Polymerization and replication of primordial RNA induced by clay-water interface dynamics

Check for updates

Carla Alejandre [1,2,4], Adrián Aguirre-Tamaral [1,3,4], Carlos Briones [1] & Jacobo Aguirre [1,2] ✉

In the study of life's origins, a key challenge is understanding how RNA could have polymerized and subsequently replicated in early Earth. We present a theoretical and computational framework to model the non-enzymatic polymerization of ribonucleotides and the template-dependent replication of primordial RNA molecules, at the interfaces between the aqueous solution and a clay mineral. Our results demonstrate that systematic polymerization and replication of single-stranded RNA polymers, sufficiently long to fold and acquire basic functions (>15 nt), were feasible under these conditions. Crucially, this process required a physico-chemical environment characterized by large-amplitude oscillations with periodicity compatible with spring tide dynamics, suggesting that large moons may have played a role in the emergence of RNA-based life on planetary bodies. Interestingly, the theoretical analysis presents rigorous evidence that RNA replication efficiency increases in oscillating environments compared to constant ones. Moreover, the versatility of our framework enables comparisons between different genetic alphabets, showing that a four-letter alphabet —particularly when allowing non-canonical base pairs, as in current RNA— represents an optimal balance of replication speed and sequence diversity in the pathway to life.

The origin of life on Earth is a fundamental question that remains elusive when we aim to fully describe it in physico-chemical terms. For a complex enough molecular system to be considered alive, it must exhibit and combine a number of features such as compartmentalization, energy dissipation, replication of a genetic material and metabolism, thus allowing its self-reproduction and open-ended evolution. In the earlier stages of the pre-cellular world, simple mechanisms at the molecular level, still not catalyzed by enzymes, must have driven the increasing complexity of prebiotic molecules. Different sets of chemical species needed to be concentrated and separated from the external environment through some form of surface- or membrane-based compartmentalization. A small fraction of them could have been used as monomers to build (by means of non-enzymatic polymerization mechanisms) progressively longer and more complex polymeric molecules, some of which were endowed with heritable information that provided a substrate for Darwinian evolution to act upon[1–4].

Since the pioneering work of Aleksandr I. Oparin a century ago[5] and of John B.S. Haldane 5 years later[6], particularly over the past four decades, various hypotheses have emerged to try to understand the transition from chemistry to biology. One model that gained significant attention since the 1980s is the RNA World hypothesis[7–11]. According to its current version, in

the framework of prebiotic systems chemistry, the first self-reproducing entities contained single-stranded RNA (ssRNA) as genetic material, as well as a proto-metabolism based on ribozymes: structured ssRNA molecules endowed with catalytic activities (including RNA ligase and RNA polymerase), likely assisted by short abiotic peptides and low molecular weight compounds as cofactors[1,12,13].

However, before the establishment of an RNA (or RNA-peptide) world, the available and, somehow, chemically activated ribonucleotides (from now on, termed nts) should have polymerized into ssRNA oligomers in absence of the catalytic activity of ribozymes. Additionally, a non-enzymatic process of template-dependent RNA replication could have ensured the availability of populations of RNA molecules with related sequences, on which selection could act[14]. Indeed, the minimum length of an RNA polymerase ribozyme (i.e., a structured RNA molecule capable of replicating other RNA molecules used as templates) has been estimated in the range of 165 nts[15] to 182 nts[16–18], which is far longer than the up to 50-mer ssRNA molecules that can be experimentally obtained by means of non-enzymatic polymerization using activated nts and montmorillonite clay as a catalyst[19–21]. Therefore, random RNA polymerization of short oligomers, which is known to be favored by clays and other heterogeneous

[1]Centro de Astrobiología (CAB), CSIC-INTA, Torrejón de Ardoz, Madrid, Spain. [2]Grupo Interdisciplinar de Sistemas Complejos (GISC), Madrid, Spain. [3]Department of Biology, University of Graz, Graz, Austria. [4]These authors contributed equally: Carla Alejandre, Adrián Aguirre-Tamaral. ✉e-mail: jaguirre@cab.inta-csic.es

media[22–25], could have been followed by a ligation-based, stepwise process of modular evolution of RNA that allowed the assembly of a template-dependent RNA polymerase ribozyme[26,27]. Importantly, non-enzymatic, template-dependent RNA replication has been explored both experimentally[28–31] and through theoretical models, most of them focusing on liquid environments[32–44]. However, little work has focused on the entire process, connecting RNA polymerization and template-dependent replication in the same heterogeneous environment and making use of a global, systems chemistry approach[1,45,46]. Furthermore, the holistic perspective of complexity science is just starting to be exploited in this context, and our work is a step in this promising direction[47,48].

The vast diversity of geological environments that coexisted in early Earth opens up a multitude of possible chemical reactions and physical scenarios, not all equally conducive to the establishment of an RNA world. Under specific conditions, ssRNA polymers can grow when wet/dry cycles are present[49,50] or in environments subjected to extreme temperature changes[51,52]. Interestingly, extensive work suggests that oscillating environments were crucial for the emergence of life[53–58]. These fluctuating conditions were common in early Earth and could be of various types. For example, analyses of tidal interactions in the Earth-Moon system since its formation suggest that, when life first emerged on our planet, a day lasted ~10–12 h, resulting in tidal maxima at ocean shores every 5–6 h[59]. While Lathe proposed two decades ago that tides could have provided the oscillating driving force for the origin of replicating biopolymers[60,61], Fernando et al. and Chamanian et al. noted that linear templates subjected to external oscillations might provoke a biopolymer elongation that would hinder its accurate replication[35,43]. Anyway, temperature oscillations and wet/dry cycles have received the most attention: temperature cycles are known to promote the hybridization/denaturation of complementary ssRNA strands[62], while dehydration periods are essential to allow condensation reactions, such as the formation of 3′-5′ phosphodiester bonds necessary for RNA polymerization[55,63,64].

Numerous studies have experimentally demonstrated that the surfaces of certain minerals and rocks, in particular clays such as montmorillonite, exhibit chemical and crystallographical properties (including their microscopic structure, distribution of cations and spacing of the phyllosilicate interlayers) that can favor the adsorption and correct orientation of nts while providing a locally anhydrous environment, thus catalyzing the abiotic polymerization of RNA[19–23,56,65–67]. The importance of including clay surfaces in the prebiotic context is further highlighted by studies showing that they protect molecules from degradation[1] and concentrate organic molecules[68]. Also, clays catalyze different key processes, including the synthesis of nucleobases from their precursors[1], the formation of vesicles from fatty acid micelles[69], and the polymerization of amino acids into peptides[53]. Furthermore, the sedimentation of ssRNA oligomers on clay surfaces or other minerals induces a phase separation into oligonucleotide-dense and dilute phases that can facilitate the selection of specific RNA sequences with prebiotic functions[70].

Despite the aforementioned research conducted on the comparison among potential physico-chemical contexts for the origin of life, we are still far from unveiling the precise mechanisms that led to the emergence and replication of the first RNA polymers, around 4 Gyr ago[3,71,72]. To shed light on this subject, we propose an integrative numerical and theoretical approach that considers both non-enzymatic RNA polymerization and template-dependent replication in an environment where nts and ssRNA polymers interact through a clay-water interface. We introduce EarlyWorld, a computational framework supported by a theoretical model, which allows us to test how a wide variety of environmental conditions could affect the polymerization and replication of ssRNA oligomers. This approach differs from most existing models of non-enzymatic RNA replication in fluctuating conditions, which take the existence of two phases for granted: a "cold phase" where elongation reactions take place, and another "warm phase" where all RNA duplexes denature. EarlyWorld has been developed as a toy-model representation of the chemical processes that could have occurred in

a clay-water microenvironment in early Earth, and it is not pretended to be chemically detailed at this stage, but to capture the essential features of the processes modeled. By exploring and refining our understanding of the specific fluctuating environments that could effectively promote the increase in biochemical complexity, we aim to certify whether an heterogeneous mineral-water interface in early Earth was suitable for the emergence of RNA populations, as a prerequisite for the origin of life.

Methods: Description and rules of the model

We have developed a computational environment called EarlyWorld to simulate the non-enzymatic RNA polymerization and template replication in primordial heterogeneous media, such as on the interface between an aqueous solution and a clay surface, in the absence of RNA polymerase ribozymes. Clays are known to present a complex structure of internal channels and interlayers, where water can get in and promote an intricate set of clay-water interfaces of around 100 m² per g of clay[73]. The RNA polymerization and replication processes computationally simulated here take place in two compartments of the clay showing different geochemical properties. Figure 1 schematizes the dynamics described by EarlyWorld: Fig. 1A, B sketch the natural compartments of the model, and Fig. 1C, D show example realizations of the evolution with time of the system in each compartment. A rigorous explanation of the algorithm is provided in Supplementary Note S1, and information on the public availability of the code is found in section "Code Availability".

Compartment I

Random polymerization of nts takes place in compartment I (comp.I from now on) due to the catalytic action of clays. It represents an internal channel or interlayer that is wide enough to allow ssRNA, but not double-stranded RNA (dsRNA) oligomers to pass through. This assumption is supported by experimental data showing that the A-helix of dsRNA has a diameter of ~2.4 nm[74], while the interlayer space of montmorillonite and other clays typically varies between 0.9 and 1.0 nm (in the dried state) to 1.3 nm (hydrated with one water layer) and 1.9 nm (fully hydrated with 2–3 water layers), which also depends on their ionic composition, pH and presence of different cations in the medium[75,76]. In consequence, complementary base pairing (and, thus, template-dependent replication) is not possible in the interlayers due to steric hindrance (Fig. 1A). We have focused on interlayers of this kind as the preferred location for ssRNA polymerization, following the assumption that the potential presence of catalytic sites in the montmorillonite interlayers enhances the rate and regioselectivity of polymerization[22,77,78], though alternative scenarios are also considered (see "Discussion").

