
Ligation of random pools and autonomous ribozymatic RNA replication in thermal non-equilibria.

Anna Magdalena Salditt



München 2023

Ligation of random pools and autonomous ribozymatic RNA replication in thermal non-equilibria.

Anna Magdalena Salditt

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Anna Magdalena Salditt
aus München

München, den 02.11.2023

Erstgutachter: Prof. Dr. Dieter Braun
Zweitgutachter: Prof. Dr. Hannes Mutschler
Tag der mündlichen Prüfung: 14.12.2023

Zusammenfassung

Eine zentrale Eigenschaft des Lebens ist seine Fähigkeit, Informationen zu speichern, zu replizieren und zu übersetzen. Dies wirft zwangsläufig die Frage auf, wie die Replikation auf der frühen Erde entstanden ist. Aufgrund der Fähigkeit der RNA, Sequenzinformationen zu speichern und sich zu katalytisch aktiven Komplexen zu falten, stehen ihre Replikation und die Suche nach Ribozymen seit Jahrzehnten im Mittelpunkt der Forschung zum Ursprung des Lebens. Effiziente RNA-Replikation, die möglicherweise zu einer darwinschen Evolution führt, und RNA-Welt-Szenarien, in denen Ribozyme mehrere chemische Reaktionen katalysieren, stehen jedoch vor dem gleichen Hindernis, nämlich der Blockierung der Produkte nach einem erfolgreichen Kopierschritt in einem Produkt-Templat Komplex.

Wie können thermische Nichtgleichgewichte helfen, dieses Hindernis der RNA Replikation zu überwinden? Ziel dieser Arbeit ist es, zu zeigen, dass thermische Nichtgleichgewichtssysteme bestimmte Replikationsprozesse nicht nur unterstützen, sondern sie überhaupt erst ermöglichen. Der erste Teil befasst sich mit der Frage, wie thermische Nichtgleichgewichtsbedingungen den Übergang von Pools kurzer zufälliger Oligomere zu längeren Produktsträngen mit einem reduzierten Sequenzraum ermöglichen. Hier werden die Auswirkungen zyklischer Temperaturschwankungen auf die Ligationsdynamik zufälliger Pools von 12 nt langen DNA Oligomeren in Gegenwart einer DNA *Taq* Ligase untersucht. Diese Experimente zeigen, wie unterschiedliche Temperaturbedingungen die Ausdehnung von zufälligen Pools mit verschiedenen Sequenzräumen entweder verhindern oder ermöglichen können. Darüber hinaus wird gezeigt, dass wenige höher konzentrierte Sequenzen, die jeweils ihr eigenes Replikationsnetzwerk bilden können, das gesamte Replikationsverhalten und die Replikationsrate erheblich beeinflussen können. Dies zeigt, wie nur kleine Unterschiede im Ausgangspool verstärkt werden und zu unterschiedlichen langen Sequenzen führen.

Im zweiten Teil dieser Arbeit werden zwei verschiedene geophysikalische Szenarien der frühen Erde experimentell imitiert. In ihnen wird RNA mit Hilfe von Ribozymen repliziert. Im ersten Fall wird eine zylindrische, komplett mit Wasser gefüllte Kammer verwendet. In dieser Kammer werden Dank einer punktförmigen Hitzequelle Temperaturzyklen erzeugt, indem Moleküle konvektiv zwischen dem heißen Temperaturpunkt und dem kalten Außenbereich der Kammer hin- und hergeschoben werden. So können mit Hilfe der 24-3 Ribozym Polymerase mehrere Runden PCR-ähnlicher Replikation angetrieben werden. Gleichzeitig wird die längere RNA selektiv weg von höheren Temperaturen akkumuliert, wodurch die Polymerase vor hohen Temperaturen und der damit einhergehenden Hydrolyse geschützt wird. Das zweiten Szenario besteht aus einer Gas-Wasser-Grenzfläche in einem Temperaturgradienten. Innerhalb desselben Reaktionskompartiments können sich die Moleküle nun in verschiedene Phasen, Salzkonzentrationen und Temperaturgradienten aufhalten. Regionen mit niedriger und hoher Magnesiumkonzentration führen dazu, dass sowohl Replikation als auch Trennung von RNA-Strängen ohne künstliches Eingreifen möglich werden. Selbst für Sequenzen, die komplementär zu dem replizierendem Ribozym *sunY* sind, finden voll-

ständige Replikationszyklen statt - ein erster Schritt in Richtung autokatalytischer Selbstreplikation. Darüber hinaus wird gezeigt, dass so nicht nur die Replikation eines aktiven Hammerhead-Ribozyms, sondern auch die darauffolgende Separation von Templat und Kopie und die sekundäre katalytische Reaktion innerhalb der selben Kammer ermöglicht werden. Dies ist ein entscheidender Schritt für ein RNA-Welt Szenario, bei dem mehrere Ribozyme diverse katalytische Funktionen parallel erfüllen müssen.

Abstract

A central quality of life is its ability to store, replicate, and translate (sequence) information. Naturally, this provokes the question of how replication emerged on early Earth. Due to the ability of RNA to store sequence information as well as to fold into catalytically active complexes (ribozymes), its replication and the search for ribozymes receive abiding interest. Efficient RNA replication is a prerequisite for open-ended evolution and RNA world scenarios, where ribozymes catalyze multiple chemical reactions. However, they face the same obstacle - the product-template inhibition after a successful copying step.

How can thermal non-equilibria help overcome this obstacle? This thesis aims to show that thermal non-equilibrium systems not only help certain replication processes but even make them possible in the first place. In the first part, the templated ligation of random or semi-random pools of 12 nt long DNA oligomers by a DNA *Taq* polymerase is used as a model system for early non-enzymatic replication systems. Thermal cycling, a most simple non-equilibrium, enables templated ligation of these random pools to elongate to longer product strands with a reduced sequence space. The employed cycling conditions critically influence the outcome and can either hinder or promote the extension of random pools depending on the size of their sequence space. Further, spiking the system with subsets of sequences, each capable of forming its own replication network, alters the elongation behavior significantly. This demonstrates how biases in the starting pool are amplified and lead to different long sequences.

In the second part of this work, two different geophysical scenarios of the early Earth are experimentally mimicked to accommodate replication with ribozymes. In the first setting, a cylindrical water-filled chamber with a point-like temperature source induces temperature cycles by convectively shuttling molecules between the hot temperature spot and the cold outer region. In this chamber, multiple rounds of PCR-like replication can be driven using 24-3 ribozyme polymerase. At the same time, the longer RNA is selectively accumulated away from higher temperatures, protecting the polymerase from high temperatures and associated hydrolysis. In the second setting, an air-water interface in a temperature gradient adds another level of complexity to the system. Within the same reaction compartment, molecules can now reside in different phases, salt concentrations, and temperatures, allowing different reactions, such as replication and separation of RNA strands without human intervention. Even complete replication cycles of RNA sequences complementary to the replicating *sunY* ribozyme are thus possible, indicating a pathway to autocatalytic self-replication. Moreover, this setting allows for one-pot reactions that enable replication of an active hammerhead ribozyme, subsequent separation of template and copy, and secondary cleavage reaction. This is a critical step in an RNA world scenario where ribozymes must perform multiple catalytic functions in parallel.

Contents

1	Motivation: Emergence of life as an interdisciplinary challenge	1
2	Fundamentals	7
2.1	Thermal non-equilibrium settings	7
2.1.1	Water-filled compartments in temperature gradient	7
2.1.2	Air-Water Interfaces in temperature gradients	8
2.2	Replication and sequence space of nucleic acids	9
2.3	Ribozymatic replication	10
3	Sequence space influences the ligation dynamics of random pools	13
3.1	Introduction	14
3.2	Results	15
3.2.1	Influence of sequence space on the ligation dynamics	15
3.2.2	ATGC random pools	18
3.2.3	Reduced complexity system	20
3.2.4	Short sequence subsets drive elongation behavior	24
3.3	Discussion and Outlook	32
3.4	Materials and Methods	35
3.4.1	Nucleic Acids	35
3.4.2	Reaction conditions	37
3.4.3	Temperature cycling protocols	37
3.4.4	Melting curves	37
3.4.5	<i>Taq</i> DNA Ligase specifications	38
3.4.6	PAGE analysis	39
3.4.7	Illumina sequencing	39
3.4.8	Sample Pearson Correlation Coefficient Matrix	40
3.5	Appendix - Supplementary Figures	41
4	A thermal habitat for RNA amplification and accumulation	45
4.1	Introduction	46
4.1.1	RNA model system for early replication scenarios	46
4.1.2	Thermal habitat for RNA replication	48
4.2	Results	50
4.2.1	Convective temperature oscillations	50
4.2.2	Amplification of RNA by riboPCR in convection chamber	51
4.2.3	Thermophoretic accumulation of RNA polymerase	52
4.2.4	Finite-element simulation of the thermophoretic accumulation	54
4.2.5	Probability density of Polymerase and RNA degradation	57
4.3	Discussion	60
4.4	Conclusion	61

4.5	Materials and Methods	62
4.5.1	Nucleic acids	62
4.5.2	RNA-catalyzed replication of RNA	62
4.5.3	Convection system	63
4.5.4	Finite-element simulations	64
4.5.5	Random walk model	66
4.5.6	Sigmoidal model for exponential replication	66
4.6	Supplementary Figures	67
5	RNA synthesis and replication by the <i>sunY</i> ribozyme driven by a non-equilibrium microenvironment	71
5.1	Introduction	72
5.1.1	The replication problem in the RNA world	72
5.1.2	Exploiting the non-equilibrium of the AWI-setting	73
5.1.3	<i>sunY</i> as model system for early ribozymatic replication	76
5.2	Results	78
5.2.1	<i>sunY</i> -catalyzed RNA ligation at heated air-water interfaces	78
5.2.2	<i>sunY</i> -mediated ligation of <i>sunY</i> fragments	80
5.2.3	Coupled sense and antisense synthesis of <i>sunY</i> fragments	82
5.2.4	Synthesis of the hammerhead ribozyme	85
5.2.5	One-pot synthesis and folding of active hammerhead ribozyme	89
5.3	Discussion and Outlook	92
5.4	Materials and Methods	95
5.4.1	Nucleic acids	95
5.4.2	<i>sunY</i> transcription protocol	96
5.4.3	Reaction details of the ligation of M123	97
5.4.4	Reaction details of synthesis of <i>sunY</i> fragments	98
5.4.5	Reaction details of HH-min synthesis	98
5.4.6	Polyacrylamide gel electrophoresis (PAGE)	98
5.4.7	Preparation of AWI chambers and sample injection	99
5.4.8	Measurement of bulk pH changes in AWI	101
5.4.9	Temperature simulations with COMSOL	101
5.4.10	Melting temperatures	103
5.4.11	Stability analysis of the ligation systems	105
5.5	Appendix - Supplementary Figures	108
6	Discussion and Outlook	109
7	Bibliography	113
8	List of Publications	129
9	Acknowledgements	157

1 Motivation: Emergence of life as an interdisciplinary challenge

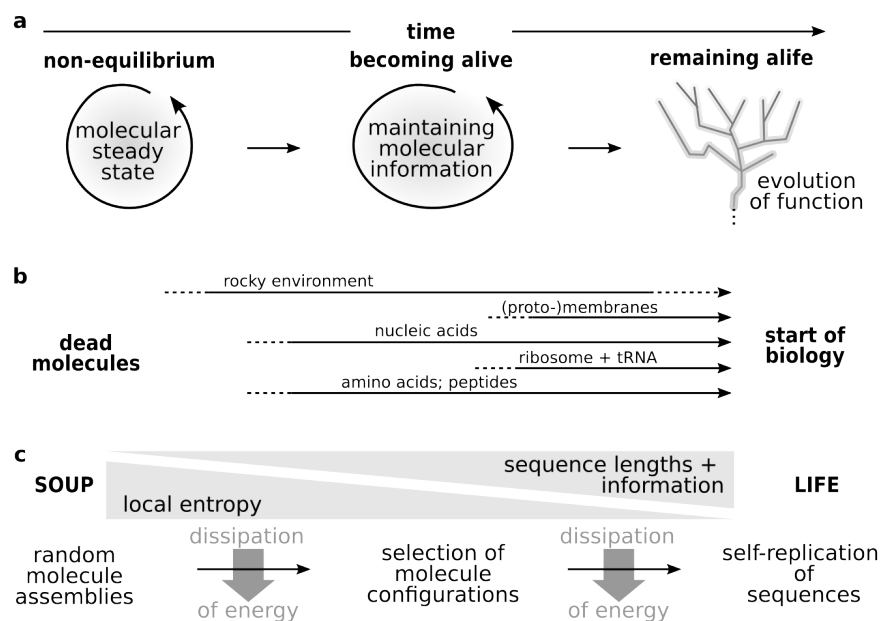


Figure 1.1: Becoming alive - Extrapolating back from life: **a** A pool of essential molecules is needed to find a steady state of synthesis and decay and to create information as a sequence of building blocks through evolution. This mechanism was likely driven by non-equilibrium conditions. Next, the sequence information must be sustained and selected by replication before the strands decay. This is the basis of Darwinian evolution and essential for the speciation as we know it. **b** Extrapolating back from biology gives us information as to what early life converged to. At the same time, the geology of early Earth provides energy sources and environments to drive prebiotic chemistries. Investigation of the plausible building blocks, their synthesis pathways, and their interplay with the environments provides the basis for understanding the road towards life. **c** Every reduction in entropy must be driven by the dissipation of energy. This applies to an early mixture of molecules suffering from degradation and dilution as well as to the information encoded in more complex molecules.

Traditionally, the emergence of life is told along a time axis. In short, extrapolation shows that early forms of life presumably date back to before 3.8 billion years, giving an upper bound for the emergence of life [1–3]. On the other hand, the moon forming impact 4.5 billion years ago induced such a drastic change of the early Earth that this event is considered the lower bound for the emergence of life [4–7]. The subsequent cool-down of Earth over the span of 200 million years provided reducing conditions in the atmosphere [8, 9] for the emer-

gence of the first life forms about 600 million years later [10]. However, instead of focusing on a time axis, one can also collect the main lessons that can be learned from disciplines contributing to the origin of life field. The overall aim is to explain how to advance from high entropy diluted species to low entropy information-rich species to start the first steps in creating the autocatalytic networks of living systems (see Figure 1.1).

Let us start with **biology**. Phylogenetic trees are a powerful tool of biology to correlate genetic information between species and are even used to trace back some core features of *LUCA*, the last universal common ancestor [11]. However, extrapolations using sequencing [12–14] by definition cannot tell us anything about molecular evolution and the time before proteins, the genetic code, and ribozymes emerged. Still, phylogenetic trees show that in particular the active site of the ribosome responsible for protein synthesis has been highly conserved. Together with the amino acid-bearing tRNA, this hints towards an important role of RNA in the emergence of life [15–17]. An often cited hypothesis in this context is the so-called RNA world hypothesis [18, 19]. It is based on the dual role of RNA in biology as it is not only able to store information, determined by the order of its four basic building blocks but can also fold into enzymatically active complexes, called ribozymes. Thus, paths towards creating open-ended molecular evolution starting from RNA [20] or similar molecules [19, 21–26] are highly investigated.

The field of prebiotic **chemistry** has achieved progress to elucidate plausible reaction pathways [27–32]. The aim is to find not only reactions with high yields from pure materials but also networks of reactions that self-select toward the desired products. Recent studies on RNA found new synthesis pathways yielding its building blocks such as purines, pyrimidines, and related molecules [33–36]. These reactions often start with pure materials at high concentrations in order to avoid unwanted side products. Besides the synthesis of RNA and protocols to drive its polymerization [37–42], the robust formation of long peptides by cyclic addition of amino acids using sulfur compounds showed possible routes for selective peptide creation [43, 44]. This is an important step towards understanding the origins of modern protein-based metabolisms. Also, ideas about translation - making proteins from RNA - have come to closer scrutiny with experiments [16, 45–47]. Following the creation of short nucleic acid pools, their replication by templated polymerization or templated ligation is another challenging question. Recent studies show that non-enzymatic RNA replication, i.e., replication without a protein or ribozyme, by imidazole activation is promising [37, 40, 48]. Another approach uses EDC¹-driven polymerization [49, 50]. However, EDC tends to trigger several unspecific side reactions that can interfere with downstream reactions [51]. Even though many approaches have been studied, one major problem with any replication mechanism (enzymatic or non-enzymatic, polymerization or ligation) remains - its limitation by strand separation [52] or product inhibition [24, 53–55].

Nevertheless, all these chemistries must be compatible with what **geology** can reveal about the conditions of the early Earth as well as with what **astrophysics** can tell about the molecules present at that time. The emergence of life must be placed in the context of planetary and space science. This research is not limited to Earth but is a universal question involving early planets [56, 57] with astrochemistry around an early protostar [58, 59]. The search

¹short for 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimid

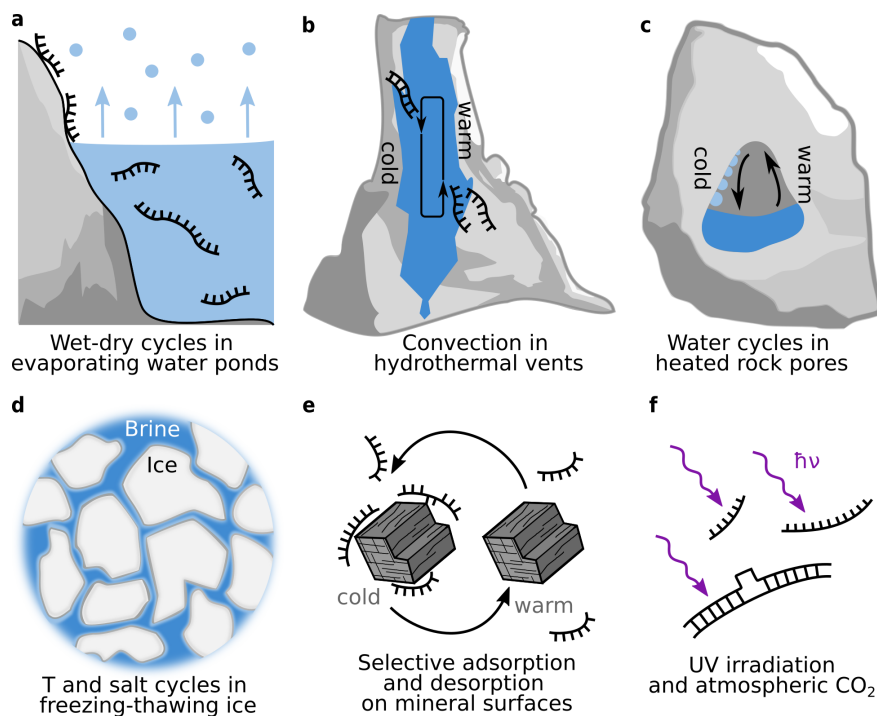


Figure 1.2: Chemical and physical non-equilibrium settings at small scales:

a The evaporation at water ponds leads to wet-dry cycles and oscillating salt concentrations. **b** Elongated, water-filled clefts can act as thermogravitational columns that selectively accumulate molecules when exposed to temperature gradients. **c** Porous rocks exposed to a temperature gradient can lead to miniature water cycles. Drying processes due to evaporation accumulate diluted molecules at the gas-water interface, while condensation droplets form in the gas phase. **d** Freeze-thaw cycles lead to changes in concentration due to volume exclusion. **e** Catalytic surfaces on particles can trigger selective adsorption and desorption. **f** UV-radiation could be a mode of early selection pressure.

for life in space also initiated missions, e.g., to Mars and the moons of Saturn [60, 61]. This not only gives us clues on possible extraterrestrial lifeforms but also helps to unravel the astro-geological history of Earth. As already mentioned, the cool-down of Earth after the moon-forming impact resulted in reducing conditions in the atmosphere [8, 9]. In addition, both fields can provide information about potential selection pressures (Figure 1.2). An example is the selective adsorption and desorption on minerals [62]. Another very prominent selection pressure is the exposure to UV radiation of the early sun [63–65].

Prebiotic chemistry must, at least at some point, operate in water. Its presence on the crust of early Earth since before 4.3 billion years was shown by oxygen isotope analysis of zircons [66]. However, water is a challenging environment, as dilution and hydrolysis imply the need for high yields. This is difficult to achieve in purely chemical systems. When employing the **physics** of far from equilibrium boundary conditions, major challenges can be circumvented. Chemical reaction pathways are controlled by the concentrations of reactants and their mixing, but also temperature, pH, and salts. However, if these quantities are not

constant but change over the reaction volume and time, the resulting non-equilibrium can be a key driver to change or maintain reaction conditions. Simple and abundant sources to generate local non-equilibria are temperature gradients, for example provided by day-night cycles (Figure 1.2 a) [67], hydrothermal [68–70] or volcanic [71, 72] activity (Figure 1.2 b,c). At millimeter scales, these lead to buoyancy-driven fluid convection resulting in temperature cycles [73, 74]. In addition to convection, temperature gradients lead to thermophoresis [75–77], i.e., the movement of molecules along the gradient. Combining both can lead to an accumulation of molecules and selection of longer strands [78]. Other phoretic transports, e.g., the movement of molecules along pH-, concentration- or salt-gradients, are expected in heterogeneous geological settings and can lead to elaborate effects [79].

Phase transitions and interfaces provide another interesting mode of accumulation and localized conditions. For instance, water-ice systems (Figure 1.2 d) have been shown to accumulate catalytic RNA and salts in their brine regions [80] and provide long-term storage and suitable reaction conditions [81, 82]. In addition, the formation of dew droplets upon evaporation and condensation cycles of air-water interfaces exposed to temperature gradients creates attractive mixing and accumulation dynamics [83, 84]. Liquid-liquid phase separation of more complex molecules, in the form of coacervates [85–88] as well as DNA- and RNA-gels [89, 90] is another mode of self-organization. These larger structures can also sediment in Earth’s gravity field due to their difference in local density [47, 91] leading to interesting effects of selection and enrichment of those species.

Finally, theoretical modeling and simulations are crucial for advancing the field. Detailed numerical and theoretical models have already helped to describe fundamental qualities of evolutionary dynamics [92–94], replication [95–98] and the emergence of homochirality [99, 100].

It is clear that the emergence of life on early Earth is one of today’s most puzzling scientific questions, which can be tackled from many different directions. Even though *life* is yet to be defined, its ability to (self-) replicate and make use of sequence information is undeniably one of its most striking features and, thus, a good starting point. This thesis aims to contribute to this by studying how thermal non-equilibrium systems can aid in various aspects of early replication and its emergence on Earth. To this end, three research questions are addressed, and their results are presented in chapters 3, 4, and 5.

Is it possible to find long functional sequences in their vast possible sequence space?

Chapter 3 makes use of the most simple thermal non-equilibrium, i.e., periodic cycling between two defined temperatures, thus leading to temporal changes of conditions. It addresses how these cycling conditions influence the ligation dynamics of random pools concerning their sequence space. To this end, pools of 12 nucleotide long DNA oligomers in the presence of a DNA *Taq* ligase are employed as a model system for any prebiotic ligation chemistry. Specifically, it is shown how cycling conditions can drive or inhibit the extension of pools with various sequence spaces. Further, it is demonstrated that spiking these pools with different subsets of sequences, each building its own replication network, can influence the starting pools’ overall replication behavior and outcome, thus limiting and biasing the resulting sequence space.

Can ribozymatic replication be driven by realistic scenarios instead of test tubes?

Periodic cycling conditions can also be implemented in a more realistic and natural way using micro-scale water-filled chambers that are exposed to a point-like temperature source. Interestingly, this not only results in temporal but also spatial changes in the conditions. In chapter 4, it is shown how such a reaction compartment can drive several rounds of PCR-like replication using the 24-3 polymerase ribozyme. In contrast to simple temporal cycling of the overall temperature of a test tube, this setting additionally selectively accumulates longer RNA away from the hot temperature regions. Thereby, it protects the polymerase from temperature-induced hydrolysis.

Can non-equilibrium conditions resolve issues of ribozymatic replication to make an RNA world more plausible?

Introducing an air-water interface in such a system leads to another level of complexity of the non-equilibrium system. Instead of only a liquid phase, this setting provides a rich environment that allows reactions to exploit not only the liquid and the dry phases but also various salt and temperature conditions. Chapter 5 highlights the potential of this setting to host an RNA world scenario. As accumulation and dew cycles within this setting provide regions of high and low magnesium concentrations at moderate temperatures, it is a promising candidate to circumvent the so-called 'strand separation problem' of early replication. Employing replication reactions catalyzed by the *sunY* ribozyme, full replication cycles of sequences that are complementary to *sunY* itself are demonstrated, showing a potential path towards autocatalytic self-replication. Furthermore, the setting allows for a one-pot reaction of the replication of an active hammerhead ribozyme and its secondary cleavage reaction. This is central for a plausible RNA world scenario, where different ribozymes must carry out several catalytic functions in parallel.

Chapter 6 then summarizes and connects the previous result chapters and provides an outlook on further research possibilities. But first, the chapter 2 provides a more detailed explanation of three concepts that are especially relevant to this thesis, i.e., thermal non-equilibrium settings, nucleic acids and their sequence space, and ribozymatic replication.

2 Fundamentals

2.1 Thermal non-equilibrium settings

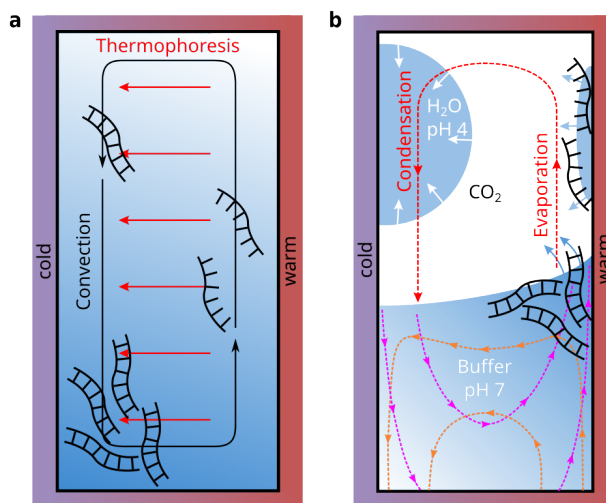


Figure 2.1: Thermal non-equilibria: Schematic representation of plausible thermal non-equilibrium settings on early Earth. **a** Water-filled clefts that are exposed to a temperature gradient can act as thermogravitational traps, based on the interplay of convective movement (black arrows) of the bulk and thermophoretic drift (red arrows) of solutes, e.g., nucleic acids. **b** Gas-water interfaces exposed to a temperature gradient exhibit a water cycle in the gaseous phase (CO_2) of evaporation at the warm and condensation at the cold side of the chamber. The accumulation of molecules at the warm side of the interface is mainly driven by the capillary flow (pink arrows). Marangoni flow (orange arrows) and convective flow are generally too weak to annihilate the accumulation.

Non-equilibrium conditions can offer an up-concentration mechanism which can compensate for low but selective yields. In addition, those geophysical settings can act as a selection pressure for emerging product species. This section provides a brief summary of the theoretical description of the effects in a water-filled compartment and a compartment containing a gas-water interface.

2.1.1 Water-filled compartments in temperature gradient

Any particle exhibits Brownian motion, i.e., random fluctuations of its position. The diffusive molecular flux \vec{j}_{dif} of dissolved particles is described by Fick's first law:

$$\vec{j}_{dif} = -D \cdot \nabla c \quad (2.1)$$

where D is the diffusion coefficient and ∇c is the concentration gradient of the particles. Considering a water-filled vessel of any sort, this implies an equal distribution of the solutes once thermal equilibrium is reached. This changes, however, if a temperature gradient is applied and the compartment is kept in an out-of-equilibrium state. The temperature-induced changes in fluid density lead to a circular motion of the fluid in Earth's gravitational field, i.e., convection. Thus, the dispersed particles are moved along with the convective flow. The molecular flux due to this transport can thus be written as:

$$\vec{j}_{conv} = \vec{v} \cdot c \quad (2.2)$$

where \vec{v} is the velocity field of the surrounding water. On a microscale, the molecular details of the interface of dissolved particles with the surrounding fluids (hydration shell) [101] lead to thermophoresis, the directed movement of the particles along the temperature gradient. Effectively, the resulting molecular flux can be described by:

$$\vec{j}_{ther} = -D_T \cdot \nabla T \cdot c = -S_T \cdot D \cdot \nabla T \cdot c \quad (2.3)$$

where D_T is the thermophoretic mobility and $S_T = D_T/D$ is the ratio between diffusive and thermophoretic mobility. Following the conservative form of the continuity equation for the concentration of the particles leads to:

$$\begin{aligned} \frac{\partial c}{\partial t} &= -\nabla \cdot (\vec{j}_{dif} + \vec{j}_{conv} + \vec{j}_{ther}) = D \cdot \Delta c + D \cdot S_T \cdot \nabla \cdot (\nabla T \cdot c) - \nabla \cdot (\vec{v} \cdot c) \\ &= D \cdot \Delta c + D \cdot S_T \cdot \nabla T \cdot \nabla c - \vec{v} \cdot \nabla c \end{aligned} \quad (2.4)$$

Here, $\Delta T = 0$ for a linear temperature gradient and $\nabla \cdot \vec{v} = 0$ for incompressible fluids was assumed.

2.1.2 Air-Water Interfaces in temperature gradients

Introducing an air-water interface to a thermal system triggers various effects (see Figure 2.1). Even so, already without a temperature gradient, the evaporation of water from a liquid-gas interface leads to the so-called 'coffee-ring' effect. It refers to an increased concentration of dried material at the edge of the liquid-gas interface on a solid boundary due to the transport of dissolved material. This transport is induced by the capillary flow towards the evaporation regime at the edges. Applying a temperature gradient across this interface leads to enhanced evaporation at the warm side of the interface. Instead of a symmetric capillary flow towards all edges of the interface, it now points from the cold to the warm side of the interface.

Several other effects are triggered by the temperature gradient. First of all, convection of the liquid bulk phase is triggered. As described before, convection is triggered by the temperature dependence of the density of any fluid in a gravitational field. In contrast to the capillary flow, the convective flow points from hot to cold at the interface.

Additionally, a gradient in surface tension at the interface induces the so-called Marangoni flow. The flow field points towards increasing surface tension, thus, leading to the same net flow from the warm to the cold side at the interface as the convective flow. Gradients in surface tension in these systems are most commonly caused by a gradient in temperature or concentration. However, the presence of surface active agents, such as surfactants, can

significantly alter the relative strength of the Marangoni flow, up to the point of suppressing the coffee-ring effect [102].

The interplay of these effects leads to a net accumulation at the warm side of the air-water interface that is specific to the parameters of the environment (dimensions, temperatures, etc.) as well as the molecules [103]. In this thesis, the setup of the air-water interface system (AWI-system) is chosen such that thermophoresis plays a minor role. However, previous work has also combined interfaces with a thermogravitational trap to boost accumulation while enabling polymerization reactions in dry-wet cycles [104].

Apart from the liquid phase, the gas phase also shows a non-equilibrium behavior in the form of evaporation-recondensation cycles when exposed to temperature gradients. Here, the evaporation of water at the warm side of the air-water interface results in an overall decrease in the level of the interface. This leads to the drying of the accumulated material on the warm side. The water vapor is shuttled again via convection to the cold side, where it condenses and forms dew droplets of pure water, i.e., droplets without salts [83]. When exposed to CO₂-rich atmosphere, CO₂ dissolves into droplets, leading to their acidification [84]. As they grow large enough, the dew droplets fall back into solution. This causes the interface to move, upon which the dried material gets dissolved and transported back into the solution. Instead of raining back into solution, a droplet, when large enough, can also rearrange due to surface tension such that it touches both sides of the compartment. In case of close proximity to the surface, these rearranged droplets can dissolve the dried material.

AWI-systems have been proven advantageous as already small temperature differences of 5 - 10 °C trigger significant non-equilibrium effects.

2.2 Replication and sequence space of nucleic acids

DNA and RNA are built up from a successive strand of nucleotide monomers consisting of a nucleobase, a five-carbon sugar (ribose for RNA or deoxyribose for DNA), and a phosphate group. The five canonical nucleobases are adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U), where T is exchanged by U in RNA, whereas A, C and G are used in both. A nucleic acid sequence is then defined as the successive chain of nucleotides of certain length l , which can bind to its complementary sequence in a 1:1 manner, i.e., G to C and A to either T or U. Thus, nucleic acids can either be present in single-stranded form (ssXNA) or as a double-stranded complex (dsXNA) appearing as the famous double helix.

In today's biology, the main role of nucleic acids is to store and translate information given by the sequence. From an origin of life perspective, it is natural to ask how information emerged and, when so, how it could be preserved. It seems plausible that some polymerization mechanisms lead to a pool of oligomers of various lengths and sequences [104, 105]. Within such a pool, the potential number of different sequences for a certain length l is z^l , where z is the number of different bases, e.g. 4 or 2. The information contained in some template sequence could presumably be copied by some non-enzymatic replication reactions employing either of the following two mechanisms: a base-by-base addition on the template (templated polymerization) [37, 38, 106] or the addition of oligomers on a template (templated

ligation) [51, 91, 107–109]. However, the copying rates of replication of these non-enzymatic replication scenarios are small compared to replication scenarios of the RNA world hypothesis employing catalytically active RNA structures (ribozymes).

To spontaneously find such a small subset of active sequences in the vast potential sequence space is highly unlikely and subject of current research. Specifically, replication by templated ligation proves to be an interesting evolutionary strategy as its cooperative behavior provides a self-selection mechanism [110] and leads to a structured sequence space with a reduced entropy for the emerging longer strands [111]. The longer strands can grow via several mechanisms defined by the interplay of dissociation and extension [98]. This collective emergence starting only from mononucleotides and short oligomers was studied *in silico* with a mean-field [98] as well as an explicit sequence dependent [112] approach. Additionally, similar reaction setups could potentially enhance the accuracy of the copied strands by kinetic error filtering in which stalling due to faulty incorporation of bases inhibits the completion of the copy when limiting the reaction times [113]. However, in the described *in silico* studies, only intermolecular base pairing was considered, leaving the emergence and influence of structure due to intramolecular base pairing an open question.

2.3 Ribozymatic replication

In addition to non-enzymatic replication mechanisms, enzymatic replication via catalytically active RNA structures (ribozymes) plays a major role in the research of the emergence of early replication and life. The research on enzymatically active nucleic acid structures was kick-started by the discovery of a ribozyme at the core of the ribosome, the key molecule of modern translation [114]. It is widely conserved across all life forms and has been called the 'smoking gun' of the RNA world hypothesis [19, 81]. Another prominent example from biology is the ribonuclease P (RNase P), which is a ribozyme that cleaves a precursor RNA sequence at the 5' end of tRNA [115, 116]. Since then, significant effort has been made to create the shortest possible catalytically active RNA [117–122]. However, ribozymes with more complex functions such as replication seem to require a sequence length above 160 nucleotides [123–126], which poses a problem for their spontaneous emergence due to the small probability of finding them in the vast, accessible sequence space, which would amount to $4^{160} = 2.1 \cdot 10^{96}$ different sequences, a number bigger than estimates for the total count of particles in the observable universe ($\approx 10^{80}$). In this regard, very interesting pre-evolutionary strategies employing ligation reactions (see previous section) are being explored [48, 81, 109, 111, 127]. Note that DNA, although less likely, is also capable of folding into catalytically active structures, so-called DNAzymes [128].

Apart from the already mentioned examples, several other classes of ribozymes have been described, one of which are small ribozymes capable of reversible RNA cleavage and ligation reactions [129]. Prominent examples of this class are the hammerhead and hairpin ribozymes. The reaction takes place at a sequence-specific cleavage (or ligation) site of the phosphodiester bond via a transition state of a cyclic 2',3'-phosphate. More specifically, for the cleavage reaction, the 2'-hydroxyl group gets 'activated' via a general base B , i.e., the general base removes the proton. The resulting 2'-O attacks the adjacent phosphate. A general acid A then donates a proton to the 5'-oxyanion leaving group, leading to a cyclic 2',3'-phosphate group

and 5'-O ends of cleaved sequence products [129, 130]. In the ligation reaction, the roles of *A* and *B* are reversed and assist the attack of the 5'-O nucleophile on the phosphate of the cyclic phosphate. It is very likely that specific nucleobases, dictated by sequence and secondary structure, act as general acids and bases, leading to the rate enhancement of the reactions.

Self-splicing introns (group I and group II), as well as RNase P, comprise a group of large metalloribozymes [81, 130]. Metalloribozymes catalyze reactions by employing site-specific metal ion cofactors, most commonly Mg^{2+} . In general, the required concentrations of magnesium are quite high. Still, they can be reduced by the presence of RNA chaperone proteins. This allows functionality in intracellular conditions even at concentrations of free Mg^{2+} of approximately 1 mM [81, 131]. Group I introns catalyze splicing reactions by a two-step transesterification mechanism. It is initiated by an exogenous molecule of guanosine or guanosine triphosphate and employs a nucleotide binding site, a nucleophilic attack, and metal-ion catalysis [132]. In the first step of this transesterification, the exogenous guanosine (or guanosine triphosphate), bound by the intron, cleaves the splice site and is covalently attached to its 5' end. It then undergoes a conformational change, in which a guanosine at the introns 3' end replaces the guanosine at the binding site. The previously cleaved-off 5' exon then attacks and cleaves the 3' splice site in the second step of the transesterification. This results in the two exons being ligated while the intron is excised from the original sequence. In the case of group II introns, the 2'-hydroxyl of specific adenosine attacks the splice site at the 5' end of the intron. This leads to an intermediate state of a branched lariat-like conformation. Also, the cleaved 5' exon attacks the splice site at the 3' end of the intron, leading to the ligated exons and an excised intron [132]. Interestingly, while the described pathway catalyzes reaction *in cis*, introns can be engineered to ribozymes that catalyze ligation or polymerization reactions *in trans*.

In general, RNA (to an extent also DNA) catalysis can manifest through several mechanisms [81, 130]. Ribozymes can, for example, catalyze general acid/base reactions either by enhancing nucleophilicity through deprotonation or by stabilizing leaving groups by protonation. Other ribozymes employ non-covalent interactions such as hydrogen bonding or electrostatic interactions to stabilize transition states or are able to guide the formation of active conformational states.

SELEX (Systematic Evolution of Ligands by Exponential Enrichment) experiments [133, 134] are *in vitro* selection experiments that have been the mode of choice for hunting long ribozymes for RNA world scenarios and for determining fitness landscapes of sequences [135–137]. Particular focus is placed on the search for ribozymes that are able to copy sequences of general templates. The 'class I ligase' was one of the first RNA replicases that was evolved by *in vitro* selection experiments [138, 139]. Ancestors of this ligase ribozyme were also able to catalyze polymerization reactions with improving fidelity and sequence generality [123, 125, 140]. Especially, the 24-3 polymerase ribozyme [125] provides a significant advancement of sequence generality in copying while maintaining a high level of fidelity. This general feature is essential for the search of replicases as substrate promiscuity of every attached monomer on template base, i.e. low fidelity, is highly undesirable, while substrate promiscuity with respect to the template itself is necessary for copying a variety of different sequences [141]. Other ways of improving sequence generality employ extension using trinucleotide triphos-

phates instead of NTPs [142] or cross-chiral polymerases [143].

So far, most *in vitro* selection experiments have been performed in optimized reaction buffers at moderate to high concentrations of magnesium, crowding agents, and close to neutral pH. However, physico-chemical parameters such as the pH, type and concentration of ions, temperature, and the presence of other molecular species can alter catalytic activity significantly [144, 145]. Not only do they play a huge role in the catalytic performance of known ribozymes (with a given sequence), but can also change the sequence requirements [146] and secondary structure [147, 148] for catalytic activity. A detailed review of ribozomatic replication with respect to the constraints provided by these parameters and the importance of exploring more reaction conditions was published by Le Vay *et al.* [81]. The following paragraphs provide a short summary of the pH, ion, and temperature dependence.

Since the phosphodiester backbones of nucleic acids are highly negatively charged, the presence of cations is required to stabilize duplexes. Many divalent ions, e.g. Ca^{2+} or Fe^{2+} , but also monovalent ions such as Na^+ are able to drive RNA folding into native-like conformations when present in high enough concentrations [149, 150]. For catalytic activity, however, site-specific metal ion cofactors, most commonly Mg^{2+} , are required [144, 151]. While enhancing the activity of ribozymes, especially divalent, and even more so trivalent, metal ions lead to two well-known problems in replication scenarios. First, they promote RNA degradation, especially at high temperatures leading to reduced RNA half-lives of down to minutes [152]. Second, they lead to increased stability of the product-template duplex, preventing multiple rounds of replications [52].

Temperature critically influences the stability of folded RNA structures and impacts the reaction rates. In principle, the reaction rate increases with increasing temperature. However, this trend is limited by denaturation at higher temperatures, first by dehybridization and, with even higher temperatures, eventually by hydrolysis. Note that the individual melting points and the temperature-dependent reaction and hydrolysis rates strongly depend on the sequence and the ions present in solution [81].

The last critical parameter discussed here is the pH of the bulk solution. The pH influences catalytic activity on multiple levels. Similar to temperature, the stability of nucleic acids depends also on the pH, where low pH leads to denaturation due to hydrolysis of RNA [153] and depurination of DNA [154], respectively. Additionally, the protonation of base pairs of G and C at low pH reduces the number of hydrogen bonds, resulting in destabilized duplexes [155]. As the 2'-OH of RNA aids in base-catalyzed hydrolysis of the backbone at high pH, RNA exhibits a stability maximum at pH 4-5 [153]. Further, depending on the employed mechanism of catalysis, ribozymes show different pH dependencies, ranging from a bell-shaped dependence [156] in the case of or a log-linear relationship in the case of the large metalloribozymes [157, 158].

Interestingly, changing the condition of the *in vitro* selection leads to rapidly adapting ribozymes. Employing directed evolution, ribozymes have been evolved to work at a different optimal pH region [159–161] or to be thermally stable at higher temperatures [146, 147]. Starting from random libraries, different sequences and structural motifs were isolated depending on the pH and the type of metal cofactor included in the selection experiments [148].

3 Sequence space influences the ligation dynamics of random pools

Summary

For the emergence of long and potentially functional sequences, pools of short oligomers had to be efficiently elongated. Instead of replicating sequences in a base-by-base manner, replication by templated ligation provides an elongation mechanism that not only elongates very efficiently but also leads to a reduction of the sequence space of the elongated pool. To this end, a simple model system for a prebiotic chemistry, i.e., the ligation of pools of DNA oligomers with a *Taq* ligase, is employed. The first part of this chapter discusses the influence of the sequence space on the parameters that govern the elongation behavior of 12 nt long pools. Depending on the sequence space, the same cycling conditions can either hinder or promote efficient elongation¹. In the second part, experiments with a reduced set of 12 nt long oligomers are characterized². Additionally to describe the influence of reaction parameters such as cycle number, ligation time and temperature, the starting pools are biased (spiked) with subsets of sequences each building different types of replication networks. PAGE analysis and Illumina sequencing show how these networks of sub-sequences can completely dominate the elongation behavior of the complete pool.

¹The first part of the work presented in this chapter contributed to the publication by Kudella et al. [111] in PNAS

Authors: Patrick Kudella, Alexei V. Tkachenko, Annalena Salditt, Sergei Maslov and Dieter Braun

²The second part is based on ongoing collaborative work together with Julio Cesar Espinoza Campos, Ulrich Gerland and Dieter Braun to match experimental results and a detailed numerical simulation, which includes sequence-specific binding and folding model.

3.1 Introduction

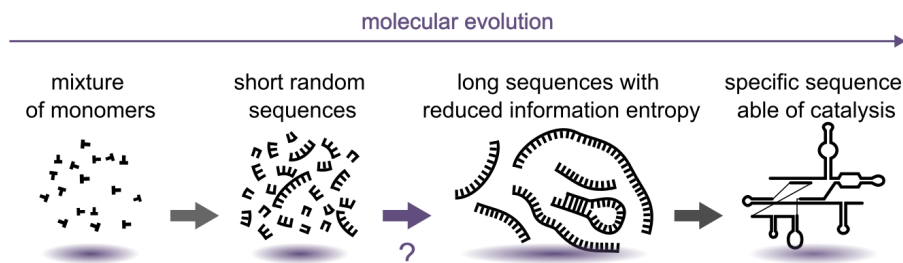


Figure 3.1: Evolution of replication: The emergence of long active RNA complexes is of critical importance for the hypothesized RNA world. A potential path of molecular evolution towards ribozymes would start from a mixture of bases, sugars, and monomers that polymerize into a pool of short random oligomers. Some selection mechanism could then filter out the active sequences capable of catalysis from a pool of pre-selected sequences. Templated ligation is a promising candidate to reduce and pre-select a pool of longer sequences.

While the primary focus of this thesis revolves around ribozymatic replication (see chapters 4 and 5), this chapter delves into the process of templated ligation of random pools. Studying the emergence of long strands from a starting pool of random or semi-random sequences is of general interest for the plausibility of an RNA-world scenario as it remains an open question how these long functional sequences emerged (see Figure 3.1). Simple abiotic polymerization reactions on early Earth would have resulted in a complex mixture of short oligomers with various sequences and lengths [162]. However, understanding the transition between the mixtures of polymerization products to specific long functional sequences remains an open question. This is mainly due to the exponential dependence of the sequence space on length.

Ribozymes that are able to replicate generic RNA sequences are generally of lengths around 180nt [123, 163]. Even though shorter ribozymes of 20-40 nt length exist [164], their functionality is highly limited. Due to the vast potential sequence space, it seems highly unlikely that nature just spontaneously 'found' these longer strands. The sequence space of a 180mer, for example, would amount to $4^{180} \approx 10^{108}$, a number exceeding the estimated number of $\approx 10^{80}$ atoms in the observable universe by several orders of magnitudes. Thus, the transition from short oligomer pools to long functional ribozymes requires an elongation mechanism with a simultaneous reduction in sequence space [96]. To this end, templated ligation is a promising replication mode as it relies on the Watson-Crick base pairing of two fragments that are complementary to a template strand, thus transmitting sequence information from the template to the copy. At the same time, it retains a memory effect of the starting pool.

A ligation reaction can occur if the fragments are bound on the template in direct proximity. Employing some condensation chemistry, they can be covalently linked, therefore replicating the sequence information of the template strand. To avoid dead-end duplexes and gain the possibility of molecular evolution, the system needs to be kept out of thermal equilibrium. Periodic temperature cycling, similar to conditions employed for polymerase chain

reactions (PCR), is a simple means to provide a thermal non-equilibrium. The cycles between a low temperature, where strands are bound and ligated, and a short temperature spike to induce melting can lead to multiple rounds of elongation. On early Earth, periodic changes between conditions for ligation and melting could have been provided by non-equilibrium settings such as laminar convection induced by temperature gradients across water-filled rock pores [165,166] or local changes of pH at air-water interfaces induced by rainfall [84,167].

Experimentally, such a reduction in sequence space has been shown recently for a starting pool of 12 nt long strands with random sequences with a binary alphabet [111]. When exposed to periodic temperature cycling, it evolved into multimer pools with a reduced entropy by means of template ligation. This chapter highlights the effect of an effective dilution in sequence space and how it can hinder the system from displaying longer sequences. Depending on the sequence space of an initial pool, the same cycling conditions can either promote an elongation behavior, transitioning the system to an autocatalytic regime, or hinder elongation, leaving the system in a monomer-dominated regime. Further, it is discussed how spiking the starting pool, i.e., increasing the concentration of specific sequences or a subset of sequences, can influence the overall ligation dynamics and outcome.

3.2 Results

3.2.1 Influence of sequence space on the ligation dynamics

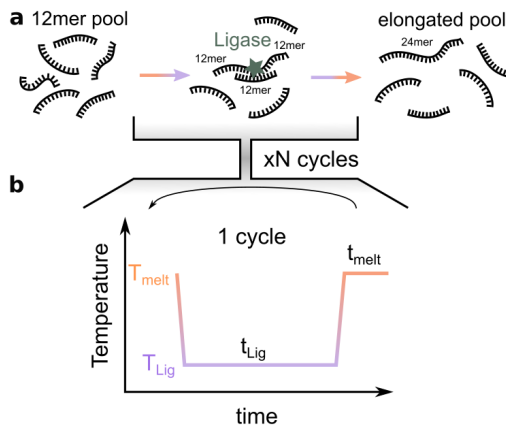


Figure 3.2: Experimental conditions for the ligation reaction starting from random pools: **a** Starting from a pool of 12mers with (semi-) random sequences, the first ligation reactions are templated by the 12mers themselves producing longer sequences. The resulting sequences can, upon temperature cycling, dissociate and further participate in ligation reactions. **b** The reaction is run for a total number of xN cycles. One cycle unit is defined by the ligation (T_{Lig} , purple) and melting temperature (T_{melt} , orange) and their respective durations (t_{Lig} and t_{melt}).

To model prebiotic ligation reactions, pools of 12 nt long DNA oligomers with various sequence spaces are ligated by a *Taq* polymerase, as described in Kudella *et al.* [111, 168] (for details on the preparation of the experiments, see section 3.4.2). The random or semi-random

starting pools are exposed to periodic temperature cycling, after which the resulting length distribution is analyzed by polyacrylamide gel electrophoresis (PAGE) (see section 3.4.6). In the initial ligation reactions, the 12mers act as both template and ligation substrates, with two strands being hybridized onto a third one, acting as a template. When the substrate strands are in direct proximity, they are covalently linked (ligated) by the *Taq* DNA ligase (see Figure 3.2 a). In subsequent rounds of replication, those elongated strands can then participate in the reaction via various elongation modes, e.g., acting as template [98]. To overcome product-template inhibition and to keep the system in an out-of-equilibrium state, the reaction undergoes repeated rounds of temperature cycling (see Figure 3.2 b) between a dehybridization temperature T_{melt} to induce melting of duplex conformations and a ligation temperature T_{lig} to allow hybridization and a subsequent ligation step. The overall runtime of the experiment is governed by the number of cycles, denoted as xN , the ligation time t_{lig} , as well as the dehybridization time t_{melt} . Evidently, an elongated pool can only emerge if ligation reactions occur, which take place with a ligation rate k_{lig} . In addition, the hybridization k_{on} and dehybridization rate k_{off} are crucially influencing the overall rate of extension as the ligation step can only occur while the two substrate strands are bound to a template in a ligation complex. For the right parameter set, this type of reaction setup in thermal non-equilibrium cycling can lead to an elongated pool of oligonucleotides with lengths of multiples of 12 nt, overcoming a monomer-dominated state.

The sequence space of the starting pool is given by z^l , where l is the length of the oligomers and z the number of bases of the used alphabet. For the case of DNA, the possible bases are adenine (A), cytosine (C), guanine (G), and thymine (T), leading to $z = 4$. This would result in a vast sequence space of $4^{12} = 16777216$ when considering 12mers (see Figure 3.3). One way of limiting the sequence space is by reducing the length. Experimentally, this approach is constrained by the requirements of the *Taq* DNA ligase, which needs approximately a 6 nt overlap in order to efficiently ligate the strands. Thus, an easy way to reduce the apparent sequence space while obeying these restrictions can be to limit z to a binary alphabet (A, T). The accessible sequence space then collapses to $2^{12} = 4096$.

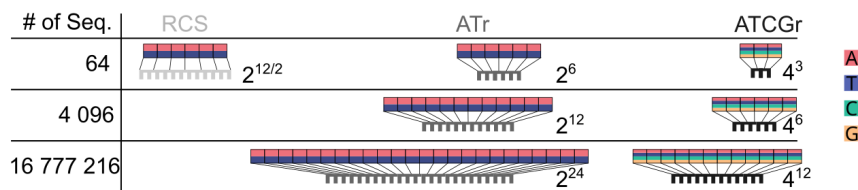


Figure 3.3: Influence of length and number of bases on the sequence space:

Even small changes in length or alphabet can have a drastic influence on the size of the sequence space of the starting pools. Considering 12 nt long oligomers, the full sequence space for DNA ($z=4$) is given by 16777216 sequences. Limiting the alphabet to a binary case (AT only) leads to a drastic reduction to only 4096 possible sequences. This number is further reduced to 64 by introducing reduced complexity sets (RCS) with double bases.

However, 4096 sequences are still a challenging number of sequences, especially when attempting to compare experimental results with detailed numerical simulations. Thus, a way to even further reduce the sequence space is to introduce the concept of 'virtual bases', i.e.,

multiples of the same nucleobase in a 12 nt long sequence. Doing so leads to several possible reduced complexity systems (RCS). If the virtual base consists of 2 actual nucleobases, $2^6 = 64$ different sequences can be obtained. Similarly, in the case of 3, 4, or 6 equal bases, the sequence space would account for $2^4 = 16$, $2^3 = 8$, or $2^2 = 4$ sequences, respectively. An example of a reduction in the case of double bases is the sequence AATTTTAATTTT in 'real space', which would correspond to ATTATT in 'virtual space'.

Whether the system can overcome the monomer-dominated state is determined by several parameters. First, for a given overall pool concentration, e.g., 10 μM , different sizes of the sequence space lead to different concentrations per strand. For 12mer pools with either a binary (AT only, $z=2$) or the full four-letter ($z=4$) alphabet, the concentration per sequence would then amount to 2.44 nM and 0.000596 nM, respectively. Similarly, this influences the timescales of hybridization, i.e., the time for two (partially) complementary sequences to find each other (see Figure 3.4). The timescale is influenced by a standard on-rate, the total concentration of the oligomer pool, and the sequence space and can then be estimated as follows:

$$t = \frac{z^l}{k_{\text{on}}^{\text{hyb}} \cdot c_{\text{total}}} \quad (3.1)$$

The standard on-rate of hybridization is assumed in the literature to be $1 (\mu\text{M s})^{-1}$ [111, 169–172]. Given a total pool concentration of 10 μM , the time a 12mer needs to find its fully complementary sequence in the random binary pool would take around 410 s. This time drastically increases to 19 days for the full four-letter random pool.

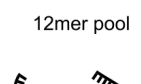


12mer pool	6nt overlap	12nt overlap
		
A A T T T T	$t = (k_{\text{on}}^{\text{hyb}} * 10 \mu\text{M})^{-1} * 2^3 = 0.8 \text{ s}$	$t = (k_{\text{on}}^{\text{hyb}} * 10 \mu\text{M})^{-1} * 2^6 = 6 \text{ s}$
A T	$t = (k_{\text{on}}^{\text{hyb}} * 10 \mu\text{M})^{-1} * 2^6 = 6 \text{ s}$	$t = (k_{\text{on}}^{\text{hyb}} * 10 \mu\text{M})^{-1} * 2^{12} = 410 \text{ s}$
A T C G	$t = (k_{\text{on}}^{\text{hyb}} * 10 \mu\text{M})^{-1} * 4^6 = 410 \text{ s}$	$t = (k_{\text{on}}^{\text{hyb}} * 10 \mu\text{M})^{-1} * 4^{12} = 19 \text{ days}$

Figure 3.4: Estimation of hybridization timescales for differently sized sequence spaces: For the same overlap length of either 6 nt or 12 nt, the estimated times for a complex to form vary drastically for pools with a binary or four-letter alphabet. Finding the fully complementary partner would take about 410 s for the binary random pool (2^{12}), compared to 19 days for the full four-letter random pool (4^{12}). However, for ligation reactions, the first collision (one fragment bound on a template) ideally is 6 nt, reducing the times to 6 s and 410 s for the binary and the four-letter pools, respectively. All calculations assume an on-rate k_{on} of $1 (\mu\text{M s})^{-1}$ following [111].

However, rather than the fully complementary sequence, only a partially complementary sequence with a binding site of 6 nt is needed to initiate the ligation step. Instead of only one

possibility, the number of sequences for potential binding partners increases to 64 in the case of the binary alphabet. This results in an approximate timescale of 6 s for two strands to find each other. Of course, the real case is significantly more complicated, as neither the binding position of the 6 nt on the template nor the complexity due to a three-body collision are considered in this calculation. Experiments have shown that ATr pools only display longer oligomers for a ligation time of $t_{\text{lig}} > 30$ s, whereas no product strands emerge for $t_{\text{lig}} = 10$ s. Still, these calculations provide an estimate for the ligation time needed in order to transition out of the monomer-dominated regime. Equivalently, the timescales for a sequence of the RCS to find its binding partner are even further reduced to 6 s for the fully complementary case and 0.8 s for a partial overlap of 6 nt.

These estimations provide a guideline for the experimental setup for the ligation time needed to obtain an elongated pool. The next sections show that by varying the conditions reported for the ATr pool [168], longer products can indeed emerge for ATGC random pools. The influence of the sequence space on the ligation behavior is further investigated by employing the reduced complexity pools.

3.2.2 ATGC random pools

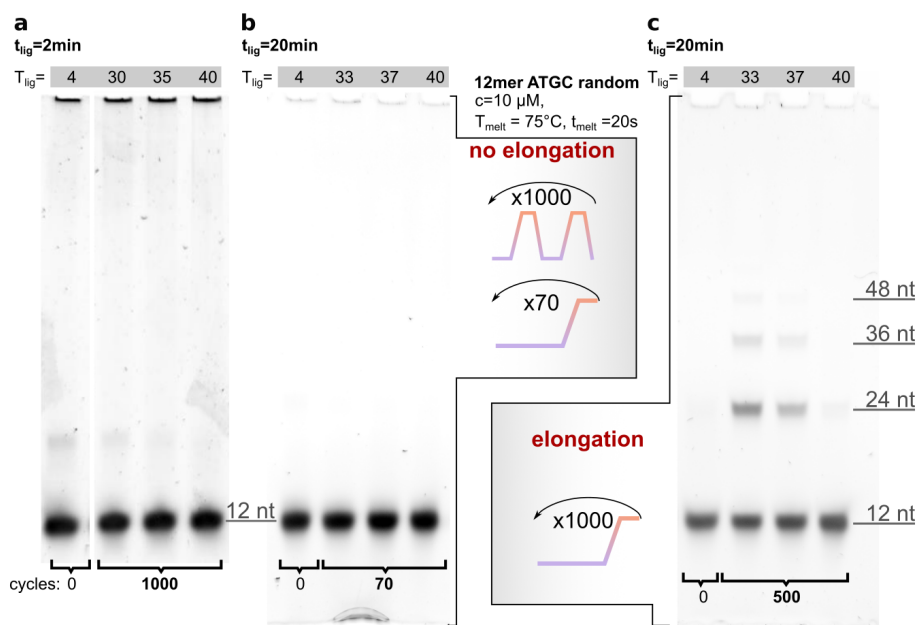


Figure 3.5: Elongation of 12mer ATGCr pools: **a** Keeping the same reaction parameters as for the elongation of the 12mer ATr pool does not lead to any elongation of the ATGCr pool. **b** Increasing t_{lig} to 20 min and running the reaction for a total of 70 cycles (approximately 1 day) still does not yield any elongated product. **c** With an increased t_{lig} of 20 min and a total runtime of 500 cycles (approximately 7 days), the ATGCr pool was able to elongate.

Maintaining identical reaction conditions as for the AT random starting pool while expanding the sequence space to the full ATGC random starting pool yielded no product formation for ligation temperatures between 25 °C to 50 °C (see Figure 3.5 a and reference [111]). This

outcome aligns with the predictions made through the calculations to estimate the hybridization timescales, as the estimated lower bound for the time it takes to form a ligation complex is 410 s, which is larger than the ligation time of 120 s. Anticipating, based on the calculations, that a longer hybridization time is needed for the ligation complexes to form, the ligation time t_{lig} was increased to 20 min. However, running the experiment for 70 cycles (approximately 1 day) still did not yield any product (see Figure 3.5 b), suggesting that a longer overall running time was needed. Indeed, increasing the running time to 500 cycles (approximately 7 days) led to the formation of longer oligomers (see Figure 3.5 c).

As the ligation reaction depends on the total concentration of the pool, concentrations of 5 μM and 20 μM were screened for the ATGC random pool. After two days of reaction time (≈ 140 cycles), only the 20 μM exhibit products, whereas the 5 μM pool does not yield any product. However, after 6 days, also the 5 μM pool shows products, suggesting that the overall concentration is high enough to find hybridization partners within the given ligation time t_{lig} of 10 min, 15 min, and 20 min.

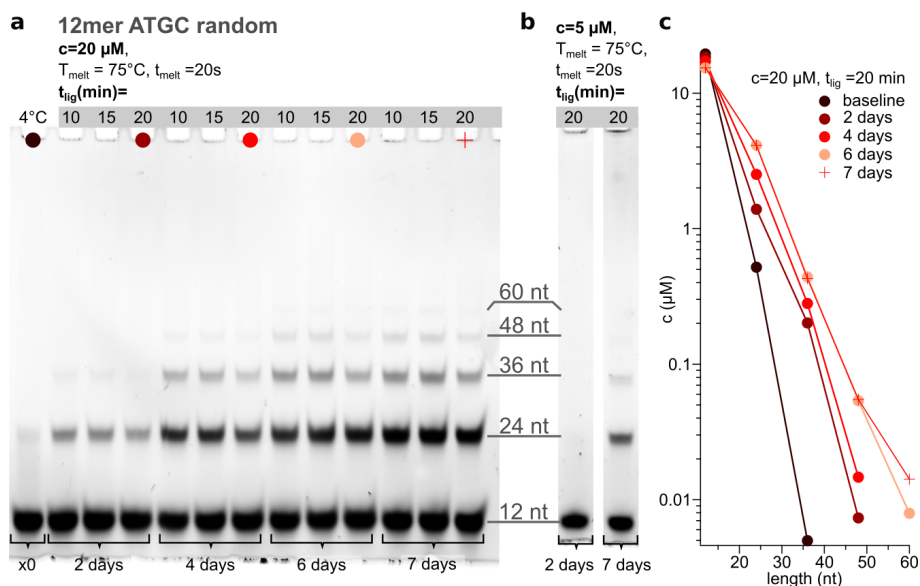


Figure 3.6: Concentration and reaction time screen for ATGCr pool: **a** PAGE analysis of the 20 μM ATGC random pool shows up to 60mers for all three ligation times (10 min, 15 min, 20 min) after 7 days of thermal cycling. The first 24 nt long product strands are already visible after 2 days. **b** The 5 μM ATGC random pool shows no elongation for total reaction times of 2 days. However, products emerge after 7 days of thermal cycling. **c** The concentration vs. length distribution of the 20 μM pool ($t_{\text{lig}}=20 \mu\text{M}$) reveals a non-monotonous length distribution for longer reaction times as the concentration builds up at the dimer level (shoulder).

These findings demonstrate the potential of the fully random ATGC pool to transition out of the monomer-dominated regime, given the right reaction parameters. However, sequencing remains challenging, as the current approach for distinguishing accurate reads from background noise relies on the unique characteristic of sequences comprising a binary AT alpha-

bet (see section 3.4.7 and [168]). Notably, PAGE analysis of the elongated ATGC pool reveals merged or shoulder peaks in the product spectra. This suggests that the product band of a specific length is divided into sub-bands enriched in subsets of bases due to the distinct running patterns of sequences of equal length, enriched in specific bases or evenly distributed in base composition. Similar to the division into A- and T-type of the product peaks in the AT random pool, this indicates a non-random distribution of product sequences in the ATGC random pool. However, further insight into the sequence space of the products is needed to confirm this hypothesis.

3.2.3 Reduced complexity system

Several reduced complexity sets are possible for 12 nt long strands, i.e. RCS-x4 with a virtual base length of 6 nt, RCS-x8 with a virtual base length of 4 nt, RCS-x16 with a virtual base length of 3 nt, and RCS-x64 with a virtual base length of 2 nt. The advantage of the reduced complexity sets is the possibility for directly comparing elongation behavior between experiment and detailed simulation (based on the simulation described in [98]) while keeping experimental and computational time at a reasonable scale. In this and the next subsection, several parameters for the different reduced complexity sets are screened to provide detailed experimental data to verify the computational model³.

Due to the ambiguous nature of the reduced complexity system, it is important to specify the employed nomenclature:

- **Pools** The reduced complexity pools are always denoted as 'RCS' followed by the 'xYY', where 'YY' is the respective size of the sequence pool, i.e. 4, 8, 16, or 64.
- **Length distribution** The monomers of the experiment are still the 12 nt long ligation substrates. Thus, the gels are labeled with the respective ('real') lengths. However, different levels of ligation are often described as 'xN', i.e., x1 for monomers, x2 for dimers, etc.
- **Sequences and Illumina sequencing** Sequences are always written in 'virtual space' for better clarity. Similarly, all sequence analysis is shown for the reduced case. Again, this course gains the information to the important features and later on allows for an easier comparison with simulated results.

As the reduced complexity sets have significantly smaller sequence spaces than the AT random pools, the concentration per strand is naturally higher. Thus, compared to the AT random and ATGC random pools, elongation products are expected even for 'harsher' reaction conditions such as short ligation times or extreme temperatures. Figure 3.7 shows the length distributions for all RCS systems with total concentrations of 10 μ M, 5 μ M, and 2.5 μ M for various T_{lig} (30 °C, 40 °C and 50 °C). The reactions were run for 200 cycles. All remaining parameters were kept similar to the AT random reaction conditions ($t_{\text{lig}}=120$ s, $t_{\text{melt}}=20$ s, $T_{\text{melt}}=75$ °C). The emergence and amount of product strands are directly connected to the ligation complexes' occurrence, concentration, and half-time. To this end, T_{lig} is a crucial parameter that directly influences the emergence of ligation complexes, as the hybridization

³The computational model is based on [98] and is currently expanded by the Gerland group to include explicit sequence and folding information.

of oligomers depends on the binding energy. Thus, the higher the temperature of the system is, the higher the number of hybridized bases that are needed to form stable complexes.

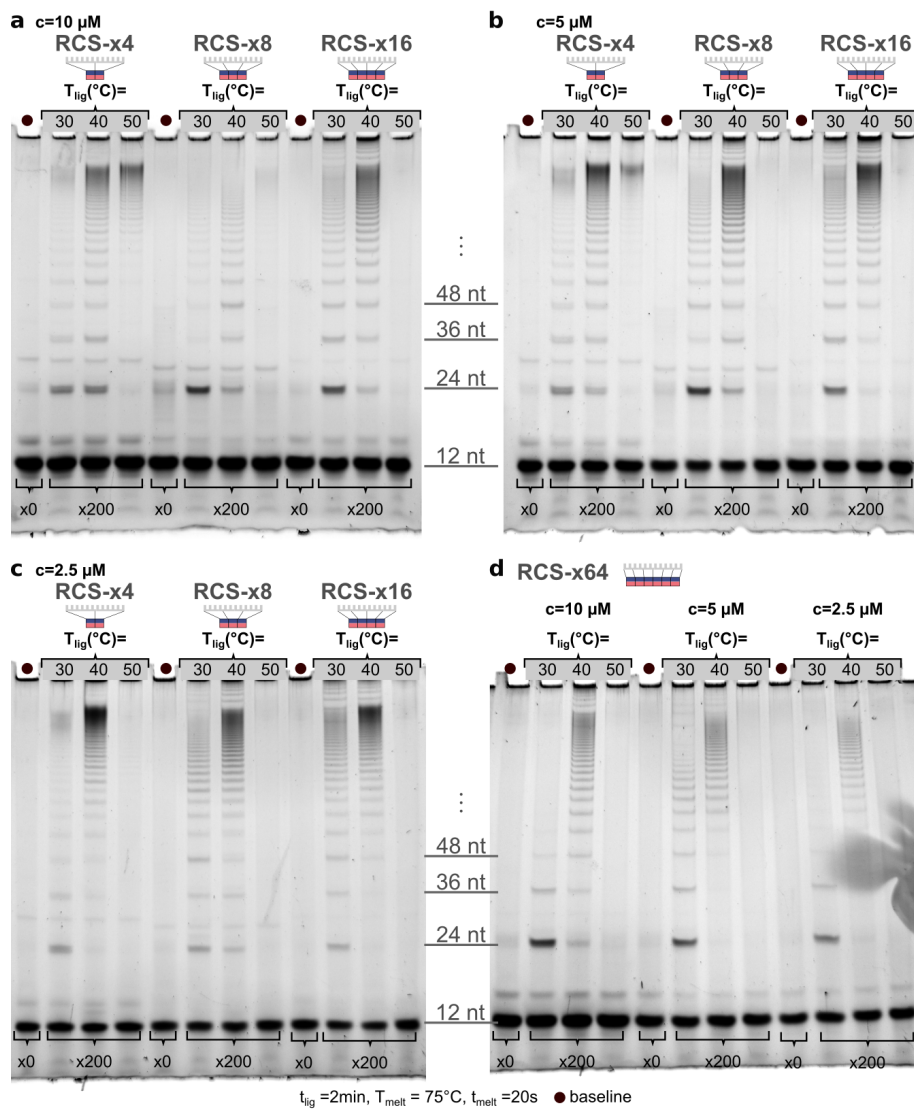


Figure 3.7: Ligation temperature and concentration screen for the reduced complexity systems: PAGE analysis for RCS-x4, RCS-x8 and RCS-x16 with a total concentration of **a** 10 μM , **b** 5 μM , **c** 2.5 μM . **d** PAGE analysis for RCS-x64 with total pool concentrations of 10 μM , 5 μM and 2.5 μM .

