the backward master equations).[13,14]

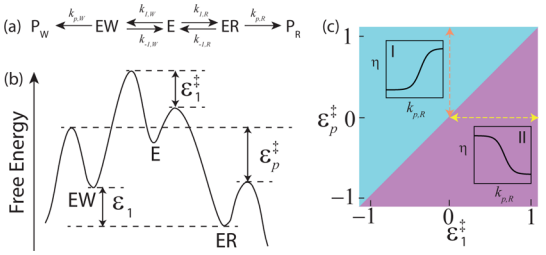


Figure 1. (a) Schematic view of the reaction network with right (R) and wrong (W) substrates being transformed into respective products by enzyme E. The catalysis is based on MM kinetics with distinct rate constants for all the steps between the R and W pathways. (b) Free-energy profile of the network in panel a with the energy differences between the pathways highlighted. The difference in transition-state energies are denoted by $\varepsilon_j^{\ddagger}$ (j = 1, p) and ε_1 is the binding stability difference of the intermediates. (c) Phase diagram showing two different regimes of error, η, variation with the catalytic rate constant $k_{p,R}$ (f_p fixed) as predicted from eq 2. The Hopfield-like case lies along the vertical orange dashed arrow, whereas the pattern for the Bennett scheme lies along the horizontal yellow dashed arrow.

We begin our analysis with the MM scheme shown in Figure 1a. For this scheme our analytical calculations give the following expression for the selectivity error in terms of the rate constants (see Figure 1a)

$$\eta = \left(\frac{k_{p,W}/K_{M,W}}{k_{p,R}/K_{M,R}}\right) = f_1 f_p \left(\frac{k_{p,R} + k_{-1,R}}{f_p k_{p,R} + f_{-1} k_{-1,R}}\right) = f_1 \left(1 + \frac{\frac{f_p}{f_{-1}} - 1}{1 + \frac{f_p}{f_{-1}}\frac{k_{p,R}}{k_{-1,R}}}\right) \quad (1)$$

where $f_i = k_{i,W}/k_{i,R}$ ($i = \pm 1, p$) and $K_{M,R/W} = \frac{k_{-1,R/W} + k_{p,R/W}}{k_{1,R/W}}$ is the MM constant. The factors f_i play an important role in our analysis. They are related to the free-energy discriminations (in units of k_BT) between the respective states of the two pathways (see Figure 1b): $f_1 = e^{-\varepsilon_1^{\ddagger}}$, $f_{-1} = e^{(\varepsilon_1 - \varepsilon_1^{\ddagger})}$, and $f_p = e^{(\varepsilon_1 - \varepsilon_p^{\ddagger})}$. The limits of error when catalysis rate is low ($k_{p,S} \ll k_{-1,S}$, S = R/W, f_p fixed) and high ($k_{p,S} \gg k_{-1,S}$, f_p fixed) are given by

$$\eta_L = \frac{f_1 f_p}{f_{-1}} = e^{-\varepsilon_p^{\ddagger}}; \ \eta_H = f_1 = e^{-\varepsilon_1^{\ddagger}}; \ \frac{\eta_L}{\eta_H} = \frac{f_p}{f_{-1}} = e^{(\varepsilon_1^{\ddagger} - \varepsilon_p^{\ddagger})} \quad (2)$$

Three important conclusions can be made by analyzing eqs 1 and 2. First, the error depends on only the values of transition-state energy differences (kinetic factors) $\varepsilon_j^{\ddagger}(j = 1, p)$ and is independent of ε_1, i.e., $\frac{\partial \eta}{\partial \varepsilon_1} = 0$. This can be seen from the fact that both numerator and denominator in the ratios $\frac{f_p}{f_{-1}}$ and $\frac{k_{p,R}}{k_{-1,R}}$ have the same ε_1 scaling making the ratios invariant to its changes (see the Supporting Information for more discussion). We point out that we have assumed equal frequency (pre-exponential) factors for all the rate constants. However, our conclusion will remain valid even if they are different but their ratio is independent of ε_1. Second, the error changes monotonically with the catalytic rate constant $k_{p,R}$ (with other parameters including f_p fixed); this can be easily seen from the last expression in eq 1. The sign of the slope is given by the sign of $\left(\frac{f_p}{f_{-1}} - 1\right)$ or equivalently the sign of $\varepsilon_1^{\ddagger} - \varepsilon_p^{\ddagger}$. Third, the limiting values of the error are given by $\eta_{min} = \eta_L(\eta_H)$ for $\varepsilon_1^{\ddagger} < \varepsilon_p^{\ddagger}$ ($\varepsilon_1^{\ddagger} > \varepsilon_p^{\ddagger}$). These properties of the error variation are shown in a phase digram in Figure 1c. It can be easily seen from eq 1 that the same monotonic behavior of error is obtained as a function of $k_{-1,R}$ (keeping f_{-1} fixed) but with a flip between increasing and decreasing behaviors and between low and high value limits.

To reconcile these results with others, we note that the Bennett scheme is reproduced from eq 2 when $\varepsilon_1^{\ddagger} > 0$, $\varepsilon_p^{\ddagger} = 0 = \varepsilon_1$; hence, η_{min} is achieved at very high $k_{p,S}$. The corresponding pattern is found along the horizontal yellow dashed arrow in Figure 1c. In contrast, one recovers the Hopfield scenario by setting $\varepsilon_1^{\ddagger} = 0$, $\varepsilon_p^{\ddagger} = \varepsilon_1(>0)$ (the vertical orange dashed arrow in Figure 1c). Then, η_{min} is obtained when $k_{p,S}$ tends to zero, as expected. It is important to note that in this special case, the kinetic discrimination $\varepsilon_p^{\ddagger}$ coincides with the thermodynamic discrimination ε_1. Here and in what follows, we use the term "thermodynamic discrimination" in the same meaning as in "energetic discrimination" used by Sartori and Pigolotti.[5] We emphasize that general independence of the error value on the energies of ER and EW complexes, and hence on their difference, indicates *kinetic discrimination always controls the selectivity error* for the MM scheme.

