Supporting Information

Rajamani et al.

- Figure S1: Rates of incorporation and mis-incorporation ……..… S2
- Figure S2: Rates of extension from matched and mismatched termini … S3
- Mathematical description of modeling ……. S4
- Discussion of modified error threshold model in a prebiotic context ..… S9
- Discussion of connection of model parameters to experimental results .. S10
- Discussion of degradation of stalled intermediates (Modeling section 2)..S11

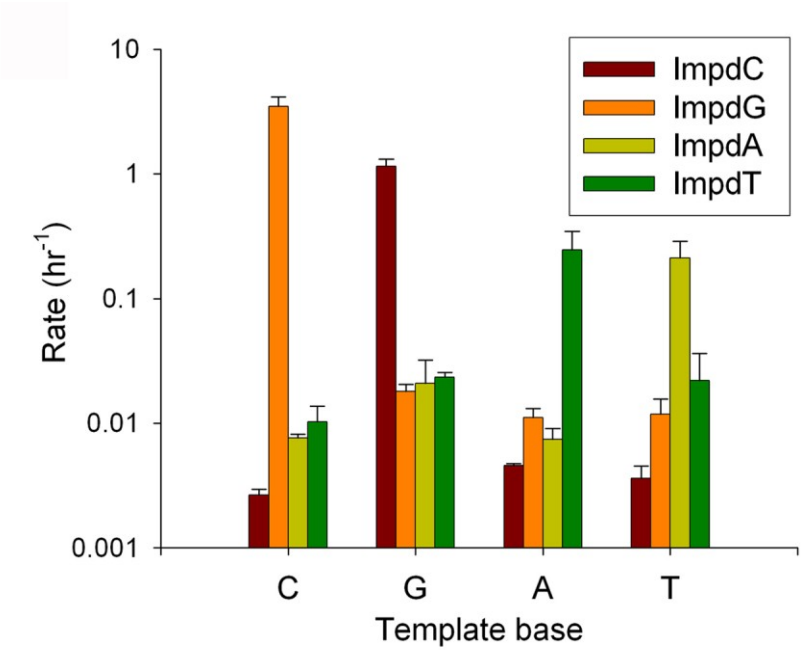

Figure S1. Rate of incorporation of different activated nucleotides (C,G,A,T) across each template base. Error bars show standard deviations calculated from duplicate sets of reactions.

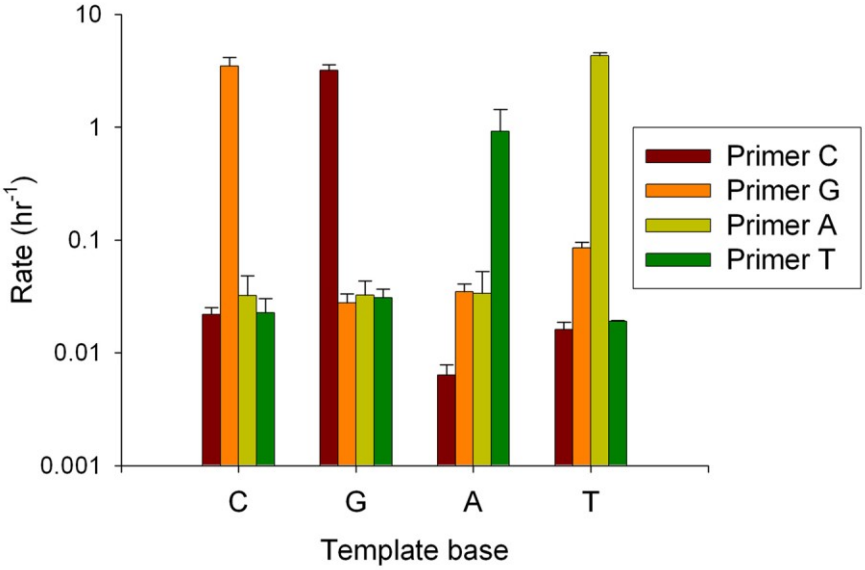

Figure S2. Rates of extension from matched or mismatched primer-template termini. Error bars show standard deviations calculated from duplicate sets of reactions.

