Non-enzymatic error correction in self-replicators without extraneous energy supply

Koushik Ghosh, Parthasarthi Sahu, Shashikanta Barik, Hemachander Subramanian*

Department of Physics, National Institute of Technology Durgapur, India

Abstract

Enzymatic error correction ensures the fidelity of DNA replication, preserving biological information flow across evolutionary timescales. Error correction, as has been envisaged until now, requires the presence of sophisticated enzymes that utilize energy and/or information to correct errors. Here, we propose a single-free-parameter theoretical model for error correction in self-replicating heteropolymers that requires neither enzymes nor an extraneous energy supply. Instead, the free-energy gradient that drives the self-replication process itself is utilized for error correction. Asymmetric cooperativity, a form of kinetic asymmetry, which has been shown to lead to unidirectional strand construction on the DNA template, is vital for this error correction functionality. Despite its simplicity, the model reproduces several experimentally observed phenomena typically associated with error correction in DNA, such as stalling, fraying, speed-accuracy trade-off, next-nucleotide effects, and adaptive tuning of error rates. The model allows us to clarify and answer some confounding issues related to the thermodynamics and kinetics of error correction: What is the energy source for the enhanced base selectivity of passive DNA polymerases? What is the role of thermodynamics and kinetics of phosphodiester bond formation in error correction? We show that the catalysis of the phosphodiester bond formation by polymerases is critical and sufficient for the error correction capability of the polymerases. This observation opens up a pathway to experimentally probe the accurate synthesis of oligomers in the primordial scenario, with the sole energy input being the thermodynamic drive for templated strand construction. We thereby demonstrate the creation of persistent order from non-equilibrium, a central requirement for emergence of life.

Keywords: Non-enzymatic error correction, DNA polymerases, error correction energy supply, phosphodiester bond catalysis, self-replication, order from non-equilibrium, Markov chain Modeling

Statement of Significance

DNA replication is very accurate, committing just one error per billion base pairs. A major, initial part of this error correction, where bases are selected for incorporation, interestingly requires no energy, apart from the energy supply required for the DNA strand construction itself. Is it possible to utilize the thermodynamic drive for templated self-replication itself for error correction? We introduce a model that answers the above question in the affirmative. Surprisingly, the model shows that, under the right conditions, simply catalyzing the covalent bond formation between nucleotides results in error correction. No structure-based discriminators, such as enzymes, are needed. This model may help us understand DNA polymerase energetics better, and elucidate primordial scenarios in which accurate self-replication was possible without enzymatic help. This model may also help explain how biological order emerged from energy flow.

1. Introduction

Heritable genetic information and the occasional alteration of this information through mutations form the bedrock over which evolutionary trajectories of organisms are carved out [1, 2]. In modern cellular systems, the error rates of biochemical reactions are significantly lowered by an energy-driven enzymatic mechanism known as "kinetic proofreading"[3, 4]. In the primordial scenario, heteropolymers that self-replicated non-enzymatically possibly accumulated mutations at an unsustainable rate to carry information (probably pertaining to autocatalysis) across generations. Such high mutation rates preclude the possibility of the emergence of error-correcting enzymes, leading to the classic chicken-and-egg

*Corresponding author: hsubramanian.phy@nitdgp.ac.in

problem. How did the primordial molecular evolution escape this fate? Below, we summarize the existing solutions to this question in the literature and point out their inadequacies before providing our solution.

I. The hypercycle approach

Manfred Eigen proposed a mechanism to evade the decay of information under high mutation rates by assuming an autocatalytically closed set of self-replicators, called quasispecies, that replicated with high fidelity, as long as the mutation rate was below a certain threshold[5, 6]. This proposal allowed for the possibility of stable information transmission across generations of quasispecies, as long as the error rates are below a threshold that depends inversely on the heteropolymer length. However, it is difficult to imagine an evolutionary pathway from this hypercyclically coupled community-based error tolerance mechanism to the emergence of a single-molecule-based information storage and error correction system. This is because the hypercyclic coupling leads to an evolutionary preference for multiple error-prone, catalytically coupled self-replicators over a single, more accurate, uncoupled self-replicator. While the quasispecies model has provided valuable insights into the highly error-prone viral evolution, especially for RNA viruses, its applicability to the evolution of stable information transfer in primordial non-viral systems is less clear. This leads us again to the following question: How did single-molecule self-replicators develop an ability to correct errors without enzymatic/community help?

II. Kinetic Proofreading

Although extraneous energy-driven, it is important to discuss the paradigmatic Hopfield-Ninio scheme of enzymatic proofreading[3, 4] here, to distinguish our model from the above scheme. Hopfield-Ninio model involves a reaction topology in which at least two discriminatory dissociation paths are available to differentiate between the correct and incorrect substrates. When both of these discriminatory paths are optimally utilized by driving the reaction flow using an external thermodynamic drive, maximal error correction, that is better than the thermodynamic limit, is achieved. The thermodynamic drive, usually the hydrolysis of ATP, drives the reaction towards not just product formation, but also towards an additional dissociation path for both the correct and incorrect substrates, resulting in the expenditure of energy even for futile reaction cycles that do not result in product formation. Because the accuracy of the substrate selection improves when the incorrect substrate cycles through the reaction a large number of times, the corresponding energy expenditure for maximal error correction becomes significantly higher, particularly in the presence of high incorrect substrate concentrations [7–10]. In addition, the rate of product formation must also be low to allow the incorrect substrate to cycle through the reaction many times, thereby introducing a time lag. This requirement in favor of accuracy reduces the rate of product yield, and results in a speed-accuracy trade-off, with the maximal error correction occurring when the product yield rate tends to zero. More complex reaction topologies that are derived from the above simple scheme inherit the same issues, although regimes in the parameter space have been found where the speed is improved significantly for a tolerable loss of accuracy[7, 11, 12]. This suggests that we need models that maximize self-replication accuracy at non-zero reaction speeds, without requiring extraneous energy sources for driving the reaction and for error correction.

III. Base selection energetics

The current understanding of energetics of *in vivo* DNA replication fidelity is unclear[13–19], at best. Error correction in DNA replication is a multistage process, with the first stage being *base selection*, where the correct bases are selected over incorrect bases by the polymerase *passively* [13], without any extraneous energy expenditure. The next stages are exonuclease-mediated proofreading and post-replicative mismatch repair, which, however, are not strictly essential for cell survival[19, 20]. At the base selection stage, the error rate goes from about 1/100, thermodynamically determined by the free energy difference between the correct and incorrect base pairs in aqueous medium, to about 1/10000, as observed in exonuclease-deficient polymerase alpha and Taq polymerases[13, 19]. It has been argued that this gain in accuracy is due to the exclusion of water inside the active center of the polymerase, which increases the enthalpy difference between the correct and incorrect base pairs, improving discrimination [21, 22]. This explanation, however, violates Landauer's principle since it suggests improvement in accuracy without concomitant energy expenditure, allowing the enzyme to function as a Maxwell's Demon[23]. For instance, during the replication of a homonucleotide DNA template, in the absence of the polymerase, the replication would result in a disordered daughter strand with a considerable number of random point mutations, whereas, in the presence of the passive polymerase, a more ordered, low entropy daughter strand with fewer errors is produced, apparently without any concomitant extraneous energy expenditure. By running the templated DNA replication experiments in two tanks to completion (equilibrium), one with the passive polymerase at very low concentrations and the other without, one can generate free energy by allowing

the chemically distinct products from the two tanks to diffuse into each other, violating the first law of thermodynamics. This raises the question of energy supply for the passive base selection step during error correction[19].

IV. Other approaches

Here, we briefly review more recent attempts at demonstrating non-enzymatic self-replication and literature exploring error correction in the models of primordial self-replication. Enzyme-free template-directed polymerization has been extensively studied experimentally as a model for primordial self-replication[24, 25]. These studies reveal a critical limitation: the high error rates, frequently exceeding 10%, that arise due to the absence of inherent error correction mechanisms [26, 27]. Some experimental and theoretical non-enzymatic error correction mechanisms have recently been proposed[28–30]. One such approach, called kinetic error filtering [28], is based on the observation that primer extension slows down after a mismatch. This slowdown is included in the theoretical model to reduce the replication of erroneous strands, which can then be selectively removed by applying a limited time window for self-replication. Another mechanism proposed in [29] utilizes thermodynamics to promote the binding of correct nucleotides to the templates. This model relies on unrealistic assumptions such as the dependence of monomer addition rates on the thermodynamic stability of the primer-template complex (which is also assumed to immediately separate after completion of primer extension), exponentially slow timescales, and extremely low yields. More recently, an experimental, non-enzymatic error correction mechanism implemented using a DNA strand displacement network has been examined in [30], wherein the system utilizes a chemical fuel to drive the reaction out of equilibrium and thus synthetically emulates Hopfield-Ninio's kinetic proofreading scheme. This approach facilitates enhanced discrimination between correct and incorrect nucleotide incorporations by allowing multiple binding and unbinding events and exploiting free energy differences between correct and incorrect base pairs, just as in the kinetic proofreading scheme. An approach to avoid using external energy for error correction has been reported in [31], where spatial separation between substrate binding and product formation events is used to impose time delays that prevent wrong substrates from reaching the product formation site, improving accuracy rates. This scheme, however, requires setting up of a concentration gradient of substrates and enzymes, which does require external energy. Despite their potential, these mechanisms predominantly focus on static thermodynamic principles while overlooking the dynamic aspects of kinetic discrimination. Furthermore, the dependence on external energy sources for error correction raises concerns regarding the viability of these mechanisms in prebiotic conditions, where such organized energy inputs were likely unavailable [32–34].

V. The kinetic origin of persistent order

More generally, the search for a model system that autonomously creates order from non-equilibrium supply of nutrients and energy, that persists on a timescale longer than the timescale of the energy supply, is still ongoing[35–38]. Demonstrating that such a system can be spontaneously created in the primordial scenario, *de novo*, would provide us confidence in our belief of spontaneous emergence of life on early Earth. It would also enable us to investigate and understand the physicochemical mechanisms for creating persistent order from free energy flow and implement them in self-repairing biomimetic systems. The model we introduce here provides such a minimal system, albeit theoretical, to explore the emergence of long-lasting order in self-replicating systems that evolved from the chemical non-equilibrium that probably existed in the primordial oceans. Just as broken symmetry in equilibrium systems introduces long-range spatial order, for example in magnets, broken symmetry in our non-equilibrium model system leads to the creation of temporally persistent order, achieved through error correction, with the assistance of kinetic control of reaction steps.

In the following, we theoretically demonstrate that it is possible to construct a non-enzymatically self-replicating molecular system that corrects errors using the thermodynamic drive for self-replication itself, without any extraneous energy supply. This model also addresses the issues raised in the above paragraphs. The central ingredient of our model is *asymmetric cooperativity*, a kinetic property introduced earlier[39, 40] to understand the counterintuitive evolutionary choice of unidirectional replication of DNA strands on a template. We take the existence of asymmetric cooperativity in the self-replicating system considered below as a *premise*, and proceed to derive the consequences of its presence for the fidelity of self-replication. Interestingly, the ability of the self-replicator to correct errors depends entirely on the *unidirectionality* of daughter strand construction on the template, which, when annulled, annuls the fidelity of self-replication. The model system improves the thermodynamically imposed error rate of 1/100 to about $1/(4 \times 10^4)$, equal to the error rates observed in passive polymerases, without using extraneous energy sources. We show that this system reproduces experimental signatures regularly associated with error correction, such as (a) stalling, where strand construction halts temporarily after the incorporation of an incorrect base pair, even during non-enzymatic DNA replication, which is shown here to be a natural outcome of our model premises[41–43]; (b) trade-off between speed and accuracy, documented in various other investigations of enzymatic error correction in biological systems [44–47]; (c) unzipping or fraying of primer end when an incorrect base pair forms at the elongation front[48]; (d) next nucleotide effects[49], where incorporation of another correct base pair after an incorrect base pair is observed

to stabilize the incorrect one and hence increase the error rate, and, additionally, an increased error rate at locations near an incorrect base pair[42], and (e) adaptive tuning of error rate, experimentally observed in model organisms under nutritional stress[50, 51]. We would like to make it clear that, by thermodynamic drive, we imply the propensity for base pair formation over base pair dissociation, due to temperatures lower than the dsDNA-to-ssDNA transition temperature and/or high nucleotide concentrations. The drive also includes the propensity for covalent bond formation between intrastrand nucleotides, which is assumed here to be irreversible, due to its high exergonicity. It has to be noted that this model is intended to capture the dynamics of base selection alone. The exploration of the origins of later stages, exonuclease-assisted proofreading and mismatch repair, is beyond the scope of this model.

2. Asymmetric Cooperativity Model

For completeness, we provide a brief summary of the asymmetric cooperativity model, introduced in[39, 40]. Asymmetric cooperativity is defined as an asymmetric *kinetic* influence of a pre-existing base pair (hydrogen bond) between a lone nucleotide and a template strand on the formation or dissociation of neighboring base pairs between other nucleotides and the template, to its left and right. Specifically, in the model, we assume that a base pair between a lone nucleotide and the template strand reduces the kinetic barrier for the formation/dissociation of a new base pair towards the $5'$-end of the template, and increases the kinetic barrier for formation/dissociation of a base pair towards the $3'$-end. This kinetic property helps to simultaneously satisfy two conflicting requirements for rapid daughter strand construction: a low kinetic barrier for the induction and easy incorporation of nucleotides to the base pair's $5'$-end, thereby enhancing the base pairing propensity with a template nucleotide, and a high kinetic barrier to the $3'$-end to retain the base-paired nucleotides on the template strand, facilitating the formation of intra-strand covalent bonds that extend the nascent daughter strand[39]. This directional asymmetry in the kinetic influence of a base pair on its left and right neighbors results in the unidirectional daughter strand construction, and has been shown to improve the rate of self-replication of template [39, 40]. This phenomenon is illustrated in Fig.1. In [40], we termed this as sequence-*independent* asymmetric cooperativity. Here, we simply call it asymmetric cooperativity, since in this article there is no need to invoke sequence-*dependent* asymmetric cooperativity. *The central premise of this article is that the above-mentioned asymmetry cooperativity, a form of kinetic asymmetry [52], is present in correct (complementary) base pairing, whereas it is absent in incorrect (non-complementary) base pairing.* To be precise, we assume that complementary, correct base pairs reduce the kinetic barrier of the $5'$-end neighbor and increase the barrier of the $3'$-end neighbor, whereas incorrect base pairs neither reduce nor increase the kinetic barriers for base pair formation/dissociation in either direction.

The mechanism of error correction within our model can be succinctly explained as follows: Let us assume that an incorrect base pair forms at the location m, as shown in Figure(2). According to the central premise of our article, this base pair does not influence the kinetic barrier of the already formed base pair at the $(m-1)^{th}$ location or for the formation of a new base pair at the $(m+1)^{th}$ location. Therefore, the growth of the new daughter strand is stalled, as the barrier towards the template's $5'$-end is not reduced. This provides more time for the incorrect base pair to dissociate, which is also catalyzed by the reduced barrier at the mth location due to the catalytic effect of the base pair at the $(m-1)^{th}$ location. Moreover, since the base pair at the $(m-1)^{th}$ location is not stabilized due to the reduction in *its* kinetic barrier caused by the $(m-2)^{nd}$ base pair and the lack of inhibition from the mth incorrect base pair, it is also prone to dissociate, which can precipitate sequential unzipping of the rest of the bonded segment due to reduced barriers at the Y-fork. This stalling and unzipping behind the growth front significantly delays the strand construction process. This time delay between the correctly and incorrectly formed base pairs can be utilized for error correction if the rate of incorporation of the correct base pair in the growing daughter strand through intrastrand covalent bond formation is fast enough. At very low covalent bond formation rates, the ratio between the probability of the covalent bond formation incorporating the wrong base pair into the growing strand and the probability of the covalent bond incorporating a correct base pair, are nearly equal, since at the long times required for the slow covalent bonds to form, the time lag between the correct and incorrect strand formation vanishes. Only when the rate of covalent bond formation approaches either of the two base pairing rates, the probabilities of correct and wrong base pair stabilization through covalent bonding become distinguishable, and error correction, possible. This idea is schematically illustrated in Fig.7. Since the covalent bond formation is a thermodynamically nearly irreversible event, if its rate of formation is just as fast or faster than the formation rate of a stretch of cooperatively coupled correct base pairs, the covalent bond kinetics will ensure that only correct base pairs are irreversibly incorporated into the daughter strand. *Thus, we come to the significant and highly consequential conclusion that catalysis of covalent bond formation is enough to introduce error correction capability in the DNA replication process, without the need for any extraneous energy supply.* This also explains and reproduces the experimentally observed improvement in accuracy during base selection that happens apparently without any extraneous energy expenditure.