DOI: 10.1021/acs.jpclett.7b00441
J. Phys. Chem. Lett. 2017, 8, 1552–1556

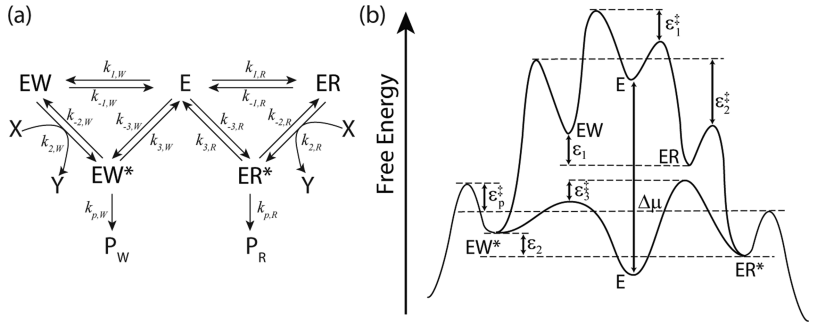


Figure 2. (a) Schematic view of the general one-loop KPR reaction network. X is some energy-currency molecule (like ATP), and Y is its hydrolyzed product (like ADP+Pi). (b) Corresponding free-energy landscape with the energy discriminations between the pathways shown. The chemical potential difference over the cycles $\Delta\mu = \mu_X - \mu_Y$.

To check the validity of these conclusions in a more general setting, we analyze a single-loop proofreading scheme that comes with two intermediates, as shown in Figure 2a. It can also be straightforwardly analyzed using the first-passage technique. The corresponding free-energy landscape highlighting the energy differences in various steps is depicted in Figure 2b. The discrimination factors defined as $f_i = k_{i,W}/k_{i,R}$, ($i = \pm 1, \pm 2, \pm 3, p$) can be expressed in terms of the discrimination energy parameters: $f_2 = e^{(\varepsilon_1 - \varepsilon_2^\ddagger)}$, $f_{-2} = e^{(\varepsilon_2 - \varepsilon_2^\ddagger)}$, $f_3 = e^{(\varepsilon_2 - \varepsilon_3^\ddagger)}$, $f_{-3} = e^{-\varepsilon_3^\ddagger}$, and $f_p = e^{(\varepsilon_2 - \varepsilon_p^\ddagger)}$ ($f_{\pm 1}$ are defined as in the MM case). Taking step 2 and step 3 to be strongly driven forward ($k_{-3,S} \ll k_{1,S}$, $k_{-2,S} \ll k_{3,S}$, S = R/W), we obtain the following expression for the error (for general expression, see the Supporting Information)

$$\eta = f_1 f_2 f_p \left(\frac{\gamma + k_{3,W}/k_{-2,W}}{\gamma + k_{3,R}/k_{-2,R}}\right)\left(\frac{k_{-1,R} + k_{2,R}}{k_{-1,W} + k_{2,W}}\right)\left(\frac{k_{3,R} + k_{p,R}}{k_{3,W} + k_{p,W}}\right) \tag{3}$$

The quantity γ is related to the chemical potential difference, $\Delta\mu$ (in units of k_BT), over the R (or W) cycle by $\gamma = \prod_{i=1}^{3} \frac{k_{i,R}}{k_{-i,R}} = \prod_{i=1}^{3} \frac{k_{i,W}}{k_{-i,W}} = e^{\Delta\mu}$. Because $\Delta\mu$ is the same for the two cycles, discrimination factors are constrained by $\prod_i \frac{f_i}{f_{-i}} = 1$.

As in MM case, both the approximate value of error (eq 3) and the exact expression (given in the Supporting Information) are invariant with respect to changes in stability of ER, EW, ER*, and EW* as long as difference in transition-state energies $\varepsilon_i^\ddagger$ and $\Delta\mu$ are fixed, i.e., $\frac{\partial\eta}{\partial\varepsilon_1} = \frac{\partial\eta}{\partial\varepsilon_2} = 0$ (see the Supporting Information for details). Moreover, it follows from eq 3 that error variation is a monotonic function of $k_{p,R}$ (compare with eq 1, f_p fixed). The sign again is determined by the sign of $\varepsilon_3^\ddagger - \varepsilon_p^\ddagger$. The ratio of the high and low $k_{p,R}$ limits comes out as (see the Supporting Information) $\eta_L/\eta_H = f_p/f_3 = e^{(\varepsilon_3^\ddagger - \varepsilon_p^\ddagger)}$. For $\varepsilon_3^\ddagger < \varepsilon_p^\ddagger$, one has $\eta_{min} = \eta_L$, whereas for $\varepsilon_3^\ddagger > \varepsilon_p^\ddagger$, $\eta_{min} = \eta_H$ (see eq 2 for comparison). Thus, only transition-state energy differences dictate the value of the error and conditions under which minimal error is achieved. In the HN model of proofreading,[3] the authors assumed $f_p = 1$, $f_3 > 1$. Thus, in this case, the minimum error can be obtained only when the catalytic rate is very low.