SUPPLEMENTARY TEXT: MODELS OF THE ERROR THRESHOLD WITH STALLING AND STRAND SEPARATION

S. RAJAMANI, J. ICHIDA, T. ANTAL, D. TRECO, K. LEU, M. NOWAK, J. SZOSTAK, I. CHEN

1. Classical error threshold

First, let us briefly discuss the classical case that Eigen solved in 1971. We have self-replicating units (e.g., RNA): master units and mutant units. Master units replicate at rate $r \geq 1$, while mutants replicate at unit rate (which fixes the time scale). The copy is perfect with probability q, and there is a mistake in the copy otherwise, with probability $1-q$. A master unit with a mistake is a mutant, and a mutant with a mistake is also considered to be mutant (i.e., we neglect back mutations). We denote master units as X and mutants as Y, so these processes can be depicted as

$$\begin{aligned} X &\xrightarrow{rq} X + X \\ X &\xrightarrow{r(1-q)} X + Y \\ Y &\xrightarrow{1} Y + Y \end{aligned} \tag{1}$$

We keep the population size constant by also removing a random unit at each replication. In the large population size limit, the frequencies x and $y = 1 - x$ of the above units evolve according to

$$\begin{aligned} \dot{x} &= rqx - \Phi x \\ \dot{y} &= r(1-q)x + y - \Phi y \end{aligned} \tag{2}$$

where $\Phi = rx + y$ is chosen to keep the total population $x + y = 1$ constant. It is easy to see that the above two equations are identical, since $y = 1 - x$. In the stationary state $\dot{x} = \dot{y} = 0$, and we can express the frequency of master units as

$$x = \frac{rq - 1}{r - 1} \tag{3}$$

which is positive as long as $q > q_E^*$, where $q_E^* = 1/r$ is the classical error threshold of Eigen. This means that the copy has to be perfect with probability $q \geq q_E^*$ in order to have master units in the population at steady state. This can be understood in simple terms by noting that the master sequence effectively replicates at rate qr, which has to be larger than the mutant fitness for the

master copy to survive, which immediately leads to $q_E^* = 1/r$. For a sequence of length L and with mutation rate μ per base pair, the probability of making a perfect copy is

$$q = (1-\mu)^L \approx e^{-\mu L} \tag{4}$$

where the last expression is valid for large sequences ($L \gg 1$) and small mutation rates ($\mu \ll 1$). Now the classical error threshold can also be formulated for the critical sequence length $L^* = (\ln r)/\mu$.

2. Strand separation and explicit degradation

Now we discuss an extension of the classical model, in which if an error occurs during the copying process, the nascent sequence is completed only with probability p. Otherwise, with probability $1-p$, it separates from the template before completion (e.g., during thermocycling) and degrades. Graphically

$$\begin{aligned} X &\xrightarrow{rq} X + X \\ X &\xrightarrow{r(1-q)p} X + Y \\ Y &\xrightarrow{q} Y + Y \\ Y &\xrightarrow{(1-q)p} Y + Y \end{aligned} \tag{5}$$

The process $Y \to Y + Y$ happens at rate $q + p(1-q)$. If we rescale time by $q + p(1-q)$, Y replicates at rate one, and X replicates at rate r. Replication produces a perfect copy with probability

$$\hat{q} = \frac{q}{q + p(1-q)} \tag{6}$$

and an imperfect copy with probability $1 - \hat{q}$. Hence our model is equivalent to the original model of Eigen but with $\hat{q}$. The critical error threshold is given by $\hat{q}^* = 1/r$, which corresponds to

$$q^* = \frac{p}{r + p - 1} \tag{7}$$

This function is depicted in Figure 1 for several values of p.

In the absence of strand separation, $p = 1$, we recover Eigen's classical result of $q_E^* = 1/r$. As the strand separation becomes stronger, $p \to 0$, the error threshold tends to zero. This means that the error frequency can be arbitrarily high if mutant copies are not produced at all. We also see

from (7) that q^* is a monotonically increasing function of p. Hence the degradation of imperfect copies always lowers the error threshold. Note also that one could generalize this model by defining two distinct probabilities of strand separation and degradation: one in the absence of a mistake p', and another one in case of a mistake p''. This model, however, would be equivalent to our present model with $p = p''/p'$.

3. Strand separation and re-annealing

In this section we assume that if a mistake happens during the copying process, the nascent sequence separates from the template (e.g., during thermocycling), and then re-anneals again to either a master or a mutant unit. Completion of the nascent sequence occurs at rate a, determined by the stalling factor. We denote a stalled, incomplete mutant unit as Z, so the model is

$$\begin{aligned} X &\xrightarrow{rq} X + X \\ X &\xrightarrow{r(1-q)} X + Z \xrightarrow{a} X + Y \\ Y &\xrightarrow{q} Y + Y \\ Y &\xrightarrow{1-q} Y + Z \xrightarrow{a} Y + Y \end{aligned} \tag{8}$$

In order to keep the total population size constant (counting all X, Y and Z units), we remove a random unit from the system each time a new unit is made. The limit $a \to \infty$ corresponds to the classical case of Eigen.