Products are visible for all reduced complexity systems for ligation temperatures of 30 °C and 40 °C. When run at 30 °C, the pools generally exhibit shorter products than at 40 °C. This is due to the higher dissociation probability of short double-stranded complexes at 40 °C, which in turn frees them to act as ligation substrates for longer products. As 50 °C is above the melting temperature (see section 3.4.4), no products are expected. However, if the overall concentration is high enough, longer products emerge at $T_{\text{lig}}=50$ °C for the RCS-x4 and RCS-x8 pools. In that case, the concentration per strand is high enough such that some ligation

complexes are formed and ligated with low probability. Due to the high ligation temperature, those complexes quickly dissociate. Similar to the 40 °C case, they are then free to produce even longer product strands.

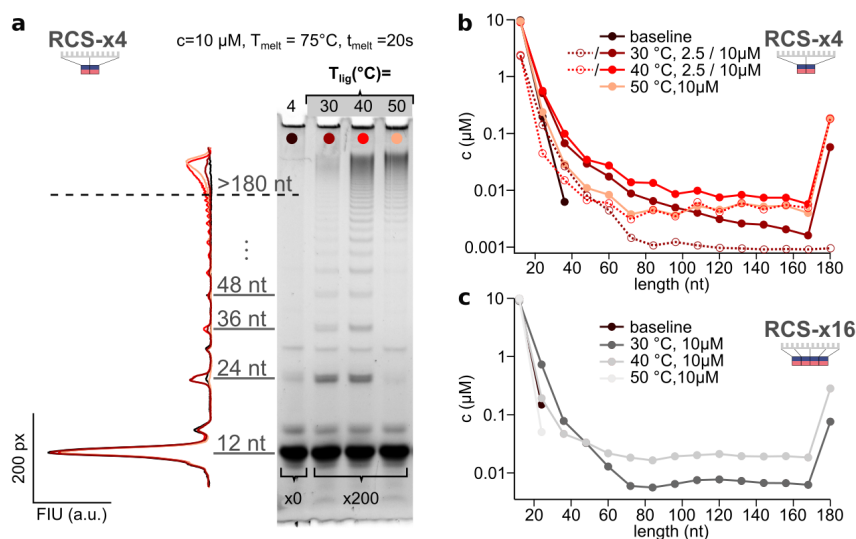


Figure 3.8: Length distributions for ligation temperature and concentration screen: **a** The exemplary fluorescence trace from PAGE analysis shows clear peaks for short product strands but a blurred product band for long oligomers. The dashed line indicates this transition (in this case 180 nt). For oligomers of shorter lengths, the peaks are integrated individually. All longer products are integrated together, leading to an increased concentration for the last peak. **b** The RCS-x4 shows products for all temperatures at a pool concentration of 10 μM . For low T_{lig} (30 °C), short to intermediate lengths, whereas for high T_{lig} (50 °C), long products dominate the distribution. Intermediate temperatures (40 °C) show the highest concentration of products for the complete length range. Reducing the pool concentration to 2.5 μM shifts the length distribution towards the high-temperature behavior with a lower concentration of short products already at 40 °C. **c** For larger sequence spaces (here RCS-x16), the temperature dependence shows the same trend. However, high temperatures show no product bands.

The concentration vs. length distributions are obtained from the integration of the peaks visible in the fluorescence trace (see Figure 3.8 a). For traces that exhibit merged product strands of long product strands, this region is integrated as a single peak. This leads to an increased concentration for the last length in the length distributions but needs to be included for normalization. The dependence on the ligation temperature is exemplary shown for the RCS-x4 system in Figure 3.8 b. The length distribution shows the highest overall concentration of products for $T_{\text{lig}}=40$ °C. Note the difference of the distributions for 30 °C and 50 °C, with 30 °C showing a significantly higher amount of (short) products than 50 °C. However, above 108 nt this behavior is inverted. The other reduced complexity systems show a similar dependence. For the RCS-x16 at a pool concentration of 10 μM , for example, this transition takes place already for $T_{\text{lig}}=30$ °C and 40 °C (see Figure 3.8 c). Below 48 nt, higher product concentrations are detected for $T_{\text{lig}}=30$ °C, whereas $T_{\text{lig}}=40$ °C shows higher product concen-

trations for products longer than 48 nt. $T_{\text{lig}}=50\text{ }^{\circ}\text{C}$ shows no products. All remaining length distributions are shown in the appendix (see Figure 3.19).

The overall pool concentration also influences the length distribution. Following equation 3.1, lower overall pool concentrations lead, on average, to longer times for complementary strands to find each other. Thus, for low pool concentrations and high ligation temperatures, the probability of complementary strands forming ligation complexes is too low, while the dissociation probability of double strands is high. Comparing the length distributions for the RCS-x4 system at a pool concentration of $10\text{ }\mu\text{M}$ and $2.5\text{ }\mu\text{M}$ shows that the elongation behavior of the lower concentrated pool at $40\text{ }^{\circ}\text{C}$ is similar to the $50\text{ }^{\circ}\text{C}$ higher concentrated pool showing a long-tailed distribution. For $30\text{ }^{\circ}\text{C}$, in contrast, $50\text{ }^{\circ}\text{C}$ exhibits no products for the $2.5\text{ }\mu\text{M}$ pool. Again, $30\text{ }^{\circ}\text{C}$ show higher concentrations for short products (below 48 nt) compared to the $40\text{ }^{\circ}\text{C}$.

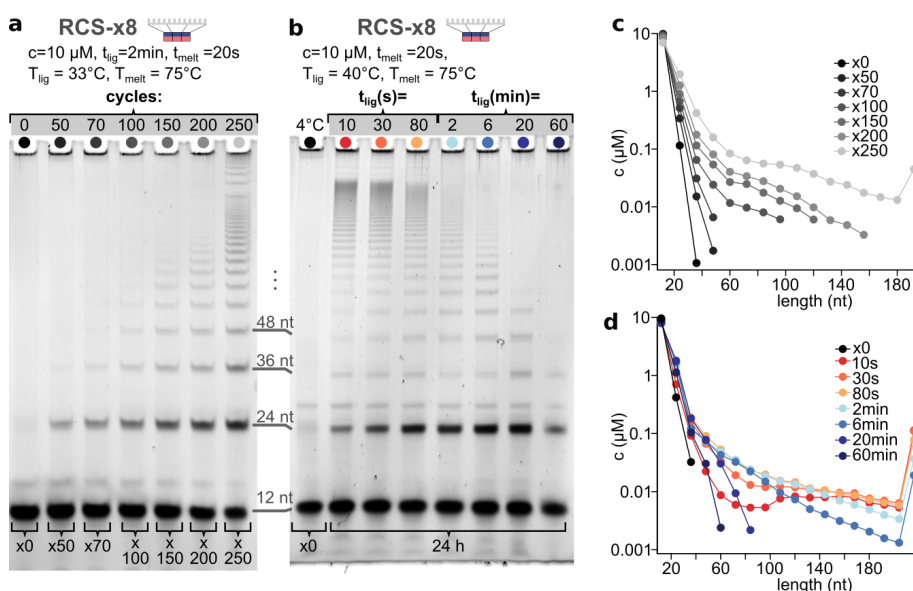


Figure 3.9: Cycle number and t_{lig} screen for RCS-x8: **a** The PAGE analysis and **c** length distribution of the cycle number screen shows that dimer (24 nt) (and to low extend also trimers (36 nt)) emerge already after 50 cycles. With an increasing number of cycles, the maximum length of product strands increases. **b** The PAGE analysis and **d** the length distribution confirms that a ligation time of 10 s is already sufficient for products to emerge.

As already discussed for the ATGC random pool, two additional parameters that critically influence the elongation behavior are the number of cycles as well as the ligation time. Generally, a higher cycle number leads to longer products (see Figure 3.9 a). After 50 cycles, all reduced complexity systems show already dimers (24 nt) that gradually lead to longer oligomers resulting in a more long-tailed distribution with a flatter slope (see Figure 3.9 c). Additionally, a higher number of temperature cycles leads to a steady increase in concentration. If the overall reaction time stays constant, the ligation time determines how many temperature cycles are performed. As long as t_{lig} is sufficiently long for strands to find their complement in a given pool, this results in long-tailed length distributions for short t_{lig} and

short-tailed distributions for long t_{lig} . Figure 3.9 b shows the resulting length distributions for a 24 h runtime with different t_{lig} ranging from 10 s to 60 min for the RCS-x8 system with a pool concentration of 10 μM . All screened ligation times lead to products with $t_{\text{lig}}=10$ s showing the highest amount of long products. This is the case for all reduced complexity systems (see Figure 3.20) and was to be expected, as 10 s are significantly higher than the 0.8 s estimated for the hybridization in the RCS-x64 pool (compare Figure 3.4), which would decrease even further for smaller sequence spaces. Interestingly, the length distributions with a short t_{lig} (10 s, 30 s, and 80 s) show a concentration dip for intermediate lengths, whereas a longer t_{lig} (2 min, 6 min, 20 min, and 60 min) leads an over-representation of dimers (see Figure 3.9 d).

3.2.4 Short sequence subsets drive elongation behavior

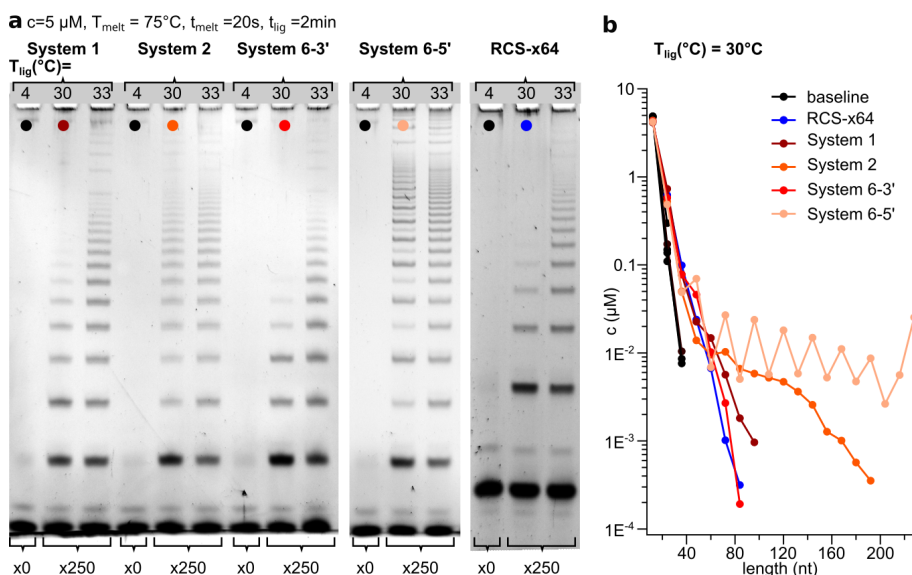


Figure 3.10: RCS-x64 spiked with sequence subsets: a PAGE analysis for the RCS-x64 system and the RCS-system spiked with the respective sequence subsets shows a clear variation in the intensity and length of the emerging products for a ligation temperature of $T_{\text{lig}}=30$ °C, whereas all systems show a long-tailed length distribution similar to the unspiked case for $T_{\text{lig}}=33$ °C.

For the AT random system, Illumina sequencing revealed that small sequence biases in the starting pool (like an overall A or T bias) get amplified by the templated ligation and are enriched in sequences of the products. Thus, the question arises whether small fluctuations of the occurrence of specific sequences instead of averaged sequence features (like an enrichment in A or T) could influence the elongation behavior. To this end, the RCS-x64 was spiked with different subsets of sequences at a constant overall concentration. The RCS-x64 was chosen to ensure a large enough sequence space to allow variability while keeping a high signal-to-noise ratio (compared to the AT random case). The following sequence subsets were chosen as spiking systems. Note that in this case, double bases (AA or TT) are denoted as one base (A or T) for clarity.

System 1	System 2
AAA TAT	AAA AAA
TTT ATA	TTT TTT
AAA ATA	AAA TTT
TTT TAT	TTT AAA
System 6-5'	System 6-3'
ATA AAT	ATA AAT
AAA TAT	AAA TAT
TAA TAT	TAA TAT
ATT TAT	TAT TTA
ATA TTT	TTT ATA
ATA TTA	ATT ATA

System 1 consists of four sequences that are poly A or poly T for the first half and ATA or TAT for the second half of the oligomer. The sequences were chosen to avoid hairpins and to potentially allow them to act as templates. System 2 also consists of four sequence strands, i.e., a completely poly A and poly T sequence as well as the symmetric AAATTT and TTTAAA sequences. The sequences of system 2 resemble the RCS-x4 system and have already been shown to efficiently elongate. Both, system 6-5' and system 6-3' consist of 6 strands. The sequences of system 6-5' were chosen to match the six most common sequences in the de Bruijn graph⁴ of the 84mer (x8) oligomers resulting from sequencing data of the RCS-x64 system [168]. Spiking the RCS-x64 pool with the most common sequences of its elongated products potentially improves the elongation behavior to be more robust at 'harsher' conditions. Interestingly, the sequences (and their reverse complements) have no straightforward templating mechanism, i.e., one sequence acting as template for three (double-)bases of each ligation fragment. Thus, this could also lead to inhibitory effects. To probe the influence of this simple templating mechanism, the direction of the 4th - 6th strand of system 6-5' was reversed in system 6-3', e.g., ATTTAT becomes TATTTA. This simple change allows templating of three (double-)bases of each ligation fragment, one example being TTTATA acting as template for AAATAT and TAATAT.

For the spiking experiments, the sequences of the systems were added to the pool with a concentration of $5 \mu\text{M}/(64 \cdot 2)$ each, which is twice the concentration any sequence would have in the equally concentrated case. The overall pool concentration was kept at $5 \mu\text{M}$. The reactions were cycled 250 times between the temperatures of $30 \text{ }^\circ\text{C}$ and $75 \text{ }^\circ\text{C}$, as well as $33 \text{ }^\circ\text{C}$ and $75 \text{ }^\circ\text{C}$ to probe short-tailed as well as long-tailed elongation conditions for the RCS-64 system. For a ligation temperature of $33 \text{ }^\circ\text{C}$, all reactions showed a long-tailed length distribution with only minor differences between the spiked systems (see Figure 3.10a). However, for the sub-optimal elongation conditions with a ligation temperature of $30 \text{ }^\circ\text{C}$, significant differences in the length distributions arise (see Figure 3.10 b). System 1 and system 6-3' show a short-tailed distribution similar to the unspiked RCS-x64 system. However, system 2 and system 6-5' each exhibit a distinct long-tailed length distribution. Interestingly, running the systems at the same concentration as in the pool experiments shows only minor product

⁴The de Bruijn graph is a tool to display networks. In this specific case, it is used to visualize the most common sub-sequences (12 nt) within the sequencing data of all 84mers as nodes. The connections between the nodes represent if and how often two sub-sequences are found in direct succession within the 84mer.

bands, although with similar characteristics of the length distribution, suggesting an interplay between the sequences of the RCS-x64 pool and the systems rather than a superposition of both elongation behaviors (see Figure 3.11).

The efficient elongation even at low ligation temperatures with the system 2 spiked pool is partly unsurprising as enriching the pool in poly A and poly T sequences allows for reciprocal elongation of both due to the suppression of hairpins in homogeneous sequences. Additionally, the results from the elongation behavior of RCS-x4, which consists of the same four sequences, already showed efficient elongation behavior of these four sequences under various conditions. As the sequences of system 6-5' were the most abundant sub-sequences of the x8 oligomers of the unspiked RCS-x64 pool, spiking the pool with these sub-sequences was expected to yield at least x8 oligomers. However, spiking leads to even longer products with higher concentrations than system 2. Interestingly, the length distribution shows an alternating high (x2, x4, x6...) and low (x3, x5, x7...) product concentration.

The differences observed in the length distributions for different spiked systems suggest that by introducing a concentration bias of a subset of sequences, different elongation behaviors can be induced. As the elongation characteristics of the total pool become similar to the elongation behavior of the spiked sequence subsets, it is likely that the total pool inherits the elongation information included in the network of the sequence subset. Whether this results in an imprinting of the information of the starting pool in the sequence information of the products can not be concluded from the length distributions alone. To this end, the three samples (RCS-x64, RCS-x64 with system 2, and RCS-x64 with system 6-5' at 30 °C) were chosen for further analysis and sequenced using Illumina sequencing. Investigating the fea-

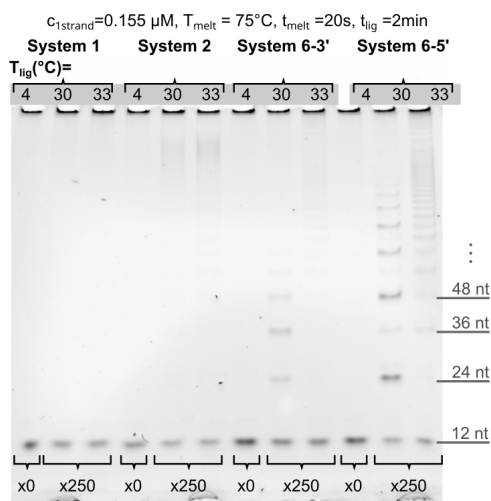


Figure 3.11: Length distribution of the systems: Running only the sequence subsets of the four different systems shows differences in the length distribution but with minor product formation. For system 1, no product can be detected for either of the temperatures. System 2 shows long product bands but no intermediate bands for both temperatures. System 6-3' and 6-5' both have a long-tailed distribution for $t_{lig}=33$ °C. For $t_{lig}=30$ °C, System 6-3' exhibits a short-tailed distribution, whereas system 6-5' shows longer products.

tures of the sequence space, especially of the product bands, allows to gain further insights into the elongation behavior and discuss differences in the product sequence space of the spiked systems. Details of the sequencing protocol and post-processing are given in section 3.4.7).

A straightforward observable for the sequence space of a pool with a binary alphabet is the A:T content. For a completely random pool, the A:T content of all sequences follows a binomial distribution. However, as previously reported for AT random 12mer pools [111], the sequencing results of the RCS-x64 pools (spiked and unspiked) showed a transition from a binomial to a bimodal distribution at the first ligation step, the x2 level (see Figure 3.12 a-c (left)). As for the AT random case, this transition is likely due to a suppression of hairpins as the replication mechanism of the system self-selects for sequences that can act as templates for the reaction. Consequently, this leads to a separation of ligated sequences into populations that are enriched in either one of the two bases, thus resulting in A-type or T-type sequences. Interestingly, the exact shape and position of the bimodal distributions vary for the different pools. The A-type fraction shows that the unspiked RCS-x64 pool is strongly T-biased. Both, the binomial distribution, as well as all ratios of the peak heights of the bimodal distribution, exhibit this bias. The maxima of the bimodal distribution are at ≈ 0.27 and ≈ 0.71 for all product lengths.

In contrast, the bimodal distributions for both spiked systems are more evenly distributed. However, the details of the distributions are very different. Spiking the pool with system 2 leads to a transition from T-biased to A-biased sequences. The maxima of the bimodal distribution at the x2 level are at ≈ 0.29 and ≈ 0.73 but move to ≈ 0.26 and ≈ 0.78 at the x4 and higher levels. As the spiked sequences include poly A and poly T sequences, which are unable to form hairpins and, therefore, efficiently participate in the ligation reaction, it is unsurprising that the maxima in the A-type fractions of the elongated sequences are further apart compared to the unspiked case. For system 6-5', the peaks of the bimodal distribution instead of a transition from T-type to A-type dominated product lengths, all products are T-biased but significantly less than for the RCS-x64 pool. Interestingly, the maxima of the bimodal distribution are significantly closer together and move from ≈ 0.38 and ≈ 0.69 at the x2 level to ≈ 0.36 and ≈ 0.65 at the x4 and higher levels. Additionally, the distribution of the monomers (x1) differs significantly from a binomial. Sequences of equal (0.5) A:T content are enriched, whereas sequences corresponding to the maxima of the bimodal distribution are depleted.

The next step is to look closer at the base-by-base sequence composition of different ligation levels. The base probabilities (see Figure 3.12 a-c (right)) show the relative probability of an A (red) or T (blue) to occur at a specific position when averaged over all sequences of the same length. The overall T-bias of the RCS-x64 pool is clearly visible by the overall blueish color (see Figure 3.12 a (right)). However, certain patterns can be detected when looking more closely. First of all, the ligation junctions (dashed lines) have a higher probability of an ...TA... pattern directly at the ligation site, with the exception of the first one, where the patterns differ. Additionally, x4, x6, and x8, as well as x3, x5, and x7 show similar patterns of their sequence probabilities. Interestingly, a similar but even more pronounced grouping of lengths is visible for the RCS-x64 spiked with system 6-5' (see Figure 3.12 c (right)). Again, uneven and even ligation levels form two different groups. The even ligation levels start with

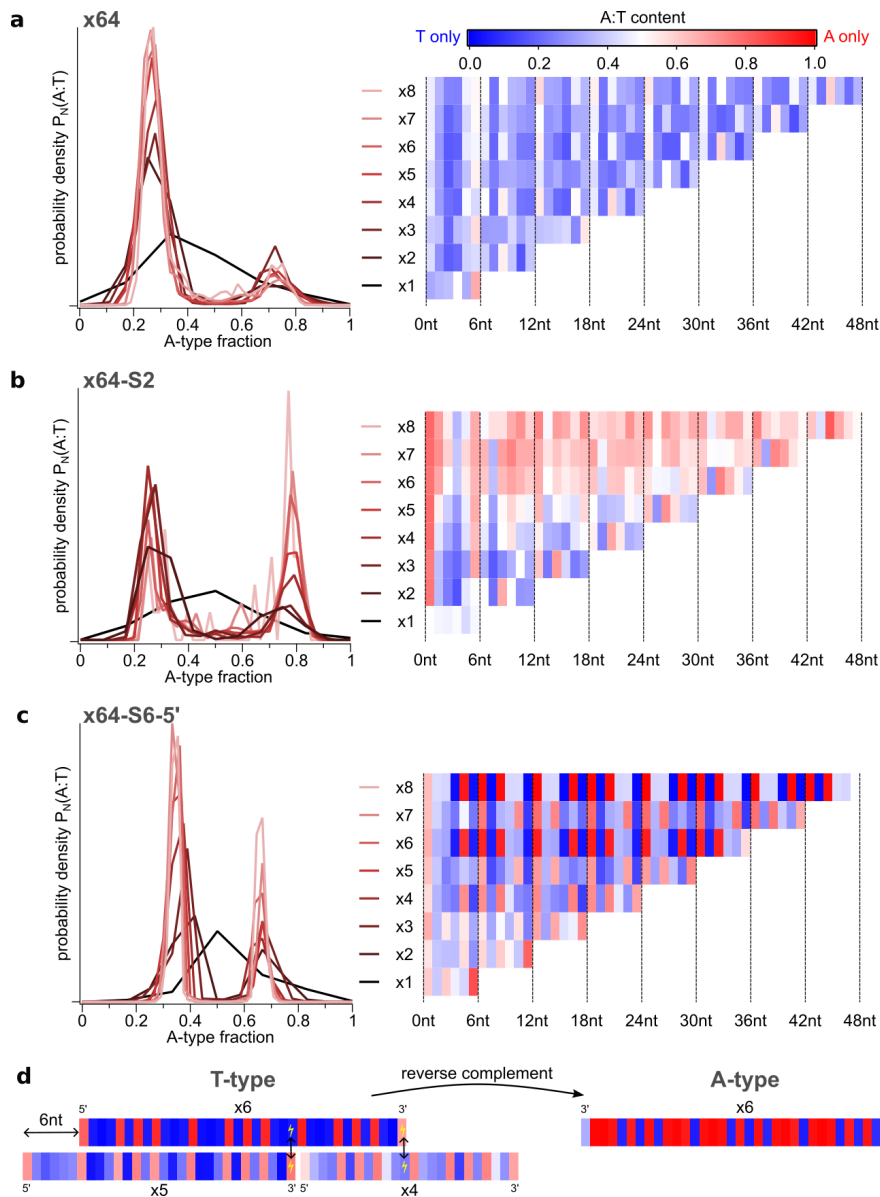


Figure 3.12: A-type vs. T-type and base probabilities for the spiked and unspiked RCS-x64 pool: All pools show a transition from a binomial distribution for the monomers to bimodal distributions for the A-type fraction (left). The base probability gives insights into sequence sub-patterns, as well as compositional biases (right). **a** The A-type fraction of the RCS-x64 is heavily T-biased. The base probability also reflects the overall T bias. **b** The A-type fraction is T-biased for short lengths and A-biased for longer lengths with a transition from x5 to x6 level. **c** The A-type fraction of the spiked system 6-5' is again slightly T-biased but much more evenly distributed than for the RCS-x64 pool. The base probability shows close resemblances of the sequence sub-patterns of the x4, x6, and x8 levels, as well as the x5 and x7 levels. **d** Proposed elongation mechanism for the spiked system 6-5'. The A- and T-types are perfect complements of each other.

the sequence ATT TAT followed by ATA TTT, which are both part of the spiked system. These two sequences are then repeated until the last sub-sequence ATA TTA, which is also part of the spike. Thus, the overall ligation pattern has a periodicity of 12 nt (two sub-sequences) with a defined end sub-sequence. Note that the last sub-sequence only differs by one nucleotide (A vs. T) at the last position from the original pattern. Looking separately at the base probabilities of A- and T-type strands shows that A-type strands of certain lengths are the perfect complements of T-type strands (see Figure 3.12 and 3.22). The two most common sub-sequences are AAA TAT and ATA AAT, which are again part of the spiked system. The most common last sub-sequence of the A-type sequences is either ATA AAT or ATA AAA. In the case of an ending T, the probability for T at this position is decreased in comparison to the middle sub-sequences. This sub-sequence is not present in the spike but is still over-represented in the base probabilities, suggesting a distinct function in the elongation behavior, i.e., as ending sub-sequence. Until here, the only unaccounted sub-sequence of the spiked system is TAA TAT. However, this sub-sequence is over-represented at the first sub-sequence position of the A-type sequences. Interestingly, when shifted by 6 nt, the uneven and even ligation levels show the same sequence patterns (see Figure 3.12 d).

To gain further insight into the correlation between different sub-sequences of two oligomers, the sample Pearson correlation coefficients (PCC) can be computed. To this end, x- and y-mers (x,y equal all oligomers with a specific length) are divided into the 6 nt long sub-sequences. Following section 3.4.8, the PCC can then be computed for the two sets of sub-sequences at a specified position and plotted as a correlation matrix (see Figure 3.13). High correlations correspond to light, while low correlations correspond to dark colors. Plotting the PCC matrix of x6 vs. x8 for the RCS-x64 and RCS-x64 spiked with system 6-5' shows a similar checkerboard-like correlation pattern (see Figure 3.13) confirming the 12 nt periodicity. The PCC matrix for x7 vs. x8 of the unspiked RCS-x64 pool also shows a checkerboard-like pattern. However, the overall contrast between correlated and uncorrelated sub-sequences is lower. Additionally, a phase shift of the checkerboard pattern is detected, resulting from a deviation in the x7 oligomers. Similar phase shifts are also detected for the x5 oligomers.

Unsurprisingly, the contrast between correlated and uncorrelated sequences is stronger for the spiked case, as the base probability plots already revealed more pronounced sequence patterns. Specifically, the separation of the three different correlation domains is more dominant in the spiked case. Figure 3.13 c shows a high correlation of both first, as well as both last sub-sequences with one another, which are significantly lower correlated with all other sub-sequences as visible by the darker outside bands. The correlation between the x7 and x8 is overall lower. Additionally, as expected from the base probability plot, the correlation pattern is shifted by 6 nt, with a low correlation between the first sub-sequence of the x7 oligomers and all sub-sequences of the x8 oligomers (see 3.13 d). However, plotting the correlation between the reverse complements of A-type x7 and T-type x8 strands recovers the high contrast correlation pattern as observed for x6 vs. x8 PCC matrix. This points to an elongator-type elongation mechanism. In this mechanism, an A-type strand would bind to a T-type strand with an overhang. A shorter oligomer can then bind and be ligated, again leading to an overhang. This type of elongation efficiently leads to long product strands as it does not rely on strand separation and can grow even when the melting temperature of the complex exceeds T_{melt} [98]. A similar correlation pattern and elongation mechanism has been observed for a specifically designed subset of 8 sequences as reported in [111], thus, fur-

ther supporting this conclusion. Notably, the last sub-sequences show a higher correlation for all x7 vs x8 compared to the reverse complements of A-type x7 and T-type x8. This is to be expected as the base probabilities revealed that the last sub-sequence of T-types resembles ATA TTA, whereas for A-types, it resembles ATA AAT.

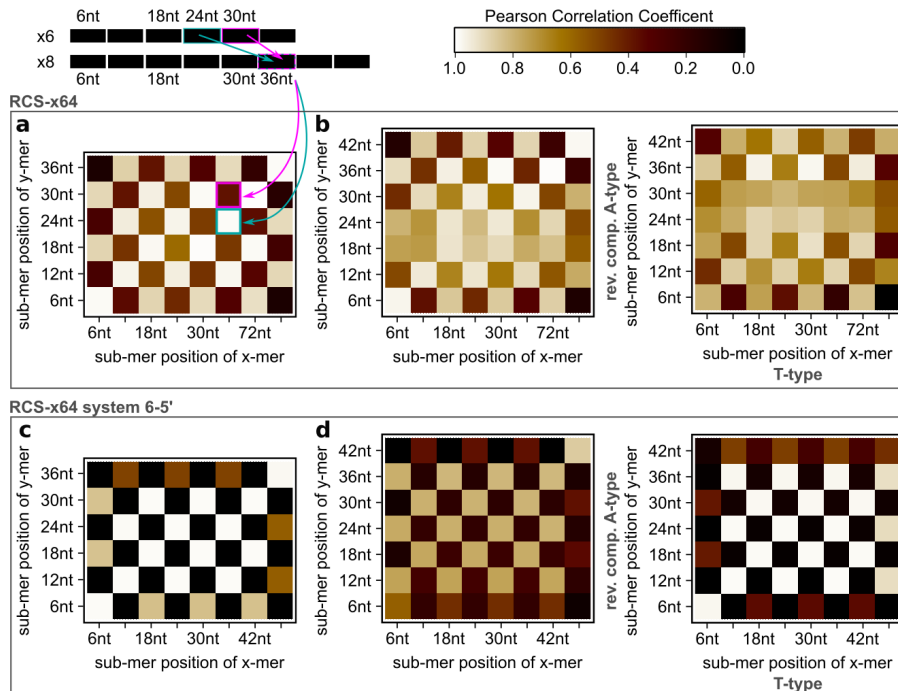


Figure 3.13: PCC matrix for the x64-RCS pool and x64-RCS pool with system 6-5': The PCC matrix plots show the correlations between each of the 6 nt long sub-sequences of two product strands. **a** The correlation matrix between the x6 and x8 products of the unspiked pool shows a checkerboard-like pattern with high correlations between alternating sub-sequences. **b** The correlation between the x7 and x8 product strands exhibits a similar checkerboard-like pattern, although with a phase shift. Separating the strands in A- and T-type populations shows no increased correlation. **c** The checkerboard-like pattern of the correlation matrix between the x6 and x8 product strands is enhanced further when RCS-x64 is spiked with system 6-5'. The first and last sub-sequences correlate significantly less with the intermediate sub-sequences. **d** The correlation strongly decreases between the x7 and x8 product strands. However, separating the strands in A- and T-type populations reveals a strong correlation between the reverse complexity strands of the x7 T-type and the x8 A-type strands suggesting that x7 strands act as templates for x8 strands.

In comparison, the base probabilities for the RCS-x64 spiked with system 2 display different sequence patterns (see Figure 3.12 b). First of all, the base probabilities show a transition from mainly bluish to reddish between the x5 and x6 oligomers. This corresponds to the transition from mainly T-type to mainly A-type strands dominating the product strands. Plotting the base probabilities for A- and T-type separately reveals a remarkable resemblance between the first and the last sub-sequences of each group (see Figure 3.23). The sub-sequence ATT

TTA dominates the first position of the T-types, whereas the sub-sequence AAA TA(A/T) dominates the first position of the A-types. At the last position, the sub-sequence ATA TTT dominates the T-types. Similarly, the sub-sequence ATA AAA dominates the A-types. Between those start and end-sequences, the sub-sequences in the middle consist mainly of poly-A and poly-T stretches for the A- and T-type, respectively.

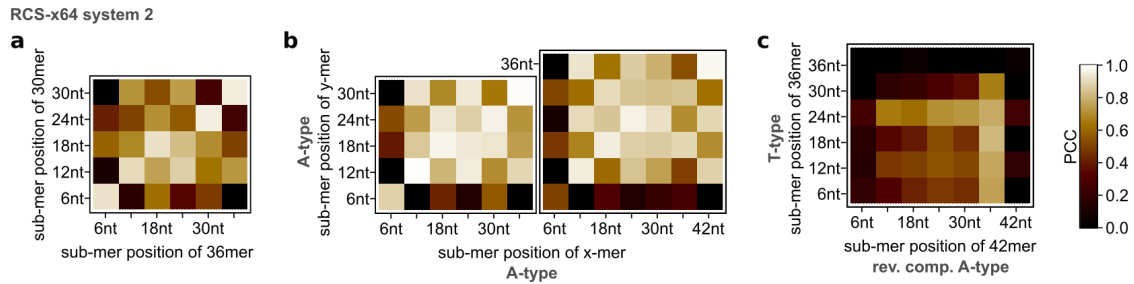


Figure 3.14: PCC matrix for the x64-RCS pool with system 2: **a** The PCC matrix for x5 and x6 show a significant correlation of the first and last sub-sequences. Overall, no clear correlation pattern can be deduced. **b** The PCC matrix for x5 and x6 of only the A-type strands again shows a significant correlation of the first sub-sequences. However, it is then followed by a separate, highly correlated part of similar sub-sequences. The PCC matrix for x6 and x7 shows a similar behavior but with a decreased correlation for the first sub-sequences. **c** No significant correlation can be detected for the x6 T-type strands and the reverse complement strands of x7 A-type strands.

The difference is also captured in the PPC matrix of the RCS-x64 spiked with system 2 (see Figure 3.14). Plotting the correlation matrix of x5 vs. x6 oligomers shows a high correlation between the first sub-sequences, which exhibit a significantly lower correlation to all other sub-sequences. However, no clear correlation pattern can be deduced. This changes when plotting the same matrix, but this time only for A-type sequences (see Figure 3.14 b). Here, the sub-sequences of oligomers of consecutive length show a high correlation throughout all sub-sequences with the exception of the first sub-sequences, which themselves correlate but strongly differ from the following sub-sequences. A similar behavior is found for the PPC matrix for the A-type of x6 vs x7 oligomers. Interestingly, the reverse complements of the A-type strands do not correlate with T-type strands (see Figure 3.14 c). Although both A- and T-types are strongly enriched in poly A and poly T stretches, the exact position and number of the poly stretches differ. Thus, the exact elongation mechanism of the pool spiked with system 2 can not be simply deduced from these results. A possible explanation could be that multiple competing sub-networks develop, each with its own elongation mechanism. Further investigation is needed, at best, in combination with a detailed simulation based on the code developed by Rosenberger *et al.* [98].

In summary, the results show a clear difference in the length distribution of pools spiked with different sequences, which translates into the sequence space. The base probabilities and PPC matrixes point to different elongation mechanisms for the random pools spiked with different subsets. Thus, spiking a random pool with different subsets of sequences not only leads to a compositional bias of the product pool but also results in a 'spiked' elongation mechanism. Notably, the base probabilities for both the initial and ending sub-sequences

exhibit comparable sequence patterns within all three investigated pools. This intriguing resemblance to the start and stop codons within the genetic code warrants further investigation.

3.3 Discussion and Outlook

The results show that the sequence space of the initial random or semi-random starting pool has an immense effect on the characteristics of the resulting pool. The sequence space dictates whether a pool can transition from a monomer-dominated into an autocatalytic regime of elongated sequences under the given thermal non-equilibrium conditions. However, not only does the sequence space influence the length distributions, but sequence biases of the starting pool can further lead to great variations of the sequence space of the elongated products. Increasing the concentration of a subset of sequences in a random background can dominate the elongation characteristics of the entire pool, thus, changing the sequence space characteristics of the product pool. The effect becomes most obvious when placing the system in unfavorable elongation conditions. For effective replication networks of the subsets of sequences, i.e., system 2 and system 6-5', the length distributions change from a short-tailed distribution for the unspiked to long-tailed distributions for the spiked pools. Illumina sequencing revealed that the product strands then contain the information of the spiked systems and inherent the elongation mechanisms of the spiked networks.

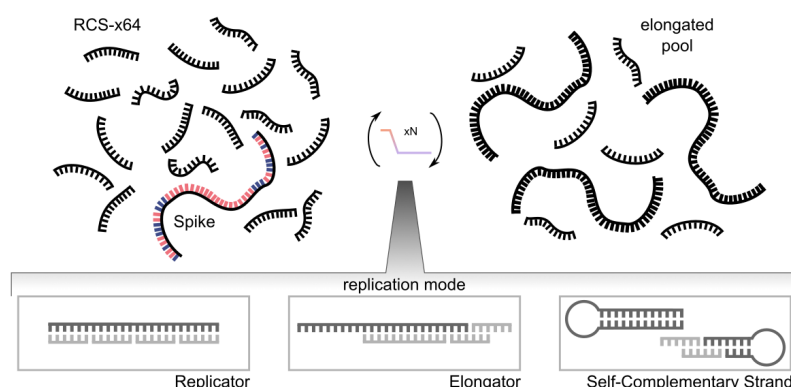


Figure 3.15: Potential elongation modes of random pools spiked with long sequences: Long oligomers could participate in the ligation reaction in multiple ways, depending on their sequences.

Similarly, one could ask whether the information contained in a long sequence can imprint some elongation behavior on the random pool. Or, from the point of view of the spike, are there certain regions of the spike where a ligation is more probable to take place, or could the spiked sequence even evolve towards a more efficient one? Generally, there are multiple ways a longer sequence could participate in the elongation (see Figure 3.15). It could simply act as a template where monomer strands hybridize and get ligated ('replicator'). Another possible interaction could be an 'elongator' type of interaction. In this case, the spike hybridizes to another strand with an overhang, where a shorter oligomer can hybridize and be ligated to the spike. Similarly, a third possible interaction could involve self-templation of the spike by forming hairpins either with or without overhangs.

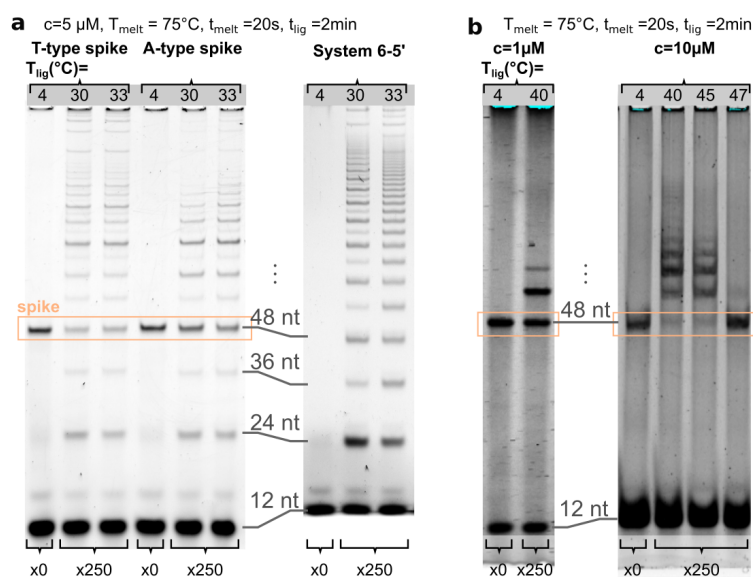


Figure 3.16: Long spiking sequences influence replication behavior: **a** Spiking the RCS-system with the prominent 48 nt long A- or T-type sequence from spiking experiment with system S6-5' leads to long-tailed distributions even for the 30 °C ligation temperature. **b** A long spike, that is able to form a hairpin induced a completely different elongation behavior.

In preliminary experiments, the RCS-x64 pool was spiked with 0.1 μM of the most common x4 A-type (or T-type) sequence from the RCS-x64 pool spiked with system 6-5'. Interestingly, both sequences recover the long-tailed length distribution at low ligation temperatures (30 °C) similar to the RCS-x64 pool spiked with system 6-5' (see Figure 3.16 a). A closer look reveals that even the alternating high and low concentrations of the product strands are similar to the elongation behavior when spiked with system 6-5'. This suggests that an analogous elongation behavior is induced by the long spike.

Likewise, a different spike, for example one that is able to fold on itself, would presumably result in a different elongation behavior. Indeed, first results show that hairpin formation of the spike can lead to an elongator-type of replication behavior (see Figure 3.16 b) by extending the spike sequence and, therefore, moving it to 'higher ligation levels'.

To understand the replication behavior of the pools and the influence of the spikes in more depth, current efforts try to match the experimental with an extended version of the numeric model developed by Rosenberger *et al.* [98]. In this new version, detailed sequence information is included in the model. Further, the simplification of DNA as stiff rods is lifted. Instead, a folding model is implemented that first calculates the full partition function and base pairing probabilities based on the base pairing energy and sequence. It outputs the most probable structure for a given sequence, which can then contribute to further ligation reaction (see Figure 3.15). With these adaptations, several experimental features can be captured, including the transition from binomial to bimodal distribution and length distributions of the reduced complexity systems.

For a detailed analysis and cross-referencing of simulation and experiment, the experimental system has been further reduced to a pool that includes only the four monomers that make up the spike and their complements. The experiments have then been performed with and without the spike at different ligation temperatures. Although the simulations are still ongoing, the first results show a promising resemblance between experimental and numerical behavior. With a successful implementation of the complete sequence information and folding characteristics in the numerical model, it can also be attempted to understand the dynamics of the spiking experiments with the short sequence networks. It is worth noting that the architecture of the numerical model is independent of the specific details of the chemistry and could be adapted to more realistic scenarios in the future.

3.4 Materials and Methods

3.4.1 Nucleic Acids

DNA oligonucleotides were obtained in dry form from biomers and then adjusted to a stock concentration of approximately 200 μ M using nuclease-free water (Ambion™ Nuclease-free water from Invitrogen). All DNA oligonucleotide sequences are provided in Table 3.1. The final concentrations of DNA were determined by measuring the absorbance at 260 nm. All stock solutions were stored at -20 °C and thawed on ice before use.

Table 3.1: List of all DNA sequences used in the experiments.

Name	nt	Sequence (5'-3')	Modification
ATr_12mer	12	www www www www	5'-POH
ATGCr_12mer	12	nnn nnn nnn nnn	5'-POH
RCS_1 x16, x8, x4	12	AAA AAA AAA AAA	5'-POH
RCS_2	12	AAA AAA AAA ATT	5'-POH
RCS_2	12	AAA AAA AAA ATT	5'-POH
RCS_3	12	AAA AAA AAT TAA	5'-POH
RCS_4 x8	12	AAA AAA AAT TTT	5'-POH
RCS_5	12	AAA AAA TTA AAA	5'-POH
RCS_6	12	AAA AAA TTA ATT	5'-POH
RCS_7	12	AAA AAA TTT TAA	5'-POH
RCS_8 x8, x4	12	AAA AAA TTT TTT	5'-POH
RCS_9	12	AAA ATT AAA AAA	5'-POH
RCS_10	12	AAA ATT AAA ATT	5'-POH
RCS_11	12	AAA ATT AAT TAA	5'-POH
RCS_12	12	AAA ATT AAT TTT	5'-POH
RCS_13 x8	12	AAA ATT TTA AAA	5'-POH
RCS_14	12	AAA ATT TTA ATT	5'-POH
RCS_15	12	AAA ATT TTT TAA	5'-POH
RCS_16 x8	12	AAA ATT TTT TTT	5'-POH
RCS_17	12	AAT TAA AAA AAA	5'-POH
RCS_18	12	AAT TAA AAA ATT	5'-POH
RCS_19	12	AAT TAA AAT TAA	5'-POH
RCS_20	12	AAT TAA AAT TTT	5'-POH
RCS_21	12	AAT TAA TTA AAA	5'-POH
RCS_22	12	AAT TAA TTA ATT	5'-POH
RCS_23	12	AAT TAA TTT TAA	5'-POH
RCS_24	12	AAT TAA TTT TTT	5'-POH
RCS_25	12	AAT TTT AAA AAA	5'-POH
RCS_26	12	AAT TTT AAA ATT	5'-POH
RCS_27	12	AAT TTT AAT TAA	5'-POH
RCS_28	12	AAT TTT AAT TTT	5'-POH
RCS_29	12	AAT TTT TTA AAA	5'-POH
RCS_30	12	AAT TTT TTA ATT	5'-POH
RCS_31	12	AAT TTT TTT TAA	5'-POH
RCS_32	12	AAT TTT TTT TTT	5'-POH

Name	nt	Sequence (5'-3')	Modification
RCS_33	12	TTA AAA AAA AAA	5'-POH
RCS_34	12	TTA AAA AAA ATT	5'-POH
RCS_35	12	TTA AAA AAT TAA	5'-POH
RCS_36	12	TTA AAA AAT TTT	5'-POH
RCS_37	12	TTA AAA TTA AAA	5'-POH
RCS_38	12	TTA AAA TTA ATT	5'-POH
RCS_39	12	TTA AAA TTT TAA	5'-POH
RCS_40	12	TTA AAA TTT TTT	5'-POH
RCS_41	12	TTA ATT AAA AAA	5'-POH
RCS_42	12	TTA ATT AAA ATT	5'-POH
RCS_43	12	TTA ATT AAT TAA	5'-POH
RCS_44	12	TTA ATT AAT TTT	5'-POH
RCS_45	12	TTA ATT TTA AAA	5'-POH
RCS_46	12	TTA ATT TTA ATT	5'-POH
RCS_47	12	TTA ATT TTT TAA	5'-POH
RCS_48	12	TTA ATT TTT TTT	5'-POH
RCS_49 x8	12	TTT TAA AAA AAA	5'-POH
RCS_50	12	TTT TAA AAA ATT	5'-POH
RCS_51	12	TTT TAA AAT TAA	5'-POH
RCS_52 x8	12	TTT TAA AAT TTT	5'-POH
RCS_53	12	TTT TAA TTA AAA	5'-POH
RCS_54	12	TTT TAA TTA ATT	5'-POH
RCS_55	12	TTT TAA TTT TAA	5'-POH
RCS_56	12	TTT TAA TTT TTT	5'-POH
RCS_57 x8, x4	12	TTT TTT AAA AAA	5'-POH
RCS_58	12	TTT TTT AAA ATT	5'-POH
RCS_59	12	TTT TTT AAT TAA	5'-POH
RCS_60	12	TTT TTT AAT TTT	5'-POH
RCS_61 x8	12	TTT TTT TTA AAA	5'-POH
RCS_62	12	TTT TTT TTA ATT	5'-POH
RCS_63	12	TTT TTT TTT TAA	5'-POH
RCS_64 x16, x8, x4	12	TTT TTT TTT TTT	5'-POH
RCS_2-x16	12	AAA AAA AAA TTT	5'-POH
RCS_3-x16	12	AAA AAA TTT AAA	5'-POH
RCS_4-x16	12	AAA AAA TTT TTT	5'-POH
RCS_5-x16	12	AAA TTT AAA AAA	5'-POH
RCS_6-x16	12	AAA AAA AAA TTT	5'-POH
RCS_7-x16	12	AAA TTT TTT AAA	5'-POH
RCS_8-x16	12	AAA TTT TTT TTT	5'-POH
RCS_9-x16	12	TTT AAA AAA AAA	5'-POH
RCS_10-x16	12	TTT AAA AAA TTT	5'-POH
RCS_11-x16	12	TTT AAA TTT AAA	5'-POH
RCS_12-x16	12	TTT AAA TTT TTT	5'-POH
RCS_13-x16	12	TTT TTT AAA AAA	5'-POH
RCS_14-x16	12	TTT TTT AAA TTT	5'-POH

Name	nt	Sequence (5'-3')	Modification
RCS_15-x16	12	TTT TTT TTT AAA	5'-POH
Spike	48	TTA AAA TTA AAA TTT TTT AAA ATT TTT TAA AAA AAA AAA AAA AAA AAA	
Spike; A-Type	48	TTA AAA TTA ATT AAT TAA AAA ATT AAA AAA TTA ATT AAT TAA AAA ATT	5'-POH
Spike; T-Type	48	AAT TTT TTA ATT AAT TAA TTT TTT AAT TTT TTA ATT AAT TAA TTT TAA	5'-POH

For the reduced complexity systems RCS-x4, -x8 and -x16, the corresponding sequences are marked in Table 3.1 in violet, teal, and purple, respectively.

3.4.2 Reaction conditions

The *Taq* DNA ligase (New England Biolabs) was used for the enzymatic templated ligation reaction. Prior to the reaction, all DNA oligomers were pipetted in a 2x concentration and mixed in equal parts with a 2x ligase in a buffer solution. The 1x buffer components were stated by the manufacturer to be 20 mM Tris-HCl, 25 mM potassium acetate, 10 mM magnesium acetate, 1 mM NAD 1, 10 mM DTT, 0.1% Triton® X-100 with a pH of 7.6 at 25 °C. The reaction volumes were 15 µl unless stated otherwise. For the reduced complexity stock, all strands were individually ordered and pipetted together to obtain a 100 µM stock. All pipetting was done on ice.

3.4.3 Temperature cycling protocols

For temperature cycling, the Thermo Fisher Scientific ProFlex PCR System was used to cycle between dissociation and ligation temperatures. As dissociation temperature, a temperature of 75 °C was chosen, ensuring melting of the 12mer monomer pool up to 36mers. Additionally, choosing 75 °C as a relatively low dissociation temperature compared to standard dissociation temperatures of 95 °C enables reaction setups of several thousand temperature cycles. Exposing the *Taq* DNA ligase to 95 °C for extended time periods leads to a significant decrease in its stability and, thus, to a loss of activity over time. The ligation time t_{lig} was varied for different reactions. The dissociation time was kept at 20 s for all experiments. The temperature protocol was initiated by a short dissociation step at 75 °C. After completion of all cycles, the reaction was kept at 4 °C, after which it was removed from the cyclor and quenched with loading buffer, and stored at -20 °C until analysis.

3.4.4 Melting curves

For information on the thermal stability of the model ligation system, melting temperatures were experimentally measured using a C1000™ Thermal Cycler with a CFX96™ detection system from Bio-Rad Laboratories (BioRad CFX Manager 3.1). Melting curves were derived from the change of fluorescence of 1x EvaGreen Plus dye (stock at 20X in Water, Biotium). They were done for several oligomer lengths and sequences spaces, i.e. a 6 nt, 12 nt, 18 nt, and 24 nt long AT random pool, as well as a 12 nt long RCS-x64, RCS-x16, and 18 nt long RCS-x16 pool, with a total concentration of 10 µM (see Figure 3.17). The melting curves were done in the 1x ligase buffer (but without the ligase).

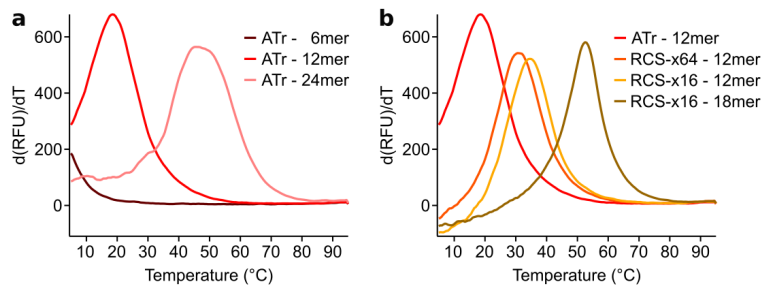


Figure 3.17: Melting temperatures for different lengths and sequence spaces: The melting temperature can be derived from the maximum change in the fluorescence signal over the whole temperature range. **a** No melting transition can be detected for the 6 nt long AT random sample, suggesting that they remain single-stranded throughout the whole temperature range. For the 12 nt and 24 nt long AT random samples, the melting temperatures are $\approx 20^\circ\text{C}$ and $\approx 48^\circ\text{C}$, respectively. **b** Not only the length but also the sequence space influences the melting temperature. For the same length (12 nt), the melting temperature varies from $\approx 20^\circ\text{C}$ to $\approx 35^\circ\text{C}$ for sequence spaces of 4096 compared to 16 different sequences.

3.4.5 *Taq* DNA Ligase specifications

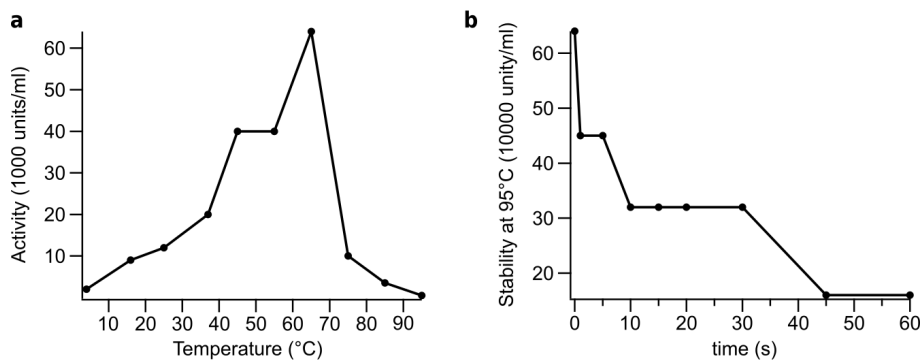


Figure 3.18: Ligase performance as given by the manufacturer: **a** The *Taq* DNA ligase activity shows a temperature dependence with an activity maximum around 70°C , good activity between 40°C - 60°C , some activity between 15°C - 40°C and little to no activity for $<10^\circ\text{C}$ and $>85^\circ\text{C}$. **b** The stability at 95°C of the *Taq* DNA ligase steadily declines. One unit corresponds to the amount of *Taq* DNA ligase that is required to give 50% ligation of the 12-base pair cohesive ends of $1\ \mu\text{g}$ of BstEII-digested λ DNA in a total reaction volume of $50\ \mu\text{l}$ in 15 minutes at 45°C . The data was adapted from the manufacturer's website [168].

The *Taq* DNA Ligase is a thermostable ligase able to covalently link a downstream 5' phosphorylated oligonucleotide to its upstream oligonucleotide ligation partner, both bound to a template strand. Its ligation activity is expressed in units/ml, where one unit is defined as the amount of enzyme required to give 50% ligation of the 12-base pair cohesive ends of $1\ \mu\text{g}$ of BstEII-digested λ DNA in a total reaction volume of $50\ \mu\text{l}$ in 15 minutes at 45°C . The activity is temperature dependent with little to no activity for low (4°C - 10°C) and very high

temperatures (85 °C-95 °C) as stated by the manufacturer (see Figure 3.18 a). At temperatures of 95 °C, its stability decreases, reducing its activity. The storage conditions are 10 mM Tris-HCl, 50 mM KCl, 1 mM DTT, 0.1 mM EDTA, 200 µg/ml BSA, 50% Glycerol with pH 7.4 at 25 °C. The manufacturer states a maximum ligation of mismatches of 1.3%.

3.4.6 PAGE analysis

Polyacrylamide gel electrophoresis (PAGE) was used to analyze and quantify the ligation dynamics and yields. The gels were composed of 15% acrylamide prepared from a gel stock of Roth Rotiphorese DNA sequencing gel stocks with an acrylamide-to-bisacrylamide ratio of 19:1. For each gel with a thickness of 0.75 mm and a 15 tooth comb about 5 ml gel mixture was needed, containing 3 ml gel concentrate, 1.5 ml gel diluent, 0.5 ml buffer concentrate, 25 µl APS and 2.5 µl TEMED. To ensure denaturing conditions, the gels ran in a solution of 50% urea and 1x TBE buffer at ~50 °C posing denaturing conditions. Before loading, a 30 min pre-run at 400 V was started, after which the gel pockets were loaded with 4 µl composed of 0.89 µl of sample and 3.11 µl of loading buffer. About 10 ml of loading buffer 9.5 ml formamide, 0.5 ml glycerol, 1 µl EDTA (0.5 M), and 100 µl Orange G dye (e.g., New England Biolabs) were mixed. The gel was then run at 50 V for 5 min, followed by a second step at 300 V for about 25-30 min. After completion of the run, the gels were post-stained using SYBR gold. To this end, they were submersed after the run into 50 ml of 1x TBE buffer with 5 µl of 10.000x SYBR Gold Nucleic Acid Gel Stain from Thermo Scientific for 5 min. Afterwards, the gel was washed twice in 1x TBE buffer and subsequently imaged with a bio-rad ChemiDoc MP System.

Analysis of the gel images is done in in-house written LabVIEW code is described in detail in [111, 168].

3.4.7 Illumina sequencing

All samples were sent to the Gene Center Munich (LMU) for library preparation and sequencing. The samples were sequenced on the NGS Illumina NextSeq 1000 machine (flow cell type P2, 2 x 50 bp with 138 cycles for 100 bp single-end reads). For each zero sample, 5 million, and for each ligation sample, 50 million reads were ordered. Before sequencing, the samples were prepped using the ACCEL-NGS 1S Plus DNA Library Kit for library preparation. Library preparation was done in four steps, starting with the addition of a random sequence CT-tail to the 3' end of the DNA, followed by the ligation of the back primer sequence to the 3' end, a single-cycle PCR step to produce double-stranded DNA with a single A overhang, the ligation of the start primer to the 5' end, and the addition of barcode indices to both ends of the DNA via PCR.

After Illumina sequencing, the demultiplexing algorithm matched the barcodes with the library DNA and produced a consolidated FASTA file, which also included read quality scores. Demultiplexing of the sequence data was conducted using standard algorithms provided by the Galaxy servers. Each read base was assigned a Phred quality score during sequencing, corresponding to the probability of an incorrectly read base at this position [173]. As an overall quality filtering led to a drastic reduction of obtained reads, low-quality segments were trimmed from the sequencing using Trimmomatic [174]. Thereby, a sliding window of 4 nt starting from the 3' to 5' end direction was used, which was trimming the sequence at the

leftmost base of the window as soon as average Phred quality dropped below 20.

Data analysis was then carried out using an in-house written LabVIEW software as reported in [111], which also allows to separate the read sequences from the attached primers by using an algorithm based on regular expressions (RegEx) pattern matching as described in the following.

RegEx for Searching AT Random Sequences.

The regexing process involved identifying multiples of the monomer length followed by at least four bases of C or T and culminating with the sequence AGAT, signifying the presence of a relevant sequence. The 3'-primer was excised, and the resulting sequence was retained for subsequent analysis. The RegEX pattern used to search for correct reads is the following:

$(^ [AT] \{108\} | [AT] \{96\} | [AT] \{84\} | [AT] \{72\} | [AT] \{60\} | [AT] \{48\} | [AT] \{36\} | [AT] \{24\} | [AT] \{12\})$
 $(? = ([CT] \{4, \} AGAT))$

On top of that a RegEx pattern for selecting a maximum of X false reads of G or C in random sequence AT samples is employed:

$^(?! (? : .?(G|C)) \{X, \}) ^([ATCG] \{12, \})$

In the case of the reduced complexity samples X=2 was generally allowed. The resulting sequences were then transcribed to virtual space, where all sequences that only contain A and T bases but not in the form of double bases (AA and TT) are removed. However, the number of those 'false reads' is relatively low.

3.4.8 Sample Pearson Correlation Coefficient Matrix

The sample Pearson Correlation Coefficient (PCC) is a measure for linear correlation, e.g. the similarity of sequences, between two data sets. To compare the similarity of sub-sequence patterns of the ligation products, the sequences are divided into sub-sequence parts of similar length as the monomers (6 nt). The sub-sequences at a specified position of a x-mer can then be compared to the sub-sequences of a y-mer. If their abundances are plotted against one another, the shape of the distribution allows to draw conclusions about their correlation. Highly (linearly) correlated samples are distributed in a line-like shape, whereas uncorrelated samples exhibit a cloud-like distribution. The PCC can be computed as follows:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.2)$$

where n refers to the size of the sample, i is the index over all individual sequences with abundances x_i, y_i , and the mean abundances \bar{x}, \bar{y} . Values close to 1 or -1 describe a positive or negative linear correlation, respectively. Values close to 0 describe no correlation.

3.5 Appendix - Supplementary Figures

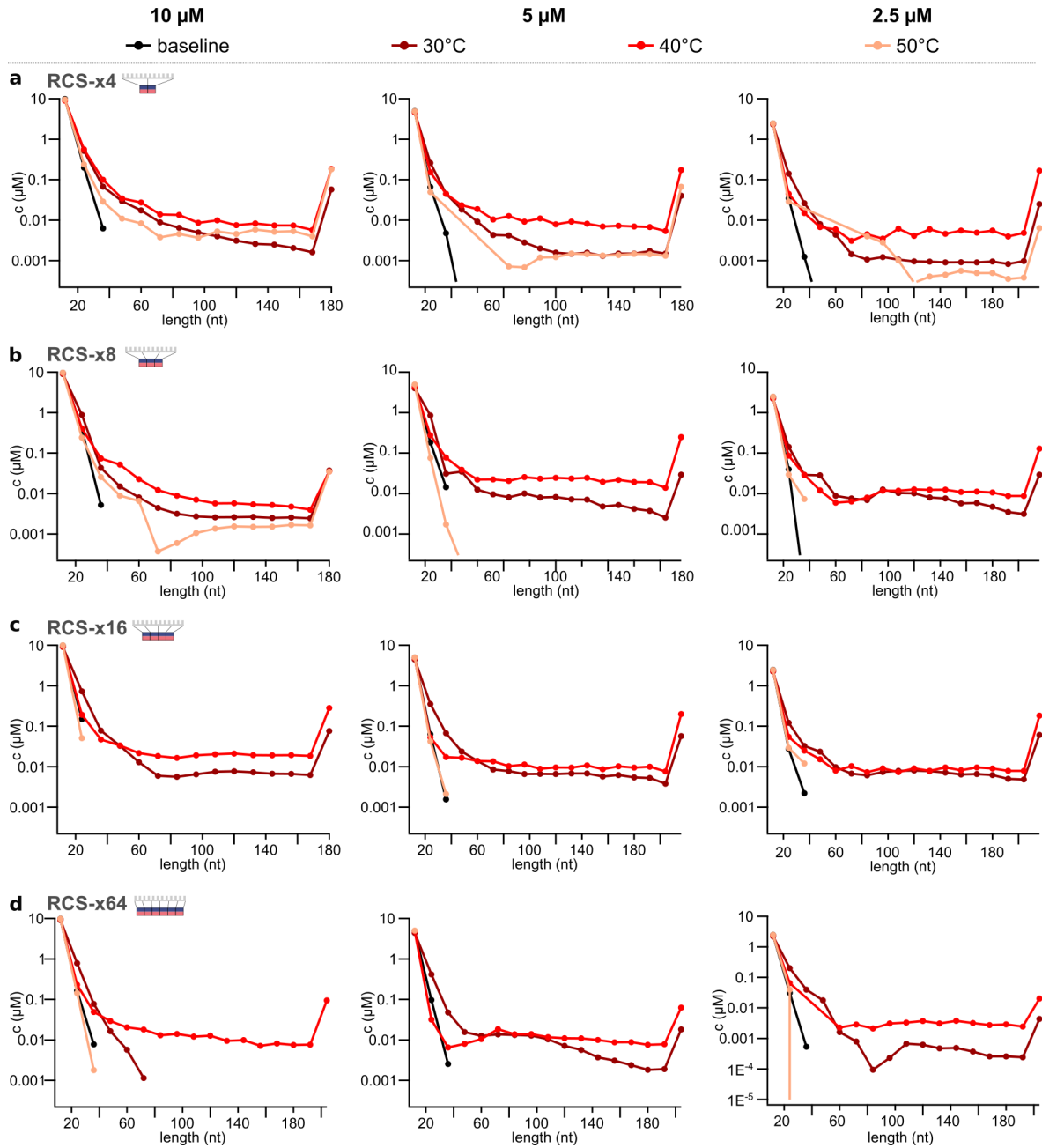


Figure 3.19: T_{ig} and concentration screen for the reduced complexity systems: PAGE analysis for RCS-x4, RCS-x8 and RCS-x16 with a total concentration of **a** 10 μM , **b** 5 μM , **c** 2.5 μM . **d** PAGE analysis for RCS-x64 with total pool concentrations of 10 μM , 5 μM and 2.5 μM .

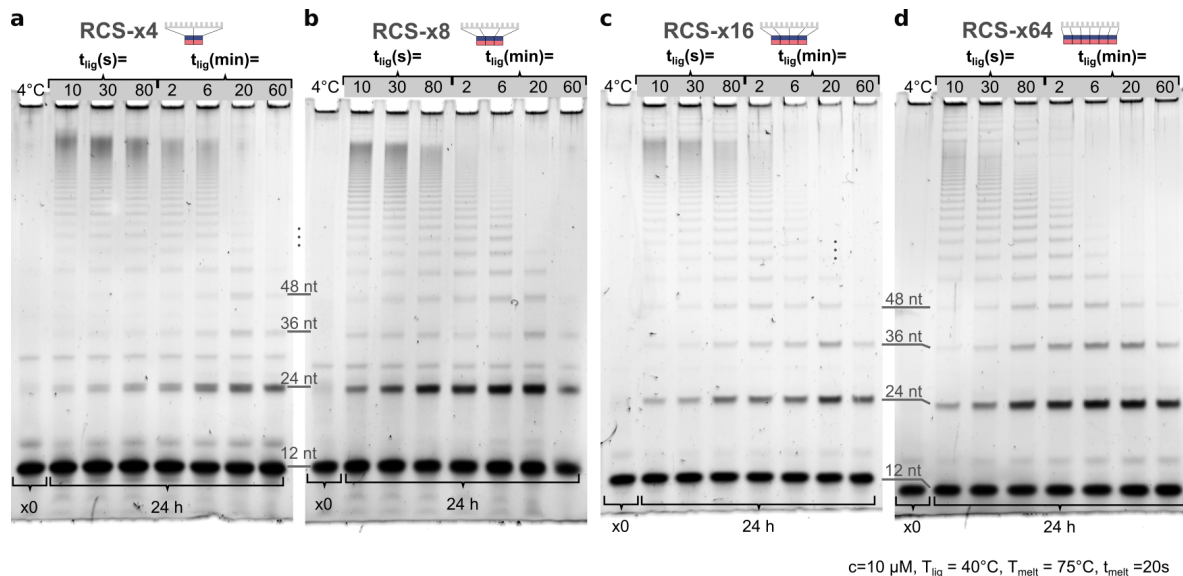


Figure 3.20: t_{lig} screen for the reduced complexity systems: All reduced complexity systems, i.e., **a** the RCS-x4 (effective 2mers), **b** the RCS-x8 (effective 3mers), **c** the RCS-x16 (effective 4mers) and **d** the RCS-x84 (effective 6mers) exhibit product formation for all screened t_{lig} . The overall reaction time was kept at 24 h, with the longest t_{lig} =60 min corresponding to the lowest number of cycles. The smallest t_{lig} showed the highest concentration of long products.

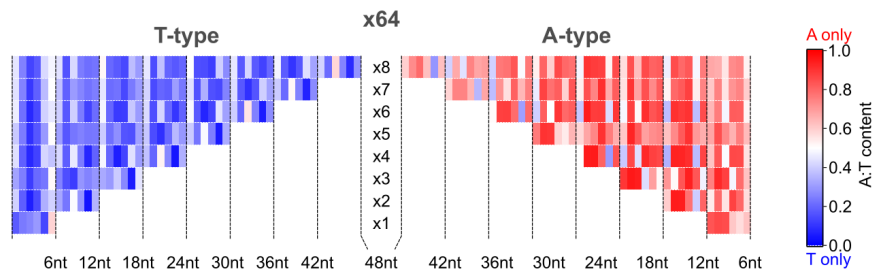


Figure 3.21: A-type vs. T-type for RCS-x64 pool: The base probabilities of A- and T-types exhibit distinct sequence motifs that show some reverse complementary between the two types.

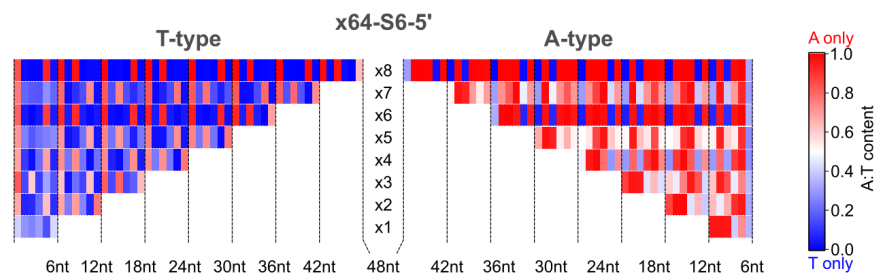


Figure 3.22: A-type vs. T-type for the RCS-x64 pool spiked with system 6-5': Separating the base probabilities into A-type and T-type shows an enhancement of the sequence motifs already present in the unspiked case. Especially for longer lengths the A-types are perfect reverse complements of the T-type.