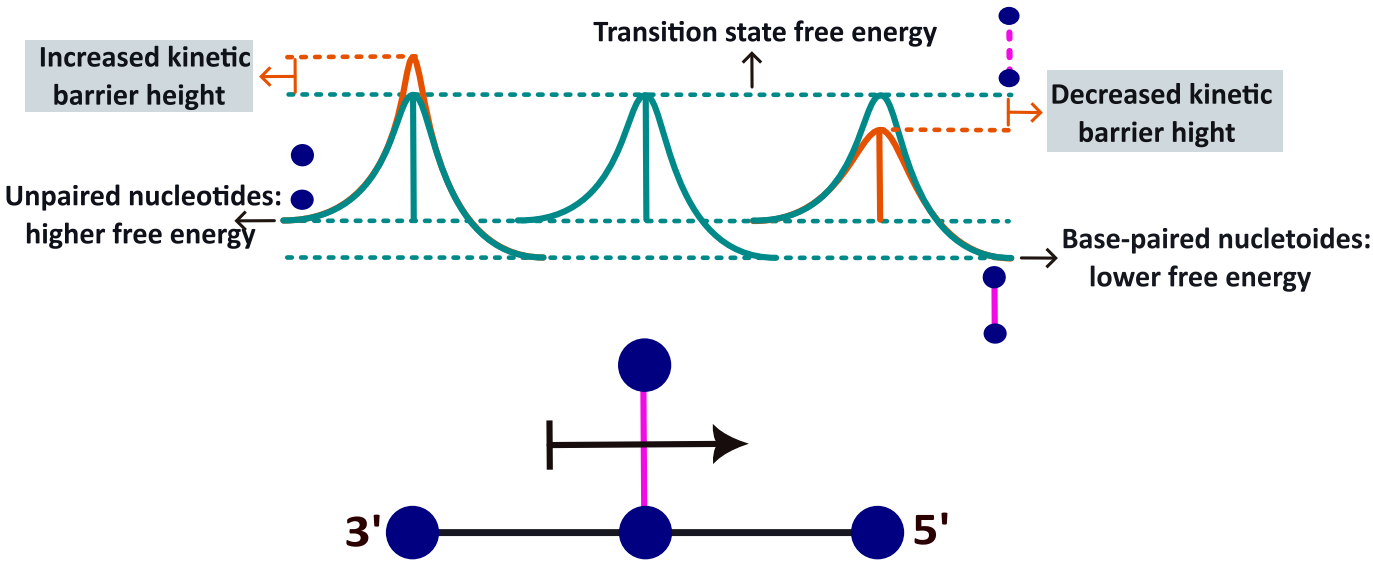


Figure 1: Illustration of asymmetric cooperativity. Nucleotides are denoted by blue circles, and vertical lines between nucleotides represent hydrogen bonds. The associated energy diagram for base pairings is presented above the illustration. The presence of a base pair lowers the kinetic barrier of its right neighboring base pair, denoted by arrow arrowhead, while increasing the barrier of the left neighbor, denoted by a barhead. This results in favorable conditions for new base pair formation to the right and increased intra-strand covalent bond formation probability to the left due to the increased kinetic barrier that stabilizes the left base pair. This kinetic asymmetry thus establishes efficient unidirectional daughter strand construction by breaking the left-right symmetry.

3. Methods

We consider a linear template strand comprising of N nucleotides. The initiation of daughter strand construction occurs when one of the free-floating nucleotides forms a base pair with a template nucleotide. An illustration of the daughter strand construction process under the influence of asymmetrically cooperative base pairing is shown in Fig.2. We model the daughter strand construction as a series of base pair formation/dissociation events between nucleotides on a template strand and free-floating nucleotides. Mathematically, these series of events can be modeled as a continuous time Markov chain [53–56] that describes each of these bonding/dissociation events as a transition between any two predefined states. These states are denoted by an N-length numerical sequence whose digits can take any of the values from the set 0,1,2. A value of 0 at any location represents the absence of a base pair between the template nucleotide and another nucleotide at that location, and a value of 1(2) represents the presence of a correct (incorrect) base pair at that location. For example, for a given template strand consisting of $N = 5$ nucleotides, the state 00000 would signify a single template strand with no base pairs present, whereas 11111 and 11211 represent states in which all nucleotides in the template strand are hydrogen-bonded to free-floating nucleotides, either correct or incorrect. To keep our analysis simpler and the size of the state space small, *we allow for an incorrect base pair formation only at a single predetermined location in the strand*. This does not affect the generality of our results and conclusions in any way.

I. Calculation of transition rate matrix

The transition rates among these states can be calculated using the base pair formation rate p between free-floating and template nucleotides, along with two distinct dissociation rates:q_r for correct base pairs and q_w for incorrect base pairs. To incorporate asymmetric cooperative interactions between neighboring base pairs on the template, we introduce two asymmetric cooperativity parameters α and $\beta = \frac{1}{\alpha}$ that modify the formation/dissociation rates of the nearest-neighbor base pairs to the right (template's $5'$-end direction) and left (template's $3'$-end direction) of a pre-existing base pair, respectively. The parameter α is the only free parameter used in the model. Fig.3 illustrates the calculation of transition rates between different states used to model the daughter strand construction on a four-nucleotide long linear template strand.

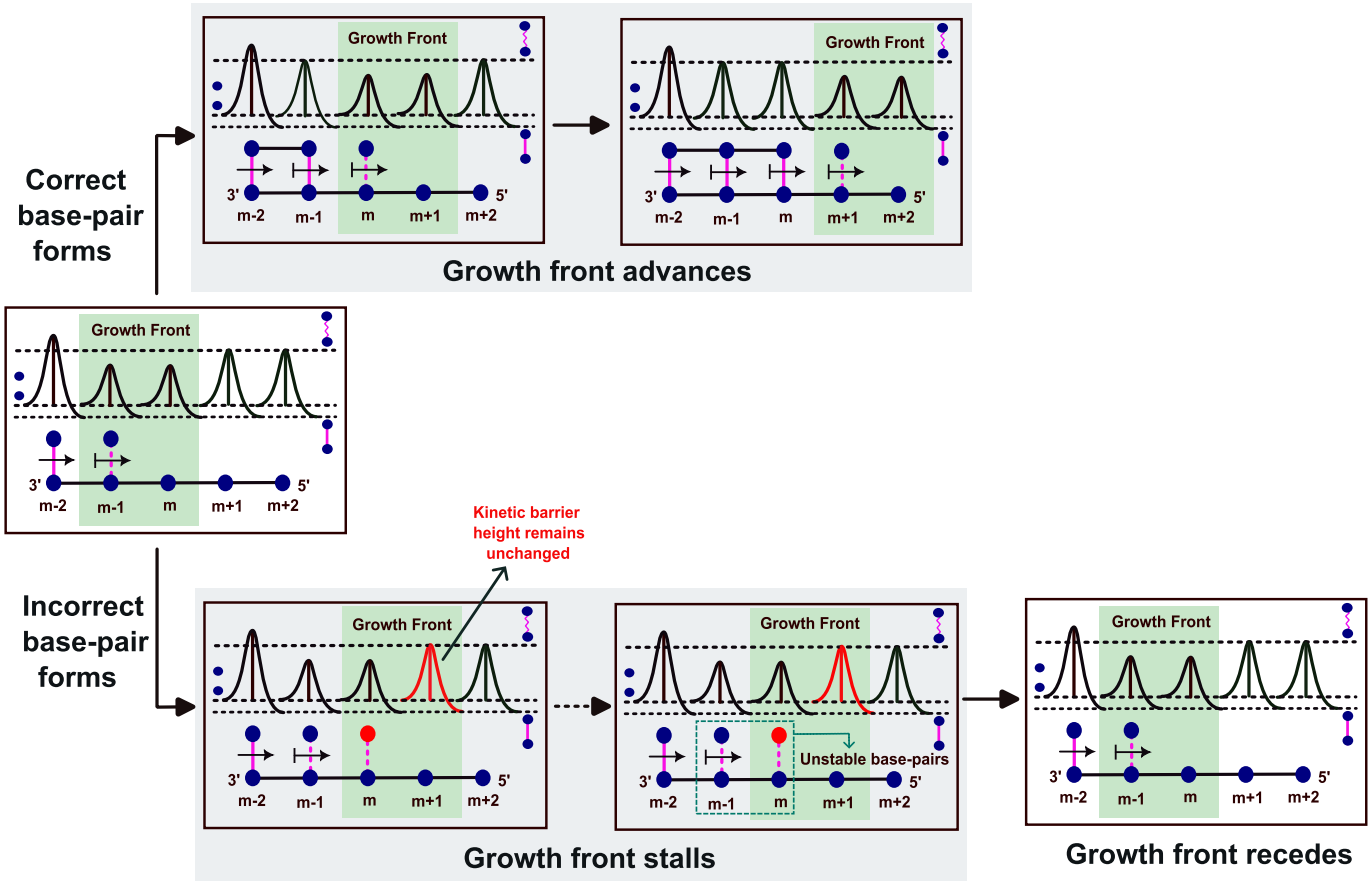


Figure 2: Illustration of error correction during daughter strand construction on a template. Both the correct (top) and wrong (bottom) base pair formation pathways are shown. The non-complementary nucleotide is represented by a red circle in the bottom panels, which is base paired with the template nucleotide at the mth location. This incorrect base pair does not catalyze the formation of a new base pair at $(m+1)^{th}$ location to its $5'$-end (right), nor does it hinder the dissociation of the $(m-1)^{th}$ base pair to the $3'$-end (left), in accordance with the model premises. This can be seen by comparing the barrier heights of these base pairs at the top and the bottom panels. The kinetic barrier of the $(m-1)^{th}$ base pair is raised due to inhibition from the mth correct base pair, stabilizing it (top), whereas, the barrier of $(m-1)^{th}$ base pair is unaffected by the presence of mth incorrect base pair, leaving it unstable (bottom). Similarly, the kinetic barrier of the $(m+1)^{th}$ location is lowered due to the presence of the mth correct base pair, catalyzing the formation of a new base pair and extending the daughter strand (top), whereas, the barrier of the $(m+1)^{th}$ location is unaffected by the presence of an incorrect base pair at m, making daughter strand extension harder, which also leaves the incorrect base pair unstable, due to its lowered kinetic barrier. The growth front advances in case of correct base pair formation at m (top), whereas, the growth front recedes in case of an incorrect base pair (bottom). This results in a time lag between the formation of correct and incorrect base pairs. If the covalent bond formation time is of the same order of the time lag, the irreversible covalent bond formation event can discriminate between the correct and incorrect base pairs, and would lead to error correction. The top panels show the covalent bond forming between the first and second daughter strand nucleotides, whereas no covalent bonds form between these two nucleotides in the bottom panels, due to the second base pair's instability, illustrated by its lowered kinetic barrier. Thus, covalent bond catalysis leads to error correction.

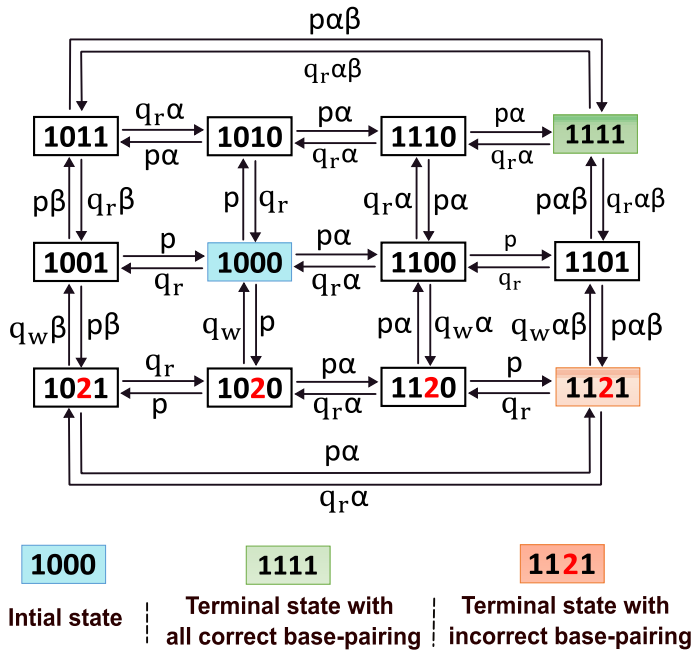


Figure 3: Illustration of calculation of the transition rates between states representing the growth of the daughter strand on a four-nucleotide-long linear template strand. The states are represented by a four-digit-long sequence containing 0, 1, or 2, with 0 denoting the absence of a base pair between the template nucleotide and a free-floating nucleotide, and 1 (2) denoting the presence of a base pair between the template strand nucleotide and a complementary (noncomplementary) nucleotide, respectively. The starting state, 1000, represents a single template strand with one base pair at the $3'$-end, whereas, the final state, 1111 (1121), corresponds to all nucleotides of the template strand base paired with complementary (noncomplementary at third position) free-floating nucleotides. The transition rates between these states are dependent on the presence or absence of adjacent base pairs to the left and right of a given base pair. The bare formation and dissociation rates, denoted by p and $q_{r(w)}$, are modulated by the local context. Specifically, for any given location on the template, if only the left neighboring correct base pair is present, it reduces the kinetic barrier for base pair formation or dissociation at that location by a factor α. If only the correct right neighboring base pair is present, it increases the kinetic barrier by a factor β. When both the neighboring base pairs are present and are correct, the kinetic barrier for base pair formation/dissociation at the given location is simultaneously reduced and enhanced and is therefore multiplied by a factor $\alpha\beta$. Incorrect neighboring base pairs do not affect the formation or dissociation rates.

II. Time evolution of state occupation probability: Absorbing and non-absorbing target states

In order to understand the daughter strand construction process in the presence of asymmetric cooperativity and intuitively understand the origin of the error correction mechanism, we need to evaluate the transition of the Markov chain across each and every state, starting at the initial state 00000, and ending at either a correctly formed daughter strand 11111 or a daughter strand with an error, 11211. We computationally evaluate the Markov chain trajectory from the initial state to the final state in two different ways: (i) as an absorbing Markov chain, where, once the chain reaches the correct/incorrect final state for the first time, it stops and never returns back. This assumption implies that once all the base pairs are formed, the chain terminates, and the base pair hydrogen bonds are not allowed to break. (ii) as a non-absorbing Markov chain, which allows for the chain to leave the final state and re-enter intermediate states. This assumption allows for breaking of base pairs even after all base pairs are formed.

The time evolution of the occupation probability $P_i(t)$ of the ith state involves solving the master equation [53–56].

$$\frac{dp_i(t)}{dt} = \sum_{i \neq j} [Q_{ji} P_j(t) - Q_{ij} P_i(t)] \tag{1}$$

In matrix notation Eq.1 can be written as

$$\frac{dP}{dt} = Q.t \tag{2}$$

whose formal solution is,

$$P(t) = e^{Qt}.P(0). \tag{3}$$

Here, $P(0)$ is a column matrix that contains the initial probability for the Markov chain to be in every state at the start of the process, and $P(t)$ is a column matrix in which each entry $P_i(t)$ is the probability for the Markov chain to be in the ith state at time t. All elements of $P(0)$ are zero, except for the initial state element, which is assigned 1.

For modeling an absorbing Markov chain time evolution, we assign $Q_{R,i} = 0, i \neq R$, and $Q_{W,i} = 0, i \neq W$, where R and W denote the indices of the correct and incorrect final states. This assignment prevents the Markov chain from traversing back to the intermediate states once it reaches the final states. For non-absorbing Markov chain, these rate matrix elements carry their usual values assigned as mentioned above.