For both the MM scheme (without proofreading) and the general KPR scheme, the error was a monotonic function of catalytic rate constant. However, for some other reaction transitions, the KPR scheme can allow for a nonmonotonic error variation. This would imply that the minimum error can occur at some intermediate value of the rate constant. Specifically, the variation of the rate constant $k_{2,R}$ (keeping f_2 and other rates fixed) leads to three limiting values of the error: η_L ($k_{2,S} \ll k_{-1,S}$, $\gamma \approx 1$), η_M ($k_{2,S} < k_{-1,S}$, $\gamma \gg 1$), and η_H ($k_{2,S} \gg k_{-1,S}$, $\gamma \gg 1$). Explicit expressions are given by

$$\eta_L = f_p f_{-3}\left(\frac{k_{3,R} + k_{p,R}}{k_{3,W} + k_{p,W}}\right); \quad \frac{\eta_L}{\eta_M} = \frac{f_3}{f_{-2}} = e^{\varepsilon_{23}^\ddagger}; \quad \frac{\eta_H}{\eta_M} = \frac{f_{-1}}{f_2} = e^{\varepsilon_{21}^\ddagger}; \quad \frac{\eta_H}{\eta_L} = \frac{f_1}{f_{-3}} = e^{\varepsilon_{31}^\ddagger} \tag{4}$$

Here, $\varepsilon_{23}^\ddagger = \varepsilon_2^\ddagger - \varepsilon_3^\ddagger$, $\varepsilon_{21}^\ddagger = \varepsilon_2^\ddagger - \varepsilon_1^\ddagger$, and $\varepsilon_{31}^\ddagger = \varepsilon_3^\ddagger - \varepsilon_1^\ddagger = \varepsilon_{21}^\ddagger - \varepsilon_{23}^\ddagger$. Values of the ratios of these bounds, which are again *controlled only by the transition-state energy differences*, lead to six basic patterns of error variation as a function of $k_{2,R}$. The resulting phase diagram of error variation is shown in Figure 3. The regions I−VI correspond to different relationships of transition-state energy differences. Similar nonmonotonic patterns are obtained as a function of $k_{-1,R}$ (with fixed f_{-1}) but, of course, in an opposite manner (not shown here).

In the HN model, it was assumed that only the dissociation rate constants of the intermediates ($k_{-1,S}$, $k_{3,S}$, S = R/W) are different between the two pathways.[3,7] This leads to $\varepsilon_1^\ddagger = 0 = \varepsilon_3^\ddagger$, $\varepsilon_2^\ddagger = \varepsilon_1 = \varepsilon_2$, and subsequently $\eta_L = \eta_H > \eta_M$. Thus, the pattern exhibited by the HN model lies along the boundary between regions I and II, i.e., along the $\varepsilon_{21}^\ddagger = \varepsilon_{23}^\ddagger > 0$ line (the black dashed double-headed arrow in Figure 3).

Next, we consider specific biological proofreading networks to explore which types of error variation patterns may be realized in nature. In the elongation stage of protein translation, ribosome decodes aminoacyl(aa)-tRNAs with high accuracy.[2,15] Detailed kinetic data are available for various steps of the corresponding reaction network in *Escherichia coli*.[10] Specifically, the scheme employed by Zaher and Green,[12] focusing on key steps that discriminate between cognate and near-cognate aa-tRNAs, maps nicely into our general KPR scheme shown in Figure 2a. Step 2 in this case is the GTP-hydrolysis step which plays a crucial role in regulating the accuracy.[16] The corresponding data for wild-type (WT) *E. coli* ribosome are listed in Table 1 in the Supporting Information. We take $k_{-2,R} = k_{-3,R} = 10^{-3}\ s^{-1}$ to ensure that both step 2 and step 3 are nearly irreversible.[12] Because $f_{-1} > f_2$ (see Table 1 in the Supporting Information), according to eq 4, η_M *is always less than* η_H. This suggests that patterns belonging to regions III−V are absent for

DOI: 10.1021/acs.jpclett.7b00441
J. Phys. Chem. Lett. 2017, 8, 1552−1556

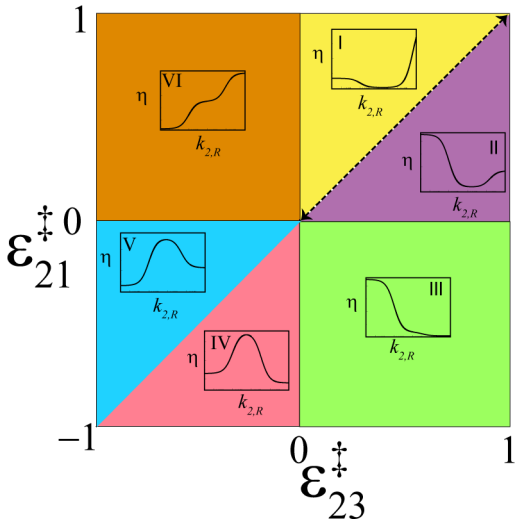


Figure 3. Phase diagram displaying six regions of distinct error variation patterns. The curves are generated as a function of $k_{2,R}$ (fixed f_2) with the following choice of model rate parameters: $k_{1,R} = 5.0$, $k_{-1,R} = 50.0$, $k_{3,R} = 1.0$, $k_{-2,R} = 10^{-3}$, $k_{-3,R} = 10^{-3}$, and $k_{p,R} = 10^{-2}$ (all in s^{-1}). The energetic discriminations are taken as (in units of k_BT) $\varepsilon_1 = 4.0$, $\varepsilon_2 = 5.0$, and $\varepsilon_2^{\ddagger} = 5.5$. The curves representing the patterns for regions I–III are determined by setting $\varepsilon_3^{\ddagger} = 3.0$ and varying the parameter $\varepsilon_1^{\ddagger}$. For the curves showing the patterns for regions IV–VI, we set $\varepsilon_3^{\ddagger} = 6.5$. The black dashed double-headed arrow along the boundary of regions I and II represents the pattern for the HN model.

this system. The limit η_L depends on the choice of the free parameter f_{-2} (f_{-3} then gets fixed by the constraint of equal $\Delta\mu$). It follows from eq 4 that, for $\frac{f_2 f_3}{f_{-1}} < f_{-2} < f_3$, the qualitative pattern of error variation belongs to region I in Figure 3. According to the data for WT ribosome, this means a broad range of f_{-2} (4×10^{-3} to 7.9) over which the system will show such a pattern. For $f_{-2} < \frac{f_2 f_3}{f_{-1}}$, the pattern shifts to that in region II. In contrast, for $f_{-2} > f_3$, the pattern belongs to region VI. These features are verified by the plots shown in Figure 4a. However, one does not expect f_{-2} to be significantly greater than 1. Therefore, the system is more likely to show patterns of either region I or II. For the sake of completeness, we mention that, as $f_p < f_3$, the minimum in error as a function of the catalytic rate constant (with fixed f_p) is obtained in the $k_{p,R} \to 0$ limit.

