

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

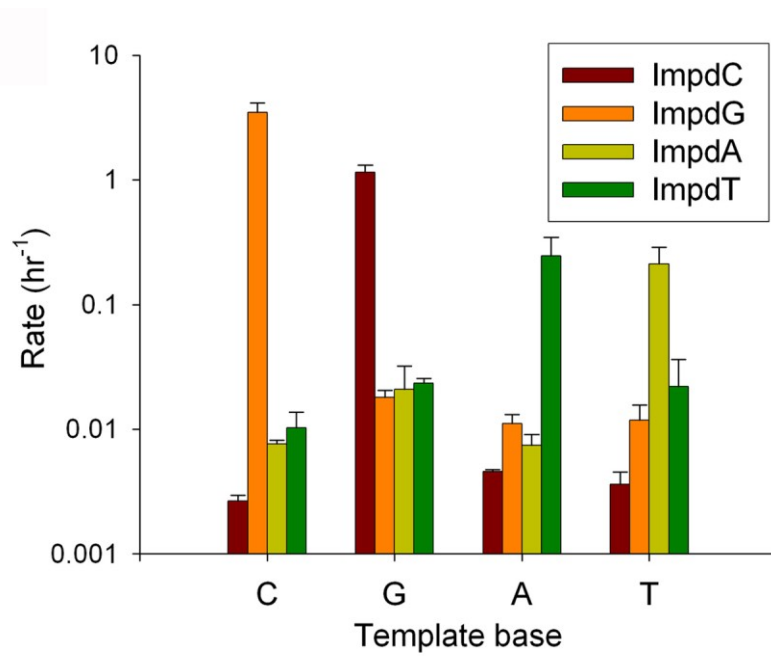


Figure S1. Rate of incorporation of different activated nucleotides (C,G,A,T) across each template base. Error bars show standard deviations calculated from duplicate sets of reactions.

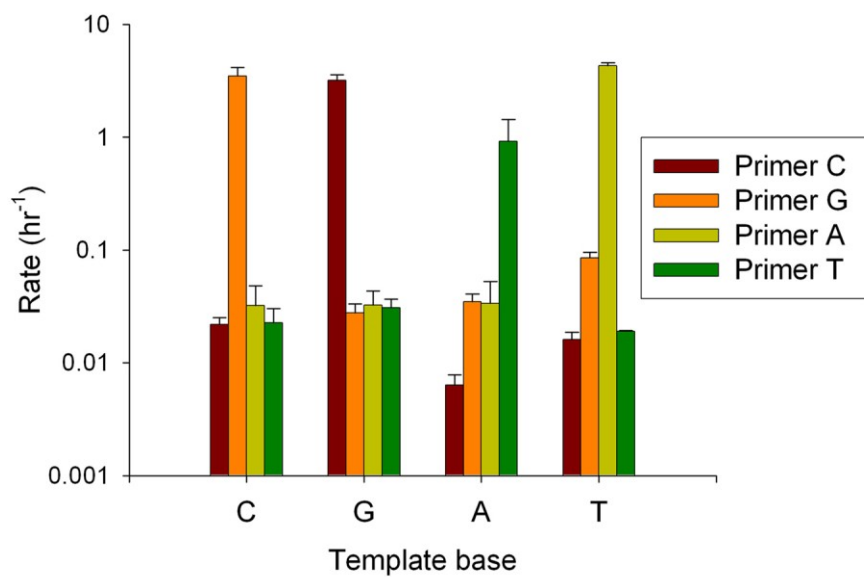


Figure S2. Rates of extension from matched or mismatched primer-template termini.

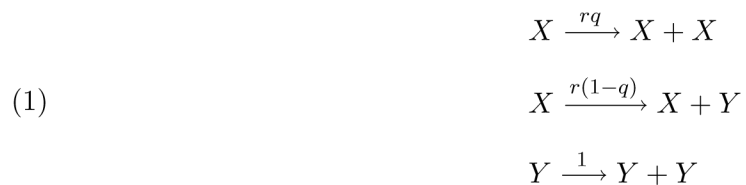
Error bars show standard deviations calculated from duplicate sets of reactions.

**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



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

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

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

$$(3) \quad x = \frac{rq - 1}{r - 1}$$

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

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 base¹, 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) might be dictated by thermal cycling. For a diurnal cycle, τ would be ~12 hours assuming a rotational period of 24 hours, or approximately ~7-10 hours for the early earth with higher angular velocity². For thermocycling in convection cells (e.g., deep sea hydrothermal vent), τ could be as short as several seconds³. 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 L mutations per sequence on average, mutant sequences would take an additional time S to copy through stalled bases. The additional time is related to the completion of Z . Therefore, in terms of measurable parameters, $a = 1/(S)$.

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.