We quantify error correction as the ratio, η, of the probability of occupation of the correct final state to the probability of occupation of the incorrect final state, after the Markov chain reaches steady state, i.e., becomes time-independent, for both absorbing and non-absorbing Markov chains.

$$\eta = \frac{P_R(t \to \infty)}{P_W(t \to \infty)}. \tag{4}$$

Thus far, we have used the 00000 as the initial state at which the Markov chain begins its sojourn. However, the transition from the state 00000 to any of the single-bonded states (10000, 01000, 00100, 00010, 00001) is independent of the kinetic parameters α and β, and therefore, sets a kinetics-independent floor value to the mean first passage times for the correct and incorrect final states. This is because of the much-lower rate of the above kinetically-unassisted transition, compared to majority of the state transitions that follow. Since our goal is to demonstrate the necessity of kinetics for error correction, this is counterproductive. Henceforth, we designate the state 10000 as the effective starting point for daughter strand construction in our analysis, and remove 00000 from the state space. This reassignment can be thought of as modeling daughter strand construction through primer extension[57]. The first base pair (primer) is assumed to be dynamically inactive or frozen and is not allowed to break.

III. Modeling of covalent bond formation

Since the error correction of an incorrect base pair at a location m depends on the presence or absence of base pairs at both $m-1$ and $m+1$ locations within our model, we impose an experimentally observed[58] constraint that the covalent bond formation between the nucleotides in the daughter strand at the locations $m-1$ and m happens *after* the hydrogen bonding of the $m+1$ base pair. However, at high nucleotide binding rates, the possibility of a base pair forming at $m+2$ without a base pair at $m+1$ becomes non-negligible. This configuration hinders the formation of a base pair at $m+1$, improving the error correction because of the catalysis of the wrong bond dissociation by the bond at $m-1$. Moreover, a base pair at $m+2$ stabilizes the base pair at m indirectly, by stabilizing $m+1$, thereby enabling covalent bond formation between m and $m-1$ nucleotides. To include these interactions, we include two base pairs to the left of the wrong base pair location, and two to the right, resulting in a template of length $N = 5$. We impose the constraint that, for a template of length $N = 5$ with a possible incorrect base pair at $m = 3$, the covalent bond would form after the completion of the base pair formation at all 5 locations, that is, after the Markov chain reaches either of the two final states, 11111 or 11211. We model only the covalent bond formation between the 2nd and 3rd base pairs, and reasonably assume that the covalent bonds between previous nucleotides on the daughter strand have already formed, while the covalent bond between 3rd and 4th, and 4th and 5th nucleotides are yet to form, and are not included in the model. We introduce a variable v to denote the rate of this covalent bond formation, and express it in terms of the rate of formation of a stretch of base pairs, starting from 10000 to 11111 or 11211. We define a dimensionless covalent bonding rate $w \equiv \frac{v}{1/T_{1\to R}}$, where $T_{1\to R}$ is the conditional mean first-passage time to the fully paired state. The transition matrix thus arrived at is solved numerically in the main article, and after introducing certain approximations, lends itself to analytical solution, which is then analytically solved for accuracy ratio, and is shown to exhibit similar dynamics as the full system, elaborated in the supplementary material.

IV. Stalling: Evaluation of Conditional Mean First Passage Time

To demonstrate the stalling of strand construction due to the formation of an incorrect base pair, we need to evaluate the time taken for the completion of strand construction. This is done by calculating the conditional Mean First Passage Time (MFPT)[53, 59]. The conditional mean MFPT, denoted as $T_{i\to k}$, represents the expected time it takes

for a system, starting from transient state i, to reach absorbing state k, given that it eventually reaches k. In a continuous-time absorbing Markov chain, the *conditional mean first passage time* (MFPT) from a transient state i to an absorbing state k is given by

$$T_{i\to k} = \frac{\sum_j \mathbb{F}(i,j)\cdot\mathbb{B}(j,k)}{\mathbb{B}(i,k)}, \tag{5}$$

where $\mathbb{F}$ is the *fundamental matrix*, and $\mathbb{B}$ is the matrix of *absorption probabilities*. The element $\mathbb{F}(i,j)$ represents the expected total time the process spends in transient state j given that it started in transient state i, while $\mathbb{B}(i,k)$ denotes the probability that the process, starting from state i, is eventually absorbed in absorbing state k.

To compute these quantities, we first partition the infinitesimal generator matrix Q of the Markov chain as follows:

$$Q = \begin{bmatrix} \mathbb{Q} & \mathbb{R} \\ 0 & 0 \end{bmatrix} \tag{6}$$

where $\mathbb{Q}$ is the submatrix corresponding to transitions among transient states, and $\mathbb{R}$ contains transition rates from transient states to absorbing states. The fundamental matrix is then given by

$$\mathbb{F} = -\mathbb{Q}^{-1} \tag{7}$$

and the matrix of absorption probabilities is computed as

$$\mathbb{B} = \mathbb{F}\cdot\mathbb{R} \tag{8}$$

Each row of $\mathbb{B}$ corresponds to a transient starting state, and each column corresponds to a distinct absorbing state.

Using Eq.5, we calculate the conditional MFPTs for two possible outcomes of strand construction: $T_{1\to R}$, the expected time to reach the correctly formed daughter strand and $T_{1\to W}$, the expected time to reach a strand containing an error. We then define the *stall quotient* τ as:

$$\tau = \frac{T_{1\to W}}{T_{1\to R}} \tag{9}$$

In this study, we examine the dependence of accuracy ratio η and stall quotient τ on the thermodynamic drive for daughter strand formation, characterized by $\Delta\mathrm{G} = \ln\left(\frac{p}{q_r}\right)$ (in units of k_BT) and, the kinetics of the daughter strand formation, characterized by the parameters α and $\beta = \frac{1}{\alpha}$.

V. Parametrization

i. Thermodynamic variables

The thermodynamic drive depends on the base pair formation and dissociation rates p and q, which are functions of temperature, monomer concentration, and other environmental variables. The Gibbs free energy that drives the base pair formation rate and hence the daughter strand construction on a template is $\Delta G = ln(\frac{p}{q_r})$ (in units of k_BT). The dependence of p and q on various environmental variables has been extensively investigated [60]. The studies of short DNA/RNA duplexes report bimolecular association rate constants on the order of $\approx 10^6 - 10^7\mathrm{M}^{-1}\mathrm{sec}^{-1}$ at room temperature and high salt concentration ($\mathrm{Na}^{++} \approx 1\mathrm{M}$) for DNA [61]. Prebiotic monomer concentrations are estimated to have ranged from micromolar to millimolar levels [62]. Based on this, we assume the base pair formation rate $p = 2\mathrm{sec}^{-1}$ at a monomer concentration of $2\mu M$. The base pair formation is entropically driven and is therefore taken to be the same for both correct and incorrect pairings. However, the dissociation rate of incorrect base pairs is two orders of magnitude faster than that of correct base pairs[14]. In our calculations, we vary the thermodynamic drive $\Delta G = \ln\left(\frac{p}{q_r}\right)$ by adjusting the base pair formation rate p, keeping the correct base pair dissociation rate $q_r = 1\mathrm{sec}^{-1}$ and the incorrect base pair dissociation rate $q_w = 100\mathrm{sec}^{-1}$ constant.

ii. Kinetic parameters

Our model requires only one free kinetic parameter, α, which quantifies the amount by which a base pair catalyzes the rate of formation/dissociation of a base pair to the $5'$-end of the template. The increase in the kinetic barrier to the $3'$-end of the template is simply assumed to be $\beta = \frac{1}{\alpha}$, to keep the number of parameters as low as possible. In our calculations, we set $\alpha = 5\times10^3$ and therefore, $\beta = \frac{1}{5\times10^3}$. These values can be motivated by the experimentally observed fact that the ratio of rates of all successive base pair formations to the rate of initial or the first base pair formation (which is unassisted by its neighboring base pairs) on a template, during the annealing of two single strands

of DNA, is of the order of 10^2 to 10^3 [63]. In other words, the rate of zippering of the two single strands is about 10^3 times faster than the initial nucleation event. Here, we have assumed that the rate enhancement due to neighborhood *kinetic* influence is of a similar order. The kinetic variable v denoting the rate of transition from the either of the two final states 11111 or 11211 to the covalent bonded states is expressed in terms of the mean first passage time for the Markov chain to reach the correct final state 11111, starting from the initial state 10000 (see "Methods" above).

4. Results

This section is organized as follows. First, in part I, we allow the kinetic parameters α and β to vary *independently* to demonstrate that asymmetric cooperativity, and hence unidirectional strand construction, is vital for error correction. We include the state 00000 as the initial state only for this demonstration, in order to maintain the left-right symmetry. We do not include the dynamics of covalent bond formation until part IV below. We show that the error correction ability is minimal when $\alpha = \beta$. We show that kinetic discrimination, defined as the presence (absence) of asymmetric cooperativity in correctly (incorrectly) formed base pairs, is most effective in correcting errors when cooperative kinetic interactions between neighboring base pairs are highly asymmetric. We then fix $\alpha = 5 \times 10^3$ and $\beta = \frac{1}{\alpha}$ for all the calculations that follow. For the rest of the analysis, we initiate the Markov chain from the 10000 state, and exclude 00000 state from the state space. In part II, to show that kinetic discrimination alone is sufficient for error correction, we first turn off thermodynamic discrimination by setting the dissociation rates of the correct and incorrect base pairs to be equal, $q_r = q_w$. We demonstrate the fact that kinetic discrimination persists into the steady state only for absorbing Markov chain, and the discrimination between correct and incorrect base pairs is lost when non-absorbing Markov chain is used to model strand construction. We then include thermodynamic discrimination in part III, along with kinetic discrimination, and show that absorbing Markov chain's error correction improves due to the addition, whereas the non-absorbing Markov chain's discrimination is limited by thermodynamics. In part IV, we provide the physical rationale for using an absorbing Markov chain by introducing covalent bond formation between the base paired nucleotides. We show that the rate of covalent bond formation is crucial for error correction, and catalysis of the covalent bond formation is sufficient for introducing the capability of error correction. In part V, we show that there is a trade-off between strand construction speed and accuracy in our model, an observation recorded in nearly every other error correction scheme proposed and tested thus far. This trade-off implies the existence of an optimal thermodynamic drive at which the error correction is maximal. We argue that this trade-off results in a built-capability for adaptive tuning of error rates in primordial self-replicators, in response to food supply variability or environmental variations, thereby enhancing the rate of evolutionary search for optimal processes and configurations.

I. Unidirectional strand construction is necessary for error correction

In our previous work[39], we argued that unidirectional strand construction arises from asymmetric cooperativity, where a base pair influences the formation and stability of neighboring base pairs unequally on either side, promoting growth in a single direction. We showed that this directional bias resolves conflicting requirements of low kinetic barrier for fast nucleotide incorporation and high kinetic barriers for base pair stability to promote covalent bonding, giving such heteropolymers an evolutionary advantage over symmetric (bidirectional) replicators.

Our contention that asymmetric cooperativity (and unidirectional daughter strand construction as its consequence) leads to error correction can be explained as follows. As shown in Fig.2, a correct base pair at m^{th} position catalyzes the formation of another base pair towards its right at $(m+1)^{th}$ position (template's $5'$ direction) by lowering the latter's kinetic barrier, thereby extending the strand construction further. This stabilizes the correct base pair at m^{th} position, due to inhibition from the $(m+1)^{th}$ base pair, resulting in a longer lifetime for the mth correct base pair. However, an incorrect base pair at m has a low probability to extend the daughter strand construction to the right, due to its inability to reduce the kinetic barrier towards its right ($5'$-end direction), according to our central premise. Therefore, the kinetic barrier and hence the stability of the incorrect base pair at m is lowered, due to the absence of a stabilizing base pair at $(m+1)$ and the presence of a destabilizing base pair at $(m-1)$. This results in higher probability for the incorrect base pair to dissociate from the template, resulting in error correction. Thus, the error correction by kinetic discrimination results from unidirectional strand construction, both originating from asymmetric cooperativity. An identical analysis holds if we switch the direction of strand construction from right to left, instead of left to right, by switching the directionality of asymmetric cooperativity.

The foregoing can be quantified by calculating the accuracy ratio η, the ratio of the probability for a fully-formed strand with a correct base pair to the probability for the strand with an incorrect base pair at *steady state* (Eq.4), at the predetermined location $m = 3$, as a function of the asymmetric cooperativity parameters α and β. These two parameters are treated as independent for this demonstration alone. The above probabilities are calculated as explained in the Methods section. We have used an absorbing Markov chain for these calculations, where, the Markov

chain is absorbed once it reaches any of the two final states (see below). Please note that covalent bond formation dynamics is not included until part IV. We calculate the accuracy ratio resulting solely from kinetic discrimination, where a correct base pair influences the kinetic barriers of its neighbors asymmetrically, while an incorrect base pair does not affect its neighbor's kinetics. We turn off the thermodynamic discrimination between correct and incorrect base pairs by setting $q_r = q_w$. Fig.4 shows the accuracy ratio, η_K, as a function of α and β. The subscript K in η_K denotes the fact that we have included kinetic discrimination alone. α and β are varied from 10^{-2} to 10^3. The values of other parameters for generating the plot are $N = 5$ with the incorrect base pair at the third location (Methods), base pair formation rate $p = 2sec^{-1}$ and the dissociation rate $q_r = q_w = 1sec^{-1}$ for both the correct and incorrect base pairs.

It is clear from Fig.4 that the accuracy ratio η_K resulting from kinetic discrimination alone is maximal when $\alpha \neq \beta$. There are two equivalent maxima, one where a base pair catalyzes its right neighboring base pair formation/dissociation and inhibits the left base pair's formation/dissociation, and the other where the left neighbor formation/dissociation is catalyzed, and the right neighbor's, inhibited. These two maxima suggest that when the symmetry of the kinetic influence on the left and right neighbors is broken, discrimination between correct and incorrect base pairs increases. Specifically, when the kinetic influence of a base pair on the left and right neighbors is equal, the error correction is minimal. This demonstrates the fundamental connection between unidirectional replication and error correction. Apart from the evolutionary advantage of unidirectional replication identified in [39], asymmetric cooperativity is shown here to provide an additional advantage of error correction. Thus, we arrive at the first central conclusion of our article: *Unidirectional daughter strand construction is essential for error correction.*

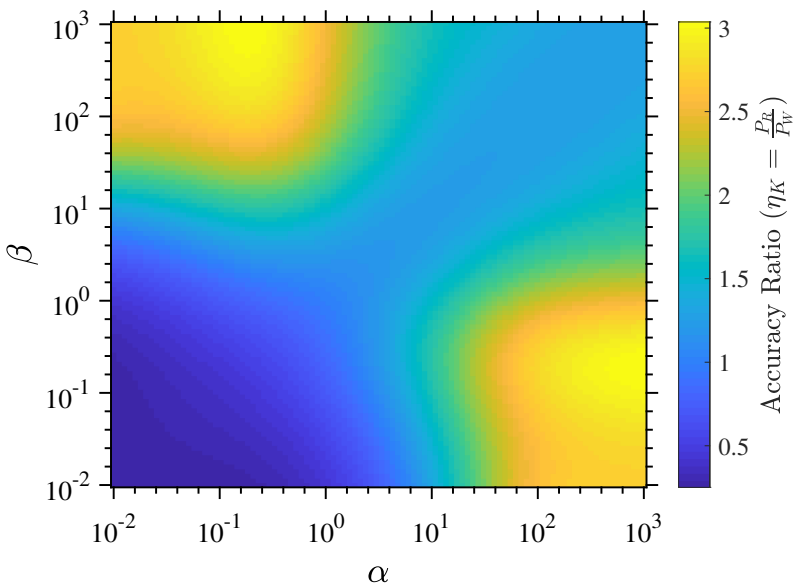


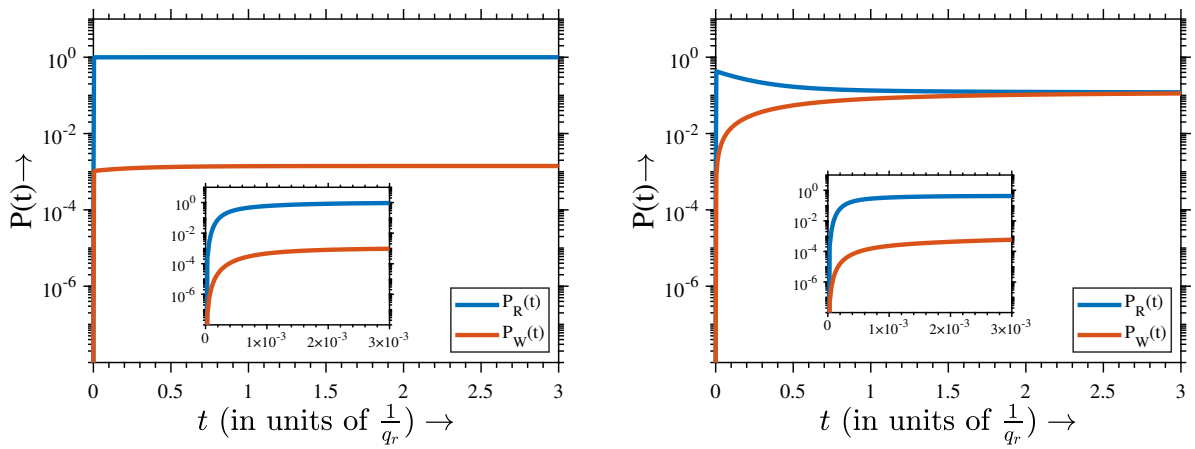
Figure 4: Accuracy ratio (η_K) as a function of the kinetic parameters α and β, for a linear template comprising $N = 5$ nucleotides and an incorrect base pair at the third location. The base pair formation and dissociation rates are set at $2sec^{-1}$ and $1sec^{-1}$,respectively. Each point on the plot corresponds to a unique pair of α and β values, with the color intensity indicating the magnitude of the accuracy ratio. The plot exhibits two distinct maxima, where $\alpha \neq \beta$, positioned symmetrically away from the diagonal. An optimal accuracy ratio is achieved when the cooperative kinetic interactions between the base pairs are asymmetric, whereas symmetric cooperativity along the diagonal ($\alpha = \beta$) results in the lowest possible accuracy ratio.