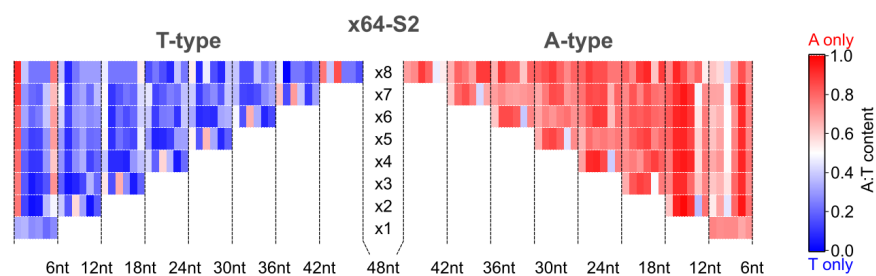


Figure 3.23: A-type vs. T-type for the RCS-x64 pool spiked with system 2: The base probabilities of the separated A-type and T-type show now obvious similarities between the two types. However, both types exhibit long stretches especially of poly A (A-type) but also poly T (T-type).

4 A thermal habitat for RNA amplification and accumulation

Summary

For the emergence of early replication, RNA-catalyzed processes are assumed to be an essential step. *In vitro* evolution provided ever better RNA polymerases, but the required strand separation as well as the emergence of these complex ~200-base catalytic sequences are unsolved questions for the Origin of Life. This chapter presents a microfluidic reaction compartment with a pointed heat source that both protected and drove laminar thermal convection in aqueous solution and allowed the autonomous, exponential RNA amplification by the RNA strand separation. Accumulation experiments with fluorescently labeled nucleic acids revealed a ring-like accumulation pattern for the long functional RNA-polymerase as well as its DNA complement. In contrast, similar experiments for dsDNA of equivalent length showed the expected central accumulation. Imaging the reaction mixture at higher resolution revealed the formation of micrometer-sized conglomerates that depended on the presence of PEG, introduced as a crowding agent in the buffer. By including a diffusiophoretic term, the experimental accumulation behavior could be matched by simultaneously simulating the accumulation of the conglomerates and PEG with a finite element simulation code. The results revealed how the thermal gradient caused RNA polymerase molecules to accumulate into a ring geometry, protecting them from hot regions of the chamber. These findings demonstrate a size-selective pathway for autonomous RNA-driven replication¹.

¹This chapter is based on the publication by Salditt, Keil, and Horning et al. [166] in Physical Review Letters. Authors: Annalena Salditt*, Lorenz M. R. Keil*, David P. Horning*, Christof B. Mast, Gerald F. Joyce and Dieter Braun (* contributed equally)
Contributions: A.S., L.K., D.H., and C.M. performed the experiments. L.K., D.H., A.S., C.M., G.J., and D.B. conceived and designed the experiments, A.S., L.K., D.H., C.M., G.J. and D.B. analyzed the data and wrote the paper.

4.1 Introduction

4.1.1 RNA model system for early replication scenarios

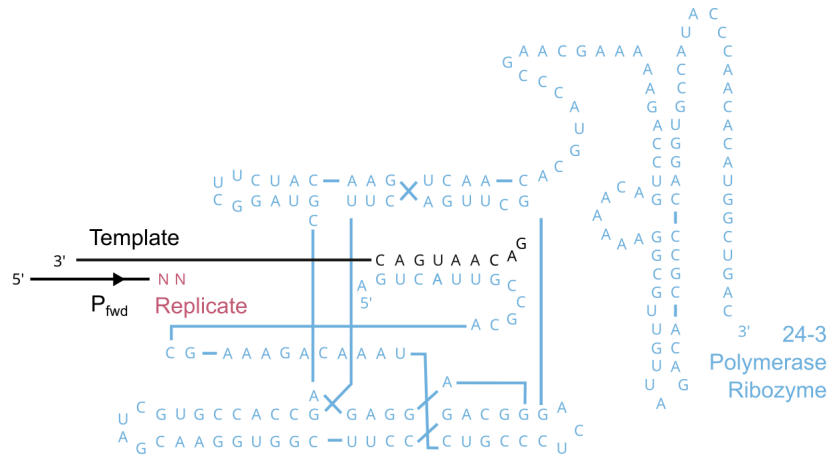


Figure 4.1: Sequence and secondary structure of the 24-3 polymerase ribozyme: The sequence and secondary structure of the 24-3 polymerase ribozyme are shown in blue. At the 5'-end of the polymerase, the template primer complex (black) can bind, leading to the elongation of the primer (pink). The figure is adapted from [125].

In today's biology, the translation of information from DNA to proteins is mediated by an RNA intermediate. This separates the requirements for the storage of genetic material and the performance of the functional entities. However, in contrast to the highly evolved and complex machinery required for replication in modern living systems, early replication scenarios must have relied on simpler means. Notably, RNA itself possesses the dual capability of storing genetic information as well as folding into catalytically active structures, including those capable of RNA replication processes, making it an attractive molecule for early replication scenarios [80, 125, 162]. This so-called RNA world hypothesis was sparked by the discovery of an RNA ribozyme in the large ribosomal subunit catalyzing the synthesis of proteins in today's biology [175]. Thus, an RNA-only replication mechanism could provide a solution for the storage and effective transfer of information on early Earth prior to the emergence of genetically encoded proteins [18, 19, 176, 177].

Any RNA copying mechanism based on templated ligation or templated polymerization relies on Watson-Crick base pairing to pass on information. This implies the necessity of an energy source to separate the two complementary strands (product-template complex) from the templated synthesis. Only then the next round of templated synthesis can begin. Separation of hybridized strands can be achieved by means of pH cycling [178, 179], evaporation-wetting cycles [103, 180, 181], oscillation of salt concentrations [83], or elevated temperatures. To ensure the viability of RNA replication mechanisms, it is essential to maintain conditions that minimize the spontaneous cleavage of RNA molecules [182, 183]. This cleavage typically occurs under high-salt conditions, which are necessary for RNA catalysis [184–186]. Therefore, it is crucial to establish a specific setting that minimizes the degradation of RNA, allowing for the successful replication and preservation of genetic information.

The RNA model system studied in this chapter is based on the 24-3 RNA polymerase ribozyme evolved by Horning and Joyce (see Figure 4.1) [125]. The selection pressure was designed to support sequence generality and accuracy in the copying mechanism. This was achieved by a selection step based on the binding of the complete synthesis of different aptamers that subsequently bind to their cognate ligand. As exponential amplification is a key feature for any replication scenario to maintain and protect copied information against degradation or dilution, Horning and Joyce showed that the 24-3 RNA polymerase can carry out PCR-like amplification solely based on RNA (riboPCR) by extending an RNA primer, either the forward (P_{fwd}) or reverse (P_{rev}) one, on an RNA template with nucleoside triphosphates (NTPs). Part of the sequence of the template is complementary to the 5' end of the 24-3 polymerase, allowing it to be bound to the polymerase by Watson-Crick base pairing. The same sequence region is included in the primers to ensure that the copied strands can act as templates and be replicated further. The complementary region is followed by a hexadenylate spacer, i.e., six consecutive adenosine nucleotides. The sequences and the binding sites are shown below, where the blue part always indicates the binding site of the respective template to the polymerase, and magenta indicates the extended sequence part:

Template

3' CGUUGAAAAG CACACCUCAC AAAAAA CAGUACAG 5'
 5' GACAAUGAC AAAAAA GCAACUUUUC GUGUGGAGUG 3'

P_{fwd}

3' GUGAGGUGUG CUUUUCAACG AAAAAA CAGUACAG 5'
 5' GACAAUGAC AAAAAA CACUCCACAC GAAAAGUUGC

P_{rev}

To minimize RNA cleavage at the high-temperature spikes, the Mg^{2+} concentration was reduced from 200 mM to 50 mM. However, this was only feasible after introducing PEG8000 to the buffer, which acted as a molecular crowding agent to improve ribozyme activity at the reduced Mg^{2+} concentration [187]. Additionally, tetrapropylammonium chloride (TPA) was included to lower the melting temperature of the duplex RNA [188, 189]. Upon temperature cycling of a denaturation step at 68 °C for 1 s followed by an extension step at 17 °C for 30 min for 72 cycles, an amplification of the template and the reverse template sequence could be observed. The following section will expand on how a thermal microenvironment can naturally drive such a replication scenario of amplification of small RNAs catalyzed by a larger polymerase ribozyme.

4.1.2 Thermal habitat for RNA replication

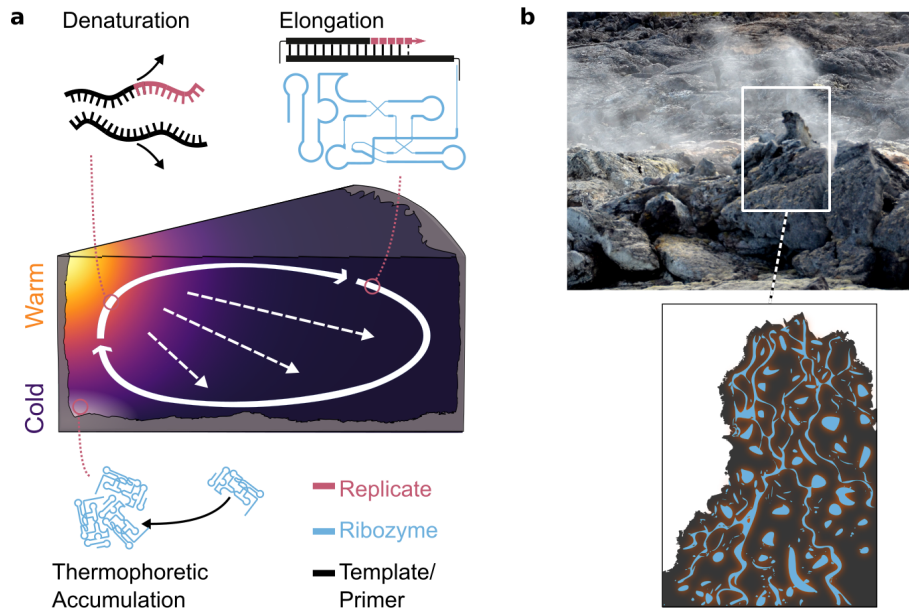


Figure 4.2: A temperature gradient across a water-filled porous rock can drive RNA-catalyzed RNA replication: **a** A punctual heat source creates a temperature gradient across a water-filled pore. The temperature gradient induces both thermophoresis of molecules (dashed arrows along the thermal gradient) and convection of water (solid arrows). Finite-element simulations predict that these non-equilibrium effects lead to a local up-concentration of the RNA polymerase toward cold regions of the cylindrical chamber, protecting it against thermal degradation. Smaller RNA molecules are shuttled through the high-temperature spot instead of being accumulated. The resulting temperature oscillations lead to strand separation of the product-template complex in the warm spot after the template-directed replication in the cold areas. **b** Volcanic or hydrothermal rocks can be comprised of water-filled porous structures exposed to natural temperature differences. These structures could be the first habitats of early replication scenarios, providing the necessary non-equilibrium conditions.

A habitat suitable for RNA replication on early Earth should be able to provide two crucial characteristics: the ability to separate double-stranded RNA and a mechanism to ensure persistent accumulation of the copied information against dilution by diffusion. Previous studies have demonstrated that the accumulation of nucleic acids can occur within closed, elongated, water-filled compartments through the interplay of thermophoresis of molecules and convection induced by a temperature gradient across the pore. This mechanism favors the retention of longer RNA strands within a replication reaction mixture [101, 165, 190]. While this replication and selection process has been shown for protein-based replication of DNA employing a *Taq* polymerase [73] also in conjunction with accumulation methods [191, 192], it remained uncertain whether the thermal instability of RNA would impede a similar approach.

This section describes a modified thermal microenvironment designed to drive the previously described replication mechanism of a PCR-like amplification of small RNAs catalyzed by the 24-3 polymerase ribozyme. By realizing a punctual heat source within a cylindrical compartment, as shown in Figure 4.2 a, two essential effects are achieved:

- (i) The laminar convection arises due to the temperature-dependent density of water upon gravity, which results in temperature cycles that meet the specific requirements for the elongation and strand separation for RNA-catalyzed RNA replication.
- (ii) The thermophoretic movement of molecules occurs along the applied temperature gradient, with molecules being driven outward from the high-temperature region in a length-dependent manner.

The phenomenon of thermophoretic movement relies on a combination of non-ionic interactions, ionic shielding, and Seebeck effects [101], where in the case of polyanionic nucleic acids at elevated temperatures, thermophoresis leads to a movement of the molecules from warmer to cooler areas [193]. For a cylindrical compartment, as presented here, the interplay of both phenomena, convective and thermophoretic transport, resulted in a length-dependent net transport of molecules away from the warm temperature spot.

The efficiency of this net transport is increased for longer RNAs such as the 24-3 polymerase ribozyme (180 nt), thereby stabilizing them against RNA cleavage that would occur at higher temperatures. In contrast, shorter replicated RNA oligomers (~35 nt) would be cycled rapidly through the hotter region of the habitat, leading to their thermally-induced denaturation. This cycling process guarantees the melting of double-stranded RNA molecules, thereby releasing templates from the copied strand. This allows subsequent rounds of polymerization of both the original template and its complement. Consequently, genetic information could be replicated and preserved within a single environment driven by thermal dynamics. Such thermal hatcheries could have occurred in porous volcanic or hydrothermal rock formations on the early Earth (Figure 4.2 b) and might have played a vital role in facilitating the emergence of an RNA-based origin of life.

A focused infrared laser was used to introduce the centralized heat flow inside a water-filled cylindrical chamber, leading to a radially symmetric temperature distribution. Within the chamber, a central hot temperature spot of approximately 80 °C was generated, gradually decreasing to 17 °C at the water-cooled bottom side of the chamber. The temperature profile was measured using fluorescent thermometry (see Materials and Methods and Figure 4.12). A similar geometric setup involving localized heating of the chamber surface has been previously reported for protein-catalyzed replication [194]. Numerical simulations confirmed the similarity of temperature profiles between these implementations (see Materials and Methods and Figure 4.12).

4.2 Results

4.2.1 Convective temperature oscillations

As the requirements for the replication conditions of the riboPRC reaction with the 24-3 polymerase have to be matched both in temperature settings as well as timing, it was essential to model the cycling condition inside the convective chamber.

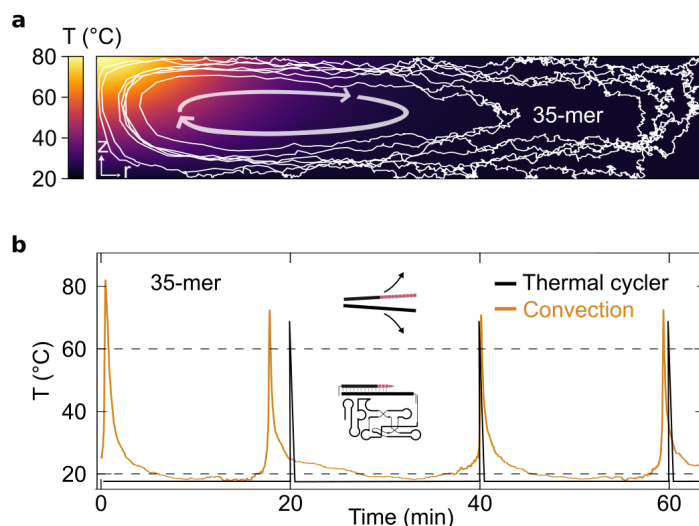


Figure 4.3: Convective temperature oscillations inside cylindrical chamber: **a** The white arrows indicate the convective flow. The thin white lines correspond to exemplary simulated stochastic trajectories for a 35mer inside the cylindrical convection chamber. Combined with the simulated temperature profile, the difference in velocity at different temperatures becomes apparent, as almost no Brownian motion occurs in the hot temperature region. The height and radius of the compartment are 500 μm and 2.25 mm, respectively. **b** In the convection chamber, RNA predominantly occupied low-temperature regions suitable for extension by polymerization while swiftly traversing high-temperature regions that allowed strand separation. Stochastic simulations indicated an average temperature cycling time of ~ 26 min for a 35mer. Thermal cycler experiments were conducted with cycles of 20 min at 17 $^{\circ}\text{C}$ followed by a 2 s temperature spike at 68 $^{\circ}\text{C}$.

The temperature within the chamber was determined by measurements of the temperature-dependent fluorescence decrease of BCECF to obtain the temperature profile (see section 4.5.3 for details). The resulting profile was used as a reference for finite-element simulations of the heat transfer inside the convective chamber of 500 μm height and 2.25 mm radius (see Figure 4.13), which was then used to derive the velocity profile $\vec{v}(x,y)$ of the laminar convection as a function of the position inside the chamber. Simulating trajectories solely based on this convection resulted in a mean cycle time of 19 min (see Figure 4.14).

To include a species-specific flow profile, the simulation was expanded, accounting for laminar convection and thermophoretic drift (refer to section 4.5.4). All finite-element simula-

tions were performed in Comsol Multiphysics 5.4. To further derive statistical data on thermal cycles of the RNA oligonucleotides, Brownian motion was additionally implemented by a randomly directed movement $\vec{\eta}(t)$. The resulting position shift $\Delta\vec{s}(x,y)$ is described by:

$$\Delta\vec{s}(x,y) = \sqrt{4D\Delta t} \cdot \vec{\eta}(t) + \Delta t \cdot (\vec{v}(x,y) + D \cdot S_T \cdot \nabla T(x,y)) \quad (4.1)$$

where Δt denotes the time-step, $T(x,y)$ the local temperatures, $\vec{v}(x,y)$ the convective flow, D the diffusion constant and S_T the Soret coefficient. Details of the simulation are described in section 4.5.5.

The molecules were considered to complete a full temperature cycle when moving from a high-temperature region (>60 °C for denaturation) to a low-temperature region (<20 °C for elongation) and back to a high-temperature region. For 35-nucleotide RNA molecules exhibited a mean cycle time of 26 min, oscillating between the threshold temperatures of 20 and 60 °C, was observed (see Figure 4.3). This finding corresponded well with the temperature cycling protocol of the thermal cycler experiments, where the bulk amplification was conducted through cycles of 68 °C for 2 s, followed by a 20 min step at 17 °C.

4.2.2 Amplification of RNA by riboPCR in convection chamber

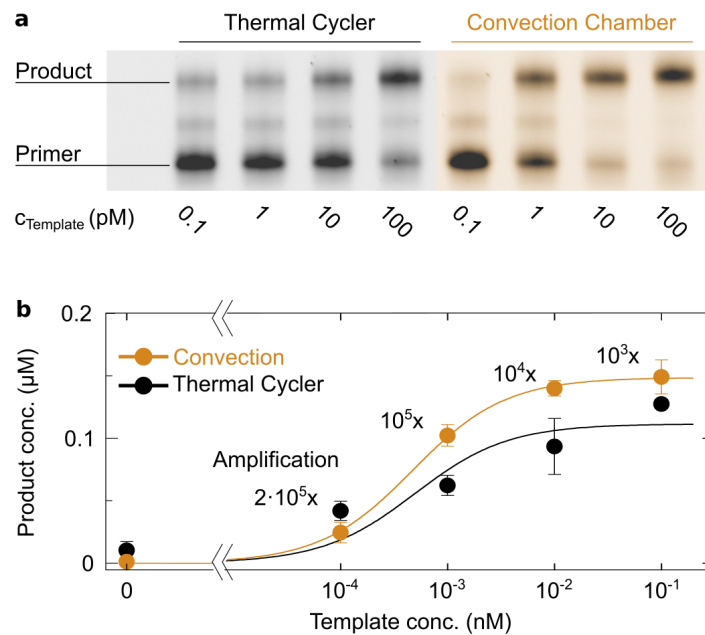


Figure 4.4: Convective RNA-catalyzed replication of RNA: **a** Similar yields of primer extension on the 35-nucleotide RNA template by the polymerase ribozyme were observed for both the convective system as well as the thermal cycler. **b** The polymerase ribozyme achieved exponential amplification, starting from template concentrations as low as 100 fM, within a run time of 24 hours. The amplification process can be effectively modeled using a two-parameter growth equation (see section 4.5.6). The convection chamber showed a maximum amplification of $2 \cdot 10^5$ -fold. The error bars in the figures represent the deviation observed in duplicate experiments.

The convection chamber indeed drove the amplification of a 35-nucleotide RNA template using 25-nucleotide short RNA primers catalyzed by the 24-3 polymerase ribozyme. Due to the slow polymerization rate of the ribozyme, the convection chamber had to be run for approximately 24 hours. Under these conditions, the convection chamber showed exponential amplification of the 35mer RNA, with starting template concentrations as low as 100 fM (see Figure 4.4). Similar results were obtained by performing thermal cycling under optimized conditions of the bulk reaction. The latter was run for 50 temperature cycles with a cycle time of 20 min. Starting from a 100 fM RNA template, the RNA-catalyzed amplification produced product yields of 2×10^5 -fold and 4×10^5 -fold for the bulk and convectively driven reactions, respectively, as shown in Figure 4.4 b.

The amplification process could be theoretically modeled using a two-parameter growth equation (section 4.5.6), yielding a similar maximum replication efficiency (E) for convection and thermal cycling. For convection, E was calculated to be 1.27, while for thermal cycling, E was determined to be 1.28. As described in section 4.1.1, the buffer used was optimized to enhance replication yields for bulk experiments and adopted as such for the convection experiments. This included the addition of PEG8000 as a molecular crowding agent.

4.2.3 Thermophoretic accumulation of RNA polymerase

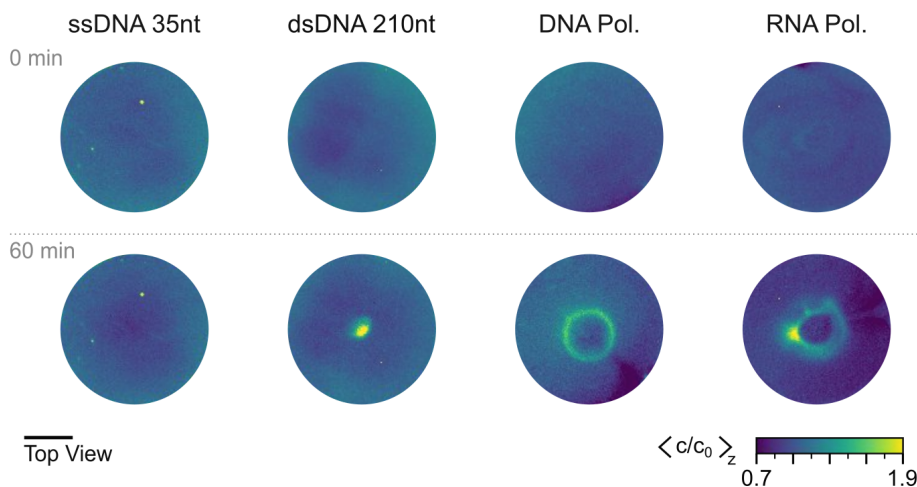


Figure 4.5: Thermophoretic accumulation of nucleic acids: The accumulation pattern of fluorescently labeled nucleic acids inside the convection chamber with a central heat source implemented via an IR-laser was investigated for 35 nt long single-stranded DNA, 210 nt long double-stranded DNA, and folded DNA and RNA complexes corresponding to the polymerase sequences (from left to right). After 60 min, the short single-stranded DNA showed no significant accumulation, whereas the 210 nt long double-stranded DNA exhibited a central 5-fold accumulation. In contrast to the purely double-stranded 210mer, both, the DNA and RNA complexes showed a ring-shaped pattern. The scale bar corresponds to 1 mm.

To gain insight into the potential protection of the long RNA from the high-temperature regime inside the convection chamber, its accumulation behavior was studied. To do so, the

accumulation behavior of different RNA components was investigated by examining fluorescently labeled single- and double-stranded DNA substitutes of the two important length regimes (35 nt and 210 nt), in addition to the RNA polymerase and its DNA analog. Previous work has shown that diluted DNA and RNA, whether in a single-stranded or double-stranded state, exhibit similar thermophoretic behavior, whereas it strongly depends on the length of the molecules [101, 192].

As anticipated, the shorter DNA molecules (35 nt) exhibited only very minor accumulation within the convection system. In contrast, fully double-stranded DNA with a length of 210 nucleotides demonstrated a prominent 5-fold accumulation at the central region of the chamber already after 60 min (see Figure 4.5). These experimental observations were consistent with finite-element simulations incorporating convection, diffusion, and thermophoresis as contributing factors. The transport of the nucleic acid species i is described by the molecular flux (\vec{j}_i) resulting from diffusion, thermophoresis, and convection. It is described by:

$$\vec{j}_i = -D_i \cdot \nabla c_i - S_{T_i} \cdot D_i \cdot \nabla T \cdot c_i + \vec{v} \cdot c_i \quad (4.2)$$

where c_i , D_i , S_{T_i} and \vec{v} denote the concentration of molecules, diffusion coefficient, Soret coefficient, and convective transport velocity. A detailed description of the finite-element simulations is provided in section 4.2.4.

Of particular interest was the accumulation behavior of the RNA polymerase. As described above, the RNA polymerase forms a ternary complex with an RNA primer and template due to complementary sequences located at the 5' ends of both the polymerase and template molecules. The same applies to the DNA analog. Thus, instead of direct labeling, this configuration allowed the molecules to be stained with a fluorescently labeled primer, enabling further investigation and analysis of their accumulation behavior. Both the RNA polymerase as well as its DNA analog showed a distinct accumulation pattern in form of a ring shape in the colder regions of the chamber (see Figure 4.5). This was unexpected, as the 210mer ds-DNA and the polymerase-primer complex should experience a comparable thermophoretic drift.

Examining the solution at a higher resolution using a 40 μm thin capillary revealed the presence of conglomerates for both the DNA and RNA polymerases at 17 $^\circ\text{C}$. In contrast, the 35mer single-stranded DNA and 210mer double-stranded DNA did not form aggregates, as imaging of the solution showed a homogeneous distribution of fluorescence (see Figure 4.6 a).

Through the systematic removal of the individual buffer components, the presence of the crowding agent PEG8000 (6% w/v) was shown to be the crucial component in inducing the aggregation phenomenon (see Figure 4.6 b). This effect was observed for both the RNA and DNA versions of the polymerase sequence. Interestingly, the conglomerates exhibited a temperature-dependent behavior, as uniform heating of the solution resulted in the melting of these aggregates (see Figure 4.6 c). Upon cooling, the conglomerates reappeared.

Despite performing finite-element simulations using experimentally determined diffusion coefficients and a broad range of Soret coefficients, the central accumulation observed in the conglomerates persisted. This suggested the presence of an additional interaction not

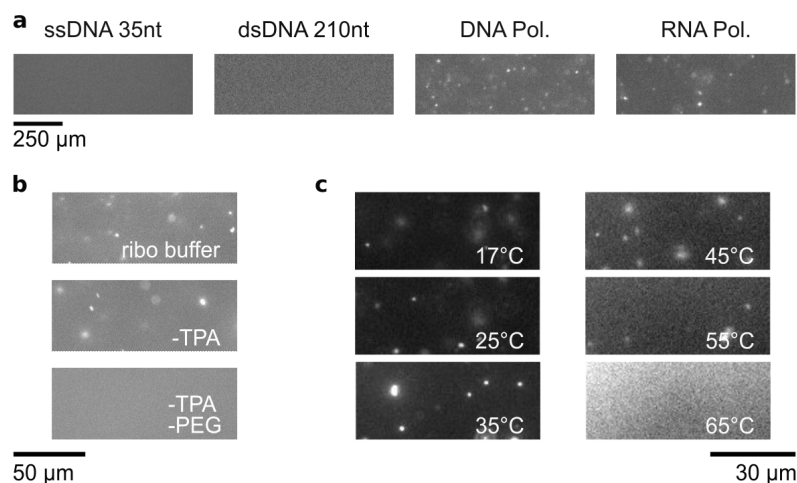


Figure 4.6: Characteristics of nucleic acids conglomerates: **a** Imaging the different nucleic acids inside the reaction buffer revealed the formation of conglomerates in the case of the polymerase sequences, whereas no conglomerates were detected for the single-stranded 35mer, nor the double-stranded 210mer. **b** Through a successive removal of the individual buffer components, exemplary shown for the RNA polymerase, it was discovered that the presence of PEG8000 is critical for driving the formation of conglomerates. **c** The conglomerates exhibit a temperature-dependent behavior, whereby they disappear and reemerge reversibly at higher and lower temperatures, respectively.

yet captured by the simulation. To address this, a diffusiophoretic interaction term was incorporated into the simulations as a possible explanation. This term would potentially account for an interaction between the concentration gradient of polyethylene glycol and the nucleic acid conglomerates, which could play a significant role in their accumulation behavior. The following section looks at this effect and describes the details of the finite-element simulation.

4.2.4 Finite-element simulation of the thermophoretic accumulation

In the previous section, PEG8000 was discovered to be the crucial buffer component to induce the formation of the polymerase conglomerates. Simulating the combined transport phenomena of convective, thermophoretic, and diffusive movement led to a central accumulation similar to the long 210mer double-stranded DNA (see Figure 4.7 a). Thus, the resulting concentration gradient of PEG as the molecular crowding agent in the discussed buffer could have a profound effect on the accumulation behavior of the conglomerates.

Similar to thermophoresis, diffusiophoresis is the emergent transport of molecules, in this case, along a concentration gradient rather than a temperature gradient. In the present case, diffusiophoresis would lead to the transport of the polymerases along the concentration gradient induced by the thermophoretic movement of the PEG8000. The molecular flux \vec{j}_i must then include the diffusive, convective, and thermophoretic as well as diffusiophoretic trans-

port and is thus given by:

$$\vec{j}_i = -D_i \cdot \nabla c_i - S_{T_i} \cdot D_i \cdot \nabla T \cdot c_i + (\vec{v} + \vec{u}_D) \cdot c_i \quad (4.3)$$

where equation 4.2 is expanded by a diffusiophoretic velocity \vec{u}_D . For better clarification, the equation can also be expressed as:

$$\vec{j}_i = -D_i \cdot \nabla c_i + (\vec{v} + \vec{u}_D - \vec{u}_T) \cdot c_i \quad (4.4)$$

Here the diffusive transport is represented by the first term, whereas the second term includes all drift velocities, i.e., \vec{v} , \vec{u}_D and \vec{u}_T as convective, diffusiophoretic and thermophoretic velocity, respectively.

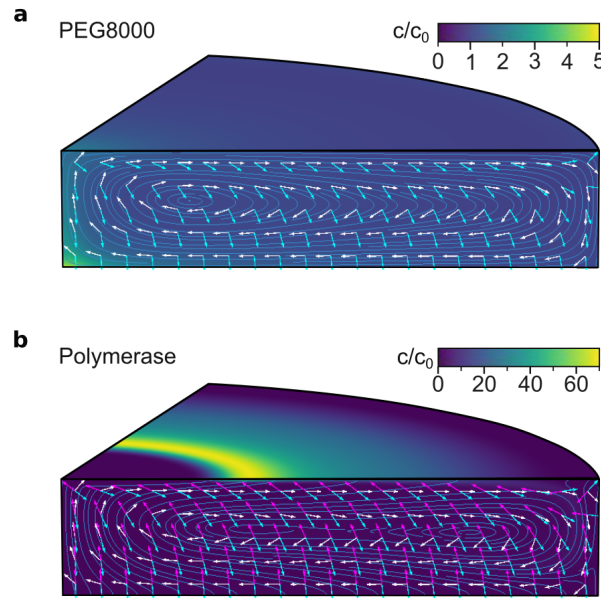


Figure 4.7: Accumulation profiles derived by finite-element simulations including diffusiophoresis:

The convective, thermophoretic, and diffusiophoretic movements are represented by the white, cyan, and magenta arrows, respectively. It is important to note that the length of the arrows is scaled for better visibility and serves only as an indicator of the respective directions of the velocity profiles. The superposition of all velocity profiles is depicted by the light blue streamlines. **a** The simulation of the accumulation behavior of PEG8000 in the convection chamber demonstrates a similar pattern to the long 210mer DNA. After 60 minutes, PEG exhibits a concentration increase of 4.5-fold at the bottom center of the compartment. **b** Only in the case of the slow diffusing conglomerates diffusiophoresis emerges as a significant driving force that opposes the thermophoretic movement. By not only including diffusive, convective, and thermophoretic but also the diffusiophoretic transport in the finite-element simulation, the ring-shaped concentration enhancement of the polymerases, as observed in the experiments, was captured by the simulated results. After 60 minutes, the initial concentration exhibits a maximum increase of 66-fold in a ring-shaped pattern. The streamlines illustrate the molecular movement towards the wall, forming the ring-shaped accumulation as observed in the experiments.

Following Madea et al. [195], the diffusiophoretic velocity \vec{u}_D is given by:

$$\vec{u}_D = \frac{0.5k_B T}{3\eta/\mu} \cdot \left(S_{T_{\text{PEG}}} - \frac{1}{T} \right) \cdot r_{G_{\text{PEG}}}^2 \cdot \frac{N_A \cdot c_i \cdot c_{0_{\text{PEG}}}}{c_{0_i}} \quad (4.5)$$

Here, $r_{G_{\text{PEG}}}$ is the radius of gyration of PEG8000, and η/μ is the corrected viscosity of water to account for changes induced by the buffer, c_{0_i} and $c_{0_{\text{PEG}}}$ are the initial concentrations of species i , i.e. the species solved for, and PEG, respectively. For more details of the simulation, refer to section 4.5.4.

By including the above changes in the molecular flux of the finite-element simulation, the ring-shaped accumulation region of the conglomerated RNA or DNA polymerases was successfully predicted. In contrast to the 35mer and 210mer (see Figure 4.15), diffusiophoresis played a significant role in the movement of the conglomerates in the temperature gradient, now pointing towards the heat source. However, this inverted force only has an effect near the boundary walls where the flow velocity does not dominate over the slower diffusion of the conglomerates. Interestingly, this led to an accumulation in a ring-shaped pattern, away from the hot temperature spot at the top chamber wall (see Figure 4.7 b).

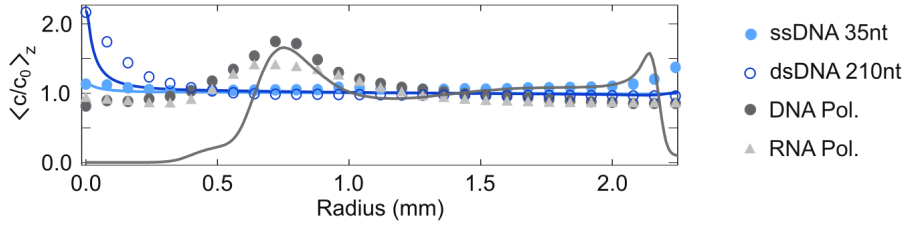


Figure 4.8: Concentration height average of nucleic acids inside the convective chamber: The height average $\langle c/c_0 \rangle_z$ of the simulated relative concentrations (lines) reproduced the experimental fluorescence signal (symbols) for all species. The data points represent the radial averaging of the background-corrected fluorescence images with respect to the central point of the chamber.

For a quantitative comparison between the experiment and simulation of the accumulation behavior for all species, the height average of the concentration $\langle c/c_0 \rangle_z$ along the radius was extracted (see Figure 4.8). This approach was chosen, as the fluorescence data of the experiment corresponds to an averaged concentration signal along the height. First, the mid-point of the chamber was identified to simplify the two-dimensional fluorescence data to a one-dimensional signal. The fluorescence was then plotted outwards along the radius and averaged. As the simulation was performed in a 2D-axisymmetric compartment, the concentration just had to be averaged over the height.

Introducing crowding agents to the buffer and their emergent concentration gradient leads to a dependence of the nucleic acids' accumulation behavior on the binding details of the molecules. This stems from a strong size dependence of the diffusiophoresis. The unsaturated, single-stranded parts of the folded polymerase molecules could potentially engage

in intermolecular interactions and therefore lead to the formation of conglomerates. This hypothesis is supported by the stark difference between double-stranded DNA and the single-stranded DNA analog of the polymerase sequence, as well as the reversible melting and formation of the conglomerates upon temperature changes.

4.2.5 Probability density of Polymerase and RNA degradation

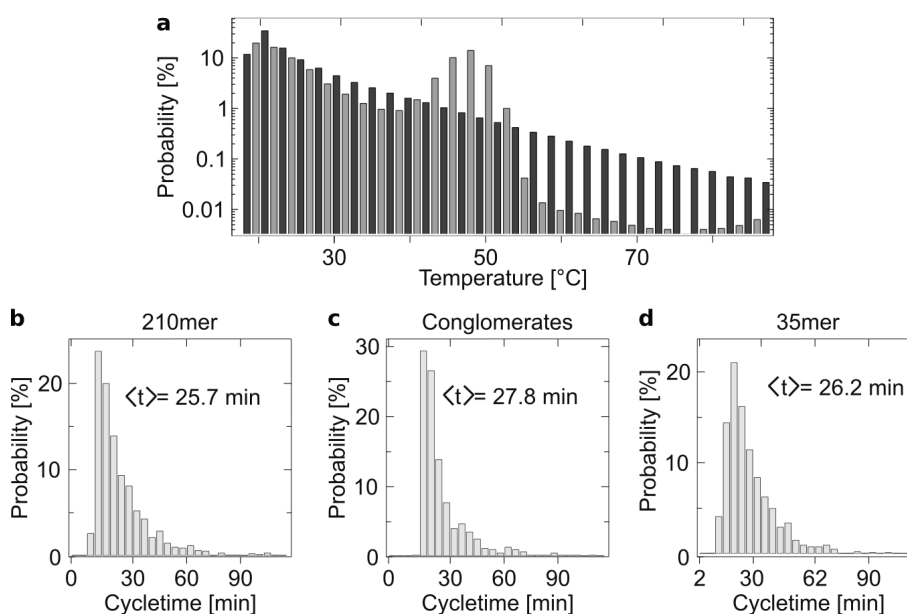


Figure 4.9: Cycling characteristics derived for nucleic acids from stochastic trajectories: The statistics are each obtained from 400 simulated trajectories corresponding to a total time span of 2 h. A full cycle is defined as a trajectory, starting from temperatures below 20 °C to temperatures above 60 °C and back to 20 °C. **a** The random walk simulations conducted for the conglomerates showed no trajectories that completed the full temperature cycle with a spike temperature exceeding 60 °C. Instead, the cycling times considered in these simulations involved trajectories between temperatures below 20 °C and temperatures above 40 °C. While the conglomerates exhibited a higher probability of residence around 45 °C than the 35mer, the probability decreased significantly for temperatures above 60 °C. **b**, **c** and **d** show the distribution of cycling times for the 210mer dsDNA, the conglomerates and the 35mer, respectively. The mean cycle times are between approximately 26 to 28 min. Again, the cycle times for the conglomerates correspond to temperature cycles between <math><20 \text{ °C}</math> and $>40 \text{ °C}$.

Apart from running the replication in a plausible setting, an additional goal was to selectively protect long RNA from spontaneous cleavage. As the cleavage reaction is enhanced at elevated temperatures, the length-dependent accumulation of the polymerase away from the high-temperature spot could lead to an enhanced protection. However, this is *a priori* not clear, as the accumulation spot exhibits intermediate temperatures of around 45 °C. This

section focuses on details of the temperature and cycling characteristics of the different nucleic acid types, as well as the impact of those characteristics on their potential cleavage.

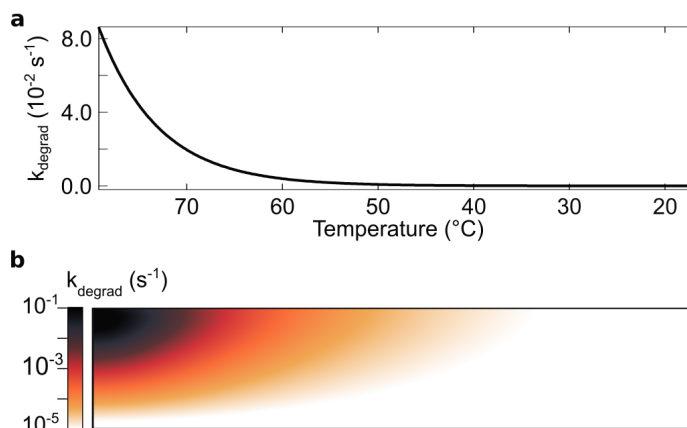


Figure 4.10: Temperature-dependent RNA cleavage rate: **a** shows the rate for RNA cleavage as implemented in the finite-element simulation. It exhibits a strong temperature dependence, with a minimal rate of $k_{\text{degrad}} = 3.81 \cdot 10^{-6} \text{ s}^{-1}$ at $17 \text{ }^\circ\text{C}$ and a maximal rate of $k_{\text{degrad}} = 0.16 \text{ s}^{-1}$ at $83.2 \text{ }^\circ\text{C}$, respectively. Note that the actual values have to be taken with caution as this corresponds only to degradation through transesterification. However, the overall trend of degradation should be reliable. **b** depicts the linear representation of the RNA cleavage rate in the 2D axisymmetric compartment of the finite-element simulation.

To this end, the trajectories of 400 particles with random starting positions were simulated similarly as described in section 4.2.1. This provided valuable insights into the statistics for the temperature distribution and cycling times for the polymerases, double-stranded DNA, and single-stranded DNA, respectively (see Figure 4.9 and Figure 4.16). A complete cycle was defined as a trajectory starting at temperatures below $20 \text{ }^\circ\text{C}$, moving to a temperature region above $60 \text{ }^\circ\text{C}$, and then back to $20 \text{ }^\circ\text{C}$. Considering the flow profile from the finite-element simulation and Brownian motion, the simulations revealed that the ring-shaped accumulation effectively maintained the conglomerates at a temperature of around $45 \text{ }^\circ\text{C}$. This ensured that the polymerases effectively remained below $60 \text{ }^\circ\text{C}$. More specific, the random walk simulations for the conglomerates yielded no trajectories, completing the full temperature cycle with spike temperatures of above $60 \text{ }^\circ\text{C}$. Instead, for comparison of the cycling times for all nucleic acid species, the trajectories for the conglomerates were analyzed for cycles between temperatures of below $20 \text{ }^\circ\text{C}$ and above $40 \text{ }^\circ\text{C}$. In contrast, molecules that did not form conglomerates had a higher probability of residing at the lowest temperatures but were frequently exposed to temperatures above $60 \text{ }^\circ\text{C}$.

For a more quantitative analysis, rates of RNA degradation were included in the finite-element simulation. According to Li and Breaker [152], the rates of RNA degradation due to transes-

terification can be predicted as follows:

$$k_{\text{degrad}}^{\text{1bond}} = k_{bg} \cdot 10^{0.983(\text{pH}-6)} \cdot 10^{-0.24(3.16-[K^+])} \cdot 10^{0.07(T-23)} \cdot 69.3 \cdot [Mg^{2+}]^{0.80} \cdot 3.57[K^+]^{-0.419} \quad (4.6)$$

where $k_{bg} = 1.3 \cdot 10^{-9} \text{ min}^{-1}$ is the background rate for pH 6, a potassium concentration of $[K^+] = 3.16 \text{ M}$ and at a temperature of 23 °C. The experimental values of $[Mg^{2+}] = 0.05 \text{ M}$, $[K^+] = 0.05 \text{ M}$, and pH 8.3 were assumed to be stationary. The rate for a cleavage reaction in at least one position along the entire length of an RNA molecule can be estimated by

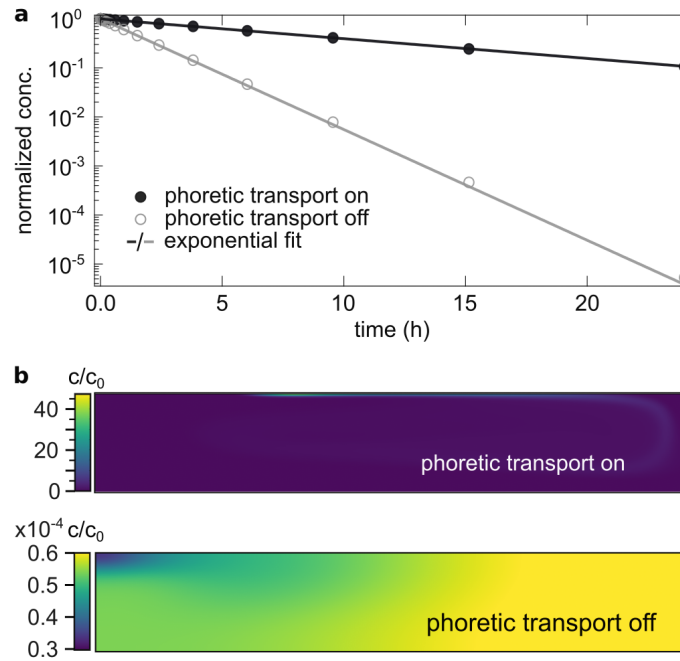


Figure 4.11: Simulation results for the effect of the temperature-dependent RNA cleavage on the concentration of the polymerase: **a** The finite-element simulation showed an exponential degradation of the surfaced averaged ribozyme concentration for both cases, with and without phoretic transport. However, fitting an effective degradation rate yielded a 5-fold more stable ribozyme degradation rate for the case including phoretic transports, as they are accumulating the ribozyme conglomerates away from the high temperature spot. Such mechanisms might have played an important role for a selection pressure towards more structured molecules. **b** shows the concentration profiles at $t=60 \text{ min}$ for both cases, with and without phoretic transports. In the case without phoretic transports, no accumulation is observed, and there are only small differences in the absolute concentration values. However, a significant fraction of the polymerase is constantly exposed to high temperatures, resulting in a higher effective cleavage rate. On the other hand, the ring-shaped aggregation behavior resulting from phoretic transports is preserved, leading to the observed accumulation and improved stability.

$$k_{\text{degrad}} = k_{\text{degrad}}^{\text{1bond}} \cdot (N - 1) \quad (4.7)$$

with N being the number of total nucleotides of the RNA (see Figure 4.10). The resulting temperature-dependent degradation rate was then included as an additional term in the modified convection-diffusion equation (see equation 4.3) of the finite-element simulation. The full equation then reads:

$$\nabla \cdot (-D_i \nabla c_i - S_{T_i} \cdot D_i \cdot \nabla T \cdot c_i + (\vec{v} + \vec{u}_D) c_i) = R_i \quad (4.8)$$

with $R_i = -c_i \cdot k_{\text{degrad}}$. To investigate whether the phoretic transports lead to the protection of the polymerase from degradation, the simulation was run following equation 4.8. Additionally, an artificial case without any phoretic forces, described by

$$\nabla \cdot (-D_i \nabla c_i + \vec{v} \cdot c_i) = R_i \quad (4.9)$$

was analyzed for comparison. In both cases, at each time point, the surface integral over the resulting concentration profile was performed to evaluate the total remaining concentration (see Figure 4.11). Both cases showed an exponential degradation of the RNA polymerase. However, including the phoretic transports resulted in a 5-fold reduced degradation rate compared to the case of no phoretic transports, i.e., frequent exposure to the high-temperature spot.

Although the conglomerates of the polymerase were not accumulated to the coldest regions of the convection chamber, the accumulation due to the phoretic transport led to some protection from temperature-induced degradation. Similar combinations of transport mechanisms could have implied an additional selection pressure towards more structured molecules that are required for catalytic activity.

4.3 Discussion

What kind of environment could have supported the emergence of life through an RNA-based system? So far, extensive research has focused on investigating the synthesis of RNA using an RNA polymerase ribozyme [123, 125, 139, 143, 193]. Still, most of these reactions have yet to be conducted in a setting that is plausible for prebiotic conditions. In previous experiments using a different non-equilibrium approach, RNA-catalyzed RNA polymerization has been performed in eutectic ice, which concentrates the reactants and reduces spontaneous RNA cleavage [82]. In future iterations of thermal habitats, incorporating such ice phases could help to reduce the overall salt concentrations while still enabling thermal strand separation and long-term localization through thermal convection and thermophoresis.

One major challenge in the replication of RNA is the spontaneous cleavage that occurs at the elevated temperatures necessary for separating template and product strands. This limits the formation and preservation of longer RNA molecules. However, longer RNAs would have been crucial on the early Earth to support robust enzymatic activities [126, 196]. The thermal habitat described in this study provides temperature conditions that drive RNA-catalyzed RNA replication, achieving similar replication kinetics to optimized bulk reactions

performed in standard thermal cyclers. An interesting aspect of the thermal habitat is its length-dependent accumulation mechanism, which favors longer and more structured RNAs. This mechanism helps to counter the threat of short parasitic sequences that tend to be copied more rapidly. In this habitat, shorter RNAs have a higher probability of exposure to elevated temperatures, where degradation is enhanced (see Figure 4.9 and 4.16) [183].

However, achieving strand separation only by means of temperature will always be limited. Already for 50mers, the melting temperature is close to the boiling point of water at monovalent salt concentrations of 500 mM in the bulk². Employing other physio-chemical conditions, such as periodic changes in salt concentrations or pH, holds several potential advantages [83, 84]. First, a combination of low pH and elevated temperature lowers the melting temperature. This unlocks the potential for longer sequences to be separated. Second, using pH or salt cycling reduces the risk of cleavage of long RNA and, thus, prolongs the lifetime of the ribozymes. This allows for overall longer reaction times and could pave the way towards a broader application of molecular evolution to RNA replication.

The experiments conducted in the thermal habitat suggest the existence of selective pathways that could guide RNA evolution toward longer and more structured sequences. This environment may provide a means to replicate not only 35-nucleotide strands but also complete RNA polymerases assembled from shorter component strands [82, 142, 143]. However, how RNA molecules longer than 200 nucleotides could have emerged from simple, non-enzymatic replication chemistries in a similar setting remains an intriguing open question.

4.4 Conclusion

The quest to uncover how life emerged involves identifying a suitable environment where informational molecules can replicate and undergo open-ended evolution despite challenges such as dilution and spontaneous degradation. The experiments described in this chapter show how a temperature gradient across a porous rock pore of millimeter scale, filled with water, could have driven both the accumulation of long RNA away from high temperatures as well as the RNA-catalyzed replication of RNA molecules. This process of strand separation and RNA accumulation is driven by thermal energy, an abundant energy source on and inside a plausible scenario for the geological conditions of the early Earth. It is worth noting that the employed form of the RNA polymerase ribozyme has limitations in replicating longer RNA strands. However, the observed thermal selection bias towards longer RNA molecules in this system holds the potential to guide molecular evolution towards greater complexity. With work being put into evolving more potent ribozymes, the next iterations of RNA-replication facilitated by thermal non-equilibrium settings hold great potential to understand the emergence of replication and eventually life.

²This is based on calculations provided from the OligoCalc biotool [] for the sequence ATG CAT GCA TGC ATG CAT GCA TGC ATG CAT GCA TGC ATG ATG CAT GCA TG with a GC content of 48%. Increasing the GC content would increase the theoretical melting temperature even further. Additionally, no divalent ions are considered in the calculation, which again leads to a strong increase in duplex stability and, thus, melting temperature.

4.5 Materials and Methods

4.5.1 Nucleic acids

Table 4.1: Nucleic acid sequences used in the experiments. The RNA sequences for the ribozymatic replication of RNA are highlighted in light grey. In the accumulation experiments, the 35 nt long template sequence, the 24-3 polymerase and its DNA sequence analog were used. For the accumulation experiment of the 210 nt long dsDNA, the sequence was amplified from the pBR322 plasmid via a standard PCR.

Forward Primer	5' GACAAUGAC AAAAAA GCAACUUUUC 3'
Reverse Primer	5' GACAAUGAC AAAAAA CACUCCACAC 3'
rPCR Template	5' GACAAUGAC AAAAAA CACUCCACAC GAAAAGUUGC 3'
24-3 Polymerase	5' AGU CAU UGC CGC ACG AAA GAC AAA UCU GCC CUC AGA GCU UGA GAA CAU CUU CGG AUG CAG AGG AGG CAG CCU UCG GUG GAA CGA UCG UGC CAC CGU UCU CAA CAC GUA CCC GAA CGA AAA AGA CCU GAC AAA AAG GCG UUG UUA GAC ACG CCC AGG UGC CAU ACC CAA CAC AUG GCU GAC 3'
dsDNA	5' TCG GTG TAG GTC GTT CGC TCC AAG CTG GGC TGT GTG CAC GAA CCC CCC GTT CAG CCC GAC CGC TGC GCC TTA TCC GGT AAC TAT CGT CTT GAG TCC AAC CCG GTA AGA CAC GAC TTA TCG CCA CTG GCA GCA GCC ACT GGT AAC AGG ATT AGC AGA GCG AGG TAT GTA GGC GGT GCT ACA GAG TTC TTG AAG TGG TGG CCT AAC TAC GGC 3' 5' GCC GTA GTT AGG CCA CCA CTT CAA GAA CTC TGT AGC ACC GCC TAC ATA CCT CGC TCT GCT AAT CCT GTT ACC AGT GGC TGC TGC CAG TGG CGA TAA GTC GTG TCT TAC CGG GTT GGA CTC AAG ACG ATA GTT ACC GGA TAA GGC GCA GCG GTC GGG CTG AAC GGG GGG TTC GTG CAC ACA GCC CAG CTT GGA GCG AAC GAC CTA CAC CGA 3'

4.5.2 RNA-catalyzed replication of RNA

The 24-3 polymerase ribozyme that was used in the experiments was obtained by *in vitro* transcription of double-stranded DNA (20 µg/mL), following the method described by Horning et al. [125]. The RNA oligomers, i.e., the polymerase, primers, and template, were purified using denaturing polyacrylamide gel electrophoresis (PAGE) and subsequent ethanol precipitation. To anneal the RNA, it was heated to 95 °C for 30 s and then slowly cooled down to 4 °C at a rate of 0.2 °C/s in a buffered solution containing 10 mM Tris at pH 8, 1 mM EDTA and 0.05 % TWEEN20. The annealed RNA was mixed with the reaction buffer and nucleoside triphosphates (NTPs) to prepare the final reaction mixture. The resulting mixture contained 400 nM of the 24-3 polymerase ribozyme, 200 nM primer, 4 mM NTPs, and varying amounts of the template. The reaction buffer consisted of 50 mM Tris (pH 8.3), 50 mM MgCl₂, 6% w/v PEG8000, 0.9 M TPA, and 0.05% Tween20. The primers were 5'-biotinylated and labeled with fluorescein for detection by polyacrylamide gel electrophoresis (PAGE) analysis. To quench

the reaction, 0.5x volumes of 500 mM EDTA (pH 8) were added. The biotinylated primers and extended products were captured using Streptavidin C1 Dynabeads (ThermoFisher Scientific, USA). The captured materials were washed four times with alkali solution (25 mM NaOH, 0.05 % TWEEN20, 1 mM EDTA). Then it was washed two times with a mixture of 8 M urea, 1 mM EDTA, 0.05 % TWEEN20, and 10 mM Tris (pH 8.0). For PAGE analysis, the materials were mixed with 98% formamide and 10 mM EDTA, and then heated at 95 °C for 10 minutes. After heating, the samples were separated using a 12.5% polyacrylamide gel.

The products were analyzed using CCD photography (Orca 03-G, Hamamatsu, Japan) through a green bandpass filter (520 nm, 10 nm FWHM, Newport, Germany) with a spectrally filtered source (LED 470 nm, filter 470 nm, 10 nm FWHM, Thorlabs, Germany) and high-quality interference filters (bandpass 692±20 nm, OD 6 blocking, Edmund Optics, USA; bandpass 700±35 nm, OD 2 blocking, Newport, USA) using a filtered source (LED 625 nm, filter 630 nm, 10 nm FWHM, Thorlabs, Germany) for fluorescein- and Cy5 labels, respectively. The bulk control experiments were conducted in a thermocycler with enhanced heating/cooling rates (Analytik Jena AG, Germany).

4.5.3 Convection system

The cylindrical convection chamber was constructed from a 500 µm thermally conductive soft-silicone film, with cut out holes of approximately 5 mm diameter (KU-TCS, Aavid Kunze GmbH, Germany). The bottom was sealed by a silica wafer (Si-Mat-Silicon Materials e.K., Germany) with a thickness of 525±25 µm and a 100 nm SiO₂ coating. The top was closed with a 170 µm borosilicate glass coverslip (Carl Roth, Germany). The stable temperature gradient was generated using an IR laser (TLR-30-Y12, IPG Laser GmbH, Germany) and Peltier element (Uwe Electronic Vertriebs GmbH, Germany) as a heat source and heat sink, respectively. A PID-controlled feedback loop was used to regulate the temperature of the Peltier element in conjunction with a water bath (CF41 Kryo-Kompakt-Thermostat, Julabo, Germany).

The accumulated nucleic acids were fluorescently imaged with a long working distance 2x objective (Mitutoyo Plan Apo Infinity 2x, 0.055 NA, Mitutoyo Corporation, Japan), equipped with a CCD camera (Stingray F-145B, Allied Vision Technologies GmbH, Germany) and illuminated with two alternating LEDs (625 nm and 470 nm, Thorlabs, Germany) in combination with a dual-band filter set (fluorescein and Cy5, AHF, Germany). Up to four sample chambers could be run in parallel employing scanning mirrors (6200-XY, Cambridge Technology, England) to sequentially direct the IR-laser onto one of the sample chambers. A cold dichroic mirror (transmission 400–700 nm, reflection 633/1940 nm, AHF, Germany) coupled the IR-laser into the optical path between the sample and objective. The fluorescence of 50 µM 2',7'-bis-(2-carboxyethyl)-5-(and 6)-carboxyfluorescein (BCECF), diluted in 10 mM Tris (pH 8.0) Was used to measure the experimental temperature profile. It was calibrated with the temperature-dependent fluorescence by uniformly varying the temperature of the heat bath.

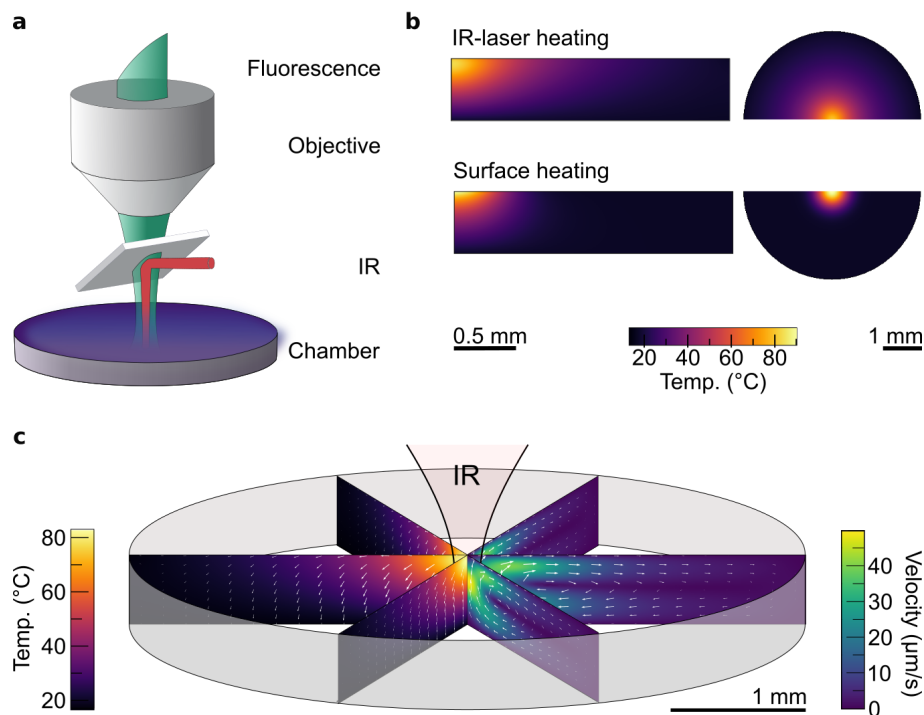


Figure 4.12: Experimental realization of the convection chamber with central heating: **a** A focused IR-laser is used to induce a radial symmetric heat profile inside a cylindrical convection chamber. **b** Finite-element simulations of centralized heating using an IR-laser (top) or surface heating (bottom), with similar temperature profiles emerging in the two cases. **c** The finite-element simulation shows that absorption of the focussed IR laser in the upper layers of the water-filled cylindrical compartment generates convective motion in the fluid. Within the temperature profile (left) and velocity profile (right), the white arrows represent the thermophoretic and convective velocities, respectively. It is important to note that the length of the arrows is adjusted for improved visibility and does not directly represent the relative strengths of the velocity fields.

4.5.4 Finite-element simulations

The finite-element simulations are performed in an axisymmetric compartment with a rectangular cross-section using Comsol Multiphysics 5.4, making use of the radial symmetry of the system in order to save computational time. The experiments are performed in a cylindrical compartment with a punctual heat source that forms a temperature gradient, pointing outwards from the given heat spot. Measuring the experimental temperature profile was done using the temperature-dependent fluorescence decrease of BCECF in Tris (see section 4.5.3). The readout of the temperature measurement equals the average over the height of the chamber. To correctly model the experimental situation, a corresponding height average of the 2D temperature calculation was implemented in Comsol. The modeled heating power and beam width of the laser were then adjusted such that numerical and experimental temperature profiles matched within a 20 % and 30 % error, respectively. Instead of running the simulation in a 3D compartment, the axisymmetric nature of the experimental setting al-

lowed the reduction of the following partial differential equations to the 2D case. This led to an increase in computational speed while still capturing the full details of the 3D experiment. The temperature profile that was obtained from the comparison to the experiment was fed into the numerical model, which then couples the contributing transport mechanisms in the following steps:

- (i) The temperature profile is determined by applying partial differential equations for transient heat transfer, incorporating the material properties of water, silicon, silicone, and glass. In the case of water, the temperature dependence of the contributing parameters is included in the modeling. The absorption of the Gaussian-shaped IR laser is modeled using the Beer-Lambert law, and the momentum of the absorbed photons is considered in the force balance of the Navier-Stokes equation. This inclusion accounts for a potential slowdown of convection. The absorption coefficient for the IR light with a wavelength of 1940 nm is calculated as 129.6 cm^{-1} based on published material constants.
- (ii) The convective flow profile inside the thermal convection chamber is determined by numerically solving the incompressible Navier-Stokes equation. The buoyancy force is calculated based on the temperature-dependent density of the fluid. Non-slip boundary conditions and a zero-pressure point boundary are imposed in the simulation. To account for the enhanced viscosity of the rPCR buffer caused by the addition of PEG, the viscosity was experimentally measured using the Brownian motion of fluorescently labeled beads with a size of $1 \mu\text{m}$. At a temperature of $20 \text{ }^\circ\text{C}$, the viscosity of the rPCR buffer is determined to be 28 mPas .
- (iii) To simulate the accumulation of the 35mer, 210mer, PEG8000, and conglomerates, a modified convection-diffusion equation is solved using the simulated flow and temperature profile as input. The thermophoresis and diffusiophoresis are implemented as an additional molecular flow. The boundaries are given as non-adsorbing. The total molecular flux (\vec{j}_i) for a nucleic acid species i arises from the combination of diffusion, thermophoresis, and convection and is stated in equation 4.2. The fully modified equation with the diffusiophoretic flow is given by equation 4.3.

The Soret coefficients S_{T_i} that were used in the simulation in order to match the experimental accumulation profile for the 35mer and the 210mer are given by 0.03 K^{-1} and 0.2 K^{-1} , respectively. Those values are about a factor of 3 higher than stated in the literature, although with different salts but at a comparable Debye length. However, deviations are somewhat expected for PEG-mixtures with high concentrations of MgCl_2 and 0.9 M TPA . This might be due to changes in the Seebeck effect, which dominates the thermophoresis in this regime, as many of the electrophoretic mobilities that determine the buildup of the electric field in the Seebeck effect, are not known. Therefore, changing the Soret coefficients within this range appears to be a plausible approach. The Soret coefficient used for the RNA and DNA polymerase conglomerates was assumed to be 0.3 K^{-1} , which is only slightly higher than the Soret coefficient for the 210mer. However, it should be noted that the diffusiophoretic forces dominate the ring formation and thermophoresis is a rather minor effect.

To model the diffusion of the different nucleic acid species, the diffusion coefficients were

adapted according to:

$$D_i = 0.5 \cdot 643 \cdot n_i^{-0.46} \cdot \frac{D_{\text{PEG}_{\text{factor}}}}{D_{\text{Viscosity}_{\text{factor}}}} \quad (4.10)$$

as reported by Mast *et al.* [165] for dsRNA in salty conditions, where n denotes the length of the strand. The slower diffusion resulting from the enhanced viscosity of a thermophoretically enhanced PEG concentration was modeled with $D_{\text{PEG}_{\text{factor}}}$.

4.5.5 Random walk model

Random walk simulations were implemented in a Labview code to derive statistical data on the thermal cycles for the RNA oligonucleotides inside the convection chamber. To gather the statistics, the trajectories of 400 particles were considered, with a total time frame of 2 hours and time-steps of $\Delta t = 1$ ms. The displacement of particles $\Delta \vec{s}(x,y)$ is stated in equation 4.1. The flow profile $\vec{v}(x,y)$ used in the simulation was derived from the finite-element simulations as described in the previous section. The Brownian motion was implemented by a randomly directed movement $\eta(t)$. Molecules are defined to undergo a complete temperature cycle as they move from a high-temperature region of above 60 °C for denaturation to a low-temperature region of below 20 °C for elongation and then back to the high-temperature region. For the polymerase conglomerates, a full temperature cycle is defined with hot temperatures above 40 °C, as 60 °C does not yield any completed temperature cycles due to the efficient accumulation. It is important to note that the random walk model used in the simulations does not take into account minor changes in molecule size that may occur during the experiment as a result of partial elongation scenarios by the polymerase.

4.5.6 Sigmoidal model for exponential replication

The exponential replication of RNA can be modeled effectively by a sigmoidal growth model [197], with the product concentration c being dependent on the number of cycles following:

$$c(\text{cycle}) = \frac{c_{\text{max}}}{1 + e^{-\frac{\text{cycle} - \text{cycle}_{1/2}}{k}}} \quad (4.11)$$

Here c_{max} denotes the final maximal concentration, k denotes the replication factor, cycle denotes the number of cycles, and $\text{cycle}_{1/2}$ denotes the cycle factor corresponding to the midpoint of the growth curve. The cycle factor is given by:

$$\text{cycle}_{1/2} = k \cdot \ln \left(\frac{c_{\text{max}}}{c_0 - 1} \right) \quad (4.12)$$

The model exhibits an initial exponential growth phase for $c \ll c_{\text{max}}$, which is then later inhibited. This can be attributed to a depletion of primers and NTPs, a decline in polymerase ribozyme activity and the presence of pyrophosphates in the reaction. In the bulk reaction, the number of cycles corresponds to 50, leading to the dependence of the final concentration c_{max} on the initial template concentration c_0 of:

$$c_{50}(c_0) = \frac{c_{\text{max}}}{1 + e^{\ln \left(\frac{c_{\text{max}}}{c_0 - 1} \right) - \frac{50}{k}}} \quad (4.13)$$

Equation 4.13 was used to fit the results shown in Figure 4.4. The replication efficiency E is given by $E = c_N/c_{N-1}$ and describes the increase of the concentration c between two consecutive cycles cycle_N and cycle_{N-1} . In the case of the initial exponential phase, i.e., $c \ll c_{max}$, the expression for E simplifies to $E = e^{1/k}$, which was used to calculate the stated values for convection and bulk, respectively.

4.6 Supplementary Figures

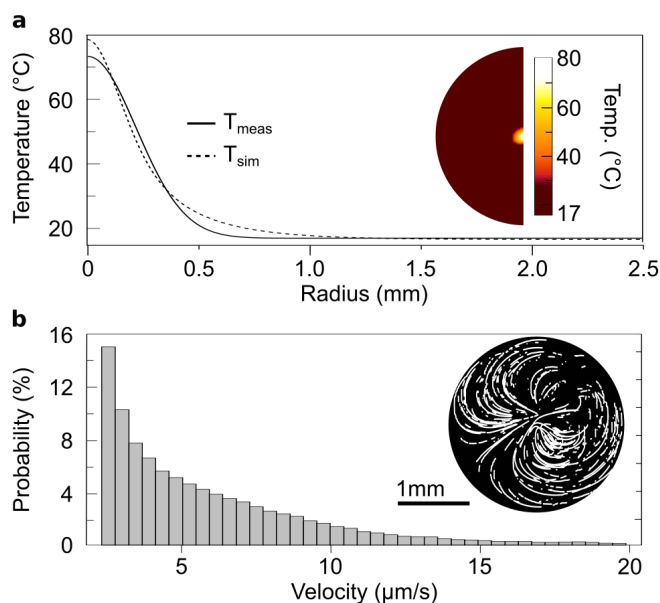


Figure 4.13: Local heating of the convection chamber: a The local heating of the compartment is achieved by an IR-laser, creating a Gaussian profile that induces a temperature difference between the heated regions (80 °C) and the water-bath cooled (17 °C) bottom side of the convective chamber. The radial temperature profile is determined by measuring the temperature-dependent fluorescence of BCECF (50 μM) diluted in TRIS buffer (10 mM). Experimental measurements match the results obtained from a finite-element simulation, which takes into account the axial symmetry of the system. The scale bar corresponds to 1 mm. **b** The established temperature gradient leads to a convective flow within the chamber. This flow effectively transports fluorescently labeled beads with a size of 1 μm between the inner and outer regions of the chamber. The highest velocities, exceeding 20 $\mu\text{m/s}$, are observed in the heated regions of the chamber.