We apply a similar approach to the KPR network for the DNA replication in bacteriophage T7 by T7 DNA polymerase (DNAP) enzyme.[11] Here, variation of the polymerization rate constant $k_{1,R}$ results in three bounds of errors similar to the GTP hydrolysis step in the tRNA selection (for detailed expressions, see the Supporting Information). Our calculations predict that for this system the transition-state energy differences also govern the pattern of error variation. Corresponding kinetic data[17] (see Table 2 in the Supporting Information) imply that the pattern should be qualitatively similar to that for region II in Figure 3. The plot in Figure 4b shows that this is indeed the case.

In this study, we developed a quantitative theoretical method to investigate the mechanisms of selectivity in biological processes. More specifically, we investigated how the free-energy landscapes control the enzymatic selectivity error in nonequilibrium steady state. To this end, the MM scheme and the Hopfield–Ninio KPR reaction scheme were analyzed for a general set of parameters without any simplifying assumptions on the rate constants. For both schemes, the values of error are shown to depend on only the values of the transition-state energies and are invariant to changes in energy of stable enzyme–substrate complexes. Moreover, we showed that the error changes monotonically as a function of the catalytic rate constants and that the sign of the change is given by the differences in the transition-state energy values. Thus, the highest selectivity limit (lowest error) is *always* determined by kinetic discrimination. Therefore, in general, the error correction, with or without proofreading, is fully determined by kinetic discrimination factors. Thermodynamic discrimination arises only as a special case when the kinetic and thermodynamic discrimination factors coincide. Furthermore, in our generalized KPR network, we find that error can be a nonmonotonic function of the rates of other reaction steps. The pattern of error variation in such cases is again governed by transition-state energy differences, and multiple behaviors can be found. Taking important biological KPR networks as guiding examples, we show which type of error variation patterns are present in living systems. Thus, this theoretical analysis clarifies some important features of the enzymatic selectivity mechanisms in biological systems.

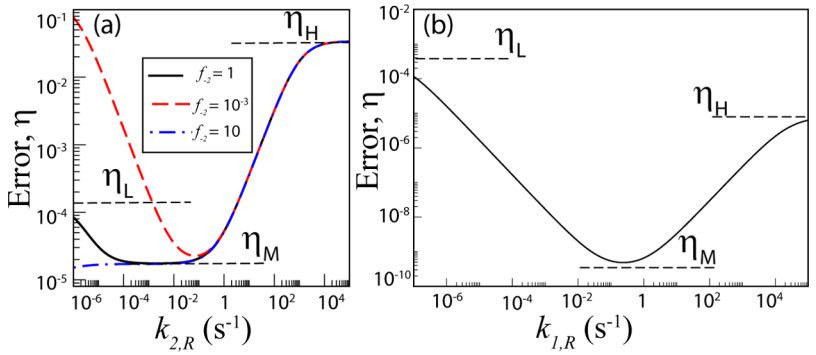


Figure 4. (a) Changes of error, η, for WT *E. coli* ribosome as a function of the hydrolysis rate constant $k_{2,R}$ for three different choices of f_{-2}. Error varies between the bounds as predicted in eq 4 (shown for the $f_{-2} = 1.0$ case). (b) Variation of error as a function of the polymerization rate constant $k_{1,R}$ for DNA replication by T7 DNAP enzyme.

DOI: 10.1021/acs.jpclett.7b00441
J. Phys. Chem. Lett. 2017, 8, 1552–1556

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.7b00441.

Additional computational methods (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: tolya@rice.edu.

*E-mail: igoshin@rice.edu.

ORCID

Anatoly B. Kolomeisky: 0000-0001-5677-6690

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is supported by Center for Theoretical Biological Physics NSF Grant PHY-1427654. A.B.K. also acknowledges the support from the Welch Foundation (Grant C-1559) and from the NSF (Grant CHE-1360979).

■ REFERENCES

(1) Kunkel, T. A.; Bebenek, K. DNA Replication Fidelity. *Annu. Rev. Biochem.* **2000**, *69*, 497−529.

(2) Zaher, H. S.; Green, R. Fidelity at the Molecular Level: Lessons from Protein Synthesis. *Cell* **2009**, *136*, 746−762.

(3) Hopfield, J. J. Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 4135−4139.

(4) Bennett, C. H. Dissipation-error tradeoff in proofreading. *BioSystems* **1979**, *11*, 85−91.

(5) Sartori, P.; Pigolotti, S. Kinetic versus energetic discrimination in biological copying. *Phys. Rev. Lett.* **2013**, *110*, 188101.

(6) Murugan, A.; Huse, D. A.; Leibler, S. Speed, dissipation, and error in kinetic proofreading. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 12034−12039.

(7) Hartich, D.; Barato, A. C.; Seifert, U. Nonequilibrium sensing and its analogy to kinetic proofreading. *New J. Phys.* **2015**, *17*, 055026.

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

(9) Hopfield, J. J.; Yamane, T.; Yue, V.; Coutts, S. M. Direct experimental evidence for kinetic proofreading in amino acylation of tRNAIle. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 1164−1168.

(10) Gromadski, K. B.; Rodnina, M. V. Kinetic Determinants of High-Fidelity tRNA Discrimination on the Ribosome. *Mol. Cell* **2004**, *13*, 191−200.