II. Kinetic discrimination leads to error correction only under irreversible conditions

We have shown above that kinetic discrimination, introduced through our premise that incorrect base pairs do not influence the kinetic barriers of their neighbors, is sufficient for error correction, provided that the asymmetric cooperativity parameters α and β have dissimilar values. Henceforth, we fix the values for α and β as 5×10^3 and $\frac{1}{5\times10^3}$ respectively, for the remainder of the article (see "Parametrization"). It has to be remembered that no thermodynamic discrimination between the correct and incorrect base pairs has been introduced yet. Here, we study the role of kinetic

discrimination alone by examining how it operates under two different modeling approaches. Specifically, we compare an absorbing and a non-absorbing Markov chain model to assess whether kinetic discrimination alone, without any thermodynamic bias, i.e., with $q_r = q_w = 1sec^{-1}$, is sufficient to sustain replication fidelity in each case.

In Fig.5, we plot the time-evolution of the occupation probability of the two final states 11111 and 11211, denoted by $P_R(t)$ and $P_w(t)$, respectively. These trajectories reveal the fundamental differences between the absorbing and non-absorbing models in the long-time limit. In the absorbing Markov chain (Fig.5a), kinetic discrimination persists even at steady state. This is because of the presence of an absorbing final state, which makes the system inherently out of equilibrium, thereby violating the detailed balance condition. As a result, kinetic asymmetries continue to influence the dynamics, due to the persistent flow of probabilities from the initial to the final states, allowing for sustained generation of accurate daughter strands. In contrast, in the non-absorbing Markov chain (Fig.5b), kinetic discrimination vanishes at steady state. Since the discrimination is based on the time difference between the correct and incorrect daughter strands, and since the steady state is time-independent, kinetic discrimination becomes unproductive. The local detailed balance [64] holds and entropy production becomes zero, and the system reaches equilibrium, despite the presence of a thermodynamic drive, suggesting that the system is unable to absorb the energy from the drive due to the possibility of reverse transitions.



(a) **Absorbing Markov Chain** (b) **Non-absorbing Markov Chain**

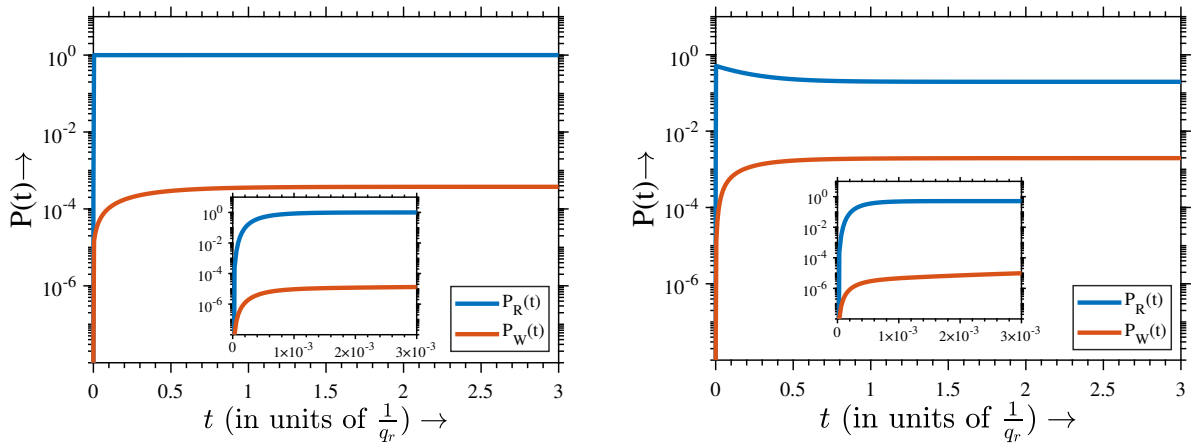
Figure 5: Time evolution of the occupation probabilities for the correct state 11111 and the incorrect state 11211, denoted by $P_R(t)$ and $P_W(t)$, respectively. (a) In the absorbing Markov chain framework, kinetic discrimination persists at steady state due to the irreversible nature of the final state, which drives the system out of equilibrium. The probabilities of all intermediate states go to zero at steady state. (b) In the non-absorbing Markov chain, the system evolves toward equilibrium, and kinetic discrimination vanishes at steady state as detailed balance is satisfied. The probabilities of intermediate states (not shown) are non-zero.

This observation indicates that kinetic discrimination can only be sustained either during the transient phase of a non-absorbing Markov chain or by rendering the final-state transition effectively irreversible, thereby transforming the system into an absorbing Markov chain. In this study, we adopt the latter approach by employing an absorbing Markov chain model. We will later justify this choice when we include the covalent bond formation, which naturally renders the Markov chain irreversible.

III. Kinetic+thermodynamic discrimination improves error correction in absorbing Markov chain

Now we proceed to examine how the integration of kinetic and thermodynamic discrimination improves error correction when the absorbing and non-absorbing Markov chain modeling approaches are used. To incorporate thermodynamic discrimination alongside kinetic discrimination, we consider the dissociation rate of the correct base pair as $q_r = 1$ sec^{-1} and that of the incorrect base pair as $q_w = 100$ sec^{-1}, hundred times larger than q_r (Methods).

As shown in Fig.6a, in the absorbing Markov chain approach, combining kinetic and thermodynamic discrimination improves the accuracy ratio $\eta_{\mathrm{K+T}}$ beyond what either mechanism can achieve on its own. The absorbing final state keeps the system out of equilibrium, allowing kinetic differences to continue influencing the dynamics and work together with thermodynamic bias to enhance fidelity. In contrast, in the non-absorbing Markov chain, Fig.6b, the system eventually



(a) **Absorbing Markov Chain** (b) **Non-absorbing Markov Chain**

Figure 6: Time evolution of the occupation probabilities for the correct state 11111 and the erroneous state 11211, denoted by $P_R(t)$ and $P_W(t)$, respectively, under combined kinetic and thermodynamic discrimination. (a) In the absorbing Markov chain, the system remains out of equilibrium, allowing both kinetic and thermodynamic discrimination to contribute to improved accuracy. (b) In the non-absorbing Markov chain, the system reaches equilibrium at long times, and only thermodynamic discrimination governs the steady-state accuracy, which is lower than the accuracy obtained in the absorbing case.

reaches equilibrium, where kinetic discrimination no longer plays a role due to detailed balance. As a result, in this case, the accuracy ratio $\eta_{\mathrm{K+T}}$ is limited by thermodynamic discrimination alone.

IV. Error correction requires rapid, irreversible covalent bond formation

We have thus far employed two approaches to model the base pairing of nucleotides on a template strand during DNA/RNA self-replication: non-absorbing and absorbing Markov chains. We showed that the latter approach leads to error correction, while the former does not. We have argued in part II that irreversible absorption of probabilities is necessary for error correction, and the absorbing Markov chain approach imposes such irreversibility, by disallowing reverse transitions from the final state. The physical justification for our mathematical choice of irreversible Markov chain is the presence of irreversible covalent bonding between the base-paired nucleotides on the template strand. Since covalent bond formation, involving hydrolysis of triphosphate bond, is highly exergonic ($\Delta\mathrm{G} = -7\mathrm{kcal/mol} = -11.8k_BT$)[65], we can safely assume that this reaction is effectively irreversible. Once covalent linkages are established, backward transitions (such as pyrophosphorolysis) become effectively irreversible and do not contribute to error correction. This assumption is supported by experimental observations, such as those reported by Wong et al.[66], who noted that "the rate of pyrophosphorolysis on a mismatched 3′-end is undetectable, indicating that pyrophosphorolysis does not play a proofreading role in replication". How is this irreversibility used for error correction?

As explained earlier in the article, reaching a final state with an error takes longer than reaching the correct final state. This time difference between wrong and correct daughter strand formations can be used for error correction if the faster correct daughter strands are immediately secured through the irreversible covalent bond formation. The rate of covalent bond formation must be high such that it can *resolve* the difference between the correct and wrong strand formation times. To count seconds, one must have a timer that ticks at least once every second, if not faster. The idea that a transient temporal difference can be converted into a persistent structural order, through irreversibility, is schematically illustrated in Fig.7. Thus, we come to the conclusion that rapid, irreversible covalent bond formation naturally imposes irreversibility and justifies our choice of absorbing Markov chain approach, and results in error correction.

To mathematically substantiate the foregoing explanation, we introduce two additional terminal states: **R** and **W**, representing covalently bonded correct and incorrect final states, respectively. We make the above two states absorbing, instead of the earlier absorbing states 11111 and 11211. Since we only care about incorporation of the incorrect/correct nucleotide at the $m = 3$ location into the daughter strand, the terminal states **R** and **W** represent daughter strands with a covalent bond between the $m = 3$ and the previous $m = 2$ nucleotides. The covalent bonds between previous nucleotides are assumed to have formed, while the bonds between nucleotides at $m = 3\&4$ and $m = 4\&5$ are assumed to have not formed yet, although nucleotides are base paired to the template at these locations.

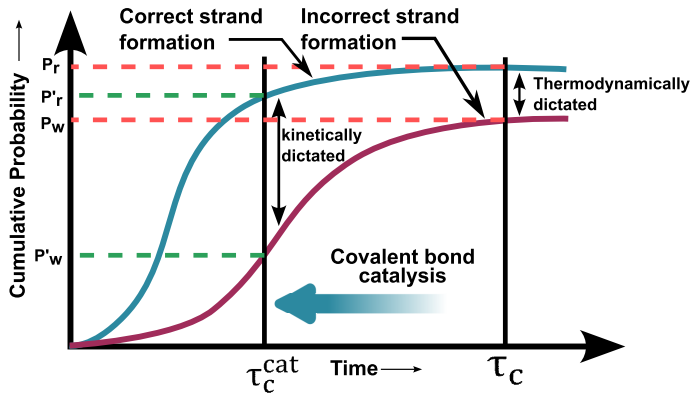
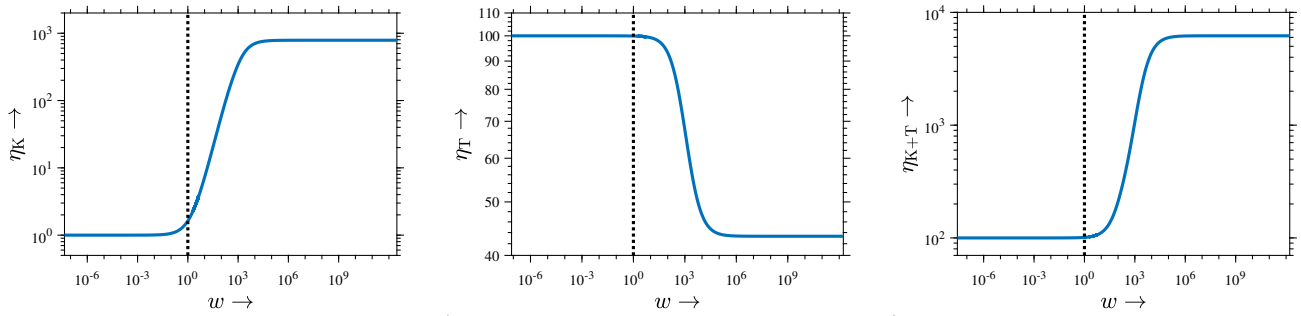


Figure 7: A schematic illustration of how covalent bond catalysis leads to error correction. The diagram also illustrates how a transient time difference between correct and incorrect strand construction is converted into persistent structural order through irreversibility-aided error correction. Two curves, a blue one depicting the cumulative probability of formation of correct strand as a function of time, and a red one depicting the time variation of cumulative probability of incorrect strand formation, are shown. These base pair formation/dissociation events are reversible, and are presumed to happen at rates much faster than the timescale of the figure. Due to asymmetric cooperativity-induced kinetic discrimination between correct and incorrect base pairs, a transient time delay arises between the correct and incorrect strand formation, with incorrect strand formation transiently lagging behind, as illustrated by the slow rise of its cumulative probability curve. When the covalent bond formation between two nucleotides on the daughter strand is not catalyzed, the expected time τ_c for the formation of the covalent bond would be very large, by which time, the above transient time difference would have ceased to exist. This uncatalyzed covalent bond formation time is denoted by a vertical line towards the right. Therefore, uncatalyzed covalent bonds irreversibly stabilize base pairs whose error frequencies, calculated as $\frac{P_r}{P_w}$, are determined by thermodynamics alone. This is the regime most non-enzymatic DNA replication experiments operate in. When the covalent bond is catalyzed by a polymerase or a metal ion, the expected time for its formation decreases to τ_c^{cat}, denoted by the left vertical line. If this time is within the time window when the transient kinetic difference between correct and incorrect base pairs persist, the irreversible covalent bond formation can discriminate between the correct and incorrect base pairs better than the thermodynamically-imposed limit, and freezes-in the correct base pair. The accuracy ratio in this regime is given by $\frac{P'_r}{P'_w}$.

The justification for these assumptions are provided in the Methods section. These terminal states are connected to the fully base paired configurations 11111 and 11211, with the dimensionless transition rates given by $w \equiv \frac{v}{1/T_{1\to R}}$, where $T_{1\to R}$ is the conditional mean first-passage time to the correctly base paired covalent bonded state $\mathbf{R}$. This implies that the absolute forward transition rate v from the non-covalent-bonded to the covalent-bonded states are measured in terms of the conditional mean first passage time to reach the correct state $\mathbf{R}$ from the initial state. we again calculate the accuracy ratio η, defined in Eq.4, to quantify the changes in error correction when covalent bonding is added to the model.

As before, we examine the accuracy ratio η, with the absorbing Markov chain now including the covalent bonded states as the terminal states, under three distinct regimes: kinetic discrimination regime, thermodynamic discrimination regime, and the combined regime. In Fig.8, the accuracy ratio is depicted as a function of the covalent bond formation rate w, across these three distinct regimes. The sharp distinction between the kinetic and thermodynamic regimes can be seen in the Figs.8a and 8b: when the covalent bonding rate increases, kinetic discrimination helps in increasing the accuracy ratio, whereas, thermodynamic discrimination hinders the accurate self-replication. When the covalent bonding rate is high, discriminating the time difference between the early correct sequence and the later-arriving wrong sequence becomes more effective. Therefore, higher covalent bonding rates lead to better error correction. This can be seen from the knee point in Fig.8a, where the accuracy ratio begins to increase, $v \approx \frac{1}{T_{1\to R}}$. As soon as v exceeds $\frac{1}{T_{1\to R}}$, the covalent step "freezes in" whichever intermediate is currently favored-so any kinetic advantage amplifies into a large accuracy gain.