The compartment has two different levels: the clay-water interface (shortened to "clay"), and the aqueous phase (or "pool"). For simplicity, the clay level is described by a 1D-vector of length L where each element/position can be either empty or contain a single nt, while the aqueous phase is a *disordered box* with space for an unlimited number of nts and ssRNA molecules. The clay and the pool will be enriched with ssRNA oligomers throughout each simulation, but every numerical realization in comp.I starts with the clay level totally empty and a number N of nts equally distributed between those composing the genetic alphabet in use (EarlyWorld accepts any genetic alphabet and base pairs of all kind, see section Results "RNA replication across different genetic alphabets"). At each time step, one of the molecules in the pool is chosen randomly and gets adsorbed to the clay in a randomly chosen place (involving one or more contiguous positions), as far as it is empty. Otherwise the molecule returns to the pool. In the clay, it will form a covalent (3′-5′ phosphodiester) bond with other nts potentially placed at any (or both) of its adjacent positions. The resulting polymer remains attached to the clay, and we assume that its covalent sugar-phosphate backbone is unbreakable (except when we explicitly study the influence of hydrolysis in the system, see section Results "Compartment II: Template-dependent RNA replication on a clay surface"). After that, every oligomer adsorbed to the clay can be released to the pool with a desorption
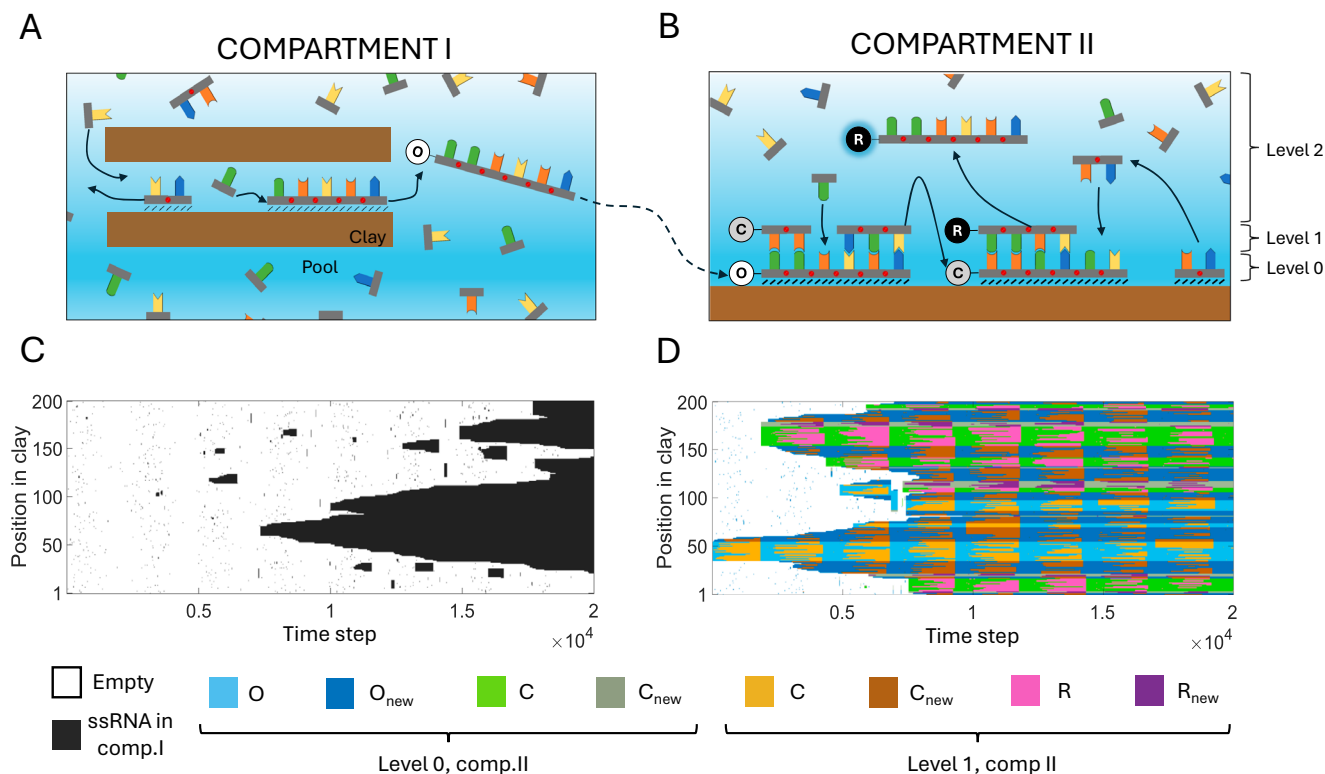
Fig. 1 | Description of the non-enzymatic RNA polymerization and replication in an early Earth's clay-water interface modeled by EarlyWorld. Top (**A**, **B**) sketch each natural compartment of the model and bottom (**C**, **D**) show example realizations of the evolution with time t of all the positions available in an oscillating clay environment. The bottom legend corresponds to the colors shown in (**C**, **D**). If not said otherwise, throughout this work we will use clays of length $L = 200$ positions, $N = 1600$ nts equally distributed between A, G, C, U types and initially placed in the aqueous phase, $C \equiv G$ and $A = U$ nucleobase complementarities, and simulation times $t_{\max} = 2 \times 10^4$ time steps. Compartment I: Random polymerization occurs due to the catalytic action of clays in an interlayer of a clay grain, where complementary base pairing is not permitted due to steric hindrance. Only clay-nt interactions (of strength α) are permitted and ssRNA can be formed at the clay surface. Compartment II: The template replication of the original strand O occurs involving complementary base pairing; strand C (complementary to O) is formed, denatured from the template, released to the pool and adsorbed to the clay. In a next step, the replication of strand C yields strand R, whose sequence is equal to the original (R = O). Once the dsRNA molecule C-R is formed, its denaturation will release the ssRNA oligomer R, completing the RNA replication process. Other polymers O_{new} spuriously formed in comp.II can also be used as template (C_{new}) and replicated (R_{new}). In comp.II both clay-nt (of strength α) and nt-nt (of strength β) interactions are allowed, and ssRNA molecules can remain adsorbed to the clay (level 0), hybridize to their complementary strand to form dsRNA (level 1), or move to the aqueous phase (level 2). In this example, $\alpha = 0.8 + 0.5 \sin(2\pi t / 2500)$ and $\beta = 6 + 5.9 \sin(2\pi t / 2500)$.

probability of

$$P_{\alpha} = e^{-l\alpha}, \qquad (1)$$

where the interaction parameter $\alpha > 0$ describes the strength of clay-nt interactions (due to Van der Waals forces and ionic bonds) in the environment that surrounds them, and l is the length (i.e., number of nts) of the polymer whose desorption is evaluated. Finally, note that only the oligomers detached from the clay into the pool are recorded as effectively produced in the process, because exclusively free and mobile ssRNA molecules can migrate to compartment II and continue the process.

Compartment II

This compartment represents a wide internal channel, an inter-particle site or an external clay surface, and, in consequence, is able to accommodate both ssRNA and dsRNA molecules and template-dependent RNA replication based on complementary base pairing. Simulations start with the same number of nts of each type available in the pool and with an initial ssRNA molecule (original strand O from now on), consisting of 20 randomly chosen nts adsorbed to the clay in a random place. This ssRNA O represents a strand that is assumed to be formerly polymerized in comp.I, was released from it, eventually entered compartment II (comp.II from now on) and attached to the clay surface. Three levels are relevant for the evolution of the system: the clay or "level 0", where nts and polymers can be

adsorbed (as in comp.I); the complementary level or "level 1", where one nt can only be placed if it is complementary to other nt included in a ssRNA oligomer which is placed in level 0 and thus contribute to the formation of dsRNA via hydrogen bonding; and the aqueous phase, pool or "level 2", where free nts, ssRNA and dsRNA stay in solution. In each time step, a randomly chosen molecule in the pool interacts with a randomly chosen place in the surface of the clay. ssRNA and dsRNA of any length can be adsorbed to the clay in level 0, and ssRNA will be hybridized to other ssRNA strand adsorbed to the clay acting as a template in level 1 if they are fully complementary. After that, every ssRNA or dsRNA of length l in level 0 will have a desorption probability (i.e., probability of being released to the pool) of $P_{\alpha} = e^{-l\alpha}$, where α is the clay-nt interaction parameter already mentioned in comp.I. Also, every dsRNA (both attached to the clay or free in the pool) can be denatured with a probability of
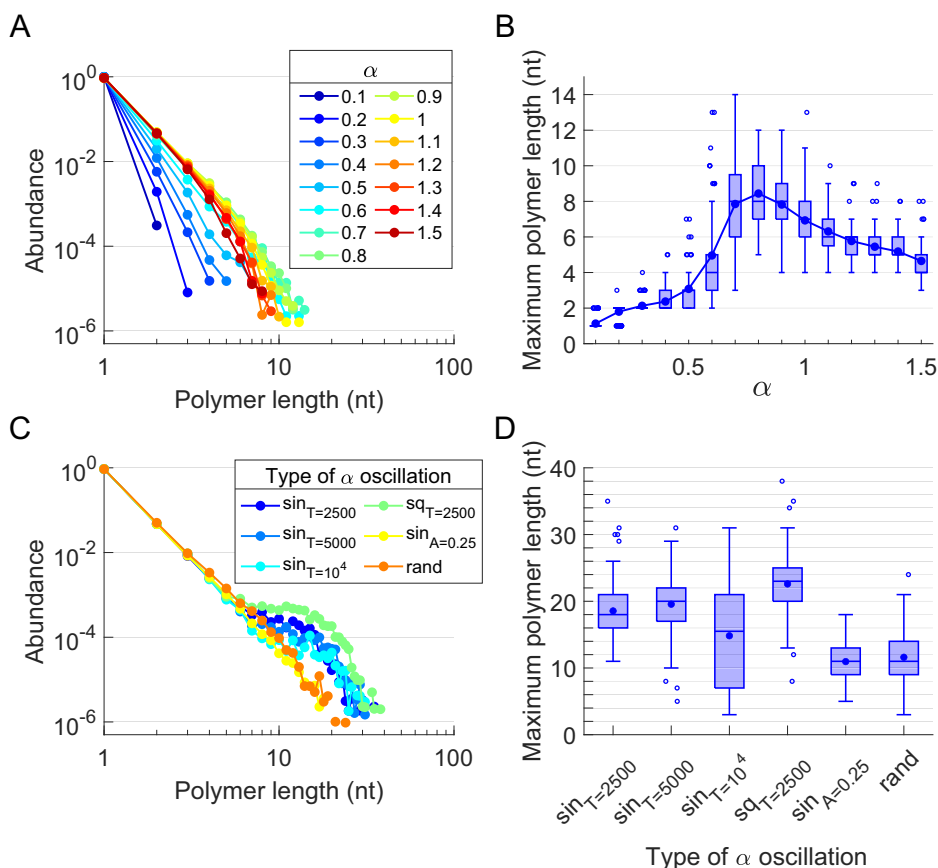
$$P_{\beta} = e^{-l\beta}, \qquad (2)$$

where l stands for the dsRNA length (i.e., number of nts), and the nt-nt interaction parameter $\beta > 0$ describes the strength of the hydrogen bond interactions established between their nucleobases in the environment that surrounds them.

Throughout each simulation, in comp.II the dynamics described might eventually release complementary C strands of the original O sequence to the pool (level 2), as well as complementary C_{new} strands of other newly-

Fig. 2 | Effect of constant and fluctuating environmental conditions on the polymerization of RNA catalyzed by a clay surface (comp.I).
A Distribution of abundances of ssRNA k-mers (monomers, dimers, etc) and **B** maximum polymer length distributions over constant environments (clay-nt parameter α is constant with time). **C** Distribution of abundances of k-mers and **D** maximum polymer length distributions over six fluctuating environments (α varies with time): "sin" stands for sinusoidal oscillation, "sq" for square wave oscillation and "rand" for random fluctuation. If not said otherwise in the axes, $\alpha_0 = 0.8$, $A = 0.5$, $T = 2500$. Every simulation was repeated 100 times.

generated O_{new} sequences randomly polymerized in the clay (level 0). Interestingly, when a polymer C falls in an empty place of the clay and is adsorbed, it might become the template for a further polymerization based on nucleobase complementarity, and the denaturation of the dsRNA formed will release to the pool a ssRNA that is "complementary to the complementary C": the strand replicate or R, equal (partially, or in its full length) to the original polymer O.

In summary, our model describes the influence of the environment in the polymerization and replication of short ssRNA polymers through two parameters, α and β, which represent the global strength of clay-nt and nt-nt interactions, respectively. Changing these parameters mimics varying external physico-chemical magnitudes such as temperature, pH, ionic strength and presence of divalent cations, to cite a few. Large values of α and β represent mild environments that facilitate said interactions (and the concomitant formation of progressively longer ssRNA and dsRNA molecules), whereas low values represent harsh environments and have the opposite effect.