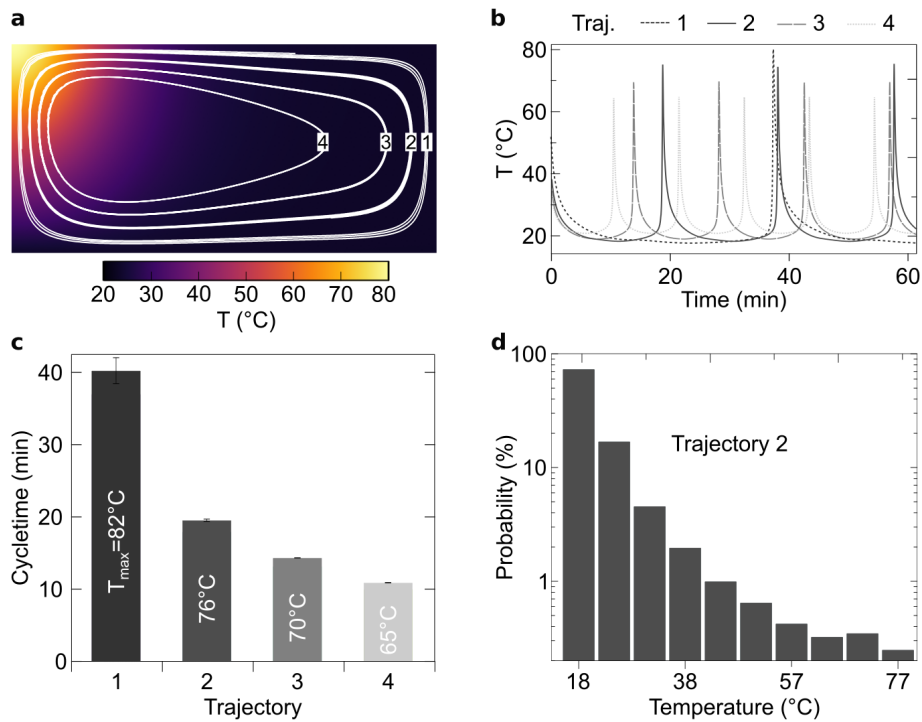


Figure 4.14: Temperature cycle conditions derived from finite-element simulation: **a** Exemplary trajectories of macro molecules solely driven by thermal convection. Various thermal scenarios were screened to find a scenario that matches the delicate temperature conditions necessary for RNA amplification. For this purpose, a finite-element simulation derives temperature cycles of molecules with respect to their initial starting position, neglecting thermophoretic and Brownian motion. **b** The temperature oscillation inside a chamber of 0.5 mm in height and 2.25 mm radius matches the temperature cycle requirement of an RNA-only replication system. Here, molecules undergo fast temperature changes while remaining for most of the time at low temperatures. **c** Simulating 400 trajectories of random starting positions yields a mean cycle time of 19.0 ± 10.3 min between 20 °C and 60 °C. More excessive temperature oscillations, such as temperature differences between 20 °C and 80 °C, require extended cycle times of 55.8 ± 18.0 min. **d** In these thermal habitats, the molecules are convectively shuttled between regions of varying temperatures but remain at low temperatures most of the time.

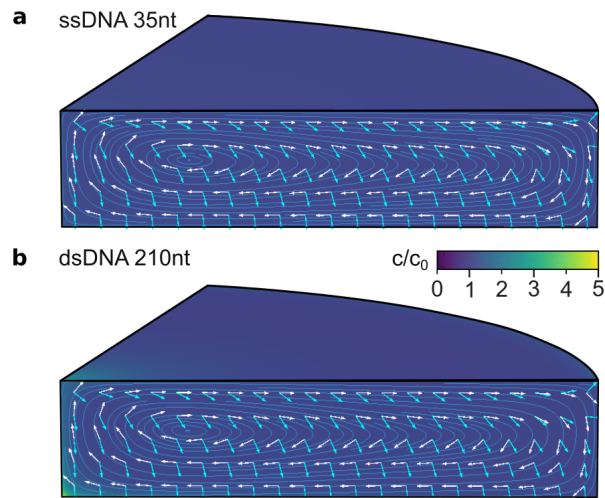


Figure 4.15: Accumulation profiles of short and long DNA derived by finite-element simulations: The arrows in white, cyan and magenta represent the direction of the convective, thermophoretic and diffusiophoretic movement, respectively. The length of the arrows is scaled for better visibility and only serves as an indicator for the respective directions of the various velocities. The light blue streamlines correspond to the superposition of all velocity profiles. The simulation compartment had a height of 0.5 mm and a radius of 2.25 mm. **a** The 35 nt ssDNA demonstrates minimal accumulation with only a 1.33-fold increase in concentration after 60 minutes. **b** The 210 nt dsDNA exhibits significant accumulation, showing a 5-fold increase in concentration after the same period, specifically at the bottom center of the compartment.

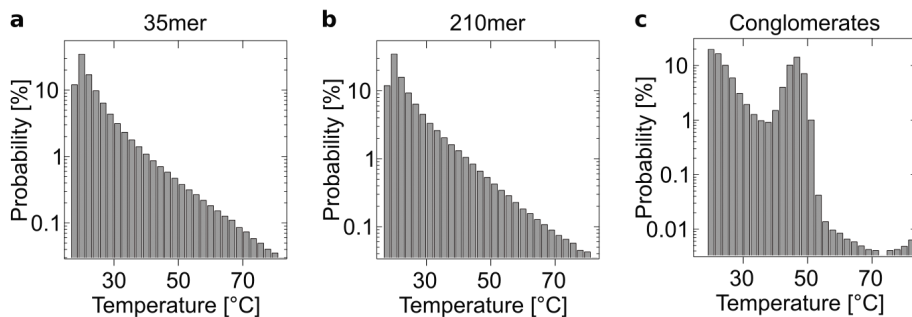


Figure 4.16: Temperature characteristics derived from stochastic cycling trajectories: The statistics are obtained from a total of 400 simulated trajectories corresponding to a total time span of 2 h. **a**, **b** show the cycle statistics for a 35mer ssRNA and a 210mer dsDNA, respectively. A cycle refers to a trajectory that begins at temperatures below 20 °C, moves to temperatures above 60 °C, and returns to temperatures below 20 °C. The strands primarily reside at lower temperatures and undergo rapid cycling through the high-temperature spots. **c** The random walk simulations for the conglomerates do not exhibit trajectories that complete the full temperature cycle with a spike temperature exceeding 60 °C. The resulting temperature distribution reveals a secondary peak around 45 °C, which corresponds to the temperature at the accumulation region.

5 RNA synthesis and replication by the *sunY* ribozyme driven by a non-equilibrium microenvironment

Summary

Enzyme-catalyzed replication of nucleic acid sequences is a prerequisite for the survival and evolution of biological entities. Before the advent of protein synthesis, genetic information was most likely stored in and replicated by RNA. However, experimental systems for sustained RNA-dependent RNA replication are difficult to realize, in part due to the high thermodynamic stability of duplex products and the low chemical stability of catalytic RNAs. Using a derivative of a group I intron as a model for an RNA replicase, it is shown that heated air-water interfaces that are exposed to a CO₂-rich atmosphere enable sense and antisense RNA replication as well as template-dependent synthesis and catalysis of a functional ribozyme in a one-pot reaction. Both reactions are driven by autonomous oscillations in salt concentrations and pH, resulting from the precipitation of acidified dew droplets, which transiently destabilize RNA duplexes. The results suggest that a microenvironment mimicking an abundant geo-physical setting on Hadean Earth could have promoted both replication and synthesis of functional RNAs¹.

¹This chapter is based on the publication by Salditt et al. [167] in Nature Communications.

Authors: Annalena Salditt, Leonie Karr, Elia Salibi, Kristian Le Vay, Dieter Braun and Hannes Mutschler
Contributions: Project design, funding acquisition, and supervision: D.B., H.M. Experiments and data analysis: A.S., L.K., E.S., K.L.V. Visualization: A.S., E.S. Paper writing: A.S., L.K., E.S., K.L.V., D.B., H.M.

5.1 Introduction

5.1.1 The replication problem in the RNA world

In the search for the origin of life, finding replication mechanisms relying only on RNA is of particular interest as RNA encodes sequential information and folds into enzymatically active structures (ribozymes) [18, 19, 198]. Specifically, identifying a replication scenario for ribozymes leading to general copying of RNA as well as autocatalytic copying of itself would be an immense advancement in understanding the emergence of replication. Over the past decades, re-engineering introns and *in vitro* evolution has led to the discovery of several catalytic RNAs, capable of accelerating RNA ligation and polymerization [126, 135, 142, 199, 200]. Although this tremendously progressed the field of protein-free RNA replication, full replication scenarios are difficult to realize.

As catalytic activity is coupled to long and more structured sequences [201], it is specifically difficult to find a polymerase or ligase ribozyme that is able to copy itself or ribozymes of similar function. As a possible solution, it was shown that fragmenting ribozymes into multiple sub-sequences can still lead to assemblies that maintain their function [82, 126]. In promising recent approaches, ligases and polymerases can successfully copy a template strand into its complement with ever-increasing lengths, given the required starting material (activated nucleotides, triplets, or oligomers) [125, 135, 142]. This suggests that accurately copying these lengths is not the limiting factor of RNA replication. Similarly, the replication of template strands containing double-stranded regions is also possible, meaning that the successful replication of a long and structured RNA, i.e., an RNA replicase, is not an impossible feat [108, 125, 166, 202–204].

However, even if a full-length product is achieved, all these approaches suffer from the same downside - the high thermodynamic stability of long RNA duplexes, which is further increased by high salt concentrations often required for folding and catalytic activity, prevents the separation of the synthesized strand from its template in equilibrium conditions [52, 81]. Specifically, high magnesium concentrations, which are nearly always required for catalytic activity, are incompatible with current predictions of the prebiotic environment, as well as the stability of RNA in these conditions [152]. Thus, template-dependent RNA synthesis results in the formation of dead-end duplexes. These duplexes not only hinder the release and folding of encoded ribozymes but also impede the reuse of templates for subsequent replication cycles. Thus, current approaches often rely on sophisticated separation and purification protocols.

Nonetheless, different strategies have been tested to circumvent this strand separation problem in a more plausible way. For example, the use of thermal denaturation in viscous solvents takes advantage of the difference in mobility between short and long oligomers, allowing the shorter oligomers to diffuse and hybridize with the template faster than its long, ligated counterpart [205]. During non-enzymatic RNA polymerization, including short invader oligomers can promote toe-hold strand displacement and subsequent invasion of the strand being synthesized [108]. A third strategy plays on the melting temperatures of chimeric double-stranded structures; RNA:DNA duplexes or RNA:XNA duplexes are less stable than their homopolymeric counterparts and thus easier to melt [202] and can help with strand

separation [203]. Another strategy leverages the strain of covalently closed circular RNA by using it as the template for copying [206, 207]. Still, each strategy has its limitations, and a definitive solution to the strand separation problem, among others, has yet to be demonstrated [20, 52].

5.1.2 Exploiting the non-equilibrium of the AWI-setting

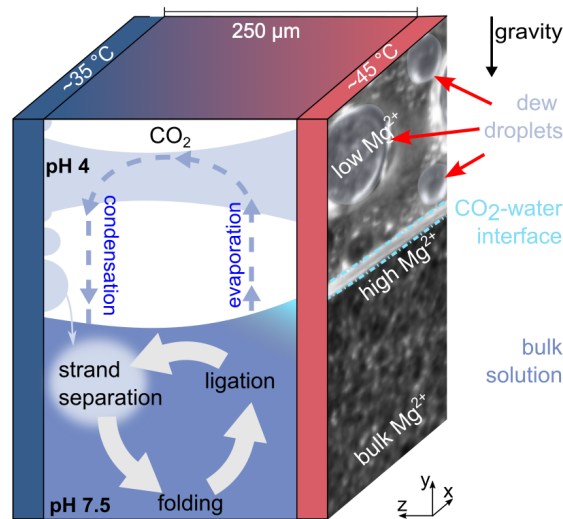


Figure 5.1: The non-equilibrium of AWI-systems enables RNA folding, ligation, and strand separation at low bulk magnesium concentrations:

The non-equilibrium of the AWI-system is maintained by asymmetric heating across the 250 μm wide compartment. This leads to the accumulation of RNA and magnesium ions at the warm side of the interface (cyan), reducing the overall required bulk magnesium concentrations c_{bulk} . In the CO_2 -rich atmosphere, the temperature gradient drives a miniature water cycle, i.e., evaporation at the warm and condensation at the cold side. Due to surface tension, the low salt and pH droplets can either fall back into the solution or rearrange upon growing to touch the opposite hot site. This locally induces strand-separating conditions separating duplexes, potentially triggering multiple rounds of replication. This setup mimics a gas bubble inside a porous volcanic rock. Its miniature water cycle creates different regions in close proximity with strong fluctuations in salt concentration and pH. AWI-settings are a plausible and abundant geological environment on Hadean Earth, i.e. porous volcanic rocks exposed to the CO_2 -rich atmosphere with a temperature gradient originating from volcanic activity and the cold ocean.

Instead of engineering smart ways of circumventing dead-end products in RNA replication, the strand-separation problem can be tackled by a more holistic approach. Physical means to separate duplexes, i.e., temperature, salt, or pH cycles, can be implemented in small reaction compartments. To achieve cycles in an autonomous and realistic way, these compartments must be kept in an out-of-equilibrium state. Rather than focusing on temporal changes in reaction conditions, such as temperature cycles in thermal cyclers, dilution series, or manual pH changes by the experimentalist, the non-equilibrium leads to locally

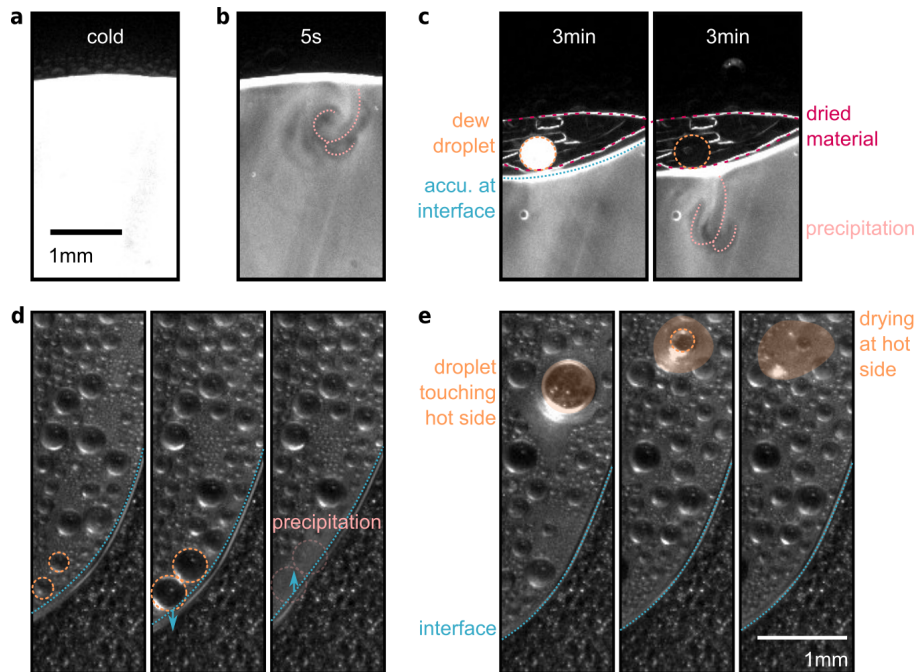


Figure 5.2: Dynamics of dew droplets in AWI-system: **a** The distribution of RNA, in particular the primer, within the half-filled AWI-system was monitored using the Cy5-channel. As the fluorescence signal indicates, the bulk shows a uniform RNA concentration before heating. **b** Upon heating, dew droplets form by condensation and precipitate back into the bulk solution (precipitation), starting already after a few seconds. Simultaneously, RNA molecules accumulate at the interface, as indicated by the increased fluorescence. The overall drop in fluorescence is due to the temperature dependence of the Cy5 molecule. **c** Evaporation, pressure changes, and precipitation droplets lead to position changes of the interface. Upon lowering of the interface, RNA can dry on the warm sapphire window (pink line). Contrarily, dried material can be dissolved again as the interface moves up. Additionally, dew drops that are large enough rearrange due to surface tension such that they touch both sides of the AWI chamber (left). Near the interface, such droplets resolve the dried material and precipitate back into the bulk solution, carrying the material with them. **d** To better visualize the droplet dynamics, the microscope can be operated in bright field mode. The cold side is covered with droplets of various sizes that can grow, merge, and precipitate back into the bulk solution. The downward movement of the interface as water evaporates and is stored in the dew droplets, as well as the upward movement as droplets precipitate, can also be tracked in bright field. **e** Droplets bridging the opposite sides of the chamber not only precipitate back but occasionally split. The droplet on the cold backside again grows and coalesces, etc. However, the droplet on the warm sapphire slowly dries up. This process can occur multiple times in direct succession, leading to fast wet-dry cycling of material on the warm sapphire.

varying reaction conditions in close proximity. Molecules will move from one location to another either by drag forces, diffusion, or other transport phenomena, consequently lead-

ing to temporal changes in reaction conditions. A general way of implementing a constant non-equilibrium is asymmetric heating, i.e., a temperature gradient across the reaction compartment. The advantage of such implementations has already been proven for DNA-based model systems of early replications by enabling elongation as well as strand separation cycles [78, 165]. Similarly, as described in chapter 4, recent results show that RNA-based replication can also be driven in similar settings [166]. However, denaturation of double-stranded RNA solely by means of temperature is limited by the respective melting temperatures, which rapidly reach $>90\text{ }^{\circ}\text{C}$ for longer lengths and elevated salt concentrations [150].

Rather than relying on completely water-filled reaction compartments, new approaches employing temperature gradients across air-water interfaces (AWI, Figure 5.1) show promising results in melting longer, more structured sequences at moderate temperatures [83, 88]. These interfaces could have occurred in porous rock systems exposed to temperature gradients resulting from volcanic activity or even just heating through the sun (as smaller gradients compared to thermogravitational traps are needed) and the cold ocean. This setting is of particular interest for RNA replication mediated by ribozymes since it should not compromise their activity but allows for lower overall salt concentrations as the temperature gradient leads to a co-accumulation of salts and nucleic acids at the warmer side of the interface. Thus, it produces a high local concentration of salts needed for catalysis [103]. This accumulation mainly results from the capillary flow created by water evaporation at the warm side of the air-water interface. The additional effects of water convection and Marangoni flow at the interface contribute for large molecular assemblies on the order of tens of micrometers but are minimal for most cases, including the ones described in the following. At the same time, the temperature gradient leads to a miniature water cycle. Water evaporates at the warm side of the interface and re-condenses at the cold side. Already the low salt conditions of the condensation (or dew) droplets could locally induce strand separating conditions [83]. This is, however, limited to shorter lengths. The strand-separating capacity can be enhanced even further by introducing a reducing CO_2 atmosphere, leading to condensation droplets that are low in salt and pH [88]. The low pH results from an equilibrium of dissolved carbonic acid, bicarbonate, and carbonate in dew droplets depending on the applied partial pressure of CO_2 (see Figure 5.21). While low salt conditions decrease the stability of double-stranded nucleic acids due to the lack of cations present acting as shielding agents that reduce the repulsion of the negatively charged backbones, low pH enables transient melting of otherwise stable nucleic acid duplexes presumably due to nucleobase protonation [178].

The dew droplets on the cold side can grow and coalesce. Large enough droplets either directly precipitate back into the bulk or rearrange due to surface tension so that they bridge both sides of the chamber. Those also precipitate back into the bulk or slowly dry up. The overall droplet dynamic depends on the pressure and type of gas, the temperature gradient, as well as the thickness of the chamber. The AWI-system can be monitored either in bright field or in fluorescence mode (see Figure 5.2). Monitoring the fluorescence of the Cy5 labeled RNA primer confirms the accumulation of RNA at the air-water interface triggered by the temperature gradient. The combination of accumulation and the sinking movement of the interface due to evaporation, dried material sticks at warm sapphire above the interface. This material can be re-dissolved and transported back into solution either by dew droplets that bridge to the warm side and subsequently precipitate or the moving interface as the pressure slightly changes or as droplets precipitate back into solution (see Figure 5.2 a-c).

The overall droplet dynamics are a crucial parameter for the respective experiment as they are responsible for the strand-separating conditions and dictate the rate of changing conditions. They can be best visualized in bright field mode (see Figure 5.2 d-e). To attain an appropriate temperature gradient, the top temperature of the gradient for all experiments was fixed to match the isothermal conditions. Then, several cold temperatures were tested by monitoring the droplet dynamics of the reaction buffer and chosen according to the best droplet dynamics.

The next sections demonstrate how ribozyme-catalyzed RNA replication benefits from the AWI-setting. By exploiting the non-equilibrium conditions, replication of sense and anti-sense strands and functional ribozymes becomes possible in a one-pot system. The combination of conditions suitable for ligation as well as strand separation enables reaction steps, which are otherwise mutually exclusive under isothermal conditions. Thus, the combined synthesis, release, and folding of active ribozymes appears to be an attainable objective.

5.1.3 *sunY* as model system for early ribozymatic replication

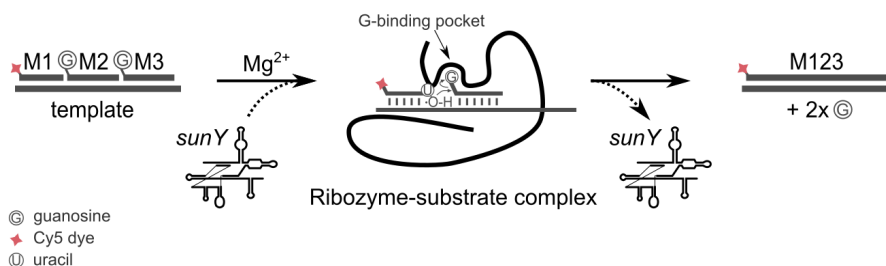


Figure 5.3: *sunY*-mediated templated ligation: The *sunY* ribozyme can catalyze the templated ligation of RNA strands using guanosine nucleosides (G) at the 5'-end as leaving group. This process is schematically shown for three ligation substrates M1-3. The red label at the M1 substrate corresponds to a Cy5 label necessary for downstream analysis. The ligation mechanism is based on the exon ligation step during RNA splicing by a group I intron. First, the ribozyme substrate complex is formed, followed by the positioning of the ligation junction of the substrate duplex in the catalytic center of the ribozyme by binding of the terminal guanosine nucleotide of the downstream exon in the G-binding pocket [208]. The 3'-OH group of the upstream exon, once aligned, initiates an inline attack on the phosphodiester bond of the bound 5'-guanosine. This results in the ligation of the adjacent upstream and downstream exons, simultaneously releasing the guanosine and the catalytic intron.

The discovery of ribozymes, i.e., functional RNA sequences, in the 1980s has boosted the search for early replication scenarios solely dependent on RNA [209, 210]. Ribozymes can either be of non-viral or viral origin and are mostly catalyzing transesterification reactions [163]. Group I or group II introns are of the first type and are found in non-coding regions of genes (introns). Group I introns are especially interesting for the search of the origin of life as they present a potential mechanism for the emergence of RNA-based catalysis and information transfer. They consist of a highly structured RNA molecule that includes a catalytic

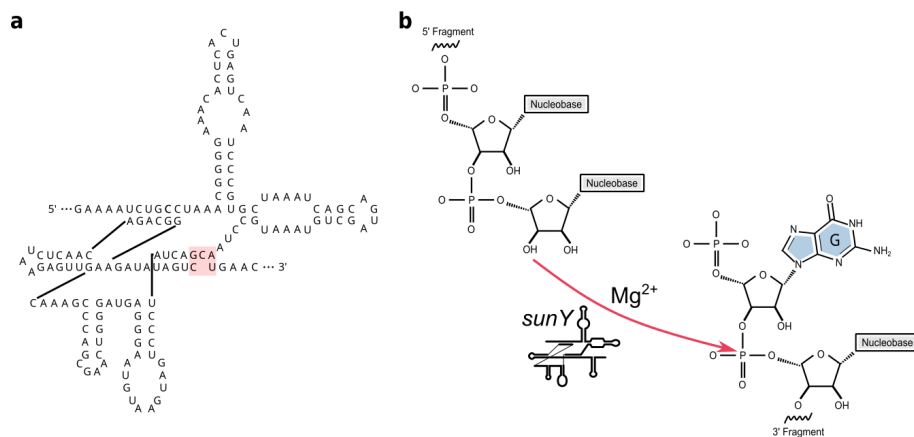


Figure 5.4: Detailed scheme of *sunY* and the reaction mechanism at the ligation site of the templated ligation: **a** Sequence and schematic representation of the secondary structure of the full length *sunY* variant (adapted from [199]). The G binding pocket, essential for substrate-binding and positioning in the catalytic center, is highlighted by the pink square. **b** The hybridization of the substrates on the template ensures accurate positioning for the attack of the 3' hydroxyl of the upstream substrate on the 3' end of the 5' end guanosine of the downstream substrate. The transesterification reaction, where the 5' guanosine acts as a leaving group, is catalyzed by *sunY*, provided that the magnesium concentration is sufficient. As a result, both strands are covalently linked via a 3'-5' phosphodiester bond.

core region and several peripheral domains and are able to self-splice, meaning that they are capable of catalyzing their own removal from pre-mRNA transcripts during gene expression. During the splicing process, the intron RNA folds into a specific three-dimensional structure that brings together the 5' and 3' splice sites. This allows the intron to catalyze its own excision from the RNA transcript and the ligation of the flanking exons. In biology, they are typically found in the genomes of bacteria, bacteriophages, and eukaryotic organelles such as mitochondria and chloroplasts.

In the origin of life research, derivatives from group I intron ribozymes are used to catalyze templated ligation reactions. One such ribozyme is the *sunY* ribozyme, a derivative of the self-splicing *sunY* intron from bacteriophage T4 (in the following only referred to as *sunY*). If the ligation substrates are bound to their template, *sunY* can catalyze template-dependent oligonucleotide ligation using 5' guanosines as leaving group [126,199] (see Figure 5.3). Thus, their catalytic activity is independent of prerequisite activation chemistries such as phosphate- or imidazole-based leaving groups. There are several slightly varying versions of the *sunY* ribozyme with minor differences in the catalytic rate. For a robust catalytic activity in isothermal ligation reactions, *sunY* requires a magnesium concentration of ≈ 50 mM and a temperature of 45 °C. Following the hybridization of ligation substrates to their respective template, *sunY* catalyzes the inline attack of the 3' hydroxyl of the upstream substrate on the 3' end of the 5' end guanosine of the downstream substrate. After this transesterification reaction, with the 5' guanosine acting as leaving group, both strands are covalently linked via the natural 3'-5' phosphodiester bond. Although the only prerequisite of the templated ligation reaction on the sequence is the additional guanosine, the ligation efficiency can be

significantly increased by introducing a G:U wobble pair at the 3' end of the attacking fragment. To ensure maximal ligation efficiency, this constraint was applied to the design of the ligation systems.

Overall, the *sunY* ribozyme provides a robust activity and only requires pure RNA without any prerequisite activation for its catalytic activity, making ligation reactions based on *sunY*-derived ribozymes an attractive model system for primitive enzymatic RNA-replication [211, 212].

5.2 Results

5.2.1 *sunY*-catalyzed RNA ligation at heated air-water interfaces

As described previously in subsection 5.1.2, the AWI-system leads to an accumulation of magnesium as well as RNA at the warm side of the interface, lowering the bulk Mg^{2+} requirements. In order to probe the dependence of the catalytic activity on the bulk magnesium concentration and explore whether the AWI-system provides a suitable environment for the ligation reaction by a 182 nt variant of the *sunY* ribozyme, a simple ligation system for the 30 nt long Mango aptamer was designed (see Table 5.2). The system consisted of a 39 nt long template, a 5'-Cy5 labeled substrate (M1, 13 nt), which allows direct fluorescence-based PAGE analysis, and two downstream 5'-guanosine 'activated' substrates, M2 (7 nt), and M3 (10 nt). Based on estimates of the concentration effect at the evaporation zone of the heated side of the air-water interface, bulk $MgCl_2$ concentrations of 1 mM, 5 mM up to 10 mM were tested

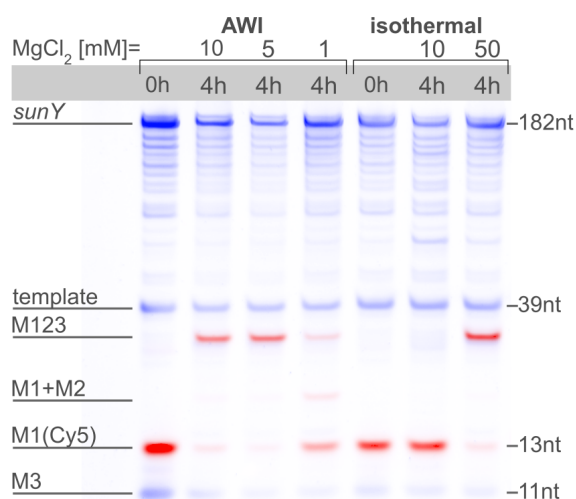


Figure 5.5: PAGE analysis of *sunY*-mediated templated ligation: Analysis by PAGE (imaging with two channels: SYBR Gold in blue and Cy5 in red) reveals that within the AWI environment, ligation of the RNA product M123 is achievable from three RNA substrates (M1-3) even with $MgCl_2$ concentrations as low as 1 mM. To attain similar yields to the AWI-system (at 5 mM) under isothermal conditions, a much higher $MgCl_2$ concentration of 50 mM is necessary. Importantly, the maximum temperature within the reaction setup in both scenarios (isothermal and AWI) remains considerably below the calculated melting temperature (76 °C) of the M123-template duplex.

in the AWI-system. As expected, the significantly lower Mg^{2+} bulk concentration (5 mM and 10 mM) were required to observe similar substrate ligation ratios (89% and 80%) compared to isothermal conditions (50 mM, 86%) after 4 h of incubation (5.5). Even at the lowest tested bulk concentrations of 1 mM $MgCl_2$, product was observed (17%). In an isothermal (equilibrated) system kept at 45 °C, no detectable ligation was observed already for 10 mM. These findings confirmed that the local up-concentration of solutes at the warm side of the air-water interface was sufficient to allow *sunY*-dependent RNA ligation at bulk magnesium concentrations considerably lower than those required under isothermal conditions.

The Mango aptamer was specifically chosen in order to investigate the strand-separating capacity of the AWI-system. Upon binding of the Mango aptamer to its ligand, the TO1b-dye, the dye starts to fluoresce (see Figure 5.6 a). Running the AWI-system with the full-length aptamer and its ligand showed the expected uniform fluorescence signal at the start of the experiment and the characteristic accumulation and droplet dynamics after the onset of the temperature gradient. However, after filling the AWI-system with the aptamer and its template in the presence of the TO1b-dye, no fluorescence signal could be detected, as the aptamer is tightly bound to its template. Thus, its secondary structure and, consequently, the fluorescence is inhibited. This is precisely the case for isothermal conditions. In contrast, turning on the temperature gradient led to an increase of fluorescence for certain regions in the chamber, more specifically, the region at and slightly above the interface (see Figure 5.6 b). To investigate whether the fluorescence is not an artifact of the concentrated and drying TO1b-dye in the presence of RNA, a control experiment including only the template and dye was run. Surprisingly, the control experiment also showed a slight fluorescence increase at the interface. This could be due to an unspecific confinement of the TO1B in the presence of high RNA concentrations. Thus, the fluorescence data is not entirely conclusive.

As an alternative approach to assess the strand-separating capability of the AWI-system, the ligation reaction was conducted in the presence of the full-length aptamer at an equal concentration to the template. The pipetting order was chosen such that first template and aptamer were pipetted, and only then the ligation fragments, to ensure maximum binding of aptamer and template. Since the aptamer would sequester the template, rendering it unavailable for the ligation reaction, no product would be expected under isothermal conditions. PAGE analysis confirmed this inhibitory effect. However, when the reaction was carried out in the AWI-system, the ligation product M123 was observed even in the presence of the full-length Mango aptamer, confirming strand separation (see Figure 5.6 d). Although in the case of the reverse complement being present, the reaction proceeded with lower efficiency, the dew droplet dynamics are sufficient to provide enough strand separation capacity to allow for the synthesis and potentially multi-turnover. The yield for the full-length product (M123) was 82% and 30% with and without the complement present, respectively. The negative control, consisting solely of the ligation fragments, does not show any ligation product. This finding confirms that, despite significant accumulation within the AWI-system, the ligation process persists in a templated manner, guaranteeing the transmission of information rather than simply elongating the sequence (see Figure 5.6 e).

These results make the AWI-system a promising candidate for RNA-based replication that would not only allow for synthesis at lower bulk magnesium concentrations but would be able to promote strand separation. The obvious next step is to employ the AWI-system for

more interesting ligation reactions that would benefit even more from the strand-separating conditions, such as self-replication, which is discussed in the next section.

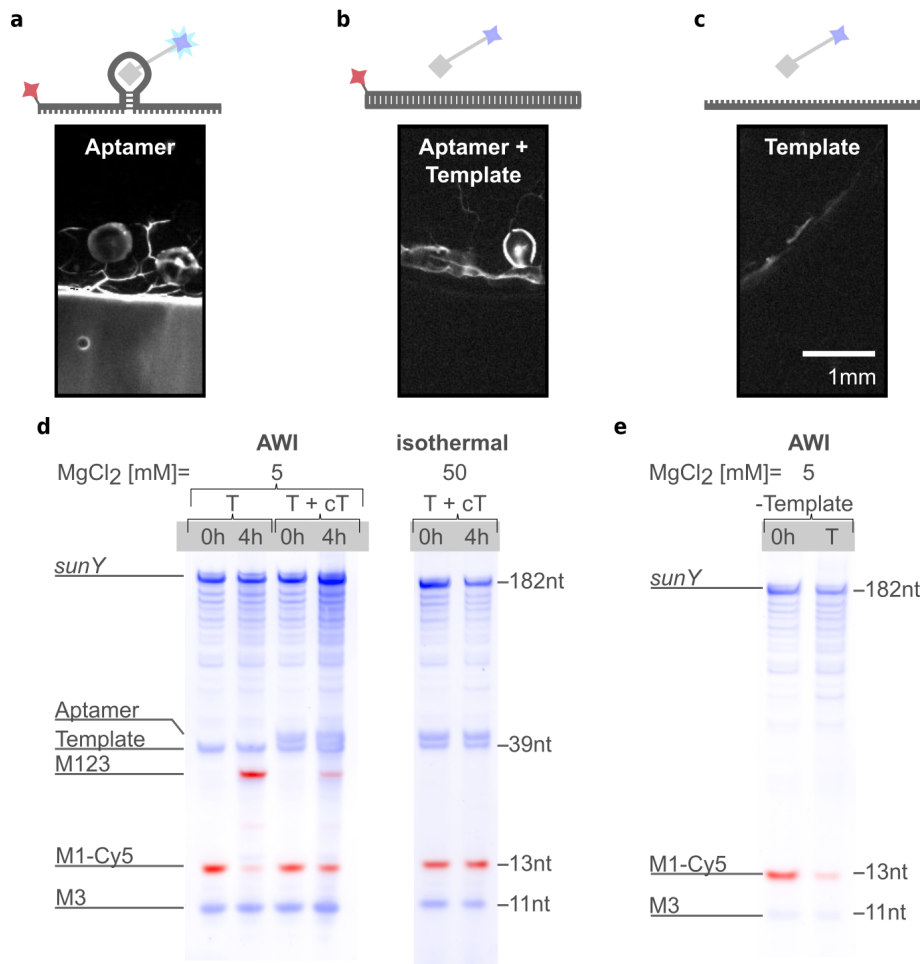


Figure 5.6: Strand separating capacity of AWI-system: Fluorescence signal of the TO1b-dye in presence of **a** only the aptamer, **b** aptamer and template and **c** only the template. The TO1b-dye shows a significant fluorescence increase upon binding to the Mango aptamer. **d** The reaction under AWI conditions proceeded both in the absence (left) and presence (right) of the reverse complement of the template. The isothermal control shows neither product nor intermediate formation. **e** The negative (no template) control for the simple ligation system shows no ligation product. This confirms that even with strong accumulation inside the non-equilibrium compartment, the ligation still proceeds in a templated manner, ensuring information transfer instead of just an elongation behavior.

5.2.2 *sunY*-mediated ligation of *sunY* fragments

One of the biggest mysteries for the onset of an RNA world is the possibility for autocatalytic self-replication of a universal replicating ribozyme, i.e., a ribozyme capable of replicating itself as well as any other RNA. For self-replication, however, template-product inhibition is

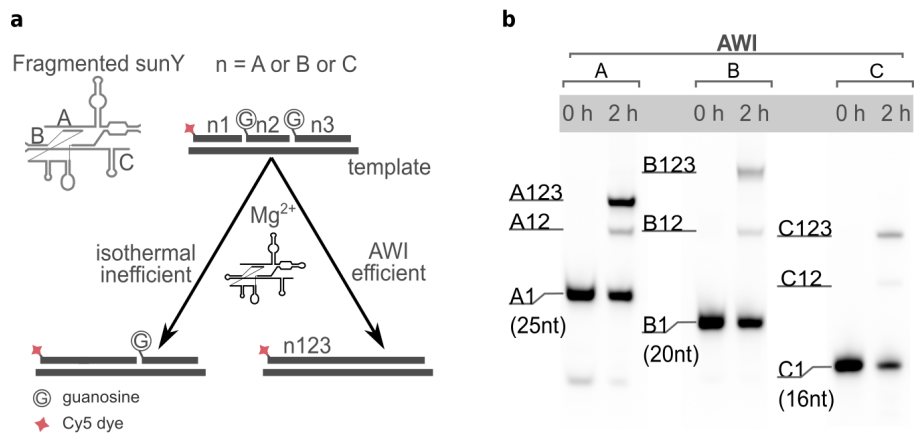


Figure 5.7: *sunY*-mediated ligation of *sunY* fragments: **a** The ligation of the fragmented *sunY* (composed of fragments A, B and C) was each individually build from three ligation substrates (n1-3, with n = A, B or C) on their respective template. Providing enough *sunY* and Mg²⁺, the reaction is hypothesized to proceed efficiently in the AWI-setting. **b** PAGE analysis of separately performed template-dependent ligations of the three *sunY* fragments A123, B123, and C123. Lanes show reactions before (0 h) and after 2 h of incubation.

especially challenging as even the first ligation step is inhibited by the complementarity of the template with the replicating ribozyme. The results from the previous section showed that the AWI-system is capable of strand separations and, therefore, prompted the attempt to synthesize strands with the same or similar sequence as *sunY*. Previous work by the Szostak group demonstrated the non-covalent assembly of an active version of *sunY* from three oligonucleotides (A, B, and C) of lengths ranging from 43 to 75 nt [126]. In their experiments, they have also employed the fragmented version of *sunY* (denoted as ABC) to ligate the complement of fragment C. However, the concentration of the strands was adapted to provide an excess of the fragment C. This way, C could act as a template for the complement while still being available to fold into the active ABC form. The objective here is to demonstrate an alternative approach by driving the synthesis of the fragments themselves, starting with equal concentrations of *sunY* (full-length version) and a complementary section that acts as the template by using the full-length version of *sunY*. The aim is to investigate whether the non-equilibrium conditions in the AWI-system are enough to promote strand separation and support replication of *sunY*-derived RNA strands.

To this end, the ligation of each of the three fragments (herein referred to as A123, B123, and C123) was started from three short oligonucleotides substrates (A1-3, B1-3, and C1-3) (see Table 5.2) with equal concentrations (2.5 μ M) of *sunY* and template and four-fold excess of each ligation fragment (see Figure 5.7). Significant yields of full-length products were observed in the AWI-system after a 2 h reaction (38.5 \pm 1.5% for A123, 19.5 \pm 2.5% for B123, and 22 \pm 7% for C123), with only minor amounts of incomplete intermediates. In contrast, isothermal conditions resulted in low yields of full-length fragments (8% for A123, 1% for B123, and 1.3% for C123 after 2 h) and higher relative amounts of intermediate products. These findings sug-

gest that the high melting temperatures of the different RNA duplexes in the system restrict processivity under bulk conditions (Figure 5.7 b).

5.2.3 Coupled sense and antisense synthesis of *sunY* fragments

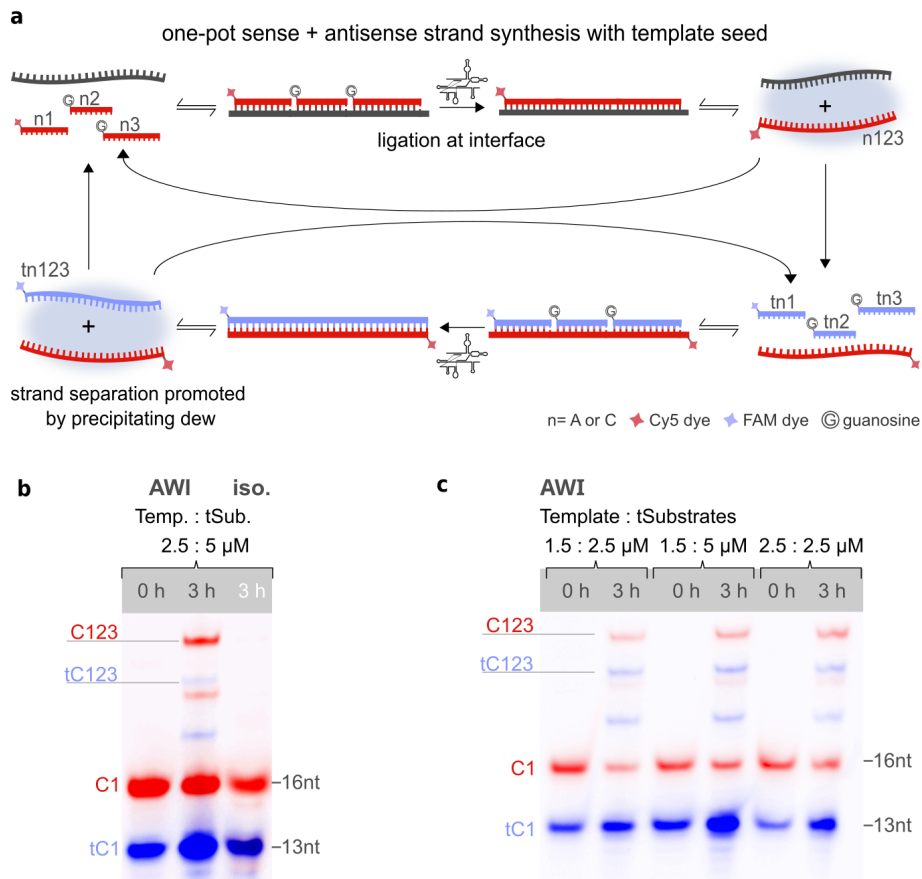


Figure 5.8: Complete replication cycles of the *sunY* fragments in the presence of template: **a** Schematic representation of the full replication cycle of the *sunY* fragments A123 or C123. The replication proceeds via ligation of the RNA substrates for the sense strand (n1-Cy5, n2, n3, red) on the template, followed by the ligation of the antisense strand (tn1, tn2, tn3-FAM, blue) on the newly synthesized n123. **b** PAGE-analysis confirms that in contrast to the isothermal conditions, both the sense and antisense strands emerged in the AWI-system with the template present. The concentrations of the ligation substrates for the antisense strand tC123 were lowered to 5 μM instead of 10 μM (as is the case of tA123), as leaving the concentration for both sense and antisense ligation fragments at 10 μM did not lead to any synthesis. **c** Several concentration ratios were tested to decrease the overall complementary pressure on *sunY*. The combined concentration of all complementary strands to *sunY* was always kept above the 2.5 μM equal concentration limit.

After successfully demonstrating the synthesis of individual *sunY* fragments by the full-length ribozyme within the AWI-system, the system's potential for supporting complete cycles of RNA replication was explored. The goal was to achieve the simultaneous synthesis of

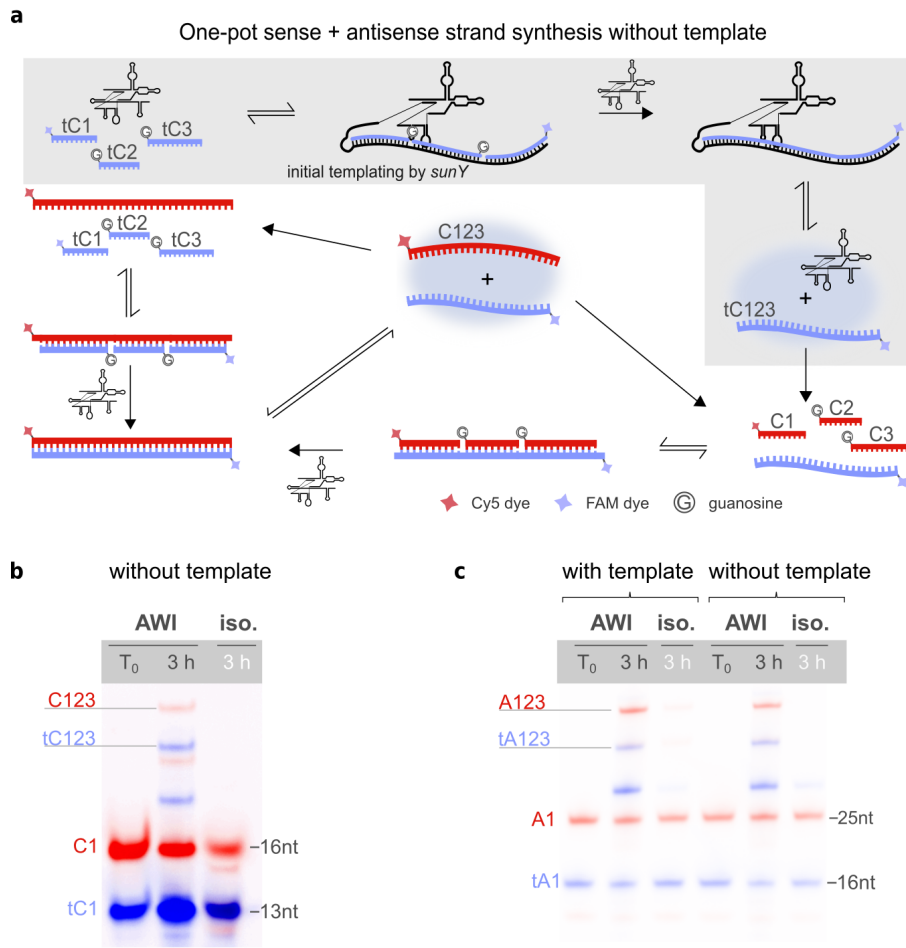


Figure 5.9: Complete replication cycles of the *sunY* C123 and A123 fragments: **a** Reaction path for the full replication cycle with absent template schematically shown for fragment C123. The hypothesized (grey box) beginning of the replication proceeds via an initial templating step by *sunY* (tC1-FAM, tC2, tC3, blue). The ligation of the sense strand from the RNA substrates (C1, C2, C3-Cy5, red) and the antisense strand proceeds then via the newly synthesized tC123. A similar reaction path is plausible for tA123 and A123, where another part of *sunY* would act as the initial template. **b** In contrast to isothermal conditions, both the sense and antisense strands of the C fragment emerged without the template present in AWI conditions. **c** For the *sunY* fragment A, PAGE analysis revealed that both sense (A123) and antisense (tA123) strands emerged in the AWI system with and without the template present. The isothermal conditions did not yield any product.

both sense and antisense strands in a single reaction environment. To accomplish this, separate ligation experiments were designed, employing three substrate oligonucleotides each for *sunY* fragments C123 and A123 (C1-3 and A1-3), as well as their corresponding templates (tA1, tA2, tA3 for tA123 and tC1, tC2, tC3 for tC123) (see Figure 5.8 a and Table 5.2). For a full replication cycle, 10 μM of ligation substrates for both the sense and antisense strands were initially included in the reaction. The concentrations of templates and *sunY* were both at 2.5 μM . The sequences and the binding sites of the ligation fragments are shown below.

Note that compared to ligation substrates for A, the ligation sites for fragment C are more blunt-ended.

Complementary ligation substrates for sense and antisense strand of A123:

tA1 | tA2 | tA3

FAM - 5' ugauuuagc gacggau|ugucucagugag ggu|uucccguuuuag3'
3' augacgacuaaaucg|uugccua acagagucacuc|uca aagggcaaaaucgcgucuaaaag5' -Cy5

A3 | A2 | A1

Complementary ligation substrates for sense and antisense strand of C123:

tC1 | tC2 | tC3

FAM - 5' cugcauguc gccu|ugcaguucag gcu|auaucuucaa3'
3' cuaggacguacag|uggu acgucaaguc|uga uauagaaguugagaa5' - Cy5

C3 | C2 | C1

To initiate the ligation of sense fragments (A123 and C123), seed amounts of tA123 and tC123 were included in the respective ligation reaction. The incubation period was extended to 3 h to allow for a sufficient reaction time. PAGE analysis revealed the formation of both sense strands as well as antisense strands (A123, tA123) for the A fragment, indicating that once formed, A123 acted as template for the synthesis of their respective antisense template, tA123 (see Figure 5.9 c). The yields in the AWI system were 40% and 19% for full-length A123 and tA123, respectively. Under isothermal conditions, no full-length product tA123 was detected. The yield for A123 under isothermal conditions was 6%. For the *sunY* fragment C, leaving the concentration for both sense and antisense ligation fragments at 10 μ M did not yield any product. However, decreasing the concentration of the overall complementary strands of *sunY* again resulted in the formation of sense strands as well as antisense strands (C123, tC123) as shown in Figure 5.8 b,c. The yields for C123 and tC123 in the AWI system were 11% and 1%, respectively. Several additional concentration ratios of the seeding template and anti-sense substrates were tested with the combined concentration of all complementary strands to *sunY*, always above the 2.5 μ M equal concentration limit. All reactions showed full-length product with yields for C123 of 26%, 27% and 34% and tC123 of 7%, 4% and 8% (from left to right, Figure 5.8 c).

Intriguingly, the formation of sense and antisense fragments for A as well as C was also detected in the absence of any seeding templates (Figure 5.9). This suggests that the *sunY* ribozyme itself is able to function as a template for the synthesis of A123 and C123 in the AWI-system. Subsequently, these newly synthesized fragments could then serve as templates for the formation of tA123 and tC123, completing the replication cycle within the AWI-system. Without the template, the yields in the AWI system were 34% and 24% for A123 and tA123, respectively. The isothermal control showed no full-length product for both A123 and tA123. The yields for C123 and tC123 were both 3% in the AWI-system. Under isothermal conditions, no full-length products were detected.

The difference in the ligation efficiency between the A and C fragments could have several possible explanations. First, the *sunY* fragments B and C make up the catalytic core and the G-binding pocket. Thus, inhibition or even partial inhibition due to partial unfolding and binding of the template to the sequence part of fragment C could have a more drastic influence on the activity of the full-length *sunY* compared to similar inhibition of the sequence part of fragment A. Additionally, the full-length and fragmented versions have slightly varying sequences of the fragment A part. This results in a less efficient binding of the template for the fragment A ligation substrates to the corresponding part of the full-length *sunY* ribozyme. A more detailed analysis of the binding behavior is discussed in section 5.4.11, where different compositions of strands are analyzed with Nupack in terms of their mean field energy structure. Last, as shown above, the overlap between the ligation substrates for fragment A and C vary in terms of number of overlap length. As the ligation site between tA1 and tA2 has a minimum of 7 nt overlap with A2, this ligation step of the complementary site can proceed without strand separation and is present to some extent in all reactions (AWI and isothermal), which in turn could kick-start the reaction.

Although it is very promising that sense and antisense strands emerge, the reaction was not yet tested for an exponential replication behavior. To do so, it would be beneficial to increase the overall reaction time and test different starting concentrations.

5.2.4 Synthesis of the hammerhead ribozyme

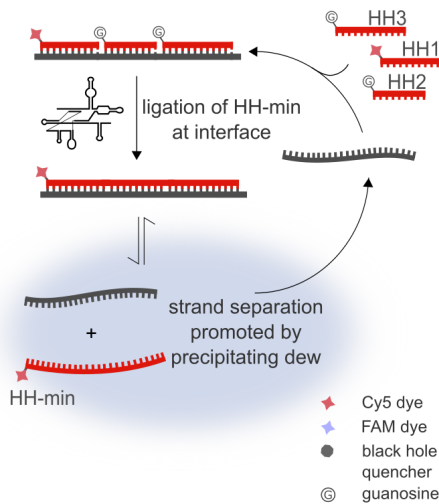


Figure 5.10: Templated synthesis of the hammerhead ribozyme (HH-min): *sunY* catalyzes the ligation of the hammerhead ribozyme (HHmin, red) from the three RNA substrates (HH1-3, red) on the template (grey). Locally varying pH and $MgCl_2$ concentration cause strand separation conditions, resulting in the release of the ligated HH-min.

Equally important for a sensible RNA world as self-amplification is the replication of a variety of ribozymes capable of diverse functions. To probe whether the AWI-system is capable of driving the *sunY*-based ligation reaction of a different ribozyme, a minimal version

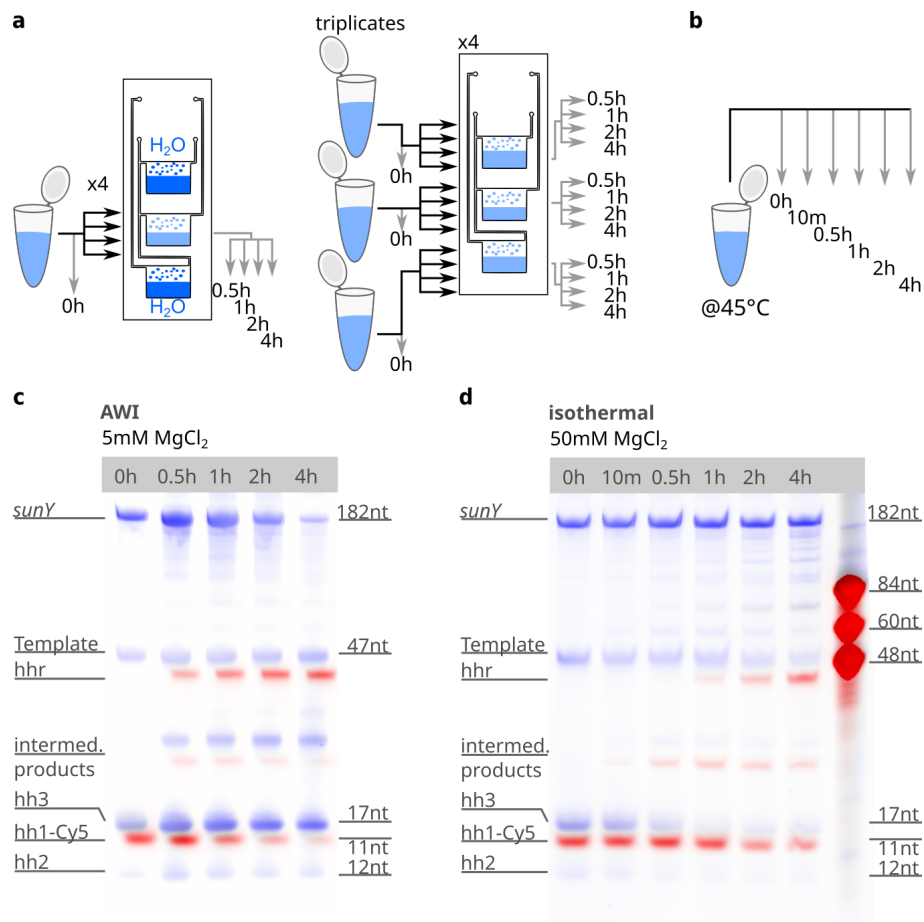


Figure 5.11: Kinetics of HH-min synthesis with a 1:2 (5 μ M:10 μ M) ratio of template to substrates: The samples with a 1:2 concentration ratio of 5 μ M:10 μ M for ligation substrates and template, respectively, were analyzed at different time points of the reaction. **a** For the AWI-system, samples were aliquoted into equal parts for the time points 30 min, 1 h, 2 h, and 4 h. For each time point, a separate AWI-system was assembled. For triplicates, the reactions were prepared individually and again divided into equal parts for the time points 30 min, 1 h, 2 h, and 4 h. Each time point corresponded to a separate AWI, with the three chambers as triplicates. **b** For reactions carried out under isothermal conditions, time points were taken from the same reaction at 0 h, 10 min, 30 min, 1 h, 2 h, and 4 h. **c** PAGE analysis of samples reacted under AWI-conditions. **d** PAGE from samples from isothermal incubations.

of the hammerhead ribozyme (HH-min) [85, 213], derived from the tobacco ring spot virus satellite RNA, was chosen as a simple test scenario. Similar to the previously described cases, the template-dependent ligation, catalyzed by *sunY*, started from three substrates (HH1-Cy5, HH2, and HH3 with 11 nt, 11 nt, and 16 nt). In the AWI-system the reduced constraint of the bulk magnesium concentration presumably leads to strand separation by the dew droplets, potentially allowing for multi-turnover (see Figure 5.10 and Table 5.5).

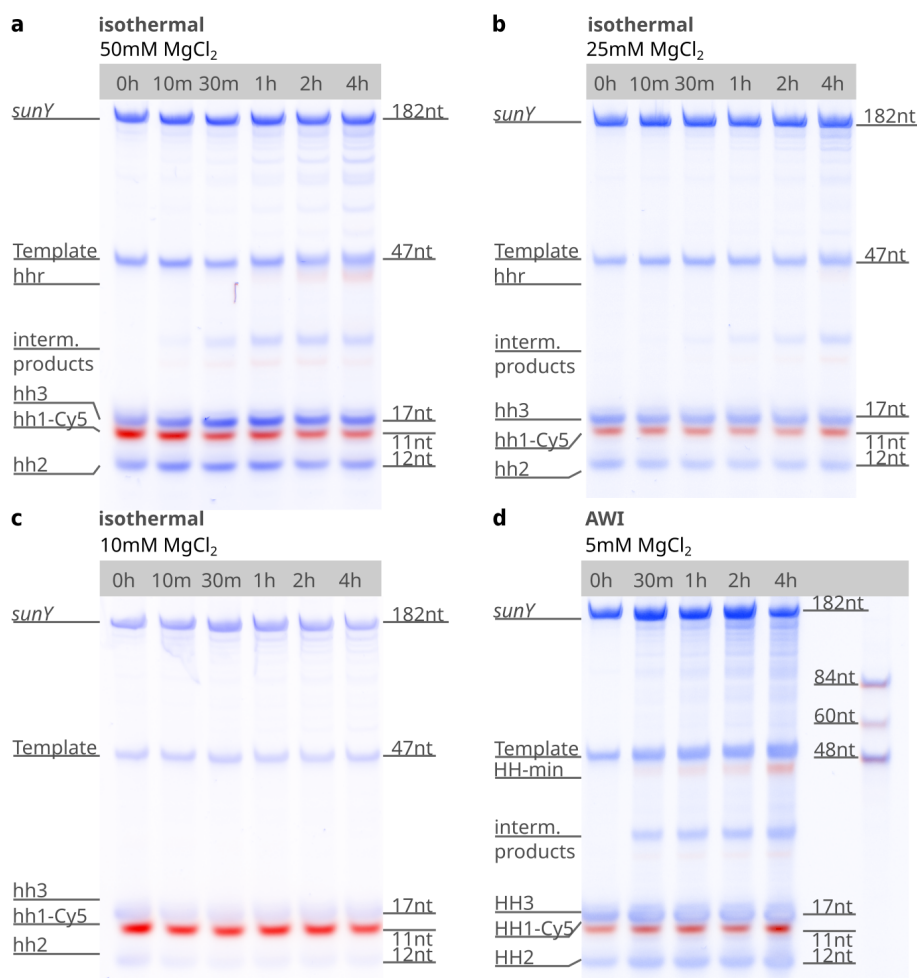


Figure 5.12: HH-min synthesis with a 1:4 (3 μ M:12 μ M) ratio of template to substrates: The reaction under isothermal conditions were analyzed by PAGE. Buffers contained a $MgCl_2$ concentration of **a** 50 mM, **b** 25 mM and **c** 10 mM. **d** In the AWI-system, the buffer contained 5 mM $MgCl_2$.

The isothermal and AWI reactions were prepared in 30 mM Tris·HCl pH 7.5, 100 mM KCl, and 50 mM or 5 mM, respectively. The isothermal reaction was again incubated at 45 °C. Samples were divided into equal aliquots for both isothermal conditions and the AWI-system. Time points of 10 min, 30 min, 1 h, 2 h, and 4 h were selected for isothermal conditions. The same time points were taken for the AWI-system, excluding the 10 min, as the heating and cooling for the temperature gradient at the beginning and end of the experiment introduce delays, which obscured results from short incubation periods. For each time point, a separate AWI-system was assembled. To ensure similar humidity conditions in the AWI-system for all experiments, empty chambers were filled with 15 μ L of water. As expected, the experiments in both the AWI-system and under standard isothermal conditions led to the synthesis of HH-min (see Figure 5.11). However, the AWI-system promoted the synthesis of HH-min even at a 10-fold lower $MgCl_2$ concentration (5 mM) than required for isothermal ligation (50 mM), supporting our previous data regarding magnesium dependence in non-equilibrium conditions.

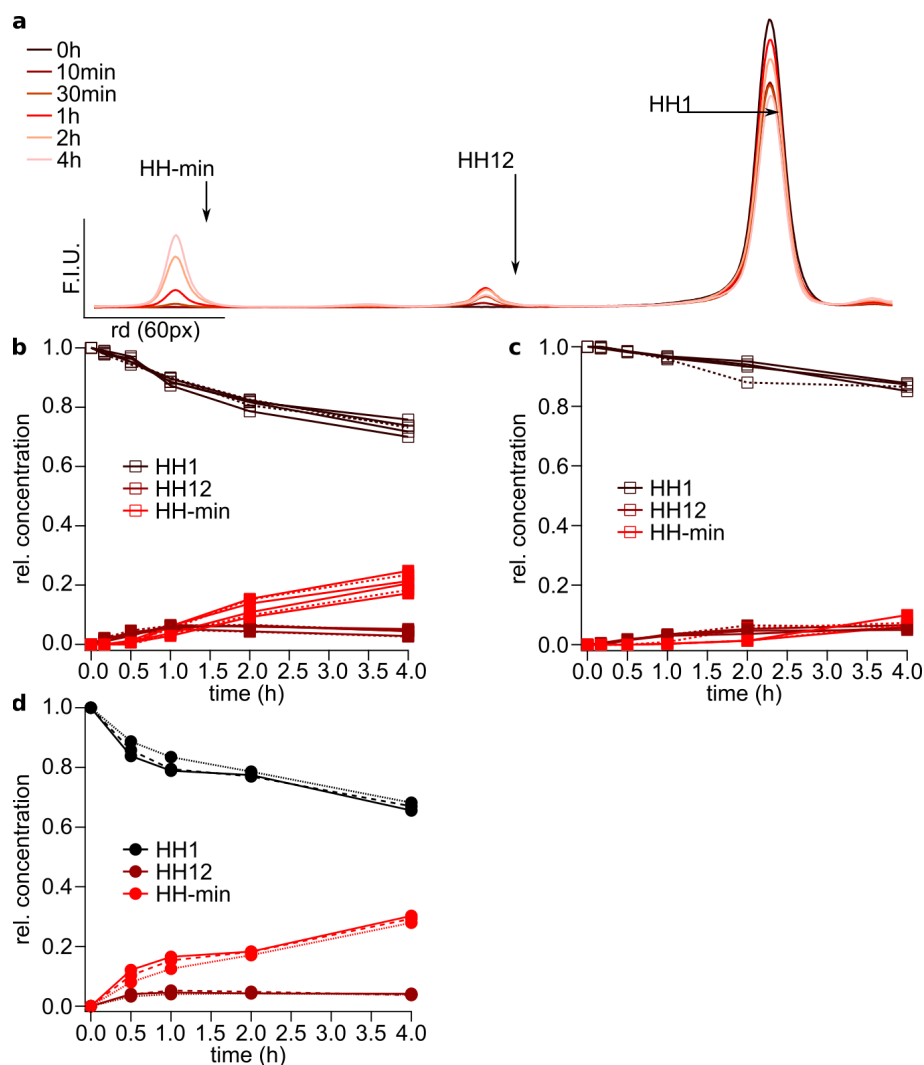


Figure 5.13: Quantification of HH-min product formation by PAGE: **a** Example traces of the fluorescence signal obtained from PAGE analysis for a time series of the 50 mM MgCl₂ isothermal reaction. Each peak was fitted with one, or in the case of broader shoulders, several Gaussians. The total area corresponding to the same peak was normalized by the total area, i.e., the sum of all Gaussians. **b** After 4 h, the results for 50 mM MgCl₂ showed an average yield of 20 ± 3% for the full-length product. **c** After 4 h, the results for 25 mM MgCl₂ showed an average yield of 8 ± 2% for the full-length product. **d** The triplicates of the AWI reaction showed an average yield of 29 ± 1% after 4 h for the full-length product.

As multi-turnover of the reaction was expected, the concentration of ligation substrates and template were changed from 10 μM and 5 μM to 12 μM and 3 μM to facilitate a clearer outcome. Kinetic measurements of the 1:4 ratio of ligation substrates to template were performed as triplicates for both, the AWI-system and isothermal conditions. To this end, each of the triplicates was individually prepared and divided into as many equal parts as time points. For the

AWI-system, a separate trap with the triple chamber Teflon cutout was assembled for each time point, with the chambers corresponding to the triplicates. PAGE-analysis revealed that the AWI-system indeed enabled multi-turnover synthesis of HH-min as the resulting yields of the full-length HH-min, which exceeded the amount of input template RNA (see Figure 5.13). After 4 h, the triplicates of the AWI reaction show an average yield of $29 \pm 1\%$. In contrast, no excess synthesis of HH-min over the template could be observed under standard isothermal conditions (average yield of $20 \pm 3\%$ for 50 mM MgCl_2), indicating that all product strands remained tightly associated with the template. This observation of multi-turnover ligation in the AWI-system suggested the presence of single-stranded, *de novo* synthesized HH-min in solution.

5.2.5 One-pot synthesis and folding of active hammerhead ribozyme

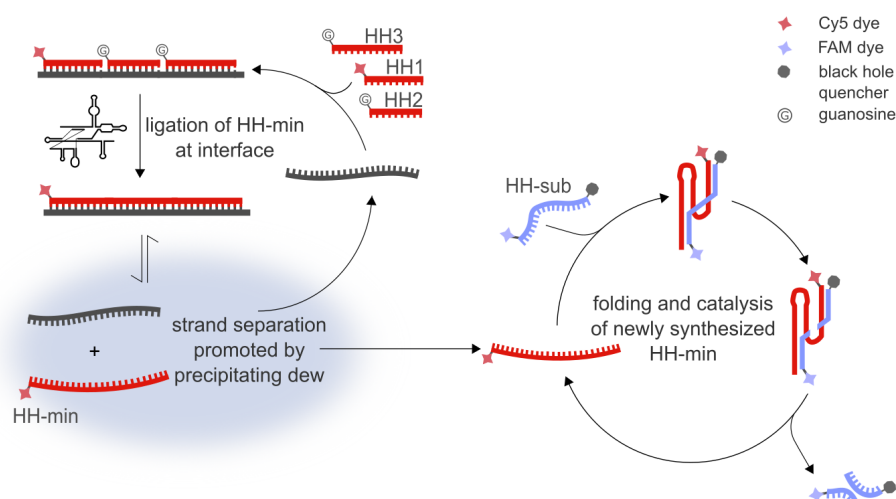


Figure 5.14: Templated synthesis and release of active hammerhead ribozyme: The template-dependent ligation by of the hammerhead ribozyme (HHmin, red) from the three RNA substrates (HH1-3, red) mediated by *sunY* can proceed in the AWI-system. Additionally, local variations in pH and salt concentration cause strand separation, potentially releasing ligated HH-min, which could then bind to and cleave its substrate (blue).