On the other hand, the behavior of η in Fig.8b can be understood by considering the role of w as the rate controlling the transition from intermediate states to the absorbing states ($\mathbf{R}$ and $\mathbf{W}$). When w is small, the absorption process occurs slowly, allowing the system to dwell longer in the transient states. This extended residence time facilitates the establishment of a quasi-equilibrium among the intermediate states, such that the occupation probabilities reflect thermodynamic biases. This regime is nearly equivalent to the non-absorbing Markov chain approach mentioned earlier.



(a) **Kinetic** (b) **Thermodynamic** (c) **Kinetic+Thermodynamic**

Figure 8: Illustration of accuracy ratio η as a function of the covalent bond formation rate w under three discrimination scenarios. In each panel, the vertical dashed line marks $v = \frac{1}{T_{1\to R}}$: (a) Kinetic discrimination: At very low rates of covalent bond formation, i.e.,($w << 1$ or $v << \frac{1}{T_{1\to R}}$), the probabilities of covalent bond formation extending the correct and incorrect daughter strands are very small, and the ratio between these probabilities is also small, because of the inability of the slow covalent bond formation event to distinguish between the much faster base pairing dynamics of the correct and incorrect nucleotides. This leads to $\eta_K \approx 1$. When the rate of covalent bonding, v, matches or exceeds the rates of base pair formation and dissociation, $\frac{1}{T_{1\to R}}$, the covalent bonding event begins to become sensitive to the time difference between the correct and incorrect base pairing dynamics. This location, $w = 1$, is highlighted by a dashed line. From this point on, the covalent bond begins to preferentially incorporate the correct nucleotide over the incorrect one, error correction rapidly improves, and saturates at a value dictated by the kinetic parameter α. (b) Thermodynamic discrimination: Increasing the covalent bonding rate w in this regime leads to reduced accuracy. When bond formation rate is slow ($v << \frac{1}{T_{1\to R}}$), the system has sufficient time to explore multiple configurations and settle into a thermodynamically favored state, leading to a discrimination factor ($q_w/q_r = 100$) determined entirely by thermodynamics. When the irreversible covalent bond formation rates increase beyond $v = \frac{1}{T_{1\to R}}$, the time for exploration of multiple configurations is minimized, the nondiscriminating kinetic factors dominate the time evolution of the Markov chain, leading to cessation of error correction. (c) Combined discrimination: When both kinetic and thermodynamic mechanisms are at play, the behavior of the system mirrors that observed in the kinetic regime with high accuracy ratio η_{K+T}. This highlights the dominant role of kinetics in error correction and suggests that rapid and irreversible covalent bond formation is sufficient to ensure high fidelity.

In contrast, when w is large, the system transitions rapidly to the absorbing states, significantly reducing the time available for equilibration among the transient states. As a result, the accuracy ratio is increasingly determined by the kinetic pathways rather than equilibrium distributions, which are non-discriminatory in this regime. In the combined regime, Fig.8c, which includes both thermodynamic and kinetic discrimination, the behavior of η is similar to that of the kinetic regime, pointing to the importance of kinetics in error correction. We provide a considerably simpler, analytically tractable model of daughter strand construction in the supplementary material, where, we mathematically demonstrate the lack of error correction for very small values of w, the relative covalent bonding rate, and show that the error correction approaches values observed in passive polymerases for high values of w.

With the above demonstration, we arrive at one of the central conclusions of this article: *Covalent bond catalysis is critical and sufficient for maintaining highly accurate daughter strand construction.* This mechanism alone would explain the ability of passive polymerases to correct errors without extraneous energy input, and thus would resolve the paradox that we pointed to in the introduction. This also accounts for the high error rates observed during non-enzymatic DNA self-replication. With a thermodynamic discrimination between the correct and wrong base pairing of 1/100, the maximum error correction achieved in our model is about $1/(4 \times 10^4)$, enough to account for the experimentally observed improvement in accuracy of 1/10000 by employing thermodynamically passive polymerases. The passive polymerases simply have to catalyze the covalent bond formation, which is their central observed function, and error correction follows as a consequence. A caveat to the above statement is that, when the covalent bond formation rates far exceeds the rates of base pair formation/dissociation, the error correction becomes infeasible again, as covalent bonds form immediately after a base pair formation event, without waiting for the wrong base pair dissociation. This can impose an upper limit on the catalytic efficiencies of extant enzymes. Our model disallows this possibility by requiring the covalent bond to form *after* all the five base pairs form, guided by experimental observations [58]. The energy supply for passive base selection comes from the irreversible covalent bond formation, within our model.

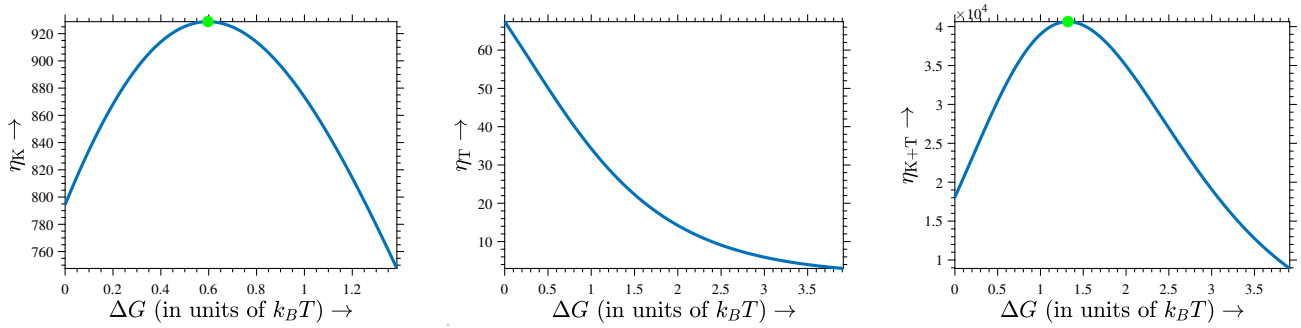
In the primordial scenario, when polymerases were not available, how was the accuracy of daughter strand con-

struction maintained? The rate of covalent bond formation in the primordial scenario was possibly very slow, of the order of minutes to hours, whereas, the rate of base pairing was of the order of seconds[67]. These values suggest that a catalyst for covalent bond formation must have been present in the primordial scenario to catalyze daughter strand polymerization and thus maintain its accuracy. Metal ions (primarily Mg^{++}), which are present in all extant DNA polymerases without exception[68] and perform crucial error correction functions, might have assisted covalent bond formation in DNA/RNA or their precursors in the primordial scenario, most probably through electrostatic interactions[68–70]. The foregoing implies that a search for efficient inorganic prebiotic catalysts for polymerization of DNA/RNA precursors may also help solve the replication accuracy problem in the primordial scenario.

V. Strand construction Speed-Accuracy trade-off: Existence of optimal thermodynamic drive

The general observation of an inverse relationship between product formation rate and product accuracy during error correction is pervasive in the literature [7, 71–75]. In the paradigmatic Hopfield-Ninio scheme, as the product formation rate (w in [3]) decreases, the error correction improves. Variations on this basic scheme show interesting behavior, such as the observation of the presence of an optimal product formation rate at which the error is minimized[44–47]. The scheme we have developed above too exhibits an optimal thermodynamic drive at which the error rate is minimized.

In our model, this thermodynamic drive, varied by varying the bond formation rate p, determines the balance between two competing processes: the *stabilization* of the incorrect base pair at m by the newly added neighboring base pair at $(m+1)$, on the $5'$ side, and the incorrect base pair's *dissociation*, catalyzed by a correct base pair at $(m-1)$, on the $3'$ side. The outcome of this competition gives rise to an inherent speed-accuracy trade-off in the model, where high base-pair formation rates favor rapid extension but increase the likelihood of retaining errors, while very low bond formation rates decrease the lifetime of base pairs, which reduces their ability to reduce the kinetic barrier of the incorrect base pair towards their $5'$-end, leading to reduced error correction. An optimum lies between these two limits, where the drive is just enough to provide long enough lifetime for the $m-1$ neighbor of the incorrect base pair at m, enhancing error correction, but is not large enough to allow the uncatalyzed formation of a stabilizing base pair neighbor of an incorrect base pair at the $(m+1)$ location. This trade-off is quantitatively illustrated in Figs.9a and 9c, which shows the accuracy ratio $\eta_{\rm K}$ as a function of the thermodynamic drive $\Delta{\rm G} = \ln\left(\frac{p}{q_r}\right)$.



(a) **Kinetic** (b) **Thermodynamic** (c) **Kinetic+Thermodynamic**

Figure 9: Accuracy ratio η as a function of the thermodynamic drive ΔG. Panels (a) and (b) correspond to purely kinetic and purely thermodynamic discrimination, respectively, while panel (c) shows the combined regime incorporating both kinetic and thermodynamic discrimination. Below the optimal thermodynamic drive ($\Delta{\rm G}_{\rm opt}$), shown in panel (c), the base pair formation is low, and the catalytic influence of the $(m-1)$ neighbor of the incorrect base pair at m is lower, leading to lower error correction. On the other hand, above the optimal thermodynamic drive, the formation of correct base pairs next to incorrect base pairs occurs rapidly, stabilizing the incorrect base pairs and increasing the error probability. An optimal error correction lies between these two regimes.

Fig.9a illustrates the dependence of η on the thermodynamic drive ΔG in the case of purely kinetic discrimination. In can be observed that the accuracy exceeds unity even at $\Delta{\rm G} = 0$. We attribute this to the non-equilibrium condition imposed by the absorbing Markov chain approach. Within this approach, the presence of an absorbing final state introduces a unidirectional bias in the dynamics, effectively enforcing irreversibility in the daughter strand construction process. Consequently, the system inherently disrupts detailed balance and maintains a nonzero net flux toward the final (absorbing) state, even in the absence of an explicit thermodynamic drive. As a result, the system

remains out of equilibrium and supports a finite degree of error correction even at $\Delta G = 0$. In Fig.9b, η_T exhibits a monotonic decrease with increasing ΔG. This indicates that in the absence of kinetic assistance, increasing the thermodynamic drive accelerates the incorporation of both correct and incorrect nucleotides indiscriminately, thereby reducing the opportunity for error correction. Fig.9c illustrates the accuracy ratio η_{K+T} under conditions of combined discrimination, where both kinetic and thermodynamic discrimination contribute to error correction. The explanation for the behavior of η_{K+T} is given above. Overall accuracy is enhanced compared to either regime independently, indicating that the two mechanisms synergistically enhance fidelity. The peak of the curve shifts to a higher ΔG, suggesting that optimal accuracy occurs under strong thermodynamic drive.

These observations highlight the distinct but reinforcing effects of kinetic and thermodynamic discrimination in enhancing the accuracy of daughter strand construction. Kinetic discrimination operates by introducing temporal delays at error-prone steps, thereby allowing incorrect incorporation to be selectively rejected before finalization. Thermodynamic discrimination, by contrast, relies on energetic differentials-quantified as $\ln\left(\frac{q_w}{q_r}\right)$ — to thermodynamically favor correct base pairings over incorrect ones. The interplay between these mechanisms gives rise to a finely tunable accuracy landscape governed by the thermodynamic drive $\Delta G = \ln\left(\frac{p}{q_r}\right)$.

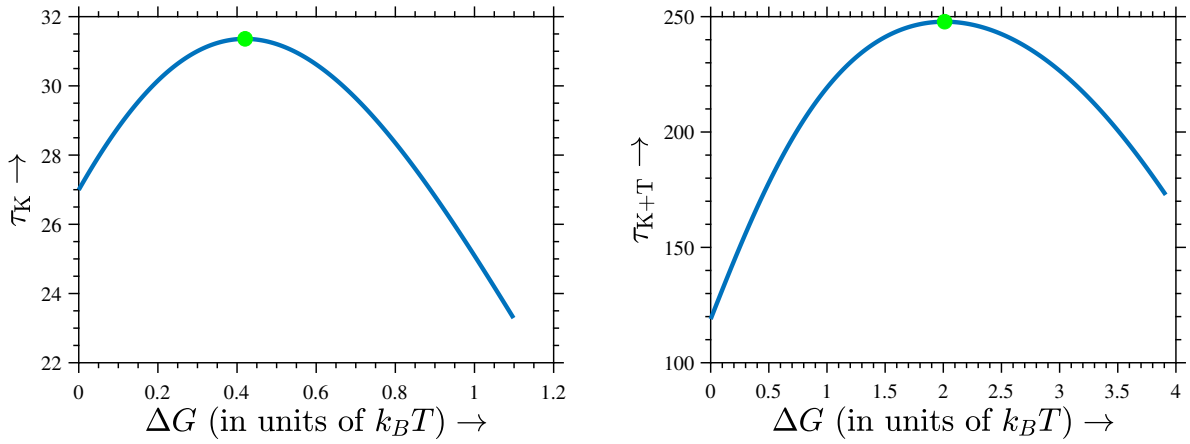
5. Experimental support

Our model captures several key phenomena observed in experimental studies of error correction in template-directed polymerization. In the following, we discuss experimental and theoretical findings in the literature that support and validate the assumptions underlying our model.

I. Stalling

Stalling denotes the slow-down of construction of daughter strands with an incorrect base pair, compared to the rate of construction of an error-free daughter strand. In our approach, stalling is quantitatively demonstrated by evaluating the conditional mean first passage time (MFPT) required to reach the final covalent bonded states - either the fully correct covalent bonded state **R** or the state containing an error **W** - starting from the initial configuration (10000). Our results show that the MFPT for reaching the fully correct final state is shorter than that for reaching the error-containing state, as the incorrect base pair inactivates the asymmetric cooperativity, making it less favorable to add new nucleotides to the growing daughter strand. Consequently, the rate of daughter strand construction decreases. This reduction in the extension rate following incorrect base pairing has been well-documented by Rajamani et al. in their study[41]. They use 3′-amino-2′, 3′-dideoxynucleotide-terminated primers to investigate the dynamics of nucleotide incorporation during *non-enzymatic* polymerization, particularly looking at what happens when there are incorrect base pairs (mismatches) at the primer-template junction. Their findings indicate that the polymerization process halts in the presence of incorrect bases, allowing accurate replication to proceed more efficiently than erroneous replication. Further support for this reduction in extension rate due to incorrect base pairing comes from the work of Leu et al.[42]. They use gel electrophoresis and mass spectrometry to quantitatively measure extension rates and mutation profiles in *non-enzymatic* nucleic acid strand synthesis. Their results demonstrate that the rate of daughter strand construction in the presence of incorrect (mismatched) base pairs is 10 to 100 times slower compared to error-free conditions. Göppel et al.[43] offer further evidence for stalling induced by incorrect base pairing through their investigation of *enzyme-free* template-directed polymerization dynamics via kinetic modeling. In their study, they introduced mismatches at the ligation site to assess the impact of non-complementary nucleotides on the polymerization process. Their findings indicate that the presence of mismatched base pairs results in a marked deceleration of strand extension, a phenomenon they termed kinetic stalling.

In Fig.10, we present the stalling quotient, defined as $\tau = \frac{T_{1\to W}}{T_{1\to R}}$, obtained under purely kinetic discrimination as well as in combination with thermodynamic discrimination, plotted as a function of the thermodynamic drive ΔG. We observe that in both the plots, τ exhibits a non-monotonic dependence on the thermodynamic drive ΔG, similar to the behavior of η in Figs.9a and 9c: it increases initially, reaches a maximum, and then decreases. Importantly, across the entire range of ΔG, the stall quotient τ remains greater than one, indicating that the conditional mean first passage time to the incorrect final state consistently exceeds that of the correct final state. Furthermore, the peak position of the stall quotient τ aligns closely with the optimum observed in the accuracy plots shown in Figs.9a and 9c. This correspondence solidifies the proposed mechanistic link between stalling induced by incorrect base pairing and the increased likelihood of subsequent dissociation of the incorrect base-pair, thereby enhancing the accuracy of the daughter strand construction, quantified by the accuracy ratio (η).



(a) **Kinetic** (b) **Kinetic+Thermodynamic**

Figure 10: Stalling quotient $\tau = \frac{T_{1\to W}}{T_{1\to R}}$ as a function of thermodynamic drive ΔG. The plots show the conditional mean first passage time (MFPT) to the error-containing final state (**W**) relative to that for the error-free final state (**R**), under (a) purely kinetic discrimination and (b) combined kinetic and thermodynamic discrimination. In both cases, τ exhibits a non-monotonic dependence on ΔG, peaking at intermediate values. The quotient τ remains greater than one throughout, indicating that incorrect base pairing leads to a slowdown in daughter strand construction. The green dot marks the value of ΔG at which τ reaches its maximum. This maximum closely corresponds to the optimal accuracy points in Figs.9a and 9c, supporting a mechanistic link between stalling and improved accuracy.