Results
Compartment I: RNA polymerization on a clay surface
The polymerization process in comp.I depends exclusively on the clay-nt parameter α. Figure 2 plots the length distribution of the ssRNA molecules (polymers from now on) that can be formed with the rules set of EarlyWorld for different environmental conditions, both constant and fluctuating. The length distribution of the polymers produced attached to the clay and desorbed back to the pool at any time step for different constant values of α reveals a steep decay in Fig. 2A, in full agreement with the assumed relationship between the abundance of polymerized ssRNA molecules and their length[40,41,79]. In this work we are specially interested in obtaining the largest polymer that can be produced in a given heterogeneous environment, as it could eventually be more structurally complex and functionally relevant than the shorter ones. Thus, Fig. 2B shows the distribution of the maximum polymer length obtained along every realization for each parameter α (the

detailed explanation of how maximum polymer lengths are obtained is provided in Supplementary Note S2). For $\alpha \to 0$ the environment is so harsh that every nt that reaches the clay is immediately desorbed from it and released back to the pool, while for $\alpha \to \infty$ the mild environment allows the adsorbed nts to polymerize fast and become strongly stuck to the clay surface. Both limits result in the presence of very short ssRNA oligomers in the pool. For intermediate α values, however, oligomers can be substantially polymerized but still be desorbed, reaching maximum polymer lengths of $\sim 5 - 12$ for $\alpha \sim 0.8$.

Fluctuating environments can be easily introduced in EarlyWorld by changing the value of the parameters α and β over time. Figure 2C, D show, respectively, the length distribution of ssRNA molecules polymerized and desorbed from the clay, and the distribution of maximum polymer lengths obtained in every realization, for several α-oscillating environments. Considering environmental oscillations of period T and amplitude A, the sinusoidal oscillatory environments studied follow $\alpha(t) = \alpha_0 + A \sin(2\pi t/T)$, square wave oscillations follow $\alpha(t) = \alpha_0 + A \, \mathrm{sgn}(\sin(2\pi t/T))$, and randomly fluctuating environments follow the continuous uniform distribution $\alpha(t) = U\{\alpha_0 - A, \alpha_0 + A\}$. The length distribution of polymers plotted in Fig. 2C shows a power-law decay for the random environment that behaves as a lower limit for the rest of fluctuating environments, which, depending on their characteristics, show a larger or smaller hump for long (>10 nt) polymer lengths. In Fig. 2D, it can be seen that, for all oscillating patterns, maximum polymer lengths notably increase with respect to those achieved in constant environments: means move from 1 to 9 nts for constant environments (with absolute maxima around 15 nts) to 10–24 nts for fluctuating ones (with absolute maxima around 40 nts).

Compartment II: Template-dependent RNA replication on a clay surface
The simplest secondary structures that ssRNA molecules can fold into are called stem-loops, consisting of a dsRNA stem closed by a terminal loop. In

turn, hairpin structures are slightly more complex and contain a bulge within the stem, which can provide an additional biochemical functionality to the molecule[80]. Thus, it is commonly assumed that ribozymes should contain at least 15–35 nts to be able to fold into complex enough structures[26]. As shown in the previous section, this polymer length interval can be typically obtained in EarlyWorld when the environment (parameter α) fluctuates. Accordingly, simulations in comp.II start with a 20 nt-long, random ssRNA polymer O adsorbed to the clay. Along each simulation, we record the complementary C and replicate R strands derived from the original O sequence, which are produced and released to the pool.

Throughout this work, two terms will be used to quantify the goodness of the overall template-dependent RNA replication process. The first one is the *efficiency* of the replication associated with each combination of interaction parameters α and β. It will be quantified through the maximum polymer length distribution obtained for these parameters. Accordingly, an environment will be said to permit efficient replication if it yields maximum polymer length distributions of large mean with several realizations reaching length 20. The second relevant term is the *accuracy* required to accept a certain copy as sufficiently pure. Remarkably, Cs and Rs can contain, in addition to the exact copy of the corresponding template, other random or spurious sequences attached to their two ends during the replication process, thus forming a longer molecule than the polymer O from which they derive[35]. For that reason, and in order to avoid the loss of identity of the RNA molecules that are being replicated, throughout this work we consider those C and R polymers that have an *accuracy* of > 75%, meaning that their sequence is formed by >75% of products of the original polymer O (in other words, these copies can have as much as 25% of their length 'contaminated' by spurious sequences at their ends, see Supplementary Note S3 and Fig. S3 for detailed information on sequence accuracy and simulations with 100% accuracy, which yield equivalent results).

Both clay-nt parameter α and nt-nt parameter β play a role in the formation of C and R polymers in comp.II. Figure 3 shows the effect of
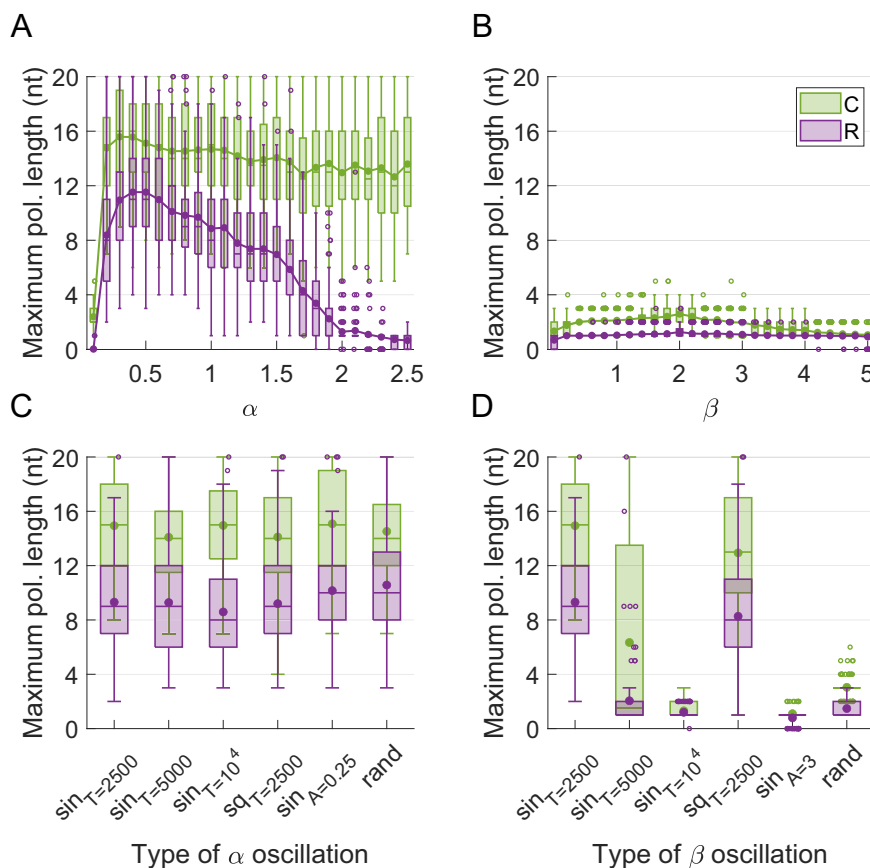
fluctuating environmental conditions (i.e., α and β) on the stepwise replication of the original 20 nt-long polymer O. Considering oscillations of period T_γ and amplitude A_γ, the sinusoidal oscillatory environments follow $\gamma(t) = \gamma_0 + A_\gamma \sin(2\pi t/T_\gamma)$, square wave oscillations follow $\gamma(t) = \gamma_0 + A_\gamma \text{sgn}(\sin(2\pi t/T_\gamma))$, and randomly fluctuating environments follow the continuous uniform distribution $\gamma(t) = U\{\gamma_0 - A_\gamma, \gamma_0 + A_\gamma\}$, where $\gamma = \{\alpha, \beta\}$. When one parameter is constant and the other oscillates (Fig. 3A, B), β becomes the most limiting factor. If β remains constant over time, only insignificant fractions of the original sequence O are replicated as R polymers. In contrast, a wide range of constant α values is compatible with the production of long C and R ssRNA molecules and allows the replication of the entire initial molecule O (Fig. 3A).

Figure 3C, D show how simultaneous (but not necessarily similar) fluctuations in both α and β parameters affect the formation and length of C and R polymers. Results in Fig. 3C indicate that, for the different cases studied, maintaining a standard sinusoidal β oscillation while varying the period and amplitude of sinusoidal α oscillations (as well as changing it to a more abrupt square wave) does not significantly affect replication efficiency. The mean length of the obtained R polymers remains close to a 50% of the original O sequence, and the maximum value after 100 realizations of all oscillation patterns recovered the whole original sequence length (20 nts). In contrast, maintaining an α oscillation compatible with high polymerization activity in comp.I and varying the period or amplitude of β oscillations drastically affects the system performance (Fig. 3D), presenting from a strong enhancement to a collapse of the replicative dynamics. In summary, EarlyWorld shows that efficient template replication of RNA is only plausible under oscillating environments, being their proficiency much more dependent on the properties of the nt-nt interactions that rule the template replication than on the clay-nt interplay that controls the initial RNA polymerization.

In Fig. 4, a more complete scan of the space of parameters for the template-dependent RNA replication that takes place in comp.II is