(11) Johnson, K. A. Conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.* **1993**, *62*, 685−713.

(12) Zaher, H. S.; Green, R. Hyperaccurate and Error-Prone Ribosomes Exploit Distinct Mechanisms during tRNA Selection. *Mol. Cell* **2010**, *39*, 110−120.

(13) van Kampen, N. G. *Stochastic processes in physics and chemistry*, 3rd ed.; North Holland Publishing Company: Amsterdam, 2007.

(14) Kolomeisky, A. B. *Motor Proteins and Molecular Motors*; CRC Press, Taylor and Francis Group: Boca Raton, FL, 2015.

(15) Johansson, M.; Zhang, J.; Ehrenberg, M. Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 131−136.

(16) Wohlgemuth, I.; Pohl, C.; Mittelstaet, J.; Konevega, A. L.; Rodnina, M. V. Evolutionary optimization of speed and accuracy of decoding on the ribosome. *Philos. Trans. R. Soc., B* **2011**, *366*, 2979−2986.

(17) Wong, I.; Patel, S. S.; Johnson, K. A. An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry* **1991**, *30*, 526−537.

Supporting information: Accuracy of substrate selection by enzymes is controlled by kinetic discrimination

Kinshuk Banerjee[1], Anatoly B. Kolomeisky[1,2*], Oleg A. Igoshin[1,3*]

1 Center for Theoretical Biological Physics, Rice University
2 Dept. of Chemistry, Rice University
3 Dept. of Bioengineering, Rice University

* tolya@rice.edu, igoshin@rice.edu

1 First-passage probability density: backward master equations

Here, we give the theoretical methodology to determine the selectivity error of the networks in terms of the first-passage process [1]. The quantity of interest is the first-passage probability density. To illustrate, we start with the simple Michaelis-Menten (MM) scheme, and then sketch how it is applied to the more complex kinetic proofreading network.

Let us denote $F_{\mathrm{R,i}}(t)$ as the first-passage probability density to reach the right(R) end-product state $\mathrm{P_R}$ at time t for the first time before reaching the wrong(W) end-product state $\mathrm{P_W}$, starting from some state i on the network at time $t = 0$ (see Fig. 1(a) in the main text). The corresponding probability density $F_{\mathrm{W,i}}(t)$ is defined in the same manner. The time-evolution equations of the first-passage probability densities are generally known as the backward master equations [1].

To formulate these for MM network, we define $\mathbf{F_R} = (F_{\mathrm{R,E}}, F_{\mathrm{R,ER}}, F_{\mathrm{R,EW}})^{\mathrm{T}}$ and $\mathbf{F_W} = (F_{\mathrm{W,E}}, F_{\mathrm{W,ER}}, F_{\mathrm{W,EW}})^{\mathrm{T}}$ It's time-evolution is given by

$$\frac{d}{dt}\mathbf{F_R} = \mathbf{P}\,\mathbf{F_R} + \mathbf{Q_R} \tag{1}$$

$$\frac{d}{dt}\mathbf{F_W} = \mathbf{P}\,\mathbf{F_W} + \mathbf{Q_W}. \tag{2}$$

Here, the transition matrix is given by the transition rates between the different states

$$\mathbf{P} = \begin{pmatrix} -(k_{1,R}+k_{1,W}) & k_{1,R} & k_{1,W} \\ k_{-1,R} & -(k_{-1,R}+k_{2,R}) & 0 \\ k_{-1,W} & 0 & -(k_{-1,W}+k_{2,W}) \end{pmatrix} \tag{3}$$

The source terms $\mathbf{Q_R} = (0, k_{p,R}\delta(t), 0)^{\mathrm{T}}$ and $\mathbf{Q_W} = (0, 0, k_{p,W}\delta(t))^{\mathrm{T}}$ are proportional to the Dirac-delta function $\delta(t)$.

We solve the above set of differential equations using the standard technique of Laplace transformation. For example, the Laplace transform of $F_{\mathrm{R,i}}(t)$ is defined as

$$\tilde{F}_{\mathrm{R,i}}(s) = \int_0^\infty \mathrm{e}^{-st} F_{\mathrm{R,i}}(t)dt. \tag{4}$$

Then, with the initial conditions $F_{\mathrm{R,i}}(t=0) = F_{\mathrm{R,i}}(t=0) = 0$, ($i$ = E, ER, or EW), Eqs.(1, 2) transform into a set of algebraic equations

$$s\tilde{\mathbf{F}}_{\mathbf{R}} = \mathbf{P}\,\tilde{\mathbf{F}}_{\mathbf{R}} + \tilde{\mathbf{Q}}_{\mathbf{R}} \tag{5}$$

$$s\tilde{\mathbf{F}}_{\mathbf{W}} \;=\; \mathbf{P}\,\tilde{\mathbf{F}}_{\mathbf{W}} + \tilde{\mathbf{Q}}_{\mathbf{W}}. \tag{6}$$

Here, $\tilde{\mathbf{F}}_{\mathbf{R}} = (\tilde{F}_{\mathrm{R,E}}, \tilde{F}_{\mathrm{R,ER}}, \tilde{F}_{\mathrm{R,EW}})^{\mathrm{T}}$ ($\tilde{\mathbf{F}}_{\mathbf{W}}$ is defined similarly), $\tilde{\mathbf{Q}}_{\mathbf{R}} = (0, k_{p,R}, 0)^{\mathrm{T}}$ and $\tilde{\mathbf{Q}}_{\mathbf{W}} = (0, 0, k_{p,W})^{\mathrm{T}}$. From Eq.(6), one gets the required first-passage probability densities $\tilde{F}_{\mathrm{R,E}}(s)$ and $\tilde{F}_{\mathrm{W,E}}(s)$. The probabilities to reach either of the end-states, also known as the splitting probabilities, are obtained directly from these densities in Laplace space as [1, 2]

$$\Pi_{\mathrm{R}} = \tilde{F}_{\mathrm{R,E}}(s=0); \quad \Pi_{\mathrm{W}} = \tilde{F}_{\mathrm{W,E}}(s=0). \tag{7}$$