II. Fraying

When the formation of the new base pair is stalled due to an incorrect base pair at m, both the $(m-1)^{th}$ and m^{th} base pairs in Fig.2 at the growing front remain unstable and are prone to dissociate, due to the low kinetic barriers at their respective locations. Dissociation of any of these two base pairs can precipitate a sequential dissociation of other base pairs in the $3'$ direction, leading to strand *fraying*. Therefore, the presence of an incorrect base pair increases the likelihood of fraying of the daughter strand from the template strand, in our model. This phenomenon is experimentally observed in [48], where Fernandez-Leiro et al. used cryo-electron microscopy (cryo-EM) and nuclear magnetic resonance (NMR) spectroscopy to study the structural dynamics of a mismatched DNA substrate during high-fidelity replication. They found that the presence of an incorrect base pair increases the fraying of the DNA duplex by one additional base pair at the $3'$ terminus of the daughter strand, leading to a distorted conformation that facilitates the transfer of the DNA strand to the exonuclease active site for proofreading. Their findings indicate that mismatch-induced fraying acts as a self-correcting mechanism, where the DNA itself enhances its error correction by increasing instability at the growth front.

Fig.11 provides quantitative evidence that a singe mismatch considerably increases the likelihood of strand fraying. The plots illustrate the probability $P_{10000}(t)$ of reaching the initial one-base-pair state "10000" over time, starting from either the all-correct intermediate state 11100 or the mismatched state 11200. The probability for the state with an error 11200 to revert back to the "10000" state is greater than that of the correct state 11100.

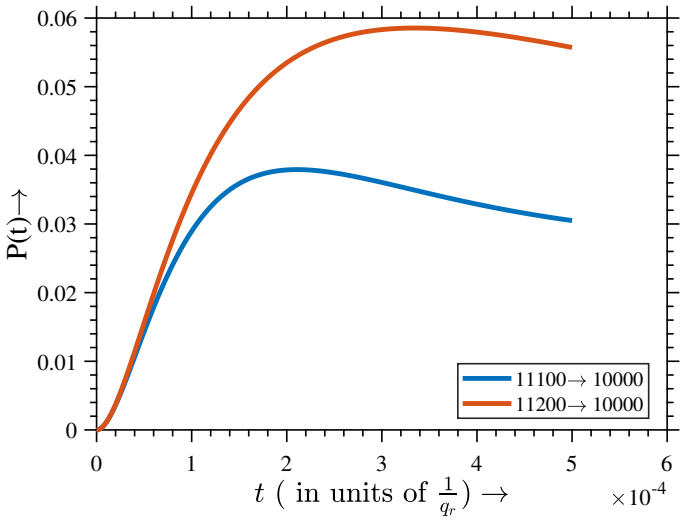


Figure 11: Time evolution of the occupation probability of the state 10000 from two initial conditions: 11100 (all three base pairs $1-3$ correct) and 11200 (base pairs $1-2$ correct, position 3 is incorrect). State 10000 represents a frayed configuration. The plot compares $P_{11100\to10000}$ and $P_{11200\to10000}$. The probability of reaching 10000 starting from 11200 is larger than from 11100. This disparity indicates that the stand frays much more readily when a mismatch present.

III. Next nucleotide effects

Upon the formation of a correct base pair at $(m+1)$, next to the incorrect base pair at m, in the $5'$ direction of the template, the incorrect base pair is kinetically stabilized, and the extension rate of daughter strand synthesis normalizes, within our model. This stabilization and recovery of the extension rate are due to the asymmetric cooperativity of the newly formed correct base pair, which inhibits the dissociation of the incorrect base pair and facilitates the formation of subsequent base pairs in the $5'$ direction. The ability to recover extension speed following the incorporation of a correct base pair next to an incorrect one (in the $5'$ direction) represents a key feature of our error correction model. This behavior is consistent with the findings of Leu et al.[42], who demonstrated that the addition of a correct nucleotide next to a mismatched base pair can restore the daughter strand extension rate to near-normal levels, indicating a localized stabilization effect that supports continued daughter strand construction.

More over, a correctly paired base immediately downstream of a mismatch tends to lock in the error and prevent its correction. This effect has been characterized in DNA replication: increasing the concentration of the "next correct" nucleotide dramatically reduces proofreading. In classic experiments with E.coli DNA pol I, Kukel et al. showed that raising the level of the next complementary dNTP increased misincorporation roughly 25-fold at a specific site[49]. They attributed this mutagenic effect to a decrease in exonucleolytic excision: once a correct base is paired after a mismatched end, Pol I's $3' \to 5'$ exonuclease is far less likely to remove the mismatch. In other words, the next nucleotide (dATP in their example) stabilized the nascent helix and suppressed proofreading, causing the error to persist.

In contrast, when an incorrect base pair forms next to an existing incorrect base pair in the $5'$ direction, the lack of asymmetric cooperativity associated with the preceding incorrect base pair does not lower the kinetic barrier for the subsequent new base pair in the $5'$ direction. Consequently, the newly added incorrect base pair is not destabilized by the local asymmetric kinetic influence of the preceding incorrect base pair. As a result, the dissociation of the newly formed incorrect base pair becomes less probable, allowing both incorrect base pairs to persist. Thus according to our model, this scenario creates a favorable environment for continued erroneous incorporation, ultimately leading to a cascade of sequential errors. This phenomenon has also been observed experimentally by Leu et al.[42], who reported that an initial mismatch is highly likely to be followed by another mismatch $54-75\%$ of the time, resulting in clusters of mismatches and significant kinetic stalling.

IV. Speed-Accuracy trade-off

Numerous models [7, 71–75] have quantitatively captured speed-accuracy trade-off: where increased accuracy typically comes at the cost of reduced speed. However, our model adds nuance to this conventional relationship . In our model, maximum accuracy is found not at the slowest product formation speed, but at an intermediate speed, determined

by the kinetic parameters (Fig.9c). The positive influence of daughter strand construction speed on accuracy below optimal thermodynamic drive (to the left of the peak in Fig.9c) is due to asymmetrically cooperative kinetics: in this regime, daughter strand construction is slow and dominated by sequential nucleotide addition at the growth front, due to catalytic influence of the most-recently formed base pair. The likelihood of forming the $(m+1)^{\text{th}}$ base pair is low when the mth base is incorrect, due to absent catalytic assistance (Fig.2). This extends the time window for the $(m-1)^{\text{th}}$ neighbor to destabilize and dissociate the incorrect base. Increasing the drive parameter p in this regime enhances error correction by increasing the $(m-1)^{\text{th}}$ base pair formation probability and stability, which increases m^{th} incorrect base pair removal efficiency. However, beyond optimal thermodynamic drive (to the right of the peak in Fig.9c), accuracy declines due to incorrect base pair stabilization by the rapidly forming $m+1$ neighbor, reducing kinetic discrimination advantage. This non-monotonic accuracy profile is consistent with the observations in [44–47].

Ravasio et al.[44] observed this optimality in their minimal non-equilibrium model for molecular order. They simulated Darwinian selection of fast replicators and found that replication speed alone, even in the absence of explicit selection for fidelity, can spontaneously drive the emergence of error-correcting pathways. The evolved networks not only replicated faster but also became more accurate, up to a point — beyond which fidelity declined. This emergent Pareto front separating speed and fidelity parallels the trade-off observed in our results.

Banerjee et al.[45] also observed this non-monotonic speed-accuracy trade-off in their proofreading model using a first-passage time framework. They demonstrated that both speed and fidelity are governed by the enzyme's internal kinetic parameters. Their analysis reveals a characteristic fidelity-speed curve, where fidelity increases with speed up to an optimum, beyond which error rates rise sharply. Remarkably, the optimal fidelity of polymerases do not coincide with the fastest possible replication; instead, polymerases operate in a biologically tuned regime where speed is favored over maximal accuracy. Their work supports our model's prediction of a fidelity maximum at intermediate speeds.

A parallel conclusion is presented by Poulton et al.[46], who studied persistent polymer copying in a minimal nonequilibrium framework. They showed that copy–template mutual information — a proxy for replication fidelity — exhibits one or more local maxima at intermediate values of chemical driving (see their Fig.4). At higher driving, fidelity declines due to the loss of kinetic discrimination: when polymerization becomes nearly irreversible, incorrect bases are incorporated as readily as correct ones, diminishing error suppression.

More corroborative evidence emerge from theoretical models of proofreading and enzymatic error correction. Galstyan et al. introduced a spatial proofreading model where diffusion-mediated delays enhance fidelity[47]. By tuning the rate of the driving motor (piston compressions), they found a "resonance" phenomenon: fidelity peaks at a finite driving rate, with lower fidelity at both slower and faster drives (Fig.6D in the paper). Thus the allosteric proofreading engine has an optimal operating frequency for maximum accuracy.

Together, these results support the generality of a speed–accuracy trade-off in polymerization systems. They reinforce the prediction from our model that accuracy is maximized not at the lowest or highest speeds, but at a finely tuned intermediate value, shaped by the interplay between stalling-induced dissociation and catalysis-driven extension.

V. Adaptive tuning of error rates

Since the base pair formation rate p is affected by environmental factors such as temperature, ionic strength, and pH, the optimal drive for error correction—and consequently, daughter strand fidelity can be modulated by the varying the environment(Fig.9c). Such adaptive tuning of error rates would have been particularly advantageous in primordial scenarios, where spatiotemporal fluctuations in thermal, chemical, and other environmental variables were likely more pronounced than they are today[76]. Any non-enzymatic self-replicator capable of adaptively tuning its error rates would have had a significant evolutionary advantage over its competitors in these unpredictable environments. The ability of extant organisms to vary the timing and genomic locations of sequence variability is reviewed extensively in [77]. The preferential localization of essential genes in the mutationally cold regions and that of communication genes that need to respond to environmental variability in the mutationally hot regions in human genome, pointing to the existence of strategies of locally adaptive variations of the mutation rates, has been observed[78]. Hypermutative phenotypes evolve independently in response to ethanol stress in *E. Coli*[50]. Hypermutative phenotypes of the clinically derived pathogen *A. Baumannii* repeatedly evolved in response to antibiotic treatment, leading to antibiotic resistance [51]. These and many other examples in the literature suggest that organisms are capable of adaptively tuning their error rates, that predictably vary across genomic locations, in response to environmentally induced stress. Our model suggests one possible way such adaptive tuning can emerge: when the environmental stress affects the thermodynamic drive for strand construction, moving it away from its optimal value, the balance between speed and accuracy gets perturbed (Fig.9c), leading to higher mutation rates. The adaptive localization of hypermutative sequences, such as inverted repeats, and within certain genomic locations, such as promoters[79, 80], cannot be modeled within the approach above. However, the model can be extended to include sequence dependence of error rates [81], which should be able to address such adaptive localization of hypervariability.

Collectively, these studies provide evidence supporting our theoretical model and validate our proposed error correction mechanism whereby a reduction in extension rate following incorrect base pairing allows for effective error correction in self-replicators.

6. Summary

we have developed a theoretical model demonstrating the possibility of non-enzymatic error correction during self-replication without any extraneous energy inputs, utilizing only the thermodynamic drive for template-directed polymerization. The central *premise* of our model is the existence of asymmetric cooperativity [39, 40], defined as unequal *kinetic* influence of a base pair on its two neighbors, which is assumed to be present in correct base pairs and absent in incorrect base pairs (Fig.2). This kinetic discrimination is shown to lead to a time lag between correct and incorrect daughter strand constructions, which is exploited by covalent bond catalysis, leading to error correction. Below, we list the central results of the model.

I. Unidirectional strand construction is necessary for error correction

Our central premise of the presence of asymmetric cooperativity in non-enzymatically self-replicating heteropolymers has been shown to lead to unidirectional daughter strand construction [39]. Thus, both unidirectionality and error correction are intimately connected and originate from the same kinetic property, asymmetric cooperativity. Abolishing unidirectionality abolishes error correction ability (Fig.4).

II. Error correction requires rapid, irreversible covalent bond formation

Time lag between correct and incorrect product formation has been central to nearly all error correction mechanisms proposed thus far [28, 29]. Our model too introduces a lag between correct and incorrect daughter strand formation times, through kinetic discrimination. This time lag is shown here to be useful for error correction only under irreversible conditions, which is provided by the highly exergonic covalent bond formation between daughter strand monomers during self-replication. We show that covalent bond catalysis is crucial, and sufficient for error correction, and, when the covalent bonding rates match or exceed the rate of the correct daughter strand assembly through base pairing, error correction ensues (Fig.7 & Fig.8). No structure-based discriminators between correct and incorrect base pairs are necessary.

III. Passive polymerases correct errors simply by catalyzing covalent bond formation

The energetic basis of base selection by passive polymerases has been unclear[13–18]. Our model resolves the issue by showing that error correction follows as a consequence of catalysis of covalent bonds by passive polymerases (Fig.7). The energy input for error correction comes from the covalent bond formation itself. This eliminates the need for sophisticated sensing of base pair geometries by passive polymerases for base selection. This would allow us to concentrate our search for simple inorganic covalent bond catalysts in order to elucidate primordial scenarios that promoted the rate, and hence the accuracy of non-enzymatic self-replication.

IV. The model reproduces several experimental signatures associated with error correction

Several phenomena have been shown to be associated with error correction in the literature: (a) stalling, the slow-down of strand construction after incorrect base pair formation [41–43] (Fig.10), (b) fraying, the unzipping of base pairs behind an incorrect base pair[48] (Fig.11), (c) next nucleotide effects, increased error rates at high thermodynamic drives and near other incorrect base pairs[42, 49], (d) speed-accuracy trade-off, a non-linear dependence of error rates on the thermodynamic drive for strand construction[44, 45, 47], and (e) adaptive tuning of error rates, the dependence of error rates on monomer supply and other environmental variables [50, 51] (Fig.6a). Our model reproduces all of the above experimentally observed phenomena.

7. Discussion

A vexing problem concerning the possibility of autonomous emergence of early life has been our inability to demonstrate error-free, non-enzymatic self-replication, and hence faithful transfer of information across generations, under plausible primordial scenarios [82]. A more immediate concern is the lack of clear understanding of the energetics of error

correction during DNA replication, specifically of base selection, which does not involve utilization of extraneous energy sources, making the Hopfield-Ninio model inapplicable [13–18]. Base selection is a major contributor to the accuracy of DNA replication, improving the thermodynamically set error rates of about 10^{-2} per base pair to about 10^{-4} per base pair[83]. Other questions include the role of covalent bond formation and of polymerases in error correction[19].

The model we introduced above clarifies the contributions of strand construction kinetics and thermodynamics to error correction, and demonstrates the possibility of non-enzymatic error correction through kinetic discrimination between correct and incorrect base pairs and covalent bond catalysis. We thus identify the source of energy for base selection process as the energy released irreversibly during covalent bond formation itself. No extraneous energy sources are needed for error correction, in contrast to existing models of error correction. This resolves, at least theoretically, the problem of identifying molecules that were capable of discriminating between correct and incorrect base pairs in the primordial scenario, using energy supply. Our answer to this problem is that we do not need a structure-based discriminator[84]: all we need to identify are molecule(s) capable of catalyzing the covalent bond formation in the primordial scenario. Error correction follows, since the asymmetric kinetics already discriminates between correct and incorrect base pairs through time delays in strand construction. This idea considerably simplifies our search for primordial scenarios for evolution of self-replicating informational heteropolymers. With reasonable kinetic and thermodynamic parameters, we reproduce the accuracy of passive polymerases[13] in the base selection process.

Our identification of the source of energy for error correction as the thermodynamic drive for templated strand construction opens up the possibility of modulating the error rates by varying the thermodynamic drive, specifically by varying monomer concentration. Although other environmental parameters, such as temperature, pH and salinity influence the thermodynamic drive, the possibility of living systems tuning the monomer supply rates to optimize the error rates appears more probable[60, 85, 86]. This might explain the tight regulation of nucleotide supply rates in extant life forms[87].