Fig. 3 | Effect of environmental fluctuations on the template-dependent replication of a ssRNA molecule adsorbed to a clay surface (comp.II).
A Maximum complementary (C, green in all panels) and replicate (R, purple in all panels) polymer length distributions for an oscillating β and a constant α parameter. **B** Maximum polymer length distributions for an oscillating α and a constant β parameter. **C** Maximum polymer length distributions for different α fluctuating patterns, with a fixed β oscillation. **D** Maximum polymer length distributions for different β fluctuating patterns, with a fixed α oscillation. Fluctuating environments: "sin" stands for sinusoidal oscillation, "sq" for square wave, "rand" for random fluctuation. If not said otherwise, $\alpha_0 = 0.8$, $A_\alpha = 0.5$, $T_\alpha = 2500$ and $\beta_0 = 6$, $A_\beta = 5.9$, $T_\beta = 2500$. Every simulation started with a random 20 nt-long ssRNA molecule, and was repeated 100 times.
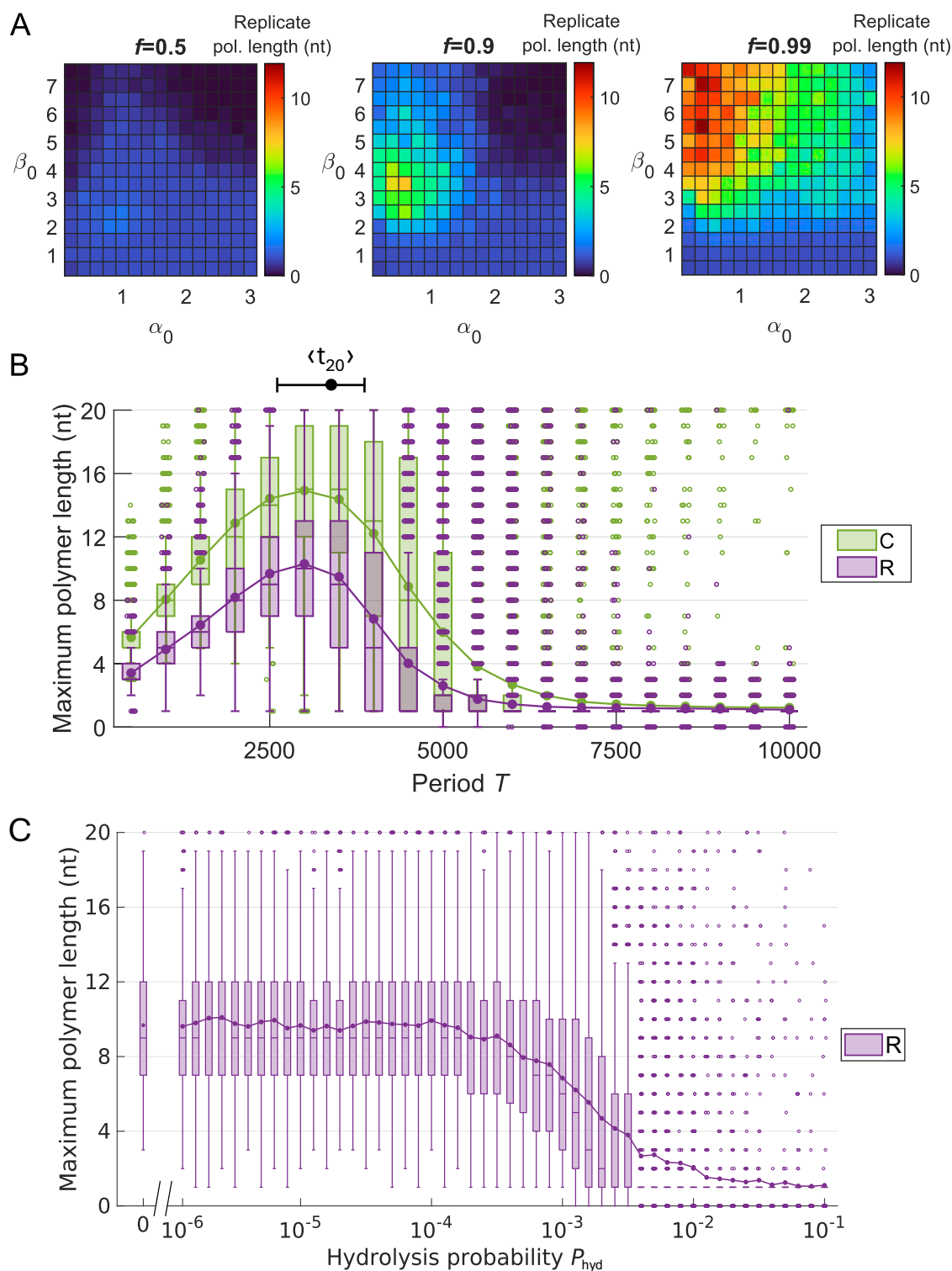
Fig. 4 | Effect of the model parameters on the template-dependent replication of a ssRNA molecule adsorbed to a clay surface (comp.II). A Dependence of the mean of maximum R polymer lengths with average interaction parameters α_0 and β_0 at three different amplitude fractions f (being $A_\alpha = f\alpha_0$ and $A_\beta = f\beta_0$). $T_\alpha = T_\beta = 2500$. **B** Dependence of the maximum length distribution of C polymers (complementary, green) and R polymers (replicates, purple) on the oscillation period $T = T_\alpha = T_\beta$. $\alpha_0 = 0.8$, $A_\alpha = 0.5$ and $\beta_0 = 6$, $A_\beta = 5.9$. The mean time to polymerize a 20-nt long ssRNA by template-dependent polymerization $\langle t_{20} \rangle$ is displayed with 25th-75th percentiles (see Supplementary Fig. S6). **C** Dependence of the maximum length distribution of R polymers on the phosphodiester bond hydrolysis probability per time step P_{hyd}. $\alpha_0 = 0.8$, $A_\alpha = 0.5$ and $\beta_0 = 6$, $A_\beta = 5.9$, $T = T_\alpha = T_\beta = 2500$. Every simulation started with a random 20 nt-long ssRNA molecule, and was repeated 20 times in (**A**), 10^4 times in (**B**) and 500 times in (**C**).

presented. Without loss of generality, it is now assumed that (i) in natural environments the whole system formed of comp. I and II of a clay grain in contact with water will be exposed to the same oscillatory environmental conditions, thus affecting interaction parameters α and β with the same period $T_\alpha = T_\beta = T$, and (ii) the amplitude of such oscillations will be proportional to the average values α_0 and β_0 by a factor $f \in [0, 1]$ (i.e., $A_\alpha = f\alpha_0$ and $A_\beta = f\beta_0$). Figure 4A shows the dependence of the replication efficiency on the average interaction parameters α_0 and β_0 and the amplitude fraction f (see Supplementary Fig. S5 for the corresponding plots on the efficiency of the formation of the complementary strand C). As already observed in Fig. 3A, relatively low α_0 is optimal because, in order to form Rs, there must be free space in the clay for Cs to adsorb and serve as Rs' templates, a condition better satisfied with an environment that hampers the attachment to the clay. On the other hand, large β_0 and very large amplitudes of the oscillations become necessary, as they represent environments that alternate periods of strong polymerization enhancement with periods of intense dsRNA denaturation.

Figure 4 B shows the dependence of the replication efficiency on the α and β oscillation period. Short periods prevent from a sufficient elongation of the polymerized or replicated strands, the lower limit $T \to 0$ being comparable to a random distribution of β values (as shown in Fig. 3D). Intermediate period values favor the replicative dynamics, and copies of all lengths are densely represented. Remarkably, the value of the oscillation period that optimizes the replication dynamics (3000 ± 500 time steps) in Fig. 4B coincides with the number of simulation time steps that in average is required to form a complete 20-nt long complementary copy attached to the original O sequence ($\langle t_{20} \rangle \approx 3300$, see Supplementary Fig. S6). This equivalence will be of the maximum importance to match simulation times with real time scales and allow EarlyWorld to generate testable predictions about prebiotic environments (see Discussion). When oscillation periods become very large, a drastic transition occurs: the environment varies so slowly that the system resembles temporarily a β-constant environment (as shown in Fig. 3B), favoring an excessive polymerization of dsRNA molecules that do not denature in such conditions, leading to the concomitant collapse of the replication process. Interestingly, as it can be observed in Supplementary Fig. S7, the maximum R polymer lengths observed for short periods follow log-normal distributions, become decreasing long-tailed distributions strongly peaked in 1 nt just after the transition, and finally collapse to the absence of any relevant replication for larger periods.

Finally, Fig. 4C shows the replication efficiency of the system when different values of the phosphodiester bond hydrolysis probability per time step, P_{hyd}, are introduced in the model. It is known that the phosphodiester bond hydrolysis rate must be much lower than the polymerization rate to allow the formation of long polymers[36], and this ratio can vary widely depending on environmental conditions[35,81,82]. In particular, montmorillonite and other clays have been shown to reduce the rate of chemical hydrolysis of RNA[22,83–85]. Accordingly, we investigated a wide range of values for the ratio between the probability of hydrolysis P_{hyd} and the probability of phosphodiester bond formation P_{form} (which is equivalent to studying the dependence on P_{hyd} because in our simulations $P_{form} = 1$). When hydrolysis probability is below a critical threshold, the replication process proceeds as if hydrolysis was absent; however, around $P_{hyd} \sim 10^{-4}$, the system begins to destabilize and replication becomes unsustainable beyond $P_{hyd} \sim 10^{-3}$. A modest enhancement in replication efficiency is observed when the hydrolysis probability varies over time, consistent with the fluctuating conditions likely present in prebiotic environments (see Supplementary Fig. S8). Notably, the system's response to changes in both the oscillation period (Fig. 4B) and the hydrolysis probability (Fig. 4C) closely resembles the first-order transition previously described by Tkachenko and Maslov[36], which distinguished a regime sustained by long, self-replicating polymer populations from one dominated by free monomers depending on the overall monomer concentration. Furthermore, optimal periods in Fig. 4B, together with the value of α_0 used ($\alpha_0 = 0.8$), also optimized polymerization in comp.I (Fig. 2). This strongly reinforces the plausibility of observing in real systems the mechanism here presented for the polymerization and

Table 1 | Parameters of the computational environment EarlyWorld and the theoretical model for different genetic alphabets, from 2 to 6 letters

Alphabet	H-bonds/nt	$s_\mathcal{A}$	Connection time	$t_\mathcal{A}$
A2$_{au}$	2	4/5	2	1/2
A2$_{gc}$	3	6/5	2	1/2
A4	5/2	1	4	1
A4*	7/3	14/15	8/3	2/3
A6$_2$	7/3	14/15	6	3/2
A6$_3$	8/3	16/15	6	3/2

$s_\mathcal{A}$ and $t_\mathcal{A}$ are normalizations (such that $s_\mathcal{A} = t_\mathcal{A} = 1$ for A4) of the average number of hydrogen bonds per nt and the connection time, respectively, used in the mathematical model.

posterior replication of RNA molecules. Also, Supplementary Fig. S9 shows that the results presented in Fig. 4 and Supplementary Fig. S6 remain robust under variations in either the available clay surface —by changing the clay length L— or in the nt concentration in the aqueous phase —by modifying the number of molecules that fall into the clay per time step—.

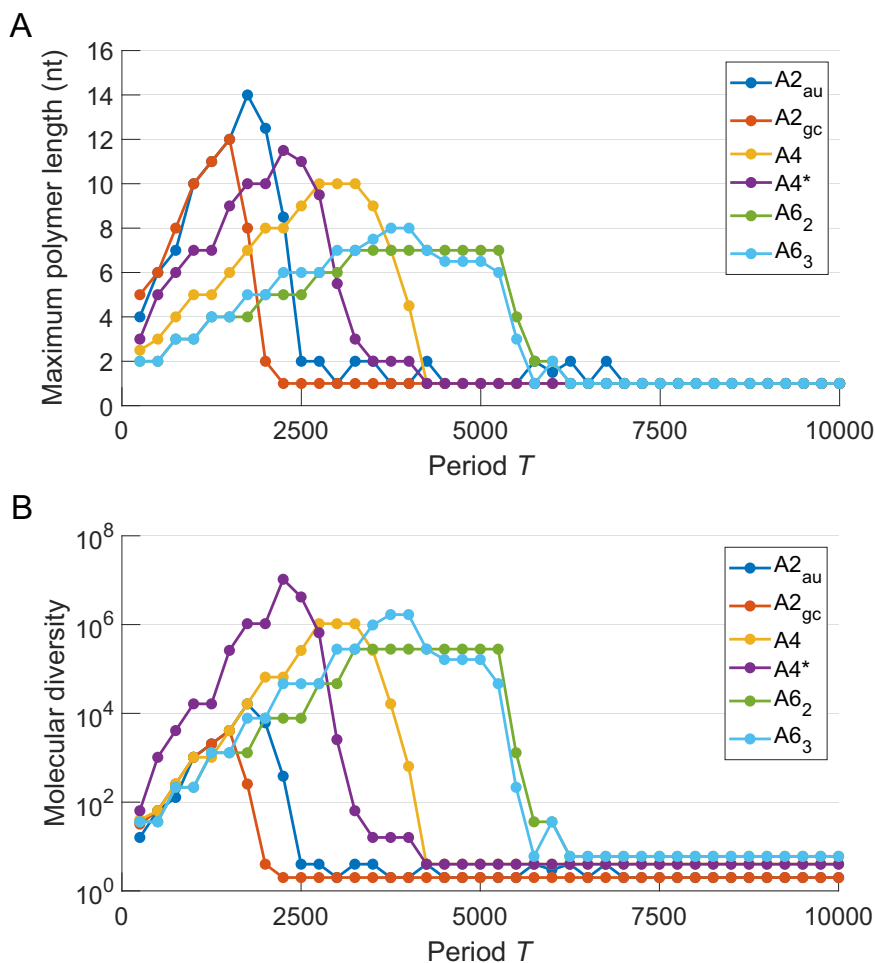## RNA replication across different genetic alphabets

The computational environment EarlyWorld can be used to test multiple physico-chemical conditions compatible with those present in the early Earth where RNA world was established. One of the most controversial questions about the origin and evolution of our genetic material deals with the number, relative concentration and pairing rules of different monomers that formed RNA and, eventually, pre-RNA polymers (including glycerol-derived nucleic acid or GNA, threose nucleic acid or TNA, pyranosyl-RNA or p-RNA, locked nucleic acid or LNA, and peptide nucleic acid or PNA)[1,86–88]. Additionally, N-(2-aminoethyl)-glycine or AEG (building block of PNA backbone) could have been the precursor of current ribonucleosides, thus promoting the synthesis of proto-RNA[89].