In the large population size limit the frequencies of different types of units change according to the differential equations

$$\begin{aligned} \dot{x} &= rqx - \Phi x \\ \dot{y} &= qy + axz + ayz - \Phi y \\ \dot{z} &= r(1-q)x - axz + (1-q)y - ayz - \Phi z \end{aligned} \tag{9}$$

where $\Phi = rx + y$ is chosen to keep the total population size $x + y + z = 1$ constant.

The stationary state of (9) is the solution of the equations

$$\begin{aligned} 0 &= rqx - (rx + y)x \\ 0 &= qy + axz + ayz - (rx + y)y \\ 0 &= r(1-q)x - axz + (1-q)y - ayz - (rx + y)z \end{aligned} \tag{10}$$

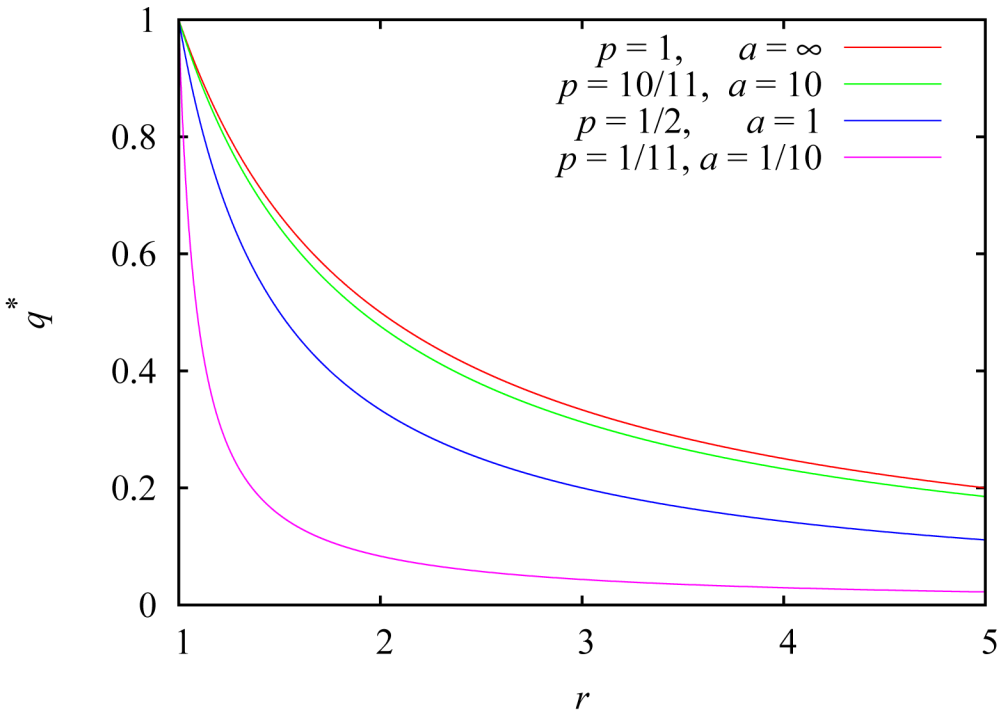

FIGURE S3. Error threshold q^* as a function of the master replication rate r, for several values of p and a. The corresponding parameter values are given by (14). The red curve is the classical Eigen result $q_E^* = 1/r$, which is recovered in the absence of strand separation $p = 1$ or $a = \infty$.

At the error threshold, $x \to 0$. In order to find the critical surface $q^*(r, a)$ we assume that $x \ll 1$ but still $x, y, z > 0$. Now the last two equations of (10) become identical in the leading order in x, which leads to

$$y = \frac{q + a}{1 + a} + \mathcal{O}(x) \tag{11}$$

Here we also used that $z = 1 - y + \mathcal{O}(x)$. Now inserting (11) into the first equation of (10), and then letting $x \to 0$, we arrive at the critical q value

$$q^* = \frac{a}{r(a + 1) - 1} \tag{12}$$

This function is depicted in Figure S3 for several values of a.