Existing theoretical models of error correction in the literature[28–30], apart from utilizing extraneous energy sources, also include multiple parameters to describe the error correction process, thereby allowing degrees of freedom in the model that may not correspond to the ground truth *in vivo*. This leads to the exasperating task of explaining how these parameters are collectively fine-tuned by evolution to achieve the incredible levels of accuracy seen in DNA replication, a task usually left unfinished. The model we introduced above includes a *single free parameter*, α, describing the asymmetrically cooperative kinetics of strand construction. The remaining are thermodynamic variables, the free energies of correct and incorrect base pair formation and dissociation, whose values and their environmental dependence have been previously determined from extensive studies. This reduction in the parameter space allows for an intuitive understanding of the error correction process without constraining ourselves into specifying regions of parameter space where such an understanding is appropriate. Thus, the model provides conceptual clarity and simplicity, and helps in deeper understanding of the error correction mechanism in DNA replication.

One of the most fundamental questions regarding origin of life concerns the ability of living systems to create order from non-equilibrium[88]. This order is persistent across time, exemplified by our ability to sequence DNA of long-extinct species, and from the persistence of seeds and spores through inhospitable environments, without any energy supply to maintain order. We contrast this persistence with the ephemerality of the order from non-equilibrium in systems such as convection cells and vortices in turbulent fluids, put forward by the Prigogine's Brussels School as precursors of the order seen in life. We simply refer the reader to P. W. Anderson's article[89] to clarify why the examples provided by the Brussels school may be inappropriate for explaining the order found in living systems, none of which show the property of temporal persistence. Here, we provide a model system capable of creating temporally persistent order from non-equilibrium, without requiring the presence of a discriminatory macromolecule, an enzyme. The irreversibility of covalent bond formation, engendered by non-equilibrium created by high initial concentrations of free nucleotide triphosphates, leads to temporal persistence, in our model. This irreversibility, when coupled with the temporal discrimination arising from a kinetic asymmetry, leads to error correction and hence to the preservation of the evolutionarily dictated "order" during DNA replication. This model system thus provides a plausible scenario for the emergence of accurate information transfer across generations without enzymatic or energetic assistance from sources extraneous to the system, requiring only simple inorganic catalysts, such as metal ions, to catalyze covalent bond formation.

It is imperative that we answer our own question raised in the "Base Selection Energetics" subsection, within the "Introduction" section. What is the energy source for the error correction? We have claimed in the article that the irreversible covalent bond formation provides the necessary energy needed for error correction, although we did not quantitatively evaluate it. The free energy change due to covalent bond formation, with our assumption of the impossibility of covalent bond breaking, is infinity. This approximation therefore disallows us from quantifying the energy needed to create and maintain order. One could include very slow reverse transitions, allowing for the possibility of covalent bond breaking, in the model, catalyzed by the very same passive polymerases that catalyze the covalent bond

formation, thereby making the model thermodynamically more accurate. However, in extant life, evolution precludes this possibility of loss of order by immediately hydrolyzing the other product of the forward reaction, pyrophosphate, before it can be used for the reverse reaction, effectively making the reaction irreversible[90]. Thus, although we can point to the energy released during covalent bond formation as the energy source for error correction, quantifying it by including the reverse transitions would be inaccurate, requiring more sophistication in the model than can be accommodated here.

The central aim of this article is to construct a highly simplified *model* that reproduces the essential features of error correcting systems, that can provide a deeper understanding of DNA replication fidelity. This necessitates a bare-bones approach, which led us to leave out many interactions that may have a bearing on the error correction ability, such as the nearest-neighbor thermodynamic interactions (stacking) between base pairs[91, 92], the dependence of error rates on local sequence characteristics[81], environmental effects on base pair bonding dynamics[93] , and a multitude of other interactions that would be present in any highly evolved biological process. Even the assumption of the relationship between the two asymmetric cooperativity parameters, α and β, is made solely for simplicity, to reduce the number of parameters. Other assumptions, that only one location on the template is error-prone, that only two neighboring base pairs on either side affect the error rate, and that the covalent bonding is completely irreversible, are also made in a similar vein. Relaxing these assumptions may provide a more accurate picture closer to reality, but might diminish the results' interpretability and the model's pedagogical utility, in our view. It will be challenging to increase the length of the sequences that can be subjected to our analysis, since the dynamics of strand construction range multiple timescales, due to asymmetric cooperativity, leading to ill-conditioned transition matrices for longer sequences. Nevertheless, substantial improvements can be made to the model without jeopardizing its simplicity, for instance, by including sequence-dependent effects on error rates and matching them to experimentally observed sequence-dependent error rates[81]. In any case, the final arbiters of truth are always experiments that directly probe for the presence of asymmetric cooperativity in DNA replication, and examine its effects on error rates, possibly by modulating the cooperativity parameters through environmental perturbations or by any other means experimentally feasible. Identification of primordially plausible inorganic catalysts of covalent bond formation between nucleotides, such as montmorillonite clay[70, 94], would also strengthen our claims above. In conclusion, we have created a single free parameter theoretical model of error correction during DNA replication demonstrating the possibility of utilizing the thermodynamic drive for self-replication itself for error correction, without needing any extraneous energy sources. We showed that the model reproduces several experimentally observed phenomena associated with error correction, and provides an explanation for the ability of thermodynamically passive polymerases to preferentially select correct base pairs. We noted that instantiations of the model are plausible in the primordial scenario because of their inherent simplicity. We showed that it is possible to create persistent order from free energy flow, provided we begin with appropriate symmetry-broken systems.

Statements and Declarations

Data availability

The codes used to simulate the Markov chain dynamics is available from the authors upon request.

Competing interests

The authors declare no competing interests.

Acknowledgments

Support for this work was provided by the Science & Engineering Research Board (SERB), Department of Science and Technology (DST), India, through a Core Research Grant with file no. CRG/2020/003555 and a MATRICS grant with file no. MTR/2022/000086.

References

[1] Charles Darwin. *On the origin of species: A facsimile of the first edition.* Harvard University Press, 1964.

[2] William Bateson and Gregor Mendel. *Mendel's principles of heredity.* Courier Corporation, 2013.

[3] John J Hopfield. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences*, 71(10):4135–4139, 1974.

[4] Jacques Ninio. Kinetic amplification of enzyme discrimination. *Biochimie*, 57(5):587–595, 1975.

[5] Manfred Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58:465–523, 1971.

[6] Manfred Eigen and Peter Schuster. *The hypercycle: a principle of natural self-organization*. Springer Science & Business Media, 2012.

[7] Arvind Murugan, David A Huse, and Stanislas Leibler. Speed, dissipation, and error in kinetic proofreading. *Proceedings of the National Academy of Sciences*, 109(30):12034–12039, 2012.

[8] T Yamane and JJ Hopfield. Experimental evidence for kinetic proofreading in the aminoacylation of tRNA by synthetase. *Proceedings of the National Academy of Sciences*, 74(6):2246–2250, 1977.

[9] JJ Hopfield, T Yamane, V Yue, and SM Coutts. Direct experimental evidence for kinetic proofreading in amino acylation of tRNAIle. *Proceedings of the National Academy of Sciences*, 73(4):1164–1168, 1976.

[10] Scott C Blanchard, Ruben L Gonzalez Jr, Harold D Kim, Steven Chu, and Joseph D Puglisi. tRNA selection and kinetic proofreading in translation. *Nature structural & molecular biology*, 11(10):1008–1014, 2004.

[11] Joel D Mallory, Anatoly B Kolomeisky, and Oleg A Igoshin. Trade-offs between error, speed, noise, and energy dissipation in biological processes with proofreading. *The Journal of Physical Chemistry B*, 123(22):4718–4725, 2019.

[12] Qiwei Yu, Joel D Mallory, Anatoly B Kolomeisky, Jiqiang Ling, and Oleg A Igoshin. Trade-offs between speed, accuracy, and dissipation in tRNAIle aminoacylation. *The journal of physical chemistry letters*, 11(10):4001–4007, 2020.

[13] Robert A Beckman and Lawrence A Loeb. Multi-stage proofreading in DNA replication. *Quarterly reviews of biophysics*, 26(3):225–331, 1993.

[14] Scott D McCulloch and Thomas A Kunkel. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell research*, 18(1):148–161, 2008.

[15] Harrison Echols and Myron F Goodman. Fidelity mechanisms in DNA replication. *Annual review of biochemistry*, 60(1):477–511, 1991.

[16] Myron F Goodman, Steven Creighton, Linda B Bloom, John Petruska, and Thomas A Kunkel. Biochemical basis of DNA replication fidelity. *Critical reviews in biochemistry and molecular biology*, 28(2):83–126, 1993.

[17] Andrew C Olson, Jennifer N Patro, Milan Urban, and Robert D Kuchta. The energetic difference between synthesis of correct and incorrect base pairs accounts for highly accurate DNA replication. *Journal of the american chemical society*, 135(4):1205–1208, 2013.

[18] Shina CL Kamerlin, Pankaz K Sharma, Ram B Prasad, and Arieh Warshel. Why nature really chose phosphate. *Quarterly reviews of biophysics*, 46(1):1–132, 2013.

[19] Keriann Oertell, Emily M Harcourt, Michael G Mohsen, John Petruska, Eric T Kool, and Myron F Goodman. Kinetic selection vs. free energy of DNA base pairing in control of polymerase fidelity. *Proceedings of the National Academy of Sciences*, 113(16):E2277–E2285, 2016.

[20] Thomas A Kunkel. Evolving views of dna replication (in) fidelity. In *Cold Spring Harbor symposia on quantitative biology*, volume 74, pages 91–101. Cold Spring Harbor Laboratory Press, 2009.

[21] John Petruska, Lawrence C Sowers, and Myron F Goodman. Comparison of nucleotide interactions in water, proteins, and vacuum: model for DNA polymerase fidelity. *Proceedings of the National Academy of Sciences*, 83(6):1559–1562, 1986.

[22] Michael S Boosalis, J Petruska, and MF Goodman. DNA polymerase insertion fidelity. Gel assay for site-specific kinetics. *Journal of Biological Chemistry*, 262(30):14689–14696, 1987.

[23] David Andrieux and Pierre Gaspard. Nonequilibrium generation of information in copolymerization processes. *Proceedings of the National Academy of Sciences*, 105(28):9516–9521, 2008.

[24] Paul G Higgs and Niles Lehman. The RNA World: molecular cooperation at the origins of life. *Nature Reviews Genetics*, 16(1):7–17, 2015.

[25] Gerald F Joyce and Jack W Szostak. Protocells and rna self-replication. *Cold Spring Harbor Perspectives in Biology*, 10(9):a034801, 2018.

[26] Patrizia Hagenbuch, Eric Kervio, Annette Hochgesand, Ulrich Plutowski, and Clemens Richert. Chemical primer extension: efficiently determining single nucleotides in DNA. *Angewandte Chemie*, 117(40):6746–6750, 2005.

[27] Jianyang Han, Eric Kervio, and Clemens Richert. High Fidelity Enzyme-Free Primer Extension with an Ethynylpyridone Thymidine Analog. *Chemistry–A European Journal*, 27(64):15918–15921, 2021.

[28] Tobias Göppel, Benedikt Obermayer, Irene A Chen, and Ulrich Gerland. A kinetic error filtering mechanism for enzyme-free copying of nucleic acid sequences. *bioRxiv*, pages 2021–08, 2021.

[29] Yoshiya J Matsubara, Nobuto Takeuchi, and Kunihiko Kaneko. Avoidance of error catastrophe via proofreading innate to template-directed polymerization. *Physical Review Research*, 5(1):013170, 2023.

[30] Rakesh Mukherjee, Aditya Sengar, Javier Cabello-García, and Thomas E Ouldridge. Kinetic proofreading can enhance specificity in a nonenzymatic DNA strand displacement network. *Journal of the American Chemical Society*, 146(28):18916–18926, 2024.

[31] Vahe Galstyan, Kabir Husain, Fangzhou Xiao, Arvind Murugan, and Rob Phillips. Proofreading through spatial gradients. *Elife*, 9:e60415, 2020.

[32] Anthonie WJ Muller and Dirk Schulze-Makuch. Thermal energy and the origin of life. *Origins of Life and Evolution of Biospheres*, 36:177–189, 2006.

[33] Laurent Boiteau and Robert Pascal. Energy sources, self-organization, and the origin of life. *Origins of Life and Evolution of Biospheres*, 41(1):23–33, 2011.

[34] Kepa Ruiz-Mirazo, Carlos Briones, and Andres de la Escosura. Prebiotic systems chemistry: new perspectives for the origins of life. *Chemical reviews*, 114(1):285–366, 2014.

[35] Ilya Prigogine and Grégoire Nicolis. Self-organisation in nonequilibrium systems: towards a dynamics of complexity. *Bifurcation analysis*, pages 3–12, 1985.

[36] Todd R Gingrich, Jordan M Horowitz, Nikolay Perunov, and Jeremy L England. Dissipation bounds all steady-state current fluctuations. *Physical review letters*, 116(12):120601, 2016.

[37] Jordan M Horowitz, Kevin Zhou, and Jeremy L England. Minimum energetic cost to maintain a target nonequilibrium state. *Physical Review E*, 95(4):042102, 2017.

[38] Jordan M Horowitz and Jeremy L England. Spontaneous fine-tuning to environment in many-species chemical reaction networks. *Proceedings of the National Academy of Sciences*, 114(29):7565–7570, 2017.

[39] Hemachander Subramanian and Robert A Gatenby. Evolutionary advantage of directional symmetry breaking in self-replicating polymers. *Journal of theoretical biology*, 446:128–136, 2018.

[40] Hemachander Subramanian and Robert A Gatenby. Evolutionary advantage of anti-parallel strand orientation of duplex DNA. *Scientific Reports*, 10(1):9883, 2020.

[41] Sudha Rajamani, Justin K Ichida, Tibor Antal, Douglas A Treco, Kevin Leu, Martin A Nowak, Jack W Szostak, and Irene A Chen. Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. *Journal of the American Chemical Society*, 132(16):5880–5885, 2010.

[42] Kevin Leu, Eric Kervio, Benedikt Obermayer, Rebecca M Turk-MacLeod, Caterina Yuan, Jesus-Mario Luevano Jr, Eric Chen, Ulrich Gerland, Clemens Richert, and Irene A Chen. Cascade of reduced speed and accuracy after errors in enzyme-free copying of nucleic acid sequences. *Journal of the American Chemical Society*, 135(1):354–366, 2013.

[43] Tobias Göppel, Joachim H Rosenberger, Bernhard Altaner, and Ulrich Gerland. Thermodynamic and kinetic sequence selection in enzyme-free polymer self-assembly inside a non-equilibrium RNA reactor. *Life*, 12(4):567, 2022.

[44] Riccardo Ravasio, Kabir Husain, Constantine G Evans, Rob Phillips, Marco Ribezzi, Jack W Szostak, and Arvind Murugan. A minimal scenario for the origin of non-equilibrium order. *arXiv preprint arXiv:2405.10911*, 2024.

[45] Kinshuk Banerjee, Anatoly B Kolomeisky, and Oleg A Igoshin. Elucidating interplay of speed and accuracy in biological error correction. *Proceedings of the National Academy of Sciences*, 114(20):5183–5188, 2017.

[46] Jenny M Poulton, Pieter Rein Ten Wolde, and Thomas E Ouldridge. Nonequilibrium correlations in minimal dynamical models of polymer copying. *Proceedings of the National Academy of Sciences*, 116(6):1946–1951, 2019.

[47] Vahe Galstyan and Rob Phillips. Allostery and kinetic proofreading. *The Journal of Physical Chemistry B*, 123(51):10990–11002, 2019.

[48] Rafael Fernandez-Leiro, Julian Conrad, Ji-Chun Yang, Stefan MV Freund, Sjors HW Scheres, and Meindert H Lamers. Self-correcting mismatches during high-fidelity DNA replication. *Nature structural & molecular biology*, 24(2):140–143, 2017.

[49] Thomas A Kunkel, RM Schaaper, RA Beckman, and LA Loeb. On the fidelity of DNA replication. Effect of the next nucleotide on proofreading. *Journal of Biological Chemistry*, 256(19):9883–9889, 1981.

[50] Toon Swings, Bram Van den Bergh, Sander Wuyts, Eline Oeyen, Karin Voordeckers, Kevin J Verstrepen, Maarten Fauvart, Natalie Verstraeten, and Jan Michiels. Adaptive tuning of mutation rates allows fast response to lethal stress in Escherichia coli. *Elife*, 6:e22939, 2017.