We note that in steady states the fluxes into the right and wrong states will be scaled proportional to the respective splitting probabilities. The selectivity error is then can be computed as the ratio of these splitting probabilities

$$\eta = \Pi_{\mathrm{W}}/\Pi_{\mathrm{R}}. \tag{8}$$

For the generalized kinetic proofreading network, the backward master equations can be similarly expressed in the form of Eqs.(1, 2). For this network (denoted by the subscript kpr),

$$\mathbf{F}^{\mathbf{kpr}}_{\mathbf{R}} = (F_{\mathrm{R,E}}, F_{\mathrm{R,ER}}, F_{\mathrm{R,ER^*}}, F_{\mathrm{R,EW}}, F_{\mathrm{R,EW^*}})^{\mathrm{T}}$$

and

$$\mathbf{F}^{\mathbf{kpr}}_{\mathbf{W}} = (F_{\mathrm{W,E}}, F_{\mathrm{W,ER}}, F_{\mathrm{R,ER^*}}, F_{\mathrm{W,EW}}, F_{\mathrm{W,EW^*}})^{\mathrm{T}},$$

$\mathbf{Q}^{\mathbf{kpr}}_{\mathbf{R}} = (0, 0, k_{p,R}\delta(t), 0, 0)^{\mathrm{T}}$, $\mathbf{Q}^{\mathbf{kpr}}_{\mathbf{W}} = (0, 0, 0, 0, k_{p,W}\delta(t))^{\mathrm{T}}$. The transition matrix in this case becomes

$$\mathbf{P}^{\mathbf{kpr}} = \begin{pmatrix} -D_1 & k_{1,R} & k_{-3,R} & k_{1,W} & k_{-3,W} \\ k_{-1,R} & -D_2 & k_{2,R} & 0 & 0 \\ k_{3,R} & k_{-2,R} & -D_3 & 0 & 0 \\ k_{-1,W} & 0 & 0 & -D_4 & k_{2,W} \\ k_{3,W} & 0 & 0 & k_{-2,W} & -D_5 \end{pmatrix} \tag{9}$$

where $D_1 = (k_{1,R} + k_{1,W} + k_{-3,R} + k_{-3,W})$, $D_2 = (k_{-1,R} + k_{2,R})$, $D_3 = (k_{-2,R} + k_{3,R} + k_{p,R})$, $D_4 = (k_{-1,W} + k_{2,W})$, $D_5 = (k_{-2,W} + k_{3,W} + k_{p,W})$. The same methodology as for the MM scheme above can be used to compute splitting probabilities and the steady-state error.

2 Invariance of the error with changes in the stability difference parameters ε_i

The splitting probabilities for the Michaelis-Menten (MM) scheme follows from Eq.(7) as

$$\Pi_{\mathrm{R}} = \frac{k_{p,R}/K_{M,R}}{k_{p,R}/K_{M,R} + k_{p,W}/K_{M,W}} \tag{10}$$

and

$$\Pi_{\mathrm{W}} = \frac{k_{p,W}/K_{M,W}}{k_{p,R}/K_{M,R} + k_{p,W}/K_{M,W}}. \tag{11}$$

Here, $K_{M,R} = \frac{k_{-1,R}+k_{p,R}}{k_{1,R}}$ and $K_{M,W} = \frac{k_{-1,W}+k_{p,W}}{k_{1,W}}$ are the MM constants for the right and wrong product formation. The rate constants $k_{1,R}$ ($k_{1,W}$) are assumed to be pseudo-first-order, i.e. defined as $k_{1,R} = k^0_{1,R}[R]$ ($k_{1,W} = k^0_{1,W}[W]$), where $k^0_{1,R}$ and $k^0_{1,W}$ are the second-order binding rate constants. Then, using Eq.(8), Eq.(10) and Eq.(11), one gets the expression of error for the MM scheme

$$\eta = f_1 f_p \left(\frac{k_{-1,R} + k_{p,R}}{f_{-1}k_{-1,R} + f_p k_{p,R}} \right) = f_1 \left(\frac{1 + \frac{k_{-1,R}}{k_{p,R}}}{1 + \frac{f_{-1}}{f_p}\frac{k_{-1,R}}{k_{p,R}}} \right) \tag{12}$$

where $f_i = k_{i,W}/k_{i,R}$, $(i = \pm 1, p)$. The factors f_i play important roles in our analysis. They are related to the free energy discriminations (in units of $k_{\mathrm{B}}T$) between the respective states of the two pathways (see Fig. 1(b)): $f_1 = e^{-\varepsilon_1^{\ddagger}}$, $f_{-1} = e^{(\varepsilon_1 - \varepsilon_1^{\ddagger})}$, $f_p = e^{(\varepsilon_1 - \varepsilon_p^{\ddagger})}$ where we assume equal frequency (pre-exponential) factors for all the rate constants. However, our conclusions about the invariance of error against the variation of stability difference parameters ε_i will be still valid even when these pre-factors are different but their ratio is independent of ε_i. Then, f_{-1}/f_p is independent of ε_1, the difference in stability of the complexes EW and ER. Further, as $k_{-1,R/W}$ and $k_{p,R/W}$ correspond to transitions starting from the same state ER/EW, their ratio and hence η are also independent of the energy of the complex ER/EW. In other words, if we vary the parameter ε_1, *keeping all transition state energy differences the same*, then discrimination factors f_{-1} and f_p would change in the same way, *i.e.,* if f_{-1} becomes some αf_{-1}, f_p will become αf_p. At the same time, because there is no change in the transition state energies, this would lead to changes in the rate constants. For simplicity, let us first assume that we only modified the energy level of the state EW. Then the rate

constants of the right pathway are unaffected and the value of error does not change with ε_1. If in addition the energy of the state ER is also changing, this will affect $k_{-1,R}$ and $k_{p,R}$ proportionally, i.e. $k_{-1,R}$ will become $k_{-1,R}/\alpha$ and $k_{p,R}$ will become $k_{p,R}/\alpha$ (see Fig. 1(b)). Thus, it follows from Eq.(12) that, the error remains constant with the changes only in the thermodynamic discrimination, *i.e.* $\frac{\partial \eta}{\partial \varepsilon_1} = 0$.