The previous section showed that the *sunY* mediated templated ligation proceeds within the AWI-system at low bulk magnesium conditions. Additionally, it was described how the periodic local decreases in pH, caused by acidified dew droplets precipitating back into the bulk reservoir (Figure 5.2), transiently lowers the melting temperatures of RNA. Thus, the AWI-system allows for both the ligation as well as the strand strand-separating conditions. Together, these results are promising for the replication of ribozymes, as the AWI-system could not only drive the synthesis but facilitate the release of ligation products from their templates, thus enabling their intramolecular folding into functional RNAs, a process similar to transcription (see Figure 5.14).

The observed multi-turnover suggests that free HH-min is present in the solution. If this dissociated HH-min was catalytically active, it should be able to bind to and subsequently cleave

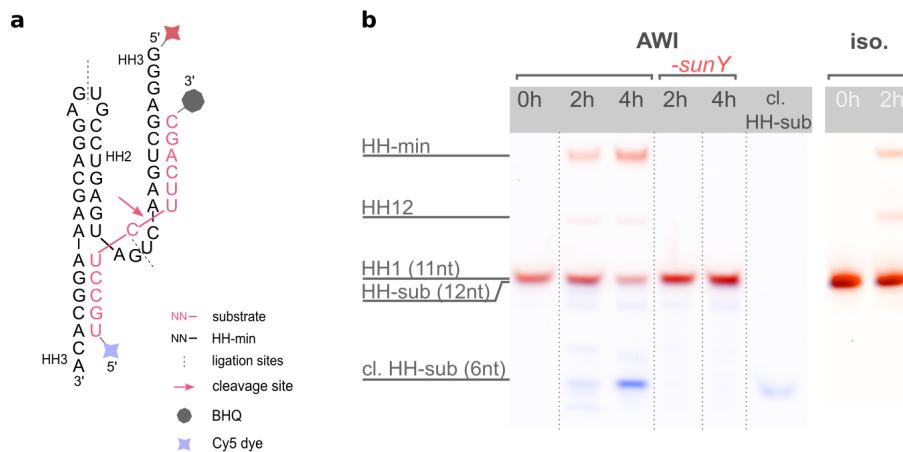


Figure 5.15: Synthesis of active HH-min in one-pot reaction in AWI-system:

a HH-min (grey) is ligated from three fragments (HH1-3). The two ligation sites are indicated with a dashed line. The active structure binds to its substrate (pink) and cleaves it at the specified cleavage site, indicated by the pink arrow. To distinguish the fluorescence signal, HH-min has a 5' Cy5 tag, whereas HH-sub has a 5' FAM- and 3' BHQ-tag. **b** After a 2 h incubation period, the presence of ligated HH-min (red, Cy5) was observed in both the AWI-system and the isothermal environment. Activity, i.e., cleaved hammerhead substrate (HH-sub, blue, FAM), was exclusively observed in the non-equilibrium (AWI) conditions. Control experiments conducted without *sunY* ribozyme demonstrated no assembly of HH-min or cleavage of HH-sub, indicating that the unligated HH-min fragments HH1-3 lacked residual RNA-cleavage activity. The isothermal control also exhibited no activity. The individual gel lanes have been rearranged and outlined with dotted lines to enhance clarity.

its cognate substrate (HH-sub) (see Figure 5.15). To test for activity, the ligation reaction of HH-min was repeated as described in the previous subsection - this time in the presence of HH-sub to observe catalytic substrate cleavage. The concentration ratio of ligation substrates to template was again 2:1 with concentrations of 10 μ M and 5 μ M, respectively. The results obtained from PAGE analysis indeed confirmed that the formation of HH-min was accompanied by the cleavage of HH-sub under non-equilibrium (AWI) conditions. For quantification of the cleaved substrate, the extracted PAGE traces were fitted with Gaussian peaks. The area corresponding to the uncleaved HH-sub (first peak) was multiplied with a factor of 1.69 to account for the decrease in fluorescence due to the BHQ. The yield was then calculated as the fraction of the area of the cleaved product divided by the total area (including, if present, bands due to hydrolysis). The cleavage ratios after 2 h and 4 h were measured to be $40 \pm 13\%$ and $60 \pm 7\%$, respectively, indicating a significant occurrence of *sunY*-catalyzed ribozyme assembly followed by product dissociation, folding, binding, and cleavage within the one-pot system (see Figure 5.16 and Table 5.1).

In stark contrast, no cleavage of HH-sub was observed under isothermal conditions, suggesting that the newly synthesized HH-min was unable to dissociate from the template within

Table 5.1: Cleavage yields of HH-sub in the AWI-system from replicates in the one-pot scenario. Experiments were performed at a stoichiometry of 1:2 template to ligation substrates.

											mean	sdv
2 h	0.31	0.34	0.37	0.39	0.60	0.61	0.28	0.49	0.46	0.19	0.40	0.13
4 h	0.57	0.50	0.64							0.68	0.60	0.07

this environment. This observation aligns well with the predicted melting temperature of the HH-min-template complex at the given magnesium concentrations in the reaction, which was approximately 90 °C (see sections 5.4.10 and 5.4.11). Furthermore, as expected, control

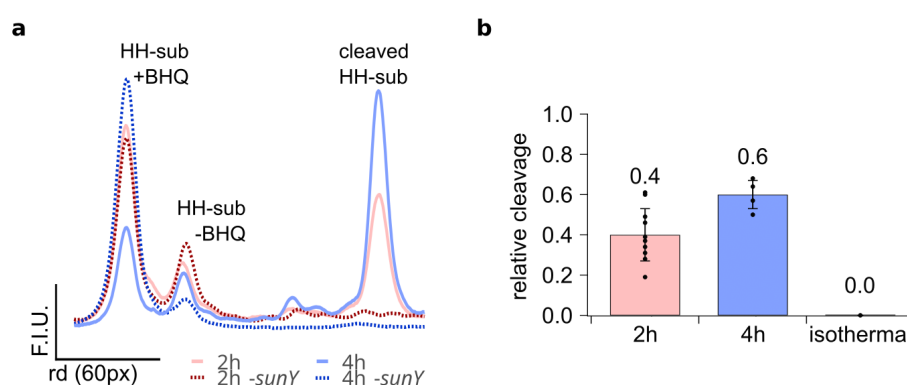


Figure 5.16: Yields for cleaved HH-sub in one-pot reaction: **a** Extracted traces of the PAGE analysis for HH-sub cleavage reaction in the AWI-system. **b** Bar graph showing the calculated yields of HH-sub cleavage after 2 h and 4 h in the AWI-system and under isothermal conditions. The data is presented as mean values \pm SEM, with the black circles representing the values from independent replicates.

experiments conducted in the AWI-system without the presence of *sunY* showed no observable assembly of HH-min or cleavage of HH-sub, further supporting the pivotal role of *sunY* in mediating the ligation reaction.

This is an exciting result as it shows the potential of the AWI-system to host several ribozymes. Not only could they perform different independent functions, but they could even be newly synthesized, an important prerequisite for an RNA world.

5.3 Discussion and Outlook

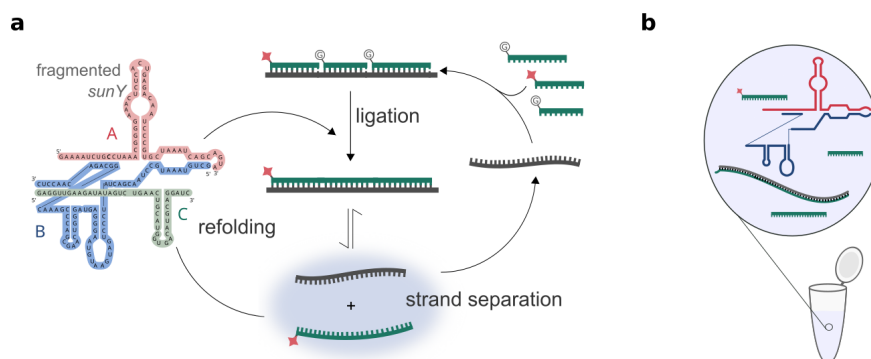


Figure 5.17: Hypothesis for an autocatalytic replication scenario of fragmented *sunY* : **a** Driving the ligation reaction of *sunY* with the fragmented version of *sunY* in the AWI-system could potentially lead to an exponential autocatalytic self-replication as the separated, newly synthesized fragments could actively participate in the replication reaction. **b** In contrast, in isothermal conditions, the template likely inhibits the fragmented *sunY* even before the first ligation reaction, not to mention the lack of strand separation and, thus, lack of potential for exponential amplification.

This chapter discussed a compelling model for a prebiotic reaction environment at an air-water interface (AWI-system) that supports RNA-catalyzed replication of short functional RNAs. It is plausible that similar non-equilibrium microenvironments could have hosted early autocatalytic nucleic acid replicators, which presumably played a crucial role in the emergence of molecular evolution towards greater complexity.

The AWI-systems, as shown in this chapter, are well-equipped to facilitate both ligation reactions and subsequent strand-dissociation of RNA strands ranging from <20 to 75 nucleotides that were built from much shorter precursor oligonucleotides. Especially attractive is the potential of the *sunY* fragments to assemble through non-covalent interactions into functional replicating ribozymes, as this reduces the length requirements of the replicating ribozyme. However, fragmented ribozyme variants typically exhibit lower activity compared to their full-length counterparts. This is likely due to a reduced stability and folding defects. First experiments confirmed that no ligation of HH-min through the fragmented *sunY* variant could be detected under standard AWI or isothermal conditions (see Figure 5.18). However, lowering the temperature on the warm side of the AWI-system from 45 °C to 40 °C showed that the fragmented *sunY* ribozyme was able to synthesize moderate amounts of HH-min. Intriguingly, this suggests that milder conditions could partially compensate for the lower activity of the fragmented variant (see Figure 5.18).

Since the AWI-system can drive the synthesis of the fragmented *sunY* version, the naturally following question is whether the fragmented version could synthesize itself, eventually leading to an exponential autocatalytic self-replication. To explore the autocatalytic potential of the system, the synthesis of each of the three *sunY* fragments (A123, B123, and C123) using the

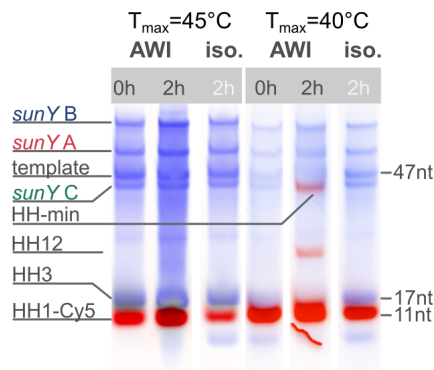


Figure 5.18: Ligation of HH-min by fragmented *sunY* : Running the ligation reaction of HH-min with the fragmented *sunY* did not yield any product at the same temperature conditions as the full-length version. However, reducing the temperature of the hot side of the AWI-system to 40 °C leads to the complete synthesis of HH-min with a yield of 7%, which is significantly lower compared to the full-length version with a mean yield of 40%. Reducing the temperature in isothermal conditions did not lead to any product.

fragmented *sunY* ribozyme was tested. Synthesizing the individual fragments using the fragmented *sunY* version proves to be considerably more challenging. In addition to reducing the temperature of the hot side, the total reaction time was increased to 18 h with the CO₂ supply temporarily halted overnight. Note that this reduces the capacity of the AWI-system to separate strands. Although the yields were considerably lower compared to ligations catalyzed by the full-length ribozyme (see Figure 5.19), full-length products were detected for each fragment.

Further optimization of the experimental setup, such as increasing the reaction times and automatizing the time-dependent CO₂ inflow, will be necessary to improve the reaction yields. Still, these results provide an optimistic outlook on the system's capacity for self-replication from a pool of shorter oligonucleotide substrates. The next goal would be the simultaneous ligation of all three *sunY* fragments by the fragmented version. To begin with, one could start with the same total concentration of complementary strands as fragmented *sunY*, i.e., reducing the concentration of the individual templates instead of matching it with each strand. It would be worth collecting kinetic data of this reaction as an increase in reaction efficiency would be expected as more *sunY* fragments are ligated. This would already be the first evidence of a positive feedback loop expected from autocatalysis.

Additionally, improving the chamber and heater design could improve the strand-separating capability of the AWI-system, allowing for multi-turnover reactions in all replication scenarios. Instead of a uniform heating of the front sapphire, moving to a point-like heat source, similar to the one described in chapter 4, could reduce the thermal stress of the system while maintaining or even increasing strand separating conditions. This proves especially interesting for AWI-systems, customized to meet the requirements of other ribozyme systems such as polymerase ribozymes reported by Holliger *et al.* [207] and Horning *et al.* [125] variants. Testing RNA polymerase ribozymes in the AWI-system is exciting, as non-equilibrium conditions

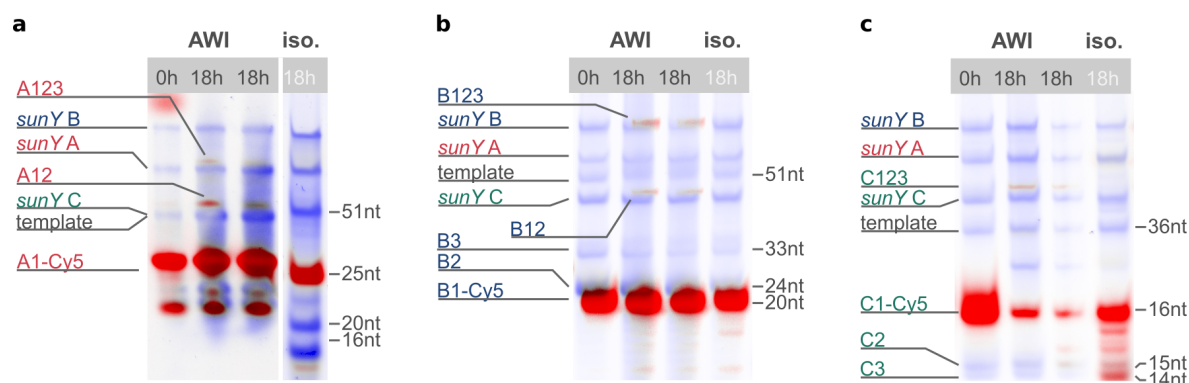


Figure 5.19: Ligation of individual fragments by fragmented *sunY* : For all fragments, i.e., **a** fragment A, **b** fragment B and **c** fragment C, traces of the full-length product are detected in the AWI-system. No full-length product is detected in isothermal conditions.

are not only a route towards self-replication but also help to avoid Spiegelman's dilemma of the 'tyranny of the shortest' [214].

However, several potential challenges need to be considered. First, it is essential to assess these ribozymes' compatibility with the system's dry phase environment. Each ribozyme may have specific requirements for the presence of water or particular buffer conditions to maintain its activity. Therefore, it is crucial to investigate whether the dry phase conditions provided by the setup can support the functionality of these ribozymes effectively. If the dry phase inhibits catalytic activity, automatized pH cycles could also be achieved in different ways, e.g., by acid-base mixtures in thermogravitational traps [215].

Additionally, the buffer conditions may need to be tailored to meet the specific needs of the ribozymes under investigation. Each ribozyme has distinct optimal pH ranges, salt concentrations, or cofactor requirements. Consequently, evaluating and optimizing the buffer conditions is necessary to ensure the ribozymes' stability and catalytic efficiency within the modified chamber design. Thus, it remains to be seen how these replication systems can benefit from salt and pH oscillations and if yields can be enhanced by the template release and reduced degradation.

Interestingly, similar interface settings can drive other prebiotically relevant processes, such as the phosphorylation of mono-nucleosides, RNA polymerization, nucleic acid encapsulation in membrane vesicles, as well as the fusion and fission of coacervate-based proto-cells [84, 88, 103, 105]. Combining these key aspects of biological systems into a unified scenario converges towards the molecular evolution of replication.

In conclusion, the presented data strongly suggest that the heated AWI-system has the potential to host fragmented RNA replicators capable of catalyzing both general and autocatalytic template-directed RNA synthesis. This finding further supports the hypothesis that dynamic, non-equilibrium environments played a crucial role in the emergence of higher-order biomolecular complexity.

5.4 Materials and Methods

5.4.1 Nucleic acids

RNA oligonucleotides were obtained in dry form from Biomers or Integrated DNA Technologies (IDT) and then reconstituted to a stock concentration of 200 μ M using nuclease-free water (Ambion™ Nuclease-free water from Invitrogen). HH-sub was synthesized by Eurofins Genomics. All RNA oligonucleotide sequences are provided in Table 5.2. The final concentrations of RNA were determined by measuring the absorbance at 260 nm. All stock solutions were stored at -80 °C and thawed on ice before use. The filling and extraction of AWI-systems were carried out on ice to maintain the integrity of the components. The *sunY* DNA template for *in vitro* transcription was ordered as a gBlock with the following sequence:

GAT CGA TCT CGC CCG CGA AAT TAA TAC GAC TCA CTA TAG GGA AAA TCT GCC TAA ACG
GGG AAA CAC TCA CTG AGT CAA TCC CGT GCT AAA TCA GCA GTA GCT GTA AAT GCC TAA
CGA CTA TCC CTG ATG AAT GTA AGG GAG TAG GGT CAA GCG ACC CGA AAC GGC AGA CAA
CTC TAA GAG TTG AAG ATA TAG TCT GAA CTG CAT GGT GAC ATG CAG GAT C

The gBlock was amplified using the following primers:

P_{rev} = GATCCTGCATGTCACCATGCAGTTCAGACT

P_{fwd} = GATCGATCTCGCCCGCAAATTAATACGAC

Table 5.2: List of all RNA sequences used in the experiments. Each of the second and third ligation substrates was ordered with an additional 5' G base (lowercase), which acts as a leaving group during the RNA ligation reaction catalyzed by *sunY*.

Name	nt	Sequence (5'-3')
A Template	43	GAU UUU GCC CUU UGG GAG UGA CUC UGU UAG GCA GCG AUU UAG U
A1 (5'-Cy5)	25	GAA AAU CUG CCU AAA ACG GGA AAC U
A2	20	gCUC ACU GAG ACA AUC CGU U
A3	16	gGCU AAA UCA GCA GUA
tA123	43	UGA UUU AGC GAC GGA UUG UCU CAG UGA GGG UUU CCC GUU UUA G
tA1 (5'-FAM)	16	UGA UUU AGC GAC GGA U
tA2	16	gUGU CUC AGU GAG GGU
tA3	13	gUUC CCG UUU UAG
B Template	51	ACG GGU UGC UGG UAG GGA CUA CUU ACA UUC CCU CGU CCC AGU UCG CUG GGC
B1 (5'-Cy5)	20	GCU GUA AAU GCC CAA CGA CU
B2	24	gAUC CCU GAU GAA UGU AAG GGA GU
B3	33	gAGG GUC AAG CGA CCC GAA ACG GCA GAC AAC UC
tB123	52	UCG GGU CGC UUG ACC CUG CUC CCU UAC AUU CAU CAG GGA UGGU CGU UGG GCA
tB1 (5'-FAM)	20	UCG GGU CGC UUG ACC CUG CU
tB2	20	gCCC UUA CAU UCA UCA GGG A

Name	nt	Sequence (5'-3')
tB3	14	gUGG UCG UUG GGC A
C Template	36	AAC UUC UAU AUC GGA CUU GAC GUA CCG CUG UAC GUC
C1 (5'-Cy5)	16	GAG UUG AAG AUA UAG U
C2	15	gCUG AAC UGC AUG GU
C3	14	gGAC AUG CAG GAU C
tC1 (5'-FAM)	13	CUG CAU GUC GCC A
tC2	14	gUGC AGU UCA GGC U
tC3	11	gAUA UCU UCA A
tC123	36	CUG CAU GUC GCC AUG CAG UUC AGG CUA UAU CUU CAA
HH-sub (5'-FAM, 3'-BHQ)	12	UGC CUC UUC AGC
HH-min	38	GGA GCU GAA CUG AUG AGU CCG UGA GGA CGA AAG GCA CA
HH Template	47	GGG AAA UGU GCC UUU CGU CCU CGC GGA CUC AUC GGU UCA GCU CCA AA
HH1 (5'-Cy5)	11	GGA GCU GAA CU
HH2	12	gGAU GAG UCC GU
HH3	17	gGAG GAC GAA AGG CAC A
M123	40	GCG CGC GAA GGA AGG AUU GGU AUG UGG UAU AUU CGC GCG C
M Template	39	GGG AAA CGA AUA UAC CGC AUA CCG AUC CUU CCU UCG AAA
M1 (5'-Cy5)	13	CGA AGG AAG GAU U
M2	8	gGGU AUG U
M3	11	gGGU AUA UUC G
<i>sunY</i>	182	GGG AAA AUC UGC CUA AAC GGG GAA ACA CUC ACU GAG UCA AUC CCG UGC UAA AUC AGC AGU AGC UGU AAA UGC CUA ACG ACU AUC CCU GAU GAA UGU AAG GGA GUA GGG UCA AGC GAC CCG AAA CGG CAG ACA ACU CUA AGA GUU GAA GAU AUA GUC UGA ACU GCA UGG UGA CAU GCA GGA UC

5.4.2 *sunY* transcription protocol

The *sunY* DNA template was amplified by PCR using the Q5 High Fidelity 2x Master Mix (NEB). The PCR reaction included 10-25 ng of the gBlock template and 1 μ M of both primers, P _{fwd} and P _{rev}. The PCR protocol consisted of an initial denaturation at 98 °C for 30 s, followed by 12 cycles of denaturation at 98 °C for 7 s and annealing/extension at 72 °C for 50 s. A final extension step was performed at 72 °C for 2 min. The PCR product was then purified using the Monarch PCR cleanup kit (NEB) and eluted in 6-10 μ L of nuclease-free water.

The transcription buffer consisted of 40 mM Tris·HCl pH 8, 20 mM MgCl₂, 10 mM DTT, and 2 mM spermidine. The template was at an approximate final concentration of 0.5 μ M. Additionally, 3.75 mM of each NTP, 5 U/mL or 2 U/400 μ L of *E. coli* Inorganic Pyrophosphatase (100 U/mL stock), and 1500 U of T7 polymerase were added. The transcription reaction was

incubated at 37 °C for 3-4 h. After transcription, the RNA product was purified using the Monarch RNA cleanup kit (NEB).

The resulting material was gel purified using a 2 mm two-component, 6-well comb gel consisting of 15% PAGE gel (30 mL) and 10% stacking gel (10 mL). The gel was run at 25 W constant power for approximately 2 h. After the finished run, it was wrapped in plastic wrap, placed on TLC plate, and illuminated under UV (254 nm). Under illumination, the ribozyme band was marked, excised, placed in a 2 mL tube, and weighed. The gel slice was dry crushed in the tube using a syringe plunger. 0.3 M sodium acetate pH 4.8–5.2 was added to the tube (2× mg gel weight in μL) and soaked overnight at 4 °C on a rotator.

The supernatant was recovered the next day using Spin-X centrifuge tube filters (Corning, 0.45 μm , sterile) and centrifuged for 5 min at maximum speed. Subsequently, 20 ng glyco-gen together with either 1 volume of isopropanol or 2.5 volumes of ethanol were added. The sample was then frozen for 45-60 min (-80 °C in the case of ethanol or -20 °C in the case of isopropanol) and centrifuged for 45–90 min at -9°C and maximum speed. Subsequently, the supernatant was removed and the pellet was washed with 80% ethanol. It was then centrifuged again for 5 min at maximum speed. This step was repeated twice before drying the pellet for 10–15 min. In the final step, it was dissolved in deionized H₂O.

The final concentrations were determined by measuring the absorption at 260 nm using a Nanodrop (Thermo). The respective extinction coefficients for sequences were calculated using OligoCalc [216]. All *sunY* aliquots were then stored in the -80 °C freezer, only thawed on ice prior to the experiment, and returned to -80 °C after usage. Multiple different aliquots were used for the ligation experiments.

5.4.3 Reaction details of the ligation of M123

RNA ligation reactions were conducted using the following final concentrations: 20 μM for all ligation substrates, 5 μM for the template, 30 mM Tris·HCl pH 7.5, 100 mM KCl, and varying bulk magnesium concentrations as specified. The ligation process was initiated by adding *sunY* ribozyme to a final concentration of 2.5 μM . For the experiments under AWI conditions, a reaction volume of 17 μL was prepared, while a volume of 12 μL was prepared for the isothermal control. As soon as *sunY* was added, 2 μL of each sample was quenched with 2× gel loading buffer (90% formamide, 5% glycerol, 0.015% Orange G). The isothermal control samples were placed in a T100 thermal cycler (Bio-Rad) set to 45 °C. To conduct the AWI experiments, the remaining reaction volume of 15 μL was injected into the fully assembled AWI-system through a Teflon tubing (Techlab) using a Hamilton syringe. The reaction time for both the isothermal controls and the AWI experiments was 4 h. After the reaction time, the sample was extracted from the AWI-system and immediately quenched and analyzed with PAGE. The analysis was done with a two-channel protocol enabled by the sybr stained gel and the 5' Cy5 labeled first substrate (M1). For more detailed information on the PAGE protocol, please refer to the subsection 5.4.6. Additionally, the subsection 5.4.7 provides further details on the assembly and filling process of the AWI-system.

5.4.4 Reaction details of synthesis of *sunY* fragments

The synthesis reactions were conducted for 2 h at a final concentration of 10 μM for each ligation substrate, including the 5'-Cy5 labeled first ligation substrate and 2.5 μM of template. The reaction buffer consisted of 30 mM Tris·HCl, pH 7.5, 100 mM KCl, and 10 mM MgCl₂. The conditions for the reactions under both AWI and isothermal conditions followed the previously described procedure.

For the complete replication of *sunY* sense and antisense fragments, both the sense and antisense ligation substrates were added. The syntheses of A123 and tA123 were carried out with a final concentration of 10 μM for each substrate (A1, A2, A3) and their complementary substrates (tA1, tA2, tA3). The reactions were incubated for 3 h with and without 2.5 μM of template. The first substrate of the A123 and C123 ligation reactions (A1/C1) were 5'-Cy5 labeled, while the 5'-terminal substrates of the tA123 and tC123 ligation reactions (tA1/tC1) were 5'-FAM labeled. To improve yields and reduce incubation times, the concentrations for the synthesis of C123 and tC123 were adjusted by decreasing the overall concentration of strands complementary to the full-length *sunY* ribozyme. Four concentration ratios were tested, denoted as 'Template: complementary substrates: substrates (T:tS:S)': 2.5 μM : 5 μM : 10 μM , 1.5 μM : 2.5 μM : 10 μM , 1.5 μM : 5 μM : 10 μM , and 2.5 μM : 2.5 μM : 10 μM . All replication reactions were performed in a buffer containing 30 mM Tris·HCl, pH 7.5, 100 mM KCl, and 10 mM MgCl₂. The ligation of substrates was initiated by adding *sunY* to a final concentration of 2.5 μM . For the ligation of C123 and tC123, the experiments ran for 3 h or 18 h, as indicated.

5.4.5 Reaction details of HH-min synthesis

The HH-min synthesis reactions were conducted at different final concentrations of either 10 μM and 5 μM or 12 μM and 3 μM of all ligation substrates and template, respectively. As specified for the reactions, the first RNA ligation substrate, HH1, was either unlabeled or labeled with a 5'-FAM or 5'-Cy5. The buffer used in the AWI reactions and isothermal controls consisted of 30 mM Tris·HCl, pH 7.5, 100 mM KCl and varying bulk magnesium concentrations (C_{bulk} , as specified). The reactions were initiated by adding *sunY* to a final concentration of 2.5 μM . To verify the activity of the *in situ* synthesized HH-min hammerhead ribozyme, 0.4 μM of HH-sub, which contained a 5'-FAM tag and a 3'-BHQ quencher, was added. For a T_0 sample, 2 μL of the prepared reaction was quenched immediately with 18 μL of gel loading buffer (94% formamide, 5% glycerol, 0.01% Orange G) right after the reaction initiation. All isothermal controls were incubated at 45 °C. To ensure that the three unligated HH-min ligation substrates, HH1-3, had no residual RNA-cleavage activity triggered by the accumulation process at the water-air interface, control reactions without *sunY* were performed for 2 hours and 4 hours.

5.4.6 Polyacrylamide gel electrophoresis (PAGE)

All samples were subjected to analysis using polyacrylamide gel electrophoresis (PAGE) with a 15% acrylamide gel. The gel stock was prepared using Roth Rotiphorese DNA sequencing gel stocks with an acrylamide-to-bisacrylamide ratio of 19:1. For a 0.75 mm thick gel with a 15-tooth comb, approximately 5 mL of gel mixture was required, consisting of 3 mL of gel concentrate, 1.5 mL of gel diluent, 0.5 mL of buffer concentrate, 25 μL of APS, and 2.5 μL

of TEMED. Denaturing conditions were employed for gel electrophoresis, utilizing 50% urea and 1x TBE buffer at temperatures around 50-55 °C. Each runtime started with a 30 min pre-run at 400 V, after which 4-5 μL of the samples were loaded per well. For PAGE analyses of T_0 samples and samples from isothermal experiments, a 1:9 dilution with gel loading buffer (94% formamide, 5% glycerol, 0.01% Orange G) was used. For PAGE analyses of samples from AWI experiments, a 3:7 dilution was employed. Electrophoresis was conducted with a two-step protocol consisting of 5 min at 50 V, followed by 18-24 min at 300 V. After the finished run, the gels were directly imaged in the case of quantifying the reaction for both Cy5 and FAM labels. When only quantifying a Cy5 label, they were stained in 50 mL of 1x TBE buffer with 5 μL of 10,000x SYBR Gold Nucleic Acid Gel Stain for 5 min before imaging. Gel images were captured using a two-channel protocol with the Bio-Rad ChemiDoc MP System and analyzed using an in-house LabView routine (NI LabVIEW 2014 14.0.1f11 (64 bit)). The background correction of the recorded lane intensities was done with the respective average intensity of the inter-lane spaces. Each band was fitted with one or multiple Gaussian peaks using Igor Pro 6.37. Since extraction from AWIs and quenching during imaging can introduce artifacts, only relative intensities r for each lane were utilized for quantification. They were calculated following equation 5.1:

$$r = \frac{I_{\text{product}}}{I_{\text{total}}} = \frac{I_{\text{product}}}{I_{\text{substrates}} + I_{\text{intermediate}} + I_{\text{rest}} + I_{\text{product}}} \quad (5.1)$$

where I_{product} and I_{total} refer to the intensity of the product and to total intensity, respectively. For the quantification of the activity, the intensities from the uncleaved substrate HH-sub were multiplied by a factor of 1.69 to account for the quenching effect caused by the presence of the Black Hole Quencher at the 3' end.

5.4.7 Preparation of AWI chambers and sample injection

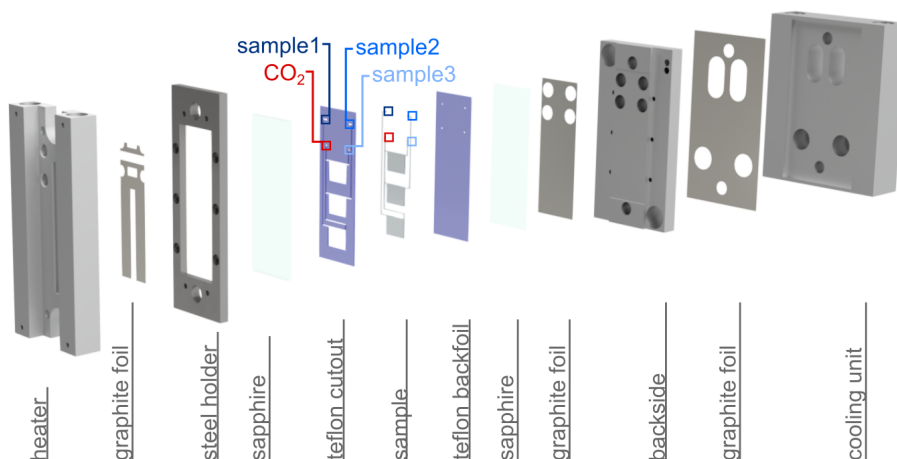


Figure 5.20: Components for the assembly of the AWI-system setup: The whole AWI setup was designed in Autodesk Inventor. The assembled trap consists of the 12 parts shown here. As the setup is modular, only slight changes have to be implemented when adapting to the experimental need. Such as the shape of the Teflon cutout or the heater.

The AWI chambers were constructed following a similar design to the trap system described in Matreux *et al.* [150]. Figure 5.20 shows a schematic image of such a chamber. The desired shape was cut out from a 250 μm thick Teflon foil (Holscot Europe) using a GRAPH-TEC CE6000-40 Plus cutter. This Teflon slide defined the outline of the non-equilibrium AWI compartments. One block was designed to have three reaction compartments, each constituting an AWI-system. The compartments each had dimensions of 11 mm width and 8 mm height, allowing for a maximum capacity of 22 μL with a 250 μm foil thickness. A second foil with a thickness of 100 μm was cut to serve as the hydrophobic backside. Both Teflon slides were sandwiched between two sapphire windows and secured to a metal block using a steel frame. To enhance heat conductivity between the cold sapphire and the metal surface, a 25 μm thick graphite foil (PANASONIC) was inserted between the two elements. All components of the assembly had dimensions of 22 mm width and 60 mm height and were cleaned with ethanol and Milli-Q water before the assembly. Additionally, dust was removed using scotch tape. A heater was attached to the top side of the chamber. To ensure optimal heat conductivity for applying the temperature gradient, 200 μm thick graphite foils (PANASONIC) were placed between the heater and the assembled AWI-system, as well as between the AWI-system and the water-cooled backside of the setup.

A reaction volume of 15 μL was injected with a Hamilton syringe through a Teflon tubing (Techlab) fixed by PTFE fittings (Techlab) prior to each experiment. Before applying the temperature gradient, the inlets were sealed with soft silicone and PTFE fittings. All compartments were connected to the same CO_2 inflow to ensure similar pressure conditions. A pressure controller (RIEGLER), directly at the setup, was used to adjust the pressure. Additionally, a manometer (Bourdon Instruments, MEX 3) was attached to the CO_2 supply of the chamber to cross-verify the applied pressure. The influence of the CO_2 atmosphere on the bulk pH is described in section 5.4.8.

The temperature gradient was controlled by heating the front using an electrical heater regulated by an Arduino microcontroller board and cooling the backside of the chamber with a refrigerated water bath (CF41 and 300 F, Julabo). Temperatures were monitored throughout the experiment using a GTH 1170 thermometer (Greisinger) at the back sapphire by screwing it into one of the holes in the backside using one of the fittings. At the beginning and end of each experiment, temperatures were also measured using a heat camera (Seek Thermal, SQ-AAA) on the front sapphire to monitor the applied gradient. The measured temperatures were used as input parameters for a finite-element simulation to simulate the thermal gradient inside the chamber. Details of the temperature simulation are provided in section 5.4.9.

Two types of setups were used to create the non-equilibrium conditions in the AWI-system, either with or without the possibility of optical readout. For fluorescence excitation, the THORLAB LEDs M470L2-C4 (470 nm emission) and M625L2 (625 nm emission) were used for the dyes FAM and Cy5, respectively. For the readout of the TO1-b a 505 nm wavelength Cyan-LED from THORLABS was used. Imaging was done with the Axio Scope.A1 microscope (Zeiss) employing the infinity-corrected long working distance objective (2 \times lens Mitotoyo, M Plan Apo) and a 0.5 adapter, coupled to a stingray camera F145B ASG (Allied). Temperature and microscope were controlled with code written in NI LabVIEW 2014 14.0.1f11 (64 bit).

5.4.8 Measurement of bulk pH changes in AWI

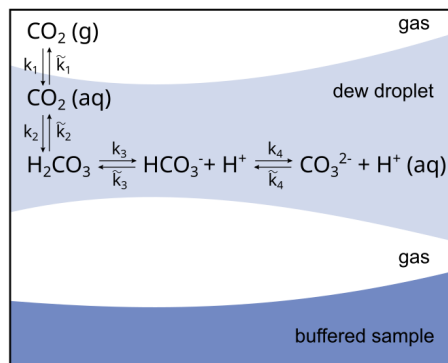


Figure 5.21: pH change due to CO_2 atmosphere: The pH of the dew droplets gets reduced due to the dissolution of gaseous carbon dioxide. The pH reduction depends on the applied pressure. The scheme of the reaction was adapted from [84].

The pH of the dew droplets acidifies in a CO_2 -rich atmosphere, as the gaseous CO_2 dissolves in water (see Figure 5.21). The total pH drop depends on the applied pressure of CO_2 inside the chamber. In the reported experiments, the applied CO_2 pressure was 0.5-0.7 bar, leading to mean pH values between 4 and 4.5 [84]. In case the buffer capacity of the bulk solution is strong enough, only the dew droplets exhibit a pH reduction. To assess the impact of the CO_2 atmosphere on the bulk solution, the pH of the sample solution was measured before and after the experiment. Initially, a 45 μL buffer solution was prepared, and its pH was determined using a Thermo Scientific™ Orion™ 9826BN Micro pH Electrode (Thermo Fisher Scientific, USA). The solution was then divided into three 15 μL aliquots and filled into an AWI-system with a triple chamber design. The AWI-system was subjected to the same heat gradient and CO_2 flux as the experimental conditions for RNA ligation. After a runtime of 4 h, the volumes from the three chambers were combined, and the pH was measured once again. These control experiments were repeated on three different days, and although the CO_2 influx was manually controlled, slight variations in the intervals and lengths of the influx may have occurred. The median of the initial pH of the solution was 7.61, with an average pH decrease of 0.57 ± 0.32 , resulting in a median pH of 7.07 after 4 h of incubation. Based on these control experiments, it was concluded that the buffer in the bulk solvent was sufficient to prevent significant overall acidification.

5.4.9 Temperature simulations with COMSOL

Finite-element simulations (COMSOL Multiphysics 5.4) were used to model the temperature gradient inside the chamber. A digital model of the system was created using 3D CAD software (Autodesk Inventor 2021) and was imported into the finite element software for a realistic 3D simulation. The thermal conductivities of the different materials in the 3D model were obtained from the COMSOL internal database or the product data sheets. The thermal conductivities k at room temperature are shown in Table 5.3.

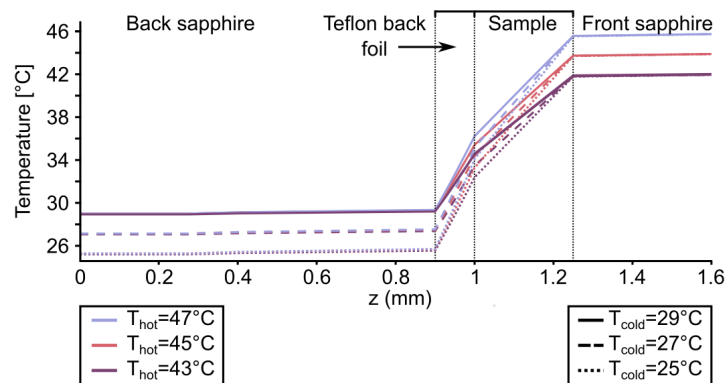


Figure 5.22: Cross-section of temperature simulation: The temperature gradient across the non-equilibrium chambers was simulated employing finite-element simulations for different sets of experimentally measured temperatures at the back (cold) and front (hot) side of the sapphire. The temperature gradient plays a crucial role in influencing the accumulation behavior and dew droplet dynamics within the system. Experimental optimization was performed to determine the optimal temperature gradient that would result in favorable droplet dynamics, specifically ensuring the formation of sufficiently large condensation droplets for rearrangement towards the hot side. Good overall droplet dynamics were found for temperatures between 45-47 °C and 26-28 °C at the front and back sapphire, respectively.

Table 5.3: Thermal Conductivity of Materials

	Water (H ₂ O)	Steel	Aluminum	Sapphire	FEP Foils
k [W/(m K)]	0.62	44.5	237	35	0.2

The heat flow through the outer surfaces of the model was assumed to be thermally insulating, i.e., $Q_{\text{external surface}} = 0$. The temperature profile was calculated using the following formula:

$$\nabla \cdot (k\nabla T) = -Q \quad (5.2)$$

where T represents the temperature, k is the thermal conductivity, and Q is the heat source in (W/m³). The temperature of the back side of the cold sapphire and the front side of the warm sapphire were fixed as boundary conditions based on measurements obtained from a thermometer (Greisinger, GTH 1170) and with a heat camera (Seek Thermal, SQ-AAA), respectively.

The choice of the hot side temperature for the sapphire was intentionally made to match the isothermal temperature. The cold temperature was experimentally optimized to ensure favorable dew droplet dynamics. It was found that temperature combinations within the range of 45-47 °C for the front sapphire and 26-28 °C for the back sapphire resulted in good overall dynamics. Variations within this temperature range did not have a significant impact on the reaction outcomes.

5.4.10 Melting temperatures

The melting temperatures for the ABC and hammerhead system were calculated for various divalent ion concentrations based on data from [84].

Table 5.4: Predicted Melting Temperatures for the ABC System. Melting temperatures were predicted for different pH, 100 mM of monovalent ions, adapted length and GC content, and varying divalent ion concentrations. The last two columns represent melting temperatures with low concentrations of monovalent (1 mM) and divalent (0.2 mM) metal ions at pH 4 and pH 3.6. If no melting transition was observed, the melting temperature is reported as NaN. The calculations considered the longest possible double-stranded region and assumed stoichiometric conditions.

strands					len. (nt)	GC%	c_{MgCl_2} (mM)	T_M (°C) pH7	T_M (°C) pH4	T_M (°C) pH3.6	low salt	
T	A123	A1	A2	A3							T_M (°C) pH4	T_M (°C) pH3.6
yes	yes				57	43%	50 10 5 1	85.72 84.53 83.56 79.23	57.98 56.88 55.99 51.97	29.64 28.63 27.81 24.13	42.61	15.53
yes		yes			25	40%	50 10 5 1	72.98 71.83 70.89 66.68	45.1 44.04 43.17 39.25	17.37 16.39 15.6 12	30.06	3.51
yes			yes		19	47%	50 10 5 1	69.89 68.74 67.81 63.61	41.57 40.51 39.64 35.73	13.72 12.75 11.95 8.35	26.47	NaN
yes				yes	15	40%	50 10 5 1	53.1 52 51.11 47.08	25.13 24.12 23.28 19.51	NaN NaN NaN NaN	10.48	NaN
T	B123	B1	B2	B3	len. (nt)	GC%	c_{MgCl_2} (mM)	T_M (°C) pH7	T_M (°C) pH4	T_M (°C) pH3.6	T_M (°C) pH4	T_M (°C) pH3.6
yes	yes				75	52%	50 10 5 1	95.03 93.81 92.81 88.37	66.34 65.21 64.29 60.18	37.35 36.32 35.48 31.71	50.61	22.93
yes		yes			20	50%	50 10 5 1	73.72 72.56 71.62 67.38	45.24 44.16 43.29 39.34	17.16 16.17 15.37 11.74	30.01	3.1
yes			yes		23	43%	50 10 5 1	73.06 71.91 70.97 66.75	45.01 43.94 43.07 39.15	17.18 16.2 15.4 11.79	29.91	3.26

yes				yes	32	59%	50 10 5 1	90.56 89.35 88.36 83.95	61.62 60.5 59.59 55.49	32.66 31.64 30.8 27.04	45.91	18.21
T	C123	C1	C2	C3	len. (nt)	GC%	c_{MgCl_2} (mM)	T_M (°C) pH7	T_M (°C) pH4	T_M (°C) pH3.6	T_M (°C) pH4	T_M (°C) pH3.6
yes	yes				43	44%	50 10 5 1	83.88 82.69 81.73 77.41	55.77 54.68 53.78 49.78	27.51 26.51 25.69 22.02	40.44	13.43
yes		yes			16	31%	50 10 5 1	51.66 50.57 49.69 45.69	24.21 23.2 22.38 18.64	NaN NaN NaN NaN	9.74	NaN
yes			yes		14	50%	50 10 5 1	53.71 52.61 51.71 47.65	25.15 24.13 23.3 19.49	NaN NaN NaN NaN	10.33	NaN
yes				yes	13	54%	50 10 5 1	48.1 47.01 46.12 42.11	19.32 18.31 17.48 13.7	NaN NaN NaN NaN	4.55	NaN

Table 5.5: Predicted Melting Temperatures for the HH-min System. Melting temperatures were predicted for different pH, 100 mM of monovalent ions, adapted length and GC content, and varying divalent ion concentrations. The last two columns represent melting temperatures with low concentrations of monovalent (1 mM) and divalent (0.2 mM) metal ions at pH 4 and pH 3.6. If no melting transition was observed, the melting temperature is reported as NaN. The calculations considered the longest possible double-stranded region and assumed stoichiometric conditions.

strands					len. (nt)	GC%	c_{MgCl_2} (mM)	T_M (°C) pH7	T_M (°C) pH4	T_M (°C) pH3.6	low salt	
T	HH -min	HH 1	HH 2	HH 3							T_M (°C) pH4	T_M (°C) pH3.6
yes	yes				38	55%	50 10 5 1	90.11 88.91 87.93 83.52	61.4 60.28 59.38 55.3	32.58 31.56 30.73 26.98	45.77	18.22
yes		yes			11	55%	50 10 5 1	14.91 13.92 13.11 9.41	NaN NaN NaN NaN	NaN NaN NaN NaN	NaN	NaN
yes			yes		11	55%	50 10	14.91 13.92	NaN NaN	NaN NaN	NaN	NaN

							5	13.11	NaN	NaN		
							1	9.41	NaN	NaN		
yes			yes	16	56%		50	67.67	38.17	10.16		
							10	66.53	37.11	9.19	NaN	NaN
							5	65.6	36.25	8.39		
							1	61.4	32.33	4.78		

5.4.11 Stability analysis of the ligation systems

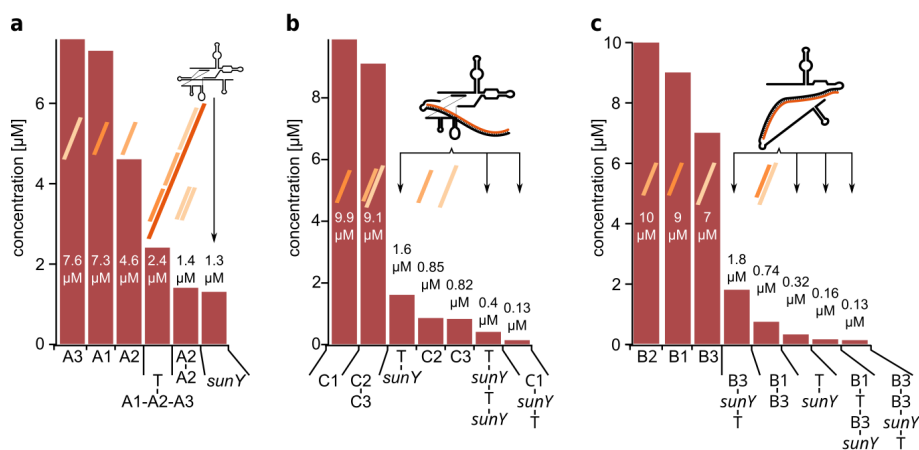


Figure 5.23: MFS simulated with NUPACK for fragmented *sunY* : NUPACK simulations were conducted to analyze the mean free energy structures for the synthesis of fragmented *sunY*. The concentrations of the ligation substrates was set to 10 μM , while the concentrations of the template and *sunY* were set to 2.5 μM each. **a** The mean free energy structures of the A fragment. As fragment A differs slightly from the corresponding region in the full-length version, the inhibitory effect of the template appears to be weak. The mean free energy structure suggests that almost all of the template is correctly bound to its ligation fragments. **b** Similarly, mean free energy structures were simulated for the synthesis of fragment C. In this case, the template and *sunY* form a complex, inhibiting the folding of *sunY* into its active structure. **c** The synthesis of fragment B was also simulated with NUPACK. Once again, the template and *sunY* form a complex, preventing the proper folding of *sunY* into its active structure.

The equilibrium base pairing for the sequence design of all systems was evaluated using the NUPACK analysis tool [217]. This algorithm is based on free energy calculations incorporating empirical data from Serra and Turner for nearest neighbor interactions [218]. It is important to note that in the simulation, the parameter for the salt concentration was set to 1 mM Na^+ , which typically underestimates the binding stability of the experimental systems due to the higher duplex stability conferred by magnesium ions. The temperature was maintained at 25 $^{\circ}\text{C}$, and the maximum complex size was limited to 4. NUPACK predicts the minimal free energy structures (MFS) and base-pairing probabilities for RNA sequences, considering the specified input parameters, i.e., salt concentration, temperature, maximum complex size, sequences, and strand concentrations. Additionally, the prediction provides

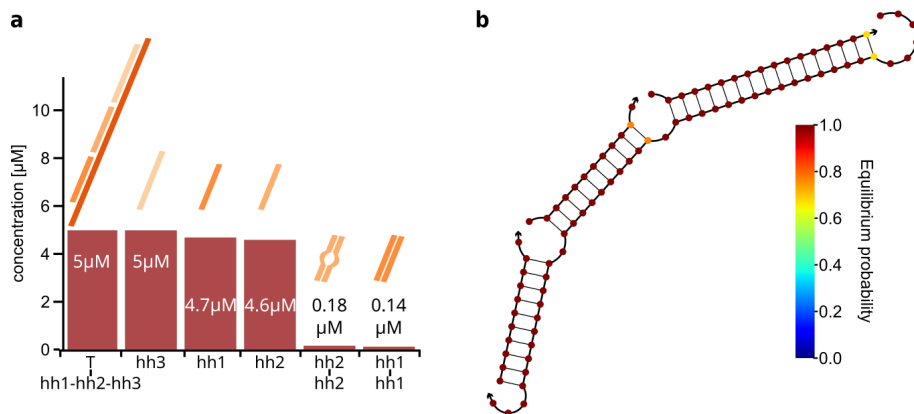


Figure 5.24: MFS simulated with NUPACK for the hammerhead system: a For concentrations of 10 μM of fragments and 5 μM of templates, all template is occupied by the three fragments allowing for the templated ligation at the simulated conditions, indicating efficient ligation conditions. **b** The MFS of the T-hh1-hh2-hh3 complex has a mean free energy of the secondary structure of -82.89 kcal/mol. The color code represents the equilibrium probability for each base to be in the shown configuration.

the equilibrium concentrations of the MFS complexes and their corresponding free energy of folding.

Several scenarios were simulated for the ABC system, taking into account the experimental conditions. The input concentrations of 2.5 μM for the template and 10 μM for each of the three ligation substrates were selected to align with the experimental setup. Additionally, the sequence of the *sunY* ribozyme was included in the simulation with an input concentration of 2.5 μM to estimate the inhibitory effect of complementary templates on *sunY*. The cutoff for the fraction of maximum concentration was set to 0.005.

Initially, the equilibrium concentrations were calculated for each ligation substrate of *sunY* individually, namely (tA123, A1, A2, A3, *sunY*), (tB, B1, B2, B3, *sunY*), and (tC123, C1, C2, C3, *sunY*). It should be noted that the A123 fragment, as reported in [126], of *sunY* used in this study has slight sequence variations compared to the full-length version used (see Table 5.2). This discrepancy is reflected in the outcome of the NUPACK simulation. The simulation shows that almost all templates are bound to all three ligation substrates, resulting in a mean free energy of -80.27 kcal/mol. On the other hand, the simulation suggests that *sunY* is mostly present in its active form and is barely inhibited. However, when simulating the mean free energy structures for tA123 and *sunY*, as well as tA123, tB123, tC123, and *sunY*, structures with free energies of -114.47 kcal/mol and -254.74 kcal/mol were obtained, respectively. This indicates that at least some of the *sunY* is inhibited by tA123. For fragments B123 and C123, the inhibitory effect, as indicated by the mean free energy structures, is significantly stronger, with a large portion of *sunY* being bound to either tB123 or tC123. These results could also explain the difference in the isothermal ligation efficiency of A123 compared to B123 and C123. To access this further, the experiments should be repeated with a *sunY* fragment A, which is similar in sequence to the full-length version.

For the hammerhead system, the input concentrations of 5 μM for the template and 10 μM for each of the three ligation substrates were selected to align with the experimental conditions. The cutoff for the fraction of maximum concentration was set to 0.005. At a temperature of 25 $^{\circ}\text{C}$, it was observed that all templates were bound to all three ligation substrates, resulting in a complete occupancy of the binding sites. The mean free energy of this fully bound structure was calculated to be -82.8 kcal/mol. This indicates a strong binding affinity between the template and the ligation substrates, suggesting favorable conditions for the ligation reaction to occur efficiently.

5.5 Appendix - Supplementary Figures

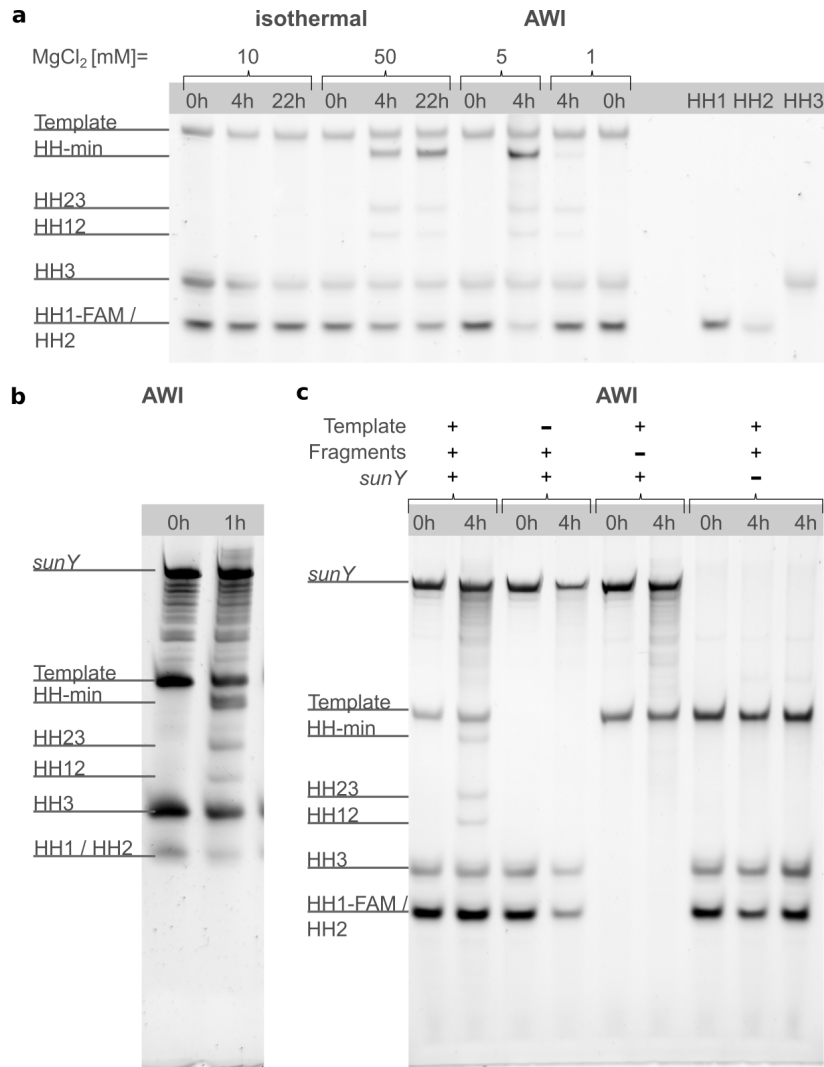


Figure 5.25: Ligation of HH-min by *sunY* : To exclude any dye artifacts, the HH-min synthesis was run with a 5' FAM label at HH1 for **a** and **c** or without any label for **b**. All reactions yielded the full-length product (HH-min). **a** Again, the reaction in the AWI-system produces significant HH-min full-length product at 5 mM MgCl₂ and even at lower magnesium concentrations of 1 mM, whereas no product was observed under isothermal conditions below 50 mM MgCl₂. **b** Synthesis using unlabeled HH1 (5 mM MgCl₂). **c** To ensure that the accumulation in the non-equilibrium did not lead to any unintended side reactions, each RNA component was omitted individually. Removing either template, substrates or *sunY* completely abolishes ligation (1 mM MgCl₂).

6 Discussion and Outlook

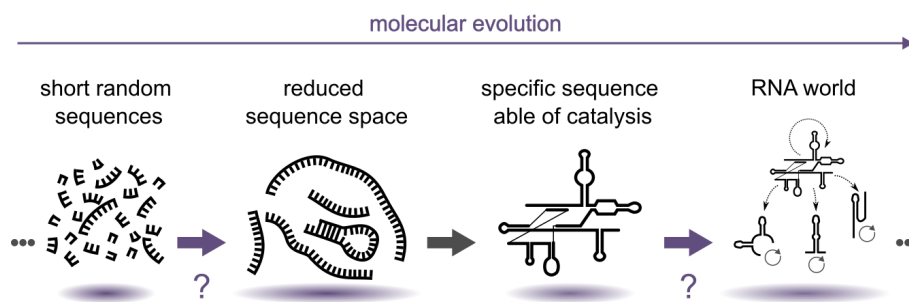


Figure 6.1: Evolution of replication: What could be the path towards an RNA world? Several evolutionary steps would potentially lead up to such a scenario. In this thesis, two intermediate steps are investigated. The first result chapter studies the emergence of long sequences with reduced sequence space compared to a random pool. It was shown how spiking short pools with subsets of sequences can completely bias the outcome. The molecular evolution of active RNA sequences is of great importance for the hypothesized RNA world. The second part of this thesis highlighted how thermal non-equilibrium settings are essential to achieve a realistic one-pot scenario of RNA replication. Physical transport effects lead to different reaction conditions within close proximity. This results in the protection of RNA from degradation. It further enables a one-pot scenario of replication and translation. Both effects are essential for an RNA world, underlining the importance of combining prebiotic chemistry with geophysical conditions in future research.

Although life certainly has more to it, a central quality is its ability to store, replicate, and translate (sequence) information. Starting from simple building blocks that polymerize into short oligomers of random sequences, molecular evolution eventually leads to the emergence of the genetic code. However, it remains heavily debated how this path could have occurred. This thesis aims to help unravel this mystery by providing experimental non-equilibrium approaches highlighting potential routes of how replication could have proceeded. One key aspect is the potential of physical non-equilibria (in this case, thermal non-equilibria) to not only aid replication chemistries but enable them in the first place. They are an essential driving force. In this chapter, the main results are summarized, connections between the three result chapters are discussed, and potential routes for future research are highlighted.

Figure 6.1, a slightly expanded version of Figure 3.1, shows how molecular evolution of replication presumably underwent several steps. Chapter 3 has addressed the question of how pools of short random sequences can lead to longer product sequences when exposed to the

right temperature cycling conditions. The sequence space of the synthesized products deviates from what is expected of random pools. The sequence space of the products, as well as the elongation behavior, can be influenced by spiked random pools either with short sequence networks or long spikes. This shows that even though major characteristics of the product sequence space, such as the separation into A- and T-types, are conserved, biases are amplified and lead to significant differences between the pools. Thus, a more realistic scenario with random pools of slightly different initial compositions that have evolved, e.g., in different rock pores, could lead to multiple parallel evolutionary experiments, yielding different product pools. Even more interesting results would emerge if mixing events occur, for example, with two spikes competing for the same resources. Further, chapter 3 has shown that given enough time to ligate and overall cycles, pools of sequences with all four bases can extend to longer products. This is exciting because all ribozymes known to date consist of sequences with all four bases, suggesting that they are required for the emergence of functional sequences. Although similar types of experiments might provide a starting pool that potentially contains a catalytically active sequence, different types of experiments need to be employed for its selection and amplification.

Similar to other recent DNA evolution experiments that use DNA pools with a *BST* polymerase [219], the results reported in chapter 3 on DNA pools and the *Taq* ligase can only serve as a model system for a prebiotic replication chemistry. Although these model systems provide valuable expertise on potential replication dynamics, experimental methods, and data analysis, the specific experimental insights must be translated and reinterpreted once a more realistic and efficient replication chemistry is identified. Promising candidates are the RNA ligation chemistry reported by [48], employing a N-alkyl imidazole organocatalyst at the 5'-end and a 3'-amino group, as well as the RNA ligation chemistry reported by [91], based on a ligation chemistry mediated by a 2',3' cyclic phosphate. Both are non-enzymatic ligation chemistries and, thus, face similar problems. First, they naturally exhibit significantly lower ligation rates compared to enzymatic ligation reactions, making cycling experiments with hundreds of cycles a long endeavor. As with any RNA replication, they are confronted with a much higher rate of backbone hydrolysis, especially at the high temperatures needed for strand separation. Additionally, high temperatures might lead to a deactivation of the respective reactive groups. Thus, evolution experiments, ideally starting from random pools, could make use of not just thermal cycling but more realistic settings such as the ones described in chapters 4 and 5. Especially the air-water interface (AWI) system could provide significant advantages, as it provides a mechanism allowing strand separation at lower temperatures.

However, experiments employing the imidazole-based ligation chemistry in the AWI-system showed that a high concentration of ligation substrates, induced by the accumulation, leads to untemplated ligation [220]. While it is unclear to what extent sequence information can still be copied in a templated manner, similar untemplated ligation reactions would at least lead to a high background reaction rate, producing longer oligomer strands. Even though experiments similar to the random DNA *Taq* ligase are impossible to achieve with this ligation chemistry, it might still be a promising route from shorter to longer strands.

Analogous replication experiments with the ligation chemistry based on the activation of strands with a 2',3' cyclic phosphate could be run in the AWI-system. It has been previously shown that polymerization experiments based on the same activation chemistry can benefit

from the AWI-settings [105]. This provides an appealing connection between polymerization reactions of 2',3' cyclic monomers to short sequences and templated ligation reactions that allow the transfer of sequence information. The key to these experiments' success is to run them in AWI-systems for multiple days to counteract the slow ligation rate. Recent updates of the experimental setting now incorporate an automatized CO₂ influx, which makes long-term experiments possible. Additionally, enough active ligation substrates need to be provided throughout the experiment. On the one hand, this could be achieved by adapting the AWI-system to a point-like heat source at the interface, i.e., combining the advantages of a small hot temperature region (chapter 4) and an interface setting (chapter 5). On the other hand, the system could be expanded by an in- and outflow of fresh ligation substrates and 'burned' material, respectively. Further, it could be interesting to build a ribozyme, e.g., *sunY*, from a non-enzymatic replication chemistry such as the 2',3' cyclic phosphate ligation and exploit the strand separation capacity of the AWI-system as shown in chapter 5. One possibility is to start with a fragmented *sunY* and activate the ends with 2',3' cyclic phosphates, which could connect the fragments to a full version. While the fragmented version can already catalyze ligation reactions, ligating it to a full version would likely result in a rate enhancement.

From another perspective, a realistic replication scenario for any ribozyme most likely has to function within a random background of sequences. Preliminary experiments have demonstrated that the *sunY* ribozyme is able to act on random pools within AWI-systems [221]. In these experiments, a Cy-5 labeled primer and its corresponding template were placed into a pool of random sequences (G-N₇-U; with N=A, T, G, C). Indeed, it was possible to detect the extension of the primer with sequences of the random pool. However, it remains uncertain how much of the template information could be copied in such a scenario as the concentration of complementary sequences of the random pool to the template is very low. However, including a random pool in ligation experiments of the hammerhead ribozyme as described in chapter 5, where all ligation substrates needed for the ligation are present in higher concentration, would allow to infer the robustness and fidelity of these replication reactions. Further, it would be worth exploring whether experiments similar to the ones described in chapter 3 are possible with this system. Especially interesting are potential sequence biases and different sequence selection criteria introduced by the *sunY* ribozyme that would, for example, provide insight into the sequence generality of the ribozyme.

To expand on the concept of an RNA world hosted by an AWI-system as discussed in chapter 5, an obvious next step is to ligate a range of different ribozymes within the same compartment, each subsequently performing their respective reaction. One could even imagine a reaction setup where two ribozymes of significantly different lengths are ligated from the same number of ligation substrates, which consequently also vary in length. At low overall temperatures, all ligation substrates could bind and ligate, but only the shorter ribozyme could separate from its template. On the other hand, increasing the overall temperatures would only allow the longer ligation substrates to bind, resulting in the ligation and strand separation of the longer ribozyme. Thus, variations in the temperatures of the gradient could provide means to regulate which ribozyme is 'expressed'. Although feedback loops, as present in the regulatory effects of biology, would still be missing, this would further attest to the potential of non-equilibrium systems in the environment of the early Earth to enable multiple concepts of modern biology.

In conclusion, the results of this thesis underline the potential of thermal non-equilibrium systems to enable evolutionary processes of early replication. As life is undoubtedly a far from equilibrium process, energy will always be necessary for its emergence and maintenance. It seems highly plausible that nature had to make use of abundant energy sources and non-equilibria during the course of molecular evolution. At the same time, geophysical constraints can provide the first selection pressures. Thus, employing more and various kinds of non-equilibrium settings in evolution experiments, combining the expertise of multiple disciplines, and progressively adding complexity to the experiments seems to be the most promising path towards unraveling the history of life on Earth.

7 Bibliography

- [1] Matthew S. Dodd, Dominic Papineau, Tor Grenne, John F. Slack, Martin Rittner, Franco Pirajno, Jonathan O’Neil, and Crispin T.S. Little. Evidence for early life in Earth’s oldest hydrothermal vent precipitates. *Nature*, 543(7643):60–64, 2017.
- [2] Allen P. Nutman, Vickie C. Bennett, Clark R.L. Friend, Martin J. Van Kranendonk, and Allan R. Chivas. Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. *Nature*, 537(7621):535–538, 2016.
- [3] S. J. Mojzsis, G. Arrhenius, K. D. McKeegan, T. M. Harrison, A. P. Nutman, and C. R. L. Friend. Evidence for life on Earth before 3,800 million years ago. *Nature*, 384(6604):55–59, 1996.
- [4] H. Genda, R. Brasser, and S.J. Mojzsis. The terrestrial late veneer from core disruption of a lunar-sized impactor. *Earth and Planetary Science Letters*, 480:25–32, dec 2017.
- [5] R. Brasser, S.C. Werner, and S.J. Mojzsis. Impact bombardment chronology of the terrestrial planets from 4.5 Ga to 3.5 Ga. *Icarus*, 338:113514, mar 2020.
- [6] Simon J. Lock, Sarah T. Stewart, Michail I. Petaev, Zoë Leinhardt, Mia T. Mace, Stein B. Jacobsen, and Matija Cuk. The Origin of the Moon Within a Terrestrial Synestia. *Journal of Geophysical Research: Planets*, 123(4):910–951, apr 2018.
- [7] M. Cuk and Sarah T. Stewart. Making the Moon from a Fast-Spinning Earth: A Giant Impact Followed by Resonant Despinning. *Science*, 338(6110):1047–1052, nov 2012.
- [8] Xiaozhi Yang, Fabrice Gaillard, and Bruno Scaillet. A relatively reduced Hadean continental crust and implications for the early atmosphere and crustal rheology. *Earth and Planetary Science Letters*, 393:210–219, 2014.
- [9] George L. Hashimoto, Yutaka Abe, and Seiji Sugita. The chemical composition of the early terrestrial atmosphere: Formation of a reducing atmosphere from CI-like material. *Journal of Geophysical Research E: Planets*, 112(5):1–12, 2007.
- [10] Holly C Betts, Mark N Puttick, James W Clark, Tom A Williams, C J Philip, and Davide Pisani. Europe PMC Funders Group Integrated genomic and fossil evidence illuminates life’s early evolution and eukaryote origins. 2018(10):1556–1562, 2019.
- [11] Madeline C. Weiss, Filipa L. Sousa, Natalia Mrnjavac, Sinje Neukirchen, Mayo Roettger, Shijulal Nelson-Sathi, and William F. Martin. The physiology and habitat of the last universal common ancestor. *Nature Microbiology*, 1(9):1–8, 2016.
- [12] Madeline C. Weiss, Martina Preiner, Joana C. Xavier, Verena Zimorski, and William F. Martin. The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLoS Genetics*, 14(8):1–19, 2018.
- [13] Eugene V. Koonin. The Biological Big Bang model for the major transitions in evolution. *Biology Direct*, 2(1):21, 2007.
- [14] David A. Liberles. *Ancestral Sequence Reconstruction*. Oxford University Press, may 2007.
- [15] Anton S. Petrov, Burak Gulen, Ashlyn M. Norris, Nicholas A. Kovacs, Chad R. Bernier, Kathryn A. Lanier, George E. Fox, Stephen C. Harvey, Roger M. Wartell, Nicholas V. Hud, and Loren Dean Williams. History of the ribosome and the origin of translation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(50):15396–15401, 2015.
- [16] Helmut Griesser, Peter Tremmel, Eric Kervio, Camilla Pfeffer, Ulrich E. Steiner, and Clemens Richert. Ribonucleotides and RNA Promote Peptide Chain Growth. *Angewandte Chemie - International Edition*, 129(5):1239–1243, 2017.
- [17] Moran Frenkel-Pinter, Jay W. Haynes, Martin C, Anton S. Petrov, Bradley T. Burcar, Ramnarayanan Krishnamurthy, Nicholas V. Hud, Luke J. Leman, and Loren Dean Williams. Selective incorporation of proteinaceous over non-proteinaceous cationic amino acids in model prebiotic oligomerization reactions. *Proceedings of the National Academy of Sciences*, 116(33):16338–16346, 2019.