[51] Troy G Hammerstrom, Kathryn Beabout, Thomas P Clements, Gerda Saxer, and Yousif Shamoo. Acinetobacter baumannii repeatedly evolves a hypermutator phenotype in response to tigecycline that effectively surveys evolutionary trajectories to resistance. *PloS one*, 10(10):e0140489, 2015.

[52] Raymond Dean Astumian. Kinetic Asymmetry and directionality of nonequilibrium molecular systems. *Angewandte Chemie*, 136(9):e202306569, 2024.

[53] NG Van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland Publishing Co, 1992.

[54] William J Anderson. *Continuous-time Markov chains: An applications-oriented approach*. Springer Science & Business Media, 2012.

[55] Oliver Ibe. *Markov processes for stochastic modeling*. Newnes, 2013.

[56] Robert P Dobrow. *Introduction to stochastic processes with R*. John Wiley & Sons, 2016.

[57] Peter M Burgers. Solution to the 50-year-old Okazaki-fragment problem. *Proceedings of the National Academy of Sciences*, 116(9):3358–3360, 2019.

[58] Sean P Fagan, Purba Mukherjee, William J Jaremko, Rachel Nelson-Rigg, Ryan C Wilson, Tyler L Dangerfield, Kenneth A Johnson, Indrajit Lahiri, and Janice D Pata. Pyrophosphate release acts as a kinetic checkpoint during high-fidelity DNA replication by the Staphylococcus aureus replicative polymerase PolC. *Nucleic Acids Research*, 49(14):8324–8338, 2021.

[59] Paul C Bressloff. Search processes with stochastic resetting and multiple targets. *Physical Review E*, 102(2):022115, 2020.

[60] Kingsley L Wong and Juewen Liu. Factors and methods to modulate DNA hybridization kinetics. *Biotechnology journal*, 16(11):2000338, 2021.

[61] Stefan Howorka, Liviu Movileanu, Orit Braha, and Hagan Bayley. Kinetics of duplex formation for individual DNA strands within a single protein nanopore. *Proceedings of the National Academy of Sciences*, 98(23):12996–13001, 2001.

[62] Elizaveta Guseva, Ronald N Zuckermann, and Ken A Dill. Foldamer hypothesis for the growth and sequence differentiation of prebiotic polymers. *Proceedings of the National Academy of Sciences*, 114(36):E7460–E7468, 2017.

[63] Thomas E Ouldridge, Petr Šulc, Flavio Romano, Jonathan PK Doye, and Ard A Louis. DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic acids research*, 41(19):8886–8895, 2013.

[64] Jianshu Cao. Michaelis- Menten Equation and Detailed Balance in Enzymatic Networks. *The Journal of Physical Chemistry B*, 115(18):5493–5498, 2011.

[65] Jithesh Kottur and Deepak T Nair. Pyrophosphate hydrolysis is an intrinsic and critical step of the DNA synthesis reaction. *Nucleic Acids Research*, 46(12):5875–5885, 2018.

[66] Isaac Wong, Smita S Patel, and Kenneth A Johnson. An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry*, 30(2):526–537, 1991.

[67] John W Brandis, Sydney G Edwards, and Kenneth A Johnson. Slow rate of phosphodiester bond formation accounts for the strong bias that Taq DNA polymerase shows against 2 ‘, 3 ‘-dideoxynucleotide terminators. *Biochemistry*, 35(7):2189–2200, 1996.

[68] Bret D Freudenthal, William A Beard, David D Shock, and Samuel H Wilson. Observing a DNA polymerase choose right from wrong. *Cell*, 154(1):157–168, 2013.

[69] Anuraag Aithal, Shikha Dagar, and Sudha Rajamani. Metals in prebiotic catalysis: a possible evolutionary pathway for the emergence of metalloproteins. *ACS omega*, 8(6):5197–5208, 2023.

[70] James P Ferris, Aubrey R Hill Jr, Rihe Liu, and Leslie E Orgel. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, 381(6577):59–61, 1996.

[71] Pablo Sartori and Simone Pigolotti. Kinetic versus energetic discrimination in biological copying. *Physical review letters*, 110(18):188101, 2013.

[72] Arvind Murugan, David A Huse, and Stanislas Leibler. Discriminatory proofreading regimes in nonequilibrium systems. *Physical Review X*, 4(2):021016, 2014.

[73] Chunhong Long and Jin Yu. Balancing Non-Equilibrium driving with nucleotide selectivity at kinetic checkpoints in polymerase fidelity control. *Entropy*, 20(4):306, 2018.

[74] Brian Munsky, Ilya Nemenman, and Golan Bel. Specificity and completion time distributions of biochemical processes. *The Journal of chemical physics*, 131(23), 2009.

[75] Riccardo Rao and Luca Peliti. Thermodynamics of accuracy in kinetic proofreading: dissipation and efficiency trade-offs. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(6):P06001, 2015.

[76] Kristian Le Vay, Elia Salibi, Emilie Y Song, and Hannes Mutschler. Nucleic acid catalysis under potential prebiotic conditions. *Chemistry–An Asian Journal*, 15(2):214–230, 2020.

[77] Oliver J Rando and Kevin J Verstrepen. Timescales of genetic and epigenetic inheritance. *Cell*, 128(4):655–668, 2007.

[78] Jeffrey H Chuang and Hao Li. Functional bias and spatial organization of genes in mutational hot and cold regions in the human genome. *PLoS biology*, 2(2):e29, 2004.

[79] Marjan W Van Der Woude and Andreas J Bäumler. Phase and antigenic variation in bacteria. *Clinical microbiology reviews*, 17(3):581–611, 2004.

[80] J David Barry and Richard McCulloch. Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Advances in Parasitology*, 49:1–70, 2001.

[81] Eric Kervio, Annette Hochgesand, Ulrich E Steiner, and Clemens Richert. Templating efficiency of naked DNA. *Proceedings of the National Academy of Sciences*, 107(27):12074–12079, 2010.

[82] Eörs Szathmáry. The origin of replicators and reproducers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1474):1761–1776, 2006.

[83] Thomas A Kunkel. DNA replication fidelity. *Journal of Biological Chemistry*, 279(17):16895–16898, 2004.

[84] William A Beard and Samuel H Wilson. Structural insights into the origins of DNA polymerase fidelity. *Structure*, 11(5):489–496, 2003.

[85] Alexandra M De Paz, Thaddeus R Cybulski, Adam H Marblestone, Bradley M Zamft, George M Church, Edward S Boyden, Konrad P Kording, and Keith E J Tyo. High-resolution mapping of DNA polymerase fidelity using nucleotide imbalances and next-generation sequencing. *Nucleic acids research*, 46(13):e78–e78, 2018.

[86] Rais A Ganai and Erik Johansson. DNA replication—a matter of fidelity. *Molecular cell*, 62(5):745–755, 2016.

[87] Thomas W Traut. Physiological concentrations of purines and pyrimidines. *Molecular and cellular biochemistry*, 140:1–22, 1994.

[88] Erwin Schrödinger. *What is life?: With mind and matter and autobiographical sketches*. Cambridge university press, 1992.

[89] Philip W Anderson and Daniel L Stein. Broken symmetry, emergent properties, dissipative structures, life: Are they related? In *Basic notions of condensed matter physics*, pages 263–277. CRC Press, 2018.

[90] Jessica LE Wimmer, Karl Kleinermanns, and William F Martin. Pyrophosphate and irreversibility in evolution, or why PPi is not an energy currency and why nature chose triphosphates. *Frontiers in microbiology*, 12:759359, 2021.

[91] John Petruska and MF Goodman. Influence of neighboring bases on DNA polymerase insertion and proofreading fidelity. *Journal of Biological Chemistry*, 260(12):7533–7539, 1985.

[92] Peter Yakovchuk, Ekaterina Protozanova, and Maxim D Frank-Kamenetskii. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic acids research*, 34(2):564–574, 2006.

[93] Kristin A Eckert and Thomas A Kunkel. DNA polymerase fidelity and the polymerase chain reaction. *Genome research*, 1(1):17–24, 1991.

[94] James P Ferris. Montmorillonite-catalysed formation of RNA oligomers: the possible role of catalysis in the origins of life. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1474):1777–1786, 2006.

SUPPLEMENTARY MATERIAL

The aim of this supplementary material is to provide an approximate, analytically tractable Markov chain model for error correction, that captures the key observation that "covalent bond catalysis leads to error correction" in the main article. The variables and parameters used here are the same as defined in the Methods section of the main article. The analytical procedure followed here closely parallels the one used in [45].

We reduce the length of the template to $N = 3$, and assume that the construction of the daughter strand proceeds in a *strictly unidirectional manner*. This constraint reduces the dimensionality of the state space, facilitating analytical treatment. We characterize the dynamics in terms of a first-passage framework, wherein the principal quantity of interest is the first-passage time probability density function[53] to measure the temporal characteristics of reaching a target state for the first time.

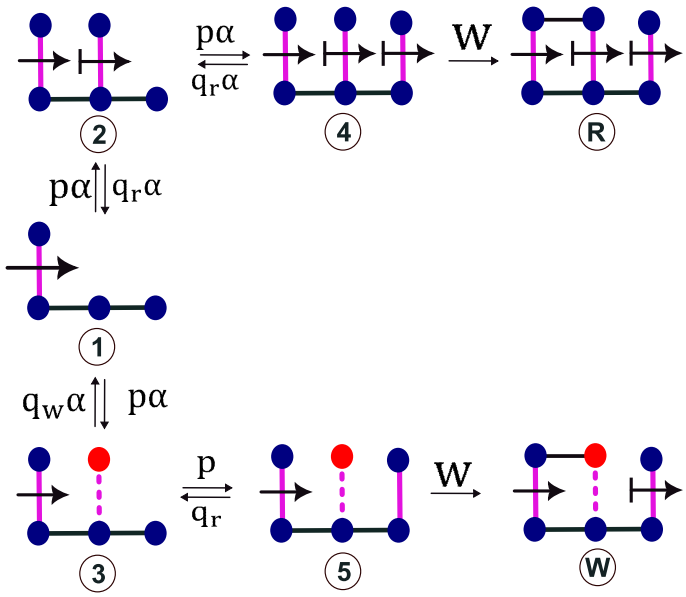


Figure 12: Minimal, approximate continuous-time Markov model for error correction with asymmetric cooperativity. Transient states represent intermediate hydrogen-bonded base-pairing configurations, while absorbing states correspond to the covalent incorporation of either a correct (R) or incorrect (W) nucleotide. The dynamics are governed by the base-pair formation rates(p), dissociation rates($q_{r(w)}$),the covalent locking rate w, and kinetic parameter α that modulates transition rates.

We use the backward (adjoint) master-equation approach to compute first-passage probability density. Let $f_{i\to\mathbf{R}}$ denote the first-passage probability density that, starting in state $\mathbf{i}$ at $t = 0$, the process is first absorbed in $\mathbf{R}$ at time t. Similarly define $f_{i\to\mathbf{W}}$ for absorption in $\mathbf{W}$. To solve them we introduce, $\mathbf{f_R(t)} = (F_{1\to R}, F_{2\to R}, F_{3\to R}, F_{4\to R}, F_{5\to R})^T$ and $\mathbf{f_W(t)} = (F_{1\to W}, F_{2\to W}, F_{3\to W}, F_{4\to W}, F_{5\to W})^T$. These satisfy the backward master equations:

$$\frac{d\mathbf{f_{R/W}(t)}}{dt} = \mathbf{Q}.\mathbf{f_{R/W}(t)} + \mathbf{M_{R/W}} \tag{10}$$

Here the term $\mathbf{Q}.\mathbf{f_{R/W}(t)}$ propagates probability through the hydrogen-bonded intermediates according to the usual adjoint chain dynamics, while the source vectors $\mathbf{M_R}$ and $\mathbf{M_W}$ inject probability corresponding to an immediate covalent-bond transition into $\mathbf{R}$ and $\mathbf{W}$, inject probability corresponding to an immediate covalent-bond transition into $\mathbf{R}$ and $\mathbf{W}$.

Now, from the network in Fig.12 $\mathbf{M_R} = [0, 0, 0, w\delta(t), 0]^T$ and $\mathbf{M_W} = [0, 0, 0, 0, w\delta(t)]^T$, $\delta(t)$ being the Dirac delta function. The transition matrix is given by

$$\mathbf{Q} = \begin{bmatrix} -2p\alpha & p\alpha & p\alpha & 0 & 0 \\ q_r\alpha & -(q_r\alpha + p\alpha) & 0 & p\alpha & 0 \\ q_w\alpha & 0 & -(q_w\alpha + p) & 0 & p \\ 0 & q_r\alpha & 0 & -(q_r\alpha + w) & 0 \\ 0 & 0 & q_r & 0 & -(q_r + w) \end{bmatrix} \tag{11}$$

The Laplace transform fo $f_{i\to\mathbf{R}/\mathbf{W}}(t)$ is defined as:

$$\tilde{f}_{i\to\mathbf{R}/\mathbf{W}}(s) = \int_0^\infty e^{-st} f_{i\to\mathbf{R}/\mathbf{W}}(t)\, dt \tag{12}$$

Taking the Laplace transform of the backward master Eq.10, and $\mathbf{f_{i\to R/W}}(t=0) = 0$ for all $(i \neq \mathbf{R}/\mathbf{W})$, we obtain:

$$s\tilde{\mathbf{f}}_{\mathbf{R}/\mathbf{W}} = \mathbf{Q}.\tilde{\mathbf{f}}_{\mathbf{R}/\mathbf{W}} + \tilde{\mathbf{M}}_{\mathbf{R}/\mathbf{W}} \tag{13}$$

Here, $\tilde{\mathbf{f}}_{\mathbf{R}/\mathbf{W}} = (\tilde{\mathbf{f}}_{\mathbf{1}\to\mathbf{R}/\mathbf{W}}, \tilde{\mathbf{f}}_{\mathbf{2}\to\mathbf{R}/\mathbf{W}}, \tilde{\mathbf{f}}_{\mathbf{3}\to\mathbf{R}/\mathbf{W}}, \tilde{\mathbf{f}}_{\mathbf{4}\to\mathbf{R}/\mathbf{W}}, \tilde{\mathbf{f}}_{\mathbf{5}\to\mathbf{R}/\mathbf{W}})^T$, $\tilde{\mathbf{M}}_{\mathbf{R}} = [0,0,0,w,0]^T$ and $\tilde{\mathbf{M}}_{\mathbf{W}} = [0,0,0,0,w]^T$.

The Laplace-transformed first-passage probability densities $\tilde{\mathbf{f}}_{\mathbf{i}\to\mathbf{R}/\mathbf{W}}(s)$ are then obtained by solving the algebraic set of Eqs.13. The *splitting probabilities*, which describe the likelihood of eventually reaching either absorbing state, are obtained by evaluating $\tilde{\mathbf{f}}_{\mathbf{R}/\mathbf{W}}$ at $s = 0$:

$$\Pi_{\mathbf{R}/\mathbf{W}} = \tilde{\mathbf{f}}_{\mathbf{i}\to\mathbf{R}/\mathbf{W}}(s)|_{s=0} \tag{14}$$

From the analytical solution of the model, the ratio of splitting probabilities is given by:

$$\frac{\Pi_{\mathbf{R}}}{\Pi_{\mathbf{W}}} = \frac{pw + \alpha q_w(q_r + w)}{\alpha q_r^2 + (p + q_r)w} \tag{15}$$

This expression exhibits two distinct limiting behaviors:

Thermodynamic regime, w is low:

$$\frac{\Pi_{\mathbf{R}}}{\Pi_{\mathbf{W}}} \approx \frac{aq_w q_r}{aq_r^2} = \frac{q_w}{q_r} = 100 \tag{16}$$

Therefore, the accuracy of the daughter strand construction is purely controlled by the base-pair dissociation rates. The system has sufficient time to equilibrate before covalent locking occurs, so accuracy is entirely thermodynamically controlled.

Kinetic regime, w is high and $p << \alpha q_w$:

$$\frac{\Pi_{\mathbf{R}}}{\Pi_{\mathbf{W}}} \approx \frac{\alpha q_w}{p + q_r} = 1.6 \times 10^5 \tag{17}$$

Here, rapid covalent bond formation prevents equilibration. Instead, accuracy depends on the kinetic accessibility of each pathway, and thus incorporates dynamic discrimination factors.