Current RNA is composed of four different nts, whose nucleobases interact specifically through hydrogen bonding: adenine and uracil (paired by means of two H bonds, A = U), as well as guanine and cytosine (linked by three H bonds, G ≡ C). In addition, non-canonical base pairs, such as Wobble base pairs (the most common type of these being G = U) are not infrequent in RNA, though they are less thermodynamically stable than canonical pairs. In order to shed light on why prebiotic chemistry chose one genetic alphabet among all possible ones, we used our computational model to test the replication efficiency of six different options: two-letter alphabet [A=U] (A2$_{au}$), two-letter [G ≡ C] (A2$_{gc}$), canonical four-letter [A = U, G ≡ C] (A4), four-letter with additional Wobble base pair [A = U, G ≡ C, G = U] (A4*, representing current RNA alphabet and pairing rules), and six-letter with a new base pair involving two unknown nucleotides (termed X and Y) linked by either two [A = U, G ≡ C, X = Y] (A6$_2$) or three hydrogen bonds [A = U, G ≡ C, X ≡ Y] (A6$_3$). The main properties of these alphabets when introduced in EarlyWorld are shown in Table 1.

Two factors play a role when comparing different genetic alphabets in EarlyWorld. The first is the connection time, that is, the average number of attempts (or time steps) that a nt takes to find a complementary nt to attach to, which is equal to the alphabet size $\mathcal{A}$ (with the exception of A4*, which allows 3 kinds of base pairs with only 4 letters). The second one is the number of hydrogen bonds involved in each base pair: we now assume that the probability of denaturation of a dsRNA is proportional to the total number of hydrogen bonds formed between a template and its complementary strand, and not just to the length of the copied molecule, as considered before. Consequently, the denaturation probability of hybridized dsRNA (Eq. (2)) becomes $P = e^{-n\beta/n_{A4}}$, where n represents the number of hydrogen bonds in the molecule and $n_{A4} = 2.5$ is the average number of bonds per nt in the A4 alphabet.

In order to compare the replication efficiency of RNA composed of the 6 different alphabets under study, Fig. 5A shows the dependence of the maximum length of the replicated strand R obtained by EarlyWorld with the

Fig. 5 | Efficiency of template-dependent RNA replication is evaluated across six different alphabets. A Median of the distribution of maximum R polymer lengths, and **B** median of the molecular diversity computed as $\mathcal{A}^l$, where $\mathcal{A}$ stands for the alphabet size and l is the median of the distribution of maximum R polymer lengths, for different values of the period $T = T_\alpha = T_\beta$. $\alpha_0 = 0.8$, $A_\alpha = 0.5$ and $\beta_0 = 6$, $A_\beta = 5.9$. Every simulation started with a random 20-nt RNA sequence, and was repeated 100 times.

oscillation period in an $\alpha -$ and $\beta -$ sinusoidal environment (with system parameters equivalent to those in Fig. 4B). The absolute maximum in replication length decreases with the alphabet size because of the slowing down of replicative dynamics (i.e., growth of connection time) for large alphabets. The effect related to the different number of hydrogen bonds between complementary nts makes $G \equiv C$-rich and $X \equiv Y$-rich dsRNA molecules more robust under denaturations and faster in polymerizing, thus accelerating the system dynamics and in consequence shifting the curves towards smaller periods (as it can be observed when comparing A2$_{gc}$ vs. A2$_{au}$ and A6$_3$ vs. A6$_2$). Remarkably, A4* reaches significantly longer maximum polymer lengths than A4 because of its faster polymerization rate due to the formation of the non-canonical base pair G = U, performing in fact similar to A2$_{gc}$.

Additionally, if the total number of different ssRNA sequences that can be replicated with a length l of $\mathcal{A}$ nts, $\mathcal{A}^l$, is used as a proxy for the molecular diversity—or, if preferred, the information storage capacity—associated with a population of RNA polymers of a certain genetic alphabet, Fig. 5B shows that A4* represents an optimum combination in terms of diversity and replication speed, its absolute maximum outperforming the rest of alphabets in at least one order of magnitude. In turn, the absolute maximum molecular diversity associated with the canonical four-letter alphabet A4 is two orders of magnitude larger than that of two-letter alphabets, and similar to the much more energetically consuming six-letters ones.

Theoretical approach to RNA replication in EarlyWorld

We propose here a simplified theoretical description of the RNA polymerization and replication process where isolated nts can adsorb to, polymerize on and be released from the clay (level 0 in EarlyWorld, comp.I) or from a ssRNA already adsorbed to the clay (level 1 in EarlyWorld, comp.II). For simplicity, the model entails a unique interaction parameter γ that plays

the role of both α and β, depending on the case. It is assumed that the probability P_i of forming and releasing a polymer of length i to the aqueous phase or pool (level 2 in EarlyWorld) is equal to the probability of a single nt of remaining adsorbed to the surface (level 0 in comp.I and level 1 in comp.II) for one time step, receiving a new nt that lands adjacent to it (which we assume with probability 1 for simplicity), remaining the new 2-mer ssRNA adsorbed for one time step, and so on until the polymer reaches length i after i time steps and is released to the pool. In summary, this iterative calculation yields the probability P_i of generating the sequences of length i in a constant environment, as

$$P_1 = e^{-\gamma}; \quad P_i = e^{-i\gamma} \prod_{k=1}^{i-1}(1 - e^{-k\gamma}) \quad \text{for } 2 \leq i \leq i_{max}, \quad (3)$$

where i_{max} should be large enough to make $P_{i_{max}}$ negligible: $i_{max} = 50$ in all calculations. The mean length of the polymers produced for a given interaction parameter γ is
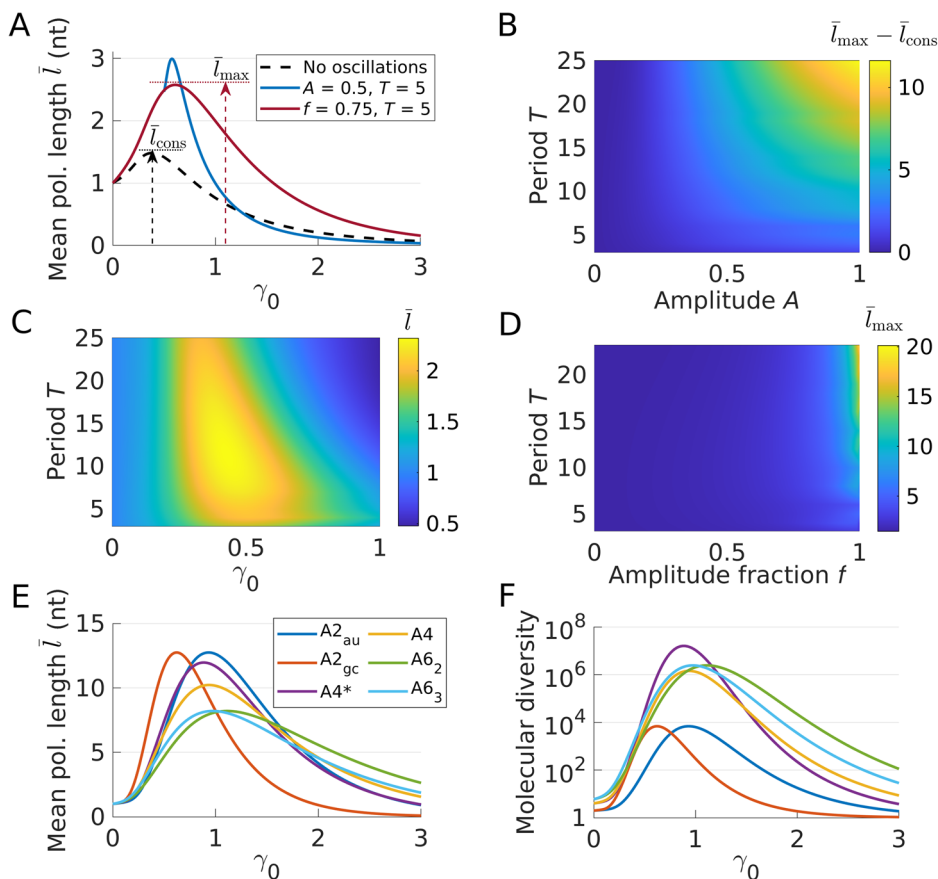
$$\bar{l} = \sum_{i=1}^{i_{max}} i P_i, \quad (4)$$

a complex but totally explicit mathematical expression. To model the clay-nt interaction in both comp.I and II in fluctuating environments, $\gamma = \alpha$ oscillates with time as

$$\gamma = \gamma_0 + A \sin(2\pi i/T), \quad (5)$$

Fig. 6 | Simplified mathematical approach to the polymerization and replication of a ssRNA molecule on clay surfaces. A Dependence of the mean polymer length $\bar{l}$ produced on the average interaction parameter γ_0 in three scenarios: a constant environment ($\gamma = \gamma_0$), an oscillating environment with amplitude $A = 0.5$ ($\gamma_0 \geq A$ to avoid meaningless negative γ) and period $T = 5$, and an oscillating environment of relative amplitude $f = 0.75$ and $T = 5$. The maximum mean length of the polymers produced in an oscillating environment $\bar{l}_{\max}$ and in the constant environment $\bar{l}_{\text{cons}} = 1.4823$ are remarked. **B** Difference between the maximum mean length produced in oscillating environments and in the constant environment for an exhaustive range of system parameters. Interestingly, all values of $\bar{l}_{\max} - \bar{l}_{\text{cons}}$ are > 0. **C** Dependence of the mean length of polymers produced in oscillating environments $\bar{l}$ with the average interaction parameter γ_0 and the period T, $f = 0.6$. **D** Dependence of the maximum mean length of the polymers $\bar{l}_{\max}$ produced in oscillating environments with the amplitude fraction f and the period T. **E** Mean length $\bar{l}$ of the polymers produced in oscillating environments for the six different genetic alphabets considered in EarlyWorld and a wide range of γ_0, $f = 0.95$, and $T = 25$. **F** Molecular diversity $\mathcal{A}^{\bar{l}}$ associated with the polymers produced for different genetic alphabets in (**E**).

where γ_0 is the average interaction parameter, A is the oscillation amplitude and T the oscillation period. T is an integer number between $T_{\min} = 3$ (to avoid $\sin(2\pi i/T) = 0$ for all i) and $T_{\max} = i_{\max}/2$ (to make sure that the system reproduces at least two full oscillations). Note, however, that the nt-nt interactions in comp.II are influenced by the genetic alphabet in use due to complementary base pairing. In consequence, the evolution of $\gamma = \beta$ in fluctuating environments becomes

$$\gamma = s_{\mathcal{A}}(\gamma_0 + A\sin(2\pi t_{\mathcal{A}} i/T)), \tag{6}$$

where $s_{\mathcal{A}}$ is proportional to the average number of hydrogen bonds between two nts for each alphabet, and $t_{\mathcal{A}}$ is proportional to the connection time that a nt in the pool would take in average to find its complementary one (see Table 1).