- [18] Walter Gilbert. Origin of life: The RNA world. *Nature*, 319(6055):618–618, feb 1986.
- [19] Michael P. Robertson and Gerald F. Joyce. The Origins of the RNA World. *Cold Spring Harbor Perspectives in Biology*, 4(5):a003608–a003608, may 2012.
- [20] Kristian Le Vay and Hannes Mutschler. The difficult case of an RNA-only origin of life. *Emerging Topics in Life Sciences*, 3(5):469–475, nov 2019.
- [21] Gerald F. Joyce. RNA evolution and the origins of life. *Nature*, 338(6212):217–224, mar 1989.
- [22] Nicholas V. Hud, Brian J. Cafferty, Ramanarayanan Krishnamurthy, and Loren Dean Williams. The origin of rna and "my grandfather's axe", 2013.
- [23] Albert C. Fahrenbach, Constantin Giurgiu, Chun Pong Tam, Li Li, Yayoi Hongo, Masashi Aono, and Jack W. Szostak. Common and Potentially Prebiotic Origin for Precursors of Nucleotide Synthesis and Activation. *Journal of the American Chemical Society*, 139(26):8780–8783, 2017.
- [24] Jack W. Szostak. An optimal degree of physical and chemical heterogeneity for the origin of life? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1580):2894–2901, 2011.
- [25] Ramanarayanan Krishnamurthy. Experimentally investigating the origin of DNA/RNA on early Earth. *Nature Communications*, 9(1):5175, 2018.
- [26] Brian J. Cafferty, David M. Fialho, Jaheda Khanam, Ramanarayanan Krishnamurthy, and Nicholas V. Hud. Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. *Nature Communications*, 7(1):11328, sep 2016.
- [27] John D. Sutherland. The Origin of Life - Out of the Blue. *Angewandte Chemie - International Edition*, 55(1):104–121, 2016.
- [28] Matthew W. Powner and John D. Sutherland. Prebiotic chemistry: A new modus operandi. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1580):2870–2877, 2011.
- [29] John D. Sutherland. Opinion: Studies on the origin of life — the end of the beginning. *Nature Reviews Chemistry*, 1(2):0012, feb 2017.
- [30] Mahipal Yadav, Ravi Kumar, and Ramanarayanan Krishnamurthy. Chemistry of abiotic nucleotide synthesis, 2020.
- [31] Hyo-Joong Kim and Steven A. Benner. A Direct Prebiotic Synthesis of Nicotinamide Nucleotide. *Chemistry - A European Journal*, 24(3):581–584, jan 2018.
- [32] Steven A. Benner, Hyo-Joong Kim, and Elisa Biondi. Prebiotic chemistry that could not not have happened. *Life*, 9(4):84, nov 2019.
- [33] Hidenori Okamura, Sidney Becker, Niklas Tiede, Stefan Wiedemann, Jonas Feldmann, and Thomas Carell. A one-pot, water compatible synthesis of pyrimidine nucleobases under plausible prebiotic conditions. *Chemical Communications*, 55(13):1939–1942, 2019.
- [34] Sidney Becker, Christina Schneider, Antony Crisp, and Thomas Carell. Non-canonical nucleosides and chemistry of the emergence of life. *Nature Communications*, 9(1):5174, 2018.
- [35] Sidney Becker, Jonas Feldmann, Stefan Wiedemann, Hidenori Okamura, Christina Schneider, Katharina Iwan, Antony Crisp, Martin Rossa, Tynchtyk Amatov, and Thomas Carell. Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science*, 366(6461):76–82, 2019.
- [36] Jennifer S. Teichert, Florian M. Kruse, and Oliver Trapp. Direct Prebiotic Pathway to DNA Nucleosides. *Angewandte Chemie - International Edition*, 58(29):9944–9947, 2019.
- [37] Noam Prywes, J. Craig Blain, Francesca Del Frate, and Jack W Szostak. Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *eLife*, 5(e17756), jun 2016.
- [38] Christopher Deck, Mario Jauker, and Clemens Richert. Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nature Chemistry*, 3(8):603–608, 2011.
- [39] Liqin Zhang, Zunyi Yang, Kwame Sefah, Kevin M. Bradley, Shuichi Hoshika, Myong Jung Kim, Hyo Joong Kim, Guizhi Zhu, Elizabeth Jiménez, Sena Cansiz, I. Ting Teng, Carole Champanhac, Christopher Mclendon, Chen Liu, Wen Zhang, Dietlind L. Gerloff, Zhen Huang, Weihong Tan, and Steven A. Benner. Evolution of functional six-nucleotide DNA. *Journal of the American Chemical Society*, 137(21):6734–6737, 2015.
- [40] Travis Walton, Wen Zhang, Li Li, Chun Pong Tam, and Jack W. Szostak. The Mechanism of Nonenzymatic Template Copying with Imidazole-Activated Nucleotides. *Angewandte*

- Chemie - International Edition*, 58(32):10812–10819, aug 2019.
- [41] Matthias Morasch, Christof B. Mast, Johannes K. Langer, Pierre Schilcher, and Dieter Braun. Dry polymerization of 3,5-cyclic GMP to long strands of RNA. *ChemBioChem*, 15(6):879–883, 2014.
- [42] M. S. Verlander, R. Lohrmann, and L. E. Orgel. Catalysts for the self-polymerization of adenosine cyclic 2,3-phosphate. *Journal of Molecular Evolution*, 2(4):303–316, 1973.
- [43] Kamila B. Muchowska and Joseph Moran. Peptide synthesis at the origin of life. *Science*, 370:767–768, 11 2020.
- [44] Günter Wächtershäuser. On the chemistry and evolution of the pioneer organism, 4 2007.
- [45] Long-Fei Wu and John D. Sutherland. Provisioning the origin and early evolution of life. *Emerging Topics in Life Sciences*, 3(5):459–468, 2019.
- [46] James M. Smith, Vinciane Borsenberger, Jim Raftery, and John D. Sutherland. Exploratory studies to investigate a linked prebiotic origin of RNA and coded peptides: Derivation and reactivity of xylose phosphates. *Chemistry and Biodiversity*, 1(10):1418–1451, 2004.
- [47] Alexandra Kühnlein, Simon A Lanzmich, and Dieter Braun. tRNA sequences can assemble into a replicator. *eLife*, 10:1–22, mar 2021.
- [48] Lijun Zhou, Derek K. O’Flaherty, and Jack W. Szostak. Template-Directed Copying of RNA by Non-enzymatic Ligation. *Angewandte Chemie - International Edition*, 59(36):15682–15687, 2020.
- [49] Marilyne Sosson, Daniel Pfeffer, and Clemens Richert. Enzyme-free ligation of dimers and trimers to RNA primers. *Nucleic Acids Research*, 47(8):3836–3845, may 2019.
- [50] Markus Kramer and Clemens Richert. Enzyme-Free Ligation of 5-Phosphorylated Oligodeoxynucleotides in a DNA Nanostructure. *Chemistry Biodiversity*, 14(9):e1700315, sep 2017.
- [51] Evgeniia Edeleva, Annalena Salditt, Julian Stamp, Philipp Schwintek, Job Boekhoven, and Dieter Braun. Continuous nonenzymatic cross-replication of DNA strands with in situ activated DNA oligonucleotides. *Chemical Science*, 10(22):5807–5814, 2019.
- [52] Jack W. Szostak. The eightfold path to non-enzymatic RNA replication. *Journal of Systems Chemistry*, 3(1):1–14, 2012.
- [53] Wim Hordijk, Shira Shichor, and Gonen Ashkenasy. The Influence of Modularity, Seeding, and Product Inhibition on Peptide Autocatalytic Network Dynamics. *ChemPhysChem*, 19(18):2437–2444, 2018.
- [54] Natasha Paul and Gerald F. Joyce. Minimal self-replicating systems. *Current Opinion in Chemical Biology*, 8(6):634–639, dec 2004.
- [55] Günter von Kiedrowski. Minimal Replicator Theory I: Parabolic Versus Exponential Growth. pages 113–146. 1993.
- [56] A. J. Cridland, Ralph E. Pudritz, Tilman Birnstiel, L. Ilesdore Cleeves, and Edwin A. Bergin. Composition of early planetary atmospheres – II. Coupled Dust and chemical evolution in protoplanetary discs. *Monthly Notices of the Royal Astronomical Society*, 469(4):3910–3927, aug 2017.
- [57] Christian T. Lenz, Hubert Klahr, Tilman Birnstiel, Katherine Kretke, and Sebastian Stammer. Constraining the parameter space for the solar nebula. *Astronomy Astrophysics*, 640:A61, 8 2020.
- [58] Jennifer B. Bergner, Rafael Martín-Doménech, Karin I. Öberg, Jes K. Jørgensen, Elizabeth Artur De La Villarmois, and Christian Brinch. Organic Complexity in Protostellar Disk Candidates. *ACS Earth and Space Chemistry*, 3(8):1564–1575, aug 2019.
- [59] Karin I. Öberg and Edwin A Bergin. Astrochemistry and compositions of planetary systems. *Physics Reports*, 893(3):1–48, jan 2021.
- [60] John P. Grotzinger, Joy Crisp, Ashwin R. Vasavada, Robert C. Anderson, Charles J. Baker, Robert Barry, David F. Blake, Pamela Conrad, Kenneth S. Edgett, Bobak Ferdowski, Ralf Gellert, John B. Gilbert, Matt Golombek, Javier Gómez-Elvira, Donald M. Hassler, Louise Jandura, Maxim Litvak, Paul Mahaffy, Justin Maki, Michael Meyer, Michael C. Malin, Igor Mitrofanov, John J. Simmonds, David Vaniman, Richard V. Welch, and Roger C. Wiens. Mars Science Laboratory Mission and Science Investigation. *Space Science Reviews*, 170(1-4):5–56, sep 2012.
- [61] G. Israël, C. Szopa, F. Raulin, M. Cabane, H. B. Niemann, S. K. Atreya, S. J. Bauer, J. F. Brun, E. Chassefière, P. Coll, E. Condé, D. Coscia, A. Hauchecorne, P. Millian, M. J. Nguyen, T. Owen, W. Riedler, R. E. Samuelson, J. M. Sigurier, M. Steller, R. Sternberg, and C. Vidal-Madjar. Complex organic matter in Titan’s atmospheric aerosols from in situ pyrolysis and analysis. *Nature*, 438(7069):796–799, 2005.

- [62] H. James Cleaves. Prebiotic Chemistry: What We Know, What We Don't. *Evolution: Education and Outreach*, 5:342–360, sep 2012.
- [63] Sukrit Ranjan and Dimitar D. Sasselov. Constraints on the Early Terrestrial Surface UV Environment Relevant to Prebiotic Chemistry. *Astrobiology*, 17(3):169–204, mar 2017.
- [64] Carlos E Crespo-Hernández and Rafael Arce. Formamidopyrimidines as major products in the low- and high-intensity UV irradiation of guanine derivatives. *Journal of Photochemistry and Photobiology B: Biology*, 73(3):167–175, 2004.
- [65] Zoe R. Todd, Albert C. Fahrenbach, Sukrit Ranjan, Christopher J. Magnani, Jack W. Szostak, and Dimitar D. Sasselov. Ultraviolet-Driven Deamination of Cytidine Ribonucleotides under Planetary Conditions. *Astrobiology*, 20(7):878–888, 2020.
- [66] Stephen J Mojzsis, T Mark Harrison, and Robert T Pidgeon. Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature*, 409(6817):178–181, jan 2001.
- [67] Hartmut Follmann and Carol Brownson. Darwin's warm little pond revisited: From molecules to the origin of life. *Naturwissenschaften*, 96(11):1265–1292, 2009.
- [68] J. B. Corliss, J.A. Baross, and S. E. Hoffman. An hypothesis concerning the relationship between submarine hot springs and the origin of life on earth. *Oceanol. Acta*, special issue:59–69, 1981.
- [69] Michael J. Russell, Roy M. Daniel, and Allan J Hall. On the emergence of life via catalytic iron-sulphide membranes. *Terra Nova*, 5(4):343–347, jul 1993.
- [70] William Martin, John Baross, Deborah Kelley, and Michael J. Russell. Hydrothermal vents and the origin of life, 9 2008.
- [71] Fabian B. Wadsworth, Jérémie Vasseur, Bettina Scheu, Jackie E. Kendrick, Yan Lavallée, and Donald B. Dingwell. Universal scaling of fluid permeability during volcanic welding and sediment diagenesis. *Geology*, 44(3):219–222, 2016.
- [72] Mathieu Colombier, Fabian B. Wadsworth, Lucia Gurioli, Bettina Scheu, Ulrich Kueppers, Andrea Di Muro, and Donald B. Dingwell. The evolution of pore connectivity in volcanic rocks. *Earth and Planetary Science Letters*, 462:99–109, 2017.
- [73] Dieter Braun, Noel L Goddard, and Albert Libchaber. Exponential DNA Replication by Laminar Convection. *Physical Review Letters*, 91(15):158103, oct 2003.
- [74] Dieter Braun. PCR by Thermal Convection. *Modern Physics Letters B*, 18(16):775–784, 2004.
- [75] Maren Reichl, Mario Herzog, Alexandra Götz, and Dieter Braun. Why Charged Molecules Move Across a Temperature Gradient: The Role of Electric Fields. *Physical Review Letters*, 112(19):198101, may 2014.
- [76] D. Niether and S. Wiegand. Thermophoresis of biological and biocompatible compounds in aqueous solution. *Journal of Physics: Condensed Matter*, 31(50):503003, dec 2019.
- [77] Jérôme Burelbach, Daan Frenkel, Ignacio Pagonabarraga, and Erika Eiser. A unified description of colloidal thermophoresis. *The European Physical Journal E*, 41(1):7, jan 2018.
- [78] Moritz Kreysing, Lorenz Keil, Simon Lanzmich, and Dieter Braun. Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nature Chemistry*, 7(3):203–208, mar 2015.
- [79] Friederike M. Möller, Franziska Kriegel, Michael Kieß, Victor Sojo, and Dieter Braun. Steep pH Gradients and Directed Colloid Transport in a Microfluidic Alkaline Hydrothermal Pore. *Angewandte Chemie International Edition*, 56(9):2340–2344, feb 2017.
- [80] James Attwater, Aniela Wochner, Vitor B. Pinheiro, Alan Coulson, and Philipp Holliger. Ice as a protocellular medium for RNA replication. *Nature Communications*, 1(1):76, sep 2010.
- [81] 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, jan 2020.
- [82] Hannes Mutschler, Aniela Wochner, and Philipp Holliger. Freeze-thaw cycles as drivers of complex ribozyme assembly. *Nature Chemistry*, 7(6):502–508, jun 2015.
- [83] Alan Ianeselli, Christof B. Mast, and Dieter Braun. Periodic Melting of Oligonucleotides by Oscillating Salt Concentrations Triggered by Microscale Water Cycles Inside Heated Rock Pores. *Angewandte Chemie*, 131(37):13289–13294, sep 2019.

- [84] Alan Inaneselli, Miguel Atienza, Patrick W Kudella, Ulrich Gerland, Christof B Mast, and Dieter Braun. Water cycles in a Hadean CO₂ atmosphere drive the evolution of long DNA. *Nature Physics*, 18(5):579–585, mar 2022.
- [85] Björn Drobot, Juan M. Iglesias-Artola, Kristian Le Vay, Viktoria Mayr, Mrityunjoy Kar, Moritz Kreysing, Hannes Mutschler, and T-Y Dora Tang. Compartmentalised rna catalysis in membrane-free coacervate protocells. *Nature communications*, 9(1):3643, 2018.
- [86] T. Y. Dora Tang, C. Rohaida Che Hak, Alexander J. Thompson, Marina K. Kuimova, D. S. Williams, Adam W. Perriman, and Stephen Mann. Fatty acid membrane assembly on coacervate microdroplets as a step towards a hybrid protocell model. *Nature Chemistry*, 6(6):527–533, 2014.
- [87] Dirk van Swaay, T.-Y. Dora Tang, Stephen Mann, and Andrew de Mello. Microfluidic Formation of Membrane-Free Aqueous Coacervate Droplets in Water. *Angewandte Chemie International Edition*, 54(29):8398–8401, jul 2015.
- [88] Alan Inaneselli, Damla Tetiker, Julian Stein, Alexandra Kühnlein, Christof B. Mast, Dieter Braun, and T.-Y. Dora Tang. Non-equilibrium conditions inside rock pores drive fission, maintenance and selection of coacervate protocells. *Nature Chemistry*, 14(1):32–39, jan 2022.
- [89] Matthias Morasch, Dieter Braun, and Christof B. Mast. Heat-Flow-Driven Oligonucleotide Gelation Separates Single-Base Differences. *Angewandte Chemie International Edition*, 55(23):6676–6679, jun 2016.
- [90] Giacomo Bartolucci, Adriana Calaça Serrão, Philipp Schwintek, Alexandra Kühnlein, Yash Rana, Philipp Janto, Dorothea Hofer, Christof B Mast, Dieter Braun, and Christoph A Weber. Sequence self-selection by cyclic phase separation. *Proceedings of the National Academy of Sciences*, 120(43):e2218876120, 2023.
- [91] Adriana Calaça Serrão, Sreekar Wunnava, Avinash V Dass, Lennard Ufer, Philipp Schwintek, Christof B. Mast, and Dieter Braun. High-Fidelity Templated Ligation of RNA via 2',3'-cyclic Phosphate. *ChemRxiv*, pages 1–7, 2023.
- [92] Tobias Reichenbach, Mauro Mobilia, and Erwin Frey. Noise and correlations in a spatial population model with cyclic competition. *Physical Review Letters*, 99(23):238105, 2007.
- [93] Erwin Frey. Evolutionary game theory: Theoretical concepts and applications to microbial communities. *Physica A: Statistical Mechanics and its Applications*, 389(20):4265–4298, 2010.
- [94] Kevin Leu, Benedikt Obermayer, Sudha Rajamani, Ulrich Gerland, and Irene A. Chen. The prebiotic evolutionary advantage of transferring genetic information from RNA to DNA. *Nucleic Acids Research*, 39(18):8135–8147, 2011.
- [95] Alexei V Tkachenko and Sergei Maslov. Spontaneous emergence of autocatalytic information-coding polymers. *The Journal of Chemical Physics*, 143(4):045102, jul 2015.
- [96] Alexei V. Tkachenko and Sergei Maslov. Onset of natural selection in populations of autocatalytic heteropolymers. *The Journal of Chemical Physics*, 149(13):134901, oct 2018.
- [97] Sara Imari Walker, Martha A. Grover, and Nicholas V. Hud. Universal sequence replication, reversible polymerization and early functional biopolymers: A model for the initiation of prebiotic sequence evolution. *PLoS ONE*, 7(4):31–37, 2012.
- [98] Joachim H. Rosenberger, Tobias Göppel, Patrick W. Kudella, Dieter Braun, Ulrich Gerland, and Bernhard Altaner. Self-Assembly of Informational Polymers by Templated Ligation. *Physical Review X*, 11(3):031055, sep 2021.
- [99] Farshid Jafarpour, Tommaso Biancalani, and Nigel Goldenfeld. Noise-Induced Mechanism for Biological Homochirality of Early Life Self-Replicators. *Physical Review Letters*, 115(15):158101, oct 2015.
- [100] Galen T. Pickett, Mark Gross, and Hiroko Okuyama. Spontaneous Chirality in Simple Systems. *Physical Review Letters*, 85(17):3652–3655, oct 2000.
- [101] S. Duhr and D. Braun. Why molecules move along a temperature gradient. *Proceedings of the National Academy of Sciences*, 103(52):19678–19682, dec 2006.
- [102] Hua Hu and Ronald G. Larson. Marangoni effect reverses coffee-ring depositions. *The Journal of Physical Chemistry B*, 110(14):7090–7094, 4 2006.
- [103] Matthias Morasch, Jonathan Liu, Christina F. Dirscherl, Alan Inaneselli, Alexandra Kühnlein, Kristian Le Vay, Philipp Schwintek, Saidul Islam, Mérina K. Corpinot, Bettina Scheu, Donald B. Dingwell, Petra Schwille, Hannes Mutschler,

- Matthew W. Powner, Christof B. Mast, and Dieter Braun. Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules. *Nature Chemistry*, 11(9):779–788, sep 2019.
- [104] Christina F. Dirscherl, Alan Ianeselli, Damla Tetiker, Thomas Matreux, Robbin M. Queener, Christof B. Mast, and Dieter Braun. A heated rock crack captures and polymerizes primordial DNA and RNA. *Physical Chemistry Chemical Physics*, 25(4):3375–3386, 2023.
- [105] Avinash Vicholous Dass, Sreekar Wunnava, Juliette Langlais, Beatriz von der Esch, Maik Krusche, Lennard Ufer, Nico Chrisam, Romeo C. A. Dubini, Florian Gartner, Severin Angerpointner, Christina F. Dirscherl, Petra Rovó, Christof B. Mast, Judit E. Šponer, Christian Ochsenfeld, Erwin Frey, and Dieter Braun. RNA Oligomerisation without Added Catalyst from 2,3-Cyclic Nucleotides by Drying at Air-Water Interfaces**. *ChemSystemsChem*, 5(1):1–8, 2023.
- [106] Daniel Duzdevich, Christopher E. Carr, and Jack W. Szostak. Deep sequencing of non-enzymatic RNA primer extension. *Nucleic Acids Research*, 48(12):E70–E70, 2020.
- [107] D. Sievers and G. von Kiedrowski. Self-replication of complementary nucleotide-based oligomers. *Nature*, 369(6477):221–224, may 1994.
- [108] Lijun Zhou, Seohyun Chris Kim, Katherine H. Ho, Derek K O’Flaherty, Constantin Giurgiu, Tom H. Wright, and Jack W. Szostak. Non-enzymatic primer extension with strand displacement. *eLife*, 8, nov 2019.
- [109] Lijun Zhou, Derek K. O’Flaherty, and Jack W. Szostak. Assembly of a Ribozyme Ligase from Short Oligomers by Nonenzymatic Ligation. *Journal of the American Chemical Society*, 142(37):15961–15965, 2020.
- [110] Shoichi Toyabe and Dieter Braun. Cooperative Ligation Breaks Sequence Symmetry and Stabilizes Early Molecular Replication. *Physical Review X*, 9(1):011056, 2019.
- [111] Patrick W. Kudella, Alexei V. Tkachenko, Analena Salditt, Sergei Maslov, and Dieter Braun. Structured sequences emerge from random pool when replicated by templated ligation. *Proceedings of the National Academy of Sciences*, 118(8):e2018830118, feb 2021.
- [112] 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):1–39, 2022.
- [113] Tobias Göppel, Benedikt Obermayer, Irene A Chen, and Ulrich Gerland. A kinetic error filtering mechanism for enzyme-free copying of nucleic acid sequences. *bioRxiv*, 8 2021.
- [114] Thomas R. Cech. The Ribosome Is a Ribozyme. *Science*, 289(5481):878–879, aug 2000.
- [115] Bradley P. Klemm, Nancy Wu, Yu Chen, Xin Liu, Kipchumba J. Kaitany, Michael J. Howard, and Carol A. Fierke. The diversity of ribonuclease P: Protein and RNA catalysts with analogous biological functions. *Biomolecules*, 6(2):27, 2016.
- [116] Olga Esakova and Andrey S. Krasilnikov. Of proteins and RNA: The RNase P/MRP family. *RNA*, 16(9):1725–1747, 2010.
- [117] William G. Scott, James B. Murray, John R P Arnold, Barry L. Stoddard, and Aaron Klug. Capturing the structure of a catalytic RNA intermediate: The hammerhead ribozyme. *Science*, 274(5295):2065–2069, 1996.
- [118] Klara R. Birikh, Paul A. Heaton, and Fritz Eckstein. The structure, function and application of the hammerhead ribozyme. *European Journal of Biochemistry*, 245(1):1–16, 1997.
- [119] Natasha Paul and Gerald F. Joyce. A self-replicating ligase ribozyme. *Proceedings of the National Academy of Sciences*, 99(20):12733–12740, oct 2002.
- [120] E. J. Hayden, C. A. Riley, A. S. Burton, and N. Lehman. RNA-directed construction of structurally complex and active ligase ribozymes through recombination. *RNA*, 11(11):1678–1687, nov 2005.
- [121] Elena Puerta-Fernández, Cristina Romero-López, Alicia Barroso-delJesus, and Alfredo Berzal-Herranz. Ribozymes: recent advances in the development of RNA tools. *FEMS Microbiology Reviews*, 27(1):75–97, apr 2003.
- [122] Steven M. Nesbitt, Heidi A. Erlacher, and Martha J. Fedor. The internal equilibrium of the hairpin ribozyme: temperature, ion and pH effects. *Journal of Molecular Biology*, 286(4):1009–1024, mar 1999.
- [123] Wendy K. Johnston, Peter J. Unrau, Michael S. Lawrence, Margaret E. Glasner, and David P. Bartel. RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. *Science*, 292(5520):1319–1325, 5 2001.

- [124] Gerald F Joyce. Directed evolution of nucleic acid enzymes. *Annual review of biochemistry*, 73(1):791–836, 2004.
- [125] David P. Horning and Gerald F. Joyce. Amplification of RNA by an RNA polymerase ribozyme. *Proceedings of the National Academy of Sciences*, 113(35):9786–9791, aug 2016.
- [126] J. Doudna, S Couture, and J. Szostak. A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science*, 251(5001):1605–1608, mar 1991.
- [127] Lijun Zhou, Dian Ding, and Jack W. Szostak. The virtual circular genome model for primordial RNA replication. *RNA*, 27(1):1–11, jan 2021.
- [128] Marcel Hollenstein. DNA catalysis: The chemical repertoire of DNAzymes. *Molecules*, 20(11):20777–20804, 2015.
- [129] E Westhof and A. Lescoute. Ribozymes. In *Encyclopedia of Virology*, number 4, pages 475–481. 2008.
- [130] David M.J. Lilley. Mechanisms of RNA catalysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1580):2910–2917, 2011.
- [131] Kathleen A. Leamy, Sarah M. Assmann, David H. Mathews, and Philip C. Bevilacqua. Bridging the gap between in vitro and in vivo RNA folding. *Quarterly Reviews of Biophysics*, 49:e10, 2016.
- [132] Jennifer A Doudna and Thomas R Cech. The chemical repertoire of natural ribozymes. *Nature*, 418(6894):222–228, 2002.
- [133] Andrew D. Ellington and Jack W. Szostak. In vitro selection of rna molecules that bind specific ligands. *Nature*, 346(6287):818–822, 8 1990.
- [134] Andrew D. Ellington and Jack W. Szostak. Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature*, 355(6363):850–852, 2 1992.
- [135] Katrina F. Tjhung, Maxim N. Shokhirev, David P. Horning, and Gerald F. Joyce. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proceedings of the National Academy of Sciences of the United States of America*, 117(6):2906–2913, 2020.
- [136] José I. Jiménez, Ramon Xulvi-Brunet, Gregory W. Campbell, Rebecca Turk-MacLeod, and Irene A. Chen. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 110(37):14984–14989, 2013.
- [137] Abe D. Pressman, Ziwei Liu, Evan Janzen, Celia Blanco, Ulrich F. Müller, Gerald F. Joyce, Robert Pascal, and Irene A. Chen. Mapping a Systematic Ribozyme Fitness Landscape Reveals a Frustrated Evolutionary Network for Self-Aminoacylating RNA. *Journal of the American Chemical Society*, 141(15):6213–6223, 2019.
- [138] D. Bartel and J. Szostak. Isolation of new ribozymes from a large pool of random sequences. *Science*, 261(5127):1411–1418, sep 1993.
- [139] Eric H Eklund and David P Bartel. Erratum: RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature*, 383(6596):192–192, sep 1996.
- [140] Aniela Wochner, James Attwater, Alan Coulson, and Philipp Holliger. Ribozyme-catalyzed transcription of an active ribozyme. *Science*, 332(6026):209–212, apr 2011.
- [141] Evan Janzen, Celia Blanco, Huan Peng, Josh Kenchel, and Irene A. Chen. Promiscuous ribozymes and their proposed role in prebiotic evolution. *Chemical Reviews*, 120(11):4879–4897, 6 2020.
- [142] James Attwater, Aditya Raguram, Alexey S. Morgunov, Edoardo Gianni, and Philipp Holliger. Ribozyme-catalysed RNA synthesis using triplet building blocks. *eLife*, 7:1–25, may 2018.
- [143] Jonathan T Sczepanski and Gerald F Joyce. A cross-chiral RNA polymerase ribozyme. *Nature*, 515(7527):440–442, nov 2014.
- [144] Margaret E. Glasner, Nicholas H. Bergman, and David P. Bartel. Metal ion requirements for structure and catalysis of an RNA ligase ribozyme. *Biochemistry*, 41(25):8103–8112, 2002.
- [145] Stephanie Kath-Schorr, Timothy J. Wilson, Nan Sheng Li, Jun Lu, Joseph A. Piccirilli, and David M.J. Lilley. General acid-base catalysis mediated by nucleobases in the hairpin ribozyme. *Journal of the American Chemical Society*, 134(40):16717–16724, 2012.
- [146] Feng Guo and Thomas R. Cech. Evolution of Tetrahymena ribozyme mutants with increased structural stability. *Nature Structural Biology*, 9(11):855–861, 2002.
- [147] Vanvimon Saksmerprome, Manami Roychowdhury-Saha, Sumedha Jayasena, Anastasia Khvorova, and Donald H Burke. Artificial tertiary motifs stabilize trans-cleaving hammerhead ribozymes under conditions of submillimolar divalent ions and high temperatures. *RNA*, 10(12):1916–1924, 2004.

- [148] Milena Popović, Palmer S. Fliss, and Mark A. Ditzler. In vitro evolution of distinct self-cleaving ribozymes in diverse environments. *Nucleic Acids Research*, 43(14):7070–7082, aug 2015.
- [149] Walter R. Farkas. Depolymerization of ribonucleic acid by plumbous ion. *BBA Section Nucleic Acids And Protein Synthesis*, 155(2):401–409, 1968.
- [150] T. Matreux, K. Le Vay, A. Schmid, P. Aikkila, L. Belohlavek, A. Z. Çalışkanoglu, E. Salibi, A. Kühnlein, C. Springsklee, B. Scheu, D. B. Dingwell, D. Braun, H. Mutschler, and C. B. Mast. Heat flows in rock cracks naturally optimize salt compositions for ribozymes. *Nature Chemistry*, 13(11):1038–1045, nov 2021.
- [151] Vinod K. Misra and David E. Draper. On the role of magnesium ions in RNA stability. *Biopolymers*, 48(2-3):113–135, 1998.
- [152] Yingfu Li and Ronald R. Breaker. Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2'-Hydroxyl Group. *Journal of the American Chemical Society*, 121(23):5364–5372, jun 1999.
- [153] Mikko Oivanen, Satu Kuusela, and Harri Lönnberg. Kinetics and mechanisms for the cleavage and isomerization of the phosphodiester bonds of RNA by bronsted acids and bases. *Chemical Reviews*, 98(3):961–990, 1998.
- [154] John A Zoltewicz, D Fred Clark, Thomas W Sharpless, and Gerwalt Grahe. Kinetics and mechanism of the acid-catalyzed hydrolysis of some purine nucleosides. *Journal of the American Chemical Society*, 92(6):1741–1750, mar 1970.
- [155] Ramanarayanan Krishnamurthy. On the Emergence of RNA. *Israel Journal of Chemistry*, 55(8):837–850, 2015.
- [156] Philip C. Bevilacqua. Mechanistic Considerations for General AcidBase Catalysis by RNA: Revisiting the Mechanism of the Hairpin Ribozyme. *Biochemistry*, 42(8):2259–2265, mar 2003.
- [157] James L. Hougland, Alexander V. Kravchuk, Daniel Herschlag, and Joseph A. Piccirilli. Functional Identification of Catalytic Metal Ion Binding Sites within RNA. *PLoS Biology*, 3(9):e277, aug 2005.
- [158] Daniel Herschlag and Mala Khosla. Comparison of pH Dependencies of the Tetrahymena Ribozyme Reactions with RNA 2'-Substituted and Phosphorothioate Substrates Reveals a Rate-Limiting Conformational Step. *Biochemistry*, 33(17):5291–5297, may 1994.
- [159] Henriette Kühne and Gerald F. Joyce. Continuous in vitro evolution of ribozymes that operate under conditions of extreme pH. *Journal of Molecular Evolution*, 57(3):292–298, 2003.
- [160] Yoshitaka Miyamoto, Naozumi Teramoto, Yukio Imanishi, and Yoshihiro Ito. In vitro adaptation of a ligase ribozyme for activity under a low-pH condition. *Biotechnology and Bioengineering*, 75(5):590–596, 2001.
- [161] Vineetha K. Jayasena and Larry Gold. In vitro selection of self-cleaving RNAs with a low pH optimum. *Proceedings of the National Academy of Sciences of the United States of America*, 94(20):10612–10617, 1997.
- [162] Gerald F Joyce. The antiquity of RNA-based evolution. *Nature*, 418(6894):214–221, jul 2002.
- [163] Falk Wachowius and Philipp Holliger. Rna replication and the rna polymerase ribozyme. *Ribozymes*, 1:359–386, 2021.
- [164] Maxine J McCall, Philip Hendry, Alain A Mir, Jason Conaty, Glenn Brown, and Trevor J Lockett. Small, efficient hammerhead ribozymes. *Molecular biotechnology*, 14:5–17, 2000.
- [165] Christof B Mast, Severin Schink, Ulrich Gerland, and Dieter Braun. Escalation of polymerization in a thermal gradient. *Proceedings of the National Academy of Sciences*, 110(20):8030–8035, may 2013.
- [166] Annalena Salditt, Lorenz M.R. Keil, David P. Horning, Christof B. Mast, Gerald F. Joyce, and Dieter Braun. Thermal Habitat for RNA Amplification and Accumulation. *Physical Review Letters*, 125(4):048104, jul 2020.
- [167] Annalena Salditt, Leonie Karr, Elia Salibi, Kristian Le Vay, Dieter Braun, and Hannes Mutschler. Ribozyme-mediated RNA synthesis and replication in a model Hadean microenvironment. *Nature Communications*, 14(1):1495, mar 2023.
- [168] Patrick W. Kudella. *Sequence self-selection by the network dynamics of random ligating oligomer pools*. PhD thesis, LMU Munich, 2021.
- [169] Ingmar Schoen, Hubert Krammer, and Dieter Braun. Hybridization kinetics is different inside cells. *Proceedings of the National Academy of Sciences*, 106(51):21649–21654, 2009.
- [170] Masataka Kinjo and Rudolf Rigler. Ultrasensitive hybridization analysis using fluorescence correlation spectroscopy. *Nucleic Acids Research*, 23(10):1795–1799, 1995.

- [171] J. G. Wetmur and N. Davidson. Kinetics of renaturation of DNA. *Journal of molecular biology*, 31(3):349–370, 1968.
- [172] Ibrahim I. Cisse, Hajin Kim, and Taekjip Ha. A rule of seven in watson-crick base-pairing of mismatched sequences. *Nature Structural and Molecular Biology*, 19(6):623–627, 6 2012.
- [173] Brent Ewing, Ladeana Hillier, Michael C Wendl, and Phil Green. Base-calling of automated sequencer traces using phred. i. accuracy assessment. *Genome research*, 8(3):175–185, 1998.
- [174] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 8 2014.
- [175] Nenad Ban, Poul Nissen, Jeffrey Hansen, Peter B Moore, and Thomas A Steitz. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science*, 289(5481):905–920, aug 2000.
- [176] Orgel Leslie E. Prebiotic Chemistry and the Origin of the RNA World. *Critical Reviews in Biochemistry and Molecular Biology*, 39(2):99–123, jan 2004.
- [177] Abe Pressman, Celia Blanco, and Irene A. Chen. The RNA World as a Model System to Study the Origin of Life. *Current Biology*, 25(19):R953–R963, oct 2015.
- [178] Angelica Mariani, Claudia Bonfio, Christopher M. Johnson, and John D. Sutherland. pH-Driven RNA Strand Separation under Prebiotically Plausible Conditions. *Biochemistry*, 57(45):6382–6386, nov 2018.
- [179] Lorenz M.R. Keil, Friederike M. Möller, Michael Kieß, Patrick W. Kudella, and Christof B. Mast. Proton gradients and pH oscillations emerge from heat flow at the microscale. *Nature Communications*, 8(1):1897, 2017.
- [180] David Ross and David Deamer. Dry/Wet Cycling and the Thermodynamics and Kinetics of Prebiotic Polymer Synthesis. *Life*, 6(3):28, jul 2016.
- [181] Jay G. Forsythe, Sheng-Sheng Yu, Irena Majanov, Martha A. Grover, Ramanarayanan Krishnamurthy, Facundo M. Fernández, and Nicholas V. Hud. Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. *Angewandte Chemie International Edition*, 54(34):9871–9875, aug 2015.
- [182] Matthew Levy and S. L. Miller. The stability of the RNA bases: Implications for the origin of life. *Proceedings of the National Academy of Sciences*, 95(14):7933–7938, jul 1998.
- [183] Mounir G. AbouHaidar and Ivan G. Ivanov. Non-Enzymatic RNA Hydrolysis Promoted by the Combined Catalytic Activity of Buffers and Magnesium Ions. *Zeitschrift für Naturforschung C*, 54(7-8):542–548, aug 1999.
- [184] Sueann C Dahm and Olke C Uhlenbeck. Role of divalent metal ions in the hammerhead RNA cleavage reaction. *Biochemistry*, 30(39):9464–9469, oct 1991.
- [185] Sueann C Dahm, Wesley B Derrick, and Olke C Uhlenbeck. Evidence for the role of solvated metal hydroxide in the hammerhead cleavage mechanism. *Biochemistry*, 32(48):13040–13045, dec 1993.
- [186] V Dange, R. Van Atta, and S. Hecht. A Mn²⁺-dependent ribozyme. *Science*, 248(4955):585–588, may 1990.
- [187] Ravi Desai, Duncan Kilburn, Hui-Ting Lee, and Sarah A. Woodson. Increased Ribozyme Activity in Crowded Solutions. *Journal of Biological Chemistry*, 289(5):2972–2977, jan 2014.
- [188] James M Collins and Kenneth S Rogers. Melting point depression of DNA by tetraalkylammonium bromides. *Chemico-Biological Interactions*, 19(2):197–203, 1977.
- [189] Teresa Golaś, Maria Miller, and David Shugar. The effects of tetraalkylammonium salts on helix-coil transition parameters in natural and synthetic ribo- and deoxyribo-polynucleotides. *Chemico-Biological Interactions*, 30(2):209–222, 1980.
- [190] P. Baaske, F. M. Weinert, Stefan Duhr, Kono H. Lemke, Michael J. Russell, and Dieter Braun. Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proceedings of the National Academy of Sciences*, 104(22):9346–9351, may 2007.
- [191] Christof B. Mast and Dieter Braun. Thermal trap for DNA replication. *Physical Review Letters*, 104(18):188102, may 2010.
- [192] Maren Reichl, Mario Herzog, Ferdinand Greiss, Manuel Wolff, and Dieter Braun. Understanding the similarity in thermophoresis between single- and double-stranded DNA or RNA. *Physical Review E*, 91(6):062709, jun 2015.

- [193] Poul Nissen, J. Hansen, N. Ban, P. B. Moore, and T. A. Steitz. The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289(5481):920–930, aug 2000.
- [194] Martin Hennig and Dieter Braun. Convective polymerase chain reaction around micro immersion heater. *Applied Physics Letters*, 87(18):183901, oct 2005.
- [195] Yusuke T. Maeda, Tsvi Tlusty, and Albert Libchaber. Effects of long DNA folding and small RNA stem–loop in thermophoresis. *Proceedings of the National Academy of Sciences*, 109(44):17972–17977, oct 2012.
- [196] Katarzyna Adamala and Jack W. Szostak. Nonenzymatic Template-Directed RNA Synthesis Inside Model Protocells. *Science*, 342(6162):1098–1100, nov 2013.
- [197] Stéphane Swillens, Barbara Dessars, and Hakim El Housni. Revisiting the sigmoidal curve fitting applied to quantitative real-time PCR data. *Analytical Biochemistry*, 373(2):370–376, feb 2008.
- [198] Gerald F. Joyce and Jack W. Szostak. Protocells and RNA Self-Replication. *Cold Spring Harbor Perspectives in Biology*, 10(9):a034801, sep 2018.
- [199] Jennifer A. Doudna, Nassim Usman, and Jack W. Szostak. Ribozyme-Catalyzed Primer Extension by Trinucleotides: A Model for the RNA-Catalyzed Replication of RNA. *Biochemistry*, 32(8):2111–2115, 1993.
- [200] Razvan Cojocar and Peter J. Unrau. Processive RNA polymerization and promoter recognition in an RNA World. *Science*, 371(6535):1225–1232, mar 2021.
- [201] Fabio Chizzolini, Luiz F.M. Passalacqua, Mona Oumais, Armine I. Dingilian, Jack W. Szostak, and Andrej Lupták. Large Phenotypic Enhancement of Structured Random RNA Pools. *Journal of the American Chemical Society*, 142(4):1941–1951, 2020.
- [202] Jesse V. Gavette, Matthias Stoop, Nicholas V. Hud, and Ramanarayanan Krishnamurthy. RNA-DNA Chimeras in the Context of an RNA World Transition to an RNA/DNA World. *Angewandte Chemie International Edition*, 55(42):13204–13209, oct 2016.
- [203] Seohyun Chris Kim, Derek K. O’Flaherty, Constantin Giurgiu, Lijun Zhou, and Jack W. Szostak. The Emergence of RNA from the Heterogeneous Products of Prebiotic Nucleotide Synthesis. *Journal of the American Chemical Society*, 143(9):3267–3279, mar 2021.
- [204] James Attwater, Teresa Augustin, Joseph F Curran, Samantha Kwok, Edoardo Gianni, and Philipp Holliger. Trinucleotide building blocks enable exponential ribozyme-catalysed rna replication and open-ended growth of diverse rna sequence pools. *bioRxiv*, pages 2023–03, 2023.
- [205] Christine He, Adriana Lozoya-Colinas, Isaac Gállego, Martha A. Grover, and Nicholas V. Hud. Solvent viscosity facilitates replication and ribozyme catalysis from an RNA duplex in a model prebiotic process. *Nucleic Acids Research*, 47(13):6569–6577, jul 2019.
- [206] Andrew S. Tupper and Paul G. Higgs. Rolling-circle and strand-displacement mechanisms for non-enzymatic RNA replication at the time of the origin of life. *Journal of Theoretical Biology*, 527:110822, oct 2021.
- [207] Emil Laust Kristoffersen, Matthew Burman, Agnes Noy, and Philipp Holliger. Rolling circle RNA synthesis catalysed by RNA. *eLife*, 11:e75186, feb 2022.
- [208] Zhaoming Su, Kaiming Zhang, Kalli Kappel, Shanshan Li, Michael Z. Palo, Grigore D. Pintilie, Ramya Rangan, Bingnan Luo, Yuquan Wei, Rhiju Das, and Wah Chiu. Cryo-EM structures of full-length Tetrahymena ribozyme at 3.1 Å resolution. *Nature*, 596(7873):603–607, 8 2021.
- [209] Kelly Kruger, Paula J. Grabowski, Arthur J. Zaugg, Julie Sands, Daniel E. Gottschling, and Thomas R. Cech. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, 31(1):147–157, nov 1982.
- [210] Cecilia Guerrier-Takada, Katheleen Gardiner, Terry Marsh, Norman Pace, and Sidney Altman. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35(3):849–857, dec 1983.
- [211] J A Doudna and J W Szostak. Miniribozymes, small derivatives of the sunY intron, are catalytically active. *Molecular and Cellular Biology*, 9(12):5480–5483, 1989.
- [212] Rachel Green and Jack W. Szostak. Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science*, 258(5090):1910–1915, 1992.

- [213] Gerry A. Prody, John T. Bakos, Jamal M. Buzayan, Irving R. Schneider, and George Bruening. Autolytic Processing of Dimeric Plant Virus Satellite RNA. *Science*, 231(4745):1577–1580, mar 1986.
- [214] S. Spiegelman, I. Haruna, I. B. Holland, G. Beaudreau, and D. Mills. The synthesis of a self-propagating and infectious nucleic acid with a purified enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, 54(3):919–927, 1965.
- [215] Thomas Matreux, Bernhard Altaner, Johannes Raith, Dieter Braun, Christof B. Mast, and Ulrich Gerland. Formation mechanism of thermally controlled pH gradients. *Communications Physics*, 6(1):14, 2023.
- [216] Warren A. Kibbe. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, 35(Web Server):W43–W46, may 2007.
- [217] Joseph N. Zadeh, Conrad D. Steenberg, Justin S. Bois, Brian R. Wolfe, Marshall B. Pierce, Asif R. Khan, Robert M. Dirks, and Niles A. Pierce. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, 32(1):170–173, jan 2011.
- [218] Martin J Serra and Douglas H Turner. [11] *Predicting thermodynamic properties of RNA*, volume 259, pages 242–261. Elsevier, 1995.
- [219] Adriana Calaca Serrao, Felix T Daenekamp, Zsofia Meggyesi, and Dieter Braun. Replication elongates short dna, reduces sequence bias, and develops trimer structure. *bioRxiv*, pages 2023–04, 2023.
- [220] Leonie Karr. *Increasing efficiency of RNA based ligation reactions driven by microscale water cycles under CO2 atmosphere*, Master thesis, 7 2021.
- [221] Patrick Haußmann. *Templated RNA ligation catalyzed by sunY ligase ribozyme from random sequence pools*, Bachelor thesis, 9 2022.

List of Figures

1.1	Becoming alive - Extrapolating back from life	1
1.2	Chemical and physical non-equilibrium settings at small scales	3
2.1	Thermal non-equilibria	7
3.1	Evolution of replication	14
3.2	Experimental conditions for the ligation reaction starting from random pools	15
3.3	Influence of length and number of bases on the sequence space	16
3.4	Estimation of hybridization timescales for differently sized sequence spaces .	17
3.5	Elongation of 12mer ATGCr pools	18
3.6	Concentration and reaction time screen for ATGCr pool	19
3.7	T_{lig} and concentration screen for the reduced complexity systems	21
3.8	Length distributions for T_{lig} and concentration screen	22
3.9	Cycle number and t_{lig} screen for RCS-x8	23
3.10	RCS-x64 spiked with sequence subsets	24
3.11	Length distribution of the systems	26
3.12	A- vs T-type and base probabilities for spiked and unspiked RCS-x64 pool . . .	28
3.13	PCC matrix for the x64-RCS pool and system 6-5'	30
3.14	PCC matrix for the x64-RCS pool with system 2	31
3.15	Potential elongation modes of random pools spiked with long sequences . . .	32
3.16	Long spiking sequences influence replication behavior	33
3.17	Melting temperatures for different lengths and sequence spaces	38
3.18	Ligase performance as given by the manufacturer	38
3.19	T_{lig} and concentration screen for the reduced complexity systems	41
3.20	t_{lig} screen for the reduced complexity systems	42
3.21	A-type vs T-type for RCS-x64 pool	42
3.22	A-type vs T-type for the RCS-x64 pool spiked with system 6-5'	43
3.23	A-type vs T-type for the RCS-x64 pool spiked with system 2	43
4.1	Sequence and secondary structure of the 24-3 polymerase ribozyme	46
4.2	Temperature gradients across a rock pores drive RNA-catalyzed RNA replication	48
4.3	Convective temperature oscillations inside cylindrical chamber	50
4.4	Convective RNA-catalyzed replication of RNA	51
4.5	Thermophoretic accumulation of nucleic acids	52
4.6	Characteristics of nucleic acid conglomerates	54
4.7	Accumulation profiles from finite-element simulations with diffusiophoresis .	55
4.8	Concentration height average of nucleic acids inside the convective chamber	56
4.9	Cycling characteristics derived for nucleic acids from stochastic trajectories .	57
4.10	Temperature-dependent RNA cleavage rate	58
4.11	Effect of temperature-dependent RNA cleavage on the polymerase concentration	59
4.12	Experimental realization of the convection chamber with central heating . .	64

4.13	Local heating of the convection chamber	67
4.14	Finite-element simulation derived temperature cycle conditions	68
4.15	Accumulation profiles as derived by finite-element simulations	69
4.16	Temperature characteristics derived from stochastic cycling trajectories	69
5.1	AWI-systems enable RNA folding, ligation, and strand separation at low Mg^{2+}	73
5.2	Dynamics of dew droplets in AWI-system	74
5.3	<i>sunY</i> -mediated templated ligation	76
5.4	<i>sunY</i> and the reaction mechanism at the ligation site of the templated ligation	77
5.5	PAGE analysis of <i>sunY</i> -mediated templated ligation	78
5.6	Strand separating capacity of AWI-system	80
5.7	<i>sunY</i> -mediated ligation of <i>sunY</i> fragments	81
5.8	Complete replication cycles of the <i>sunY</i> fragments in presence of template	82
5.9	Complete replication cycles of the <i>sunY</i> C123 and A123 fragments	83
5.10	Templated synthesis of the hammerhead ribozyme (HH-min)	85
5.11	Kinetics of HH-min synthesis with a 1:2 ratio of template to substrates	86
5.12	HH-min synthesis with a 1:4 (3 μ M:12 μ M) ratio of template to substrates	87
5.13	Quantification of HH-min product formation by PAGE	88
5.14	Templated synthesis and release of active hammerhead ribozyme	89
5.15	Synthesis of active HH-min in one-pot reaction in AWI-system	90
5.16	Yields for cleaved HH-sub in one-pot reaction	91
5.17	Hypothesis for an autocatalytic replication scenario of fragmented <i>sunY</i>	92
5.18	Ligation of HH-min by fragmented <i>sunY</i>	93
5.19	Ligation of individual fragments by fragmented <i>sunY</i>	94
5.20	Components for the assembly of the AWI-system setup	99
5.21	pH change due to CO_2 atmosphere	101
5.22	Cross-section of temperature simulation	102
5.23	MFS simulated with NUPACK for fragmented <i>sunY</i>	105
5.24	MFS simulated with NUPACK for the hammerhead system	106
5.25	Ligation of HH-min by <i>sunY</i>	108
6.1	Evolution of replication	109

List of Tables

3.1	Sequences used in experiments for <i>Taq</i> ligation of random pools	35
4.1	Sequences used in experiments for ribo PCR in thermal habitats	62
5.1	Cleavage yields of HH-sub in the AWI-system in one-pot scenario	91
5.2	Sequences used in experiments for <i>sunY</i> mediated ligation in AWI-systems . .	95
5.3	Thermal Conductivity of Materials	102
5.4	Predicted Melting Temperatures for the ABC System	103
5.5	Predicted Melting Temperatures for the HH-min System	104

8 List of Publications

List of publications published during the course of this PhD and associated (shared-) first author publications:

Annalena Salditt, Leonie Karr, Elia Salibi, Kristian Le Vay, Dieter Braun and Hannes Mutschler, Ribozyme-mediated RNA synthesis and replication in a model Hadean microenvironment, *Nature Communications* (**2023**) doi.org/10.1038/s41467-023-37206-4

Alan Ianeselli*, **Annalena Salditt***, Christof Mast*, Barbara Ercolano, Corinna Kufner, Bettina Scheu and Dieter Braun, Physical non-equilibria for prebiotic nucleic acid chemistry, *Nature Review Physics* (**2023**) doi.org/10.1038/s42254-022-00550-3

Emre Tekin, **Annalena Salditt**, Philipp Schwintek, Sree Wunnava, Juliette Langlais, James Saenz, Dora Tang, Petra Schwillie, Christof Mast, Dieter Braun, Prebiotic foam environments to oligomerize and accumulate RNA, *ChemBioChem* (**2022**) doi.org/10.1002/cbic.202200423

Patrick W. Kudella, Alexei V. Tkachenko, **Annalena Salditt**, Sergei Maslov and Dieter Braun, Structured sequences emerge from random pool when replicated by templated ligation, *Proc Natl Acad Sci USA*, (**2021**) <https://doi.org/10.1073/pnas.2018830118>

Annalena Salditt*, Lorenz M. R. Keil*, David P. Horning*, Christof B. Mast, Gerald F. Joyce, and Dieter Braun, Thermal Habitat for RNA Amplification and Accumulation, *Phys. Rev. Lett.* (**2020**) <https://doi.org/10.1103/physrevlett.125.048104>

Preiner, M.; Asche, S.; Becker, S.; Betts, H.C.; Boniface, A.; Camprubi, E.; Chandru, K.; Erastova, V.; Garg, S.G.; Khawaja, N.; Kostyrka, G.; Machné, R.; Moggioli, G.; Muchowska, K.B.; Neukirchen, S.; Peter, B.; Pichlhöfer, E.; Radványi, Á.; Rossetto, D.; **Salditt, A.**; Schmelling, N.M.; Sousa, F.L.; Tria, F.D.K.; Vörös, D.; Xavier, J.C. The Future of Origin of Life Research: Bridging Decades-Old Divisions. *Life* (**2020**) <https://doi.org/10.3390/life10030020>

Evgeniia Edeleva, **Annalena Salditt**, Julian Stamp, Philipp Schwintek, Job Boekhoven and Dieter Braun, Continuous nonenzymatic cross-replication of DNA strands with in situ activated DNA oligonucleotides, *Chem. Sci.* (**2019**), <https://doi.org/10.1039/c9sc00770a>

Thermal Habitat for RNA Amplification and Accumulation

Annalena Salditt^{1,*}, Lorenz M. R. Keil^{1,*}, David P. Horning^{2,*}, Christof B. Mast¹,
Gerald F. Joyce², and Dieter Braun^{1,†}¹*Systems Biophysics, Physics Department, Center for Nanoscience, Ludwig-Maximilians-Universität München, Amalienstraße 54, 80799 Munich, Germany*²*The Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA* (Received 23 December 2019; accepted 2 July 2020; published 24 July 2020)

The RNA world scenario posits replication by RNA polymerases. On early Earth, a geophysical setting is required to separate hybridized strands after their replication and to localize them against diffusion. We present a pointed heat source that drives exponential, RNA-catalyzed amplification of short RNA with high efficiency in a confined chamber. While shorter strands were periodically melted by laminar convection, the temperature gradient caused aggregated polymerase molecules to accumulate, protecting them from degradation in hot regions of the chamber. These findings demonstrate a size-selective pathway for autonomous RNA-based replication in natural nonequilibrium conditions.

DOI: 10.1103/PhysRevLett.125.048104

Introduction.—In modern living systems, the translation of information from DNA to proteins is performed by an RNA intermediate, separating the requirements for the replication of genetic polymers and the production of functional enzymes. However, RNA itself is capable of both storing genetic information and folding into catalytically active structures, including those that enable copying of RNA molecules [1–3]. Thus, a solution for storage and effective transfer of information on early Earth prior to the invention of genetically encoded proteins could have been achieved with an RNA-only replication mechanism [4–8].

Any RNA copying mechanism, based on either templated ligation or templated polymerization, relies on Watson-Crick base pairing. An energy source is needed to separate the two complementary strands to begin the next round of templated synthesis, which would need to have been provided by a plausible geochemical mechanism on the early Earth. Separation of hybridized strands could be achieved by *pH* cycling [9,10], evaporation-wetting cycles [11–13], oscillation of salt concentrations [14], or elevated temperatures. In all cases, a setting is needed that minimizes the spontaneous cleavage of RNA [15,16] that occurs at high temperatures for the high-salt conditions required for RNA catalysis. In the temperature range from 20 to 90 °C, degradation increases over 4 orders of magnitude, dictating a minimal exposure time at high temperatures.

A habitat for RNA replication on the early Earth should not only be able to separate double-stranded RNA, but also provide a mechanism for its persistent accumulation against dilution by diffusion. Previous studies have shown that a localized heat flux across closed, elongated compartments can accumulate nucleic acids. This mechanism favors the retention of longer strands in a replication reaction mixture [17,18]. While the combination of replication and selection

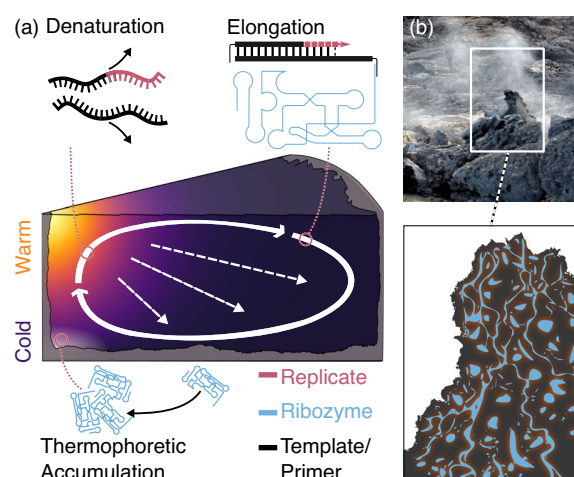


FIG. 1. Heat flux across water-filled pores drives RNA-catalyzed RNA replication. (a) A heat flow is used to create temperature differences across a water-filled pore. The temperature differences induce both thermophoresis of molecules (dashed arrows), moving them along thermal gradients, and convection of water (solid arrows). Simulations predict that the interplay of the two physical nonequilibrium effects locally concentrates the RNA polymerase toward cold regions of the cylindrical chamber, where it is protected against thermal degradation. Because of the strongly length-dependent thermophoretic properties of RNA [22], shorter RNA molecules are not accumulated, but are subjected to temperature oscillations to achieve the necessary strand separation in the warm spot after their template-directed replication in the cold areas. (b) A natural setting for such a heat flow could be the dissipation of heat across volcanic or hydrothermal rocks. This leads to temperature differences over porous structures of various shapes and lengths.

was shown for the protein-based replication of DNA using Taq DNA polymerase [19], also in combination with accumulation [20,21], it was unclear whether the thermal instability of RNA would prevent a similar approach.

Here, we describe a modified thermal microenvironment that is able to drive the replication of small RNAs catalyzed by a larger polymerase ribozyme. A punctual heat source inside a cylindrical compartment gives rise to two effects: (i) laminar gravitational convection due to the temperature-dependent density of water, resulting in temperature cycles that meet the delicate requirements for the elongation and strand separation for RNA-catalyzed RNA replication; and (ii) thermophoretic movement of molecules along a temperature gradient, pointing outward from the high temperature spot [Fig. 1(a)]. This thermophoretic movement has been shown to depend on a combination of nonionic interactions, ionic shielding, and Seebeck effects [22]. In the case of polyanionic nucleic acids at elevated temperatures, thermophoresis drives the molecules from warmer to cooler areas [23]. For a cylindrical compartment, the interplay of convective and thermophoretic transport resulted in a length-dependent net transport of molecules away from the warm temperature spot. The efficiency of this transport increased for longer RNAs, stabilizing them against cleavage that would occur at higher temperatures.

In contrast, the replicated shorter RNA oligomers cycled quickly through the hotter areas of the habitat, where they could undergo thermally induced denaturation. This cycling ensured the melting of double-stranded molecules, providing templates for new polymerization reactions. As a result, the replication and preservation of genetic information could be accomplished within a single, thermally driven environment. On the early Earth, similar thermal hatcheries could have driven RNA-based replication in natural conditions, provided confinement and temperature gradients are present, which is a common setting in volcanic or hydrothermal environments [Fig. 1(b)].

Results.—Convective temperature oscillations: The heat flow was implemented via a focused infrared laser that was absorbed in a water-filled cylindrical chamber. The resulting radial symmetric temperature profile consisted of a hot temperature spot ($\sim 80^\circ\text{C}$), decreasing across the compartment to 17°C at the bottom side, as measured by fluorescent thermometry (see Supplemental Material, Materials and Methods, and Supplemental Material Fig. 1 [24]). The emerging temperature gradient not only accounted for the accumulation, but also for the repetitive temperature cycling of the oligonucleotides.

A similar geometry driving protein-catalyzed replication has previously been reported based on localized heating of the chamber surface [25]. Both implementations showed comparable temperature fields in numerical simulations (Supplemental Material Fig. 2 [24]). However, protein-catalyzed replication mimics only later stages of evolution, after the emergence of genetically encoded proteins. Here,

the amplification of sequence information relied solely on the enzymatic activity of an RNA molecule.

The convection chamber drove RNA-catalyzed amplification of a 35-nucleotide RNA template, employing 25-nucleotide short RNA primers. The slow polymerization rate of the ribozyme required the convection to run for ~ 24 h. Under these conditions, the convection chamber showed exponential amplification of the 35-nucleotide RNA, with starting template concentrations as low as 100 fM [Figs. 2(a) and 2(b)]. Thermal cycling of a bulk reaction mixture under optimized conditions achieved similar results with 50 temperature cycles and a cycle time of 20 min. Starting with 100 fM RNA template, RNA-catalyzed amplification resulted in product yields of 2×10^5 -fold and 4×10^5 -fold for the bulk and convectively driven reactions, respectively [Fig. 2(b)]. This amplification could be described theoretically by a two-parameter growth equation [27] (see Supplemental Material [24]), deriving a similar maximum replication efficiency E , for convection with $E = 1.27$ and for thermal cycling with $E = 1.28$. In both cases, the buffer conditions were optimized to increase replication yield [3], i.e., by adding PEG8000 to serve as a molecular crowding agent, reducing the concentration of Mg^{2+} to 50 mM to reduce spontaneous cleavage of RNA, and adding tetrapropylammonium chloride to lower the melting temperature of the RNA.

To gain access to the temperature cycling conditions within the chamber, a fluorescence measurement of the temperature profile was used as the basis for a numerical model. The temperature cycling of the molecules inside the chamber was based on a combination of laminar convection, Brownian motion, and thermophoretic drift (Supplemental Material Fig. 3). With a chamber thickness of $500 \mu\text{m}$, we obtained a mean cycle time of 26 min for the 35-nucleotide RNA to oscillate between the threshold temperatures of 20 and 60°C [Figs. 2(c) and 2(d)]. This matched the 20 min cycling protocol in the thermal cyclers (68°C for 2 s, then 17°C for 20 min) used in the bulk amplification protocols in the homogeneously mixed experiments [3]. In this way, the molecules evade the fast degradation at high temperatures [26] shown in Fig. 2(e).

Thermophoretic accumulation of RNA polymerase: We characterized the accumulation behavior of the different RNA components of the system by monitoring fluorescently labeled single- and double-stranded DNA substitutes containing 35 and 210 nucleotides, in addition to the RNA polymerase itself and its DNA analog. The thermophoretic properties of diluted DNA and RNA, either single- or double-stranded, have been shown to be very similar [23]. However, the thermophoretic drift was found to be strongly length dependent [28].

As expected, we found that the shorter DNA barely accumulated in the convection system. Fully double-stranded 210mer DNA showed a central fivefold

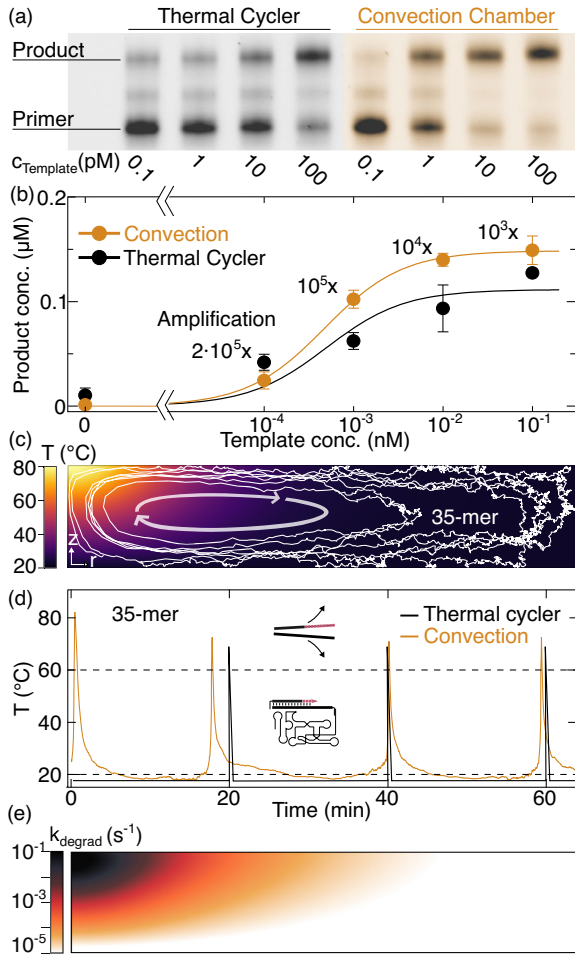


FIG. 2. Convective RNA-catalyzed replication of RNA. (a) The convection system and a thermal cycler showed similar yields of primer extension on the 35-nucleotide RNA template. (b) The polymerase ribozyme exponentially amplified starting template concentrations as low as 100 fM in 24 h. The amplification can be described theoretically by a two-parameter growth equation (Supplemental Material [24]). A maximum amplification of 2×10^5 -fold was observed in the convection chamber. Error bars indicate the deviation from duplicate experiments. (c) Thin lines show simulated stochastic trajectories for a 35mer inside the convection chamber of 500 μm height and 2.25 mm radius. (d) In the convection chamber, RNA mostly resided at low-temperature regions, where polymerization could occur, and passed quickly through high-temperature regions that enabled strand separation. Stochastic simulations found a mean temperature cycling time of 26 min. Thermal cycler experiments were performed with cycles of 17 $^{\circ}\text{C}$ for 20 min, then 68 $^{\circ}\text{C}$ for 2 s. (e) Degradation of RNA is almost 5 orders of magnitude faster at higher temperatures [26].

accumulation after 60 min at the bottom of the chamber [Fig. 3(a)]. These findings are in agreement with finite-element simulations that took into account convection, diffusion, and thermophoresis [Fig. 3(c)].

Interestingly, the active RNA polymerase showed a different, ring-shaped accumulation pattern in colder

regions of the chamber. To understand this effect, we performed accumulation experiments for the RNA polymerase, as well as for a DNA analog of the polymerase ribozyme. The RNA polymerase or its DNA analog forms a ternary complex with RNA primer and template via complementary sequences at the 5' ends of both polymerase and template [3], enabling the molecules to be stained with a fluorescently labeled primer. Imaging the solution with higher resolution in a 40 μm thin capillary revealed that a majority of the RNA polymerase and its DNA analog were present in form of conglomerates at $T = 17^{\circ}\text{C}$ [Fig. 3(b)].

By individually removing buffer components, the 6% w/v PEG8000 was found to be the crucial component that induced aggregation (Supplemental Material Fig. 4 [24]), both for the RNA and DNA version of the polymerase sequence. Additionally, the conglomerates exhibited a temperature dependence, where heating the solution led to melting of the conglomerates (see Supplemental Material and movie [24]). For the 35mer single-stranded DNA and 210mer double-stranded DNA, no aggregates were found, as imaging the solution showed homogeneous fluorescence [Fig. 3(b)].

For the conglomerates, we could predict the ring-shaped accumulation region at the top of the chamber with finite element simulations after we included the thermophoretic accumulation of PEG and its diffusiophoretic effect on DNA/RNA as reported by Madea *et al.* [29] [Supplemental Material Figure 5(c)]. For the conglomerates, diffusiophoresis dominated the movement in the temperature gradient, now pointing toward the heating source. However, this inverted force only has an effect near the boundary walls where flow velocity does not dominate over the slower diffusion of the conglomerates. As a result, the conglomerated RNA or DNA accumulated into a ring, away from the hot temperature spot at the top chamber wall (Supplemental Material Fig. 5 [24]).

The diffusiophoretic interaction between PEG and RNA made the accumulation dependent on the binding details of the molecules. Unsaturated, single-stranded molecules could engage in intermolecular interactions and therefore form conglomerates. This behavior is supported by the stark difference between double-stranded DNA and the single-stranded DNA analog of the polymerase sequence.