In the present model a mistake always leads to strand separation, but a separated copy can re-anneal and be completed at rate a. In the model of Section 2 a mistake leads to strand separation only with probability $1 - p$, but the separated copy degrades and never re-anneals. Despite the very different nature of these two models, they lead to equivalent error thresholds. When an imperfect

copy re-anneals in the model of this section, it happens at rate $a(x+y)$, and the copy dies at rate $\Phi = rx + y$ [see (9)]. Hence the imperfect copy gets completed with probability

$$p = \frac{a(x+y)}{a(x+y) + rx + y} \tag{13}$$

Close to the error threshold $x \to 0$, the survival probability becomes

$$p = \frac{a}{a+1} \tag{14}$$

Indeed, for such parameter values the error threshold q^* is the same for the two models (7) and (12). For example, we recover the classical result of Eigen $q^*_E = 1/r$ for $a \to \infty$ or as $p \to 1$. The optimal case for the master units is the zero error threshold, which is reached for $a \to 0$ or for $p \to 0$. One can see the critical error threshold in Figure 1 for several corresponding values (14) of p and a. We see from (7) and (12) that q^* is an increasing function of either p or a for all $r > 1$. Hence stalling and strand separation always lower the error threshold.

Discussion of modified error threshold model in a prebiotic context

Whether stalling affects the error threshold depends on the timescales of replication. In a prebiotic world, if strand separation is infrequent enough that all products, both perfect copies and mutants, would be completed within a single replication cycle (e.g., before the strands melt), then stalling would not affect the products. But if strand separation occurs before mutant copies are completed, then stalling could potentially reduce the effective rate of production of mutants relative to perfect copies, which continue to propagate while mutants are stalled. In experimental models of templated non-enzymatic polymerization of nucleic acids, the half-times range from hours to days per base [1], suggesting that the copying time for a short ribozyme (e.g., 30 bases) would be >18 hours. Prebiotically, the length of time available for replication before strand separation (τ_r) might be dictated by thermal cycling. For a diurnal cycle, τ_r would be ~12 hours assuming a rotational period of 24 hours, or approximately ~7-10 hours for the early earth with higher angular velocity [2]. For thermocycling in convection cells (e.g., deep sea hydrothermal vent), τ_r could be as short as several seconds [3]. We therefore modeled polymerization with stalling, assuming that strands separate more quickly than mutant copies are completed.

Discussion of connection of model parameters to experimental results

Close to the error threshold, master sequences constitute a very small proportion of the sequence pool. Therefore, most sequences (Y) will be replicated with time $1/L$ per base, with stalled bases requiring additional time S/L per mutation (master sequences replicate faster with an average time of $1/(rL)$ per base). Given μL mutations per sequence on average, mutant sequences would take an additional time μS to copy through stalled bases. The additional time is related to the completion of Z. Therefore, in terms of measurable parameters, $a = 1/(\mu S)$.

Discussion of degradation of stalled intermediates (Modeling section 2)

To examine the dependence of the modified error threshold on the details of our model, we also solved a model of stalling in which degradation of stalled sequences is explicit during the copying process (Modeling section 2). In this second model, we assume that imperfect copies have a higher probability of degradation during polymerization, because their copying time is longer. This corresponds to longer exposure to chemical damage or simply a higher chance of washing out from the system before copying is complete. The relative probability of degradation of imperfect copies would be the ratio of exposure times. In the terminology of the main model (Modeling section 3), the time taken to complete a perfect copy would be 1, and the time taken to complete an imperfect copy would be $1+1/a$. Therefore, the relative probability of survival for an imperfect copy would be $1/(1+1/a)$, or $a/(a+1)$. Indeed, L_s^* in the two models agreed when $p = a/(a+1)$ (Modeling section 3), indicating that the error threshold is robust to differences in the details of modeling.

(1) Kanavarioti, A.; Monnard, P. A.; Deamer, D. W. *Astrobiology* **2001**, *1*, 271-281; Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 317-322; Kawamura, K.; Ferris, J. P. *J. Am. Chem. Soc.* **1994**, *116*, 7564-7572.

(2) Laskar, J.; Joutel, F.; Robutel, P. *Nature* **1993**, *361*, 615-617.

(3) Krishnan, M.; Ugaz, V. M.; Burns, M. A. *Science* **2002**, *298*, 793; Braun, D.; Libchaber, A. *Phys. Biol.* **2004**, *1*, P1-8.