Figure 6A plots the dependence of the mean length $\bar{l}$ of the polymers produced by the system on the average interaction parameter γ_0 (i.e., α_0 in comp.I and II and β_0 in comp.II) for a constant environment (given by Eqs. (3)–(4) with $\gamma = \gamma_0$) and two different oscillating environments (given by Eqs. (3)–(5) for $\gamma_0 = \alpha_0$ and (3)–(6) for $\gamma_0 = \beta_0$). All mean polymer lengths show a maximum $\bar{l}_{\max}$ for intermediate values of γ_0, as it happened in the EarlyWorld simulations shown in Figs. 2B and 3A. Note that the maximum of the mean lengths of both oscillating environments are larger than that of the constant one $\bar{l}_{\text{cons}} = 1.4823$. In order to verify theoretically whether this is a general behavior, Fig. 6B shows the difference between the maximum of the mean lengths of the polymers produced in oscillating environments and constant environments (i.e., $\bar{l}_{\max} - \bar{l}_{\text{cons}}$) for an exhaustive range of all the system parameters. This subtraction results to be always positive, proving that oscillatory environments are always more favorable than constant ones regarding the abiotic polymerization and replication of RNA sequences.

Despite the simplicity of this mathematical model in comparison to the EarlyWorld computational environment, it captures several of its main

findings, reinforcing its generality. As an example, Fig. 6C shows the dependence of the mean length of the polymers produced by the system with the interaction parameter γ_0 and the oscillation period T, where it is clear that for any meaningful value of the environment (i.e., of the physico-chemical parameters that influence the clay-nt and nt-nt molecular interactions), the system shows the maximum replication efficiency for intermediate values of T, as it was shown in Fig. 4B for EarlyWorld. Figure 6D is focused on the influence of the amplitude fraction f, remarking that only strongly oscillatory environments give rise to significant replication dynamics, in agreement with Fig. 4A for the numerical model. Finally, Fig. 6E, F compare the proficiency of six genetic alphabets in the polymerization and replication of primordial RNA molecules. Here, the mathematical model recovers precisely the results obtained for EarlyWorld in Fig. 5, as: (i) faster hybridization dynamics of small alphabets (composed of 2 or 4 nts) make them generically outperform large alphabets; and (ii) regarding molecular diversity, the absolute maximum of alphabet A4 coincides in performance with any larger alphabet and drastically outperforms 2-letter ones, while A4* shows an unbeatable combination of alphabet length and speed of replication dynamics.

Discussion

The transition from astrochemistry to prebiotic chemistry and then to life on Earth involved critical steps toward increasing molecular complexity, with the origin of genetic information and its template-dependent replication being a key, unresolved question[47,90]. Motivated by this challenge, we developed EarlyWorld, a computational model that simulates primordial RNA polymerization and replication, driven solely by the interactions at the clay-water interfaces. The model features two interconnected compartments within the clay, enabling the separation of polymerization and replication processes, which is difficult to achieve in other heterogeneous, plausibly prebiotic environments such as lipid-water or air-water interfaces,

and ice-water eutectic phases[1,58]. The system's performance depends on two parameters that represent the strength of clay-nt (α) and nt-nt (β) interactions. Given that fluctuating, out-of-equilibrium conditions likely played a key role in life's origins[58], we explored a range of oscillatory patterns by varying α and β both periodically and aperiodically.

Despite random elongation of some RNA sequences[35,43], our results show that 20-mer RNA oligonucleotides can polymerize and replicate with a 100% accuracy under moderate phosphodiester bond hydrolysis rates (Fig. 4C) and specific oscillating environmental conditions (Supplementary Fig. S3), whose compatibility across both compartments of the clay system suggests that all steps of polymerization and replication could occur within a single physico-chemical environment. In particular, temperature, pH, salinity, or humidity should fluctuate with (i) large amplitudes (Fig. 4A), and (ii) periods matching the time necessary to polymerize a complementary ssRNA of the same length as the template (Fig. 4B, Supplementary Fig. S6), being the results compatible with a diversity of nucleotide concentrations in solution and available clay surfaces (Supplementary Fig. S9). Furthermore, there is still an open debate on whether RNA polymerization on different clay substrates, as described by the groups of Ferris and others, preferentially occurs within the phyllosilicate interlayers or on the more accessible external surfaces[19,20,22,56,65,91–93]. The EarlyWorld model produces similarly efficient polymerization and replication if the polymerization scenario corresponds to a sterically unrestricted surface (e.g., a wide internal channel, interparticle site or external clay surface, see Supplementary Fig. S10A and B), provided that it is sufficiently separated from the replication site. In contrast, when polymerization and replication occur at the same location on the clay surface, competition among RNA polymers for limited surface sites significantly reduces replication efficiency (Supplementary Fig. S10C).

The ability of our model to generate testable predictions about environments that could optimize primordial RNA replication hinges on result (ii) above, combined with recent geochemical insights into Hadean Earth conditions[59] and the estimates of template-dependent RNA polymerization times at clay-water interfaces. Our results yield that the optimal environmental oscillation period for the prebiotic template-dependent replication of a ssRNA of length l would be that which coincides with the time to obtain a polymer complementary to the original one. Therefore, if we obtained an experimental estimate of the template-dependent RNA polymerization time per nt in a clay-water environment t_{pol}, the ideal prebiotic environment for the polymerization and replication of a ssRNA of length l would be that with an oscillation period equal to

$$T_{opt} = l \times t_{pol}, \tag{7}$$

assuming a linear relationship between RNA polymerization time and sequence length. Though precise experimental data for such polymerization time per nt remain yet unavailable, relevant experiments offer insight. Extensive work by Ferris and collaborators[19,20,22,65] demonstrates that RNA polymerization time on montmorillonite clay varies significantly, from 0.6 to 11 h per nt, depending on nt types, concentrations and presence of different activating leaving groups linked to the phosphates[20]. Additionally, nonenzymatic template-dependent RNA polymerization in aqueous solution without mineral substrates yields a polymerization time of ~1 h/nt[94–96], which may extend to 1 day/nt under conditions that increase their prebiotic plausibility[97]. Furthermore, we can consider a hydrolysis time of ~20–50 h/nt[95], which helps us to fix an upper bound for t_{pol} because polymerization times larger than the hydrolysis time would not permit a neat sequence polymerization. Taking into account these data, we can expect an overall clay-catalyzed template-dependent RNA polymerization time in prebiotic conditions of t_{pol} in the range ~1 h/nt–1 day/nt. In summary, based on the available experimental data and Eq. (7), we propose that plausible oscillation periods would range from around one current day (i.e., 24h-long) to a few weeks to support efficient clay-catalyzed replication of ssRNA oligomers long enough to be functional, though short enough to be polymerized and replicated in prebiotic conditions (i.e., ~15–50 nt) at the dawn of the RNA world.

Recent and precise analyses of tidal interactions in the Earth-Moon system since its formation[59] suggest that the planet-satellite distance was 50–62% of its current value (Supplementary Fig. S11A) and that a day lasted about 10–12 h when life emerged on Earth, ~4.2–3.8 Gyr ago[3,71,72] (Supplementary Fig. S11B). The optimal oscillation periods for RNA replication, as discussed earlier, would probably exceed the timescales of daily temperature fluctuations from early Earth's day-night cycles or tidal wet-dry dynamics but would be shorter than seasonal variations in temperature and humidity. These oscillations, however, align well with spring tides: the maximal tides that occur twice each synodic month (the period between two Sun-Earth-Moon alignments), with a spring tide period of around 125–175 h when life emerged (Supplementary Fig. S11C). Unlike other potential fluctuating prebiotic scenarios, tidal pools benefit from the strictly periodical behavior of tidal forces, and have already been considered potentially relevant for prebiotic chemistry[60,61]. These pools would have facilitated interactions between a wide array of reactants carried by rivers, oceans, or the atmosphere[3], on a planet where isotopic evidence from detrital zircons indicates that continental crust may have formed as early as 4.35 Gyr ago[98,99]. Given that —to first order— tidal amplitudes decrease with the cube of the Earth-Moon and Earth-Sun distances[100], the Moon's contribution was 4 to 8 times greater than today, while the Sun's contribution has remained constant, as the Earth-Sun distance has not substantially changed since then.

Therefore, in the model we envisage, every 10–18 early Earth's days warm seawater would flood dry clay-bottomed ponds covered by salt precipitates, located far enough from the shoreline to be unaffected by daily tides. This seawater influx could have three key effects: (i) delivering fresh nts from the sea or nearby sources; (ii) increasing the temperature at clay-water interfaces; and (iii) reducing the local concentration of monovalent and divalent cations that were present in the salt precipitates. Together, (ii) and (iii) increased the medium's astringency (i.e., reduced α and β in the EarlyWorld model), weakening nt hydrogen bonding and promoting RNA strand separation. These findings are consistent with previous results[58], where temperatures above 60°C, Na^+ concentrations below 100 mM, and Mg^{2+} concentrations below 5 mM were shown to promote dsRNA denaturation. After several days of no further seawater input, the (perhaps, warm little) pond would dry out again (i.e., increasing α and β in the EarlyWorld model), allowing nts to polymerize on the clay surfaces as well as on the adsorbed ssRNA templates, thus contributing to the overall template-dependent replication. Thanks to the periodic behavior of spring tides, this cyclic process of RNA polymerization and replication could continue without interruption. Supplementary Fig. S11D presents our model's prediction for the RNA sequence lengths that would be optimally replicated in a spring tidal pool, based on the polymerization time (a yet very uncertain quantity) and the Earth's time. The range 4-11 h/nt for polymerization times matches the 15–40 nt lengths typically obtained in EarlyWorld. These results encourage experimental efforts to determine precise template-dependent RNA polymerization times in clay-containing prebiotic environments, as they would provide relevant insight into the expected length of primordial RNA oligomers.