Simulating the trajectories of 400 particles with random starting positions gave access to the temperature distribution and cycling times for polymerases, double-stranded DNA, and single-stranded DNA, respectively (Supplemental Material Fig. 6 [24]). These simulations showed that the ring-shaped accumulation maintained the conglomerates at a temperature of 45 $^{\circ}\text{C}$ and efficiently restricted them to temperatures below 60 $^{\circ}\text{C}$. Although molecules not forming conglomerates have a higher residence probability at the lowest temperatures, they are frequently subjected to temperatures above 60 $^{\circ}\text{C}$. Based on the experiments of Li and Breaker [26] the RNA cleavage rate can be predicted for

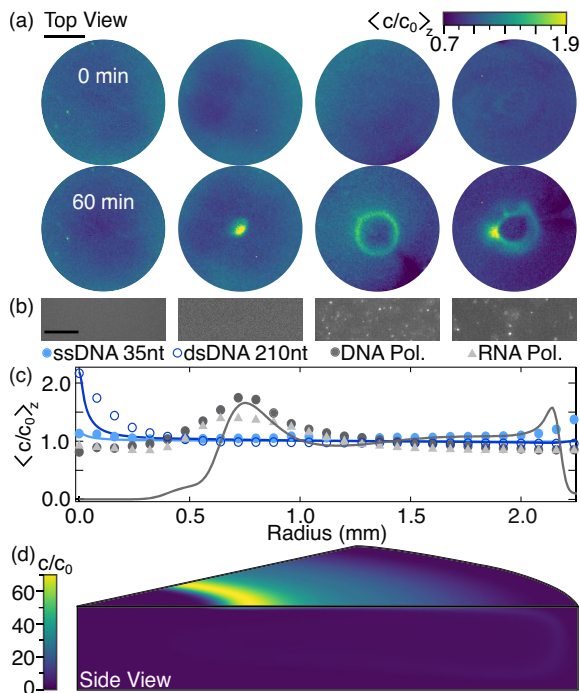


FIG. 3. Thermophoretic accumulation of nucleic acids. (a) Accumulation of nucleic acids inside the convection chamber with a central heat source implemented via an IR laser is shown for 35mer single-stranded DNA, 210mer double-stranded DNA, and folded single-stranded DNA and RNA corresponding to the polymerase sequences (from left to right). The scale bar corresponds to 1 mm. (b) Imaging the different strands inside the reaction buffer revealed the formation of conglomerates in the case of the polymerase sequences. The scale bar corresponds to 250 μm . (c) The height average $\langle c/c_0 \rangle_z$ of the simulated relative concentrations (lines) reproduced the experimental fluorescence signal (symbols) for all species. The data points represent the radial averaging of the background corrected fluorescence images with respect to the central point of the chamber. (d) By including diffusive, convective, thermophoretic, and diffusiophoretic transport in the finite element simulation, the model could capture the ring-shaped concentration enhancement of the polymerases observed in the experiments, with a maximum of 66-fold concentration increase after 60 min. The simulation was executed in a compartment of 500 μm and 2.25 mm height and radius, respectively.

varying temperature, pH and ions (see Supplemental Material methods [24]), ranging from $3.81 \times 10^{-6} \text{ s}^{-1}$ for 17 $^{\circ}\text{C}$ and 0.22 s^{-1} for 85 $^{\circ}\text{C}$. To investigate whether the accumulation indeed protects the ribozyme, we included the cleavage rate in our simulation. The total ribozyme concentration decreased exponentially (Supplemental Material Fig. 7) with a fivefold reduced degradation rate if the phoretic forces were activated in the simulation and therefore accumulated the ribozyme conglomerates in the ring pattern. These simulation capabilities will allow us to engineer optimal rock geometries, a process that would have been performed by natural selection on early Earth.

Discussion.—What habitat could provide conditions for an RNA world implementation of the emergence of life? Considerable efforts have been made to investigate the synthesis of RNA by an RNA polymerase ribozyme [3,30–34], but only very few reactions have been operated in a prebiotically plausible setting.

The spontaneous cleavage of RNA at the elevated temperatures necessary to separate template and product strands limits the formation and preservation of longer nucleic acids. On the early Earth, however, longer RNAs would have been required to provide robust enzymatic activities [35,36]. The thermal habitat described here provides temperature conditions that can drive RNA-catalyzed RNA replication. It achieved comparable replication kinetics to the optimized bulk reaction within a standard thermal cyclers. Additionally, the interplay of phoretic forces and convection concentrated the RNA polymerase away from the central heating spot. This length-dependent accumulation mechanism biased toward longer and more structured RNAs also could help to overcome the threat of short parasitic sequences that are generally copied more quickly. In the thermal habitat, shorter RNAs have a higher probability of exposure to elevated temperatures where degradation is enhanced [Fig. 3(a); Supplemental Material Fig. 5] [16]. Moreover, the reaction will be able to be fed with nucleotides by adding a flow through the chamber. The conditions present in the compartment can be tuned to match multiple reaction conditions, which allows us to adapt to future versions of early replication scenarios as well as other reactions.

One such example is the RNA-catalyzed polymerization of RNA, carried out in eutectic ice, which both concentrates the reactants and reduces spontaneous RNA cleavage [34]. Moving forward, including ice phases in future versions of the shown thermal habitat could help to reduce the bulk salt concentrations, while still achieving thermal strand separation and long term localization by thermal convection and thermophoresis.

The experiments indicate the existence of selective pathways in thermal habitats, which could guide RNA evolution toward longer and more structured sequences. This setting could provide a way to replicate not only 35 mers, but a complete RNA polymerase that is assembled from several shorter component strands [33,34,37]. How such >200 -nucleotide RNAs could have emerged starting from simple, nonenzymatic replication chemistries in a similar setting remains an open question.

Conclusion.—The search for the origin of life implies finding a location for informational molecules to replicate and undergo Darwinian evolution against entropic obstacles such as dilution and spontaneous degradation. The experiments described here demonstrate how a heat flow across a millimeter-sized, water-filled porous rock can lead to spatial separation of molecular species resulting in different reaction conditions for different species. The

conditions inside such a compartment can be tuned according to the requirements of the partaking molecules due to the scalable nature of this setting. A similar setting could have driven both the accumulation and RNA-based replication in the emergence of life, relying only on thermal energy, a plausible geological energy source on the early Earth. Current forms of RNA polymerase ribozymes can only replicate very short RNA strands. However, the observed thermal selection bias toward long RNA strands in this system could guide molecular evolution toward longer strands and higher complexity.

This work was supported by the Simons Collaboration on the Origin of Life (Grant No. 327125) and the German Research Foundation (DFG) starting with the SFB 1032 Nanoagents Project A4 and finishing with the TRR 235 Emergence of Life (Project-ID 364653263). This work was further supported by the NanoSystems Initiative Munich. We thank Christina Dirscherl, Patrick Kudella, and Alexandra Kühnlein for helpful comments on the manuscript. L. K., D. P. H., A. S., and C. B. M. performed the experiments. L. K., D. P. H., A. S., C. B. M., G. F. J., and D. B. conceived and designed the experiments, L. K., D. P. H., A. S., C. B. M., G. F. J., and D. B. analyzed the data and wrote the Letter.

The authors declare no competing financial interests.

*These authors contributed equally to the work.

†Corresponding author.
dieter.braun@lmu.de

- [1] G. F. Joyce, Toward an alternative biology, *Science* **336**, 307 (2012).
- [2] J. Attwater, A. Wochner, V. B. Pinheiro, A. Coulson, and P. Holliger, Ice as a protocellular medium for RNA replication, *Nat. Commun.* **1**, 76 (2010).
- [3] D. P. Horning and G. F. Joyce, Amplification of RNA by an RNA polymerase ribozyme, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9786 (2016).
- [4] W. Gilbert, The RNA world, *Nature (London)* **319**, 618 (1986).
- [5] L.E. Orgel, Prebiotic chemistry and the origin of the RNA World, *Crit. Rev. Biochem. Mol. Biol.* **39**, 99 (2004).
- [6] M. P. Robertson and G. F. Joyce, The origins of the RNA World, *Cold Spring Harbor Perspect. Biol.* **4**, a003608 (2012).
- [7] A. Pressman, C. Blanco, and I. A. Chen, The RNA World as a model system to study the origin of life, *Curr. Biol.* **25**, R953 (2015).
- [8] N. Ban, P. Nissen, J. Hansen, P. B. Moore and T. A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution, *Science* **289**, 905 (2000).
- [9] A. Mariani, C. Bonfio, C. M. Johnson, and J. D. Sutherland, pH-driven RNA strand separation under prebiotically plausible conditions, *Biochemistry* **57**, 6382 (2018).
- [10] L. M. R. Keil, F. M. Möller, M. Kieß, P. W. Kudella, and C. B. Mast, Proton gradients and pH oscillations emerge from heat flow at the microscale, *Nat. Commun.* **8**, 1897 (2017).
- [11] D. Ross and D. Deamer, Dry/wet cycling and the thermodynamics and kinetics of prebiotic polymer synthesis, *Life* **6**, 28 (2016).
- [12] J. G. Forsythe, S.-S. Yu, I. Mamajanov, M. A. Grover, R. Krishnamurthy, F. M. Fernández, and N. V. Hud, Ester-mediated amide bond formation driven by wet-dry cycles: A possible path to polypeptides on the prebiotic Earth, *Angew. Chem., Int. Ed. Engl.* **54**, 9871 (2015).
- [13] M. Morasch *et al.*, Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules, *Nat. Chem.* **11**, 779 (2019).
- [14] A. Ianeselli, C. B. Mast, and D. Braun, Periodic melting of oligonucleotides by oscillating salt concentrations triggered by microscale water cycles inside heated rock pores, *Angew. Chem.* **131**, 13289 (2019).
- [15] M. Levy and S. L. Miller, The stability of the RNA bases: Implications for the origin of life, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7933 (1998).
- [16] M. G. AbouHaidar and I. G. Ivanov, Non-enzymatic RNA hydrolysis promoted by the combined catalytic activity of buffers and magnesium ions, *Z. Naturforsch. C* **54**, 542 (1999).
- [17] P. Baaske, F. M. Weinert, S. Duhr, K. H. Lemke, M. J. Russell, and D. Braun, Extreme accumulation of nucleotides in simulated hydrothermal pore systems, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9346 (2007).
- [18] C. B. Mast, S. Schink, U. Gerland, and D. Braun, Escalation of polymerization in a thermal gradient, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8030 (2013).
- [19] D. Braun, N. L. Goddard, and A. Libchaber, Exponential DNA Replication by Laminar Convection, *Phys. Rev. Lett.* **91**, 158103 (2003).
- [20] M. Kreysing, L. Keil, S. Lanzmich, and D. Braun, Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length, *Nat. Chem.* **7**, 203 (2015).
- [21] C. B. Mast and D. Braun, Thermal Trap for DNA Replication, *Phys. Rev. Lett.* **104**, 188102 (2010).
- [22] M. Reichl, M. Herzog, A. Götz, and D. Braun, Why Charged Molecules Move Across a Temperature Gradient: The Role of Electric Fields, *Phys. Rev. Lett.* **112**, 198101 (2014).
- [23] M. Reichl, M. Herzog, F. Greiss, M. Wolff, and D. Braun, Understanding the similarity in thermophoresis between single- and double-stranded DNA or RNA, *Phys. Rev. E* **91**, 062709 (2015).
- [24] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.125.048104> for experimental details and further information.
- [25] M. Hennig and D. Braun, Convective polymerase chain reaction around micro immersion heater, *Appl. Phys. Lett.* **87**, 183901 (2005).
- [26] Y. Li and R. R. Breaker, Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2'-hydroxyl group, *J. Am. Chem. Soc.* **121**, 5364 (1999).

- [27] S. Swillens, B. Dessars, and H. El. Housni, Revisiting the sigmoidal curve fitting applied to quantitative real-time PCR data, *Anal. Biochem.* **373**, 370 (2008).
- [28] S. Duhr and D. Braun, Why molecules move along a temperature gradient, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19678 (2006).
- [29] Y. T. Maeda, T. Tlustý, and A. Libchaber, Effects of long DNA folding and small RNA stem-loop in thermophoresis, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17972 (2012).
- [30] P. Nissen, J. Hansen, N. Ban, P. B. Moore, and T. A. Steitz, The structural basis of ribosome activity in peptide bond synthesis, *Science* **289**, 920 (2000).
- [31] E. H. Eklund and D. P. Bartel, RNA-catalysed RNA polymerization using nucleoside triphosphates, *Nature (London)* **383**, 192 (1996).
- [32] W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, and D. P. Bartel, RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension, *Science* **292**, 1319 (2001).
- [33] J. T. Szcepanski and G. F. Joyce, A cross-chiral RNA polymerase ribozyme, *Nature (London)* **515**, 440 (2014).
- [34] H. Mutschler, A. Wochner, and P. Holliger, Freeze-thaw cycles as drivers of complex ribozyme assembly, *Nat. Chem.* **7**, 502 (2015).
- [35] J. A. Doudna, S. Couture, and J. W. Szostak, A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis, *Science* **251**, 1605 (1991).
- [36] K. Adamala, A. E. Engelhart, and J. W. Szostak, Generation of functional RNAs from inactive oligonucleotide complexes by non-enzymatic primer extension, *J. Am. Chem. Soc.* **137**, 483 (2015).
- [37] J. Attwater, A. Raguram, A. S. Morgunov, E. Gianni, and P. Holliger, Ribozyme-catalysed RNA synthesis using triplet building blocks, *eLife* **7**, e35255 (2018).

Physical non-equilibria for prebiotic nucleic acid chemistry

Alan Ianeselli^{1,2,6}, Annalena Salditt^{1,6}, Christof Mast^{1,6}, Barbara Ercolano³, Corinna L. Kufner⁴, Bettina Scheu⁵ & Dieter Braun¹✉

Abstract

The prebiotic replication of DNA and RNA is a complex interplay between chemistry and the environment. Factors that have direct and indirect effects on prebiotic chemistry include temperature, concentration of monovalent and bivalent ions, the pH of water, ultraviolet irradiation and the presence of gaseous CO₂. We discuss various primordial conditions to host the first replication reactions on the early Earth, including heated rock pores, hydrothermal vents, evaporating water ponds, freezing–thawing ice compartments, ultraviolet irradiation and high CO₂ concentrations. We review how the interplay of replication chemistry with the strand separation and length selectivity of non-equilibrium physics can be provided by plausible geo-environments. Fast molecular evolution has been observed over a few hours in such settings when a polymerase protein is used as replicator. Such experimental findings make us optimistic that it will soon also be possible to probe evolution dynamics with much slower prebiotic replication chemistries using RNA. Our expectation is that the unique autonomous evolution dynamics provided by microfluidic non-equilibria make the origin of life understandable and experimentally testable in the near future.

Sections

Introduction

Hadean Earth geological conditions

Non-equilibria to drive molecular evolution

Outlook

¹Systems Biophysics, Center for Nanoscience, Ludwig-Maximilians-Universität München, Munich, Germany.

²Faculty of Computer Science, Free University of Bozen-Bolzano, Bolzano, Italy. ³University Observatory Munich, Ludwig-Maximilians-Universität München, München, Germany. ⁴Harvard-Smithsonian Center for Astrophysics, Department of Astronomy, Harvard University, Cambridge, MA, USA. ⁵Earth and Environmental Sciences, Ludwig-Maximilians-Universität München, Munich, Germany. ⁶These authors contributed equally: Alan Ianeselli, Annalena Salditt, Christof Mast. ✉e-mail: dieter.braun@lmu.de

Introduction

To understand possible mechanisms for the origin of life, it is essential to study the environmental conditions that could generate and sustain the first stages of molecular evolution. These boundary conditions were defined by the geological conditions and processes of the very early Earth, probably in the Hadean eon after the appearance of liquid water about 4.4–4.0 billion years ago^{1–3}. Various geological settings with and without the presence of limited surficial landmass are considered plausible environments at the late Hadean and early Archaean. Some of the most studied settings are shown in Fig. 1. Each of these can host one or more non-equilibrium mechanisms that make it particularly appealing in the context of Hadean molecular evolution.

Such geological systems are sources of local non-equilibrium conditions, driven by the slowly cooling early Earth under a young faint Sun. These non-equilibria are essential because, according to entropy arguments, life could not have originated under equilibrium conditions^{4,5}. The origin of life needs a high level of self-organization, high concentration of reagents and a continuous influx of energy to be able to evolve towards more complex systems. Without a means of lowering the entropy, this process becomes impossible, and the system slowly reaches equilibrium and decays into a dead soup. For example, under equilibrium, RNA and DNA oligomers dilute by diffusion, reduce their length by hydrolysis and cannot replicate sequence information by polymerization through activated molecules as these cannot be replenished. The necessary non-equilibria – such as concentration gradients, salt and pH cycles, irradiation or temperature differences – actively drive the system towards a continuous and dynamic self-organization^{6–8} and could be provided by external forces such as geothermal heating, solar and isotopic irradiation, day–night cycles and atmospheric phenomena. The search for settings that are able to deliver continuous forces to maintain the chemical system out of equilibrium is therefore important.

Investigating the early Earth environment that may have provided the initial conditions for the emergence of life also offers inspiration for research into the question of extraterrestrial life. Although some astrophysical space missions, such as NASA's Transiting Exoplanet Survey Satellite (TESS⁹), focus on searching for Earth's analogues, there is increasing evidence that the sort of out-of-equilibrium environments reviewed in this Perspective may also be common on celestial bodies that are far from being Earth analogues. For instance, even if most rocky planets in the habitable zone around cooler (M) stars that were targeted by the Kepler mission¹⁰ are expected to be tidally locked (that is, always showing the same side to their parent star), they are still likely to have something resembling day–night cycles driven by weather systems, which could in turn drive wet–dry cycles in some regions on their surface. Extreme examples of possible habitable worlds that have conditions far from those of Earth are moons around giant planets, like those in our Solar System (Io, Europa, Titan), in extrasolar systems or, even more exotically, moons around free-floating planets¹¹. The huge investment in space missions aiming at atmospheric studies of extraterrestrial and extrasolar bodies, due to be launched within the next two decades, will strongly benefit from origins of life research that studies this complex process on Earth.

In this Perspective, we survey commonly studied geological settings in the Hadean Earth, before discussing how the non-equilibrium conditions hosted by these settings could drive molecular evolution.

Hadean Earth geological conditions

Wet–dry cycles in evaporating water ponds

In the past, the most often cited setting for the origin of life has been Darwin's "warm little pond" of fresh water on land^{12,13}, where reagents in the organic soup can become concentrated enough for chemical reactions to occur (Fig. 1a). Day–night, seasonal temperature, or weather oscillations can be imagined to lead to fluctuating water levels and

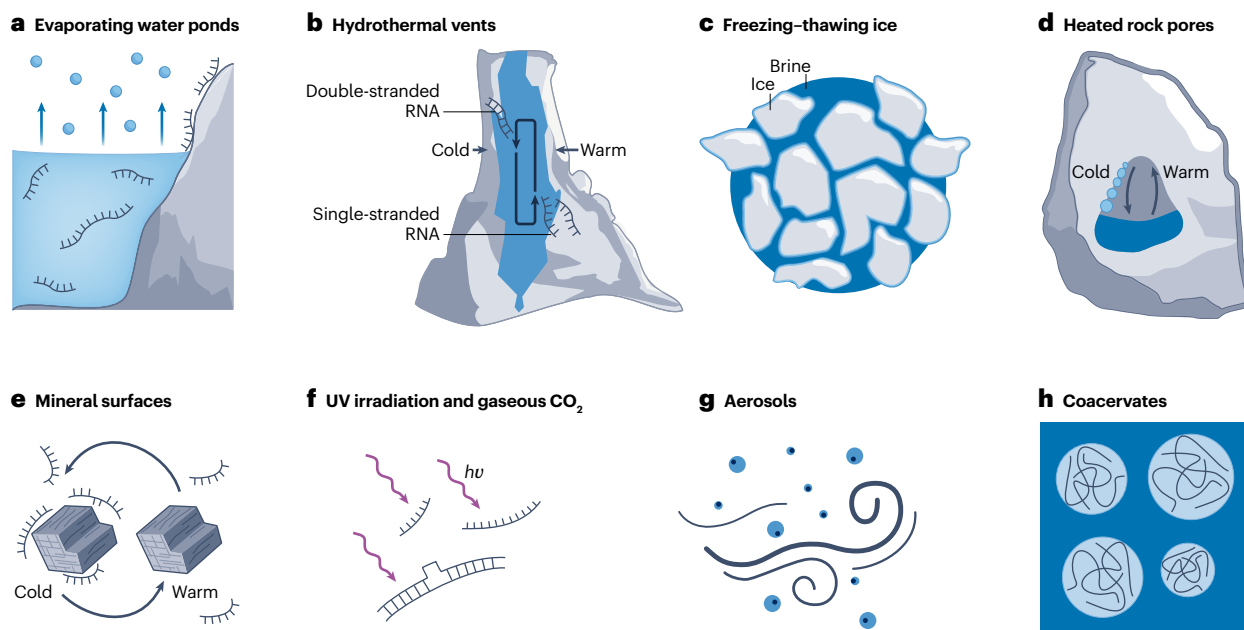


Fig. 1 | Proposed primordial geological settings for fluctuations in temperature, salts and pH, ultraviolet irradiation and high CO₂ concentrations. **a**, Evaporating water ponds on surficial land. **b**, Submarine hydrothermal vents. **c**, Freezing–thawing cycles in ice. **d**, Heated rock pores.

e, Mineral surfaces. **f**, Atmospheric ultraviolet radiation and gases. **g**, Aerosols. **h**, Liquid–liquid phase separation, forming coacervates. These systems are chemically rich and offer a variety of non-equilibrium forces that could have enhanced the prebiotic chemical reactions and driven the molecular evolution.

wet–dry cycles. This system continuously generates fluctuations in the concentrations of molecules, reagents and salts, which can become exceedingly high in the completely dry state. In addition, the pH, which depends on the chemical properties of the concentration-varying species, changes during evaporation. The chemical complexity of wet–dry cycles can enhance those chemical reactions that benefit from high concentrations, including the synthesis of canonical and non-canonical nucleosides^{14–16}, the phosphorylation of prebiotic molecules¹⁷, the formation of peptide bonds between amino acids¹⁸, the ester bond formation of sugar monomers into polymers¹⁹, the polymerization of RNA from single nucleotides^{20,21}, and the emergence of ribozymes and transfer-RNA-like structures²², among others. However, because nucleobases rapidly photodissociate during illuminated dry periods and are lost into pond seepage during the wet periods, it is possible that the synthesis of nucleotides and the subsequent RNA polymerization might require a few wet–dry cycles²³.

Chemical and temperature gradients in hydrothermal vents

Later, submarine hydrothermal vents (Fig. 1b) could have played a role as habitats for the first forms of life^{24–27}. These systems are chemically rich and contain dissolved chemicals from interactions with the newly formed oceanic crust and the underlying hydrothermal system. They are expected to host pH gradients and drive a variety of redox reactions²⁸. Moreover, rock cavities in their proximity are subject to steep temperature gradients, which induce convection and thermophoresis, cycling the solutes between hot and cold temperatures and inducing their accumulation^{29,30}. For DNA, thermal replication and thermal accumulation could be combined^{31,32}. A considerable number of chemical reactions of smaller metabolic molecules have been demonstrated to occur in these systems: redox reactions involving iron, sulfate and methane²⁸; the interconversion and formation of amino acids^{33,34}; RNA oligomerization³⁵; and DNA accumulation and polymerization^{31,36,37}. However, the absence of wet–dry cycles within hydrothermal vents has limited RNA polymerization in these systems to short oligomers^{23,35}. Moreover, the high temperatures of the hydrothermal vents pose a problem for hydrolysis reactions, which may degrade the newly formed polymers³⁸, but could also open new pathways for protometabolic reactions³⁹.

Freezing–thawing ice in brine microchannels

Another geological setting to consider is ice (Fig. 1c), which should not be excluded at the poles or at exposed landmass on the early Earth, or in the scenarios of variable CO₂ partial pressure and the faint Sun hypothesis^{40,41}. During the formation of ice from seawater (or other solute-rich solutions) water freezes to create ice crystals of pure water, while the remaining solute-rich liquid brine progressively concentrates in interstitial microchannels with a size of several micrometres, subjecting the molecules to oscillatory conditions under freeze–thaw cycles⁴². At lower temperature, when the salt solubility limits are reached, salts can precipitate and create crystals. Degassing leads to gas bubbles, which are incorporated into the brine channels and the ice. Gradients of pH between the brine and the ice are created⁴³. The ice's eutectic phase and freezing–thawing seawater offer a large chemical variety that has been shown to accelerate many chemical reactions involving RNA. Some notable examples are the templated polymerization of aminonucleotides^{42,44}; the assembly of active ribozymes⁴⁵; the ribozymatic replication of RNA⁴⁶; and a one-pot pathway from nucleotide activation to non-enzymatic RNA extension after repeated freeze–thaw cycles⁴⁷.

Microscale water cycles in rock pores

Heated rock pores containing gas and water (Fig. 1d) are another geological setting that was probably abundant when life emerged, and the prebiotic relevance of which has been explored^{48,49}. This system exists in the proximity of heat sources (such as magmatic activity in general, active volcanic regions or hydrothermal systems), with gas bubbles arising from magma degassing or from the incorporation of air. The differentially heated surfaces inside the pore trigger a water cycle at the microscale, generating temporary salt and pH fluctuations, as well as wet–dry cycles. This setting could host the crystallization of ribose, RNA phosphorylation and gelation, ribozyme catalysis and the formation of lipid vesicles⁴⁹. The pH and salt fluctuations of the water cycle also promote strand separation at low temperatures^{48,50}. Moreover, it has been shown that CO₂-acidified water drives DNA replication (albeit in an enzymatic manner), providing evidence for sequence evolution towards longer strands with an AT:GC composition that depends on the physico-chemical conditions of the environment^{51,52}.

Charged surfaces and leaching on mineral surfaces

The interaction of mineral surfaces with water on the early Earth (Fig. 1e) is almost inevitable, and part of all the systems discussed above. A variety of minerals form the walls and surfaces of hydrothermal vents, build the rocks hosting pores, and form the sediments within water ponds. The minerals offer a variety of effects on prebiotic molecules and chemistries and probably played a major role in molecular evolution⁵³. Different types of minerals, rocks and glasses have been reported to enhance various chemical reactions: borate minerals for the synthesis of RNA components such as ribose⁵⁴; phosphate minerals for the phosphorylation of nucleosides⁵⁵; the synthesis of DNA and RNA oligomers on montmorillonite and hydroxyapatite^{56–58}; the regulation of ribozymatic catalytic activity by tholeiitic basalt surfaces⁵⁹; RNA polymerization promoted by montmorillonite and amino acids⁶⁰; the synthesis of polyribonucleic acid promoted by rock glasses⁶¹; and the preferential accumulation of long RNA molecules⁶², among other examples. Mineral–water interactions also result in ions and molecules leaching out of the minerals over time by dissolution, ion exchange and surface complexation reactions⁶³; thus minerals can act as buffers for aqueous solutions⁶⁴ or provide ions to ribozymes and oligonucleotides⁵⁹, depending on the type of mineral used and the ions it contains. In addition, clay mineral surfaces, in particular those of montmorillonite, have been shown to interact with hairpin ribozymes to protect them from degradation by ultraviolet (UV) irradiation, retaining the catalytic activity of the ribozymes for longer times⁶⁵.

Irradiation by high-energy photons

Atmospheric factors such as UV irradiation and the concentration of gaseous CO₂ also have major effects on prebiotic chemistry (Fig. 1f). Without access to molecular oxygen and the resulting ozone layer, the surface of the early Earth was reached by UV radiation with wavelengths as low as 200 nm (ref. ⁶⁶). At 260 nm, the solar photon flux on the surface reached intensities of 10¹²–10¹³ photons per square centimetre per second (ref. ⁶⁶). Owing to the high photon energy in the range of 200–300 nm, more complex organic compounds such as nucleobases could be damaged⁶⁷. Although in aqueous solutions dissolved metal ions can shield this radiation (such as bromide for wavelengths $\lambda < 220$ nm, or Fe²⁺ for 220 $< \lambda < 300$ nm), in natural waters on the early Earth the penetration depth of 200–360 nm light reached up to several metres⁶⁸. Therefore, it is essential to consider the influence of UV light as a physical selection pressure on prebiotic chemistry.

The canonical nucleobases (A, C, G, T and U) are particularly resistant to UV irradiation⁶⁹ because they can convert the absorbed energy into heat on sub-picosecond timescales^{70,71}. In contrast, other precursors, such as 2-aminooxazole or 2-aminoimidazole, are more susceptible to UV damage^{72,73}. Ultraviolet radiation is not only destructive but can also catalyse the interconversion of nucleosides through photo-induced hydrolysis⁷⁴, and drive redox reactions to produce nucleoside, amino acids and lipid precursors^{75,76}.

Besides UV radiation, other ionizing radiations such as solar X-rays⁷⁷, solar protons⁷⁸ and gamma rays from radioisotopes⁷⁹ (for example, ⁴⁰K, ²³²Th, ²³⁵U and ²⁴⁴Po) were abundant on the Hadean Earth surface and inside nuclear geysers⁸⁰ (underground natural nuclear reactors). Ionizing radiation has been historically shown⁸¹ to drive various chemical reactions, such as the formation of amines, amino acids and a variety of organic compounds (such as alanine, glycine, formic acid, ammonium acetate and urea). More recently, it has been shown how ionizing radiation can be an important energy source in prebiotic synthesis, for the formation of formamide⁸², sugars and nucleosides^{78,83}, and to generate chemical diversity by radiolytically produced prebiotic precursors⁸⁴. Ionizing radiation can also aid the genetic mutations that are necessary for evolution⁸⁵. However, it can also disintegrate biomolecules by radiolysis and trigger the formation of radicals that have detrimental effects on the origin of life⁸⁶.

Gaseous CO₂ leading to water acidification and mineralization

The pressure of gaseous CO₂ in the Hadean atmosphere has been estimated to have been in the range of 0.01–10 bar, much higher than today (-3.5×10^4 bar)^{40,41,87,88}. When CO₂ is absorbed into water, it leads to the formation of carbonic acid, bicarbonate and carbonate, and therefore to water acidification⁸⁹. In the absence of additional buffers (for example in condensation water formed during the water cycle), the pH is directly proportional to the partial pressure of CO₂ (ref. ⁹⁰). It has been argued that surface waters on the Hadean Earth were therefore probably acidic, which would have profound implications for the prebiotic chemistries^{91,92}, for the molecular evolution of oligonucleotides⁵¹ and for the climate of the early Earth⁹³. An atmosphere rich in CO₂ could also have had a role in the release of phosphate from phosphate minerals such as apatite; the increased solubility of apatite in CO₂-acidified water could strongly increase the availability of phosphate at concentrations and pH levels relevant for prebiotic syntheses⁹⁴.

Compartmentalization and drying of aerosols

Another system of prebiotic relevance is atmospheric aerosols (Fig. 1g), generated by wind upon the breaking of waves at the ocean surface, bubble-bursting or geysers⁹⁵. The aerosol particles, the size of which is determined by aerodynamic drag, surface tension and gravity, have a high surface-to-volume ratio. The organic molecules and long-chain surfactants rearrange at the aerosol–wind interface and form a spherical monolayer that encloses an aqueous core, a configuration known as the ‘inverted micelle’ model^{96,97}. This structural rearrangement implies that most of the organic content of the particle resides at the surface, increasing the concentration of the molecules in that region. During their passage through fluctuating fields of humidity and temperature in the atmosphere, the aerosols rapidly lose water from the central core until a full external organic monolayer is formed⁹⁸. The chemical variety, size and the surfactant wall give aerosols strong similarities to simple cells. Reactions such as polymerization, OH oxidation reactions, compartmentalization and the formation of various organic molecules have been demonstrated to be enhanced in the crowded

and dehydrated interior of aerosol particles, which also experience pH changes during their lifetime^{96,99–101}.

Liquid–liquid phase separation and compartmentalization

Coacervation, the liquid–liquid phase separation of oppositely charged polyelectrolytes in water (Fig. 1h), has been hypothesized to have played a major role during the origin of life. The membrane-free protocells produced by coacervation can spatially localize molecules, aid fatty acid bilayer assembly¹⁰², selectively concentrate a range of different molecules^{103–105} and tolerate catalytic reactions involving RNA and ribozymes^{106,107}. Coacervate droplets of polypeptides, polysugars and oligonucleotides are highly dynamic and form with heterogeneous chemical identity¹⁰⁸. Interesting effects arise when they are put in a fluctuating environment, such as microfluidic water cycles inside heated rock cavities. There, the fluxes generated by moving water rapidly make the coacervate droplets fuse together, divide and fragment¹⁰⁹.

Non-equilibria to drive molecular evolution

To test the physical non-equilibrium constraints described above in a chemical context, many example systems are available. These examples are not necessarily limited to canonical ribonucleotides, as non-canonical systems could have been precursors to an RNA world or other hypothetical scenarios^{14,110–112}. However, because the properties of canonical ribonucleotides are exceptionally well studied and many deep analytical methods such as high-throughput sequencing have been established, in this Perspective we focus on how combinations of physical non-equilibrium systems can enable RNA-related reactions and push their sequence space by applying external selection pressures. That said, it is still a matter of debate as to which molecule came first: RNA, DNA, proteins, RNA–protein hybrids, or their coexistence; and which function they would carry^{113–117}.

Strand separation under temperature, salts and pH synergism

One of life’s most important features is the ability to continuously copy its genetic blueprint, stored in DNA and RNA. Its replication requires conditions in which the copied product strand can detach from its template, which is typically problematic under standard conditions. This difficulty is also known as the strand separation problem¹¹⁸. The issue arises from the fact that long RNA duplexes have a very high melting temperature (sometimes even higher than the boiling point of water), and their accurate copying would generate a dead-end duplex product. Moreover, the rate of strand reannealing at high strand concentrations is orders of magnitude faster than the current copying chemistries. Therefore, it is considered central to understand how the previously discussed geophysical systems can offer kinetic, non-equilibrium solutions to this problem.

The temperature, salts and pH fluctuations that occur in geological settings are subject to various geological processes of the early Earth. As discussed below, they can act synergistically to induce strand separation at moderate physico-chemical conditions. Figure 2a–d summarizes the main determinants of the duplex stability of oligonucleotides. Temperature, concentration of monovalent and bivalent ions¹¹⁹, and pH¹²⁰ determine whether a specific oligonucleotide sequence is in the single-stranded or double-stranded conformation¹²¹. A single one of these factors is usually not enough to trigger strand separation without reaching extreme values, which could damage RNA and chemical reagents (as is the case with very high temperatures), or could be incompatible with prebiotic chemistries (such as for very low salts or pH). Instead, when all these factors act in synergy, a substantial

Perspective

strand separation can be achieved in moderate regimes that can be compatible with many replication chemistries (Fig. 2e).

The environment drives sequence evolution

Several reports have shown how the environment can influence the final sequence composition and length of replicated, ligated or polymerized sequences^{51,52,122–124}. In a simulated replicative environment of the RNA world, the sequences adapted to the melting conditions of an acidic microscale water cycle⁵¹. The pH, salts and temperature fluctuations created specific regimes of melting conditions. Only the sequences with intermediate duplex stability could continuously melt and reanneal, and thereby undergo repeated replication cycles. This effect strongly biased the resulting sequence length and ATGC content, creating long sequences whose AT/GC fraction correlated with the melting conditions. For example, mild denaturing conditions promoted the replication of sequences rich in A and T, whereas stronger ones promoted the creation of sequences richer in G and C (Fig. 3a). This effect derives from the fact that the AT/GC ratio of a duplex oligonucleotide sequence is a major factor that determines its stability against denaturation¹²⁵.

On an early Earth, therefore, different environments could host pools of sequences enriched in specific nucleotides. This differentiation would strongly reduce the occupied states in the sequence space¹²⁶ and would render the spontaneous emergence of particular sequence biases or motifs far more possible: an essential feature for the emergence of ribozymes in an evolving RNA world^{127,128}. The sequence space of long oligonucleotides is immense. Without a mechanism to selectively reduce the sequence space, any sequence bias is likely to remain hidden and not emerge. The non-equilibrium features discussed above are able to circumvent this limitation, selecting specific subsets of sequences of defined nucleotide composition^{51,129}. Moreover, these observations could give an explanation to the paradox of the dual roles of RNA in the RNA world, as both information carrier and catalyst¹³⁰. These dual roles require the RNA to have good templating ability and stable folding, which are in conflict with each other. However, an open-ended replicating system drives the evolution of sequences with intermediate stability, whose melting temperature (T_m) is close to the environmental temperature. In that temperature regime, sequences can repeatedly fold and unfold, satisfying the dual role of RNA in the RNA world.

Ribozymes under non-equilibria

The catalytic activity of ribozymes is strongly influenced by salts, temperature and pH, because these factors directly affect their folding^{131–134}. For example, RNA-stabilizing conditions (that is, high salts and low temperature) can shift the equilibrium of the hairpin ribozyme towards ligation. Conversely, destabilizing and melting conditions shift the equilibrium towards cleavage¹²³. It is therefore an attractive hypothesis that fluctuating environmental conditions could tune the activity of ribozymes and introduce biases into a pool of replicating sequences, to drive molecular evolution towards a net direction (Fig. 3b).

In addition, the interactions of ribozymes with mineral surfaces insert an additional layer of regulation to their catalytic activity. As ions are liberated from minerals, heat flows actively alter the ionic salt ratios (for instance, the ratio of Mg^{2+} to Na^+) to an extent that enables key ribozyme activities in otherwise challenging solution conditions⁵⁹.

Kinetics-mediated sequence evolution

Salts, temperature and pH also affect the kinetics of annealing of complementary strands. The annealing process is slower for lower salts and pH^{48,135–140}. Therefore, denaturing conditions increase the time

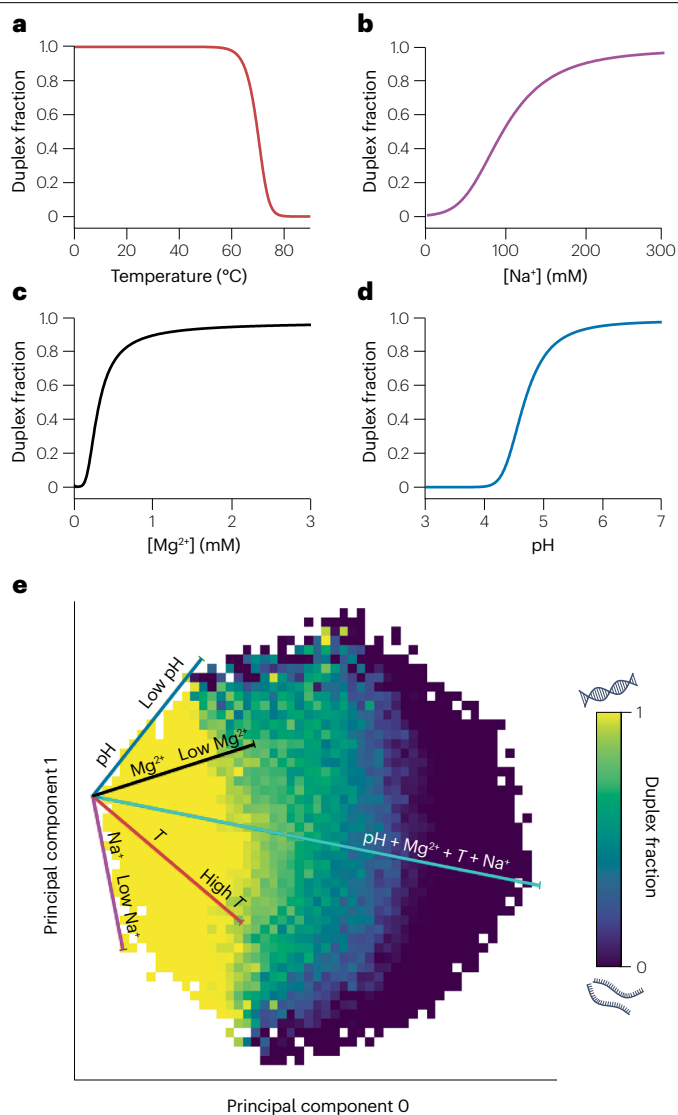


Fig. 2 | The chemico-physical determinants of oligonucleotide stability.

a–d, Melting curves of a 35-nucleotide RNA oligonucleotide (30% GC) as a function of temperature (**a**), concentration of monovalent ions (Na^+) (**b**), concentration of bivalent ions (Mg^{2+}) (**c**) and pH (**d**). **e**, Synergistic effects of temperature (T), concentration of Na^+ , concentration of Mg^{2+} and pH, visualized at reduced dimensionality (2D) via principal component analysis (kernel PCA). Combinations of factors provide enhanced strand separation efficiency. Details are explained in Supplementary Section 1.

required for strand annealing. With a limited time window available, only the sequences with a faster and more accurate annealing (such as those with fewer mutations) would be able to remain active during the replication process. The sequences that are unable to anneal in that time stall and die out.

This mechanism can trigger sequence-selective processes that bias the sequence pool of a replicative system (Fig. 3c). One example is the selection of sequences containing fewer mutations (kinetic error filtering¹²⁴), a sort of non-enzymatic proofreading method. Another example is a self-enhancing sequence selection that arises

Perspective

from kinetic stalling of mismatching sequences; this sequence selection can strongly reduce the occupied states in the sequence space and promote self-amplification¹²².

Sequence-selective UV photodamage

In addition to the impact on prebiotic chemistry discussed above, UV radiation can also photodamage oligonucleotides in a sequence-selective manner¹⁴¹. In particular, the excitation energy obtained by the absorption of UV photons can lead to the formation of dimeric lesions from neighbouring pyrimidine nucleotides (TT^{142,143}, TC¹⁴⁴) or adenosine nucleotides¹⁴⁵, which can deform the oligomer¹⁴⁶. Furthermore, the formation of charge transfer states^{147,148} in the vicinity of guanosine nucleotides can prevent damage formation, resulting in a strong anticorrelation between the GC content and the susceptibility to UV light¹⁴⁹. These factors add up to a complex sequence dependence of UV resistance in oligonucleotides¹⁵⁰. Comparing the UV damage of DNA sequences of 7 nt length (Fig. 3d) indicates that strands that are more resistant to UV damage are enriched in G and C, which also results in a higher melting temperature T_m of the strands. In addition, double-stranded DNA is more resistant to UV damage than single-stranded oligonucleotides¹⁵¹. Both effects thus reinforce each other and select more stable, GC-rich DNA duplexes, which are particularly well protected against UV. These observations pose stringently measurable selection and regulation pressures on early sequence pools such as RNA aptamers¹⁵².

Non-canonical sequence evolution

Over the years, many additional approaches have been proposed that use complex chemistry and specific RNA or DNA sequences to trigger processes of replication, recombination and strand separation.

In rolling circle RNA synthesis^{153,154}, a circular template aids continuous replication by overcoming the problem of the fast reannealing of the complementary strands, which occurs at high concentrations. The model is valid to describe the downstream evolution of polymerase-catalysed replication. It requires the loop to close after the replication in order to offer an exponentially growing replication, which is considered important to maintain information against the exponentially decaying dynamics of degradation. Note that the rather long persistence length of double-stranded RNA of 63 nm (or 250 bases) would probably restrict rolling circle replication under physiological conditions to longer sequences¹⁵⁵.

Another way to provide strand displacement can be through invading RNA strands¹⁵⁶. Short complementary RNA fragments, derived from degradation, partial replication or non-templated polymerization, are able to attack a double-stranded region and partially open it to allow this short primer to be elongated. This type of strand displacement could allow RNA replication cycles to be completed when the reannealing of complementary strands becomes too fast.

Based on the observation that there are no clear prebiotic mechanisms to discriminate between DNA and RNA, RNA–DNA chimeras have

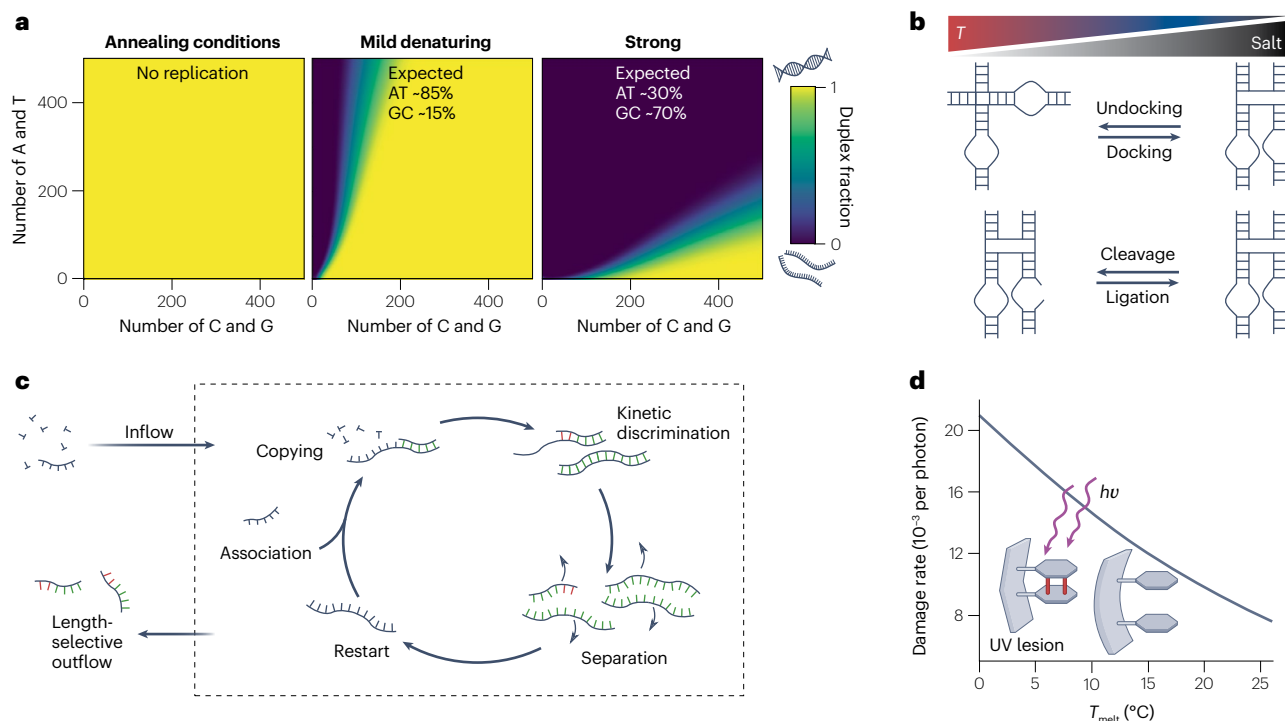


Fig. 3 Non-equilibrium environments to shape oligonucleotide evolution.

a, The nucleotide composition of replicating sequences is determined by the chemico-physics of water. Denaturing conditions to form single strands correlate with an increased GC content. Left: annealing conditions (10 $^{\circ}C$, pH 7, 300 mM Na^+ , 3 mM Mg^{2+}); middle: mild denaturing (65 $^{\circ}C$, pH 4.5, 20 mM Na^+ , 0.3 mM Mg^{2+}); right: strong denaturing (75 $^{\circ}C$, pH 4, 1 mM Na^+ , 0.1 mM Mg^{2+}). **b**, The catalytic activity of the hairpin ribozyme is regulated by temperature and concentration of salts, which

influence its folding and conformation. **c**, Kinetic selection mechanism for sequence length and error discrimination. The hybridization kinetics of oligonucleotides is also affected by the chemico-physical parameters. **d**, Sequence dependence of ultraviolet damage, indicated as damage rate per photon, for 7-nucleotide single-stranded DNA after 2.5 h of ultraviolet irradiation. Higher melting temperature (T_{melt}) corresponds to stable duplex sequences (GC-rich), which are less susceptible to ultraviolet damage. Details are explained in Supplementary Section 2.

Perspective

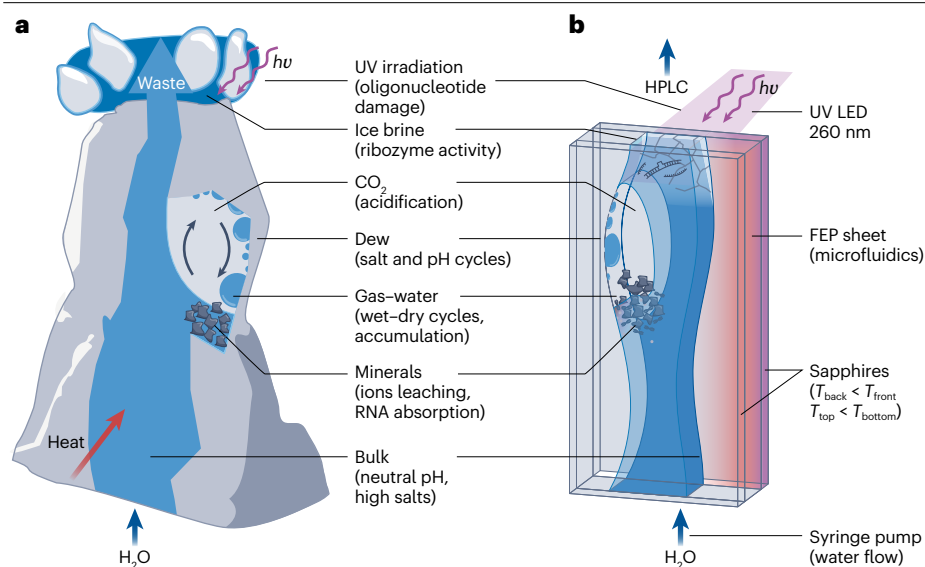


Fig. 4 | Heterogeneous chemico-physical possibilities to drive the origin of life. **a**, Multiple non-equilibrium features in the same geological setting are prebiotically plausible. Such a complex system would generate unexplored scenarios for the evolution of the first oligonucleotides on the early Earth. **b**, It is experimentally feasible to investigate such a scenario. A microfluidic implementation at the millimetre scale is possible using standard components: electric components for heating and cooling, fluorinated ethylene propylene (FEP) sheets as containers, transparent sapphire windows for imaging, syringe pumps to flow water, and light-emitting diodes (LEDs) for illumination and ultraviolet irradiation. *T*, temperature; HPLC, high-performance liquid chromatography.

been proposed to have played a role during molecular evolution^{157–159}. These chimeras have weaker base pairing than the standard duplexes, indicating they would be less affected by the strand separation problem under non-denaturing conditions. Moreover, they add a new possibility to the RNA–DNA–protein world transition model, suggesting the coexistence and co-evolution of RNA and DNA systems during the initial stages. Adding the more stable DNA to the mix for long-term storage of sequences offers many advantages, but it remains to be seen if they are outweighed by the disadvantages associated with the co-replication and co-synthesis of DNA together with RNA. Moreover, it remains unclear what chemico-physical conditions are suited for the replication of a hybrid system containing both RNA and DNA, as they have very different pH, salt and temperature preferences. For example, RNA is more susceptible to hydrolysis than DNA at high temperatures. Moreover, the conditions that enhance RNA catalytic activity¹⁶⁰ (high salts, neutral pH, moderate temperature) are those that better stabilize the DNA double-stranded conformation, posing the problem of how strand separation is to occur. Therefore, salt-poor wet–dry cycling could provide a new perspective¹⁶¹; or non-equilibrium environments hosting fluctuating pH, salts and temperatures could offer an interesting environment in which to test the molecular evolution of these types of hybrid systems.

Many of the above ideas to replicate RNA have not yet been tested under geological non-equilibria. Based on experience, we believe that interesting and unexpected dynamics will be found by this connection of geoscience, physics and chemistry. Moreover, the combination of the discussed geophysical features, when coupled with more complex prebiotic chemistries, are expected to offer synergistic results and are now becoming experimentally feasible.

Primordial scenarios and experimental analogues

It is plausible that heterogeneous physico-chemical microscale properties would have been present in close proximity. For example, a porous rock containing water and CO₂ gas inclusions could exist near to a volcanic heat source, bombarded at its top by UV irradiation and undergoing day–night or seasonal freeze–thaw cycles (Fig. 4a). The

rock part could vary between micrometres and tens of metres. In such a setting, ions are leached from the minerals to buffer the pH of the bulk and selectively absorb RNA; convection and thermophoresis in the bulk and the gas–water boundaries of the bubble exponentially and selectively accumulate molecules; microscale water cycles in the gas bubble induce pH and salt fluctuations that facilitate strand separation; the UV irradiation induces selective damage to oligonucleotides; and freeze–thaw cycles create ice brines and locally enhance the activity of ribozymes.

This system can be efficiently transferred to a lab experiment and implemented in millimetre-sized microfluidic experiments (Fig. 4b). Heat flow can be provided by differentially heating and cooling the surfaces of the microfluidic chamber, CO₂ gas bubbles can be introduced microfluidically, a continuous liquid flow can be controlled with syringe pumps, realistic geomaterials can be introduced as powdered minerals, and the sample chamber can be locally irradiated with UV-light-emitting diodes. However, the combinatorial space of the prebiotic physico-chemical and environmental conditions is immense, and so the identification of the key factors is not generalizable but depends on each specific scenario. This means that the results are likely to be unpredictable, and mutual, synergistic or inhibitory effects cannot be determined a priori. One possible approach is to systematically study the effects of each non-equilibrium feature separately. Multiple non-equilibrium features can then be merged, once the single effects have been quantified.

Researchers have started to make use of heterogeneous geo-physico-chemical conditions to study and enhance prebiotic chemical reactions in such more complex physical boundary conditions^{162–165}. One example is the production of nucleobases, ribose, and nucleotide precursors in a comprehensive model of the early Earth atmosphere coupled with an evaporating water pond¹⁶⁶. The system includes lightning, UV, impact degassing, volcanism, ocean geochemistry and water evaporation. The UV irradiation, lightning and wet–dry cycles in an atmosphere with plausible levels of water, CO₂, H₂ and CH₄ enable a series of complex chemical reactions, from the formation of HCN to sugars and nucleotide precursors. Such a level of complexity would

have not been possible without the interaction of multiple geological features. The interplay between such geological settings and prebiotic chemical reactions is likely to create unpredictable results, with mutual, synergistic or inhibitory effects in a complex reaction network.

Enzymatic replication

Whereas the chemistry for prebiotic RNA replication will probably remain a puzzle for some time, it might be a good exercise to probe the physics of the environment with polymerase or ligase proteins. And indeed, this approach has shown first signs of Darwinian-like evolution. For example, the three-body reaction of templated ligation was observed to create a strong self-selection of initially random sequences, creating a unusually complex sequence space from the random pools¹²⁹. With this self-selection, a sparse exploration of the sequence space becomes possible without having to scan the masses of the Universe that are theoretically necessary to hold all 4^{100} sequences for 100-nucleotide oligomers, as would be required to perform evolution in a very systematic way.

As seen by following the same approach, enzymatic replication at air–water interfaces has interesting strand separation dynamics⁵¹. Gas bubbles inside thermal gradients, which experimenters normally avoided at great cost in microfluidic experiments, are now hot spots for replicative polymerization. The speed of creating longer strands in these experiments, enzymatically, and protocolled by deep sequencing is compelling. Within a few hours, sequences more than 1,300 nucleotides in length evolved from 50-nucleotide initial templates, adapting to the length-selective non-equilibrium setting at the air–water interface and tuning the AT/GC ratio to fit the conditions of strand separation⁵¹. Central to this experiment is the local salt and pH cycling at the heated air–water interface, coupled with the capillary flows in the liquid that preferentially accumulate longer oligonucleotides at the gas–water boundaries. Moreover, the replication efficiency of a cold air–water interface⁵¹ or a local heated spot¹⁶⁷ is much higher than ever achieved in bulk polymerase chain reaction (PCR) or systematic evolution of ligands by exponential enrichment (SELEX) experiments¹⁶⁸. PCR run under non-equilibria on a sequence mix can amplify specific sequences locally in micrometre-sized spots, without depleting the bulk's reagents, and reaching extremely high amplification factors. However, standard bulk amplification typically runs out of resources after 20–30 cycles.

Prebiotic polymerization

A widely investigated topic in the origin of life field is the prebiotic polymerization of RNA. To form RNA from its nucleotides – or peptides from amino acids – water must be removed^{169,170}. For nucleotides, modern biology uses triphosphates as an activation group to drive this reaction. When it comes to prebiotic scenarios, the approaches differ on whether to start with a strong activation group and drive the reaction mainly chemically^{171–173} or to use a weak activation group and mainly put the burden on the physical non-equilibria, for example by temporarily using dry or semi-dry conditions and active accumulation mechanisms. An example of the latter approach is the polymerization of 2',3'-cyclic monophosphate nucleotides at elevated pH. Under conditions of temporary drying at a heated air–water interface, RNA oligomers form¹⁶¹. One could also set templated replication reactions to operate in similar semi-dry conditions and use the salt-poor wet conditions to separate the RNA strands. This inverted mechanism could offer a number of new perspectives for replicative systems, including longer lifetimes of RNA in the predominantly dry state.

Outlook

In our view, future generations of experiments will include an increasing number of the non-equilibria discussed above, with many more to be discovered. So we think that much can be expected from the above-mentioned non-equilibrium possibilities that geo-environments offer at the microscale (Fig. 4).

We have discussed the multiple ways in which environmental factors are not mere boundary conditions, but instead play an active role in driving and shaping the molecular evolution of DNA and RNA. Based on the positive cooperative effects seen in experiments using this combination of origin of life chemistry with the geoscience of the early Earth in the past, it is likely that scenarios will emerge that can make prebiotic processes much more plausible. The purpose of this Perspective is to encourage the use of non-equilibria in origin of life experiments. We have discussed and demonstrated many intrinsic advantages of combining the physics of selection and strand separation with the chemistry of replication in a natural localized setting. Apart from these biotechnological advantages, we argue that the merging of physics and chemistry is the best route to understand, through lab experiments, that the emergence of life might not have been as rare and implausible as is commonly thought. Many more interesting molecular machines recreating the emergence of life may be found at the crossroads of replication chemistry and non-equilibrium physics. The future of cross-disciplinary experiments has only just begun in this field.

Published online: 18 January 2023

References

1. Mojzsis, S. J., Harrison, T. M. & Pidgeon, R. T. Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* **409**, 178–181 (2001).
2. Wilde, S. A., Valley, J. W., Peck, W. H. & Graham, C. M. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* **409**, 175–178 (2001).
3. Valley, J. W. et al. Hadean age for a post-magma-ocean zircon confirmed by atom-probe tomography. *Nat. Geosci.* **7**, 219–223 (2014).
4. Schrödinger, E. *What is Life? The Physical Aspect of the Living Cell* (Cambridge Univ. Press, 1944).
5. Goldenfeld, N. & Woese, C. Life is physics: evolution as a collective phenomenon far from equilibrium. *Annu. Rev. Condens. Mater. Phys.* **2**, 375–399 (2010).
6. Michaelian, K. Thermodynamic dissipation theory for the origin of life. *Earth Syst. Dyn.* **2**, 37–51 (2011).
7. Marsh, M. G. E. Thermodynamics and the origin of life. *Can. J. Phys.* <https://doi.org/10.1139/CJP-2020-0013> (2022).
8. Barge, L. M. et al. Thermodynamics, disequilibrium, evolution: far-from-equilibrium geological and chemical considerations for origin-of-life research. *Orig. Life Evol. Biosph.* **47**, 39–56 (2016).
9. Ricker, G. R. et al. Transiting exoplanet survey satellite. *J. Astron. Telesc. Instrum., Syst.* **1**, 014003 (2014).
10. Borucki, W. J. et al. Kepler planet-detection mission: introduction and first results. *Science* **327**, 977–980 (2010).
11. Ávila, P. J. et al. Presence of water on exomoons orbiting free-floating planets: a case study. *Int. J. Astrobiol.* **20**, 300–311 (2021).
12. Damer, B. & Deamer, D. The hot spring hypothesis for an origin of life. *Astrobiology* **20**, 429–452 (2020).
13. Föllmann, H. & Brownson, C. Darwin's warm little pond revisited: from molecules to the origin of life. *Naturwissenschaften* **96**, 1265–1292 (2009).
14. Becker, S. et al. Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. *Nat. Commun.* **9**, 163 (2018).
15. Kim, H. J. & Benner, S. A. Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. *Proc. Natl Acad. Sci. USA* **114**, 11315–11320 (2017).
16. Okamura, H. et al. A one-pot, water compatible synthesis of pyrimidine nucleobases under plausible prebiotic conditions. *Chem. Commun.* **55**, 1939–1942 (2019).
17. Gull, M. Prebiotic phosphorylation reactions on the early earth. *Challenges* **5**, 193–212 (2014).
18. Orsythe, J. et al. Ester-mediated amide bond formation driven by wet–dry cycles: a possible path to polypeptides on the prebiotic Earth. *Angew. Chem. Int. Ed.* **54**, 9871–9875 (2015).

19. Mamajanov, I. et al. Ester formation and hydrolysis during wet-dry cycles: generation of far-from-equilibrium polymers in a model prebiotic reaction. *Macromolecules* **47**, 1334–1343 (2014).
20. Da Silva, L., Maurel, M. C. & Deamer, D. Salt-promoted synthesis of RNA-like molecules in simulated hydrothermal conditions. *J. Mol. Evol.* **80**, 86–97 (2015).
21. 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).
22. 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).
23. Pearce, B. K. D., 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. USA* **114**, 11327–11332 (2017).
24. Corliss, J. B., Baross, J. & Hoffman, S. An hypothesis concerning the relationships between submarine hot springs and the origin of life on Earth. *Oceanol. Acta Special Issue* 59–69 <https://archimer.ifremer.fr/doc/00245/35661/> (1981).
25. Russell, M. J., Daniel, R. M. & Hall, A. J. On the emergence of life via catalytic iron-sulphide membranes. *Terra Nov.* **5**, 343–347 (1993).
26. Martin, W. et al. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Phil. Trans. R. Soc. Lond. B* **358**, 59–85 (2003).
27. Martin, W. & Russell, M. J. On the origin of biochemistry at an alkaline hydrothermal vent. *Phil. Trans. R. Soc. B* **362**, 1887–1926 (2007).
28. Martin, W., Baross, J., Kelley, D. & Russell, M. J. Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **6**, 805–814 (2008).
29. Agerschou, E. D., Mast, C. B. & Braun, D. Emergence of life from trapped nucleotides? Non-equilibrium behavior of oligonucleotides in thermal gradients. *Synlett* **28**, 56–63 (2017).
30. Baaske, P. et al. Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proc. Natl Acad. Sci. USA* **104**, 9346–9351 (2007).
31. Mast, C. B. & Braun, D. Thermal trap for DNA replication. *Phys. Rev. Lett.* **104**, 188102 (2010).
32. 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).
33. Vallentyne, J. R. Biogeochemistry of organic matter — II Thermal reaction kinetics and transformation products of amino compounds. *Geochim. Cosmochim. Acta* **28**, 157–188 (1964).
34. Aubrey, A. D., Cleaves, H. J. & Bada, J. L. The role of submarine hydrothermal systems in the synthesis of amino acids. *Orig. Life Evol. Biosph.* **39**, 91–108 (2009).
35. Burcar, B. T. et al. RNA oligomerization in laboratory analogues of alkaline hydrothermal vent systems. *Astrobiology* **15**, 509–522 (2015).
36. 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).
37. Braun, D., Goddard, N. L. & Libchaber, A. Exponential DNA replication by laminar convection. *Phys. Rev. Lett.* **91**, 158103 (2003).
38. White, R. H. Hydrolytic stability of biomolecules at high temperatures and its implication for life at 250°C. *Nature* **310**, 430–432 (1984).
39. do Nascimento Vieira, A., Kleinermanns, K., Martin, W. F. & Preiner, M. The ambivalent role of water at the origins of life. *FEBS Lett.* **594**, 2717–2733 (2020).
40. Kadoya, S., Krissansen-Totton, J. & Catling, D. C. Probable cold and alkaline surface environment of the hadean Earth caused by impact ejecta weathering. *Geochim. Geophys. Geosyst.* **21**, e2019GC008734 (2020).
41. Zahnle, K., Schaefer, L. & Fegley, B. Earth's earliest atmospheres. *Cold Spring Harb. Perspect. Biol.* **2**, a004895 (2010).
42. Trinks, H., Schröder, W. & Biebricher, C. K. Ice and the origin of life. *Orig. Life Evol. Biosph.* **35**, 429–445 (2005).
43. Bronshteyn, V. L. & Chernov, A. A. Freezing potentials arising on solidification of dilute aqueous solutions of electrolytes. *J. Cryst. Growth* **112**, 129–145 (1991).
44. Lohrmann, R. & Orgel, L. E. Template-directed reactions of aminonucleosides. *J. Mol. Evol.* **9**, 323–328 (1977).
45. Mutschler, H., Wochner, A. & Holliger, P. Freeze–thaw cycles as drivers of complex ribozyme assembly. *Nat. Chem.* **7**, 502–508 (2015).
46. Attwater, J., Wochner, A., Pinheiro, V. B., Coulson, A. & Holliger, P. Ice as a protocellular medium for RNA replication. *Nat. Commun.* **1**, 76 (2010).
47. Zhang, S. J., Duzdevich, D., Ding, D. & Szostak, J. W. Freeze–thaw cycles enable a prebiotically plausible and continuous pathway from nucleotide activation to nonenzymatic RNA copying. *Proc. Natl Acad. Sci. USA* **119**, e2116429119 (2022).
48. Ianeselli, A., Mast, C. B. & Braun, D. Periodic melting of oligonucleotides by oscillating salt concentrations triggered by microscale water cycles inside heated rock pores. *Angew. Chem. Int. Ed.* **58**, 13155–13160 (2019).
49. Morasch, M. et al. Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules. *Nat. Chem.* **11**, 779–788 (2019).
50. Lathe, R. Fast tidal cycling and the origin of life. *Icarus* **168**, 18–22 (2004).
51. Ianeselli, A. et al. Water cycles in a Hadean CO₂ atmosphere drive the evolution of long DNA. *Nat. Phys.* **18**, 579–585 (2022).
52. Ianeselli, A. *Hadean Water-dew Cycles Drive the Evolution of DNA and Protocells* (Ludwig Maximilian Univ., 2022).
53. Cleaves, H. J. et al. Mineral–organic interface processes: potential roles in the origins of life. *Chem. Soc. Rev.* **41**, 5502–5525 (2012).
54. Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. Borate minerals stabilize ribose. *Science* **303**, 196 (2004).
55. Costanzo, G., Saladino, R., Crestini, C., Ciciriello, F. & Di Mauro, E. Nucleoside phosphorylation by phosphate minerals. *J. Biol. Chem.* **282**, 16729–16735 (2007).
56. Ferris, J. P., Hill, A. R., Liu, R. & Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**, 59–61 (1996).
57. 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 (2006).
58. Gibbs, D., Lohrmann, R. & Orgel, L. E. Template-directed synthesis and selective adsorption of oligoadenylates on hydroxyapatite. *J. Mol. Evol.* **15**, 347–354 (1980).
59. Matreux, T. et al. Heat flows in rock cracks naturally optimize salt compositions for ribozymes. *Nat. Chem.* **13**, 1038–1045 (2021).
60. Namani, T. et al. Amino acid specific nonenzymatic montmorillonite-promoted RNA polymerization. *ChemSystemsChem* **3**, e2000060 (2021).
61. Jerome, C. A., Kim, H. J., Mojszis, S. J., Benner, S. A. & Biondi, E. Catalytic synthesis of polyribonucleic acid on prebiotic rock glasses. *Astrobiology* **22**, 629–636 (2022).
62. Mizuchi, R. et al. Mineral surfaces select for longer RNA molecules. *Chem. Commun.* **55**, 2090–2093 (2019).
63. Seibert, S. et al. Identification and quantification of redox and pH buffering processes in a heterogeneous, low carbonate aquifer during managed aquifer recharge. *Water Resour. Res.* **52**, 4003–4025 (2016).
64. Lacroix, E., Brovelli, A., Holliger, C. & Barry, D. A. Evaluation of silicate minerals for pH control during bioremediation: application to chlorinated solvents. *Water Air Soil Pollut.* **223**, 2663–2684 (2012).
65. Biondi, E., Branciamore, S., Maurel, M. C. & Gallori, E. Montmorillonite protection of an UV-irradiated hairpin ribozyme: evolution of the RNA world in a mineral environment. *BMC Evol. Biol.* **7**, S2 (2007).
66. Ranjan, S. & Sasselov, D. D. Constraints on the early terrestrial surface UV environment relevant to prebiotic chemistry. *Astrobiology* **17**, 169–204 (2017).
67. Crespo-Hernández, C. E. & Arce, R. Formamidopyrimidines as major products in the low- and high-intensity UV irradiation of guanine derivatives. *J. Photochem. Photobiol. B* **73**, 167–175 (2004).
68. Ranjan, S. et al. UV transmission in natural waters on prebiotic earth. *Astrobiology* **22**, 242–262 (2022).
69. Boldissar, S. & De Vries, M. S. How nature covers its bases. *Phys. Chem. Chem. Phys.* **20**, 9701–9716 (2018).
70. De Vries, M. S. & Hobza, P. Gas-phase spectroscopy of biomolecular building blocks. *Annu. Rev. Phys. Chem.* **58**, 585–612 (2007).
71. Beckstead, A. A., Zhang, Y., De Vries, M. S. & Kohler, B. Life in the light: nucleic acid photoproperties as a legacy of chemical evolution. *Phys. Chem. Chem. Phys.* **18**, 24228–24238 (2016).
72. Todd, Z. R., Szabla, R., Szostak, J. W. & Sasselov, D. D. UV photostability of three 2-aminoazoles with key roles in prebiotic chemistry on the early earth. *Chem. Commun.* **55**, 10388–10391 (2019).
73. Todd, Z. R., Szostak, J. W. & Sasselov, D. D. Shielding from UV photodamage: implications for surficial origins of life chemistry on the early earth. *ACS Earth Sp. Chem.* **5**, 239–246 (2021).
74. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).
75. Rimmer, P. B. et al. The origin of RNA precursors on exoplanets. *Sci. Adv.* **4**, eaar3302 (2018).
76. Roberts, S. J. et al. Selective prebiotic conversion of pyrimidine and purine anhydronucleosides into Watson-Crick base-pairing arabino-furanosyl nucleosides in water. *Nat. Commun.* **9**, 4073 (2018).
77. Nossen, I. et al. Habitat of early life: solar X-ray and UV radiation at Earth's surface 4–3.5 billion years ago. *J. Geophys. Res. Planets* **112**, 2008 (2007).
78. Saladino, R. et al. Proton irradiation: a key to the challenge of N-glycosidic bond formation in a prebiotic context. *Sci. Rep.* **7**, 14709 (2017).
79. Baú, J. P. T. et al. Adenine adsorbed onto montmorillonite exposed to ionizing radiation: essays on prebiotic chemistry. *Astrobiology* **20**, 26–38 (2020).
80. Ebisuzaki, T. & Maruyama, S. Nuclear geyser model of the origin of life: driving force to promote the synthesis of building blocks of life. *Geosci. Front.* **8**, 275–298 (2017).
81. Miller, S. L. & Urey, H. C. Organic compound synthesis on the primitive earth. *Science* **130**, 245–251 (1959).
82. Pastorek, A. et al. Primordial radioactivity and prebiotic chemical evolution: effect of γ radiation on formamide-based synthesis. *J. Phys. Chem. B* **124**, 8951–8959 (2020).
83. Saladino, R. et al. Meteorite-catalyzed syntheses of nucleosides and of other prebiotic compounds from formamide under proton irradiation. *Proc. Natl Acad. Sci. USA* **112**, E2746–E2755 (2015).
84. Adam, Z. R., Fahrenbach, A. C., Jacobson, S. M., Kacar, B. & Zubarev, D. Y. Radiolysis generates a complex organosynthetic chemical network. *Sci. Rep.* **11**, 1–10 (2021).
85. Parke, W. C. Ionizing radiation and life. *Biophysics* https://doi.org/10.1007/978-3-030-44146-3_8 (2020).
86. Baú, J. P. T. et al. Solid adenine and seawater salts exposed to gamma radiation: an FT-IR and EPR spectroscopy analysis for prebiotic chemistry. *Heliyon* **5**, e01584 (2019).
87. Kasting, J. F. Earth's early atmosphere. *Science* **259**, 920–926 (1993).
88. Walker, J. C. G. Carbon dioxide on the early Earth. *Orig. Life Evol. Biosph.* **16**, 117–127 (1985).
89. Byck, H. T. Effect of dissolved CO₂ on the pH of water. *Science* **75**, 224 (1932).