The splitting probabilities for the generalized kinetic proofreading (KPR) network can be determined following the same methodology used for MM scheme above. The exact expression of error in the generalized KPR network, defined according to Eq.(8), is given by

$$\eta = f_p \frac{\Big((k_{p,R}+k_{3,R})(k_{-1,R}+k_{2,R})+k_{-1,R}k_{-2,R}\Big)\Big(k_{2,W}(k_{1,W}+k_{-3,W})+k_{-1,W}k_{-3,W}\Big)}{\Big((k_{p,W}+k_{3,W})(k_{-1,W}+k_{2,W})+k_{-1,W}k_{-2,W}\Big)\Big(k_{2,R}(k_{1,R}+k_{-3,R})+k_{-1,R}k_{-3,R}\Big)}$$

$$= \frac{\Big[(1+\frac{k_{p,R}}{k_{3,R}})(1+\frac{k_{2,R}}{k_{-1,R}})+\frac{k_{-2,R}}{k_{3,R}}\Big]\Big[f_{-3}\Big(1+\frac{f_2}{f_{-1}}\frac{k_{2,R}}{k_{-1,R}}(1+\frac{f_1}{f_{-3}}\frac{k_{1,R}}{k_{-3,R}})\Big)\Big]}{\Big[(\frac{f_3}{f_p}+\frac{k_{p,R}}{k_{3,R}})(1+\frac{f_2}{f_{-1}}\frac{k_{2,R}}{k_{-1,R}})+\frac{f_{-2}}{f_p}\frac{k_{-2,R}}{k_{p,R}}\Big]\Big[1+\frac{k_{2,R}}{k_{-1,R}}(1+\frac{k_{1,R}}{k_{-3,R}})\Big]} \tag{13}$$

Here, the discrimination factors are $f_1 = e^{-\varepsilon_1^\ddagger}$, $f_{-1} = e^{(\varepsilon_1-\varepsilon_1^\ddagger)}$, $f_2 = e^{(\varepsilon_1-\varepsilon_2^\ddagger)}$ $f_{-2} = e^{(\varepsilon_2-\varepsilon_2^\ddagger)}$, $f_3 = e^{(\varepsilon_2-\varepsilon_3^\ddagger)}$, $f_{-3} = e^{-\varepsilon_3^\ddagger}$, $f_p = e^{(\varepsilon_2-\varepsilon_p^\ddagger)}$. Then, it follows that all the ratios involving f's are independent of the parameters $\varepsilon_i\,(i = 1, 2)$. ε_1 gives the stability difference between the intermediates EW and ER; ε_2 denotes the same for EW* and ER* (see Fig. 2 in main text). Further, the ratio of the rate constants that correspond to transitions starting from the same state are independent of the energy of that state. For example, $k_{2,R}/k_{-1,R}$ is independent of the energy of state ER. This implies that all the ratios involving rate constants are also independent of the energy of the respective states. In this context, we clarify that the ratio $k_{1,R}/k_{-3,R} = e^{\delta_{31}^\ddagger}e^{\Delta\mu}$ where $\delta_{31}^\ddagger$ is the difference in transition state energies for step-3 and step-1 and $\Delta\mu$ is the chemical potential difference over the cycle (see Fig. 2 in main text). Now, changes in ε_1, keeping all the transition state energy differences and $\Delta\mu$ fixed, changes f_2 and f_{-1} by the same factor, say α_1, while the rates $k_{2,R}$ and $k_{-1,R}$ will be modified by the same factor $1/\alpha_1$. Then, according to Eq.(13), the error remains unchanged. Variation in the parameter ε_2 in a similar fashion alters the parameters f_3, f_{-2}, f_p by the same factor, say α_2, and correspondingly rates $k_{-2,R}$, $k_{3,R}$ and $k_{p,R}$ will change by $1/\alpha_2$. This again keeps the error constant. Thus, error in both the networks under study does not depend on thermodynamic discrimination parameters.

Table 1: Model parameters for aa-tRNA selection by WT *E. coli* ribosome

Parameter	Value (s^{-1})	Parameter	Value
$k_{1,R}$	40	f_1	0.675
$k_{-1,R}$	0.5	f_{-1}	94
$k_{2,R}$	25	f_2	4.8×10^{-2}
$k_{3,R}$	8.5×10^{-2}	f_3	7.9
$k_{p,R}$	8.415	f_p	4.2×10^{-3}

3 The limits of error for the catalytic step in the kinetic proofreading network

The two limits of the error, η for the proofreading network follow from Eq. (3) of the main text

$$\eta_{\rm L} = \left(\frac{f_1 f_2 f_p}{f_3}\right)\left(\frac{k_{-1,R}+k_{2,R}}{k_{-1,W}+k_{2,W}}\right)\left(\frac{\gamma + k_{3,W}/k_{-2,W}}{\gamma + k_{3,R}/k_{-2,R}}\right);$$

$$\eta_{\rm H} = f_1 f_2 \left(\frac{k_{-1,R}+k_{2,R}}{k_{-1,W}+k_{2,W}}\right)\left(\frac{\gamma + k_{3,W}/k_{-2,W}}{\gamma + k_{3,R}/k_{-2,R}}\right);\ \frac{\eta_{\rm L}}{\eta_{\rm H}} = \frac{f_p}{f_3} = e^{(\varepsilon_3^{\ddagger} - \varepsilon_p^{\ddagger})}. \quad (14)$$

4 Parameters for aminoacyl(aa)-tRNA selection case

We list the values of the parameters used in our model in Table 1. These are based on the experimental kinetic data of Zaher *et al.* [3].