In addition, our numerical and theoretical results reveal that the four-letter alphabet with pairing rules [A = U, G $\equiv$ C, G = U] (A4*, representing current RNA) exhibits superior replication performance than any other alphabet in terms of molecular diversity and sequence space navigability. The canonical four-letter alphabet [A = U, G $\equiv$ C] (A4) achieves molecular diversity two orders of magnitude greater than two-letter alphabets, and comparable to the more energetically costly six-letter systems. This strongly supports the primordial selection of four-letter alphabets (especially A4*) for genetic information storage in RNA. Interestingly, A4* introduces sequence mutations through base pair degeneracy (e.g., an A copied as U in the complementary polymer may become G in the replicated one) that expands dramatically the sequence exploration ability, while remaining constrained by the entropy induced by the degeneration bias (see Supplementary Note S4 for an analysis of the expansion of sequence space exploration enabled by

Article

non-canonical base pairs through the lens of information theory). Note that, while this mutation potential could enhance molecular diversity in early RNA populations, it also risks an error catastrophe (according to quasispecies theory[101]) in case non-canonical base pairs became too prevalent. Actually, G = U interactions are weaker than canonical ones, which may have helped to mitigate this risk.

The precise copies of the original ssRNA produced through the process analyzed in this work should undergo subsequent selection pressures, which recent findings suggest could occur even for 22-nt long oligonucleotides[70]. The study of such selection on the basis of the potential functionality of their molecular structures becomes a promising direction for future research. This will be addressed in the next version of EarlyWorld, thanks to the inclusion of an in silico RNA 2D-folding routine[102]. Additional enhancements, such as varying the relative initial nt abundances, differential clay adsorption of purine and pyrimidine nts, allowance for non-productive 2′-5′ and 5′-5′ phosphodiester bonds (besides the regioselective formation of 3′-5′ ones), and the introduction of replication errors (mismatches, insertions, deletions, and recombination), will help to bring the computational model closer to the rules of real (bio)chemistry.

Recent studies have modeled similar phenomenologies using chemical kinetics, a perspective that aligns with our effective approach. In our framework, the desorption and denaturation probabilities, P_α and P_β, serve as simplified proxies for the kinetic constants of RNA desorption from the clay and dsRNA denaturation, respectively. For instance, the kinetic constant for RNA strand denaturation k_{off} is commonly assumed to depend on the binding free energy per base pair, ΔG, as $k_{off} \propto \exp(-\frac{\Delta G}{k_B T} l)$, where k_B is Boltzmann's constant, T is the temperature, and l is the number of base pairs in the RNA duplex[43,44]. In our model, denaturation probability is $P_\beta = \exp(-\beta l)$, establishing an explicit connection with k_{off}, where the parameter β encapsulates the thermodynamic contribution to strand separation. A similar thermodynamic interpretation applies to α, which depends on the binding free energy of nucleotide-clay interactions governing desorption. These simplifications enable a tractable model that incorporates environmental variability in a thermodynamically consistent manner, while inevitably omitting chemical details which are beyond the scope of this study.

Finally, we have developed a theoretical study of a toy model of the system that, despite its simplicity, captures most of the key phenomenological features observed in the computational framework, suggesting the generality of our results beyond specific numerical details. The model highlights the importance of intermediate oscillatory periods and strong fluctuations in nt-nt hydrogen bonding to facilitate RNA replication, as well as the dominance of four-letter alphabets in a potential primordial soup analogous to that envisioned by A.I. Oparin a century ago. Notably, the model provides strong evidence that RNA polymerization and replication are enhanced in oscillating environments compared to constant conditions at aqueous-clay interfaces. The Moon's role in providing such oscillations in early Earth may have been crucial for the emergence of heritable information, and suggests that exoplanets with surface liquid water in interaction with clays (or, eventually, other rocks that provide catalytic surfaces) and orbited by large moons could be promising targets in the search for extraterrestrial life.

Data availability

The authors declare that the relevant data supporting the findings of this study are available within the paper and its supplementary information files. All data are available from the corresponding author upon reasonable request.

Code availability

An implementation of the computational framework EarlyWorld developed in Matlab(R2020b), with description, documentation and instructions of use is available at the following public open source repository: https://github.com/c-alejandrev/EarlyWorld.

References

1. Ruiz-Mirazo, K., Briones, C. & de la Escosura, A. Prebiotic systems chemistry: new perspectives for the origins of life. *Chem. Rev.* **114**, 285–366 (2014).
2. Sutherland, J. D. The origin of life—out of the blue. *Angew. Chem. Int. Ed.* **55**, 104–121 (2016).
3. Kitadai, N. & Maruyama, S. Origins of building blocks of life: a review. *Geosci. Front.* **9**, 1117–1153 (2018).
4. Preiner, M. et al. The future of origin of life research: bridging decades-old divisions. *Life* **10**, 20 (2020).
5. Oparin, A. I. Proiskhozhdenie zhizny. *Moscow: Izd. Moskovhii Rabochil* (1924). [First translation into English: Oparin, A. I. The Origin of Life (S. Morgulis, Trad.). New York: The Macmillan Company (1938)].
6. Haldane, J. B. S. The origin of life. *Ration. Annu.* **148**, 3–10 (1929).
7. Gilbert, W. Origin of life: the RNA world. *Nature* **319**, 618–618 (1986).
8. Robertson, M. P. & Joyce, G. F. The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, a003608 (2012).
9. Pearce, B. K., Pudritz, R. E., Semenov, D. A. & Henning, T. K. Origin of the RNA world: the fate of nucleobases in warm little ponds. *Proc. Natl Acad. Sci.* **114**, 11327–11332 (2017).
10. Cech, T. R. The RNA worlds in context. *Cold Spring Harb. Perspect. Biol.* **4**, a006742 (2012).
11. Bose, T. et al. Origin of life: protoribosome forms peptide bonds and links RNA and protein dominated worlds. *Nucleic Acids Res.* **50**, 1815–1828 (2022).
12. Szostak, J. W. An optimal degree of physical and chemical heterogeneity for the origin of life? *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2894–2901 (2011).
13. Joyce, G. F. & Szostak, J. W. Protocells and RNA self-replication. *Cold Spring Harb. Perspect. Biol.* **10**, a034801 (2018).
14. Leslie E, O. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).
15. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319–1325 (2001).
16. Portillo, X., Huang, Y.-T., Breaker, R. R., Horning, D. P. & Joyce, G. F. Witnessing the structural evolution of an RNA enzyme. *Elife* **10**, e71557 (2021).
17. Papastavrou, N., Horning, D. P. & Joyce, G. F. RNA-catalyzed evolution of catalytic RNA. *Proc. Natl. Acad. Sci. USA* **121**, e2321592121 (2024).
18. Chen, I. A. RNA life on the edge of catastrophe. *Proc. Natl. Acad. Sci. USA* **121**, e2402649121 (2024).
19. Huang, W. & Ferris, J. P. Synthesis of 35–40 mers of RNA oligomers from unblocked monomers. A simple approach to the RNA world. *Chem. Commun.* **12**, 1458–1459 (2003).
20. Huang, W. & Ferris, J. P. One-step, regioselective synthesis of up to 50-mers of RNA oligomers by montmorillonite catalysis. *J. Am. Chem. Soc.* **128**, 8914–8919 (2006).
21. Kloprogge, J. T. & Hartman, H. Clays and the origin of life: the experiments. *Life* **12**, 259 (2022).
22. Ferris, J. P. Montmorillonite-catalysed formation of RNA oligomers: the possible role of catalysis in the origins of life. *Philos. Trans. R. Soc. B: Biol. Sci.* **361**, 1777–1786 (2006).
23. Jerome, C. A., Kim, H.-J., Mojzsis, S. J., Benner, S. A. & Biondi, E. Catalytic synthesis of polyribonucleic acid on prebiotic rock glasses. *Astrobiology* **22**, 629–636 (2022).
24. Himbert, S., Chapman, M., Deamer, D. W. & Rheinstädter, M. C. Organization of nucleotides in different environments and the formation of pre-polymers. *Sci. Rep.* **6**, 31285 (2016).

25. Attwater, J., Wochner, A., Pinheiro, V. B., Coulson, A. & Holliger, P. Ice as a protocellular medium for RNA replication. *Nat. Commun.* **1**, 76 (2010).

26. Briones, C., Stich, M. & Manrubia, S. C. The dawn of the RNA world: toward functional complexity through ligation of random RNA oligomers. *RNA* **15**, 743–749 (2009).

27. Wachowius, F. & Holliger, P. Non-enzymatic assembly of a minimized RNA polymerase ribozyme. *ChemSystemsChem* **1**, 1–4 (2019).

28. Deck, C., Jauker, M. & Richert, C. Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat. Chem.* **3**, 603–608 (2011).

29. Adamala, K. & Szostak, J. W. Nonenzymatic template-directed RNA synthesis inside model protocells. *Science* **342**, 1098–1100 (2013).

30. Adamala, K., Engelhart, A. E. & Szostak, J. W. Generation of functional RNAs from inactive oligonucleotide complexes by non-enzymatic primer extension. *J. Am. Chem. Soc.* **137**, 483–489 (2015).

31. O'Flaherty, D. K., Zhou, L. & Szostak, J. W. Nonenzymatic template-directed synthesis of mixed-sequence 3'-NP-DNA up to 25 nucleotides long inside model protocells. *J. Am. Chem. Soc.* **141**, 10481–10488 (2019).

32. Anderson, P. W. Suggested model for prebiotic evolution: the use of chaos. *Proc. Natl. Acad. Sci. USA* **80**, 3386–3390 (1983).

33. Rokhsar, D. S., Anderson, P. & Stein, D. Self-organization in prebiological systems: simulations of a model for the origin of genetic information. *J. Mol. Evol.* **23**, 119–126 (1986).

34. Orgel, L. E. Molecular replication. *Nature* **358**, 203–209 (1992).

35. Fernando, C., Von Kiedrowski, G. & Szathmáry, E. A stochastic model of nonenzymatic nucleic acid replication: "Elongators" sequester replicators. *J. Mol. Evol.* **64**, 572–585 (2007).

36. Tkachenko, A. V. & Maslov, S. Spontaneous emergence of autocatalytic information-coding polymers. *J. Chem. Phys.* **143**, 045102 (2015).

37. Fellermann, H., Tanaka, S. & Rasmussen, S. Sequence selection by dynamical symmetry breaking in an autocatalytic binary polymer model. *Phys. Rev. E* **96**, 062407 (2017).

38. Toyabe, S. & Braun, D. Cooperative ligation breaks sequence symmetry and stabilizes early molecular replication. *Phys. Rev. X* **9**, 011056 (2019).

39. Tupper, A. S. & Higgs, P. G. Rolling-circle and strand-displacement mechanisms for non-enzymatic RNA replication at the time of the origin of life. *J. Theor. Biol.* **527**, 110822 (2021).

40. Rosenberger, J. H. et al. Self-assembly of informational polymers by templated ligation. *Phys. Rev. X* **11**, 031055 (2021).

41. Zhou, L., Ding, D. & Szostak, J. W. The virtual circular genome model for primordial RNA replication. *RNA* **27**, 1–11 (2021).

42. Juritz, J., Poulton, J. M. & Ouldridge, T. E. Minimal mechanism for cyclic templating of length-controlled copolymers under isothermal conditions. *J. Chem. Phys.* **156**, 074103 (2022).

43. Chamanian, P. & Higgs, P. G. Computer simulations of template-directed rna synthesis driven by temperature cycling in diverse sequence mixtures. *PLoS Comput. Biol.* **18**, e1010458 (2022).

