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.