90. Al-Hindi, M. & Azizi, F. Absorption and desorption of carbon dioxide in several water types. *Can. J. Chem. Eng.* **96**, 274–284 (2018).
91. Wigley, T. M. L. & Brimblecombe, P. Carbon dioxide, ammonia and the origin of life. *Nature* **291**, 213–215 (1981).
92. Rimmer, P. B. & Shorttle, O. Origin of life's building blocks in carbon-and nitrogen-rich surface hydrothermal vents. *Life* **9**, 12 (2019).
93. Feulner, G. The faint young Sun problem. *Rev. Geophys.* <https://doi.org/10.1029/2011RG000375> (2012).
94. Toner, J. D. & Catling, D. C. A carbonate-rich lake solution to the phosphate problem of the origin of life. *Proc. Natl Acad. Sci. USA* **117**, 883–888 (2020).
95. Donaldson, D. J., Tervahattu, H., Tuck, A. F. & Vaida, V. Organic aerosols and the origin of life: an hypothesis. *Orig. Life Evol. Biosph.* **34**, 57–67 (2004).
96. Ellison, G. B., Tuck, A. F. & Vaida, V. Atmospheric processing of organic aerosols. *J. Geophys. Res. Atmos.* **104**, 11633–11641 (1999).
97. Murphy, D. M., Thomson, D. S. & Mahoney, M. J. In situ measurements of organics, meteoritic material, mercury, and other elements in aerosols at 5 to 19 kilometers. *Science* **282**, 1664–1669 (1998).
98. Tuck, A. The role of atmospheric aerosols in the origin of life. *Surv. Geophys.* **23**, 379–409 (2002).
99. Trueblood, J. V. et al. The old and the new: aging of sea spray aerosol and formation of secondary marine aerosol through OH oxidation reactions. *ACS Earth Sp. Chem.* **3**, 2307–2314 (2019).
100. Griffith, E. C., Carpenter, B. K., Shoemaker, R. K. & Vaida, V. Photochemistry of aqueous pyruvic acid. *Proc. Natl Acad. Sci. USA* **110**, 11714–11719 (2013).
101. Dobson, C. M., Ellison, G. B., Tuck, A. F. & Vaida, V. Atmospheric aerosols as prebiotic chemical reactors. *Proc. Natl Acad. Sci. USA* **97**, 11864–11868 (2000).
102. Dora Tang, T. Y. et al. Fatty acid membrane assembly on coacervate microdroplets as a step towards a hybrid protocell model. *Nat. Chem.* **6**, 527–533 (2014).
103. Koga, S., Williams, D. S., Perriman, A. W. & Mann, S. Peptide–nucleotide microdroplets as a step towards a membrane-free protocell model. *Nat. Chem.* **3**, 720–724 (2011).
104. Tena-Solsona, M. et al. Kinetic control over droplet ripening in fuel-driven active emulsions. Preprint at <https://doi.org/10.26434/CHEMRXIV.9978539.V1> (2019).
105. Crosby, J. et al. Stabilization and enhanced reactivity of actinorhodin polyketide synthase minimal complex in polymer–nucleotide coacervate droplets. *Chem. Commun.* **48**, 11832 (2012).
106. Drobot, B. et al. Compartmentalised RNA catalysis in membrane-free coacervate protocells. *Nat. Commun.* **9**, 3643 (2018).
107. Poudyal, R. R. et al. Template-directed RNA polymerization and enhanced ribozyme catalysis inside membraneless compartments formed by coacervates. *Nat. Commun.* **10**, 490 (2019).
108. Priftis, D., Laugel, N. & Tirrell, M. Thermodynamic characterization of polypeptide complex coacervation. *Langmuir* **28**, 15947–15957 (2012).
109. Ianeselli, A. et al. Non-equilibrium conditions inside rock pores drive fission, maintenance and selection of coacervate protocells. *Nat. Chem.* **14**, 32–39 (2021).
110. Fialho, D. M., Roche, T. P. & Hud, N. V. Prebiotic syntheses of noncanonical nucleosides and nucleotides. *Chem. Rev.* **120**, 4806–4830 (2020).
111. Hud, N. V. Searching for lost nucleotides of the pre-RNA world with a self-refining model of early Earth. *Nat. Commun.* **9**, 5171 (2018).
112. Okamura, H. et al. Proto-Urea-RNA (Wöhler RNA) containing unusually stable urea nucleosides. *Angew. Chem. Int. Ed.* **58**, 18691–18696 (2019).
113. Müller, F. et al. A prebiotically plausible scenario of an RNA–peptide world. *Nature* **605**, 279–284 (2022).
114. Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).
115. Alberts, B. et al. The RNA World and the Origins of Life. In *Molecular Biology of the Cell* 4th edn (Garland Science, 2002).
116. Cojocaru, R. & Unrau, P. J. Transitioning to DNA genomes in an RNA world: the unexpected ability of an RNA polymerase ribozyme to copy RNA into DNA has ramifications for understanding how DNA genomes evolved. *eLife* **6**, e32330 (2017).
117. Xu, J., Green, N. J., Gibard, C., Krishnamurthy, R. & Sutherland, J. D. Prebiotic phosphorylation of 2-thiouridine provides either nucleotides or DNA building blocks via photoreduction. *Nat. Chem.* **11**, 457–462 (2019).
118. Szostak, J. W. The eightfold path to non-enzymatic RNA replication. *J. Syst. Chem.* **3**, 2 (2012).
119. Gruenwedel, D. W. & Hsu, C.-H. Salt effects on the denaturation of DNA. *Biopolymers* **7**, 557–570 (1969).
120. Bunville, L. G. & Geiduschek, E. P. DNA composition and stability to acid denaturation. *Biochem. Biophys. Res. Commun.* **2**, 287–292 (1960).
121. Thomas, R. The denaturation of DNA. *Gene* **135**, 77–79 (1993).
122. 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).
123. Nesbitt, S. M., Erlacher, H. A. & Fedor, M. J. The internal equilibrium of the hairpin ribozyme: temperature, ion and pH effects. *J. Mol. Biol.* **286**, 1009–1024 (1999).
124. Göppel, T., Obermayer, B., Chen, I. A. & Gerland, U. A kinetic error filtering mechanism for enzyme-free copying of nucleic acid sequences. *bioRxiv* <https://doi.org/10.1101/2021.08.06.455386> (2021).
125. Wada, A. & Suyama, A. Local stability of DNA and RNA secondary structure and its relation to biological functions. *Proc. Biophys. molec. Biol.* **47**, 113–157 (1986).
126. De Duve, C. The onset of selection. *Nature* **433**, 581–582 (2005).
127. Wachowiak, F. & Holliger, P. Non-enzymatic assembly of a minimized RNA polymerase ribozyme. *ChemSystemsChem* **1**, 1–4 (2019).
128. Zhou, L., O'Flaherty, D. K. & Szostak, J. W. Assembly of a ribozyme ligase from short oligomers by nonenzymatic ligation. *J. Am. Chem. Soc.* **142**, 15961–15965 (2020).
129. Kudella, P. W., Tkachenko, A. V., Salditt, A., Maslov, S. & Braun, D. Structured sequences emerge from random pool when replicated by templated ligation. *Proc. Natl Acad. Sci. USA* **118**, e2018830118 (2021).
130. Ivica, N. A. et al. The paradox of dual roles in the RNA world: resolving the conflict between stable folding and templating ability. *J. Mol. Evol.* **77**, 55–63 (2013).
131. Cottrell, J. W., Scott, L. G. & Fedor, M. J. The pH dependence of hairpin ribozyme catalysis reflects ionization of an active site adenine. *J. Biol. Chem.* **286**, 17658 (2011).
132. Anella, F. & Danelon, C. Prebiotic factors influencing the activity of a ligase ribozyme. *Life* **7**, 17 (2017).
133. Peracchi, A. Origins of the temperature dependence of hammerhead ribozyme catalysis. *Nucleic Acids Res.* **27**, 2875–2882 (1999).
134. Wilson, T. J., Nahas, M., Ha, T. & Lilley, D. M. J. Folding and catalysis of the hairpin ribozyme. *Biochem. Soc. Trans.* **33**, 461–465 (2005).
135. Studier, F. W. Effects of the conformation of single-stranded DNA on renaturation and aggregation. *J. Mol. Biol.* **41**, 199–209 (1969).
136. Hinckley, D. M. et al. Coarse-grained modeling of DNA oligomer hybridization: length, sequence, and salt effects. *J. Chem. Phys.* **141**, 035102 (2014).
137. Tsuruoka, M. & Karube, I. Rapid hybridization at high salt concentration and detection of bacterial DNA using fluorescence polarization. *Comb. Chem. High. Throughput Screen.* **6**, 225–234 (2003).
138. Pörschke, D. & Eigen, M. Co-operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic–oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.* **62**, 361–381 (1971).
139. Wong, K. L. & Liu, J. Factors and methods to modulate DNA hybridization kinetics. *Biotechnol. J.* **16**, 2000338 (2021).
140. Wetmur, J. G. & Davidson, N. Kinetics of renaturation of DNA. *J. Mol. Biol.* **31**, 349–370 (1968).
141. Kufner, C. L. et al. Sequence dependent UV damage of complete pools of oligonucleotides. *bioRxiv* <https://doi.org/10.1101/2022.08.01.502267> (2022).
142. Johns, H. E., Pearson, M. L., LeBlanc, J. C. & Helleiner, C. W. The ultraviolet photochemistry of thymidyl-(3'→5')-thymidine. *J. Mol. Biol.* **9**, 503 (1964).
143. Sugiyama, T., Keinard, B., Best, G. & Sanyal, M. R. Biochemical and photochemical mechanisms that produce different UV-induced mutation spectra. *Mutat. Res. Mol. Mech. Mutagen.* **823**, 111762 (2021).
144. Lemaire, D. G. E. & Ruzsicska, B. P. Quantum yields and secondary photoreactions of the photoproducts of dTpdT, dTpdC and dTpdU. *Photochem. Photobiol.* **57**, 755–757 (1993).
145. Kumar, S. et al. Adenine photodimerization in deoxyadenylate sequences: elucidation of the mechanism through structural studies of a major d(ApA) photoproduct. *Nucleic Acids Res.* **19**, 2841–2847 (1991).
146. Lee, J. H. et al. NMR structure of the DNA decamer duplex containing double T-G mismatches of cis-syn cyclobutane pyrimidine dimer: implications for DNA damage recognition by the XPC-hHR23B complex. *Nucleic Acids Res.* **32**, 2474 (2004).
147. Pilles, B. M. et al. Identification of charge separated states in thymine single strands. *Chem. Commun.* **50**, 15623–15626 (2014).
148. Bucher, D. B., Pilles, B. M., Carell, T. & Zinth, W. Dewar lesion formation in single- and double-stranded DNA is quenched by neighboring bases. *J. Phys. Chem. B* **119**, 8685–8692 (2015).
149. Kufner, C. L., Zinth, W. & Bucher, D. B. UV-induced charge-transfer states in short guanosine-containing DNA oligonucleotides. *ChemBioChem* **21**, 2306–2310 (2020).
150. Lu, C., Gutierrez-Bayona, N. E. & Taylor, J. S. The effect of flanking bases on direct and triplet sensitized cyclobutane pyrimidine dimer formation in DNA depends on the dipyrimidine, wavelength and the photosensitizer. *Nucleic Acids Res.* **49**, 4266–4280 (2021).
151. Bucher, D. B., Schlueter, A., Carell, T. & Zinth, W. Watson–Crick base pairing controls excited-state decay in natural DNA. *Angew. Chem. Int. Ed.* **53**, 11366–11369 (2014).
152. Saha, R. & Chen, I. A. Effect of UV radiation on fluorescent RNA aptamers' functional and templating ability. *ChemBioChem* **20**, 2609–2617 (2019).
153. Kristoffersen, E. L., Burman, M., Noy, A. & Holliger, P. Rolling circle RNA synthesis catalysed by RNA. *eLife* **11**, e75186. (2022).
154. 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).
155. Abels, J. A., Moreno-Herrero, F., Van Der Heijden, T., Dekker, C. & Dekker, N. H. Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys. J.* **88**, 2737 (2005).
156. Zhou, L. et al. Non-enzymatic primer extension with strand displacement. *eLife* **8**, e51888 (2019).
157. Gavette, J. V. et al. RNA–DNA chimeras in the context of an RNA world transition to an RNA/DNA world. *Angew. Chem. Int. Ed.* **55**, 13204–13209 (2016).
158. Trevino, S. G., Zhang, N., Elenko, M. P., Lupták, A. & Szostak, J. W. Evolution of functional nucleic acids in the presence of nonheritable backbone heterogeneity. *Proc. Natl Acad. Sci. USA* **108**, 13492–13497 (2011).
159. Krishnamurthy, R. On the emergence of RNA. *Isr. J. Chem.* **55**, 837–850 (2015).
160. Hampel, A. & Cowan, J. A. A unique mechanism for RNA catalysis: the role of metal cofactors in hairpin ribozyme cleavage. *Chem. Biol.* **4**, 513–517 (1997).

161. Dass, A. V. et al. RNA oligomerisation without added catalyst from 2',3'-cyclic nucleotides by drying at air-water interfaces. *ChemSystChem* <https://doi.org/10.1002/syst.202200026> (2022).
162. Altair, T., Borges, L. G. F., Galante, D. & Varela, H. Experimental approaches for testing the hypothesis of the emergence of life at submarine alkaline vents. *Life* **11**, 777 (2021).
163. Fox, S., Pleyer, H. L. & Strasdeit, H. An automated apparatus for the simulation of prebiotic wet–dry cycles under strictly anaerobic conditions. *Int. J. Astrobiol.* **18**, 60–72 (2019).
164. Vincent, L. et al. Chemical ecosystem selection on mineral surfaces reveals long-term dynamics consistent with the spontaneous emergence of mutual catalysis. *Life* **9**, 80 (2019).
165. Vincent, L. et al. The prebiotic kitchen: a guide to composing prebiotic soup recipes to test origins of life hypotheses. *Life* **11**, 1221 (2021).
166. Pearce, B. K. D., Molaverdikhani, K., Pudritz, R. E., Henning, T. & Cerrillo, K. E. Towards RNA life on early Earth: from atmospheric HCN to biomolecule production in warm little ponds. *Astrophys. J.* **932**, 9 (2022).
167. Salditt, A. et al. Thermal habitat for RNA amplification and accumulation. *Phys. Rev. Lett.* **125**, 048104 (2020).
168. Schroeder, R. Soups & SELEX for the origin of life. *RNA* **21**, 729 (2015).
169. Borsook, H. Peptide bond formation. *Adv. Protein Chem.* **8**, 127–174 (1953).
170. Kavdia, M. Phosphodiester bond formation. *Encycl. Syst. Biol.* https://doi.org/10.1007/978-1-4419-9863-7_1598 (2013).
171. Walton, T., Zhang, W., Li, L., Tam, C. P. & Szostak, J. W. The mechanism of nonenzymatic template copying with imidazole-activated nucleotides. *Angew. Chem. Int. Ed.* **58**, 10812–10819 (2019).
172. Kaddour, H. & Sahai, N. Synergism and mutualism in non-enzymatic RNA polymerization. *Life* **4**, 598–620 (2014).
173. Blain, J. C., Ricardo, A. & Szostak, J. W. Synthesis and nonenzymatic template-directed polymerization of 2'-amino-2'-deoxythreose nucleotides. *J. Am. Chem. Soc.* **136**, 2033–2039 (2014).

Acknowledgements

Financial support came from the European Research Council (ERC Evotrap, grant number 787356), the Simons Foundation (grant number 327125), the CRC 235 Emergence of Life

(Project-ID 364653263), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy — EXC-2094 — 390783331, and the Center for NanoScience (CeNS). Funding by the Volkswagen Initiative 'Life? – A Fresh Scientific Approach to the Basic Principles of Life' is gratefully acknowledged.

Author contributions

A.I., A.S., C.M., B.E., C.L.K., B.S. and D.B. structured the manuscript and wrote the text. A.I., A.S. and C.M. designed the figures and analysed the data.

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/s42254-022-00550-3>.

Correspondence should be addressed to Dieter Braun.

Peer review information *Nature Reviews Physics* thanks Saúl Villafaña-Barajas, Philipp Holliger and the other, anonymous, referee(s) for their contribution to the peer review of this work.

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

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature Limited 2023

Ribozyme-mediated RNA synthesis and replication in a model Hadean microenvironment

Received: 1 September 2022

Accepted: 6 March 2023

Published online: 17 March 2023

 Check for updates

Annalena Salditt¹, Leonie Karr¹, Elia Salibi², Kristian Le Vay², Dieter Braun¹  & Hannes Mutschler² 

Enzyme-catalyzed replication of nucleic acid sequences is a prerequisite for the survival and evolution of biological entities. Before the advent of protein synthesis, genetic information was most likely stored in and replicated by RNA. However, experimental systems for sustained RNA-dependent RNA-replication are difficult to realise, in part due to the high thermodynamic stability of duplex products and the low chemical stability of catalytic RNAs. Using a derivative of a group I intron as a model for an RNA replicase, we show that heated air-water interfaces that are exposed to a plausible CO₂-rich atmosphere enable sense and antisense RNA replication as well as template-dependent synthesis and catalysis of a functional ribozyme in a one-pot reaction. Both reactions are driven by autonomous oscillations in salt concentrations and pH, resulting from precipitation of acidified dew droplets, which transiently destabilise RNA duplexes. Our results suggest that an abundant Hadean microenvironment may have promoted both replication and synthesis of functional RNAs.

Identifying a physicochemical environment suitable for autocatalytic template-dependent self- and cross-replication cycles of RNAs would significantly advance our understanding of plausible scenarios that could explain the emergence of life on Earth^{1–3}. The development of catalytic RNAs (ribozymes) capable of accelerating RNA ligation and polymerization over the last decades has tremendously progressed the field of protein-free RNA replication^{4–8}. However, the high thermodynamic stability of the RNA duplexes resulting from these reactions, which is further enhanced by salts present in the solution required to fold the ribozymes and often for catalysis, usually prevents the dissociation of the newly synthesized strands from their template^{3,9}. Thus, template-dependent RNA synthesis frequently leads to the formation of dead-end duplexes, which are incompatible not only with the release and folding of encoded ribozymes but also with the recycling of templates for further rounds of replication.

Temperature-induced melting appears as an attractive solution to separate product from template. However, high temperatures lead to substantial RNA degradation, especially at high magnesium concentrations^{10,11}. These factors limit thermal denaturation to short strands in specific buffers and short heating times¹², possibly implemented by a centralized convection flow¹³. Strategies including toehold strand displacement¹⁴ or nucleic acids with heterogeneous backbone chemistries^{15,16} have been studied as potential solutions to this dilemma. Recently, RNA synthesis from a circular genome has been discussed as a solution to the strand separation problem¹⁷ and limited synthesis of a catalytic RNA micromotif using rolling circle replication has been demonstrated¹⁸. Nonetheless, demonstrations of key reaction steps, particularly the general synthesis and release of more complex functional RNAs, are currently missing^{11,19}.

We previously explored geochemically plausible non-equilibrium systems that could overcome the template-inhibition challenge in the

¹Systems Biophysics and Center for NanoScience (CeNS), Ludwig Maximilian University Munich, Geschwister-Scholl-Platz 1, 80539 Munich, Germany.

²Department of Chemistry and Chemical Biology, TU Dortmund University, Otto-Hahn-Str. 4a, 44227 Dortmund, Germany. ✉ e-mail: dieter.braun@lmu.de; hannes.mutschler@tu-dortmund.de

absence of excessively high temperatures²⁰. Of particular interest are porous rock systems comprising heated air-water interfaces (hereafter referred to as AWI-systems), which can be experimentally reproduced in a defined manner by microfabrication. In this work, we demonstrate how AWI-systems allow ribozyme-catalyzed RNA replication of sense and antisense strands followed by subsequent strand-dissociation in a one-pot system. The combination of these reaction steps, which are otherwise mutually exclusive under isothermal conditions, enables the combined synthesis, release, and folding of active ribozymes. Overall, these results infer that abundant geothermal microenvironments had the potential to support replication and thus evolution of early biological systems.

Results

sunY-catalyzed RNA ligation at heated air-water interfaces

In CO₂-rich AWI-systems, nucleic acids are exposed to periodic changes in Mg²⁺ concentration and pH level, the latter of which originates

from an equilibrium of dissolved carbonic acid, bicarbonate, and carbonate in dew droplets depending on the applied partial pressure of CO₂^{20,21} (Fig. 1a, Supplementary Fig. 1). While the low pH enables transient melting of otherwise stable nucleic acid duplexes presumably due to nucleobase protonation²², the co-accumulation of RNA and magnesium ions at water-gas interfaces promotes folding and catalysis^{23,24}. This accumulation is the result of the capillary flow created by water evaporation at the warm side of the air-water interface. The effect of water convection and Marangoni flow at the interface is minimal, but contribute if molecular assemblies grow to the size of tens of micrometers²³. We, therefore, speculated about the ability of AWI-systems to provide a suitable environment for repeated RNA-dependent RNA replication. Derivatives of the self-splicing *sunY* intron from bacteriophage T4 can catalyze template-dependent oligonucleotide ligation using 5'-guanosines as a leaving group^{4,5} (Supplementary Fig. 2a, b). Due to their robust activity and independence from prerequisite activation chemistries such as phosphate- or imidazole-based leaving groups, *sunY*-derived ribozymes are an attractive model system for primitive enzymatic RNA-replication^{25,26}. Initially, we explored the Mg²⁺ concentration requirements for RNA ligation by a 182 nucleotide (nt) variant of the *sunY* ribozyme (Supplementary Fig. 2c) in CO₂-rich AWI-systems (Supplementary Figs. 1, 3, 4). To this end, we probed the template-dependent ligation of a 30 nt RNA from three oligonucleotide substrates (Fig. 1b, Supplementary Table 1, Supplementary Note 1): a 5'-Cy5 labeled substrate (M1, 13 nt), which allows direct fluorescence-based PAGE analysis, and two downstream 5'-guanosine 'activated' substrates, M2 (7 nt), and M3 (10 nt). Due to the concentration effect at the evaporation zone of the heated side of the air-water interface, a significantly lower Mg²⁺ bulk concentration (5 mM) was required to observe near-complete substrate ligation ratios (89%) compared to isothermal conditions (50 mM, 86%) after 4 h of incubation (Fig. 1c). We observed ligation in the AWI-system even at bulk concentrations as low as 1 mM MgCl₂ (17%), whereas no detectable ligation occurred in a comparable equilibrated system at constant temperature (Fig. 1c, Supplementary Fig. 5). These findings confirmed that the local up-concentration of solutes at the warm side of the AWI interface was sufficient to allow *sunY*-dependent RNA ligation at bulk magnesium concentrations considerably lower than those required under isothermal conditions.

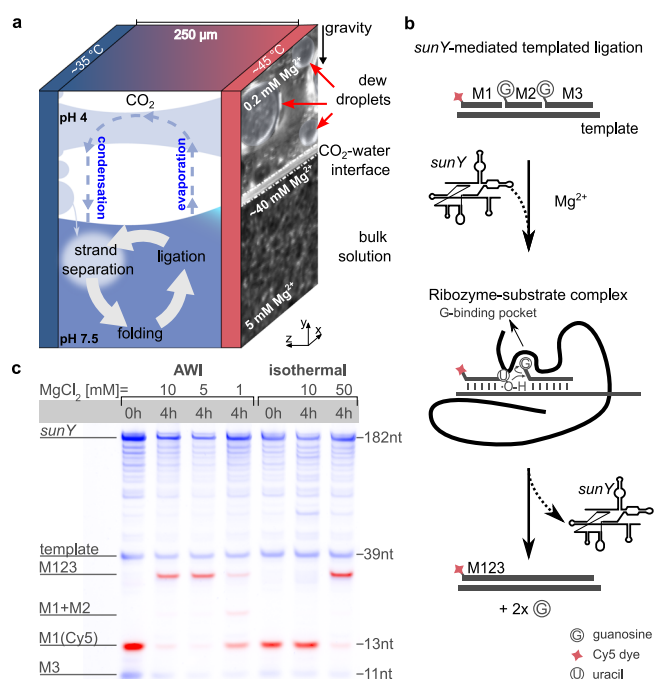


Fig. 1 | Heated AWI-systems create micro water cycles that enable RNA folding, ligation, and strand separation at low magnesium concentrations.

a An asymmetric heating triggers a temperature gradient across AWI compartments. This setup mimics a gas bubble inside a porous volcanic rock. RNA and ions are concentrated at the warm side of the interface by the evaporation of water. The evaporated water condenses as dew droplets on the cold side of the chamber. The droplets grow through surface tension and form a connection to the warm surface of the chamber, creating a miniature water cycle with strong fluctuations in salt concentration and pH similar to what has been described previously^{20,21}. **b** The *sunY* ribozyme catalyzes templated ligation of RNA strands using guanosine nucleosides (G) as leaving group as schematically shown for the ligation substrates M1-Cy5 (red label), M2 and M3. The mechanism of ligation is based on the exon ligation step during RNA splicing by group I intron: after formation of the ribozyme substrate complex, the ligation junction of the substrate duplex is positioned in the catalytic center of the ribozyme by binding of the terminal guanine nucleotide of the downstream exon in the G-binding pocket³². The aligned 3'-OH group of the upstream exon performs an inline attack on the phosphodiester bond of the bound 5'-guanosine, leading to ligation of the adjacent upstream and downstream exons and release of the guanosine and the catalytic intron. **c** PAGE analysis (two channels: SYBR Gold, blue, and Cy5, red) shows that the AWI environment enables ligation of an RNA product M123 from three RNA substrates (M1-3) at MgCl₂ concentrations as low as 1 mM. To produce yields similar to 5 mM MgCl₂ in the AWI-system, a MgCl₂ concentration of 50 mM is required under isothermal conditions. Source data are provided as a Source Data file.

AWI-systems enable *sunY*-catalyzed ligation of *sunY* fragments

We hypothesized that the periodic local pH decreases, resulting from the precipitation of acidified dew droplets back into the bulk reservoir (Supplementary Fig. 6), could lead to a transient decrease in RNA melting temperatures thereby promoting the release of ligation products from their template and allowing intramolecular folding into functional RNAs. This encouraged us to attempt to synthesize strands with the same or similar sequence as *sunY*. The Szostak group previously demonstrated that an active version of *sunY* can assemble non-covalently from three oligonucleotides (A, B, and C) between 43 and 75 nt in length⁴. We reasoned that the AWI-system might circumvent template inhibition during templated RNA ligation and therefore explored if AWI-based non-equilibrium environments could also support replication of *sunY*-derived RNA strands. We initially explored if AWI-systems enable *sunY* to catalyze the synthesis of each of the three fragments A, B, and C from three short oligonucleotides substrates (A1-3, B1-3, and C1-3) (Supplementary Table 1, Supplementary Note 1). To this end, we probed the ligation of each fragment (herein referred to as A123, B123, and C123) from three substrate strands starting from equal concentrations (2.5 μM) of *sunY* and template and fourfold excess of each ligation fragment (Fig. 2a). We observed good yields of full-length products (38.5 ± 1.5% A123, 19.5 ± 2.5% B123 and 22 ± 7% C123) after 2 h of reaction in the AWI-system and only minor amounts of incomplete intermediates. In contrast, we observed higher relative amounts of intermediate products and only low full-length yields for the three

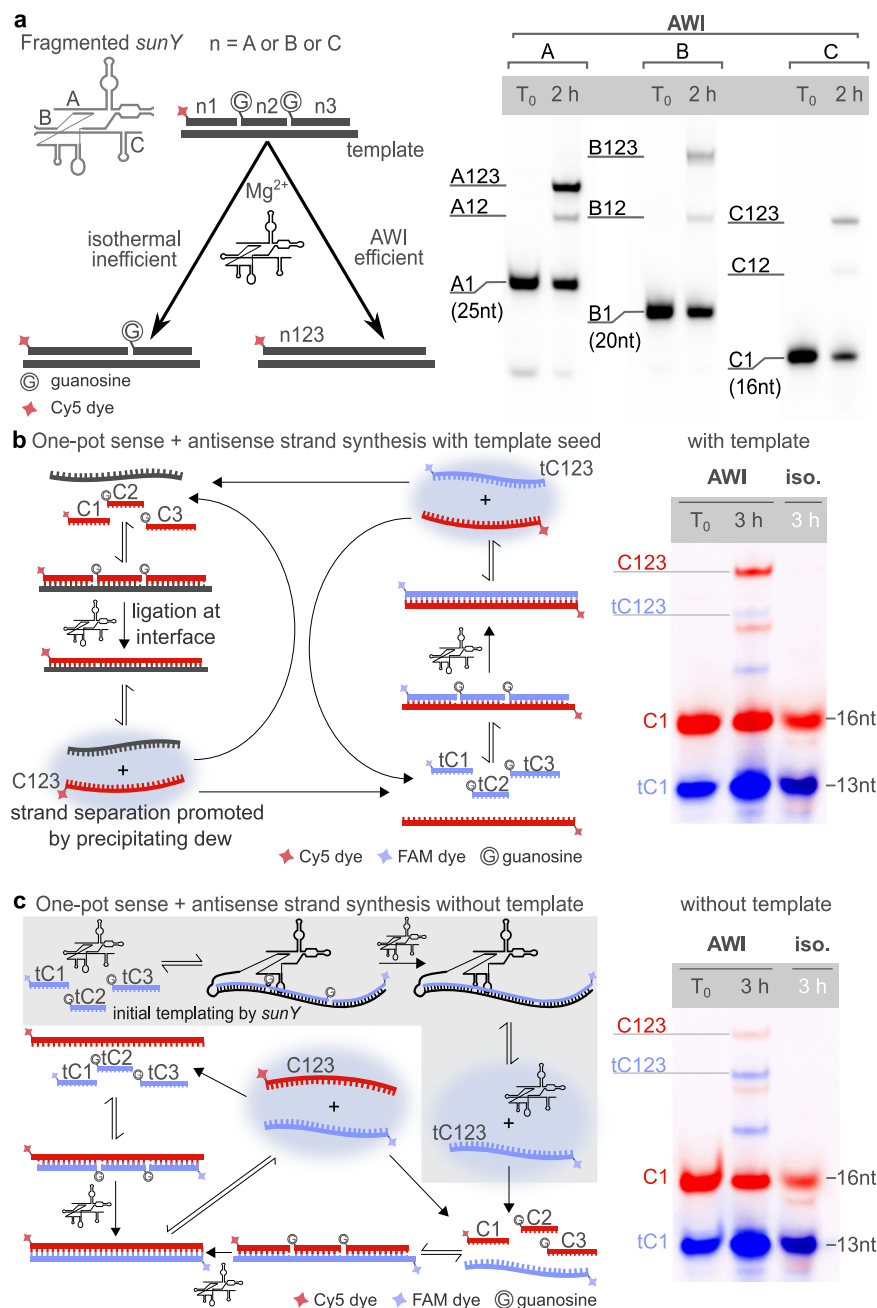


Fig. 2 | Complete replication cycles of *sunY* fragments. **a** Illustration (left) and PAGE analysis (right) of separately performed template-dependent ligations of the three *sunY* fragments A123, B123, and C123. Lanes show reactions before (T_0) and after 2 h of incubation. **b–c** Full replication cycle of fragment C123. Replication proceeds via ligation of the RNA substrates for the sense strand (C1, C2, C3-Cy5,

red) and the antisense strand (tC1-FAM, tC2, tC3, blue). In the AWI-systems, both the sense and antisense strands emerged **b** with the template present and **c** without the template present, suggesting that *sunY* partially acted as a template for antisense strand synthesis (gray box). Under isothermal conditions no products were observed after PAGE analysis. Source data are provided as a Source Data file.

fragments under isothermal conditions (~8% A123, ~1% B123, and ~1.3% C123 after 2 h), indicating that the high melting temperatures of the different RNA duplexes in the system might limit processivity under bulk conditions (Fig. 2a, Supplementary Fig. 7, Supplementary Table 2).

Coupled sense and antisense synthesis of *sunY* fragments

After demonstrating that all three *sunY* fragments can be synthesized by the full-length ribozyme in situ, we sought to explore if AWI-systems can also support full cycles of RNA replication, synthesizing both sense and antisense strands of its fragments in a single reaction environment. To this end, we carried out separate ligation experiments using

substrate oligonucleotides for the *sunY* fragments C123 and A123 as well as their corresponding templates, e.g., tA1, tA2, tA3 for tA123 (Supplementary Table 1, Supplementary Note 2). To initiate ligation of the sense fragments, we included seed amounts of tA123 and tC123 in each reaction. We detected the formation of both sense (C123, A123) and antisense strands (tC123, tA123) after 3 h of incubation suggesting that once formed, both A123 and C123 serve as templates for the synthesis of their respective antisense templates tA123 and tC123. (Fig. 2b, Supplementary Fig. 8a, b). Intriguingly, we also detected the formation of both sense and antisense fragments even in the absence of seeding template (Fig. 2c, Supplementary Fig. 8), suggesting that the *sunY* ribozyme itself can act as template for the synthesis of A123 and

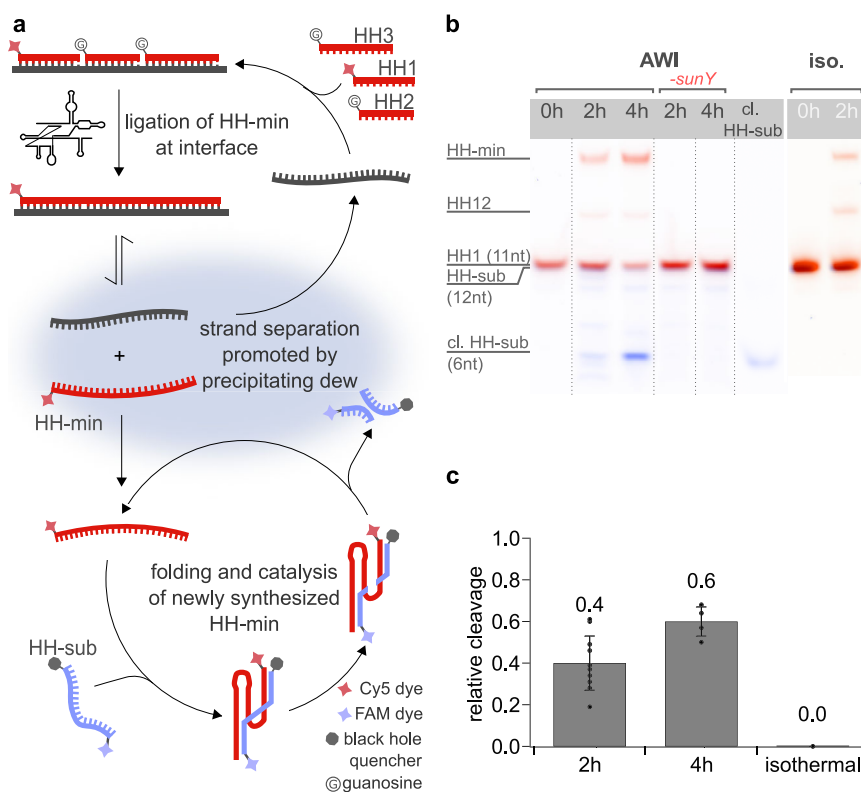


Fig. 3 | Templated synthesis and release of active hammerhead ribozyme.

a sunY catalyzes template-dependent ligation of the hammerhead ribozyme (HH-min, red) from the three RNA substrates (HH1-3, red). Local variations in pH and salt concentration cause strand separation, releasing ligated HH-min, which can then bind to and cleave its substrate (blue). **b** Ligated HH-min (red, Cy5) was observed in both the non-equilibrium and isothermal environments after 2 h of incubation. Formation of cleaved hammerhead substrate (HH-sub, blue, FAM) was observed only after incubation under non-equilibrium (AWI) conditions. Control experiments without *sunY* ribozyme exhibited no HH-min assembly or HH-sub cleavage,

demonstrating that the combined unligated HH-min fragments HH1-3 have no residual RNA-cleavage activity. The isothermal control also showed no activity. Individual gel lanes (outlined with dotted lines) have been rearranged for clarity. **c** Bar graph showing the calculated yields of HH-sub cleavage after 2 h and 4 h in the AWI-system and under isothermal control conditions. Data are presented as mean values \pm SEM. Black circles represent the values from independent replicates with $n = 10$ for 2 h incubation and $n = 4$ for 4 h incubation. Source data are provided as a Source Data file.

CI23, which subsequently can serve as template for the formation of tA123 and tCI23.

One-pot synthesis and folding of active hammerhead ribozyme

We speculated that release of the ligation products from their template might lead to folding of the products into functionally active RNAs, a process similar to transcription. To test this hypothesis, we designed a *sunY*-based assembly assay for a minimal version of the hammerhead ribozyme (HH-min) derived from the tobacco ring spot virus satellite RNA (Fig. 3a)^{27,28}. In this system, HH-min is synthesized by *sunY* from three RNA substrates (HH1-3, 11 nt, 11 nt, 16 nt), with ligation junctions interrupting critical catalytic motifs of HH-min, such that catalysis of the fragmented HH-min was prevented (Supplementary Fig. 2d, Supplementary Table 1, Supplementary Note 1). While experiments in both the AWI-system and under standard isothermal conditions led to the synthesis of HH-min, the AWI-system promoted the synthesis of HH-min even at a 10-fold lower $MgCl_2$ concentration (5 mM) than required for isothermal ligation (50 mM, Supplementary Fig. 9), corroborating our previous data regarding magnesium dependence under non-equilibrium conditions. Kinetic measurements for both conditions further revealed that the AWI-system indeed enabled multi-turnover synthesis of HH-min as the resulting yields of the full-length HH-min, which peaked after about 4 h of incubation, exceeded the amount of input template RNA (Supplementary Figs. 10–12). In contrast, no excess synthesis of HH-min over template could be observed under isothermal conditions, indicating that all product strands remained tightly associated with the template. The

observation of multi-turnover ligation in the AWI-system suggested the presence of single-stranded, de novo synthesized HH-min in solution. To probe if this dissociated RNA species was catalytically active, we repeated the experiment in presence of the cognate HH-min substrate (HH-sub) with the aim of observing catalytic substrate cleavage. Satisfyingly, PAGE analysis of the reaction mixture confirmed that HH-min formation was accompanied by cleavage of HH-sub under non-equilibrium conditions (cleavage ratio of $40 \pm 13\%$ and $60 \pm 7\%$ after 2 h and 4 h, respectively), validating that product dissociation, folding, binding and cleavage occur in addition to *sunY*-catalyzed ribozyme assembly in a one-pot system (Fig. 3b, c, Supplementary Table 3). In contrast, no HH-sub cleavage was observed under isothermal conditions, suggesting that the newly synthesized HH-min was unable to dissociate from the template under these conditions (Supplementary Fig. 13). This interpretation agrees with the predicted melting temperature of the HH-min-template complex at the magnesium concentrations used in the reaction (90 °C, Supplementary Table 4). As expected, control experiments in the AWI-system without *sunY* resulted in no observable HH-min assembly or HH-sub cleavage.

Discussion

This work revealed a credible model prebiotic reaction environment that is compatible with purely RNA-catalyzed replication of short functional RNAs. Similar non-equilibrium microenvironments might have been able to host the first autocatalytic nucleic acid replicators that marked the onset of biological evolution towards higher complexity. In its current form, AWI-systems are well suited to enable

combined replication and strand-dissociation of RNA strands between 20 and 75 nt from much shorter precursor oligonucleotides, which have the potential to assemble into more complex ribozymes. Typically, fragmented ribozyme variants are in most cases less active than their full-length counterparts presumably due to their lower stability and folding defects. In line with this, the fragmented *sunY* variant formed by noncovalent assembly of A123, B123, and C123, unlike the full-length variant, did not show detectable ligation of HH-min under standard AWI or isothermal conditions (Supplementary Fig. 14a).

Interestingly, however, we found that lowering the temperature on the warm side of the AWI-system from 45 °C to 40 °C allowed the fragmented *sunY* ribozyme to synthesize moderate amounts of HH-min, illustrating that milder conditions could compensate to some extent for the lower activity of the split variant (Supplementary Fig. 14a). To explore the autocatalytic potential of the system, we demonstrated the synthesis of all three *sunY* fragments (A123, B123, and C123) directly by the fragmented *sunY*, although the yields were considerably lower than for ligations catalyzed by the full-length ribozyme (Supplementary Fig. 14b–d). Further optimization will be required to improve reaction yields, but the results provide an optimistic outlook on the capability of the system to undergo self-replication from a pool of shorter oligonucleotide substrates.

AWI-systems can be adapted to the requirements of other ribozyme systems by altering temperatures or chamber geometries. Of particular interest are RNA polymerase ribozymes that catalyze templated primer extension reactions of equal sequence length to themselves. As previously mentioned, the problem of separating de novo RNA from its template is exacerbated by salts in solution. This is even more the case for RNA polymerase ribozymes that have been characterized under magnesium concentrations ranging from 50–200 mM^{6–8}. It remains to be seen how these systems also benefit from the salt and pH oscillations to enhance the yields by the positive feedback of template release and by reduced metal ion- and pH-mediated degradation of catalytic RNAs.

In conclusion, the data presented suggest that the heated air-water interface (AWI) system has the potential to host fragmented RNA replicators that catalyze both general and autocatalytic template-directed RNA synthesis, further strengthening the hypothesis that dynamic, non-equilibrium environments were cradles for the emergence of higher order biomolecular complexity. Intriguingly, similar interface settings can drive other prebiotically relevant key processes such as the phosphorylation of mononucleosides, polymerization of RNA, encapsulation of nucleic acids in membrane vesicles as well as the fusion and fission of coacervate-based protocells^{20,23,29,30}. Future work might explore the suitability of AWI-systems to combine these key aspects of biological systems into a single collective scenario providing a setting that could potentially support most, if not all, of the required capabilities for an inanimate system to adopt life-like behaviors.

Methods

Nucleic acids

RNA oligonucleotides were ordered from Biomers or Integrated DNA Technologies (IDT) in dry form and subsequently adjusted to 200 μM stock concentrations with nuclease-free water (Ambion™ Nuclease-free water from Invitrogen). HH-sub was synthesized by Eurofins Genomics. Final RNA concentrations were confirmed using 260 nm absorbance. All stocks were kept at –80 °C and thawed on ice prior to the experiment. Filling and extraction of AWI-systems were performed on ice. The *sunY* DNA template for in vitro transcription was ordered as a gBlock with the following sequence:

```
GATCGATCTCGCCCGCAAATTAACGACTCACTATAGGGAAA
ATCTGCCTAAACGGGAAACACTCACTAGTCAATCCCGTGCTAAAT
CAGCAGTAGCTGTAATGCCTAACGACTATCCCTGATGAATGTAAGG-
GAGTAGGGTCAAGCGACCCGAAACGGCAGACAACCTAAGAGTTGAA
GATATAGTCTGAAGTGCATGGTGACATGCAGGATC
```

The gBlock was amplified using the following primers:

P_{rev} = GATCCTGCATGTCACCATGCAGTTCAGACT;

P_{fwd} = GATCGATCTCGCCCGCAAATTAACGAC

All RNA oligonucleotide sequences are provided in Supplementary Table 1.

Transcription of *sunY*

The *sunY* DNA template was amplified by PCR using the Q5 High Fidelity 2x Master Mix (NEB) with 10–25 ng gBlock template and 1 μM P_{fwd} and P_{rev} . The PCR protocol was: 30 s at 98 °C, 12 cycles of 7 s at 98 °C and 50 s at 72 °C, followed by a 2 min step at 72 °C. The reaction was cleaned using the Monarch PCR cleanup oligo kit (NEB) and eluted in 6–10 μL of nuclease-free water. The transcription buffer consisted of 40 mM Tris-HCl pH 8, 20 mM MgCl₂, 10 mM DTT and 2 mM spermidine. Template (approximate final concentration of 0.5 μM) and 3.75 mM of each NTP, 5 U/mL or 2 U/400 μL of *E. coli* Inorganic Pyrophosphatase (100 U/mL stock) and 1500 U of T7 polymerase were added. The transcription was incubated for 3–4 h at 37 °C, and subsequently cleaned using the Monarch RNA cleanup kit (NEB). The resulting material was gel purified using a 2 mm 15% PAGE gel (30 mL) with 10% stacking gel (10 mL), (6 well comb, 25 W constant power for 2 h). The gel was wrapped in plastic wrap, placed on TLC plate, and illuminated under UV (254 nm). The ribozyme band was marked, excised, placed in a 2 mL tube and weighed. The gel slice was dry crushed with a syringe plunger, 0.3 M sodium acetate pH 4.8–5.2 was added (2× mg gel weight in μL) and soaked overnight at 4 °C on a rotator. The next day, the supernatant was recovered using Spin-X centrifuge tube filters (Corning, 0.45 μm, sterile) and centrifuging for 5 min at maximum speed. Subsequently, 20 ng glycogen and either 1 volume isopropanol or 2.5 volumes of ethanol were added, and the sample was frozen for 45 min to 1 h (–80 °C for ethanol, –20 °C for isopropanol) and centrifuged for 45–90 min at –9 °C, maximum speed. The supernatant was removed, and the pellet washed with 80% ethanol and centrifuged again for 5 mins at maximum speed. This step was repeated twice before drying the pellet for 10–15 min. Finally, it was dissolved in deionized H₂O and concentrations were determined by measuring the absorption at 260 nm using a Nanodrop (Thermo). Extinction coefficients for sequences were calculated using OligoCalc³¹.

Ligation of M123

RNA ligation reactions were carried out with final concentrations of 20 μM for all ligation substrates, 5 μM of template, 30 mM Tris-HCl pH 7.5, 100 mM KCl, and varying bulk magnesium concentrations (as specified). Ligation was initiated by adding *sunY* ribozyme to a final concentration of 2.5 μM. The first substrate (M1) contained a Cy5 label at the 5'-end. A reaction volume of 12 μL was prepared for the isothermal control and 17 μL were prepared for the experiments under AWI conditions. As soon as *sunY* was added, 2 μL of each sample was quenched with 2× gel loading buffer (90% formamide, 5% glycerol, 50 mM EDTA, 0.015% Orange G). For specifics regarding the PAGE protocol, please refer to subsection 'Polyacrylamide gel electrophoresis (PAGE)'. The remaining volume for the isothermal controls was placed in the T100 thermal cycler (Bio-Rad) set to 45 °C. The reaction volume of 15 μL for the AWI experiments was injected into the fully assembled AWI-system through a Teflon tubing (Techlab) using a Hamilton syringe. The reaction time was 4 h. Please refer to the subsection 'Preparation of AWI chambers and sample injection' for details of the AWI assembly and filling process.

Synthesis of *sunY* fragments A123, B123, and C123

The synthesis reactions took place for 2 h at a final concentration of 10 μM of each ligation substrate (including the 5' Cy5 labeled 5'-substrate) and 2.5 μM of template, respectively. The reaction buffer was 30 mM Tris-HCl pH 7.5, 100 mM KCl, and 10 mM MgCl₂ and all

reactions were initialized by the addition of *sunY* to a final concentration of 2.5 μM . Setup of reactions under AWI/isothermal conditions was performed as described above.

Complete replication of *sunY* sense and antisense fragments

For a full replication cycle of a *sunY* fragment, ligation substrates for both the sense and antisense were added. Syntheses of A123 and tA123 were carried out with final concentrations of 10 μM for each substrate (A1, A2, A3) and the complementary substrates (tA1, tA2, tA3). Reactions were incubated for 3 h with and without 2.5 μM of template. The first substrate of the A123 and C123 ligation reactions (A1 / C1) were 5'Cy5 labeled, and the 5'-terminal substrates of the tA123 and tC123 ligation reactions (tA1 / tC1) were 5'FAM labeled. For the synthesis of C123 and tC123 the concentrations had to be adjusted slightly to decrease the overall concentration of strands complementary to the full-length *sunY* ribozyme as this increased yields and reduced incubation times. Here, four concentration ratios were tested and are denoted in the following by "Template: complementary substrates: substrates (T:t:S)": 2.5 μM : 5 μM : 10 μM (Fig. 2b, c; Supplementary Fig. 8c); 1.5 μM : 2.5 μM : 10 μM ; 1.5 μM : 5 μM : 10 μM and 2.5 μM : 2.5 μM : 10 μM (Supplementary Fig. 8b). All replication reactions were performed in 30 mM Tris-HCl pH 7.5, 100 mM KCl and 10 mM MgCl₂. Substrate ligation was initialized by the addition of *sunY* to a final concentration of 2.5 μM . For the ligation of C123 and tC123, the experiments ran for 3 h or 18 h as indicated.

HH-min synthesis

HH-min synthesis reactions took place at a final concentration of either 10 μM and 5 μM or 12 μM and 3 μM of all ligation substrates and template, respectively. The first RNA ligation substrate (HH1) was either 5'FAM or 5'Cy5 labeled. The buffer in the AWI reactions and isothermal controls was 30 mM Tris-HCl pH 7.5, 100 mM KCl with varying bulk magnesium concentrations (c_{bulk} —as specified). Reactions were initialized by the addition of *sunY* to a final concentration of 2.5 μM . To verify the activity of the in situ synthesized HH-min hammerhead ribozyme, 0.4 μM HH-sub was added. The HH-sub contained 5'-terminal FAM tag and a 3' BHQ quencher. For a T_0 sample, 2 μL of the prepared reaction was quenched immediately with 18 μL gel loading buffer (94% formamide, 5% glycerol, 0.01% Orange G) after initialization of the reaction. All isothermal controls were incubated at 45 °C. To verify that the three unligated HH-min ligation substrates HH1-3 had no residual RNA-cleavage activity that was e.g. triggered by the accumulation process at the water-air interface, 2 h and 4 h control reactions in the absence *sunY* were performed (Supplementary Fig. 9).

Polyacrylamide gel electrophoresis (PAGE)

All samples were analyzed using polyacrylamide gel electrophoresis (PAGE) with 15% acrylamide. The gel stock was prepared from the Roth Rotiphorese DNA sequencing gel stocks (acrylamide:bisacrylamide = 19:1), where a 0.75 mm thick gel with a 15-tooth comb needed about 5 mL gel mixture, containing 3 mL gel concentrate, 1.5 mL gel diluent, 0.5 mL buffer concentrate, 25 μL APS and 2.5 μL TEMED. All gels were run in denaturing conditions of 50% urea and 1 \times TBE buffer at -50–55 °C. Each runtime consisted of a 30 min pre-run at 400 V, after which 4–5 μL were loaded per well. 1:9 dilutions of samples with gel loading buffer (94% formamide, 5% glycerol, 0.01% Orange G) were used for PAGE analyses of T_0 samples and samples from isothermal experiments and 3:7 dilutions were used for PAGE analyses of samples from AWI experiments. Electrophoresis was performed for 5 min at 50 V followed by 18–24 min at 300 V. After each run, gels were either directly imaged or stained in 50 mL of 1 \times TBE buffer with 5 μL of 10,000 \times SYBR Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific) for 5 min prior to imaging. Imaging was done with a two-channel protocol of Image Lab v6.0.1 from the Bio-Rad ChemiDoc MP System, using the blots for Cy5 and Alexa488. All gel images were imported

into an in-house written LabView routine (NI LabVIEW 2014 14.0.1f11 (64 bit)) for densitometric analysis. Recorded lane intensities were background-corrected using the average intensity of the inter-lane spaces. Each band was fitted with a Gaussian in Igor Pro 6.37. As the extraction from AWIs as well as quenching from imaging during the experiment can introduce artifacts, only relative intensities for each lane were used for quantification:

$$r = \frac{I_{\text{Product}}}{I_{\text{C}} + I_{\text{intermediate}} + I_{\text{Bg}} + I_{\text{Product}}} \quad (1)$$

Intensities from the uncleaved substrate HH-sub were multiplied by 1.69, which is the factor by which the in-gel fluorescence is quenched due to the presence of the Black Hole Quencher at the 3' end.

Preparation of AWI chambers and sample injection

AWI chambers were built similarly to the trap system reported in Matreux et al.²⁴. A schematic image of a chamber is shown in Supplementary Fig. 1. The desired shape was cut out from a 250 μm thick Teflon foil (HolscotEurope) with using a GRAPHTEC CE6000-40 Plus cutter. The resulting Teflon slide defined the shape of the non-equilibrium AWI compartments. The Teflon cut-out was designed to have three reaction compartments. The compartments are all 11 mm wide and 8 mm high. This corresponds to a maximum capacity of 22 μL for a 250 μm foil thickness. A second foil with 100 μm thickness was cut out to serve as hydrophobic backside. Both Teflon slides were sandwiched between two sapphire windows and fixed to a metal block using a steel frame. To increase heat conductivity between the cold sapphire and the metal surface, a 25 μm thick graphite (PANASONIC) foil was sandwiched in between both elements. All parts of the sandwich had a width of 22 mm and a height of 60 mm. Prior to assembly, all components were cleaned with ethanol and milliQ water and additionally cleaned from dust using scotch tape. A heater was screwed onto the top side. To achieve the best heat conductivity between the heater element and the surface of the AWI chamber, 200 μm thick graphite foils (PANASONIC) were placed between heater and the assembled AWI-system, as well as between the AWI-system and the water-cooled backside of the setup. The temperature gradient was maintained by cooling the backside with a refrigerated water bath (CF41 and 300 F, Julabo) and heating the front with an electrical heater regulated by an Arduino microcontroller board. For each experiment, a reaction volume of 15 μL was injected with a Hamilton syringe through a Teflon tubing (Techlab), which were fixed by PTFE fittings (Techlab). The inlets were then sealed with soft silicon and PTFE fittings. All compartments were connected to the same CO₂ inflow to ensure similar pressure conditions. Details of the resulting dynamics are described in Supplementary Note 3. Additionally, to monitor the temperature by the water bath and set the temperature of the heater, a temperature sensor was screwed into one of the holes in the backside with one of the fittings.

To create the non-equilibrium conditions in the AWI-system, two types of setups were used either with or without the possibility of optical readout. The air phase of each AWI-system was connected to a CO₂ supply and the pressure was adjusted by a dedicated pressure controller (RIEGLER) directly at the setup. Additionally, a manometer (Bourdon Instruments, MEX 3) was attached to the CO₂ supply to cross-verify the applied pressure. Details of the influence of the CO₂ atmosphere on the pH of the bulk reaction are described in Supplementary Methods 3.1. Temperatures were monitored throughout the experiment using a GTH 1170 thermometer (Greisinger) at the back sapphire. Temperatures were further monitored at the beginning and end of each experiment with a heat camera (Seek Thermal, SQ-AAA) on the front sapphire to monitor the applied gradient. The measured temperatures were used as input parameters for a finite-element simulation to simulate the thermal gradient inside the chamber.

Details of the temperature simulation are provided in Supplementary Fig. 4 and Supplementary Methods 3.2. For fluorescence excitation, the THORLAB LEDs M470L2-C4 (470 nm emission) and M625L2 (625 nm emission) were used for the dyes FAM and Cy5, respectively. Imaging was performed with the Axio Scope.A1 microscope (Zeiss) with infinity corrected long working distance objective (2× lens Mitutoyo, M Plan Apo) and a 0.5 adapter that was coupled to a stingray camera F145B ASG (Allied). Temperature and microscope control were realized with code written in NI LabVIEW 2014 14.0.1f11 (64 bit).

Statistics and reproducibility

The gel presented in Fig. 1c is a representative from $n = 4$, $n = 4$, and $n = 2$ independent replicates in the AWI-system for MgCl_2 concentrations of 10 mM, 5 mM, and 1 mM, respectively, and $n = 2$ and $n = 5$ for MgCl_2 concentrations of 10 mM and 50 mM in isothermal conditions. All corresponding gels are provided in the Source Data file. In Fig. 2a the shown gel is representative for $n = 2$ (A, B fragment) and $n = 3$ (C fragment) independent experiments for reactions in the AWI-system. The remaining gels are shown in Supplementary Fig. 7 and are provided in the Source Data file. The gels shown in Fig. 2b, c represent a single concentration ratio ($n = 2$ independent experiments), which was selected based on several pre-tested concentration ratios (each $n = 1$). The remaining concentration ratios are shown in Supplementary Fig. 8. All gels are provided in the Source Data file. In Fig. 3b the gel lanes were reordered for clarity as indicated by the dotted line. The unmodified gel is shown in Supplementary Fig. 13B. The isothermal control for this experiment was run on a different gel. The 2 h and 4 h lane (+sunY) of the AWI-system are representative gels from $n = 10$ and $n = 4$ independent experiments, respectively. These replicates are the source data for Fig. 3c and are provided in the Source Data file. Details on the statistics and reproducibility of the data shown in the Supplementary figures are provided in Supplementary Methods 3.3.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data generated in this study are provided in the paper and Supplementary Information file. Figs. 1–3 and Supplementary Figs. 3–5, 7–14 contain associated raw data provided in the Source Data file and deposited at <https://doi.org/10.5282/ubm/data.362>. Source data are provided with this paper.

Code availability

Codes are provided at <https://doi.org/10.5282/ubm/data.362>.

References

- Gilbert, W. Origin of life: the RNA world. *Nature* **319**, 618 (1986).
- Robertson, M. P. & Joyce, G. F. The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, 1 (2012).
- Joyce, G. F. & Szostak, J. W. Protocells and RNA Self-Replication. *Cold Spring Harb. Perspect. Biol.* **10**, a034801 (2018).
- Doudna, J. A., Couture, S. & Szostak, J. W. A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science* **251**, 1605–1608 (1991).
- Doudna, J. A., Usman, N. & Szostak, J. W. Ribozyme-catalyzed primer extension by trinucleotides: a model for the RNA-catalyzed replication of RNA. *Biochemistry* **32**, 2111–2115 (1993).
- Attwater, J., Raguram, A., Morgunov, A. S., Gianni, E. & Holliger, P. Ribozyme-catalysed RNA synthesis using triplet building blocks. *eLife* **7**, e35255 (2018).
- Tjhung, K. F., Shokhirev, M. N., Horning, D. P. & Joyce, G. F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci. USA* **117**, 2906–2913 (2020).
- Cojocaru, R. & Unrau, P. J. Processive RNA polymerization and promoter recognition in an RNA world. *Science* **371**, 1225–1232 (2021).
- Le Vay, K., Salibi, E., Song, E. Y. & Mutschler, H. Nucleic acid catalysis under potential prebiotic conditions. *Chem. - Asian J.* **15**, 214–230 (2020).
- Li, Y. & Breaker, R. R. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2′-hydroxyl group. *J. Am. Chem. Soc.* **121**, 5364–5372 (1999).
- Szostak, J. W. The eightfold path to non-enzymatic RNA replication. *J. Syst. Chem.* **3**, 1–14 (2012).
- Horning, D. P. & Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. USA* **113**, 9786–9791 (2016).
- Salditt, A. et al. Thermal habitat for RNA amplification and accumulation. *Phys. Rev. Lett.* **125**, 048104 (2020).
- Zhou, L. et al. Non-enzymatic primer extension with strand displacement. *eLife* **8**, e51888 (2019).
- Kim, S. C., O’Flaherty, D. K., Giurgiu, C., Zhou, L. & Szostak, J. W. The emergence of RNA from the heterogeneous products of prebiotic nucleotide synthesis. *J. Am. Chem. Soc.* **143**, 3267–3279 (2021).
- Gavette, J. V., Stoop, M., Hud, N. V. & Krishnamurthy, R. RNA–DNA chimeras in the context of an RNA world transition to an RNA/DNA world. *Angew. Chem.* **55**, 13204–13209 (2016).
- 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).
- Kristoffersen, E. L., Burman, M., Noy, A. & Holliger, P. Rolling circle RNA synthesis catalysed by RNA. *eLife* **11**, e75186 (2022).
- Le Vay, K. & Mutschler, H. The difficult case of an RNA-only origin of life. *Emerg. Top. Life Sci.* **3**, 469–475 (2019).
- laneselli, A. et al. Water cycles in a Hadean CO₂ atmosphere drive the evolution of long DNA. *Nat. Phys.* **18**, 579–585 (2022).
- laneselli, A., Mast, C. B. & Braun, D. Periodic melting of oligonucleotides by oscillating salt concentrations triggered by microscale water cycles inside heated rock pores. *Angew. Chem.* **58**, 13155–13160 (2019).
- Mariani, A., Bonfio, C., Johnson, C. M. & Sutherland, J. D. PH-driven RNA strand separation under prebiotically plausible conditions. *Biochemistry* **57**, 6382–6386 (2018).
- Morasch, M. et al. Heated gas bubbles enrich, crystallize, dry, phosphorylate, and encapsulate prebiotic molecules. *Nat. Chem.* **11**, 779–788 (2019).
- Matreux, T. et al. Heat flows in rock cracks naturally optimize salt compositions for ribozymes. *Nat. Chem.* **13**, 1038–1045 (2021).
- Doudna, J. A. & Szostak, J. W. Miniribozymes, small derivatives of the sunY intron, are catalytically active. *Mol. Cell. Biol.* **9**, 5480–5483 (1989).
- Green, R. & Szostak, J. W. Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science* **258**, 1910–1915 (1992).
- Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R. & Bruening, G. Autolytic processing of dimeric plant virus satellite RNA. *Science* **231**, 1577–1580 (1986).
- Drobot, B. et al. Compartmentalised RNA catalysis in membrane-free coacervate protocells. *Nat. Commun.* **9**, 3643 (2018).
- Dass, A. V. et al. RNA Oligomerisation without Added Catalyst from 2′,3′-Cyclic Nucleotides by Drying at Air–Water Interfaces. *Chem-SystemsChem* **5**, e202200026 (2023).
- laneselli, A. et al. Non-equilibrium conditions inside rock pores drive fission, maintenance and selection of coacervate protocells. *Nat. Chem.* **14**, 32–39 (2022).
- Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**, W43–W46 (2007).
- Su, Z. et al. Cryo-EM structures of full-length Tetrahymena ribozyme at 3.1 Å resolution. *Nature* **596**, 603–607 (2021).

Acknowledgements

The authors thank Christof B. Mast for help with the setup and for providing his expertise on the non-equilibrium compartments and Alan Ianeselli for help with the CO₂ installation. We also thank Borislav Mladenov for their experimental support. Financial support was provided by the European Research Council (ERC Evotrap, grant no. 787356, D.B. and RiboLife, grant agreement no. 802000, H.M.), the Simons Foundation (grant no. 327125, D.B.), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant TRR 235 (CRC 235) (Project-ID 364653263, A.S., D.B.), the Excellence Cluster ORIGINS (D.B.) funded by Germany's Excellence Strategy EXC-2094-390783311, and the Center for NanoScience (A.S., D.B.). H.M., D.B., and K.L.V. acknowledge support from the Volkswagen Foundation. AS is supported by the Add-on Fellowship of the Joachim Herz Foundation.

Author contributions

Project design, funding acquisition, and supervision: D.B., H.M.; experiments and data analysis: A.S., L.K., E.S., K.L.V.; visualization: A.S., E.S.; writing: A.S., L.K., E.S., K.L.V., D.B., H.M.

Funding

Open Access funding enabled and organized by Projekt DEAL.

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/s41467-023-37206-4>.

Correspondence and requests for materials should be addressed to Dieter Braun or Hannes Mutschler.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) 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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

9 Acknowledgements

First of all, I would like to thank Dieter. From the beginning, you encouraged and supported me. You provided so many opportunities for me to explore and discuss the origin of life. I am equally thankful for the way you handled the 'Annika-situation'. Many people talk about family compatibility; you really made it possible.

Christof, without you, the Braunlab would not be the same. The tale of the Christof-bubble really is true - if you are close, stuff magically works, if you are not, everything crumbles into pieces. Thank you for the countless times you rescued whatever decided to fail.

I would like to thank all the great people I got to collaborate with, especially Elia, Kris, Hannes, Julio, and Uli. I thank Patta and Alan for the many hours we spent writing and discussing the first version of the review that never got published. This helped tremendously when writing the introduction. I owe special gratitude to my student assistants, Patrick, Bobby, Emma, and Katharina. It was great to work with you.

Thanks to all 'Braunies' throughout the years - you are a special group of people and I will miss spending my time in the cellar. Patta, the morning coffees (even the digital ones) are very special to me. I have learned so much during our discussions and laughed so much during our breaks. Alex, no matter how long we couldn't talk to each other, as soon as we see each other again it feels like yesterday. I will always cherish our time in the US. Zhenya and Matze, I miss the old Braunie times. Leo, I could not have asked for a better master student. We lived through the ups and downs together, with coffee, talks, and loooooots of experiments. I am so grateful that I gained not only a great colleague, but an amazing friend. Max, you are the person one can always turn to when having a problem. And still, you are so modest about it. Thomas, we started together in the first semester and here we are. You are a powerhouse. Saroj, it is inspiring to hear you talk about your research. I will keep on steeling tea from you. Thanks to everyone in the Mensa crew. Sree, you are one of the most resourceful and creative people I know! Sometimes, several ideas already popped out while I was not even done finishing the question. Juliette, you always listened and reassured me when I was overwhelmed with being pregnant or a mom at work. We need to keep on forcing each other to go to yoga! Adriana, whether talking about politics, science or random stuff, you always say something I didn't think about before. I have the best memories of dancing to Portuguese songs. Philipp, you always had the right soundtrack for any situation, even if it was *für elise*. Thanks for teaching my daughter how to stick out her tongue. Alex, I hope you go to space one day. Paula, I will always cherish the little Moomin socks. Felix, your laughter is contagious. I cannot believe how many 'Braunie-babies' are now born. Chrissy, it is amazing to see the joy Ludwig brings to you. I am sure we keep on sharing this joy. 'Servus' Noel, your passion for your projects is contagious. I am certain this will transition into fatherhood, and I wish little Laia the most wonderful beginning. Lara, I wish you and little Lilith all the best up in the north.

Thanks to my friends and family, you are always there for me. Mareike, thank you for being the most wonderful friend. I am so happy that we both live in the same city once again. Franzi, you and your dance lessons kept me sane during the most stressful times! Thanks to my grandparents for the many invitations and the best food. Whenever I need help, you are there! Thanks to my sisters. Marie, sharing the life of being a PhD student with you made it so much easier. Whenever I was in doubt, our conversations helped to see things clearer. Clara, Munich feels even more like home since you are here. Thanks for being such a great aunt and for babysitting Annika during Zoom calls, meetings, and writing sessions. Lotti, I cannot believe that when I started studying you were still in 3rd grade and now you started studying. I am excited to get to know Frankfurt now. Thanks to my parents, you made me who I am. I am beyond grateful for your everlasting support and love.

Christoph, words can not describe how I feel about you. We share everything in life - I love you!

Annika, you are the best thing that ever happened to me.