44. Göppel, T., Rosenberger, J. H., Altaner, B. & Gerland, U. Thermodynamic and kinetic sequence selection in enzyme-free polymer self-assembly inside a non-equilibrium RNA reactor. *Life* **12**, 567 (2022).

45. Walker, S. I., Grover, M. A. & Hud, N. V. Universal sequence replication, reversible polymerization and early functional biopolymers: a model for the initiation of prebiotic sequence evolution. *PLOS ONE* **7**, e34166 (2012).

46. Kaddour, H. & Sahai, N. Synergism and mutualism in non-enzymatic RNA polymerization. *Life* **4**, 598–620 (2014).

47. García-Sánchez, M., Jiménez-Serra, I., Puente-Sánchez, F. & Aguirre, J. The emergence of interstellar molecular complexity explained by interacting networks. *Proc. Natl. Acad. Sci. USA* **119**, e2119734119 (2022).

48. Bianconi, G. et al. Complex systems in the spotlight: next steps after the 2021 nobel prize in physics. *J. Phys. Complex.* **4**, 010201 (2023).

49. Morasch, M., Mast, C. B., Langer, J. K., Schilcher, P. & Braun, D. Dry polymerization of 3',5'-cyclic GMP to long strands of RNA. *ChemBioChem* **15**, 879–883 (2014).

50. Šponer, J. E. et al. Questions and answers related to the prebiotic production of oligonucleotide sequences from 3',5' cyclic nucleotide precursors. *Life* **11**, 800 (2021).

51. Mast, C. B., Schink, S., Gerland, U. & Braun, D. Escalation of polymerization in a thermal gradient. *Proc. Natl. Acad. Sci. USA* **110**, 8030–8035 (2013).

52. Kreysing, M., Keil, L., Lanzmich, S. & Braun, D. Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nat. Chem.* **7**, 203–208 (2015).

53. Lahav, N., White, D. & Chang, S. Peptide formation in the prebiotic era: thermal condensation of glycine in fluctuating clay environments. *Science* **201**, 67–69 (1978).

54. Kompanichenko, V. Inversion concept of the origin of life. *Orig. Life Evol. Biosph.* **42**, 153–178 (2012).

55. Damer, B. & Deamer, D. Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. *Life* **5**, 872–887 (2015).

56. Kaddour, H. et al. Nonenzymatic RNA oligomerization at the mineral–water interface: an insight into the adsorption–polymerization relationship. *J. Phys. Chem. C.* **122**, 29386–29397 (2018).

57. Senatore, F., Serra, R. & Villani, M. Modelling Wet-Dry Cycles in the Binary Polymer Model. In *Artificial Life and Evolutionary Computation* 119–129 (Springer, Cham, 2023).

58. Ianeselli, A. et al. Physical non-equilibria for prebiotic nucleic acid chemistry. *Nat. Rev. Phys.* **5**, 185–195 (2023).

59. Farhat, M., Auclair-Desrotour, P., Boué, G. & Laskar, J. The resonant tidal evolution of the Earth-Moon distance. *Astron. Astrophys.* **665**, L1 (2022).

60. Lathe, R. Fast tidal cycling and the origin of life. *Icarus* **168**, 18–22 (2004).

61. Lathe, R. Tidal chain reaction and the origin of replicating biopolymers. *Int. J. Astrobiol.* **4**, 19–31 (2005).

62. Szostak, J. W. The eightfold path to non-enzymatic RNA replication. *J. Syst. Chem.* **3**, 1–14 (2012).

63. Higgs, P. G. The effect of limited diffusion and wet–dry cycling on reversible polymerization reactions: implications for prebiotic synthesis of nucleic acids. *Life* **6**, 24 (2016).

64. Ianeselli, A. et al. Water cycles in a Hadean CO2 atmosphere drive the evolution of long DNA. *Nat. Phys.* **18**, 579–585 (2022).

65. Ferris, J. P., Hill Jr, A. R., Liu, R. & Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**, 59–61 (1996).

66. Aldersley, M. F., Joshi, P. C., Price, J. D. & Ferris, J. P. The role of montmorillonite in its catalysis of RNA synthesis. *Appl. Clay Sci.* **54**, 1–14 (2011).

67. Jelavić, S. et al. Prebiotic RNA polymerisation: energetics of nucleotide adsorption and polymerisation on clay mineral surfaces. *Chem. Commun.* **53**, 12700–12703 (2017).

68. Bernal, J. D. The physical basis of life. *Proc. Phys. Soc. Sect. B* **62**, 597 (1949).

69. Hanczyc, M. M., Fujikawa, S. M. & Szostak, J. W. Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science* **302**, 618–622 (2003).

70. Bartolucci, G. et al. Sequence self-selection by cyclic phase separation. *Proc. Natl. Acad. Sci. USA* **120**, e2218876120 (2023).

71. Moody, E. R. R. et al. The nature of the last universal common ancestor and its impact on the early Earth system. *Nat. Ecol. Evol.* **8665**, 1654–1666 (2024).

72. Pearce, B. K., Tupper, A. S., Pudritz, R. E. & Higgs, P. G. Constraining the time interval for the origin of life on Earth. *Astrobiology* **18**, 343–364 (2018).

73. Macht, F., Eusterhues, K., Pronk, G. J. & Totsche, K. U. Specific surface area of clay minerals: comparison between atomic force microscopy measurements and bulk-gas (N2) and-liquid (EGME) adsorption methods. *Appl. Clay Sci.* **53**, 20–26 (2011).

74. Lipfert, J. et al. Double-stranded RNA under force and torque: similarities to and striking differences from double-stranded DNA. *Proc. Natl Acad. Sci.* **111**, 15408–15413 (2014).

75. Ervithayasuporn, V. et al. Modifying interlayer space of montmorillonite with octakis (3-(1-methylimidazolium) propyl) octasilsesquioxane chloride. *Appl. Clay Sci.* **171**, 6–13 (2019).

76. Brigatti, M. F., Galan, E. & Theng, B. Structure and mineralogy of clay minerals. In *Developments in Clay Science* Vol. 5, 21–81 (Elsevier, 2013).

77. Mathew, D. C. & Luthey-Schulten, Z. Influence of montmorillonite on nucleotide oligomerization reactions: a molecular dynamics study. *Orig. Life Evol. Biosph.* **40**, 303–317 (2010).

78. de Oliveira, L. H. et al. When RNA meets montmorillonite: Influence of the pH and divalent cations. *Appl. Clay Sci.* **214**, 106234 (2021).

79. Flory, P. J. *Principles of Polymer Chemistry* (Cornell University Press, 1953).

80. Svoboda, P. & Cara, A. D. Hairpin RNA: a secondary structure of primary importance. *Cell. Mol. Life Sci.* **63**, 901–908 (2006).

81. Bernhardt, H. S. & Tate, W. P. Primordial soup or vinaigrette: did the RNA world evolve at acidic pH? *Biol. Direct* **7**, 1–12 (2012).

82. Roy, S., Bapat, N. V., Derr, J., Rajamani, S. & Sengupta, S. Emergence of ribozyme and tRNA-like structures from mineral-rich muddy pools on prebiotic earth. *J. Theor. Biol.* **506**, 110446 (2020).

83. Aguilar-Ovando, E. & Negron-Mendoza, A. Radiation chemistry approach to the study of sedimentary microenvironments as models for the protection of bio-organic molecules on the early earth. *J. Radioanal. Nucl. Chem.* **286**, 637–642 (2010).

84. Zhang, K. et al. RNA hydrolysis at mineral–water interfaces. *Environ. Sci. Technol.* **57**, 8280–8288 (2023).

85. Saha, R., Kao, W.-L., Malady, B., Heng, X. & Chen, I. A. Effect of montmorillonite K10 clay on RNA structure and function. *Biophys. J.* **123**, 451–463 (2024).

86. Szathmáry, E. Four letters in the genetic alphabet: a frozen evolutionary optimum? *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **245**, 91–99 (1991).

87. Szathmáry, E. Why are there four letters in the genetic alphabet? *Nat. Rev. Genet.* **4**, 995–001 (2003).

88. Pinheiro, V. B. et al. Synthetic genetic polymers capable of heredity and evolution. *Science* **336**, 341–344 (2012).

89. Castanedo, L. A. & Matta, C. F. Prebiotic N-(2-Aminoethyl)-Glycine (AEG)-Assisted Synthesis of Proto-RNA? *J. Mol. Evol.* **92**, 449–466 (2024).

90. Szathmáry, E. & Smith, J. M. The major evolutionary transitions. *Nature* **374**, 227–232 (1995).

91. Franchi, M. & Gallori, E. A surface-mediated origin of the RNA world: biogenic activities of clay-adsorbed RNA molecules. *Gene* **346**, 205–214 (2005).

92. Joshi, P. C., Aldersley, M. F., Delano, J. W. & Ferris, J. P. Mechanism of montmorillonite catalysis in the formation of RNA oligomers. *J. Am. Chem. Soc.* **131**, 13369–13374 (2009).

93. Dujardin, A., Himbert, S., Pudritz, R. & Rheinstädter, M. C. The formation of RNA pre-polymers in the presence of different prebiotic mineral surfaces studied by molecular dynamics simulations. *Life* **13**, 112 (2022).

94. Wu, T. & Orgel, L. E. Nonenzymic template-directed synthesis on oligodeoxycytidylate sequences in hairpin oligonucleotides. *J. Am. Chem. Soc.* **114**, 317–322 (1992).

95. Blain, J. C. & Szostak, J. W. Progress toward synthetic cells. *Annu. Rev. Biochem.* **83**, 615–640 (2014).

96. Li, L. et al. Enhanced nonenzymatic RNA copying with 2-aminoimidazole activated nucleotides. *J. Am. Chem. Soc.* **139**, 1810–1813 (2017).

97. Ding, D., Zhou, L., Mittal, S. & Szostak, J. W. Experimental tests of the virtual circular genome model for nonenzymatic RNA replication. *J. Am. Chem. Soc.* **145**, 7504–7515 (2023).

98. Harrison, T. M., Schmitt, A. K., McCulloch, M. T. & Lovera, O. M. Early (≥ 4.5 ga) formation of terrestrial crust: Lu-Hf, $\delta^{18}O$, and Ti thermometry results for Hadean zircons. *Earth Planet. Sci. Lett.* **268**, 476–486 (2008).

99. Stüeken, E. E. et al. Did life originate from a global chemical reactor? *Geobiology* **11**, 101–126 (2013).

100. MacDonald, G. J. Tidal friction. *Rev. Geophys.* **2**, 467–541 (1964).

101. Eigen, M. & Schuster, P. The hypercycle: a principle of natural self-organization Part B: the abstract hypercycle. *Naturwissenschaften* **65**, 7–41 (1978).

102. Lorenz, R. et al. ViennaRNA package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).

Acknowledgements

Author contributions

C.A., A.A., C.B., and J.A. designed the study, conceived the computational environment EarlyWorld, discussed the results and wrote the paper; C.A., A.A., and J.A. implemented the algorithm; C.A. and A.A. performed the numerical experiments; J.A. and A.A. developed the analytical work; and C.B. explored the connections with experimental prebiotic chemistry.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42004-025-01632-w.

Correspondence and requests for materials should be addressed to Jacobo Aguirre.

Peer review information *Communications Chemistry* thanks Chérif F. Matta and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.