5 Non-monotonic error variation with polymerization rate in DNA replication by T7 DNA polymerase

DNA replication in bacteriophage T7 by T7 DNA polymerase (DNAP) enzyme is a prime example where kinetic proofreading (KPR) mechanism enhances the selectivity [4]. T7 DNAP employs KPR to rectify errors due to dNTP mis-incorporation during the elongation of DNA primer over a template in the polymerase(Pol) site of the enzyme. Mismatched nucleotides

are rejected by proofreading in the exonuclease(Exo) site and subsequent hydrolysis. The corresponding proofreading network is similar to the one used for tRNA selection in main text with an important difference: after the incorporation of one dNTP molecule in step-1, the system can go on to add another dNTP or can reset via the proofreading mechanism comprising of step-2 (Pol-Exo sliding) and step-3 (hydrolysis) [5]. In other words, the steps with rate constants $k_{p,R/W}$ are now connected to state-ER/EW instead of ER*/EW*. This makes $f_p = e^{\varepsilon_1 - \varepsilon_p^\ddagger}$. All the other f_is have the same expressions as in the tRNA selection case discussed in the main text. Now, kinetic data for this system indicate that the polymerization rate constants are the same for the cognate dNTP molecule, *i.e.*, $k_{1,R} = k_{p,R}$ (see Table 2). Variation of the polymerization rate constant results in three bounds of error at low, intermediate and high $k_{1,R}$ (with f_1, f_p fixed). Analytical expressions follow

$$\eta_{\rm L} = \frac{f_p f_1}{f_{-1}}\left(\frac{1+k_{-1,R}K_{M,R}/k_{3,R}}{1+k_{-1,W}K_{M,W}/k_{3,W}}\right);\ \frac{\eta_{\rm L}}{\eta_{\rm M}} = \frac{f_2}{f_{-1}}\left(\frac{1+k_{-2,R}/k_{3,R}}{1+k_{-2,W}/k_{3,W}}\right);$$

$$\frac{\eta_{\rm L}}{\eta_{\rm H}} = \frac{f_p}{f_{-1}}\left(\frac{1+k_{-1,R}K_{M,R}/k_{3,R}}{1+k_{-1,W}K_{M,W}/k_{3,W}}\right). \tag{15}$$

Here, $K_{M,S} = \frac{k_{-2,S}+k_{3,S}}{k_{2,S}}$ $(S = R/W)$. Although the expressions in Eq.(15) are a bit complicated, a careful inspection based on the free energy profile yields the following:

$$\frac{\eta_{\rm L}}{\eta_{\rm M}} = e^{\varepsilon_{12}^\ddagger}\left(\frac{1+e^{\delta_{32,R}^\ddagger}}{1+e^{\delta_{32,W}^\ddagger}}\right);\quad \frac{\eta_{\rm L}}{\eta_{\rm H}} = e^{\varepsilon_{1p}^\ddagger}\left(\frac{1+e^{\delta_{21,R}^\ddagger}(1+e^{\delta_{32,R}^\ddagger})}{1+e^{\delta_{21,W}^\ddagger}(1+e^{\delta_{32,W}^\ddagger})}\right). \tag{16}$$

Here, $\delta_{ij,R/W}^\ddagger$ denotes the difference in energies of the transition states of step-i and step-j for R/W pathway. Therefore, again *the transition state energy differences govern the nature of error variation as a function of the polymerization rate*. For fixed $\delta_{ij}^\ddagger$s, one can get the same six patterns as shown in Fig. 3 of the main text but now as a function of $\varepsilon_{12}^\ddagger$ and $\varepsilon_{1p}^\ddagger$. The kinetic data for T7 DNAP (see Table 2) imply that $\eta_{\rm L} > \eta_{\rm M}, \eta_{\rm H}$ and $\eta_{\rm M} < \eta_{\rm H}$. Hence, the error variation pattern should be similar to that for region II in Fig. 3 of the main text. The error vs. polymerization rate constant curve plotted in Fig. 4(b) of the main text confirms this prediction.

Table 2: Model parameters for DNA replication by T7 DNAP based on the experimental kinetic data of Wong *et al.*[6]

Parameter	Value (s^{-1})	Parameter	Value
$k_{1,R}$	250	f_1	8×10^{-6}
$k_{-1,R}$	1	f_{-1}	1×10^{-5}
$k_{2,R}$	0.2	f_2	11.5
$k_{-2,R}$	700	f_{-2}	1
$k_{3,R}$	900	f_3	1
$k_{p,R}$	250	f_p	4.8×10^{-5}

References

[1] N. G. van Kampen. *Stochastic processes in physics and chemistry.* North Holland Publishing Company, Amsterdam, 3rd edition, 2007.

[2] A. B. Kolomeisky. *Motor Proteins and Molecular Motors.* CRC Press, Taylor and Francis Group, 2015.

[3] H. S. Zaher and R. Green. Hyperaccurate and Error-Prone Ribosomes Exploit Distinct Mechanisms during tRNA Selection. Molecular Cell, **39**(1), 110-120, 2010.

[4] T. A. Kunkel and K. Bebenek. DNA Replication Fidelity. Annu Rev Biochem, **69**, 497-529, 2000.

[5] K. A. Johnson. Conformational coupling in DNA polymerase delity. Annu Rev Biochem, **62**, 685-713, 1993.

[6] I. Wong, S. S. Patel, and K. A. Johnson. An induced fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. Biochemistry, **30**, 526-537, 1991.